

Alveolar epithelial progenitor cells require Nkx2-1 to maintain progenitor-specific epigenomic state during lung homeostasis and regeneration

Received: 22 September 2022

Accepted: 4 December 2023

Published online: 19 December 2023

 Check for updates

Andrea Toth^{1,2,3,4,5}, Paranthaman Kannan^{1,2,3}, John Snowball^{1,2,3},
Matthew Kofron^{3,6,7}, Joseph A. Wayman^{8,9}, James P. Bridges^{10,11},
Emily R. Miraldi^{7,8,9}, Daniel Swarr^{1,2,7} & William J. Zacharias^{1,2,3,7,12,13} ✉

Lung epithelial regeneration after acute injury requires coordination cellular coordination to pattern the morphologically complex alveolar gas exchange surface. During adult lung regeneration, Wnt-responsive alveolar epithelial progenitor (AEP) cells, a subset of alveolar type 2 (AT2) cells, proliferate and transition to alveolar type 1 (AT1) cells. Here, we report a refined primary murine alveolar organoid, which recapitulates critical aspects of in vivo regeneration. Paired scRNAseq and scATACseq followed by transcriptional regulatory network (TRN) analysis identified two AT1 transition states driven by distinct regulatory networks controlled in part by differential activity of Nkx2-1. Genetic ablation of Nkx2-1 in AEP-derived organoids was sufficient to cause transition to a proliferative stressed Krt8⁺ state, and AEP-specific deletion of Nkx2-1 in adult mice led to rapid loss of progenitor state and uncontrolled growth of Krt8⁺ cells. Together, these data implicate dynamic epigenetic maintenance via Nkx2-1 as central to the control of facultative progenitor activity in AEPs.

The pulmonary alveolar gas exchange surface is frequently challenged by pathogens, environmental toxins, and inhaled irritants. These challenges cause chronic, recurrent stress and/or injury to the epithelial, endothelial, and mesenchymal lineages that constitute the alveolus¹. Homeostatic turnover and regenerative capacity in the

alveolus must be sufficient to maintain adequate oxygenation and ventilation throughout life^{2,3}. Therefore, extensive recent attention has focused on epithelial progenitor capacity and plasticity in the repair of the alveolar epithelium^{1,4}—a topic whose importance has been further emphasized by the COVID-19 pandemic^{5,6}. Given the dearth of

¹Perinatal Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ²Division of Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ³Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ⁴Medical Scientist Training Program, University of Cincinnati College of Medicine, Cincinnati, OH, USA. ⁵Molecular and Developmental Biology Graduate Program, University of Cincinnati College of Medicine, Cincinnati, OH, USA. ⁶Bio-Imaging and Analysis Facility, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ⁷Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. ⁸Division of Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ⁹Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ¹⁰Department of Medicine, Division of Pulmonary Sciences and Critical Care Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA. ¹¹Department of Medicine, Division of Pulmonary and Critical Care Medicine, National Jewish Health, Denver, Colorado, USA. ¹²Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, OH, USA. ¹³Center for Stem Cell and Organoid Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ✉e-mail: William.Zacharias@cchmc.org

therapies to promote alveolar epithelial regeneration, models to define epithelial cell relationships and identify pro-regenerative pathways are an area of high-priority research interest.

Organoids provide one promising avenue for modeling regenerative biology using adult cells. Multiple lung organoid approaches have been reported in recent years, derived both from primary lung epithelium⁷ and induced-pluripotent stem cells (iPSC)^{8–10}. Several themes emerge from these reports. First, the term “lung organoid” encompasses a broad, heterogeneous set of cultures with different compositions and morphologies. Second, advancements in the purity of epithelial components and/or removal of mesenchymal supportive cells have been reported^{11,12}, generally at the expense of complexity. Co-culture systems are characteristically higher in cellular heterogeneity, which is an advantage in replicating the complex cellular composition of the alveolus, but the reproducibility of these co-cultures has been challenged¹¹. Third, while iPSC-derived alveolar cells have advanced understanding of human alveolar type 2 (AT2) differentiation and biology^{9,13}, it is difficult to model complex adult lung epithelial phenotypes and pathologies using human iPSC cultures due to differences between immature and mature lung epithelium. Finally, lung regeneration involves complex *in vivo* morphogenesis occurring in tandem with cellular differentiation¹. A major barrier to building an “alveolus in a dish” is the lack of morphological similarity between *in vitro* and *in vivo* models. These challenges have limited the utility of organoid cultures as a method for studying alveolar regeneration *in vitro*.

To address these challenges, we refined and standardized the conditions and inputs of co-culture of murine AT2 cells and alveolar fibroblasts^{2,3,14,15}. Recent data demonstrate that Wnt-responsive AT2 cells, also called alveolar epithelial progenitors (AEPs), harbor extensive progenitor capacity^{3,16,17}. Following injury, AEPs expand rapidly, differentiate into new AT1 and AT2 cells, and repair regions of alveolar injury following epithelial loss or infectious stress. Herein, we describe the high-dimensional characterization of these organoids across multiple time points. We find that AEP-derived organoids, or AEP-O, develop via clonal expansion of single progenitor cells and undergo progressive cellular differentiation and spontaneous cavity formation *in vitro*. Using multistage single-cell transcriptomics and epigenomics, we defined the organoid cellular milieu and identified separable AEP, AT2, AT1, and transitional states^{18–21} in organoids. We validated these states based on published *in vivo* scRNAseq and derived cellular trajectories and lineage relationships from a known progenitor root state. Comparative transcriptional regulatory network (TRN) analysis^{22,23} along these trajectories identified several known and unknown regulators of alveolar epithelial biology and highlighted a role for the lineage transcription factor Nkx2-1^{24–26} in progenitor and transitional lineages of the adult alveolus. Genetic ablation of Nkx2-1 in AEPs *in vitro* and *in vivo* caused irreversible acquisition of a stressed transitional state, with uncontrolled proliferative growth and disruption of organoid morphology *in vitro* and alveolar structure *in vivo*. These findings highlight the utility of the AEP-O assay as a high-fidelity model of *in vivo* alveolar biology and implicate Nkx2-1 as a central regulator of alveolar epithelial progenitors.

Results

AEP-O recapitulate critical structural aspects of alveolar regeneration

We generated AEP-O by combining FACS-sorted AEPs derived from Axin2^{CreERT2-tdT} mice (Supplementary Fig. S1) at a 1:10 ratio with lung mesenchymal cells isolated by selective adhesion in a 1:1 mix of Matrigel (Corning) and small airway growth media (SAGM, Lonza). To evaluate the internal growth and maturation of these organoids, we adapted methods from iPSC culture for whole-mount immunohistochemistry of organoids²⁷ and performed a time series evaluation

using high content confocal imaging paired with longitudinal single-cell sequencing (Fig. 1A).

AEP-O grow clonally from a single AEP in definable stages. First, AEPs expand into small clusters of SFTPC⁺ cells during the first week of culture (Fig. 1C). By day 14, differentiation of RAGE⁺ AT1 cells was observed within the central portion of the organoids (Fig. 1D), consistent with prior reports^{2,3,14,15,18,20,28}. During the third week of culture, these developing AT1 cells elongated and polarized (Fig. 1E), and by d28 *in culture*, cavities were present within the organoids (Fig. 1B, F). These cavities matured into a network of alveolar-like structures during the 4th and 5th weeks of differentiation (Fig. 1F, G). AT1 cells intermixed with AT2 cells within the central portion of the organoid (Fig. 1H, M), with minimal apoptosis detectable by TUNEL staining in AEP-O during cavity formation (Supplementary Fig. S2A–D). Continued proliferation was evident by EdU incorporation at the periphery of mature organoids (Fig. 1I). Electron microscopy demonstrated that the epithelial lining of these cavities includes AT2 cells containing lamellar bodies with the apical surface directed toward the internal lumen and evidence of active surfactant secretion adjacent to elongated AT1 cells (Fig. 1J, K). These features of the mature cavities within AEP-O bear a striking similarity to the epithelial structure of mature murine alveoli (Fig. 1L, M), suggesting an unprecedented degree of morphological maturation in AEP-O compared to other reported lung organoids.

Cavity maturation requires mesenchymal signaling not mechanical function

To better characterize the progressive cellular maturation occurring during paired cell differentiation and cavity formation in AEP-O, we performed single-cell RNA sequencing at 14, 21, and 28 days after culture initiation. We identified epithelial and mesenchymal populations, as well as an unexpected immune fraction (Fig. 2A–C)²⁹. Epithelial contribution increases from d14 to d28, with progressive maturation and an increase in the relative proportion of alveolar type 1 cells (Fig. 2B, Supplementary Fig. S3A–D). All epithelial cell types were present at each time point, with more proliferative cells at early time points and more differentiated cells at later time points. To evaluate the determinants of epithelial maturation in AEP-O, we examined the supportive cells comprising the signaling niche in these organoids.

Within the mesenchyme, two major PDGFR α ⁺ populations^{2,14,30–32} were identified, corresponding to murine alveolar fibroblasts and myofibroblasts in LungMAP datasets³⁰. To localize these mesenchymal cells in complex culture, PDGFR α ^{EGFP} lung fibroblasts³³ were obtained via selective adhesion (Supplementary Fig. S4A, B). Use of these PDGFR α ^{EGFP} fibroblasts did not affect organoid growth (Supplementary Fig. S4C, D). PDGFR α ^{EGFP} cells localize predominantly in two locations, with the minority of cells surrounding the epithelial organoids suspended in Matrigel and the majority growing on the Transwell filter in a monolayer (Fig. 2F–H, Supplementary Fig. S4E, F). Few PDGFR α ^{EGFP} cells were detected within organoids (Fig. 2F, G) and no clear deposition of fibrillar collagen was seen within organoids (Supplementary Fig. S4H), suggesting that the morphological maturation and complex structural organization of AEP-O did not require direct mesenchymal cell localization within the organoid itself. CD45⁺ immune cells were scattered throughout the fibroblast stocks and within the Matrigel but never found in large clusters or within organoids (Supplementary Fig. S4I); we concluded that these represented a minor contaminant of the supportive fibroblasts. Therefore, we hypothesized that the mesenchymal cells were predominantly involved in providing a signaling niche in AEP-O.

Ligand-receptor analysis³⁴ suggested extensive signaling between epithelial and mesenchymal cells within AEP-O (Fig. 2D, E). Major signal producers in AEP-O included AT1 cells, alveolar fibroblasts and myofibroblasts. AT1 cells expressed WNT ligands with predicted receptivity in both WNT-responsive AT2 cells and multiple mesenchymal

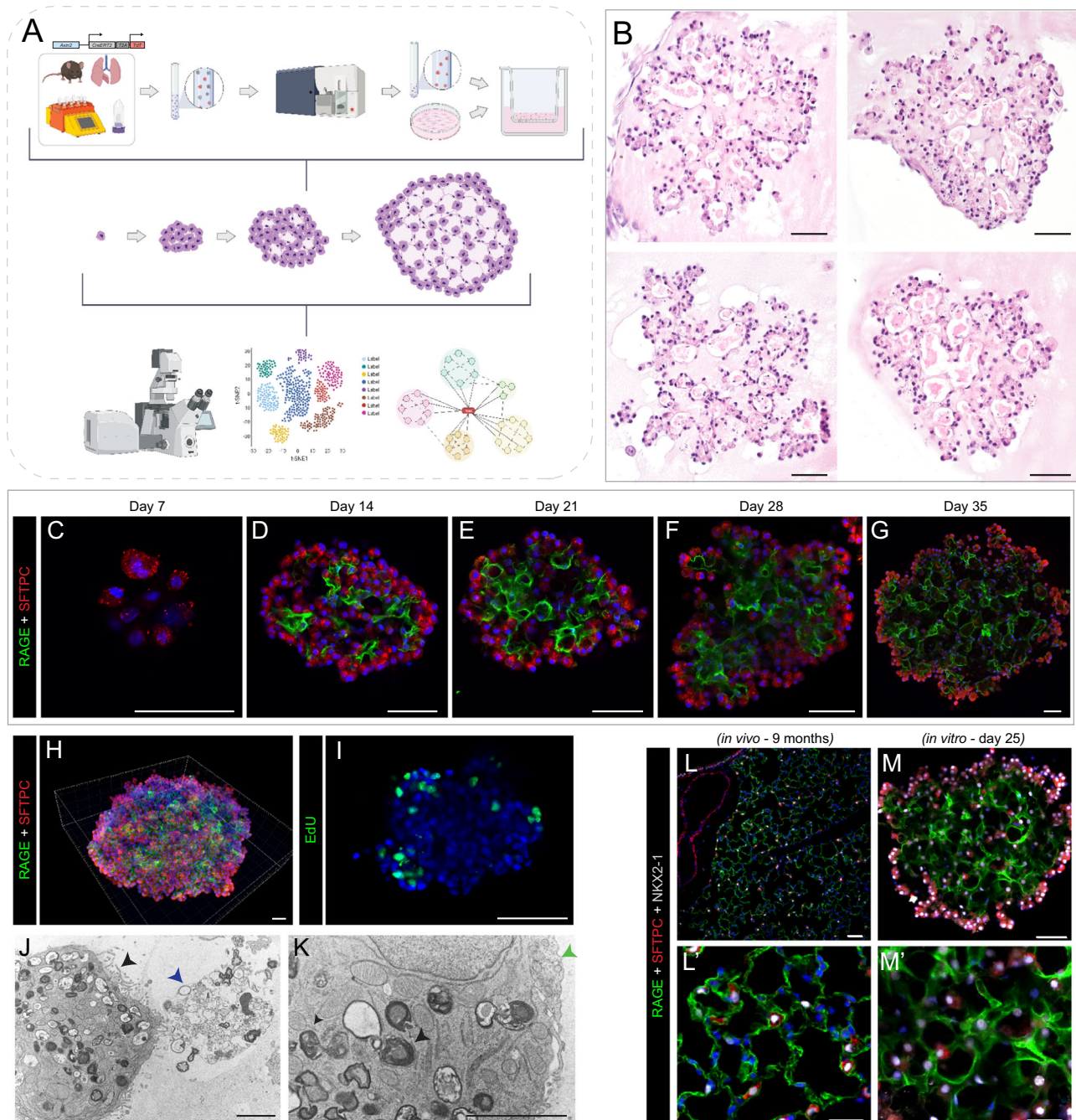
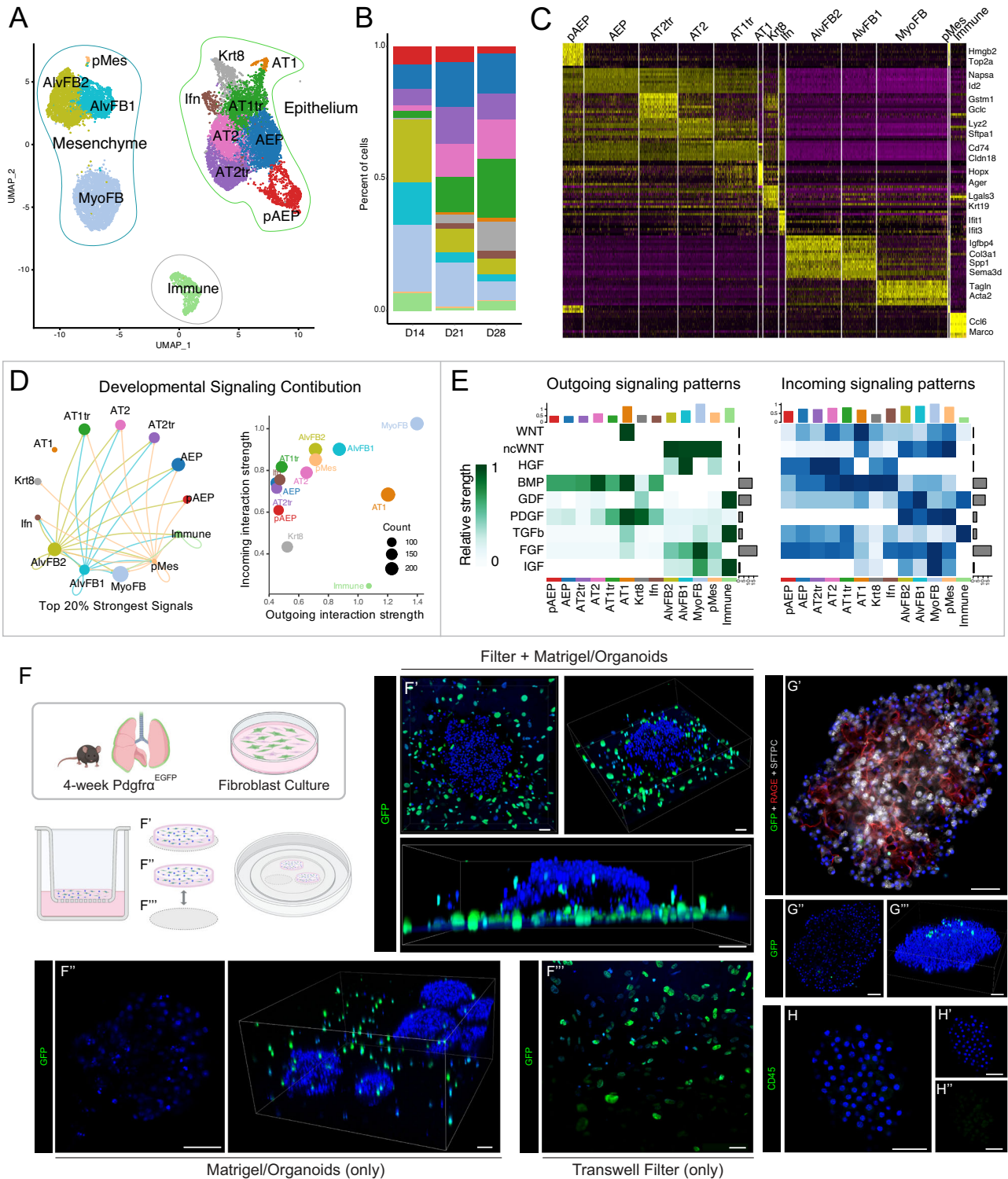


Fig. 1 | AEP-derived alveolar organoids clonally expand and pattern complex, polarized alveolar-like cavities. **A** Schematic of experimental design and overview. Live/CD31⁺/CD45⁺/CD326⁺(EpCAM⁺)/TdtTomato⁺(Axin2⁺) cells (AEPs) were mixed with mouse lung fibroblasts from P28 mice and cultured for up to 35 days, followed by analysis via high content imaging. **B** H&E of 5 μ m sections of FFPE day 35 Axin2⁺ organoids, showing cellular morphologies typical of both AT1 and AT2 cells. **C–G** Whole-mount immunofluorescence time course of Axin2⁺ organoids showing expansion of SFTPC⁺ AT2 cells (red), increased differentiation into RAGE⁺ AT1 cells (green) and increased structural complexity. **H** Imaris 3D reconstruction of day 35 Axin2⁺ organoid (z-depth = 174.13 μ m) showing cellular arrangement/organization within mature organoids. **I** Click-iT EdU (green) whole-mount day 25 Axin2⁺ organoids, with proliferating cells primarily on outer edges or ‘buds’

growing outward from the organoid. **J, K** Electron microscopy of day 28 organoids. **J** Image of properly polarized AT2 cell with apical microvilli (black arrowhead) secreting surfactant (blue arrowhead) into a lumen. **K** Image of AT2 cell with lamellar bodies (black arrowhead) adjacent to an AT1 cell (green arrowhead, right). **L, M** Comparison of in vivo mouse lung (9-month C57BL/6J mouse) and in vitro day 25 Axin2⁺ organoids. Data throughout the figure represents at least 5 biological replicates with 3 technical replicates per experiment. [Scale bars = 50 μ m, except for electron microscopy (**J, K**) scale bars = 2.5 μ m]. (RAGE Receptor for Advanced Glycation End-products [AT1 cell marker], SFTPC Surfactant Protein C [AT2 cell marker], EdU 5-ethynyl-2'-deoxyuridine, FFPE formalin-fixed, paraffin-embedded). Schematics created with Biorender.com.

populations. AT1 also produced PDGF ligands predicted to signal to the PDGFR α ⁺ mesenchyme. Mesenchymal cells expressed HGF, non-canonical WNT, and FGF ligands, consistent with published data describing the roles of these pathways in alveolar regeneration^{14,25,35–40}.

Together, these data suggested that the AEP-O signaling milieu resembles key aspects of the in vivo regenerative niche and that the mesenchymal cells provided a supportive paracrine signaling niche required for alveolar cavity formation. To directly evaluate the



requirement for mesenchymal signaling support, we cultured PDGFR α ^{EGFP} cells on the basal side of the Transwell filter (Supplementary Fig. S4G). The absence of fibroblasts in the Matrigel plug led to a complete loss of organoid formation (Supplementary Fig. S4G), confirming a requirement for paracrine mesenchymal signaling in the establishment and maturation of AEP-O.

scRNAseq defines multiple epithelial maturation trajectories of AEPs

Integrated analysis of data from d14, 21, and 28 identified eight separable epithelial cell states via graph-based clustering in Seurat²⁹

(Fig. 2A–C, Supplementary Fig. S3A–D). Consistent with our whole-mount IHC results, both AT1 and AT2 cells were identifiable within organoids. We noted a clear AEP state defined by expression of the AEP-enriched markers *Id2*, *Ctnnb1*, *Lrp5*, *Lrp2*, *Napsa*, *Bex2*, *Hdc*, and *Fgfr2*^{3,16}. Cells bearing the AEP signature were also found in a second state partially defined by high-level expression of cell cycle genes. We called these cells AEPs and proliferative AEPs (pAEPs), respectively (Fig. 2A). The proliferative state is readily detectable in day 14 cultures and decreases by day 28 (Supplementary Fig. S3A–D); EdU staining of organoids confirmed a proliferative Sftpc⁺ population at day 25 (Fig. 1I).

Fig. 2 | Single-cell composition and epithelial-mesenchymal interactions in alveolar organoids over time course of differentiation. **A** UMAP of all cell populations combining d14, d21 and d28 AEP-derived organoid scRNAseq datasets. **B** Cell population proportions at each time point with increasing proportion of epithelial cells. **C** Heatmap showing expression of top 10 most differentially expressed genes in each population. **D, E** Ligand-receptor analysis of organoid culture demonstrating extensive mesenchymal-epithelial communication in organoids. **F** Schematic of experimental set-up of live imaging and 3D reconstruction of live day 20 organoids generated using PDGFR α ^{EGFP} fibroblasts stained with Hoechst, with data shown in (F, G). **F** 3D reconstruction of confocal z-stacks of whole wells including Transwell filter, showing the majority of GFP⁺ fibroblasts are growing on the filter; **F'**–**F'''** Whole-mount immunofluorescence showing lack of PDGFR α ⁺ cells within day 20 organoids, with scattered cells found throughout the surrounding

Matrigel. **G** Whole-mount IHC showing few GFP⁺ fibroblasts inside of organoids. **H** CD45 staining of organoids; see also Supplementary Fig. S3I. Data in (F–H) represents 2 biological replicates with 3 technical replicates per experiment. [Scale bars = 50 μ m]; (PDGFR α = platelet-derived growth factor receptor alpha, GFP = Green Fluorescent Protein; RAGE = Receptor for Advanced Glycation End-products [AT1 cell marker]; SFTPC = Surfactant Protein C [AT2 cell marker]); AEP = alveolar epithelial progenitor, pAEP = proliferative AEP, AT2tr = AT2 transitional cell, AT2 = alveolar type 2 cell, AT1tr = AT1 transitional cell, AT1 = alveolar type 1 cell, Krt8 = Krt8/DATP/PATS-like transitional cell, lfn = Interferon responsive alveolar cell, pMes = proliferative mesenchymal cell, AlvFB1 = alveolar fibroblast type1, AlvFB2 = alveolar fibroblast type 2, SM = smooth-muscle like mesenchyme. Schematics created with Biorender.com.

We then examined the initiating cells of these organoids in more detail. The starting epithelial fraction of AEP-O is comprised of sorted AEPs (as described in ref. 3 and shown in Supplementary Fig. S1). scRNAseq of AEPs immediately after sorting showed two distinct fractions of epithelial cells and a small fraction of contaminating mesenchyme (Supplementary Fig. S5A). The larger epithelial fraction corresponded to AT2 cells bearing the AEP signature and a second smaller population containing cells in a Krt8⁺ stressed transitional state²¹ (Supplementary Fig. S5A), suggesting that the process of digestion and sorting of Axin2⁺ AT2 cells causes cell stress⁴¹. This signature is consistent with published reports in other organ systems^{42–45}, with prior data specifically suggesting a clear impact on cellular oxygenation⁴⁶ and metabolism^{42,44} during tissue digestion for FACS. Evaluation of scATACseq analysis at d7 of AEP-O culture (Supplementary Fig. S5B) demonstrated that the relatively small fraction of epithelium at this time point is a single population most similar to pAEPs, suggesting that the AT2/AEP fraction of the sorted cells seen by scRNAseq following FACS is the likely initiating epithelial cell in AEP-O, though a contribution of Krt8⁺ cells which then revert to the AEP state cannot be excluded by our data.

To dissect pathways of differentiation from the AEP state, we compared lineage predictions generated by both trajectory analysis⁴⁷ and RNA velocity⁴⁸ (Fig. 3A, B). Differentiation of AEPs toward AT2 cells was detected through an AT2 transitional state (AT2tr). The AT2tr state is defined by high-level expression of glutathione pathway genes and shift toward lipid metabolism, while the AT2 state expressed high levels of mature AT2 markers, including *Sftpa1* and *Lys2* (Fig. 3C). Other AT2 markers, including *Sftpc* and *Sftpb*, were expressed broadly within multiple differentiating AEP states (Figs. 2C and 3F). These predictions correspond to the lineage trajectories implied by the progression of the organoid composition shown in the time course data underlying the combined set (Supplementary Fig. S3A–D).

We next examined AT1 differentiation in AEP-O. Recent attention has focused on differentiation of AT2 progenitor cells to AT1 cells, with multiple reports describing a distinct transitional state (variously called PATS¹⁸, DATP²⁰, ADI¹⁹, or Krt8⁺ cells²¹) marked by high-level expression of cell stress markers including *Krt8*, *Lgals3*, *Tp53*, *Nupr1*, *Ddit3*, and *Cldn4*; we will call this stressed cell state Krt8⁺ in this manuscript. Both trajectory analysis and RNA velocity suggested two trajectories from AEP to AT1 cells; one through this Krt8⁺ transitional state and another through a second, less apparently stressed state we denoted AT1 transition (AT1tr). RNA markers of the AT1tr state included *Hes1* and *Igfbp7*, and RNA velocity analysis showed decreasing AT2 gene expression and intermediate expression of AT1 markers within this population (Fig. 3D–F), supporting the transitional nature of this cell state. Label transfer and integration with published organoid datasets^{18,20} showed both AT1tr and Krt8⁺ cell states were detectable in published data (Supplementary Fig. S6A–C).

These findings suggested the model that AEPs could differentiate toward AT1 cells through either the Krt8⁺ state or the AT1tr state

(Fig. 3G) rather than through an obligate intermediate state. However, given the challenges of cell state predictions and transitions based solely on RNA transcriptome⁴⁹, we proceeded to evaluate the epigenomic state and chromatin topography of the cells comprising AEP-O via scATACseq.

Progenitor and transitional states regulators defined by TRN

We performed scATAC sequencing at d14, d21, and d28 in AEP-O and performed unbiased cell clustering of epigenomic states using ArchR⁵⁰. The same number of epithelial cell states were detected by scRNAseq and scATAC using default parameters, supporting the conclusion that clustering parameters were appropriate in both assays (Fig. 4A). Concordant with published bulk ATACseq data³, AEP and AT2 are distinct at the scATAC level (Fig. 4B); AT1tr and Krt8⁺ also showed differential chromatin accessibility, with distinct regions of open chromatin (Fig. 4C). Integrated analysis combining both scRNA and scATACseq identified separable clusters of regulated genes within AEPs, for AT2 differentiation, and for AT1 differentiation (Fig. 4D). scATACseq-based pseudotime inference confirmed the presence of two separate differentiation trajectories from AEPs to AT1 cells, one passing through the Krt8⁺ state and the second passing through the AT1tr state (Fig. 4E, F). Comparison of expression and chromatin accessibility using combining gene expression with ATAC suggested overlapping gene sets shared by Krt8⁺ and AT1tr cells but also substantive differences (Fig. 4D), implying independent regulatory inputs to these two states.

To better define the differences between Krt8⁺ and AT1tr states, we turned to transcriptional regulatory network (TRN) inference^{22,23,51}. Combination of RNA expression and chromatin accessibility can improve TRN predictions by reducing both false positive and false negative regulatory predictions;⁵² application to single-cell techniques have extended the power of these approaches to estimate TF regulators of individual cell states²³. Therefore, we performed TRN inference comparing regulatory networks of various cell states within AEP-O to identify differential TF activity and predict regulators of state transitions (Fig. 4G–K); our method combines DNA accessibility, predicted DNA binding motifs per TF, and expression of associated target genes. The combination of these factors creates a high threshold for the inclusion of a TF in the core factors controlling a particular cell state. Focusing on the core TFs within each TRN, we found that AEP regulators included *Nkx2-1* and *Tfcp2l1*⁵³, both enriched in expression in bulk RNAseq from AEPs³, as well as TFs modulating Wnt (*Tcf4*) and BMP (*Smad1/4*) activity concordant with known AT2 progenitor signaling response^{54–56} (Fig. 4H). Within AT1 cells, we identified *Nkx2-1*, *Gata6*⁵⁷, and *Foxa2*^{58,59}, all well-known modulators of AT1 gene expression, and signaling response from the AT1-associated Yap/Taz^{24,60,61} (*Tead1*), Notch (*Rbpj*)^{62,63}, and Tgfb^{19,28} (*Smad3/4*) signaling pathways (Fig. 4K). These results provided evidence that TRN inference had successfully identified known regulators of the AEP and AT1 cell states, supporting the notion that TRN inference could distinguish factors driving AT1 differentiation.

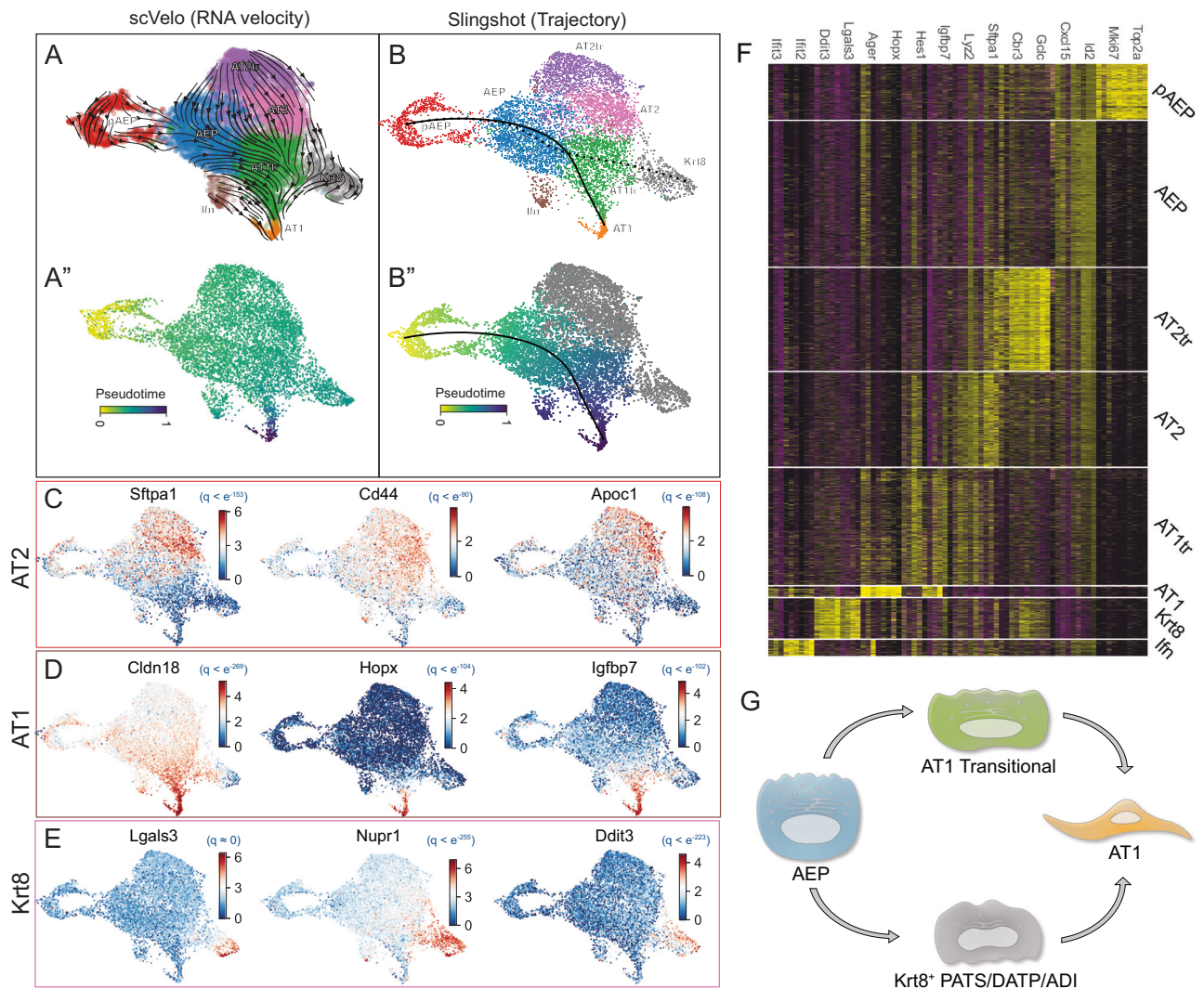


Fig. 3 | AEP-derived organoids elucidate dynamics of alveolar epithelial differentiation. **A** scVelo RNA velocity UMAP showing differentiation dynamics (A) and pseudotime inferred from RNA velocity (A'') in AEP organoids. **B** Slingshot trajectory analysis and pseudotime inference of AEP organoids demonstrate similar lineage relationships to RNA velocity. **C–E** Lineage drivers defined by CellRank for

differentiation of pAEP/AEP to AT2 cells (C), AT1 cells via AT1tr (D), and Krt8 cells (E). **F** Heatmap showing major cell markers differentiating cell states in the alveolar epithelium. **G** Model of cellular relationships and AT1 differentiation inferred from alveolar organoids.

We therefore focused our attention on differential regulators of transitional cells. We calculated transcription factor activity scores based on enrichment in predicted cell type. Both states showed overlap with AT1-enriched TFs. The TRN predictions were otherwise distinct, suggesting differential inputs to the two transitional states. The top regulators in Krt8⁺ cells were *Atf4* and its target *Ddit3*, the gene encoding C/EBP homologous protein (CHOP)⁶⁴ (Fig. 4J); together, these factors are activated by the multiple inputs of the integrated stress response (ISR)^{65,66}, with CHOP implicated in regulation of checkpoints in apoptosis vs cellular differentiation in other systems⁶⁴. The ISR is activated in the alveolar epithelium following ventilator-induced lung injury⁶⁷, and ISR activation contributes to lung fibrosis;⁶⁸ *Atf4* and CHOP activation may underlie Krt8⁺ cell accumulation in fibrosis¹⁸. The AT1tr TRN showed multiple distinct factors, which are shown for comparison (Fig. 4K and Supplementary Fig. S7); given the overlap of these factors with both AT2 and AT1 cells, specific regulators of the AT1tr state were not clearly identifiable in our dataset.

Loss of Nkx2-1 activity defines the Krt8⁺ transitional cell TRN

Notably, we identified a surprising absence of predicted Nkx2-1 activity in Krt8⁺ cells; TRNs from all other epithelial cell states in AEP-O

included Nkx2-1. Recent epigenomic profiling of the activity of Nkx2-1 during AT2 to AT1 transitions demonstrated Nkx2-1 occupancy at different genomic regions in each AT2 vs AT1 cells²⁴, suggesting a role for Nkx2-1 dis-engagement and re-engagement in the genome during differentiation. Nkx2-1 expression is lowest in stressed transitional epithelial cells at the time when Krt8 expression is highest during *in vivo* lung regeneration;²¹ similar expression dynamics are present in our organoids. Together, these observations supported the hypothesis that reduction in Nkx2-1 activity could directly promote the transition to the stressed Krt8⁺ transitional state.

To test this concept, we developed an approach to genetically manipulate AEPs during the formation of AEP-O (Fig. 5). Using an AAV6.2FF-Cre⁶⁹, which has recently been described as a high-fidelity reagent for genetic manipulation of AT2 cells *in vivo*⁷⁰, we infected AEPs harboring a R26R-lox-stop-lox-EYFP allele⁷¹ (from R26R^{EYFP} mice) immediately after FACS sorting (Fig. 5A). AAV6.2FF-Cre efficiently targeted AEPs which produced morphologically complex organoids (Fig. 5B) expressing the EYFP lineage label (Fig. 5C). Titration experiments indicated that infection at a multiplicity of infection (MOI) of 1000 was sufficient to induce significant recombination and label the majority of AEP-O (Fig. 5D); higher MOI mildly

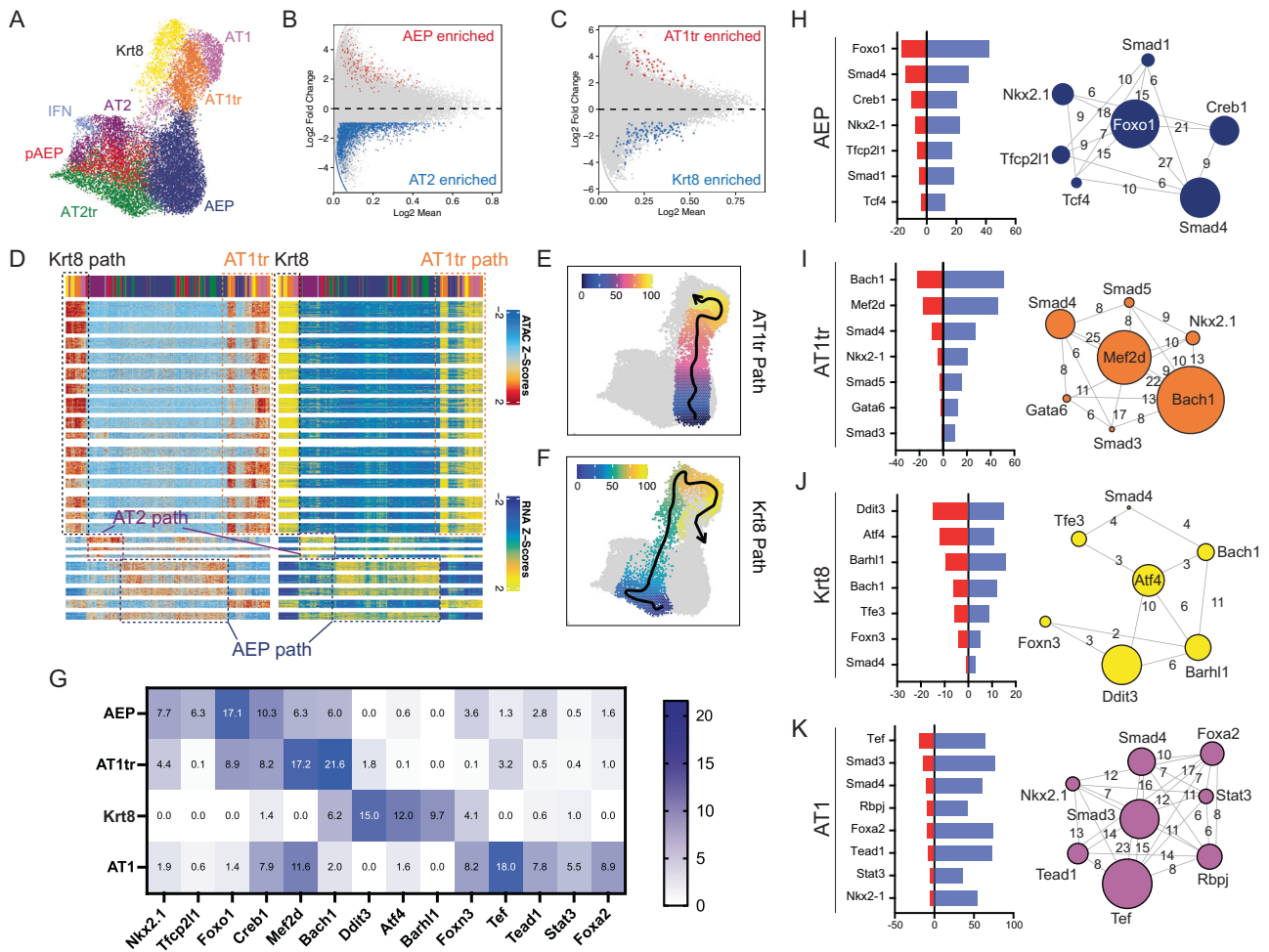


Fig. 4 | scATACseq analysis of AEP-derived organoid formation. **A** UMAP of cellular populations within organoids, named as in RNaseq integration. **B**, **C** Volcano plots showing differential chromatin accessibility regions between AEP and AT2 cells (**B**) and AT1tr and Krt8⁺ transitional cells (**C**). **D** Paired heatmap of differentially accessible genomic loci in ATAC (left) and RNA expression of nearest-neighbor gene production (right) showing an overview of regulators of AT1 cell differentiation (AT1 path), AT2 cell differentiation (AT2 path), and AEP state (AEP path) derived from integrated analysis. Cell populations shown along top bar, with colors the same as in (**A**). **E**, **F** Pseudotime prediction of separate AT1 differentiation

trajectories from AEPs to AT1 cells through AT1tr path (**E**) and Krt8⁺ path (**F**). **G** Transcriptional activity score (TAS, negative -log two-sided *p*-value of TF enrichment per cell type by Fisher exact test) for TFs per cell type in AT1 differentiation trajectories. **H–K** Top transcriptional regulators of AEP (**H**), AT1tr (**I**), Krt8⁺ (**J**), and AT1 (**K**) cell states. Red bar = TAS, and blue bar = # of regulated genes expressed in given cell type. Network diagram shows core regulator relationship, with circle size indicating TAS and numbers of bars showing co-regulated gene networks per cell type. Source data are provided within the Source Data file.

increased targeting, but at the expense of reduction in colony formation. Multiple biological replicates confirmed that MOI of 1000 led to targeting of approximately 60% of organoids with no change in colony formation efficiency or size of organoids (Fig. 5E); whole-mount IHC confirmed EYFP expression with no reduction in internal complexity of AEP-O (Fig. 5F). These results indicated that AAV6.2FF-Cre was capable of efficiently targeting AEPs in vitro for genetic manipulation without perturbing the system and confirmed the clonal nature of AEP-O.

Nkx2-1 deficient AEPs adopt a Krt8⁺ transitional cell state

To test the hypothesis that Nkx2-1 deficiency contributed to the Krt8⁺ state transition, we applied AAV6.2FF-Cre to AEPs from *Axin2^{CreERT2-tDTR} × R26R^{EYFP} × Nkx2-1^{fllox/fllox}* animals immediately after sorting. This generated Nkx2-1 knockout AEPs, which were used to initiate organoid formation (Fig. 6A). Morphology in EYFP⁺ (Nkx2-1^{KO}) organoids was noticeably different from EYFP⁺ organoids in the same well, and EYFP⁺ organoids were substantially larger by day 28 of culture (Fig. 6B). Because AAV6.2FF-Cre only targeted ~60% of organoids per well, we were able to directly compare EYFP⁺ and EYFP⁻ organoids grown in the same well to assess the impact of Nkx2-1 knockout in AEP-O. As

expected, we noted robust Nkx2-1 protein expression in EYFP⁻ organoids (Fig. 6C–E) and complete loss of Nkx2-1 protein in EYFP⁺ organoids. Nkx2-1^{KO} AEPs lost expression of AT2 markers, including *Sftpc*, with associated increased expression of E-cadherin (*Cdh1*) and change in cell shape and organoid morphology (Fig. 6F–H). Diverse morphological types were visible in Nkx2-1^{KO} AEP-O, with loss of alveolar-like cavities, prominence of one or a small number of large cavities full of debris, and pseudostratified epithelial lining with some organoids exhibiting a glandular appearance (Fig. 6F, I). These structures were reminiscent of other endoderm-derived organs, including the esophagus, stomach, and intestine. Nkx2-1 knockout in differentiated distal lung lineages has been associated with the expression of foregut endoderm genes^{24,72–74}, so we examined protein expression of a large panel of proximal lung and foregut endoderm markers including *Sox2*, *Sox9*, *Cdx2*, *Gata4*, and *Pdx1*. No substantial protein expression of any of these markers was detectable, suggesting Nkx2-1 KO AEP-O epithelium did not adopt foregut endodermal fate from proximal lung or GI organs (Supplementary Fig. S8A–Q).

To clarify Nkx2-1^{KO} AEP-O composition, we turned to unbiased profiling. We performed scRNAseq in Nkx2-1^{KO} AEP-O at 28d of culture

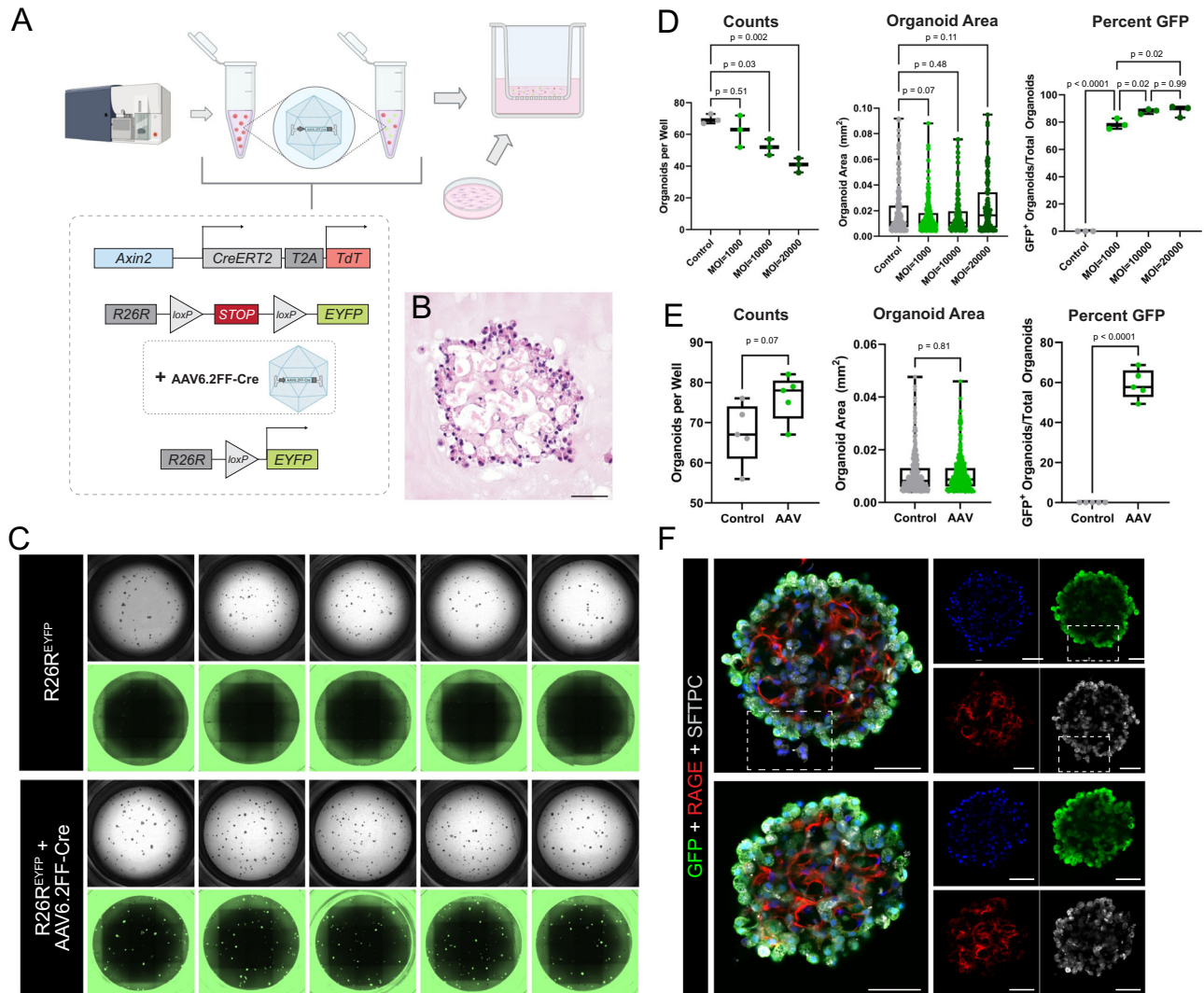


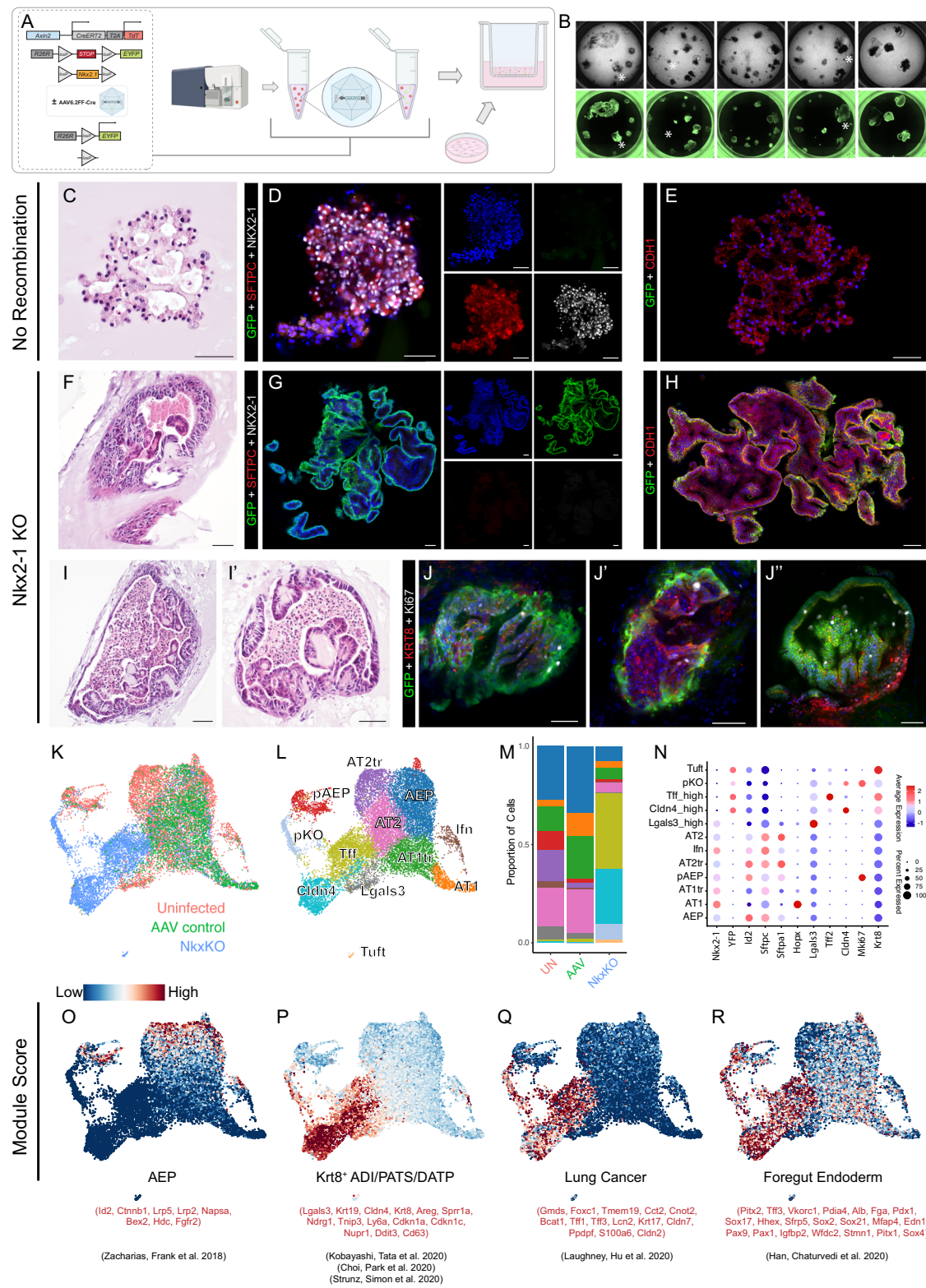
Fig. 5 | In vitro gene editing of AEP-derived alveolar organoids via AAV6.2FF-Cre. A AAV6.2FF-Cre experimental set-up. Live/CD31⁺/CD45⁺/CD326⁺ (EpCAM⁺)/TdtTomato⁺ (Axin2⁺) cells (AEPs) sorted from mice with the R26R^{EYFP} allele (Axin2^{CreERT2-tDT}; R26R^{EYFP}) were treated with AAV6.2FF-Cre and plated with wild-type fibroblasts. **B** H&E of 5 μ m sections of FFPE day 29 AAV6.2FF-Cre-treated organoids, exhibiting morphology and structural complexity similar to untreated/control organoids (Fig. 1B). **C** Whole-well brightfield and GFP images of day 29 organoids (untreated vs. AAV6.2FF-Cre-treated) at MOI = 1000. **D** Comparison of cells treated with an MOI of 1000, 10000, 20000. Quantification of day 32 organoids ($n = 3$ wells per condition) showing that an MOI of 1000 causes recombination in organoids without major effects on colony-forming efficiency (CFE). **E** Quantification ($n = 5$ wells per condition) showing that an MOI = 1000 induces significant levels of

recombination without effect on organoid number or size. **F** Whole-mount immunofluorescence of day 32 AAV6.2FF-Cre-treated AEP-derived organoids (same experimental set-up as Fig. 2). White box highlighting untargeted epithelial cells (YFP⁺/GFP⁻) next to a targeted (YFP⁺/GFP⁺) organoid in the same well, supporting clonal expansion of AAV6.2FF-Cre-treated cells. Note: EYFP was stained for using anti-GFP antibodies and imaged (whole-well images) using GFP filter cubes. Data throughout the figure represents greater than 5 biological replicates with 3 technical replicates per experiment. [Scale bars = 50 μ m]. AAV Adeno-associated virus, MOI multiplicity of infection. Box and whiskers plot shows box from 25–75% with central line at median and whiskers to the maximum and minimum points; all data points are graphed. Source data are provided within the Source Data file. Schematics created with Biorender.com.

and compared the expression profile of Nkx2-1^{KO} cells with Nkx2-1 expressing cells in our WT organoid scRNAseq time series. We added a control condition of AAV6.2FF-Cre treatment in AEP-O from Axin2^{CreERT2-tDT} \times R26R^{EYFP} (as in Fig. 5) to rule out any AAV6.2FF-Cre-specific effects. We then integrated scRNAseq expression data from Nkx2-1^{KO} and Nkx2-1^{+/+} AEP-O and compared gene expression profiles (Fig. 6K–R, Supplementary Fig. S9). Consistent with IHC, Nkx2-1 and EYFP RNA expression were mutually exclusive. Nkx2-1 was undetectable in EYFP⁺ epithelial cells at the RNA level (Fig. 6N, Supplementary Fig. S10). EYFP⁺ Nkx2-1^{KO} cells form multiple distinct clusters separated from control cell types (Fig. 6K). Nkx2-1^{KO} cells clustered near WT Krt8⁺ transitional epithelial cells by UMAP (Fig. 6L) and expressed high levels of transcripts enriched in the Krt8⁺/PATS/DATP/ADI

state, including *Cldn4*, *Tff2*, and intermediate levels of *Lgals3* (Fig. 6N, Supplementary Fig. S9Q–T). A distinct proliferative cluster was present by scRNAseq among *Cldn4*-high Nkx2-1^{KO} cells (Fig. 6K). Ki67 expression in Krt8⁺ Nkx2-1^{KO} AEP-O demonstrated ongoing proliferation at 40d of culture despite the large organoid size (Fig. 6J).

Given the close association of Nkx2-1^{KO} epithelial cells to the Krt8⁺ state in Nkx2-1 WT AEP-O, we used Seurat module scoring to compare Nkx2-1^{KO} and WT cells. Cells in Nkx2-1^{KO} organoids lost AEP-associated gene expression (Fig. 6O) while activating cell stress markers associated with Krt8⁺ cells in WT AEP-O. Cells from Nkx2-1^{KO} AEP-O were highly enriched for gene sets associated with human lung adenocarcinoma⁷⁵, concordant with previous findings implicating Nkx2-1 loss in the pathogenesis of lung cancer^{72,74}. While



we did not detect the protein expression of non-lung foregut endoderm markers by IHC, Nkx2-1 KO epithelial cells did express low levels of non-lung endodermal genes at the RNA level⁷⁶, consistent with loss of the instructive activity of Nkx2-1 in constraining lung fate. Finally, we evaluated the overall chromatin organization of Nkx2-1 KO epithelial cells using scATACseq (Supplementary Fig. S11A, B). Nkx2-1 knockout epithelium clusters in a distinct state from other epithelial cells in organoids (Supplementary Fig. S10A),

with evidence of large-scale changes in open chromatin compared to other organoid cell types (Supplementary Fig. S10B). These regions corresponded to increased chromatin opening near genes comprising the highly expressed Krt8⁺/PATS/DATP/ADI genes identified in scRNAseq. Taken together, these data support the concept that Nkx2-1 loss in AEPs led to the acquisition of the Krt8⁺/PATS/DATP/ADI-like cell state, validating the prediction of the Krt8⁺ cell state TRN analysis.

Fig. 6 | In vitro Nkx2-1 KO of AEP-derived alveolar organoids drives irreversible transition to a Krt8 stressed transitional /PATS/DAPT-like state. **A** AAV6.2FF-Cre experimental set-up. Live/CD31⁻/CD45⁺/CD326⁺(EpCAM⁺)/TdTomato⁺(Axin2⁺) cells (AEPs) sorted from Axin2^{CreERT2-tet}, R26R^{EYFP} mice and Axin2^{CreERT2-tet}, R26R^{EYFP}, Nkx2-1^{fl/fl} mice were treated with AAV6.2FF-Cre and plated with wild-type fibroblasts. **B** Comparison of brightfield and GFP whole-well images of organoids grown from control (AAV6.2FF-Cre-treated sorted R26R^{EYFP} AEPs) and Nkx2-1 KO AEPs (AAV6.2FF-Cre-treated sorted R26R^{EYFP}; Nkx2-1^{fl/fl} AEPs) at day 28 of culture. Control (non-GFP) organoids with normal morphology are marked with a white asterisk. **C–J** H&E and immunofluorescence images of R26R^{EYFP}; Nkx2-1^{fl/fl} AEP-derived organoids that did (**F–J**) or did not (**C–E**) undergo recombination via AAV6.2FF-Cre. **C–E** Non-recombined organoids (**D**) express SPC (red) and Nkx2-1 (white) but do not express the YFP lineage label (green), whereas **G** recombined organoids do not express SPC or Nkx2-1 but do express the YFP lineage label. Non-recombined (**E**)

and recombined (**H**) organoids maintain epithelial identity expressing CDH1. Nkx2-1^{KO} organoids express KRT8 and many proliferate and express Ki67 as late as day 40 of culture (**J–J'**). Data (**C–J**) represents 4 biological replicates with 3 technical replicates per experiment. **K–R** Integrated scRNAseq comparing epithelial cells from day 28 control organoids (Uninfected), AAV6.2FF-Cre-treated control organoids (AAV control), and AAV6.2FF-Cre-treated Nkx2-1 KO organoids (Nkx^{KO}). Nkx^{KO} cells cluster separately from Uninfected and AAV control cells near Krt8⁺ cells (**K, L**), which make up a majority of cells in the Nkx^{KO} condition (**M**). Marker genes for normal alveolar epithelium are lost and different markers gained (**N**) in Nkx^{KO}. **O–R** Module scoring using published gene sets for AEPs (**O**), Krt8/PATS/DATP/ADI cells (**P**), lung cancer cells (**Q**), and foregut endoderm (**R**). Compare to Supplementary Fig. S9 for marker gene analysis. [Scale bars = 50 μ m]. Schematics created with Biorender.com.

Nkx2-1 deletion in AEPs in vivo causes loss of progenitor state

To further evaluate the hypothesis that Nkx2-1 loss is sufficient to induce the Krt8⁺ cell state in AEPs, we performed in vivo lineage tracing of Nkx2-1^{KO} AEPs in adult mice at homeostasis. While Axin2^{CreERT2} functions as an effective lineage tracing reagent for lung epithelial cells after high dose tamoxifen treatment^{3,17}, the relative CreERT2 recombination inefficiency with this line prevented full deletion of floxed alleles in lineage labeled cells in adult lung epithelium at high doses of tamoxifen (Supplementary Fig. S12). We therefore turned our attention to the recently reported epithelial-specific^{77,78} Tfcp2l1^{CreERT2} mouse line⁷⁹. We performed lineage tracing of Tfcp2l1^{CreERT2} x R26R^{EYFP} animals and detected lineage label in solitary AT2 cells (SFTPC⁺/NKX2-1⁺) scattered throughout individual alveoli after administration of tamoxifen with similar distribution to Axin2^{CreERT2} (Supplementary Fig. S13A–D). We then performed scRNAseq on Tfcp2l1^{CreERT2} x R26R^{EYFP}; EYFP labeled cells constituted a subset of AT2 cells with a molecular signature indistinguishable from sorted Axin2^{CreERT2-tet} AT2 cells (Supplementary Fig. S13E–L). Sorted Tfcp2l1-lineage AT2 cells efficiently form complex organoids indistinguishable from those generated by sorted Axin2-positive AEPs (Supplementary Fig. S14). In combination with recent reports showing a role for Tfcp2l1 in AT2 cells during injury repair³³, our data suggested that Tfcp2l1^{CreERT2} constituted an AEP-enriched inducible Cre line suitable for epithelial-specific knockout of Nkx2-1.

We therefore generated Tfcp2l1^{CreERT2} x R26R^{EYFP} x Nkx2-1^{fl/fl}/lox animals, enabling Nkx2-1 knockout in AEPs during adult alveolar homeostasis (Fig. 7A–N). At 2 weeks following Nkx2-1 deletion in AEPs, we detected multi-cellular clones of lineage labeled EYFP⁺, Nkx2-1^{KO} epithelial cells throughout the lung (Fig. 7F–I). Nkx2-1^{KO} Tfcp2l1-lineage cells lost expression of AT2 markers, including Sftpc, changed shape, and acquired an E-cadherin high, Krt8-expressing, proliferative state (Fig. 7J–M), consistent with changes seen in Nkx2-1^{KO} organoids. At 4 weeks of age, these Krt8⁺ clones had grown substantially, with persistent shape change and high-level proliferation (Fig. 7N–U). Disruption of alveolar morphology was present in some areas near larger clones (Fig. 7M).

Next, we compared the molecular state of Nkx2-1^{KO} cells generated in vitro using AAV6.2FF-Cre in Axin2^{CreERT2-tet} x R26R^{EYFP} x Nkx2-1^{fl/fl}/lox and in vivo in Tfcp2l1^{CreERT2} x R26R^{EYFP} x Nkx2-1^{fl/fl}/lox. We performed whole-lung scRNAseq in Tfcp2l1^{CreERT2} x R26R^{EYFP} x Nkx2-1^{fl/fl}/lox animals (Fig. 8A–C) and identified a Nkx2-1^{KO} specific, distinct epithelial cell population which was lineage labeled with EYFP and expressed multiple markers of Krt8⁺-like cells (Fig. 8D–F). We then evaluated the molecular signature of this population compared to those found in Nkx2-1^{KO} organoids and confirmed that this population clustered with Nkx2-1^{KO} cells from AEP-O and share extensive molecular similarity (Fig. 8G, H). This similarity is mirrored at the chromatin level, with extensive overlap of chromatin openness in Nkx2-1 KO cells generated in either organoid culture or in vivo. Taken together, these findings confirm that, as in AEP-O, Nkx2-1 loss in AEPs in vivo drives the acquisition of the Krt8⁺/PATS/DATP/ADI molecular state, with

spontaneous proliferative growth and disruption of alveolar architecture (Fig. 9).

Discussion

In summary, combined scRNAseq and scATACseq of AEP-derived alveolar organoids allowed the definition of regulatory networks along multiple differentiation trajectories of lung epithelial progenitor cells toward differentiated alveolar epithelium. Transcriptomic and epigenomic data demonstrated two distinct AEP to AT1 differentiation trajectories, one through a known stressed transitional cell state and another through a state defined by signaling integration and transcriptional regulators of AT1 cell gene programs. A major difference in the regulation of these two cell states was the activity of Nkx2-1. Nkx2-1 expression in AEPs is required to maintain the AEP state, and loss of Nkx2-1 activity is sufficient for AEPs to enter the Krt8⁺ stressed transitional state. Nkx2-1 therefore plays a crucial, previously unrecognized role in the maintenance of progenitor function in the adult lung. These findings were enabled by the characteristics of the AEP-O system, allowing close interrogation of progenitor dynamics in a model with similar cellular and morphological complexity to alveolar regeneration in vivo.

Nkx2-1 loss is sufficient to cause the transition of AEPs to a proliferative stressed transitional state, a finding that emphasizes the need for active maintenance of adult AT2 alveolar progenitor capacity. At a basic level, the centrality of Nkx2-1 in lung progenitors is not surprising, as prior reports have demonstrated a requirement for Nkx2-1 expression in the maintenance of lung epithelial fate in adult differentiated cells^{24,59,72–74}. Loss of one or more alleles of Nkx2-1 is a common mutation found in lung adenocarcinoma⁸⁰. At a deeper level, however, it is quite provocative that the loss of a single transcription factor in a facultative progenitor lineage is sufficient to drive not only a fate transition but the conversion of a poised quiescent lineage to unconstrained proliferative growth. It is increasingly understood that transcription factors exert influence on gene expression through changes in chromatin state at regulatory elements, especially so-called “pioneer factors”. Nkx family factors are known to function as pioneer transcription factors in diverse contexts⁸¹, and pioneer factors catalyze changes in chromatin structure that maintain epigenetic stability⁸². There is a clear need to define with precision the mechanisms underlying the dynamic Nkx2-1 activity in chromatin reorganization²⁴. Important first steps include addressing the interactions of common chromatin regulatory complexes with known lung lineage factors, including Nkx2-1, Foxa, Gata, and Hopx, and carefully evaluating changes in and maintenance of the epigenomic state and chromatin topology across time during the lifespan of the lung.

The identification of Nkx2-1 as a regulator of the AEP progenitor state relied on the ability to distinguish similar cell states in a dynamic system. To that end, we utilized a clonal system with a known input/initial cell state obtained by sorting enriched AEPs, providing a clear ability to identify the starting point of differentiation. Identification of

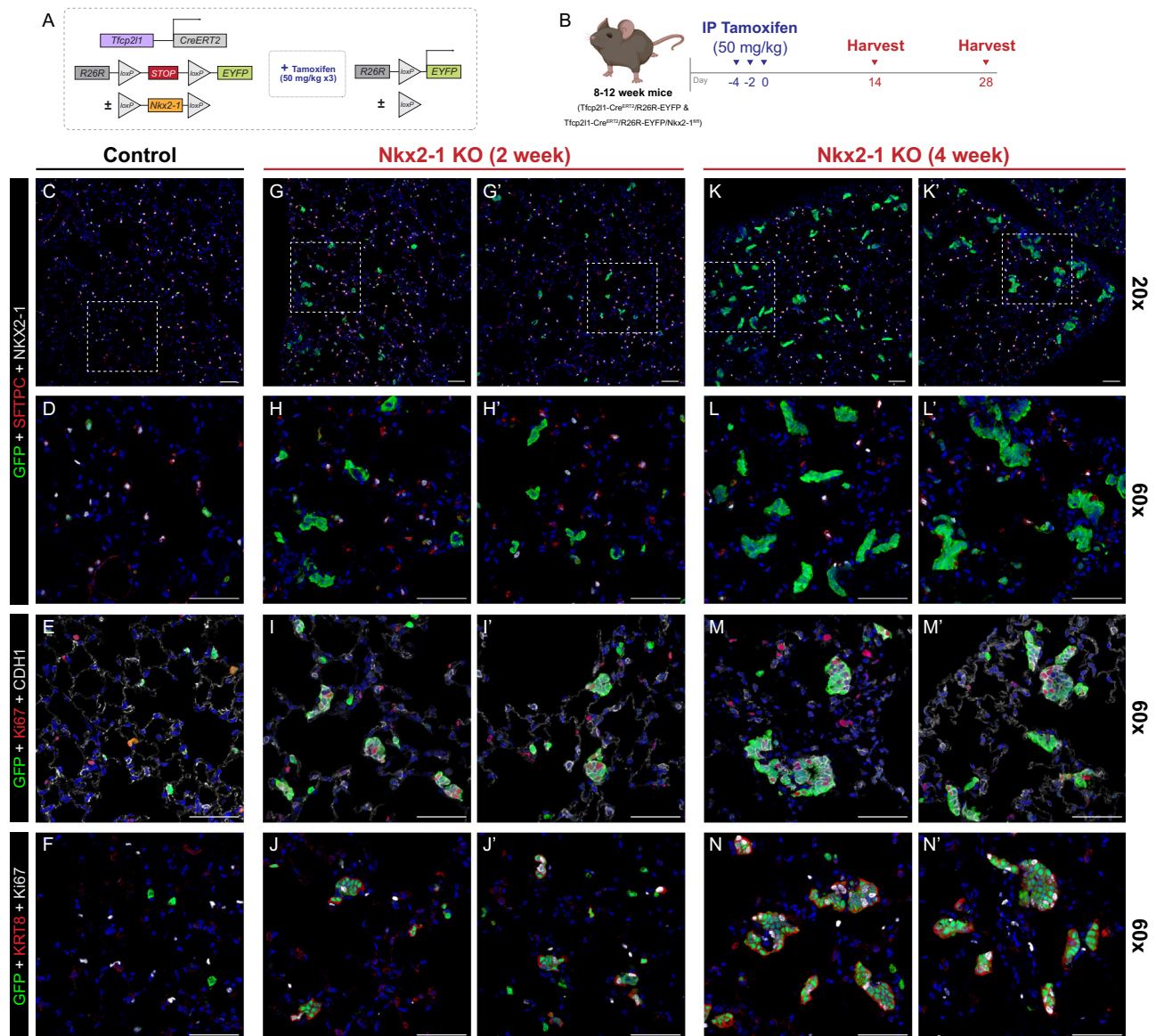


Fig. 7 | Genetic deletion of Nkx2-1 in vivo leads to loss of distal lung fate and acquisition of PATS/Krt8⁺ state. A, B Experimental design of in vivo genetic ablation of Nkx2-1 in AEPs. Mouse genetic construct (A) and experimental treatment plan and schematic (B). 8–12-week Tfc2p211-CreERT2; R26R^{EYFP} (C–F) and Tfc2p211-CreERT2; R26R^{EYFP}; Nkx2-1^{fl/fl} (G–N) were treated with three doses of IP tamoxifen (50 mg/kg) and harvested at 2 to 4 weeks post-treatment; control is from 2-week time point. C Control (Tfc2p211-CreERT2; R26R^{EYFP}) mice exhibited YFP induction in a subset of AT2 cells (SPC⁻ [red]/Nkx2-1⁺ [white]) with normal

histological characteristics. G, H, K, L Nkx2-1 KO (Tfc2p211-CreERT2; R26R^{EYFP}; Nkx2-1^{fl/fl}) mice exhibited clustered YFP⁺ proliferative clones negative for AT2 cell markers (SPC⁻ [red]/Nkx2-1⁻ [white], H, L), with acquisition of Ki67 and Krt8 expression (I, J, M, N). Progressive clonal enlargement by 4 weeks post-treatment (K) disrupts normal lung morphology, with continued growth and proliferation. Data throughout the figure represents 3 biological replicates with 3–5 animals per experiment. [Scale bars = 50 μm]. Schematics created with Biorender.com.

initial root cell in trajectory, pseudotime, and RNA velocity analyses can be challenging, and the shortfalls can be mitigated by controlling input^{49,83}. Knowing the input cell state of AEP-O added confidence when the unbiased methods correctly identified AEPs or proliferating AEPs as the initial state; nonetheless, multiple additional lines of evidence were required to strengthen and trust this conclusion. Together, these factors enabled downstream identification of AT1 differentiation trajectories through both the AT1tr and Krt8⁺ states. The addition of chromatin state provided a more refined signature for AEPs, allowing better identification of AEPs within the AT2 population, which has been challenging based on scRNA transcriptome alone^{16,36}. Together, scRNA and scATAC provided the resolution necessary to derive TRN inference, leading to Nkx2-1. Beyond the biological impact of Nkx2-1 deletion in AEPs, the marked phenotype provides a clear proof of principle

that the addition of chromatin topology and regulatory network inference can identify unexpected regulators of lung biology. There are several other testable predictions of the AEP-O TRN, which will need to be validated individually. The reproducibility and flexibility of an organoid system will enable these studies, as reagents such as AAV6.2FF provide a platform for rapid screening of novel regulators of alveolar regeneration.

The marked differences between AT1tr and Krt8⁺ transitional cells further the complexity of choosing cellular targets for lung regeneration. Krt8⁺ stressed transitional cells have been described in diverse models, are found in mouse and human, are increased in several disease states, and are readily identifiable based on high-level expression of enriched markers⁸⁴. However, our data suggest that permanent acquisition of a stressed transitional state, such as seen in Nkx2-1 KO

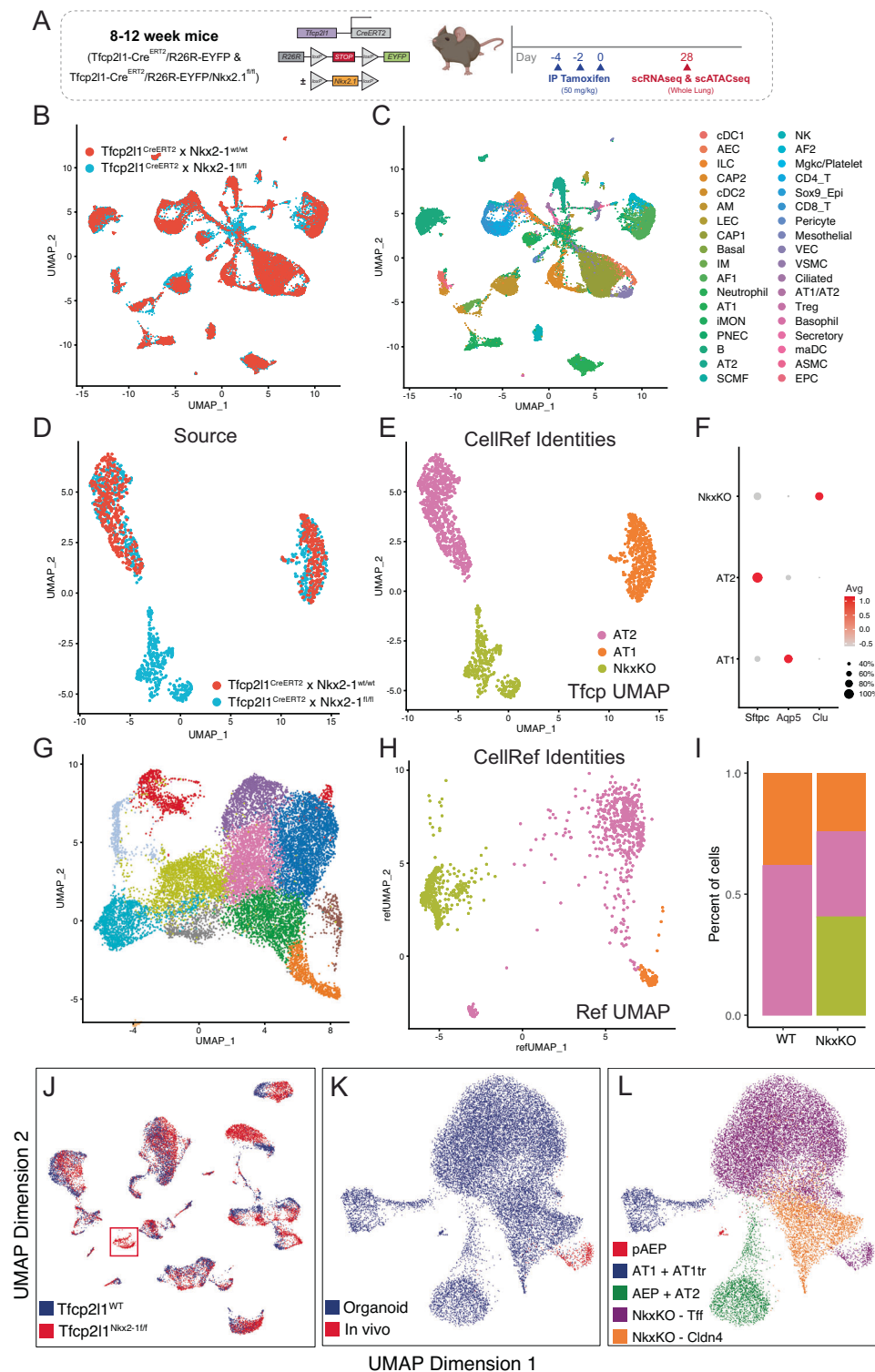


Fig. 8 | In vitro and in vivo Nkx2-1 deletion generates a common PATS/*Krt8*⁺-like molecular state. **A** Experimental design of in vivo genetic ablation of Nkx2-1 in AEPs prior to scRNAseq. **B, C** Common UMAP of whole-lung scRNAseq from WT and Nkx2-1^{KO} animals. **C** shows cell identities using LungMAP labels. **D–F** Distal epithelial cell populations in WT (red) and Nkx2-1^{KO} (blue) animals; a Nkx2-1^{KO} specific population is present, which expresses markers of *Krt8*/*PATS*/*DAPT*⁺ cells (*Clu*) but not AT1 or AT2 cells. **G–I** Overlap and label transfer of AEP-O identities to in vivo epithelial cells from **(D–F)**. **G** shows reference UMAP for comparison, reproduced from Fig. 6 with all NkxKO states concatenated together for comparison to in vivo cells. **H** shows clustering of in vivo epithelial cells on UMAP generated from

organoids as in **(G)** and colored by identity as in **(E)**. Proportions of cells in each state are shown in **(I)**. **J–L** scATACseq comparison of Nkx2-1 knockout states in vivo and in vitro. **J** shows overall UMAP of single nuclear ATACseq from WT and Nkx2-1^{KO} animals, demonstrating (as in RNA, compared to **(D)**) an Nkx2-1^{KO}-specific epithelial population (red box). **K** Integrated scATAC analysis of overlap of Nkx2-1^{KO} organoids (blue) and Nkx2-1^{KO}-specific epithelial cluster from **(J)** (red). Cells comprising the Nkx2-1^{KO}-specific epithelial cluster with Nkx2-1^{KO} cells from AEP-O, confirming epigenomic similarity of the Nkx2-1^{KO} cells generated in vivo and in vitro. Schematics created with Biorender.com.

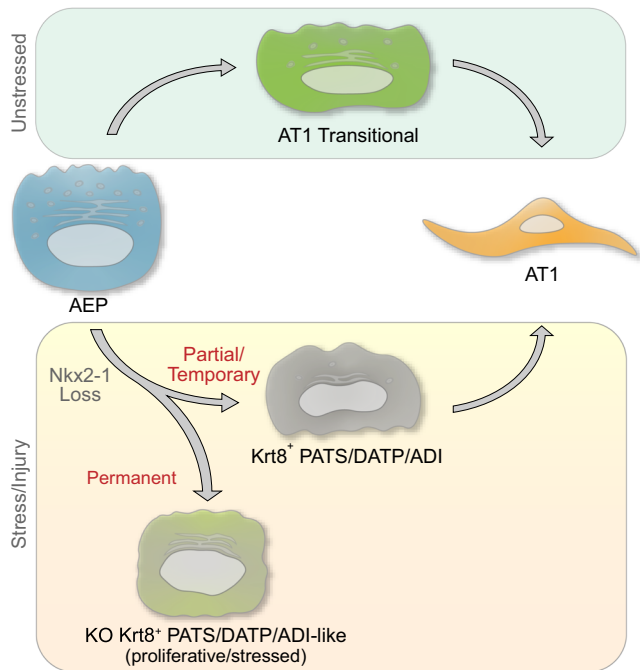


Fig. 9 | Model of Nkx2-1 activity in controlling progenitor and transitional cell state. AEP can differentiate to AT1 cells via either the AT1tr or Krt8⁺ states during homeostasis, with Nkx2-1 release from AT2 genes during transition through a Krt8⁺ state²⁴. Nkx2-1 activity and expression are lowest in Krt8⁺ cells²¹, and Nkx2-1 must re-engage chromatin to complete the AT1 transition from the Krt8⁺ state²⁴. Permanent Nkx2-1 loss in AEPs causes transition to proliferative, stressed, Krt8⁺-like state characterized by unconstrained growth in vitro and in vivo. Nkx2-1 is therefore required for the progenitor activity of AEPs.

organoids and Nkx2-1 KO AEPs in vivo, may drive aberrant proliferation, expression of lung cancer programs, and loss of lung identity. Progressive acquisition of this cell state may therefore be deleterious; even if most cells pass through or ‘recover’ from this stressed transitional state, accumulation of ‘stuck’ transitional cells may represent a risk factor for the development of lung disease. Our data suggest two potential avenues to move forward in identifying actionable therapeutic strategies. First, to identify factors, such as Nkx2-1, required for the progression of stressed transitional cells toward AT1 fate to promote progression or “rescue” stuck cells. Second, to determine mechanisms to promote progenitor cells to pass through the less stressed AT1tr transitional state as they differentiate. Future studies designed to address the determinants of these transition states during lung injury and modifiers of progression during regeneration and disease are therefore of high priority. Given the accumulation of stressed transitional cells in fibrosis¹⁸ and relative transience in acute regeneration²¹, both strategies are likely to be useful in different clinical contexts.

Restoration of a gas exchange surface through repair or replacement of injured alveoli is the central process needed to promote therapeutic regeneration. Here, we show that AEP-derived lung organoids recapitulate the major stages of the epithelial portion of the alveolar regenerative process, modeling progenitor cell expansion, alveolar epithelial differentiation, and formation of alveolar-like cavities with properly polarized and organized epithelium. Nkx2-1 deletion caused concordant changes both in organoids and in vivo, providing proof of principle that AEP-O model key aspects of adult alveolar biology and alveolar regeneration. While we characterized AEP-O only from mice in this study, prior reports have demonstrated the capacity for human AT2 progenitor culture^{3,7,11,85}, and advancing human primary alveolar progenitor culture is a high priority for

therapeutic development in tandem with continued progress in stem cell-based organoid technology.

Controversy exists regarding the similarity of alveolar regeneration and alveolar development;⁸⁶ Some have argued that regeneration is fundamentally different due to the altered milieu of the injured alveolus or even suggested that pathological remodeling rather than functional regeneration is the end state of a significant portion of lung injury⁸⁷. Our findings argue that extensive regenerative potential is encoded in the AEP epigenomic state, driving a distinct process to rebuild alveoli. AEPs contain the required information to undergo progenitor self-renewal, multilineage differentiation, and complex morphogenesis in the presence of a minimal signaling niche. This challenge bears a striking resemblance to the injured lung, where epithelial, mesenchymal, endothelial, immune lineages and the underlying matrix environment are all altered by pathogens⁸⁸. AEP-driven cavity formation occurs in the absence of mechanical contribution from myofibroblasts, quite different from during alveologenesis when the mechanical activity of myofibroblasts is required for the formation of alveoli^{89,90}. Our data therefore emphasize the difference between alveolar development and alveolar regeneration and suggest that AEP-O provide unique benefits to study regenerative biology specifically. Continued advancement of primary lung organoids with diverse cell compositions will provide models that balance fidelity and reproducibility, enabling and accelerating progress in defining molecular mechanisms of lung regeneration.

Methods

To optimize important aspects of the AEP-O culture, we undertook extensive reagent testing and standardization. The primary components of epithelial organoid co-culture assays are (1) epithelial cells; (2) supportive cells, if any; (3) matrix for three-dimensional suspension and growth; (4) media and media additives; and (5) growth surface (i.e., Transwell filter). We tested each of these components iteratively. As previously reported, AEPs form more and larger organoids than unselected AT2 cells³, so we focused on using FACS-sorted AEPs derived from Axin2^{CreERT2-tdT} mice as the epithelial starting fraction. To control for observed variability in organoid morphology and composition from different types of supportive mesenchymal cells, we used large, consistent preparations of primary lung fibroblasts obtained by selective adhesion from P28 wild-type C57BL/6 mice at passage 3-4. We then lot tested 4 available commercial matrices, all with compositions similar to Matrigel (Corning), and found that the best lots of Matrigel support organoid growth with 2-3-fold increased CFE compared to other Matrigel lots and competing products; we therefore confined our studies to a single lot of Matrigel and all data herein uses this standardized reagent. We cultured 5000 FACS-purified AEPs and 50,000 lung fibroblasts in each well on a Transwell filter in a 1:1 ratio of Matrigel and media (SAGM with 5% FBS and limited additives); we took this approach to minimize exogenous signaling modulators in the media and allow evaluation of the supportive capacity of the mesenchymal fraction. To evaluate the growth of these organoids more fully, we adapted methods from iPSC culture for whole-mount immunohistochemistry of organoids²⁷.

Ethical compliance and animals

All animal studies were conducted under the guidance and supervision of the Cincinnati Children’s Hospital Medical Center (CCHMC) Institutional Animal Care and Use Committee (IACUC) in accordance with CCHMC regulatory and biosafety protocols. Mouse lines used included: C57BL/6J mice (Jackson Strain #000664), PDGFRα^{EGFP} (B6.129S4-PDGFRα^{tm1(EGFP)Sor}/J); Jackson Strain #007669) Axin2^{CreERT2-tdT} (a gift from Edward Morrissey), Tfcp2l1^{CreERT2} (B6.129S-Tfcp2l1^{tm1(cre/ERT2)Ovi}/J); Jackson Strain #028732), Nkx2-1^{fl/fl} (a gift from Shioiko Kimura), and R26R^{EYFP} (B6.129×1-Gt(ROSA)26Sor^{tm1(EYFP)Cos}/J); Jackson Strain #006148). All

experiments for both organoids and in vivo lineage tracing included both male and female mice. For Cre recombinase induction in mouse models, 8–12-week-old mice were treated intraperitoneally (IP) with Tamoxifen (Sigma, T5648; dissolved in ethanol and resuspended in corn oil) at a dose of 50 mg/kg, one or three times (every other day), at the experimental time points indicated previously.

Mouse lung harvest

Mice were anesthetized via IP Ketamine + Xylazine, followed by euthanasia via cervical dislocation and thoracotomy. The chest cavity was opened to expose the heart and lungs. The right ventricle was perfused with 5–10 mL of cold PBS (Gibco, 10010-023) to clear blood from the lungs. For tissue dissociation for organoids, the lungs were removed and placed in cold PBS on ice. For tissue fixation for histology and immunofluorescence, the trachea was cannulated and lungs were inflated to a pressure of 30 cm H₂O using 4% paraformaldehyde (PFA). Inflated lungs were immersed in a conical of 4% PFA and then left on a rocker at 4 °C overnight.

Processing fixed lung tissue for histology and immunofluorescence

The day following inflation, fixed lung tissue was trimmed and placed in cassettes. The cassettes were washed (15 min each) 3x in DEPC-treated PBS, 1x in DEPC-treated 30% ethanol, 1x in DEPC-treated 50% ethanol, and 3x in DEPC-treated 70% ethanol. Following a standardized overnight automated processing protocol (Thermo Scientific, Excelsior ES), the samples were embedded in paraffin. Samples were sectioned at a thickness of 5 µm. Paraffin sections were incubated at 65 °C for 2 h, deparaffinized in xylene (3x for 10 min), rehydrated through an ethanol gradient, and standard H&E staining was performed. Slides were mounted with Permount Mounting Medium (Electron Microscopy Sciences, 17986-05) and cover slipped with #1.5 Gold Seal 3419 Cover Glass (Electron Microscopy Sciences, 63790-01).

Immunofluorescence on paraffin sections was performed as previously described³. Briefly, following deparaffinization, rehydration, and sodium citrate antigen retrieval (10 mM, pH 6.0), and blocking, immunofluorescence was performed on paraffin sections using antibodies in Supplementary Table 1b and the following reagents: ImmPRESS[®] HRP Horse Anti-Rabbit IgG Polymer Detection Kit (Vector Labs, MP-7401-50), ImmPRESS[®] HRP Horse Anti-Goat IgG Polymer Detection Kit (Vector Labs, MP-7405-50), and ImmPRESS[®] HRP Goat Anti-Rat IgG, Mouse adsorbed Polymer Detection Kit (Vector Labs, MP-7444-15). Following the application of TSA fluorophores (listed in Supplementary Table 1b; 1:100), sections were stained with DAPI (Invitrogen, D1306; 1:1000) and mounted using Prolong Gold antifade mounting medium (Invitrogen, P36930)."

Mouse lung digestion and single cell suspension

Clonal mouse alveolar epithelial progenitor (AEP)-based alveolar organoids were generated as previously described³ with minor modifications. Briefly, following harvest, lungs were removed from ice-cold PBS and non-pulmonary tissue and gross airways were removed via manual dissection, and lung tissue was finely chopped and transferred to a GentleMACS C tube (Miltenyi Biotec, 130-093-237) (tissue from one mouse per C tube) containing 5 mL of digestion buffer [composed of 9 mL of phosphate-buffered saline (PBS; Gibco, 10010-023) combined with 1 mL of Dispase (stock: 50 U/mL; final concentration: 5 U/mL, Corning, 354235), 50 µL of DNase (stock: 5 mg/mL; final concentration: 0.025 mg/mL or 50 U/mL, GoldBio, D-301), and 100 µL of Collagenase Type I (stock: 48,000 U/mL; final concentration of 480 U/mL, Gibco, 17100-017)]. C tubes were placed on a gentleMACS Octo Dissociator with Heaters (Miltenyi Biotec, 130-096-427), and the following protocols were run: "*m_lung_01_02*" (36 s) twice, "*37C_m_LIDK_I*" (36 min 12 s) once, and "*m_lung_01_02*" (36 s) once. Samples were passed through a 70 µm filter (Greiner Bio-One, 542070) and

centrifuged at 500 × g for 5 min at 4 °C. Following removal of the supernatant, 5 mL of RBC Lysis Buffer (Invitrogen, 00-4333-57) was added and incubated for 5 min. All centrifugation steps with this single-cell suspension were performed at 500 × g for 5 min at 4 °C for the following procedures.

Fibroblast stock preparation and maintenance

For generation of fibroblast stocks, 4-week C57BL/6J mice and/or 4-week PDGFR^α^{EGFP} mouse lungs were harvested, digested, and processed as described above. Following centrifugation, cells were washed 3x with MACS Buffer (autoMACS Rinsing Solution [Miltenyi Biotec, 130-091-222] with MACS BSA Stock Solution [Miltenyi Biotec, 130-091-376]). After removing supernatant from final wash, the cell pellet was resuspended in 10 mL fibroblast medium (DMEM/F-12 [Gibco, 11320-033], Antibiotic-Antimycotic [Gibco, 15240-062, final concentration 1x], and Heat Inactivated Fetal Bovine Serum [Corning, 35-011-CV, final concentration 10%]) and plated on a 10 cm tissue culture plate (approximately 1 mouse per plate). Non-adherent cells were removed via media change 2–12 h post-plating.

Cells were passaged at 80% confluency to P3. For passaging, media was removed from each plate and cells were washed with 5 mL of DPBS (Gibco, 14190-094). Then, 3 mL of 0.25% Trypsin-EDTA (Gibco, 25200-056) was added and plates were incubated at 37 °C for 7 min. Next, 5 mL of fibroblast medium was added to each plate, pipetted to dissociate cells, and transferred to a 15 mL conical tube. Cells were centrifuged at 500 × g for 5 min at 4 °C, the supernatant was removed, and the cell pellet was resuspended in 2 mL fibroblast medium/per plate (split 1:2 or 1:3) and transferred to plates containing 6 mL fibroblast medium. Once confluent at P3, cells were washed, trypsinized, centrifuged as above, and resuspended in 1 mL of freezing medium (90% FBS, 10% DMSO) (one plate per cryovial) and transferred to Mr. Frosty Cryogenic Freezing Container (Nalgene, 5100-0001) filled with isopropyl alcohol, which was placed in a –80 °C freezer overnight, before samples were moved to long-term liquid nitrogen storage.

For the use of frozen fibroblast stocks in organoids, 48 h prior to use in organoids, cells were rapidly thawed and resuspended in a 10 mL fibroblast medium in a 15 mL conical. Cells were centrifuged at 500 × g for 5 min at 4 °C, supernatant was removed, and cell pellet was resuspended in 2 mL fibroblast medium and transferred to a 10 cm tissue culture plate containing 6 mL of fibroblast medium. Fibroblasts used for organoids were washed, trypsinized, and resuspended (as described for passaging) before counting.

Processing for organoids, FACS, or cell sorting

Single-cell suspensions were obtained as above, and cells were resuspended in 5 mL MACS Buffer (autoMACS Rinsing Solution [Miltenyi Biotec, 130-091-222] with MACS BSA Stock Solution [Miltenyi Biotec, 130-091-376]) and passed through a 40 µm filter (Greiner Bio-One, 542040). Cells were centrifuged, the supernatant was removed, and the cell pellet was resuspended in Fc Receptor Binding Inhibitor Polyclonal Antibody (Invitrogen, 14-9161-73) diluted 1:100 in MACS buffer and incubated for 10 min at room temperature. Following centrifugation, cells were resuspended in a mixture of the following antibodies diluted 1:100 in MACS buffer and incubated for 10 min protected from light: CD31 (PECAM-1; Monoclonal Antibody [390], eFluor 450) (Invitrogen, 48-0311-82), CD45 (Monoclonal Antibody [30-F11], eFluor 450) (Invitrogen, 48-0451-82), CD326 (EpcAM; Monoclonal Antibody [G8.8], APC) (Invitrogen, 17-5791-82). Cells were washed 1x with 1–5 mL of MACS buffer and resuspended in Fixable Viability Dye eFluor 780 (Invitrogen, 65-0865-14) diluted 1:1000 in MACS buffer and incubated for 15 min protected from light. Cells were centrifuged and washed in 1–5 mL MACS buffer 3x. After the final wash/centrifugation, the cell pellet was resuspended in MACS buffer (volume adjusted for cell count) and passed through a 35 µm filter lid (Corning, 352235) into a FACS tube for sorting.

Using single-stain controls from experimental animals and wild-type littermates (TdTomato⁺) for compensation and adjusting gating to remove debris/doublets, the live/CD31⁺/CD45⁺/CD326⁺(EpCAM⁺)/TdTomato⁺ (AEP) population was sorted into a tube containing 'spiked' SAGM organoid medium (see "Organoid medium" section below) at 4 °C, using a BD FACSAria Fusion cell sorter with a 100 µm nozzle. Yield is approximately 10⁵ AEPs per mouse using this protocol. For generation of organoids from Tfcp2l1⁺ epithelial cells, Tfcp2l1-Cre^{ERT2}; R26R^{EYFP} mice were treated with tamoxifen as described above, followed by sorting of Live/CD31⁺/CD45⁺/CD326⁺[EpCAM⁺]/EYFP⁺ cells using the same sorter settings. Organoid initiation and maintenance with cells generated with both lines were performed identically.

Organoid growth medium

To generate 'spiked' SAGM medium for mouse lung alveolar organoids, SABM Small Airway Epithelial Cell Growth Basal Medium (Lonza, CC-3119) was combined with the following additives: SAGM Small Airway Epithelial Cell Growth Medium SingleQuots Supplements and Growth Factors (using only the BPE [2 mL], Insulin [0.5 mL], Retinoic Acid [0.5 mL], Transferrin [0.5 mL], and hEGF [0.5 mL] aliquots) (Lonza, CC-4124), Heat Inactivated Fetal Bovine Serum (Corning, 35-011-CV, final concentration 5%), Antibiotic-Antimycotic (Gibco, 15240-062, final concentration 1x), Cholera Toxin from *Vibrio cholerae* (Sigma, C8052, final concentration 25 ng/mL).

Standard organoid plating and maintenance

AEPs (live/CD31⁺/CD45⁺/CD326⁺[EpCAM⁺]/TdT⁺ cells) sorted from Axin2^{creERT2-tdT} mice were counted using a hemocytometer and resuspended in 'spiked' SAGM at a concentration 500 cells/µL. Fibroblasts were prepared (as described above), counted, and resuspended in 'spiked' SAGM at a concentration of 5000 cells/µL. For the remaining steps, it was extremely important that all reagents were kept cold/on ice and that bubbles were not introduced to mixtures when pipetting. It is recommended to prepare the plate (Falcon 24-well companion plates [Corning, 353504]), insert Transwells (Falcon Transwell Insert/Permeable Support with 0.4 µm membrane [Corning, 353095]), and place them on ice before use.

For each well of organoids to be plated, 10 µL AEPs (5000 total cells), 10 µL fibroblasts (50,000 total cells), and 25 µL 'spiked' SAGM were combined (create one master mix of cells and medium for all wells before adding Matrigel) and placed on ice. Corning Matrigel GFR Membrane Matrix (Corning, 356231) was added to the cell mixture (45 µL per well, 1:1 ratio of SAGM to Matrigel) and carefully mixed, then placed back on ice.

For plating, 90 µL of the combined cell/Matrigel mixture was pipetted carefully directly into the center of the Transwell (placed in the companion plate) without introducing bubbles. Organoid plates were incubated at 37 °C for 15 min, then 500 µL of 'spiked' SAGM supplemented with ROCK Inhibitor/Y-27632 Dihydrochloride (Sigma, Y0503, final concentration 0.01 mM) was added beneath the Transwell insert. After 48 h (and for subsequent media changes), media was replaced every 2 days with 'spiked' SAGM without ROCK inhibitor, and plates were maintained at 5% CO₂ and 37 °C.

AAV6.2FF-Cre organoids plating and maintenance

AAV6.2FF-Cre (titer of 2.779 × 10¹⁰ viral genomes [vg]/µL) was generated and characterized in vivo as previously described^{69,70}. Working dilutions (2.779 × 10⁹ vg/µL, 2.779 × 10⁸ vg/µL, and 2.779 × 10⁷ vg/µL) were generated via serial dilution of viral stocks in 'spiked' SAGM and frozen -80 °C in single-use aliquots. AEPs (live/CD31⁺/CD45⁺/CD326⁺[EpCAM⁺]/TdT⁺ cells) sorted from Axin2^{creERT2-tdT}; R26R^{EYFP} mice were counted using a hemocytometer and resuspended in 'spiked' SAGM at a concentration 1000 cells/µL. The total cells needed for the desired number of wells were transferred to a new 1.5 mL tube (i.e., 10 wells → 50,000 cells → 50 µL cells [1000 cells/µL]). Total cell number

per tube, desired MOI (i.e., 1000, 10000, 20000), and known viral titers were used to calculate the volume of virus needed from viral stocks. For each MOI, the calculated volume of the virus was added to each cell mixture, mixed, and incubated on ice for 60 min. Following viral incubation, 'spiked' SAGM and fibroblasts (5000 cells/µL) were added to create a mixture with the same proportions of cells as described above for standard plating of organoids (i.e., for each well – 5000 AEPs + 50,000 fibroblasts in 45 µL 'spiked' SAGM). The cell mixture was mixed with Matrigel (45 µL/well) and plated/maintained as described above for standard organoids.

Organoid plating with fibroblasts on the basolateral side of transwell

One day prior to organoid plating, fibroblasts were prepared (as described above), counted, and resuspended at a concentration of 50,000 cells in 100 µL in fibroblast medium. Transwells were placed in the wells of the companion plate, and then the plate was flipped so the Transwells rested on the inside of the lid. The companion plate was removed, exposing the basolateral side of the Transwells/filters. Then, 100 µL of the resuspended fibroblast mixture was added to the basolateral side of each Transwell filter. The plate base was placed back on top of the Transwells and was incubated (basolateral side up) at 37 °C and 5% CO₂ for 4 h. After incubation, the 100 µL of the medium was removed via gentle pipetting (without disturbing the filter) and the plate was flipped to the standard orientation. The Transwells were washed with 500 µL of DPBS (beneath the Transwell insert) and then moved to a fresh well/plate with 500 µL fibroblast medium beneath the Transwell insert. After standard isolation of epithelial cells (AEPs) for organoid plating, the Transwells were again washed with 500 µL of DPBS (beneath the Transwell insert), then the epithelial cell/Matrigel mixture (5000 AEPs in 45 µL 'spiked' SAGM + 45 µL Matrigel) was added to the apical side of the Transwell filter. Organoids were maintained as described above for standard organoids.

Fixation and processing for sections, histology, and H&E

Organoids were washed with 500 µL PBS, above and below Transwells. After removing PBS, 500 µL of 4% PFA was added above and below the Transwell for fixation overnight at 4 °C. Transwells were washed 5x (above and below) with PBS. Using a small knife or scalpel, the Transwell filter and Matrigel plug/organoids were cut out of the Transwell and placed on parafilm. Using forceps, the Transwell filter was carefully removed from the Matrigel plug/organoids [note: older organoid cultures are more likely to adhere to the filter]. Using a transfer pipet, HistoGel (Eprelia, HG400012) (pre-heated to liquid consistency) was added on top of the Matrigel plug/organoids until covered on all sides. Once solidified (-15–30 min), the sample was transferred to a tissue processing cassette (Fisher, 15-182-702 A). Once in cassettes, samples were processed as described for whole-lung processing above for paraffin embedding and sectioning, H&E, and immunofluorescence of paraffin sections.

Isolation of organoids for whole-mount immunofluorescence

Previously established whole-mount organoid staining protocols from Dekkers et al.²⁷ were adapted for mouse lung alveolar organoids using the following modifications. All steps used cut or wide-bore pipette tips. All steps after first wash and prior to fixation were performed on ice/with chilled reagents and used cut/wide-bore pipette tips coated in 1% BSA in PBS.

Briefly, Transwells were washed (above and below) with 500 µL of room-temperature PBS. Then, 500 µL of Cell Recovery Solution (Corning, 354253) was added to each Transwell, and a cut pipette tip was used to mechanically disrupt the Matrigel—the mixture was pipetted up and down and transferred to a new 24-well plate. Each Transwell was washed with an additional 250 µL of cell recovery solution and added to the new plate. The plate was incubated on ice (gel ice

packs were optimal) on an orbital/horizontal shaker for 60 min. The organoid mixture was transferred to a 15 mL conical pre-coated in 1% PBS-BSA. Each well was washed with 500 μ L of 1% PBS-BSA and added to the 15 mL conical. Wells from the same experimental condition (up to 4 wells) were combined in one conical. Conicals were filled to 10 mL with ice-cold PBS and centrifuged at 70 \times g for 5 min at 4 °C. The supernatant was removed very carefully [note: if the organoid pellet is not compact/tight, the entire pellet may be lost with suction due to loose matrix]. If Matrigel was still visible, the organoid pellet was gently resuspended in 1 mL of ice-cold 1% PBS-BSA and centrifuged again at 70 \times g for 5 min at 4 °C. After careful removal of the supernatant, the organoid pellet was resuspended in 1 mL of 4% PFA and incubated at 4 °C for 45 min (resuspending once halfway through incubation). For permeabilization, conicals were filled to 10 mL with 0.1% PBS-Tween and incubated overnight at 4 °C (alternate permeabilization option: for Click-iT protocols or shorter permeabilization, remove PFA and incubate in 0.25% Triton X-100 for 20 min at room temperature).

Whole-mount blocking and immunofluorescence

After isolation, fixation, and permeabilization, organoids were centrifuged at 70 \times g for 5 min at 4 °C, resuspended in 500 μ L of 5% Normal Donkey Serum (Jackson ImmunoResearch, 017-000-121) in 0.1% PBS-Triton X-100, and transferred to a 24-well plate for blocking. Organoids were incubated at room temperature on an orbital shaker for 1–2 h.

After blocking, the supernatant was removed from each well without disturbing organoids [note: supernatant was removed more easily when the plate was placed at a 45° angle for 5–10 min to allow organoids to settle to the bottom edge of the well]. Primary antibodies (see Supplementary Table 1) were diluted to a final concentration of 1:100 in 5% Normal Donkey Serum in 0.1% PBS-Triton X-100 (approximately 200–250 μ L total) and incubated overnight at 4 °C on an orbital shaker. For this protocol, a ‘quick wash’ was defined as adding 1 mL of organoid wash buffer (0.2% BSA, 0.1% Triton X-100 in PBS)²⁷ and immediately allowing organoids to settle/removing the wash, and a ‘long wash’ was defined as adding 1 mL of organoid wash buffer placing the plate on an orbital shaker for 1–2 h before allowing organoids to settle/removing the wash. After primary antibody staining, one ‘quick wash’ and three ‘long washes’ were performed. Then, secondary antibodies (see Supplementary Table 1) were diluted to a final concentration of 1:200 in 5% Normal Donkey Serum in 0.1% PBS-Triton X-100 (approximately 200–250 μ L total) and incubated overnight at 4 °C on an orbital shaker. For this step and all subsequent steps, samples were covered/protected from light to prevent photobleaching. After secondary antibody staining, one ‘quick wash’ was performed, then organoids were incubated in DAPI (Invitrogen, D1306, final concentration of 1:1000) in 5% Normal Donkey Serum in 0.1% PBS-Triton X-100 (approximately 200–250 μ L total) for 15 min. After removing the supernatant, one ‘quick wash’ and three ‘long washes’ were performed.

Whole-mount clearing and mounting

Following the final wash after immunostaining, as much wash buffer as possible was removed from each well, and organoids were transferred to a 1.5 mL tube. Organoids were centrifuged at 70 \times g for 5 min at 4 °C and as much supernatant as possible was removed without disturbing the organoids. Using a cut or wide-bore pipette tip, organoids were gently resuspended in room temperature fructose-glycerol clearing solution (60% vol/vol glycerol + 2.5 M fructose)²⁷. Depending on organoid volume, ~50–200 μ L of clearing solution was used. Organoids were left to clear for at least 1 day (and as long as several months) at 4 °C before mounting. Prior to preparing slides, cleared organoids were allowed to equilibrate to room temperature. Organoids were mounted as described previously²⁷—briefly, two pieces of double-sided tape were applied to a microscope slide approximately 25–30 mm apart, perpendicular to the length of the slide (for larger organoids,

additional layers of tape can be used). Using a PAP pen (Abcam, ab2601), a square was drawn between the two pieces of tape. Using a cut P200 pipette tip, approximately 20 μ L of organoids in clearing solution was placed in the middle of the drawn square, avoiding bubbles. A #1.5 Gold Seal 3419 Cover Glass (Electron Microscopy Sciences, 63790-01) was applied over the organoids, bridging the two pieces of tape. Slides were imaged immediately or stored at 4 °C.

Whole-mount Click-iT EdU staining

For whole-mount Click-iT EdU staining, the standardized commercial protocol for Click-iT EdU Cell Proliferation Kit for Imaging, Alexa Fluor 488 dye (Invitrogen, C10337) was combined with our optimized whole-mount immunofluorescence protocol. Kit reagents were prepared as directed in commercial protocols. Briefly, 48 h prior to harvest/fixation, the organoid medium was replaced with ‘spiked’ SAGM supplemented with EdU (final concentration of 10 μ M) from the commercial kit. After EdU incubation, all steps described in “Isolation of organoids for whole-mount immunofluorescence” were performed. Next, steps for EdU detection from the commercial kit’s standardized protocol with kit reagents were followed (i.e., 30-min incubation of “Click-iT Reaction Cocktail” at room temperature, followed by 1 mL wash with 3% PBS-BSA). Following EdU detection, whole-mount immunofluorescence was performed as described in the “Whole-mount blocking and immunofluorescence” and “Whole-mount clearing and mounting” sections above.

Whole-mount Click-iT TUNEL staining

For whole-mount Click-iT TUNEL staining, the standardized commercial protocol for Click-iT Plus TUNEL Assay for In Situ Apoptosis Detection, Alexa Fluor 488 dye (Invitrogen, C10617) was combined with our optimized whole-mount immunofluorescence protocol. Kit reagents were prepared as directed in commercial protocols. All steps described in “Isolation of organoids for whole-mount immunofluorescence” were performed. Organoids were washed twice with DI H₂O. Next, steps for ‘TdT Reaction’ and ‘Click-iT Plus Reaction’ from the commercial kit’s standardized protocol with kit reagents were followed (i.e., 60-min incubation of “TdT Reaction Mixture” at 37 °C, followed by 2 washes with 3% PBS-BSA, and a 30-min incubation of “Click-iT Plus TUNEL reaction cocktail” at 37 °C). Organoids were washed twice with 3% PBS-BSA, and then whole-mount immunofluorescence was performed as described in the “Whole-mount blocking and immunofluorescence” and “Whole-mount clearing and mounting” sections above.

Hoechst and live imaging preparation

For live imaging of organoids grown with PDGFR α ^{EGFP} fibroblasts, Hoechst 33342 (Invitrogen, H3570) was diluted 1:10000 in ‘spiked’ SAGM, and 500 μ L was added above and below the Transwell and incubated at 37 °C for 30–45 min. Using a small knife or scalpel, the Transwell filters and Matrigel plug/organoids were cut out of the Transwells and placed into a coverslip bottom dish (MatTek, P35G-1.5-20-C). For some samples, the entire Matrigel plug/filter was imaged, and for others, the Matrigel plug and filter were separated and imaged independently. Samples were covered in ‘spiked’ SAGM and cover slipped (MatTek, PCS-1.5-18) prior to imaging on an inverted confocal microscope.

Imaging

Brightfield H&E images were acquired on a Nikon Eclipse NiE Upright Widefield Microscope (Nikon DS-Fi3 Camera—with a Plan Apo VC 20x DIC N2 objective). Fluorescent images were acquired on Nikon A1 inverted LUNV and Nikon A1R inverted LUNV confocal microscopes using the following objectives: Plan Apo λ 10x, Plan Apo λ 20x, Apo LWD 20x WI λ S (water immersion), Apo LWD 40x WI λ S DIC N2 (water immersion), and SR HP Plan Apo λ S 100xC Sil (silicone immersion).

Second harmonic generation was performed to visualize fibrillar collagens I and II using a Nikon FN1 Upright Multiphoton microscope using the following objectives: Plan Apo VC 20x DIC N2 and Apo LWD 25x 1.10W DIC N2. Images were processed in Nikon Elements with minimal global adjustment of LUTs for acquired channels.

Organoid plate imaging/cytation imager

For whole-well imaging, plates were loaded into a Cytation 5 Imager (BioTek, CYT5PV) configured with a CO₂ gas controller (BioTek, 1210012). Plates were maintained at 5% CO₂ and 37 °C during imaging using *Cytation Gen5 Microplate Reader and Imager Software* (BioTek, version 3.08.01). Protocols specific to Falcon 24-well companion plates (Corning, 353504) and Falcon Transwell Insert/Permeable Support with 0.4 μm membrane (Corning, 353095) were established and used to take brightfield and fluorescent (GFP) 4x tile scans at 10 z-steps (-50 μm per step). 4x tile scans were used to generate z-projections. Individual tile scans and z-projections were used for further quantification.

Organoid quantification

Z-projections of stitched 4x images from each well were loaded into a custom Fiji-macro (run in Fiji/ImageJ v1.53) to count organoids per well, GFP⁺ organoids per well, and organoid area. This macro allowed for batch analysis of each experiment, reducing the subjectivity of counts. Briefly, given specific input parameters, the macro contained commands to: set the scale based on the diameter of each Transwell, subtract background, adjust image threshold, convert to mask, analyze particles/count objects meeting a specific threshold, and export data. Data was imported into GraphPad Prism 9.0 for analysis. T-tests were used for comparison of 2 groups, and ANOVA with prespecified multiple comparisons was used to compare 3 or more groups.

Electron microscopy

Fixation, sectioning, and acquisition of electron micrographs of alveolar cells were performed as previously described⁹¹.

Organoid dissociation and preparation of single-cell suspension for scRNAseq and scATACseq

Transwells were washed (above and below) with 1 mL of PBS. Then, 60 μL of organoid digest buffer (Dispase [Corning, 354235, undiluted, 50 U/mL], DNase I [GoldBio, D-301, final concentration 5 U/mL], Collagenase Type I [Gibco, 17100017, final concentration 4800 U/mL]) was added, and Matrigel plugs were gently disrupted and pipetted using a cut or wide-bore pipette tip. Organoids were incubated in a digest buffer for 30 min at 37 °C. Following incubation, the digested organoid mixture was pipetted several times and transferred to a low-binding 1.5 mL tube (3 wells of the same experimental condition combined into each tube). Then, 500 μL of cold PBS was added to each tube and incubated on ice for 5–10 min. Samples were centrifuged at 500×g for 5 min at 4 °C. Supernatant was removed carefully and the sample was washed with 1 mL of cold DPBS (Gibco, 14190-094). Following centrifugation at 500×g for 5 min at 4 °C and removal of supernatant, samples were resuspended in 60 μL of 0.25% Trypsin-EDTA (Gibco, 25200-056) and incubated for 30 min at 37 °C. Then, 1 mL of ice-cold PBS was added to each tube, samples were centrifuged at 500×g for 5 min at 4 °C, and the supernatant was carefully removed. Samples were washed 2x in 1 mL of cold 0.04% PBS-BSA and centrifuged at 500×g for 5 min at 4 °C. Following the removal of the supernatant, the samples were resuspended in 100 μL of 0.04% PBS-BSA. Prior to filtering cells, 40 μm Flowmi Cell Strainers (Bel-Art, H13680-0040) were equilibrated by passing 100 μL of 0.04% PBS-BSA through the strainer using a P1000 pipette tip. The 100 μL cell suspension was then pipetted through the 40 μm Flowmi Cell Strainer. Cells were counted manually using a hemocytometer and resuspended at a concentration of 1000 cells/μL prior to processing for scRNAseq.

Nuclei isolation from organoids for scATACseq

Using the same filtered cell suspension generated for scRNAseq, the standard 10x Genomics protocol for ‘Nuclei Isolation for Single Cell ATAC Sequencing’ (CG000212 Revision B) was followed. Briefly, the single cell suspension was centrifuged at 500×g for 5 min at 4 °C. Following removal of the supernatant, 100 μL of ATAC lysis buffer (from standard 10x Genomics protocol, CG000212 Revision B) was added, gently mixed and incubated on ice for 4–4.5 min. Immediately following incubation, 1 mL of chilled ATAC wash buffer (from standard 10x Genomics protocol) was added and gently mixed. Nuclei were centrifuged at 500×g for 5 min at 4 °C, supernatant was removed, and nuclei were resuspended in 100 μL of 1x nuclei buffer (diluted from 20x nuclei buffer [10x Genomics, 2000153/2000207]). Prior to filtering nuclei, 40 μm Flowmi Cell Strainers were equilibrated by passing 100 μL of nuclei buffer through the strainer using a P1000 pipette tip. The 100 μL of nuclei suspension was then pipetted through the 40 μm Flowmi Cell Strainer. Nuclei were counted manually using a hemocytometer and resuspended at a concentration of 5000 nuclei/μL prior to processing for scATACseq.

Mouse lung preparation for single-cell suspension for scRNAseq and scATACseq

For scRNAseq and scATACseq of adult mouse lungs, lungs were harvested, digested, and processed as described above in ‘‘Mouse lung digestion and single cell suspension.’’ Samples were then passed through a 40 μm filter (Greiner Bio-One, 542040) and centrifuged at 500×g for 5 min at 4 °C. Following removal of the supernatant, cells were resuspended in DMEM-F12 (Gibco, 11320-033), counted, and resuspended at a final concentration of 1000 cells/μL (50,000 cells total in 50 μL) for downstream processing for scRNAseq.

Cells for scATACseq were further processed as described in the ‘‘Nuclei isolation from organoids for scATACseq’’ section, with the following modifications: cells were left in lysis buffer for 120 to 150 s total before the addition of wash buffer. All other steps were completed as described for organoids. Finally, the nuclear count was adjusted to 5000 nuclei/μL for sequencing preparation.

Sequencing/library preparation

From each single cell or single nuclear preparation described above, a maximum of 16,000 cells or nuclei were loaded into a channel of a 10x Genomics Chromium system by the Cincinnati Children’s Hospital Medical Center Single Cell Sequencing Core. Libraries for RNA (v3) and ATACseq (v2) were generated following the manufacturer’s protocol. Sequencing was performed by the Cincinnati Children’s Hospital DNA Sequencing Core using Illumina reagents. Raw Sequencing data was aligned to the mouse reference genome mm10 with CellRanger 3.0.2 to generate expression count matrix files. To detect YFP expressing cells following Cre-mediated activation, a YFP contig was added to the mm10 genome following 10x Genomics ‘‘Build a Custom Reference’’ instructions (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/tutorial_mr) with modifications. Briefly, a custom EYFP.fasta file was generated using the ‘EYFP’ segment (682-1389) of the pEYFP-N1 plasmid sequence available through Addgene. This sequence was integrated into the standard mm10 assembly available from Ensembl to create a reference compatible for alignment with the CellRanger pipeline described above.

scRNAseq analysis and visualization

For RNAseq analysis, output data from CellRanger was partitioned into spliced and unspliced reads using Velocyto⁹². Velocyto output files were loaded into Seurat 4.0 using SeuratWrappers and SeuratDisk using the *ReadVelocity* command, and spliced transcripts were used as the expression input to *SCTransform*. Cells with less than 2000 or more than 8000 features were filtered and cells were clustered using the standard Seurat workflow. Putative doublets were identified and

removed using DoubletFinder⁹³, and libraries from individual time points and treatments were integrated using *SelectIntegrationFeatures* and *IntegrateData* commands in Seurat. Following integration, cells were re-clustered, UMAP project generated, and samples identified based on expression similarity to published data as described in the “Results”. Module scoring was performed using *AddModuleScore* function in Seurat for gene sets indicated in the figures. For lineage inference, these Seurat objects were directly used for Slingshot⁴⁷ pseudotime inference and were converted to a h5ad file using the *SaveH5Seurat* command. These h5ad file were used as input to scVelo⁴⁸ and CellRank⁹⁴ in Python 3.9.12 in Spyder following the standard pipeline (scVelo.readthedocs.io) to generate RNA velocity mapped to the Seurat UMAP and cell populations. For ligand-receptor analysis, we used CellChat³⁴ (<https://github.com/sqjin/CellChat>) v1.1 using the *SecretedSignaling* subset of the Mouse CellChatDB, with default parameters. Visualizations were generated with these tools and ggplot2.

scATACseq analysis and visualization

For ATACseq analysis, CellRanger output was loaded into ArchR⁵⁰, and Arrow files were generated per package defaults. Clusters were generated based on ATACseq parameters and named based on the evaluation of integrated gene expression from the paired Seurat RNA object. Peak calls for regions of open chromatin were generated from pseudobulk analysis of each cell state, followed by peak calling in MACS2. Differential open chromatin peaks were identified based on $FDR < 0.01$ and $\text{Log}_2FC \geq 1$ between cell states. Visualizations were generated using standard ArchR commands.

Transcriptional regulatory networks and visualization

For TRN inference, scRNAseq gene expression data and scATACseq chromatin accessibility data from each epithelial cell population were used as input for the Inferelator 3.0 (<https://github.com/flatironinstitute/inferelator>) with minor modifications in Python 3.9.22. Enriched transcriptional regulators were identified per cell type by performing Fisher’s exact test to compare the observed vs expected number of genes regulated in a cell type based on the TRN model, and visualizations were generated in Illustrator with details as noted in the figure legends.

Reporting summary

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

Data availability

All raw data has been uploaded to the GEO database under accession number [GSE215824](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE215824) and are also available directly by request to the contact PI. All data are available at the LungMAP web portal (<https://lungmap.net/explore-data/>) for full visualization and individual exploration. Source data are provided with this paper.

References

- Basil, M. C. et al. The cellular and physiological basis for lung repair and regeneration: past, present, and future. *Cell Stem Cell* **26**, 482–502 (2020).
- Barkauskas, C. E. et al. Type 2 alveolar cells are stem cells in adult lung. *J. Clin. Invest.* **123**, 3025–3036 (2013).
- Zacharias, W. J. et al. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* **555**, 251–255 (2018).
- Toth, A., Zhao, B. & Zacharias, W. J. *Lung Stem Cells in Development, Health and Disease*, 122–133 (European Respiratory Society, 2021).
- Grant, R. A. et al. Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. *Nature* **590**, 635–641 (2021).
- Bharat, A. et al. Lung transplantation for patients with severe COVID-19. *Sci. Transl. Med.* **12**, eabe4282 (2020).
- Barkauskas, C. E. et al. Lung organoids: current uses and future promise. *Development* **144**, 986–997 (2017).
- Hawkins, F. et al. Prospective isolation of NKX2-1-expressing human lung progenitors derived from pluripotent stem cells. *J. Clin. Invest.* **127**, 2277–2294 (2017).
- Jacob, A. et al. Differentiation of human pluripotent stem cells into functional lung alveolar epithelial cells. *Cell Stem Cell* **21**, 472–488.e410 (2017).
- Strikoudis, A. et al. Modeling of fibrotic lung disease using 3D organoids derived from human pluripotent stem cells. *Cell Rep.* **27**, 3709–3723.e3705 (2019).
- Salahudeen, A. A. et al. Progenitor identification and SARS-CoV-2 infection in human distal lung organoids. *Nature* **588**, 670–675 (2020).
- Katsura, H. et al. Human lung stem cell-based alveolospheres provide insights into SARS-CoV-2 mediated interferon responses and pneumocyte dysfunction. *Cell Stem Cell* **27**, 890–904.e8 (2020).
- Abo, K. M. et al. Human iPSC-derived alveolar and airway epithelial cells can be cultured at air-liquid interface and express SARS-CoV-2 host factors. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.06.03.132639> (2020).
- Zepp, J. A. et al. Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. *Cell* **170**, 1134–1148.e1110 (2017).
- Frank, D. B. et al. Emergence of a wave of Wnt signaling that regulates lung alveologenesis by controlling epithelial self-renewal and differentiation. *Cell Rep.* **17**, 2312–2325 (2016).
- Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
- Nabhan, A. N., Brownfield, D. G., Harbury, P. B., Krasnow, M. A. & Desai, T. J. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* **359**, 1118–1123 (2018).
- Kobayashi, Y. et al. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat. Cell Biol.* **22**, 934–946 (2020).
- Riemyndy, K. A. et al. Single cell RNA sequencing identifies TGF β as a key regenerative cue following LPS-induced lung injury. *JCI Insight* **4**, e123637 (2019).
- Choi, J. et al. Inflammatory signals induce AT2 cell-derived damage-associated transient progenitors that mediate alveolar regeneration. *Cell Stem Cell* **27**, 366–382.e367 (2020).
- Strunz, M. et al. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* **11**, 3559 (2020).
- Gibbs, C. S. et al. High-performance single-cell gene regulatory network inference at scale: the Inferelator 3.0. *Bioinformatics* **38**, 2519–2528 (2022).
- Miraldi, E. R. et al. Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells. *Genome Res.* **29**, 449–463 (2019).
- Little, D. R. et al. Differential chromatin binding of the lung lineage transcription factor NKX2-1 resolves opposing murine alveolar cell fates in vivo. *Nat. Commun.* **12**, 2509 (2021).
- Ostrin, E. J. et al. β -Catenin maintains lung epithelial progenitors after lung specification. *Development* **145**, dev160788 (2018).
- Kusakabe, T. et al. Thyroid-specific enhancer-binding protein/NKX2.1 is required for the maintenance of ordered architecture and function of the differentiated thyroid. *Mol. Endocrinol.* **20**, 1796–1809 (2006).
- Dekkers, J. F. et al. High-resolution 3D imaging of fixed and cleared organoids. *Nat. Protoc.* **14**, 1756–1771 (2019).
- Jain, R. et al. Plasticity of Hopx⁺ type I alveolar cells to regenerate type II cells in the lung. *Nat. Commun.* **6**, 6727 (2015).

29. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).
30. Sun, X. et al. A census of the lung: CellCards from LungMAP. *Dev. Cell* **57**, 112–145.e112 (2022).
31. Guo, M. et al. Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. *Nat. Commun.* **10**, 37 (2019).
32. Green, J., Endale, M., Auer, H. & Perl, A. K. Diversity of interstitial lung fibroblasts is regulated by platelet-derived growth factor receptor α kinase activity. *Am. J. Respir. Cell Mol.* **54**, 532–545 (2016).
33. Hamilton, T. G., Klinghoffer, R. A., Corrin, P. D. & Soriano, P. Evolutionary divergence of platelet-derived growth factor alpha receptor signaling mechanisms. *Mol. Cell Biol.* **23**, 4013–4025 (2003).
34. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
35. Sucre, J. M. S. et al. Hyperoxia injury in the developing lung is mediated by mesenchymal expression of Wnt5A. *Am. J. Respir. Crit. Care* **0**, 1249–1262 (2019).
36. Negretti, N. M. et al. A single-cell atlas of mouse lung development. *Development* **148**, dev199512 (2021).
37. Paris, A. J. et al. STAT3–BDNF–TrkB signalling promotes alveolar epithelial regeneration after lung injury. *Nat. Cell Biol.* **22**, 1197–1210 (2020).
38. Yuan, T., Volckaert, T., Chanda, D., Thannickal, V. J. & Langhe, S. P. D. Fgf10 signaling in lung development, homeostasis, disease, and repair after injury. *Front. Genet.* **9**, 418 (2018).
39. Lee, J. H. et al. Anatomically and functionally distinct lung mesenchymal populations marked by Lgr5 and Lgr6. *Cell* **170**, 1149–1163.e1112 (2017).
40. Ruppert, C., Henneke, I., Korfei, M., Seeger, W. & Günther, A. Role of HGF in the healthy and injured lung. *Eur. Respir. J.* **46**, PA3831 (2015).
41. Uniken Venema, W. T. C. et al. Gut mucosa dissociation protocols influence cell type proportions and single-cell gene expression levels. *Sci. Rep.* **12**, 9897 (2022).
42. Llufrío, E. M., Wang, L., Naser, F. J. & Patti, G. J. Sorting cells alters their redox state and cellular metabolome. *Redox Biol.* **16**, 381–387 (2018).
43. Box, A. et al. Evaluating the effects of cell sorting on gene expression. *J. Biomol. Tech.* **31**, 100–111 (2020).
44. Binek, A. et al. Flow cytometry has a significant impact on the cellular metabolome. *J. Proteome Res.* **18**, 169–181 (2019).
45. Pfister, G., Toor, S. M., Nair, V. S. & Elkord, E. An evaluation of sorter induced cell stress (SICS) on peripheral blood mononuclear cells (PBMCs) after different sort conditions—are your sorted cells getting SICS? *J. Immunol. Methods* **487**, 112902 (2020).
46. Neuschulz, A. et al. A single-cell RNA labeling strategy for measuring stress response upon tissue dissociation. *Mol. Syst. Biol.* **19**, e11147 (2023).
47. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
48. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
49. Gorin, G., Fang, M., Chari, T. & Pachter, L. RNA velocity unraveled. *PLoS Comput. Biol.* **18**, e1010492 (2022).
50. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
51. He, B. & Tan, K. Understanding transcriptional regulatory networks using computational models. *Curr. Opin. Genet. Dev.* **37**, 101–108 (2016).
52. Duren, Z., Chen, X., Jiang, R., Wang, Y. & Wong, W. H. Modeling gene regulation from paired expression and chromatin accessibility data. *Proc. Natl Acad. Sci. USA* **114**, E4914–E4923 (2017).
53. Cardenas-Diaz, F. L. et al. Temporal and spatial staging of lung alveolar regeneration is determined by the grainyhead transcription factor Tfcp2l1. *Cell Rep.* **42**, 112451 (2023).
54. Chung, M. I., Bujnis, M., Barkauskas, C. E., Kobayashi, Y. & Hogan, B. L. M. Niche-mediated BMP/SMAD signaling regulates lung alveolar stem cell proliferation and differentiation. *Development* **145**, dev163014 (2018).
55. Serra, M. et al. Pluripotent stem cell differentiation reveals distinct developmental pathways regulating lung- versus thyroid-lineage specification. *Development* **144**, 3879–3893 (2017).
56. Gaskill, C. F. et al. Disruption of lineage specification in adult pulmonary mesenchymal progenitor cells promotes microvascular dysfunction. *J. Clin. Investig.* **127**, 2262–2276 (2017).
57. Flodby, P. et al. Cell-specific expression of aquaporin-5 (Aqp5) in alveolar epithelium is directed by GATA6/Sp1 via histone acetylation. *Sci. Rep.* **7**, 3473 (2017).
58. Chung, C. et al. Hippo-Foxa2 signaling pathway plays a role in peripheral lung maturation and surfactant homeostasis. *Proc. Natl Acad. Sci. USA* **110**, 7732–7737 (2013).
59. Herriges, M. J. et al. Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. *Genes Dev.* **28**, 1363–1379 (2014).
60. LaCanna, R. et al. Yap/Taz regulate alveolar regeneration and resolution of lung inflammation. *J. Clin. Invest.* **130**, 2107–2122 (2019).
61. Gokey, J. J. et al. Active epithelial Hippo signaling in idiopathic pulmonary fibrosis. *JCI Insight* **3**, e98738 (2018).
62. 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.e2945 (2019).
63. Carvalho, A. D. et al. Glycogen synthase kinase 3 induces multi-lineage maturation of human pluripotent stem cell-derived lung progenitors in 3D culture. *Development* **146**, dev171652 (2019).
64. Yang, Y. et al. Transcription factor C/EBP homologous protein in health and diseases. *Front. Immunol.* **8**, 1612 (2017).
65. Beers, M. F. et al. Alveolar type 2 epithelial cell quality control responses to pulmonary fibrosis related SFTPC mutations are dysfunctional and substrate specific. *FASEB J.* **34**, 1–1 (2020).
66. Weng, A. et al. Lung injury induces alveolar type 2 cell hypertrophy and polyploidy with implications for repair and regeneration. *Am. J. Respir. Cell Mol. Biol.* **66**, 564–576 (2022).
67. Dolinay, T., Himes, B. E., Shumyatcher, M., Lawrence, G. G. & Margulies, S. S. Integrated stress response mediates epithelial injury in mechanical ventilation. *Am. J. Respir. Cell Mol. Biol.* **57**, 193–203 (2017).
68. Wu, H. et al. Progressive pulmonary fibrosis is caused by elevated mechanical tension on alveolar stem cells. *Cell* **180**, 107–121.e117 (2020).
69. Lieshout, L. P. V. et al. A novel triple-mutant AAV6 capsid induces rapid and potent transgene expression in the muscle and respiratory tract of mice. *Mol. Ther. Methods Clin. Dev.* **9**, 323–329 (2018).
70. Rindler, T. N. et al. Efficient transduction of alveolar type 2 cells with adeno-associated virus for the study of lung regeneration. *Am. J. Respir. Cell Mol. Biol.* **65**, 118–121 (2021).
71. Srinivas, S. et al. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001).
72. Maeda, Y. et al. Kras(G12D) and Nkx2-1 haploinsufficiency induce mucinous adenocarcinoma of the lung. *J. Clin. Investig.* **122**, 4388–4400 (2012).
73. Mollaoglu, G. et al. The lineage-defining transcription factors SOX2 and NKX2-1 determine lung cancer cell fate and shape the tumor immune microenvironment. *Immunity* **49**, 764–779.e769 (2018).

74. Tata, P. R. et al. Developmental history provides a roadmap for the emergence of tumor plasticity. *Dev. Cell* **44**, 679–693.e675 (2018).
75. Laughney, A. M. et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat. Med.* **26**, 259–269 (2020).
76. Han, L. et al. Single cell transcriptomics identifies a signaling network coordinating endoderm and mesoderm diversification during foregut organogenesis. *Nat. Commun.* **11**, 4158 (2020).
77. Du, Y., Guo, M., Whitsett, J. A. & Xu, Y. ‘LungGENS’: a web-based tool for mapping single-cell gene expression in the developing lung. *Thorax* **70**, 1092–1094 (2015).
78. Du, Y. et al. Lung Gene Expression Analysis (LGEA): an integrative web portal for comprehensive gene expression data analysis in lung development. *Thorax* **72**, 481–484 (2017).
79. Maruyama, E. O. et al. Cell-specific Cre strains for genetic manipulation in salivary glands. *PLoS ONE* **11**, e0146711 (2016).
80. Yamaguchi, T., Hosono, Y., Yanagisawa, K. & Takahashi, T. NKX2-1/TTF-1: an enigmatic oncogene that functions as a double-edged sword for cancer cell survival and progression. *Cancer Cell* **23**, 718–723 (2013).
81. Iwafuchi-Doi, M. & Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes Dev.* **28**, 2679–2692 (2014).
82. Mayran, A. & Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **293**, 13795–13804 (2018).
83. Gorin, G., Svensson, V. & Pachter, L. Protein velocity and acceleration from single-cell multiomics experiments. *Genome Biol.* **21**, 39 (2020).
84. Verheyden, J. M. & Sun, X. A transitional stem cell state in the lung. *Nat. Cell Biol.* **22**, 1025–1026 (2020).
85. Kathiriyai, J. J. et al. Human alveolar type 2 epithelium transdifferentiates into metaplastic KRT5⁺ basal cells. *Nat. Cell Biol.* **24**, 10–23 (2022).
86. Rodríguez-Castillo, J. A. et al. Understanding alveolarization to induce lung regeneration. *Respir. Res.* **19**, 148 (2018).
87. Barnes, P. J., Anderson, G. P., Fagerås, M. & Belvisi, M. G. Chronic lung diseases: prospects for regeneration and repair. *Eur. Respir. Rev.* **30**, 200213 (2021).
88. Kligerman, S. J., Franks, T. J. & Galvin, J. R. From the radiologic pathology archives: organization and fibrosis as a response to lung injury in diffuse alveolar damage, organizing pneumonia, and acute fibrinous and organizing pneumonia. *Radiographics* **33**, 1951–1975 (2013).
89. Li, R., Li, X., Hagood, J., Zhu, M. S. & Sun, X. Myofibroblast contraction is essential for generating and regenerating the gas-exchange surface. *J. Clin. Invest.* **130**, 2859–2871 (2020).
90. Zepp, J. A. et al. Genomic, epigenomic, and biophysical cues controlling the emergence of the lung alveolus. *Science* **371**, eabc3172 (2021).
91. Ikegami, M., Na, C. L., Korfhagen, T. R. & Whitsett, J. A. Surfactant protein D influences surfactant ultrastructure and uptake by alveolar type II cells. *Am. J. Physiol. Lung Cell Mol. Physiol.* **288**, L552–561 (2005).
92. Manno, L. G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
93. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e324 (2019).
94. Lange, M. et al. CellRank for directed single-cell fate mapping. *Nat. Methods* **19**, 159–170 (2022).

Acknowledgements

The authors would like to thank Cheng-Lun Na (CCHMC) for his expertise and assistance with electron microscopy and Tara Rindler for the preparation of the AAV6.2FF-Cre. The authors would also like to thank the Bio-Imaging and Analysis Facility (BAF) (especially Sarah McLeod), the CCHMC Single Cell Genomics Core (SCGC) (especially Kelly Rangel and Shawn Smith), the CCHMC Division of Bioinformatics (for providing/maintaining the high-performance computational cluster), the CCHMC DNA Sequencing Core, and the CCHMC Research Flow Cytometry Core (RFCC) in the Division of Rheumatology. Select figure schematics created using BioRender.com. A.T. was supported by HL007752 (NIH), E.R.M. by AI150748 (NIH), and W.J.Z. by HL140178, HL156860, HL166245, and AI150748 (NIH) and the CCHMC Procter Scholar Award.

Author contributions

Experimental design—A.T., P.K., W.J.Z.; Data acquisition—A.T., P.K.; Data analysis—A.T., J.S., M.K., J.A.W., E.R.M., D.S., W.J.Z.; Provision of critical reagents and methods expertise—J.P.B., M.K.; Writing of first draft of paper—A.T., W.J.Z.; Editing and critical revision—A.T., P.K., J.S., M.K., J.A.W., J.P.B., E.R.M., D.S., W.J.Z.; Acquisition of funding support—W.J.Z.

Competing interests

A.T. and W.J.Z. declare that they are inventors on a patent application (PCT/US2023/031312), which has been submitted and is currently pending, regarding the methods for creation and the composition of the AEP-O. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-44184-0>.

Correspondence and requests for materials should be addressed to William J. Zacharias.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023