Cell Genomics

A single-cell regulatory map of postnatal lung alveologenesis in humans and mice

Graphical abstract

Highlights

- Chromatin accessibility map of human and mouse postnatal alveolar development

- Construction of regulatory modules directing alveolar cell differentiation

- Temporal and interspecies gene regulation of mechanical stretch signaling

- Disease risk variants mapped to regulatory units in developmental cell types

Authors

Thu Elizabeth Duong, Yan Wu, Brandon Chin Sos, ..., Greg Myers, James S. Hagood, Kun Zhang

Correspondence

kzhang@bioeng.ucsd.edu

In brief

Duong et al. produce single-cell chromatin accessibility profiles from human and mouse lungs during critical alveolar development stages after birth. They present a computational template to characterize developmental cell types, construct regulatory modules to propose cell lineage-specific transcription factors directing expression of marker genes, and identify regulatory units critical to pathogenesis.

CelPress

Resource

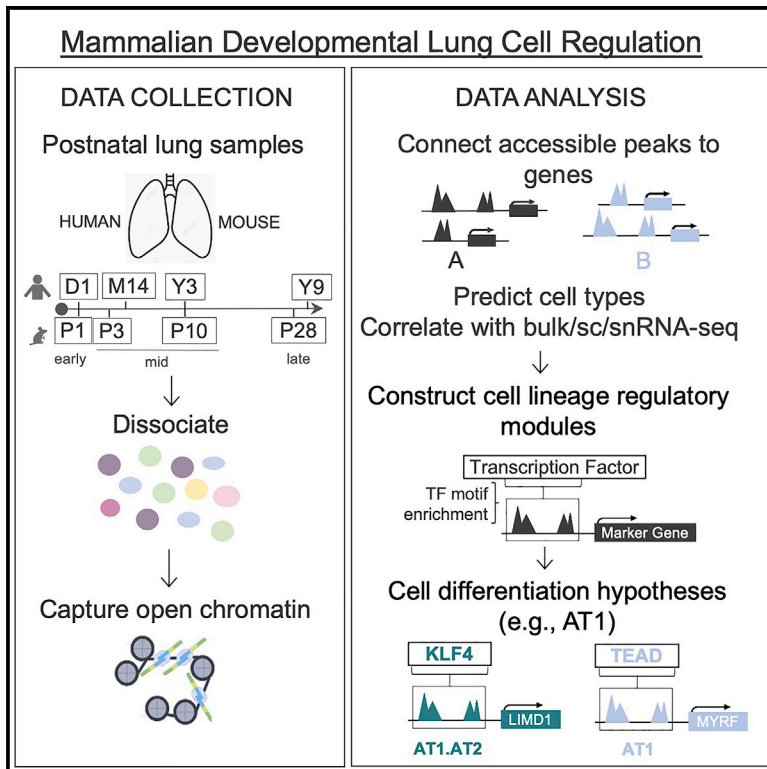# A single-cell regulatory map of postnatal lung alveologenesis in humans and mice

Thu Elizabeth Duong,[1] Yan Wu,[2] Brandon Chin Sos,[2] Weixiu Dong,[2] Siddharth Limaye,[2] Lauraine H. Rivier,[3] Greg Myers,[3] James S. Hagood,[3,4,5] and Kun Zhang[2,4,5,6,*]

[1]Department of Pediatrics, Division of Respiratory Medicine, University of California San Diego, La Jolla, CA 92093, USA
[2]Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA
[3]Department of Pediatrics, Division of Pediatric Pulmonology, University of North Carolina, Chapel Hill, NC 27599, USA
[4]Senior author
[5]These authors contributed equally
[6]Lead contact
*Correspondence: kzhang@bioeng.ucsd.edu
https://doi.org/10.1016/j.xgen.2022.100108

SUMMARY

Ex-utero regulation of the lungs' responses to breathing air and continued alveolar development shape adult respiratory health. Applying single-cell transposome hypersensitive site sequencing (scTHS-seq) to over 80,000 cells, we assembled the first regulatory atlas of postnatal human and mouse lung alveolar development. We defined regulatory modules and elucidated new mechanistic insights directing alveolar septation, including alveolar type 1 and myofibroblast cell signaling and differentiation, and a unique human matrix fibroblast population. Incorporating GWAS, we mapped lung function causal variants to myofibroblasts and identified a pathogenic regulatory unit linked to lineage marker *FGF18*, demonstrating the utility of chromatin accessibility data to uncover disease mechanism targets. Our regulatory map and analysis model provide valuable new resources to investigate age-dependent and species-specific control of critical developmental processes. Furthermore, these resources complement existing atlas efforts to advance our understanding of lung health and disease across the human lifespan.

INTRODUCTION

Embryonic and childhood pulmonary development lay the foundation for optimal lung growth, function, and responses to the environment. The origins of many adult lung diseases may be traced back to aberrations in these early processes. Our current understanding of mechanisms driving human lung development is sparse and largely stems from interrogating mouse models. Mice, however, are born with lungs at an earlier stage of development. Morphologic, cellular, and molecular differences strongly indicate that unique aspects of human pulmonary development can only be understood by directly examining human cells and tissues.[1] Limited access to tissue has been a barrier to studying human lung especially during alveologenesis, which is the most commonly disrupted process in lung disease.[2]

In humans, the formation of critical alveoli largely (~90%) occurs postnatally in two phases.[3] The first is characterized by alveolar septation, which subdivides distal saccules into alveoli lined by two epithelial cell types: flat type 1 (AT1) cells essential for gas-exchange and surfactant-secreting type 2 (AT2) cells.[4,5] The surrounding alveolar mesenchyme contains myofibroblasts (myoFB) characterized by alpha smooth muscle actin (*ACTA2*), contraction of which forms septal ridges that are vital for alveologenesis.[2,6,7] The localization of alveolar myoFBs has been shown to be dependent on the production of platelet derived growth factor subunit A (*PDGFA*) by epithelial cells to signal cognate platelet-derived growth factor receptor A+ (*PDGFRA*) mesenchymal progenitors.[8–10] In addition to epithelial-mesenchymal interactions critical to alveolar septation, mechanical forces during postnatal breathing, including stretch-dependent processes, promote alveolar cell differentiation and proliferation.[11] The second phase of alveologenesis consists of continued septation from mature septa, mechanisms for which remain unclear.

A surge in the development and successful application of high-throughput, single-cell (sc) transcriptomic approaches to profile human tissues is revolutionizing our ability to more comprehensively characterize human lung cell types. Collaborative efforts by the Human Cell Atlas (HCA, www.humancellatlas.org),[12] the Human BioMolecular Atlas Program (HuBMAP, www.hubmapconsortium.org), and the Lung Molecular Atlas Program (LungMAP, www.lungmap.net)[13] are underway to produce healthy Human Lung Cell Atlases. To augment these efforts, we utilized scTHS-seq[14] to generate a human and mouse epigenetic atlas to contribute a critical regulatory dimension during postnatal lung development. By optimizing the dissociation of archived lung tissue and processing all samples in one run, we captured 40,427 human and 41,476 mouse sc accessible chromatin datasets with minimal batch effects across four time points each beginning at postnatal day one. We assembled a computational pipeline to define key cellular states, construct critical species-specific
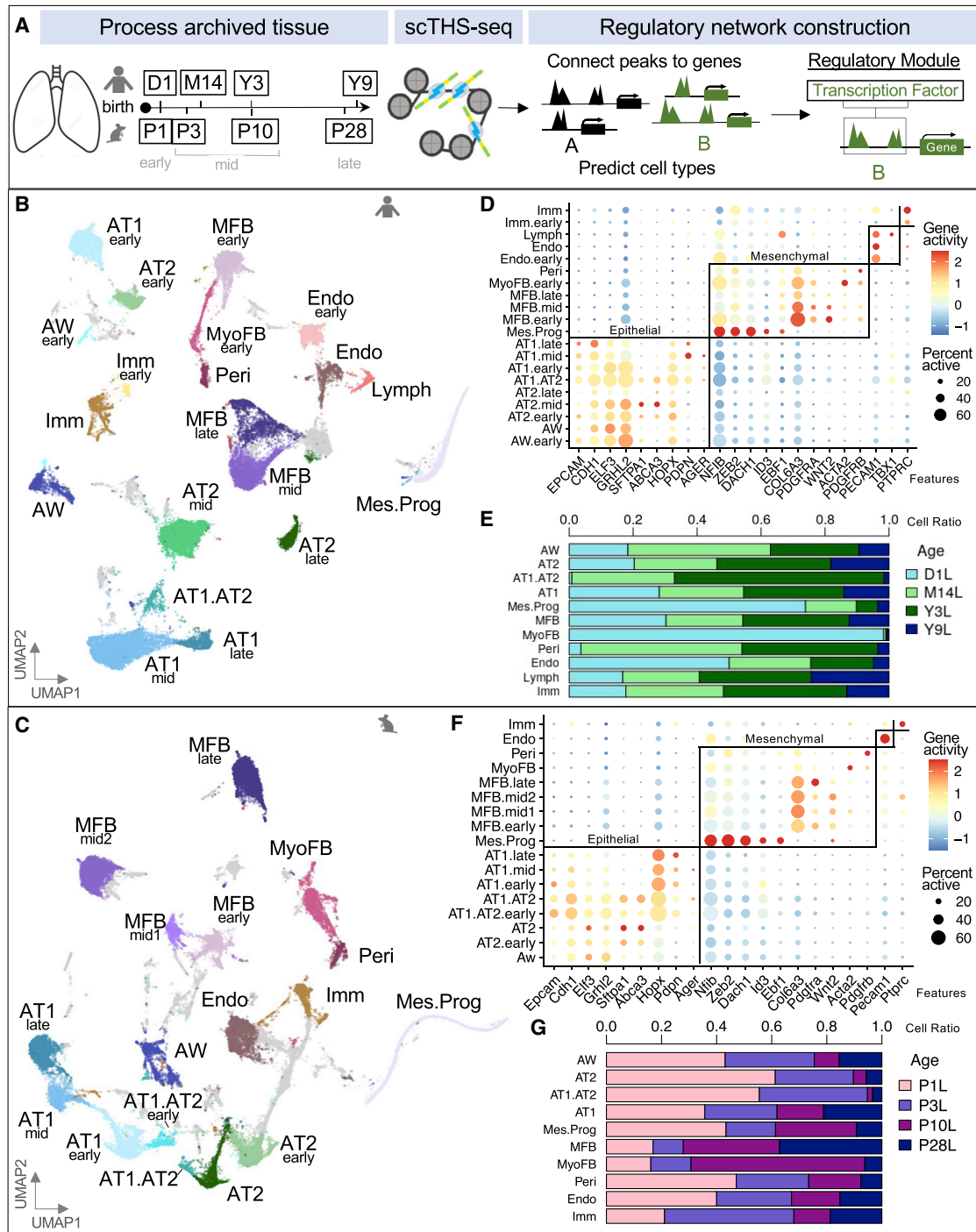
Figure 1. Lung cell types defined by scTHS-seq chromatin profiles

(A) Overview of workflow with single-cell isolation from archived lung tissue, application of scTHS-seq, and downstream de-novo cell-type identification and construction of regulatory modules. Human and mouse icons indicate corresponding species. For age alignment, a post-weaning, pre-pubertal human to mouse estimation of 1 human year equals 3.65 mouse days was used after postnatal day 1.[15]

(B and C) UMAP of human (40,427 cells) and mouse (41,476 cells) cell types predicted using peaks in promoter regions of marker genes and Cicero gene activity scores. Based on the majority of cells from different donor ages in each cluster, cell types were further annotated as early (day 1), mid (2 middle time points), or late (oldest time point). Each dot represents a cell and gray clusters represent un-classified populations.

(legend continued on next page)

regulatory modules, and locate pathogenic regulatory units enriched in lung disease causal variants. Delving into human day of life 1, we set out to better understand how the lungs respond to breathing air and uncovered regulatory mechanisms essential to alveolar septation and functionally relevant to ex-utero breathing. We then tracked these regulatory signals across development and species for comparison. This is the first study to characterize open chromatin in single cells from human neonatal and pediatric "healthy" lung tissue during the most critical period of alveolarization and demonstrates the strength of chromatin accessibility profiles to expose cell-type-specific regulation of key developmental transitions.

RESULTS

Construction of a sc open chromatin atlas profiling postnatal human and mouse lung development

We applied scTHS-seq to 4 human donor lungs collected on day 1 (D1) of life, 14 months (M14), 3 years (Y3), and 9 years (Y9) of age with history of birth after full-term gestation and no known lung disease from the Biorepository for Investigation of Neonatal Disease of the Lung (BRINDL), plus age-correlated wildtype C57 Bl/6 mouse lungs harvested at postnatal days 1, 3, 10, and 28 (P1, P3, P10, and P28) (Figure 1A; Table S1A). An optimized nuclear isolation protocol was used to process flash frozen human lung biopsies (STAR Methods). To minimize potential batch effects, all 8 samples were processed simultaneously in one experiment using combinatorial indexing and library generation as previously described.[14] For each species dataset, only cells with >1,500 unique reads were utilized resulting in a median of 7,542 and 7,159 unique mapped reads per cell for human and mouse, respectively. After additional quality filtering, our human lung dataset is composed of 40,427 single-cell profiles with 427,057 accessible peaks and 41,476 single cells with 345,486 accessible peaks for mouse.

scTHS-seq defined expected and newly described lung cell types

Applying our computational pipeline (STAR Methods), we identified 20 epigenetically distinct human clusters composed of 11 major cell types (Figure 1B) and 17 clusters with 10 major cell types for mouse (Figure 1C). Initially, major cell types were identified based on Cicero-linked gene activity scores,[16] which connect open regulatory regions to genes and serve as an estimation of gene expression, and corresponding peaks in promoter regions of canonical markers (epithelial – EPCAM, CDH1; mesenchymal – COL6A3, PDGFRA; endothelial – PECAM1; immune – PTPRC) (Figure S1). To further classify and predict cell types within compartments, we ran a likelihood-ratio test using gene activity to identify top differential marker genes for each cluster (log fold difference > 0.09; STAR Methods; Figures 1D and 1F). Annotated clusters represented expected cell types, except for a cluster we labeled as mesenchymal progenitors (Mes.Prog, NFIB+, ZEB2+, DACH1+, ID3+, EBF1+) in both human and mouse. Epithe-

lial compartments consisted of airway (AW, ELF3+, GRHL2+), alveolar type 2 (AT2, SFTPA1+, ABCA3+), alveolar type I (AT1, HOPX+, PDPN+, AGER+), and clusters with both AT1 and AT2 signatures labeled as AT1.AT2. Matrix fibroblasts (MFB, COL6A3+, PDGFRA+, WNT2+), myofibroblasts (MyoFB, ACTA2+), and pericytes (Peri, PDGRB+) were identified in the stromal populations. In human endothelial clusters (Endo, PECAM1+), we also identified lymphatic cells (Lymph, TBX1+) that did not clearly separate out in mouse at this level of resolution. Small immune (Imm, PTPRC+) clusters were present in both species.

To substantiate our cell type predictions, we correlated cluster gene activity markers with available human and mouse sorted bulk RNA-seq on the LungMAP Lung Gene Expression Analysis (LGEA) web portal (Figures S1C and S1D), published scRNA data (Figure S2),[17–21] and independently merged and analyzed LGEA human pediatric (LungMAP_humanD1, M21, Y9, Y24) scRNA-seq datasets (STAR Methods; Figure S4).[22–24] Correlation with bulk and sc gene expression supported all cell types we identified by scTHS-seq, except for Mes.prog clusters, which were only appreciated by chromatin accessibility profiling.

The mesenchymal progenitors we identified in both human and mouse datasets share high gene activity in similar genes (NFIB, ZEB2, DACH1, ID3, and EBF1) that have been connected to mesenchymal cells and stemness. Recently published work identifies a novel mouse fibroblast population expressing Ebf1.[25] Xie et al. also describes a mesenchymal progenitor (MP) population seen in adult mice similar to LGEA embryonic day 16.5 (E16.5) proliferative mesenchymal progenitors (PMPs).[26] We identified 182 human and 499 mouse differentially active (DA) genes (p value < 0.05) for our Mes.Prog clusters. Comparing differentially expressed (DE) genes in adult mouse MP and E16.5 PMPs, overlapping gene signatures were present (152/182 human and 415/499 mouse Mes.Prog DA genes in MP DE genes, 6/182 human and 13/499 mouse Mes.Prog DA genes in PMP DE genes) (Table S3). Overall, by combining the presence of peaks in promoter regions of marker genes and gene activity profiles, we classified expected cell types and a mesenchymal progenitor state in humans and mice.

Lung cellular diversity at human day 1 of life

Despite humans and mice being born with lungs at different stages of development (alveolar versus saccular), both species are capable of successful gas exchange ex-utero. To better understand cell types and transition states involved in these biological pathways, we further analyzed epithelial and mesenchymal clusters at human day of life 1 and connected active genes to developmentally relevant biological processes and disease.

Iterative clustering of D1 cells improved cell-type resolution for downstream regulatory analysis

We gathered all D1 cells in Figure 1B that belonged to an annotated cell type and performed iterative clustering to generate a higher resolution UMAP embedding (Figure S3A). Differential gene activity profiles were used to predict cell types (Figure S3B).

(D and F) Dot plots of log-normalized gene activity scores for select genes in each cell type. The size of each dot corresponds to the percent of cells in that cell type with non-zero gene activity while the color corresponds to the average gene activity.

(E and G) Cell-type contribution from human and mouse lung developmental time point captured as a ratio of total cells.

See also Figures S1 and S2 and Tables S1 and S2.

Independent analysis of corresponding LGEA scRNA-seq for D1 was also completed (Figures S4D and S4E). To provide insight on cell-type-specific regulation, we applied ChromVAR to identify differential TF activities for each cluster (STAR Methods; Figure S3C).[27,28] Several of the top TFs for each cell type aligned with previous descriptions such as regulatory factor X (*RFX*) for ciliated cells,[29,30] NK2 homeobox 1 (*NKX2-1*) for alveolar epithelial cells,[31,32] serum response factor (*SRF*) for myofibroblasts,[33] Ets variant 2 (*ETV2*) for endothelial cells,[34] and Spi-1 proto-oncogene (*SPI1*) for myeloid and lymphoid lineages,[35] along with previously undescribed cell-type-specific TFs. The same analysis completed for mouse P1 also led to improved cell-type resolution (Figure S5).

Characterization of distal alveolar progenitor and epithelial populations

We identified two epithelial progenitor states at D1: Alv.Prog and AT1.AT2. *AXIN2*+ Alv.Prog cells contained regulatory elements related to *NAV2*, which plays a role in cellular growth and migration, and *FZD2*, a receptor for bronchoalveolar stem cells with a role in stem cell maintenance and differentiation (Figure S3B).[36,37] *Fzd2* expression is seen in bipotent progenitor cells from mouse distal epithelium at E18.5.[38] Both *Axin2* and *Fzd2* can inhibit the WNT pathway promoting a progenitor-like state in lung epithelial cells and preventing differentiation.[36,39,40] AT1.AT2 cells, active at both AT1 and AT2 genes (*PDPN* and *MUC1*) and signature genes (*EGFR* and *SHH*) (Figure S3B), are defined as intermediates or bipotent progenitors of AT1 and AT2 cells at E18.5 and P1.[20,38,40,41] Our analysis additionally demonstrated notable high activity of AT1.AT2 cells near *LIMD1*, *KRT8*, *KRT19*, and *CLDN4* (Figures 2A and 2B). Expression of *LIMD1* is potentially relevant to the lungs' quick adaptation to air-inflated breathing and cyclic stretch at birth as *Limd1* is part of a tension-dependent pathway that limits Hippo activity where cell proliferation is needed.[42] Immunofluorescent staining demonstrates *LIMD1* co-staining with *SFTPC* and *AGER* in AT1.AT2 cells at day 1 in three donors, including donor 1 for which scTHS-seq data was collected (Figure 2C). Fewer triple-positive cells were present in Y3 donors and none were seen in the adult donor (Figures 2C and S3D). In Wang et al. snRNA-seq datasets, *LIMD1* gene expression is seen in AT1/AT2-like cells at D1 in ~30-week gestational age and ~3 year donors (Figure 2D). *Limd1* gene activity was also present in mouse AT1.AT2 cells during alveolar septation starting at P3 and peaking at P10 (Figure S5D). By zooming in on the epithelial compartment at D1, we uncovered a *LIMD1*+AT1.AT2 cell state linked to a biomechanical pathway possibly stimulating AT1 cell differentiation.

Dynamic AT1 cell regulation

AT1 cells separated into two subpopulations (AT1_1 and AT1_2) (Figure 2A); genomic regions enrichment of annotations tool (GREAT)[43] analysis suggested a growth state for AT1_1 (morphogenesis and insulin receptor signaling in epithelial cell differentiation) while AT1_2 cells demonstrated a more mature profile (membrane organization) (Figure 2E). AT1 clusters were active near canonical marker *PDPN* along with more contemporary markers: *RTKN2*, *VEGFA*, *GPRC5A*, and *MYRF* (Figure 2F).[5,18,38,44] Both AT1 cell states were significantly linked to establishment of cellular polarity and regulation of *PDGFR* signaling pathway. Apical-basal polarity plays a role in cell fate decisions and the posi-

tioning of lung epithelial cells.[45] AT1_2 negative regulation of *EGFR/ERBB* signaling suggested influence on the directional differentiation of AT1.AT2 cells as *EGFR/ERBB* signaling can induce AT2 renewal.[20,41] AT1_2 cell to cell junction assembly and extracellular matrix (ECM) actin interactions functionally reinforce the alveolar epithelial barrier.[46,47]

Construction of AT1 regulatory modules

To ascertain mechanisms controlling AT1 differentiation from AT1.AT2 cells and AT1 proliferation after birth, we constructed regulatory modules by aligning cell-type-specific TFs and TF motifs enriched in active peaks (regulatory regions linked to expressed marker genes) (Figure 3A). First, examination of differential ChromVAR TF activities for human D1 AT1.AT2 and AT1 clusters identified *KLF*, *NKX2*, and *TEAD* as top lineage-defining TF candidates (Figures 3B and S3C). Expression of these TFs is seen in premature (~30 week gestation age) human donors at birth (Figure 3C).[21] Next, we located TF motifs that are overrepresented in differentially accessible (DA) regions linked to cell-type-defining genes (STAR Methods; Table S4). Motif enrichment analysis suggested TF binding to gene-specific regulatory elements. Directional regulation (activation or repression) of target gene expression was inferred based on sc or single-nucleus (sn) RNA-seq from similarly aged BRINDL donors (Figure 3E).[21–24] In Figure 3D, we examined accessible regions linked to DA genes enriched in *KLF4*, *NKX2*, and *TEAD* TF motifs. Regulatory peaks for AT1.AT2 genes *LIMD1*, *KRT8*, and *EGFR* were enriched in *KLF4* motifs. *NKX2* motif enrichment was present in AT1 markers (*GPRC5A*, *CLDN18*, and *LMO7*), apicobasal polarity genes (*LAMA3*, *PATJ*, and *PRKCZ*), general epithelial marker *KRT7*, and *TEAD1*. AT1_2 DA genes *WWC2*, *LMO7*, and *MYRF* were enriched in TEAD motifs. By aligning cell-type-specific TFs and enrichment of TF binding motifs in active peaks of cell-type marker genes, we constructed regulatory modules controlling AT1 differentiation at D1 (Figure 3F).

NKX2 TFs likely play a role in human AT1 cell differentiation, which is supported by recent efforts in mice demonstrating that *Nkx2-1* is necessary for the initiation, development, and maintenance of AT1 cells.[31,48] AT1_1 cells contained open *TEAD1* peaks enriched in *NKX2* motifs, indicating that *NKX2* binding can trigger expression of TF *TEAD1* (Figure 3D). *TEAD1* is expressed by AT1 cells at LGEA D1 (Figure 3E). Temporal *TEAD1* accessibility and expression likely allows for *TEAD* binding to AT1 accessible elements, such as *MYRF*. *MYRF* is a novel, potential master regulator of adult human and mouse AT1 cells.[18] Mouse AT1 cells demonstrated similar *Nkx2* and *Tead* TF activity at P1 (Figure S5C). By assembling regulatory modules using scTHS-seq and sc/sn-gene expression, we formed new developmental mechanism hypotheses involving differentiation of *KLF4*-directed *LIMD1*+*KRT8*+ AT1 precursors and *TEAD*-directed AT1 cell maturation as well as supporting *NKX2*-directed differentiation of alveolar epithelial cells.

TF regulation of AT1 cells across time points and species

Evaluation of human AT1 clusters from all time points revealed that AT1.late cells displayed divergent TF activity compared to earlier time points; this change is not seen in later mouse AT1 cells (Figures S6A and S6B). *TEAD* TF activity was again more specific for AT1 clusters in both species. Distinctive to human AT1.late clusters, significant differential TF activity was seen involving Jun
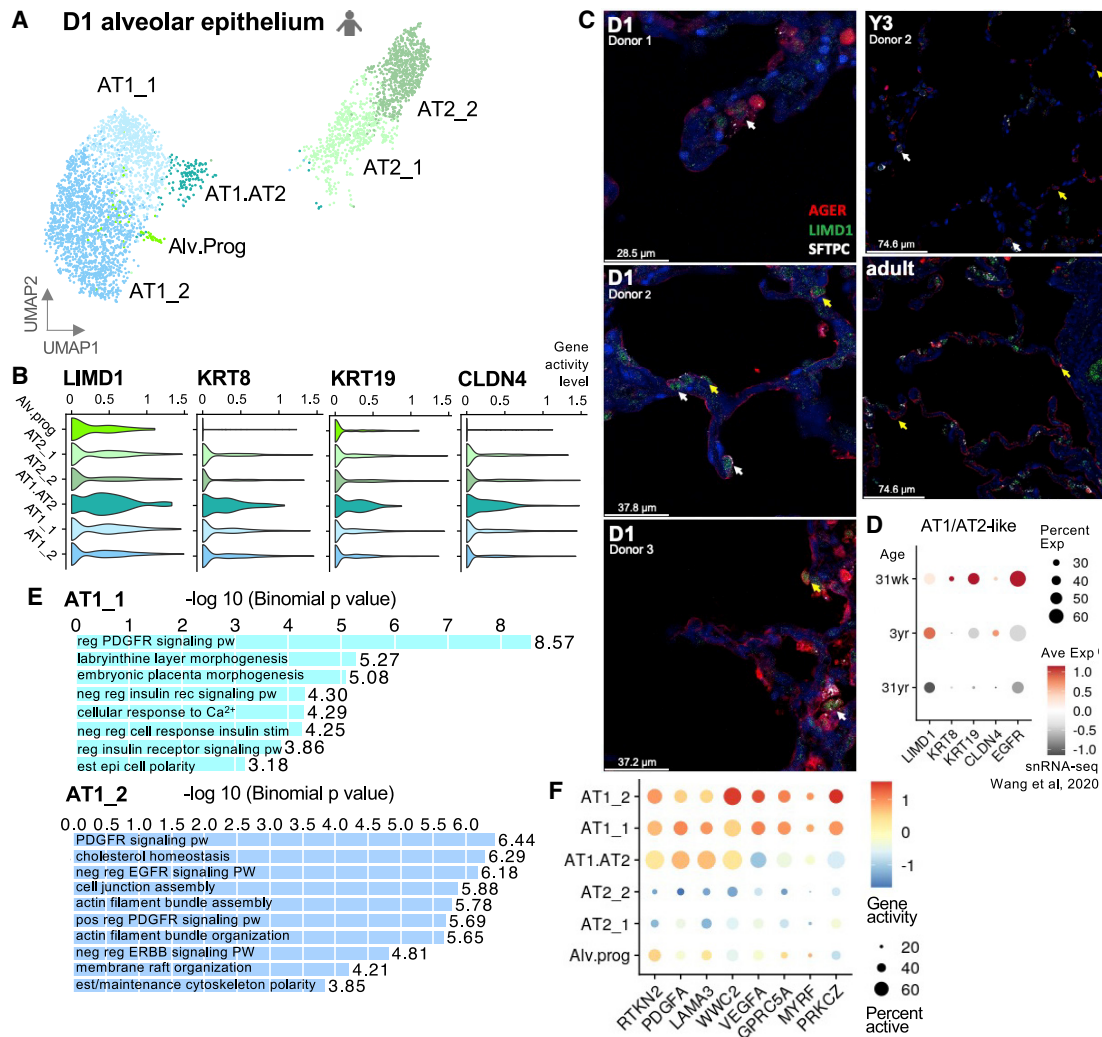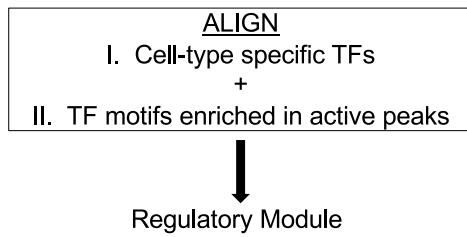
Figure 2. Characterizing D1 alveolar epithelium cell states
(A) UMAP of annotated D1 alveolar epithelial clusters.
(B) Violin plots of gene activity levels for notable AT1.AT2 cell differentially accessible genes.
(C) Immunofluorescent *LIMD1, AGER*, and *SFTPC* staining in alveolar region of human in D1, Y3, and adult lung from different donors. White arrows indicate triple positive AT1.AT2 cells. Yellow arrows indicate *AGER* and *LIMD1* in AT1 cells. scTHS-seq data collected from donor 1 in each age group except for adult sample. Scale bars units listed in bottom left corner for each image. See also Figure S3D.
(D) Dot plot of AT1.AT2 marker gene expression in AT1/AT2-like cells from Wang et al. snRNA-seq grouped by donor age, ∼30 weeks gestational age and 3 and 31 years old.
(E) GREAT significant GO biological processes for D1 AT1 clusters for top 2k differential peaks based on -log₁₀(binomal p value)[43].
(F) Dot plot of gene activity scores for AT1 marker genes in alveolar epithelial clusters.
See also Figure S3 and Table S4.

proto-oncogene (*JUN*), Fos proto-oncogene (*FOS*), and nuclear factor erythroid 2 (*NFE2*). *Jun* plays a critical role in maintaining lung alveolar homeostasis and the loss of *Jun* contributes to lung inflammation and emphysema.[49,50] *NFE2* is linked to lung host response and cytoprotection against oxidative stress.[51] Verification of inferred AT1 regulation across human time points was achieved by confirming the availability of appropriate TFs and corresponding available binding sites in AT1 populations. In Figure S6C, the temporal availability of TFs was established using TF gene activity scores. Alignment of regulatory regions in
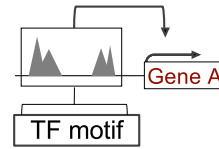
AT1 clusters with binding sites for *NKX2-1, TEAD1,* and *JUN* was examined near transcription start sites (TSS) of AT1 marker genes, *GPRC5A* and *MYRF*, using curated ChIP-seq analysis from ReMAP 2018 for various human cell lines (Figure S6C).[52] A distal *GPRC5A* element only contained accessible peaks for early and mid AT1 cells and ChIP-seq peaks for *NKX2-1* and *TEAD1*. In the AT1.late cluster, a dominant accessible element overlapped with abundant *JUN* ChIP-seq peaks. TF analysis across developmental time points demonstrated a synchronized pattern of TF accessibility and binding to direct AT1 cells.
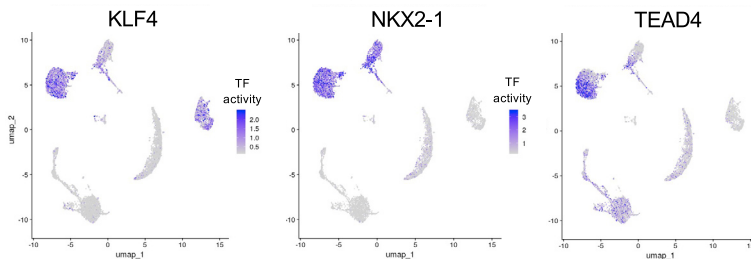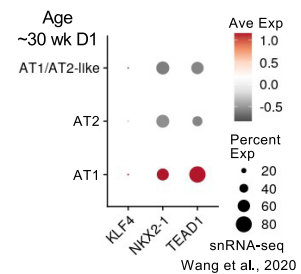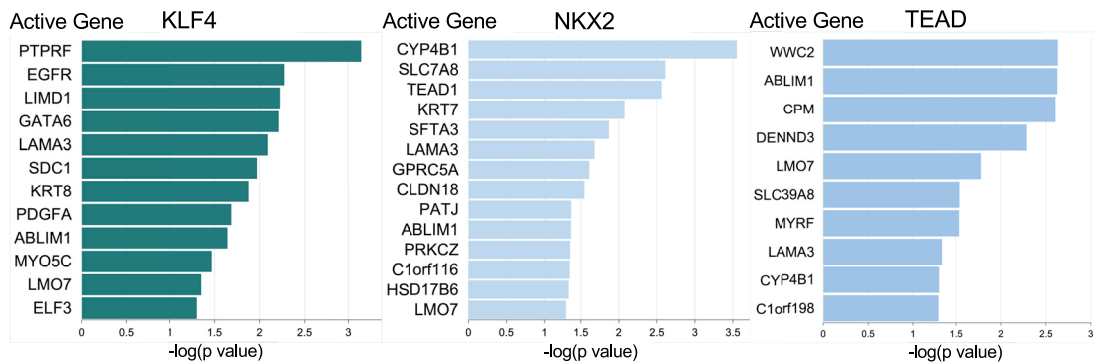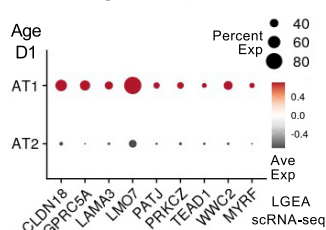
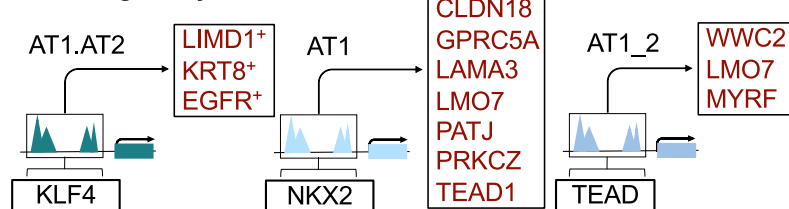**Figure 3. Identifying D1 regulatory modules driving AT1 differentiation during alveolar septation**

(A) Schematic for regulatory module construction involving I. identification of candidate TFs and II. motif enrichment analysis of peaks linked to active cell-type marker genes.

(B) UMAP from Figure S3A re-colored with TF activity levels for key alveolar epithelial TFs.

(C) Dot plot of TF gene expression regulating AT1 cell differentiation in Wang et al. snRNA-seq by cell type in ~30 week gestational age 1 day old (~30 week D1) donors.

(D) Bar plots of -$\log_{10}$(p values) for select active genes (> 1 peak linked to gene and SE in AT1.AT2 or AT1 cells) with *KLF4, NKX2,* and *TEAD* motif enrichment.

(E) Dot plot of active AT1 gene expression in LGEA D1 scRNA-seq dataset.

(F) Proposed regulatory modules for D1 AT1.AT2 and AT1 cells. All marker genes in rectangles contained peaks enriched in *KLF4, NKX2,* or *TEAD* motifs. Genes listed in red are expressed in LGEA D1 AT1 cells.

[+]Gene expression present in AT1/AT2-like cells. See Figure 2D.

See also Table S4.

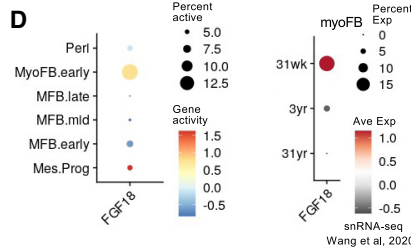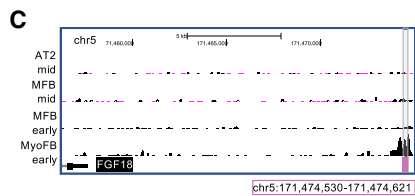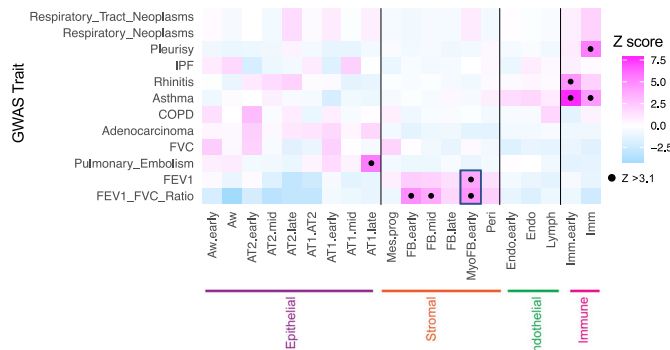## A   LGEA ACTA2 immunostaining
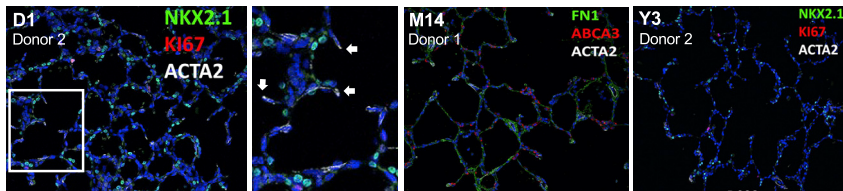
B   Causal variant enrichment

Figure 4. Alveolar myoFB regulatory units enriched in lung function causal variants

(A) Protein fluorescent immunostaining for *NKX2-1, KI67, ACTA2, FN1,* and *ABCA3* in alveolar parenchyma of human lungs at D1, M14, and Y3 from LGEA Lung Imaging provided courtesy of Dr. Jeffrey Whitsett. Arrows indicate *ACTA2*[+] staining at septal ridges. scTHS-seq data collected from donor 1 in each age group.

(B) Prediction of cell type targets in disease. Using gChromVAR, enrichment of GWAS causal variants for various lung traits and disease in scTHS-seq human lung clusters. Black dot represents significant enrichment with Z-score > 3.1.

(C) Genome browser plot displaying unique MyoFB.early regulatory region overlapping identified high causal probability lung function risk unit and nearest gene, *FGF18*.

(D) Dotplot of *FGF18* gene activity in scTHS-seq human stromal cells and gene expression in myoFBs from Wang et al. snRNA-seq.

See also Table S5.

C

D

Myofibroblasts and regulation of epithelial-mesenchymal interactions critical to alveolar septation

As seen in Figures 1E and 1G, 98% (949/971) of human myoFBs captured across all time points were from D1; however 63% (1,488/2,358) of mouse myoFBs were from P10. A substantial increase in myoFB contribution from P10 (compared to ∼15% from P1 and P3) likely reflects the surge in myoFBs needed during the peak of murine alveolar septation (P4 to P12). Subsequent myoFBs reduction to 6% (146/2,358) at P28 is consistent with prior lineage tracing[2,53] and scRNA-seq studies.[54] In humans, alveolar septation is estimated to occur from 36 weeks gestation to 3 years. *ACTA2*[+] myoFBs captured by scTHS-seq and LGEA protein fluorescent immunostaining from BRINDL healthy lung donors of similar age corroborated the presence of myoFBs at septal ridges at D1 with fewer detected at this location at M14 and Y3 (Figure 4A).[24] The postnatal cellular composition captured by scTHS-seq using our tissue processing protocol suggested a similar increase and decrease of myoFBs during human and murine alveolar septation.

Alveolar myoFB regulatory units enriched in lung function causal variants

By combining causal variants for lung traits and our scTHS-seq data, we traced pathogenic risk to specific developmental lung cell types and regulatory units. Fine-mapped causal variants for lung diseases were obtained from CAUSALdb, a comprehen-

sive database of uniformly processed GWAS summary statistics,[55] and from a large GWAS study on lung function in the UK Biobank and SpiroMeta Consortium.[56] We used two traits, FEV1 (forced expiratory volume in 1 second) and FEV1/FVC ratio (FVC, forced vital capacity), both markers of obstructive lung disease, to assess lung function. In Figure 4B, we applied g-chromVAR[57] to calculate the enrichment of causal variants in regulatory regions of the 20 characterized human cell types from Figure 1B. Early myoFBs were enriched in causal variants linked to FEV1 and FEV1/FVC (Z score > 3.1). Furthermore, evaluation of lung function single nucleotide polymorphisms (SNPs) overlapping with accessible peaks revealed a common regulatory unit in chromosome 5 (chr5:171,474,530–171,474,621) for FEV1 and FEV1/FVC in MyoFB.early with causal probability (Figure 4C and Table S5). The closest gene to this lung function regulatory risk unit was fibroblast growth factor 18 (*FGF18*) located ∼15 kb downstream. *Fgf18* labels myoFBs during alveolar septation,[53] and adult small airway myoFBs are associated with decreased lung function in chronic obstructive pulmonary disease.[58] MyoFBs expression of *FGF18* in ∼30-week human lung donors corresponded to active *FGF18* regulatory regions (Figure 4D). Enrichment analysis of lung disease causal variants revealed a genetic risk association for lung function impairment in specific regulatory units of early alveolar myoFBs.

Construction of alveolar epithelial signaling and myofibroblast regulatory modules

To further explore the role of myoFBs during alveolar septation, we identified two mesenchymal clusters with myoFB characteristics at D1 (myoFB_1, myoFB_2) (Figures 5A ands S3A). The myoFB_1 cluster demonstrated highest gene activity for *PDGFRA* and myoFB_2 for *ACTA2* (Figure 5B). In mice as
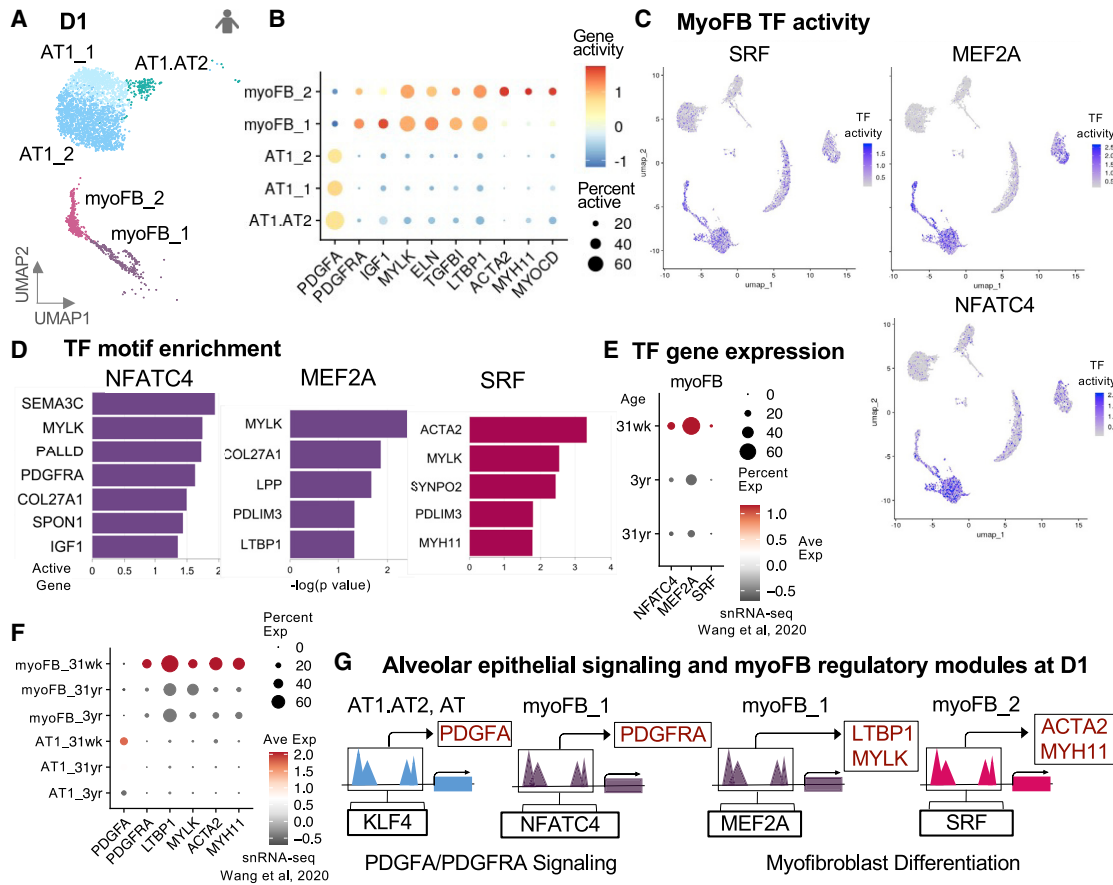
Figure 5. Exploring myofibroblast states at D1 and epithelial-mesenchymal regulatory modules guiding alveolar septation

(A) UMAP of annotated D1 AT1.AT2, AT1 and myoFB clusters.

(B) Dot plot of log-normalized gene activity scores for *PDGFA* and select myoFB DA genes.

(C) UMAP from Figure S3A re-colored with TF activity levels for key myoFB TFs.

(D) Bar plots of -log$_{10}$(p values) for select active genes in myoFB_1 or myoFB_2 cells with significant *NFATC, MEF2*, and *SRF* motif enrichment.

(E) Dot plot of TF gene expression in Wang et al. snRNA-seq by myoFBs grouped by BRINDL donor age, ~30 weeks gestational age, 3, and 31 years old.

(F) Dot plot of active gene expression in Wang et al. snRNA-seq grouped by donor age.

(G) Proposed regulatory modules for D1 *PDGFA:PDGFRA* signaling and myoFB differentiation. All marker genes in rectangles contained peaks enriched in TF motifs and are expressed in AT1 or myoFB clusters in Wang et al. sn RNA-seq at ~30 week D1.

See also Table S4.

alveologenesis begins, *Pdgfra*⁺ progenitor cells give rise to *Acta2*⁺ alveolar myoFBs.[2,7,59,60] *PDGFRA*⁺ human myoFB_1 is likely a myoFB progenitor or intermediate, which has been described previously.[61–64] Both myoFB clusters contained activity associated with myosin light chain kinase (*MYLK*), elastin (*ELN*), transforming growth factor beta induced (*TGFBI*), and latent transforming growth factor beta binding protein 1 (*LTBP1*) genes (Figure 5B). MyoFB_2 displayed additional myosin heavy chain 11 (*MYH11*) and myocardin (*MYOCD*) gene activity along with *SRF* TF activity (Figures 5C and S3C). MyoFB_1 was accessible near *IGF1* which stimulates the proliferation and differentiation of lung fibroblasts into myoFBs.[65] In the myoFB_1 cluster, differential TF analysis revealed highest activity of myocyte enhancer factor 2A (*MEF2A*) which is involved in muscle development, cell growth, and ECM remodeling (Figures 5C and S3C).[66]

PDGFRA ligand *PDGFA* was accessible in AT1 and AT1.AT2 clusters at D1 (Figure 5B). In murine development, *Pdgfa* mRNA surges at P1 and P7,[9] and sc expression is present in AT1/AT2, AT1, and AT2 cells at E18.5.[23,38] Evaluation of *PDGFA*- and *PDGFRA*-specific regulatory regions in alveolar epithelial and myoFB cells at D1 revealed significant cell-type-specific TF activity and enrichment of *KLF4* (Figures 3B and 3D) and nuclear factor of activated T cells (*NFATC*) binding motifs (Figures 5C and 5D) respectively. *Nfat* is a critical member in signaling networks governing myoFB differentiation and contraction;[67–69] *Acta2* expression in smooth muscle requires cooperative interaction between *Nfat* and *Srf*.[69] Our analysis suggested that *KLF4* binds to *PDGFA* regulatory regions in AT1.AT2 and AT1 cells activating *PDGFA* expression (Figures 3D, 5F, and 5G). In the mesenchyme, *NFATC* modulates *PDGFRA* expression in myoFB progenitors. Motif enrichment
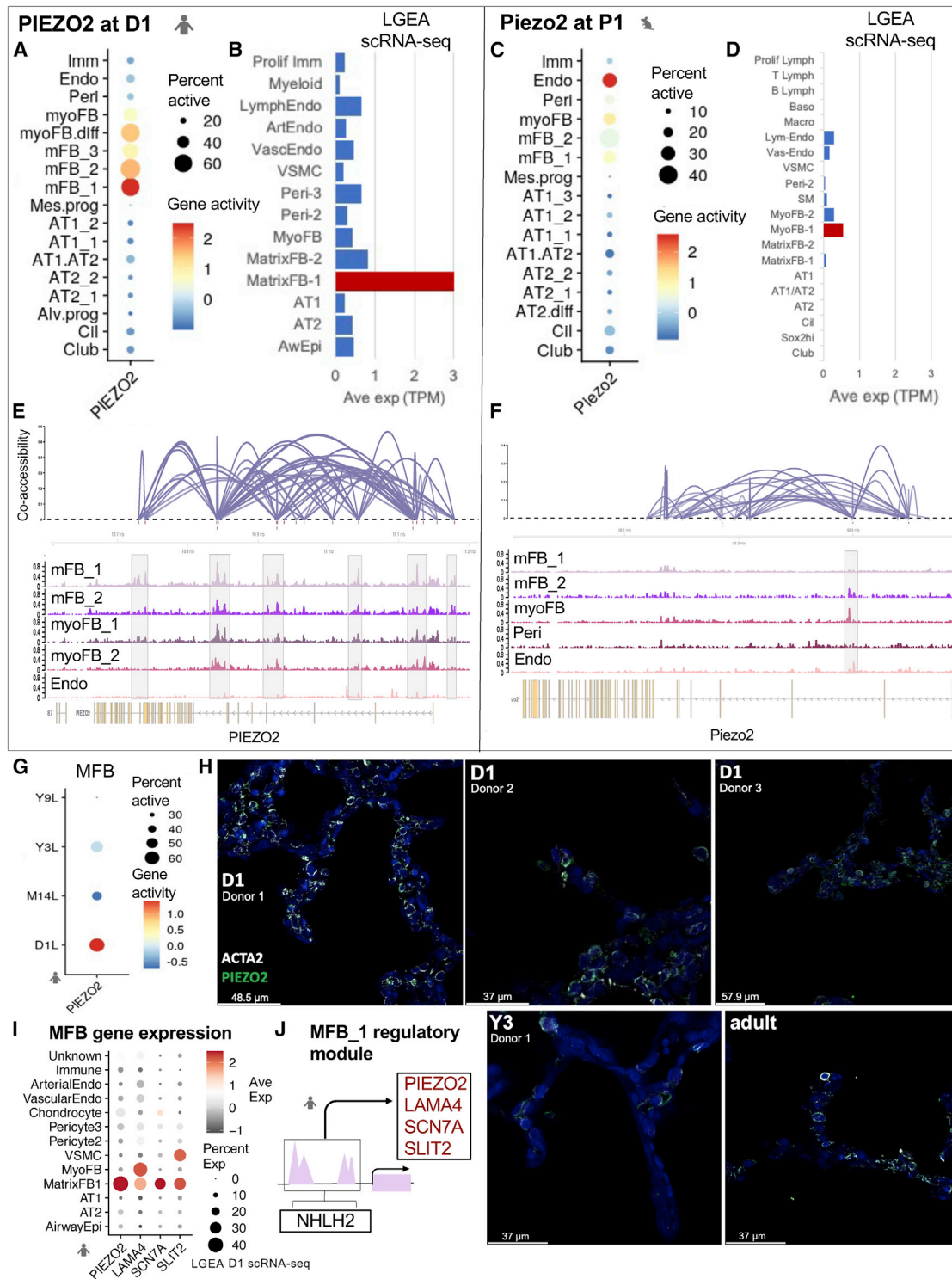
Figure 6. Divergent *PIEZO2* mechanotransduction signaling in fibroblasts at postnatal day 1 and across development

(A–D) *PIEZO2* gene activity and expression at D1 and P1. Log-normalized gene activity from all D1 (A) and P1 (C) cell types. Average expression in transcripts per million (TPM) from LGEA scRNA-seq database for D1 (B) and P1 (D) lung.

(E–F) Connectivity plots for human and mouse open chromatin regions linked to *PIEZO2* and genome browser tracks with peaks highlighted for collapsed cell-type bigwigs. Height of connections represent Cicero co-accessibility score between connected peaks.

(G) *PIEZO2* gene activity for all human matrix fibroblasts grouped by age.

(legend continued on next page)

analysis of regulatory regions linked to target genes of the two D1 myoFB states identified *MEF2* motifs overrepresented in *LTBPI* and *MYLK* regulatory elements and *SRF* motifs in *ACTA2* and *MYH11* peaks (Figure 5D), suggesting a regulatory module that first involves *MEF2* activation of *LTBPI* transcription, *TGFB* pathway stimulation, and subsequent *SRF* binding to accessible *ACTA2* regions driving expression and establishment of myoFB cell identity (Figure 5G). By honing in on *PDGFA-PDGFRA* receptor-ligand activity and myofibroblasts in D1 populations, we assembled regulatory modules for epithelial-mesenchymal communication and myofibroblast differentiation which may direct critical alveolar septal ridge formation.

Divergent mechanotransduction signaling in human stromal cells

PIEZO2, a mechanically activated channel that converts physical stress into chemical signals for cellular communication,[70–72] was the most significant DA gene in human D1 MFB_1 (Figure 6A) with LGEA SE in corresponding matrixFB-1 (Figure 6B). In mouse, highest average *Piezo2* gene activity was noted in P1 endothelial cells (Figure 6C), which did not correlate with LGEA P1 sc data (Figure 6D). A connectivity map visually summarizes Cicero co-accessibility scores between distal regulatory elements and promoters connected to *PIEZO2* (co-accessibility cutoff = 0.25) (Figure 6E). For P1, fewer cis-regulatory elements were captured (co-accessibility cutoff = 0.1 indicating lower confidence in connections) (Figure 6F). Following changes across developmental time points, *PIEZO2* accessibility peaked at D1 and waned over time in human (Figure 6G), consistent with LGEA sc and bulk RNA results (Figures S7A and S7B) and protein immunostaining (Figures 6H and S7C). In mouse endothelial and FB populations, the percent of cells containing active regions, gene activity, and bulk RNA expression remained low across time points (Figures S7D and S7E). Human D1 MFBs were characterized by *PIEZO2* accessibility and expression, which was not similarly present in mouse.

To elucidate regulation of *PIEZO2* expression, differential TF activity analysis revealed that a majority of top TFs in MFB_1 at D1 were from a diverse family of basic helix-loop-helix-zipper (bHLH-ZIP) TFs involved in fetal development.[73] Interestingly, the top identified active TF was achaete-scute family bHLH TF 2 (*ASCL2*), a marker gene for rare epithelial, chemosensing tuft or brush cells (Figure S3C).[74] Forty of the 110 TFs belonging to the bHLH TF family had significant activity in MFB_1 (p value < 0.01, ave logFC > 0.69, Table S4B).[75] Motif enrichment analysis of *PIEZO2* regulatory regions did not find significant *ASCL2* enrichment, however, seven of the 40 active bHLH TF motifs in MFB_1 were enriched in *PIEZO2* regulatory elements (p value < 0.05) with one of the seven being *NHLH2*. In Figure S7F, *NHLH2* TF activity decreased with age across all human stroma clusters. MFB_1 markers that demonstrated significant gene activity and

expression with *NHLH2* binding motif enrichment are depicted in Figures 6I, 6J, and S7G. In summary, *PIEZO2* was identified as a novel marker gene for a matrix FB subtype at human D1, its expression is likely regulated by a bHLH TF, and *Piezo2* was not active in mouse matrix FBs across all sampled time points suggesting a species-specific difference in early developmental response pathways to stretch.

DISCUSSION

In order to lay the foundation for a comprehensive sc regulatory atlas of human and mouse postnatal lung development, we produced high-quality data for ~80,000 single cells, characterized cell types present, and uncovered new mechanistic insights driving cell differentiation and cross-compartmental interactions during alveolar septation. Annotated scTHS-seq cell types identified were those expected, but we also describe a large mesenchymal progenitor-like population in both human and mouse that shares the same signature gene activity profile (*NFIB, ZEB2, DACH1, ID3,* and *EBF1*). This is not too surprising, as our understanding of mesenchymal cell types is far from complete. *Nfib* controls late gestation epithelial and mesenchymal cell differentiation.[76] *Zeb2* is involved in maintaining stemness.[77] *Id3* regulates pluripotency of stem cells and is expressed in lung mesenchyme during mouse embryogenesis.[78,79] Two studies identified *Ebf1*+ mesenchymal progenitor populations in mice.[26,80] Liu et al. also identifies a homologous *EBF1*+ mesenchymal population in LGEA human scRNA-seq datasets from a healthy adult and M21 donor.[80] Our Mes.Prog clusters may represent a similar progenitor state. Further studies are warranted to better classify these populations.

Within epithelial and mesenchymal compartments at D1, scTHS-seq allowed characterization of epithelial and mesenchymal subtypes not captured by scRNA-seq in LGEA datasets, such as those seen for AT1 cells and myofibroblasts (Figure S3). AT1.AT2 *KRT8, KRT19*, and *CLDN4* gene activity overlapped with key genes identified in a mouse AT1 transitional cell state after injury;[81–83] however, our neonatal AT1.AT2 cells lacked DNA damage response, senescence, and inflammatory signals. We uncovered regulatory modules critical to alveolar septation including AT1 and myoFB cell differentiation, AT1 to myoFB ligand-receptor interaction, and mechanotransduction signaling. During alveolar epithelial differentiation, gene and TF activity connections to Hippo signaling were noteworthy for intermediate AT1.AT2 and AT1 cells. The Hippo pathway regulates organ size, mechanotransduction, and the cellular transition to air breathing.[84,85] Hippo kinases *LATS1/2* require *Limd1* for tension-dependent adherens junctions localization. High tension and low cell density causes *Limd1*-mediated recruitment and inhibition of *LATS1/2*, which then allows *YAP/TAZ* to complex with *TEAD*.[42] Coinciding gene accessibility and expression of *LIMD1*

(H) Immunofluorescent staining for *PIEZO2* and *ACTA2* in alveolar region of human in D1, Y3, and adult lung from different donors. scTHS-seq data collected from donor 1 in each age group except for adult sample. Scale bars units listed in bottom left corner for each image. See also Figure S7C.

(I) Dot plot of average expression of active genes in LGEA D1 scRNA-seq dataset.

(J) Proposed regulatory module for D1 MFB_1. All marker genes in rectangles contained peaks enriched in *NHLH2* motifs and are expressed in MatrixFB1 cluster in LGEA D1 human.
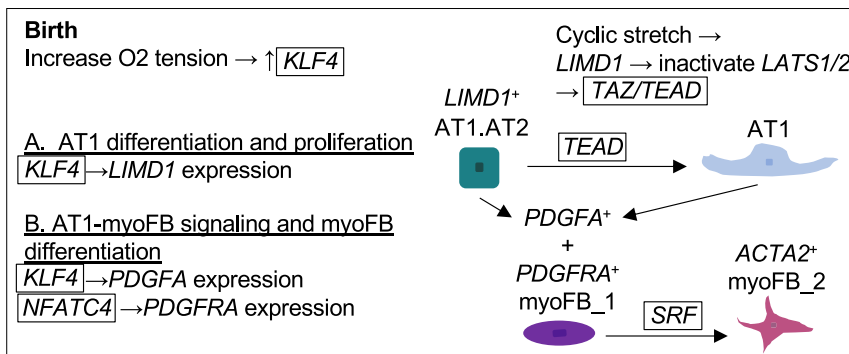
See also Figure S7 and Table S4.

Figure 7. Proposed regulation of AT1 and myoFB cells during alveolar septation at D1
Schematic summary of regulatory mechanism hypotheses after birth of (A) AT1 differentiation and (B) AT1 and myoFB cell-cell communication inferred from analysis of scTHS-seq gene and TF activity, motif enrichment, and available sc/snRNA-seq datasets.

in AT1.AT2 cells (Figure 2) and *TEAD* TF activity in AT1 cells (along with preceding *TEAD* gene expression, and ChIP-seq binding sites in AT1 gene regulatory regions) (Figures 3 and S6) established a connection between physiologic changes at birth from air inflated stretch and stimulation of AT1 cell differentiation and proliferation via inactivation of the Hippo pathway. During manuscript revision, recent publications applying genetic mouse models [48,86,87] and a human 2D primary culture model [88] further support the role of *Yap/Taz* and *Tead* in AT1 cell development.

Despite the essential role of AT1 cells in gas exchange and lung disease, AT1 cell transcriptional regulation remains unclear. We propose a novel stretch-induced pathway signaling human AT1.AT2 cells to differentiate into AT1 cells and *TEAD* as an accompanying TF family involved in AT1 lineage-specific maturation (Figure 7A). Mouse conditional knockout studies reveal that *TAZ* is essential to alveolarization while *YAP* is critical to airway morphogenesis;[89] therefore, inactivation of Hippo signaling may allow *TAZ* to complex with *TEAD*. *LIMD1*-linked differential peaks and reconstructed *TEAD*-gene accessibility modules identified potential cell-type-specific stretch-sensitive gene enhancers, which change with age and require further evaluation (Figure 3). Morever, significant *KLF* TF activity for AT1.AT2 cells and regulatory regions enriched in *KLF4* TF motifs linked to *LIMD1* (Figures 3B and 3D) suggests *KLF4* binding to open enhancers stimulating *LIMD1* expression. *KLF4*, a transcriptional regulator of cell growth arrest and differentiation, is the most significantly altered gene (p < 0.00001) with the greatest fold change at birth in mice.[90] Given the dramatic surge of *Klf4* expression at birth, the relative hyperoxic postnatal environment may serve as a physiologic signal to induce *Klf4* expression. Together, increased oxygen tension at birth may stimulate *KLF4* expression and binding to *LIMD1* enhancer regions, and mechanical stretch from air inflated breathing promotes *LIMD1* downstream inactivation of the Hippo pathway and AT1 cell differentiation (Figure 7A).

AT1 cells likely also play an important role in alveolar septation via interactions with mesenchymal cells. *PDGFA*-expressing AT2 cells are implicated in paracrine signaling to myoFBs to drive alveolar septation.[9,10,91] However, our dataset analyzed in light of sc expression studies suggests a key signaling role for AT1 cells in directing alveolar septation, which aligns with a 3D imaging study showing AT1 cells folding where myoFBs reside in septal ridges.[5] Epithelial-mesenchymal crosstalk via *PDGFA* and *PDGFRA* at D1 appears to involve AT1.AT2, AT1, and differentiating myoFBs extending our transcriptional regulatory hy-

pothesis in Figure 7 to include *KLF4* binding to accessible chromatin in AT1. AT2 and AT1 cells to activate *PDGFA* signaling. In *Klf4* null lungs, connective tissue genes and actin protein expression is absent from septal tip myoFBs.[90] In the mesenchyme, mechanical stretch from air inflation may signal *NFATC4* binding to drive *PDGFRA* expression promoting myoFB differentiation. The coordination of mechano-response and epithelial signaling likely guide the correct positioning of myoFBs for septation.[92] Applying scRNA-seq and *in situ* hybridization, a newly published study supports the spatial interaction of AT1 cells and myoFBs at P3 and P7; this AT1 signaling node is found to be conserved in human lung at 2 months.[54] Zepp at al. also hypothesizes *Klf* and *Tead* TF involvement in AT1 cell development and mesenchymal signaling utilizing scATAC-seq at P3. The group describes *PDGFRA*+ transient secondary crest myofibroblasts (SCMF) as a functionally specialized, force exerting lineage present only during early postnatal alveolar septation which aligns with our results. SCMFs are likely similar to D1 myoFB_1 and may explain the higher proportion of myofibroblasts captured during human and murine alveologenesis. Our finding of discrete lung function regulatory risk units enriched in myoFBs (Figure 4) further stresses the importance of better delineating epithelial-mesenchymal interactions as aberrant myofibroblasts may contribute to abnormal alveolar septation and lung growth with subsequent reduction in lung function.

In addition to evidence of mechanotransduction signals related to alveoli generation, we found a subset of unique human mFBs containing *PIEZO2* accessible regulatory regions that decreased with age. *Piezo2* is characterized as a stretch sensor responsible for lung volume regulation and is necessary to start effective breathing in newborn mice, however no abnormalities are detected in prenatal lung development.[72] In humans, loss-of-function variants resulted in transient respiratory insufficiency at birth.[93,94] Distinct stretch-sensing responses between humans and mice may be attributed to differences in organ absolute size, the magnitude of stretch required for successful gas exchange, and birth and adaptation to air breathing at differing stages of lung development. Regulatory and expression variations along with a milder human phenotype with non-functional *PIEZO2* indicate a species-specific difference and may have critical implications in human respiratory distress at birth, control of breathing, and developmental lung diseases.

Limitations of the study
We captured a substantial number of alveolar epithelial and mesenchymal cells, with a more limited population of immune

cells. Lower numbers of immune cells may be due to tissue preservation strategies and technical differences between whole cell and nuclear protocols, the latter of which has been described previously with snDrop-seq[95] and seen when comparing scTHS-seq to LGEA scRNA-seq from similar donor tissues (Figure S4C). Comparing the proportion of cell types captured from scTHS-seq, LGEA scRNA-seq, and Wang et al., 2020 snRNA-seq, even when utilizing tissue from the same repository, resulted in ratio differences which are not unexpected and likely multifactorial such as from donor tissue quality and clinical characteristics, dissociation protocol, and assay type. Generally, scTHS-seq and snRNA-seq produced more similar cell type proportions.

Our dataset sampled one lung for each time point, which limits our ability to generalize our findings regarding epigenetic variation between developmental stages, especially in humans. In addition to age, epigenetic differences between samples can be attributed to genetics, environmental exposures, donor characteristics, and location of tissue sampling in a large organ. To mitigate these concerns and because of the scarcity of healthy human pediatric tissues, we processed all scTHS-seq data in one experiment to reduce batch effects and substantiated our findings using available bulk RNA-seq, sc/snRNA-seq, CHIP-seq, and protein immunostaining, most of which utilize similarly aged human donors and mouse lungs from the same repository. Table S6 lists our key findings and supporting evidence. Previous published single-cell studies have identified meaningful biological insights with an n = 1,[96,97] nonetheless, developmental changes and species variation presented must be interpreted with caution and require validation such as through characterizing genetic mouse models with lineage-specific reporters or human organoids from purified progenitor cell populations. Moving forward, it will be important to integrate additional regulatory data to complement single-cell transcriptome findings, especially in pediatric and developmental samples, to enhance growing human cell atlas efforts. Employing dual and multi-omic single-cell technologies with a spatial component will further provide a more complete description of cellular states.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects
  - Animals
- METHOD DETAILS
  - Nuclei isolation
  - scTHS-seq library generation
  - Immunofluorescent staining and imaging
- QUANTIFICATION AND STATISTICAL ANALYSIS

  - scTHS-seq data processing
  - Quality filtering and peak calling
  - Dimensionality reduction and clustering
  - Gene activities and cell-type annotation
  - Differential accessible peak and gene analysis
  - Transcription factor analysis
  - Connectivity spider plots and bigwig track visualizations
  - GWAS analysis
  - LGEA scRNA-seq analysis
  - scTHS-seq pseudobulk and bulk LGEA RNA-seq comparison
  - Comparison of scTHS-seq with published data

AUTHOR CONTRIBUTIONS

Conceptualization, T.E.D., J.S.H., and K.Z.; Methodology, T.E.D., B.S., and K.Z.; Software, T.E.D., W.D., S.L., and Y.W.; Formal Analysis, T.E.D., W.D., S.L., and Y.W.; Investigation, T.E.D., B.S., L.R., and G.M.; Resources, T.E.D., J.S.H., and K.Z.; Data Curation, T.E.D., and Y.W.; Writing - Original Draft, T.E.D.; Writing - Review & Editing, J.S.H. and K.Z.; Visualization, T.E.D., W.D., S.L., and Y.W.; Supervision, J.S.H. and K.Z.; Funding Acquisition, T.E.D., J.S.H., and K.Z.

REFERENCES

1. Nikolić, M.Z., Sun, D., and Rawlins, E.L. (2018). Human lung development: recent progress and new challenges. Development *145*, dev163485.

2. Branchfield, K., Li, R., Lungova, V., Verheyden, J.M., McCulley, D., and Sun, X. (2016). A three-dimensional study of alveologenesis in mouse lung. Dev. Biol. *409*, 429–441.

3. Ochs, M., Nyengaard, J.R., Jung, A., Knudsen, L., Voigt, M., Wahlers, T., Richter, J., and Gundersen, H.J.G. (2004). The number of alveoli in the human lung. Am. J. Respir. Crit. Care Med. *169*, 120–124.

4. Haies, D.M., Gil, J., and Weibel, E.R. (1981). Morphometric study of rat lung cells. I. Numerical and dimensional characteristics of parenchymal cell population. Am. Rev. Respir. Dis. *123*, 533–541.

5. Yang, J., Hernandez, B.J., Alanis, D.M., Narvaez, O., Vila-Ellis, L., Akiyama, H., Evans, S.E., Ostrin, E.J., and Chen, J. (2015). Development and plasticity of alveolar type 1 cells. Development *143*, 54–65.

6. Boström, H., Willetts, K., Pekny, M., Levéen, P., Lindahl, P., Hedstrand, H., Pekna, M., Hellström, M., Gebre-Medhin, S., Schalling, M., et al. (1996). PDGF-A signaling is a critical event in lung alveolar myofibroblast development and alveogenesis. Cell *85*, 863–873.

7. Li, R., Li, X., Hagood, J., Zhu, M.-S., and Sun, X. (2020). Myofibroblast contraction is essential for generating and regenerating the gas-exchange surface. J. Clin. Invest. *130*, 2859–2871.

8. Boström, H., Gritli-Linde, A., and Betsholtz, C. (2002). PDGF-A/PDGF alpha-receptor signaling is required for lung growth and the formation of alveoli but not for early lung branching morphogenesis. Dev. Dyn. *223*, 155–162.

9. Gouveia, L., Betsholtz, C., and Andrae, J. (2017). Expression analysis of platelet-derived growth factor receptor alpha and its ligands in the developing mouse lung. Physiol. Rep. *5*, e13092.

10. Gouveia, L., Betsholtz, C., and Andrae, J. (2018). PDGF-A signaling is required for secondary alveolar septation and controls epithelial proliferation in the developing lung. Development *145*, dev161976.

11. Li, J., Wang, Z., Chu, Q., Jiang, K., Li, J., and Tang, N. (2018). The Strength of Mechanical Forces Determines the Differentiation of Alveolar Epithelial Cells. Dev. Cell *44*, 297–312.e5.

12. Schiller, H.B., Montoro, D.T., Simon, L.M., Rawlins, E.L., Meyer, K.B., Strunz, M., Vieira Braga, F.A., Timens, W., Koppelman, G.H., Budinger, G.R.S., et al. (2019). The human lung cell atlas: A high-resolution reference map of the human lung in health and disease. Am. J. Respir. Cell Mol. Biol. *61*, 31–41.

13. Ardini-Poleske, M.E., Clark, R.F., Ansong, C., Carson, J.P., Corley, R.A., Deutsch, G.H., Hagood, J.S., Kaminski, N., Mariani, T.J., Potter, S.S., et al.; LungMAP Consortium (2017). LungMAP: The molecular atlas of lung development program. Am. J. Physiol. Lung Cell. Mol. Physiol. *313*, L733–L740.

14. Lake, B.B., Chen, S., Sos, B.C., Fan, J., Kaeser, G.E., Yung, Y.C., Duong, T.E., Gao, D., Chun, J., Kharchenko, P.V., and Zhang, K. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat. Biotechnol. *36*, 70–80.

15. Dutta, S., and Sengupta, P. (2016). Men and mice: Relating their ages. Life Sci. *152*, 244–248.

16. Pliner, H.A., Packer, J.S., McFaline-Figueroa, J.L., Cusanovich, D.A., Daza, R.M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., et al. (2018). Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. Mol. Cell *71*, 858–871.e8.

17. Vieira Braga, F.A., Kar, G., Berg, M., Carpaij, O.A., Polanski, K., Simon, L.M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J., et al. (2019). A cellular census of human lungs identifies novel cell states in health and in asthma. Nat. Med. *25*, 1153–1163. https://doi.org/10.1038/s41591-019-0468-5.

18. Travaglini, K.J., Nabhan, A.N., Penland, L., Sinha, R., Gillich, A., Sit, R.V., Chang, S., Conley, S.D., Mori, Y., Seita, J., et al. (2020). A molecular cell atlas of the human lung from single-cell RNA sequencing. Nature *587*, 619–625.

19. Cohen, M., Giladi, A., Gorki, A.D., Solodkin, D.G., Zada, M., Hladik, A., Miklosi, A., Salame, T.M., Halpern, K.B., David, E., et al. (2018). Lung Single-Cell Signaling Interaction Map Reveals Basophil Role in Macrophage Imprinting. Cell *175*, 1031–1044.e18.

20. Guo, M., Du, Y., Gokey, J.J., Ray, S., Bell, S.M., Adam, M., Sudha, P., Perl, A.K., Deshmukh, H., Potter, S.S., et al. (2019). Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. Nat. Commun. *10*, 37.

21. Wang, A., Chiou, J., Poirion, O.B., Buchanan, J., Valdez, M.J., Verheyden, J.M., Hou, X., Kudtarkar, P., Narendra, S., Newsome, J.M., et al.; NHLBI LungMap Consortium (2020). Single-cell multiomic profiling of human lungs reveals cell-type-specific and age-dynamic control of SARS-CoV2 host genes. eLife *9*, 1–28.

22. Du, Y., Kitzmiller, J.A., Sridharan, A., Perl, A.K., Bridges, J.P., Misra, R.S., Pryhuber, G.S., Mariani, T.J., Bhattacharya, S., Guo, M., et al. (2017). Lung Gene Expression Analysis (LGEA): an integrative web portal for comprehensive gene expression data analysis in lung development. Thorax *72*, 481–484.

23. Du, Y., Guo, M., Whitsett, J.A., and Xu, Y. (2015). 'LungGENS': a web-based tool for mapping single-cell gene expression in the developing lung. Thorax *70*, 1092–1094.

24. Du, Y., Ouyang, W., Kitzmiller, J.A., Guo, M., Zhao, S., Whitsett, J.A., and Xu, Y. (2021). Lung gene expression analysis web portal version 3: Lung-at-a-glance. Am. J. Respir. Cell Mol. Biol. *64*, 146–149.

25. Liu, X., Rowan, S.C., Liang, J., Yao, C., Huang, G., Deng, N., Xie, T., Wu, D., Wang, Y., Burman, A., et al. (2021). Categorization of lung mesenchymal cells in development and fibrosis. iScience *24*, 102551.

26. Xie, T., Wang, Y., Deng, N., Huang, G., Taghavifar, F., Geng, Y., Liu, N., Kulur, V., Yao, C., Chen, P., et al. (2018). Single-Cell Deconvolution of Fibroblast Heterogeneity in Mouse Pulmonary Fibrosis. Cell Rep. *22*, 3625–3640.

27. Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978.

28. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

29. Didon, L., Zwick, R.K., Chao, I.W., Walters, M.S., Wang, R., Hackett, N.R., and Crystal, R.G. (2013). RFX3 modulation of FOXJ1 regulation of cilia genes in the human airway epithelium. Respir. Res. *14*, 70.

30. Quigley, I.K., and Kintner, C. (2017). Rfx2 Stabilizes Foxj1 Binding at Chromatin Loops to Enable Multiciliated Cell Gene Expression. PLoS Genet. *13*, e1006538.

31. Little, D.R., Gerner-Mauro, K.N., Flodby, P., Crandall, E.D., Borok, Z., Akiyama, H., Kimura, S., Ostrin, E.J., and Chen, J. (2019). Transcriptional control of lung alveolar type 1 cell development and maintenance by NK homeobox 2-1. Proc. Natl. Acad. Sci. USA *116*, 20545–20555.

32. DeFelice, M., Silberschmidt, D., DiLauro, R., Xu, Y., Wert, S.E., Weaver, T.E., Bachurski, C.J., Clark, J.C., and Whitsett, J.A. (2003). TTF-1 phosphorylation is required for peripheral lung morphogenesis, perinatal survival, and tissue-specific gene expression. J. Biol. Chem. *278*, 35574–35583.

33. Sandbo, N., Kregel, S., Taurin, S., Bhorade, S., and Dulin, N.O. (2009). Critical role of serum response factor in pulmonary myofibroblast differentiation induced by TGF-β. Am. J. Respir. Cell Mol. Biol. *41*, 332–338.

34. Casie Chetty, S., Rost, M.S., Enriquez, J.R., Schumacher, J.A., Baltrunaite, K., Rossi, A., Stainier, D.Y.R., and Sumanas, S. (2017). Vegf signaling promotes vascular endothelial differentiation by modulating etv2 expression. Dev. Biol. *424*, 147–161.

35. Pang, S.H.M., de Graaf, C.A., Hilton, D.J., Huntington, N.D., Carotta, S., Wu, L., and Nutt, S.L. (2018). PU.1 is required for the developmental progression of multipotent progenitors to common lymphoid progenitors. Front. Immunol. *9*, 1264.

36. Zhang, Y., Goss, A.M., Cohen, E.D., Kadzik, R., Lepore, J.J., Muthuku-maraswamy, K., Yang, J., DeMayo, F.J., Whitsett, J.A., Parmacek, M.S., and Morrisey, E.E. (2008). A Gata6-Wnt pathway required for epithelial stem cell development and airway regeneration. Nat. Genet. *40*, 862–870.

37. Li, C., Bellusci, S., Borok, Z., and Minoo, P. (2015). Non-canonical WNT signalling in the lung. J. Biochem. *158*, 355–365.

38. Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

39. Frank, D.B., Peng, T., Zepp, J.A., Snitow, M., Vincent, T.L., Penkala, I.J., Cui, Z., Herriges, M.J., Morley, M.P., Zhou, S., et al. (2016). Emergence of a Wave of Wnt Signaling that Regulates Lung Alveologenesis by Controlling Epithelial Self-Renewal and Differentiation. Cell Rep. *17*, 2312–2325.

40. Nabhan, A.N., Brownfield, D.G., Harbury, P.B., Krasnow, M.A., and Desai, T.J. (2018). Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. Science *359*, 1118–1123.

41. Desai, T.J., Brownfield, D.G., and Krasnow, M.A. (2014). Alveolar progenitor and stem cells in lung development, renewal and cancer. Nature *507*, 190–194.

42. Ibar, C., Kirichenko, E., Keepers, B., Enners, E., Fleisch, K., and Irvine, K.D. (2018). Tension-dependent regulation of mammalian Hippo signaling through LIMD1. J. Cell Sci. *131*, jcs214700.

43. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

44. Marconett, C.N., Zhou, B., Sunohara, M., Pouldar, T.M., Wang, H., Liu, Y., Rieger, M.E., Tran, E., Flodby, P., Siegmund, K.D., et al. (2017). Cross-species transcriptome profiling identifies new alveolar epithelial Type i Cell-specific genes. Am. J. Respir. Cell Mol. Biol. *56*, 310–321.

45. Szymaniak, A.D., Mahoney, J.E., Cardoso, W.V., and Varelas, X. (2015). Crumbs3-Mediated Polarity Directs Airway Epithelial Cell Fate through the Hippo Pathway Effector Yap. Dev. Cell *34*, 283–296.

46. Burgstaller, G., Oehrle, B., Gerckens, M., White, E.S., Schiller, H.B., and Eickelberg, O. (2017). The instructive extracellular matrix of the lung: basic composition and alterations in chronic lung disease. Eur. Respir. J. *50*, 1601805.

47. Meiners, S., Lloyd, C., and Chambers, R.C. (2018). Cell-matrix interactions in lung disease and regeneration: ERS Lung Science Conference 2018 report. Eur. Respir. Rev. *27*, 180040.

48. Little, D., Lynch, A., Yan, Y., Akiyama, H., Kimura, S., and Chen, J. (2021). Differential chromatin binding of the lung lineage transcription factor NKX2-1 resolves opposing murine alveolar cell fates in vivo. Nature Communications *12*, 2509.

49. Reddy, N.M., Vegiraju, S., Irving, A., Paun, B.C., Luzina, I.G., Atamas, S.P., Biswal, S., Ana, N.A., Mitzner, W., and Reddy, S.P. (2012). Targeted deletion of Jun/AP-1 in alveolar epithelial cells causes progressive emphysema and worsens cigarette smoke-induced lung inflammation. Am. J. Pathol. *180*, 562–574.

50. Rahman, I., and MacNee, W. (1998). Role of transcription factors in inflammatory lung diseases. Thorax *53*, 601–612.

51. Cho, H.Y., Reddy, S.P., and Kleeberger, S.R. (2006). Nrf2 defends the lung from oxidative stress. Antioxid. Redox Signal. *8*, 76–87.

52. Chèneby, J., Gheorghe, M., Artufel, M., Mathelier, A., and Ballester, B. (2018). ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic Acids Res. *46* (*D1*), D267–D275.

53. Hagan, A.S., Zhang, B., and Ornitz, D.M. (2020). Identification of a FGF18-expressing alveolar myofibroblast that is developmentally cleared during alveologenesis. Development *147*, dev181032.

54. Zepp, J.A., Morley, M.P., Loebel, C., Kremp, M.M., Chaudhry, F.N., Basil, M.C., Leach, J.P., Liberti, D.C., Niethamer, T.K., Ying, Y., et al. (2021).

55. Wang, J., Huang, D., Zhou, Y., Yao, H., Liu, H., Zhai, S., Wu, C., Zheng, Z., Zhao, K., Wang, Z., et al. (2020). CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. Nucleic Acids Res. *48* (*D1*), D807–D816.

56. Shrine, N., Guyatt, A.L., Erzurumluoglu, A.M., Jackson, V.E., Hobbs, B.D., Melbourne, C.A., Batini, C., Fawcett, K.A., Song, K., Sakornsakolpat, P., et al.; Understanding Society Scientific Group (2019). New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat. Genet. *51*, 481–493.

57. Ulirsch, J.C., Lareau, C.A., Bao, E.L., Ludwig, L.S., Guo, M.H., Benner, C., Satpathy, A.T., Kartha, V.K., Salem, R.M., Hirschhorn, J.N., et al. (2019). Interrogation of human hematopoiesis at single-cell and single-variant resolution. Nat. Genet. *51*, 683–693.

58. Eapen, M., Ward, C., Walters, H., Markos, J., Chia, C., Larby, J., Haug, G., Hills, A., Hackett, T.-L., Lu, W., et al. (2018). Myofibroblasts are associated with increase in small airway thickening, epithelial plasticity and decreased lung function in COPD. European Respiratory Journal *52*, OA2122.

59. Ntokou, A., Klein, F., Dontireddy, D., Becker, S., Bellusci, S., Richardson, W.D., Szibor, M., Braun, T., Morty, R.E., Seeger, W., et al. (2015). Characterization of the platelet-derived growth factor receptor-α-positive cell lineage during murine late lung development. Am. J. Physiol. Lung Cell. Mol. Physiol. *309*, L942–L958.

60. McGowan, S.E., Grossmann, R.E., Kimani, P.W., and Holmes, A.J. (2008). Platelet-derived growth factor receptor-alpha-expressing cells localize to the alveolar entry ring and have characteristics of myofibroblasts during pulmonary alveolar septal formation. Anat. Rec. (Hoboken) *291*, 1649–1661.

61. Moiseenko, A., Kheirollahi, V., Chao, C.M., Ahmadvand, N., Quantius, J., Wilhelm, J., Herold, S., Ahlbrecht, K., Morty, R.E., Rizvanov, A.A., et al. (2017). Origin and characterization of alpha smooth muscle actin-positive cells during murine lung development. Stem Cells *35*, 1566–1578.

62. Guo, M., Wang, H., Potter, S.S., Whitsett, J.A., and Xu, Y. (2015). SINCERA: A Pipeline for Single-Cell RNA-Seq Profiling Analysis. PLoS Comput. Biol. *11*, e1004575.

63. Chen, L., Acciani, T., Le Cras, T., Lutzko, C., and Perl, A.K.T. (2012). Dynamic regulation of platelet-derived growth factor receptor α expression in alveolar fibroblasts during realveolarization. Am. J. Respir. Cell Mol. Biol. *47*, 517–527.

64. Zepp, J.A., Zacharias, W.J., Frank, D.B., Cavanaugh, C.A., Zhou, S., Morley, M.P., and Morrisey, E.E. (2017). Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung. Cell *170*, 1134–1148.e10.

65. Wang, Z., Li, W., Guo, Q., Wang, Y., Ma, L., and Zhang, X. (2018). Insulin-Like Growth Factor-1 Signaling in Lung Development and Inflammatory Lung Diseases. Biomed. Res. Int. https://doi.org/10.1155/2018/6057589.

66. Xu, J., Gong, N.L., Bodi, I., Aronow, B.J., Backx, P.H., and Molkentin, J.D. (2006). Myocyte enhancer factors 2A and 2C induce dilated cardiomyopathy in transgenic mice. J. Biol. Chem. *281*, 9152–9162.

67. Stempien-Otero, A., Kim, D.H., and Davis, J. (2016). Molecular networks underlying myofibroblast fate and fibrosis. J. Mol. Cell. Cardiol. *97*, 153–161.

68. Rice, N.A., and Leinwand, L.A. (2003). Skeletal myosin heavy chain function in cultured lung myofibroblasts. J. Cell Biol. *163*, 119–129.

69. Davis, J., Burr, A.R., Davis, G.F., Birnbaumer, L., and Molkentin, J.D. (2012). A TRPC6-dependent pathway for myofibroblast transdifferentiation and wound healing in vivo. Dev. Cell *23*, 705–715.

70. Coste, B., Mathur, J., Schmidt, M., Earley, T.J., Ranade, S., Petrus, M.J., Dubin, A.E., and Patapoutian, A. (2010). Piezo1 and Piezo2 are essential

components of distinct mechanically activated cation channels. Science *330*, 55–60.

71. Zhong, M., Komarova, Y., Rehman, J., and Malik, A.B. (2018). Mechanosensing Piezo channels in tissue homeostasis including their role in lungs. Pulm. Circ. *8*, 2045894018767393.

72. Nonomura, K., Woo, S.H., Chang, R.B., Gillich, A., Qiu, Z., Francisco, A.G., Ranade, S.S., Liberles, S.D., and Patapoutian, A. (2017). Piezo2 senses airway stretch and mediates lung inflation-induced apnoea. Nature *541*, 176–181.

73. Ball, D.W. (2004). Achaete-scute homolog-1 and Notch in lung neuroendocrine development and cancer. Cancer Lett. *204*, 159–169.

74. Montoro, D.T., Haber, A.L., Biton, M., Vinarsky, V., Lin, B., Birket, S.E., Yuan, F., Chen, S., Leung, H.M., Villoria, J., et al. (2018). A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature *560*, 319–324.

75. Yates, B., Braschi, B., Gray, K.A., Seal, R.L., Tweedie, S., and Bruford, E.A. (2017). Genenames.org: the HGNC and VGNC resources in 2017. Nucleic Acids Res. *45 (D1)*, D619–D625.

76. Hsu, Y.C., Osinski, J., Campbell, C.E., Litwack, E.D., Wang, D., Liu, S., Bachurski, C.J., and Gronostajski, R.M. (2011). Mesenchymal nuclear factor I B regulates cell proliferation and epithelial differentiation during lung maturation. Dev. Biol. *354*, 242–252.

77. Brabletz, S., and Brabletz, T. (2010). The ZEB/miR-200 feedback loop–a motor of cellular plasticity in development and cancer? EMBO Rep. *11*, 670–677.

78. Niola, F., Zhao, X., Singh, D., Castano, A., Sullivan, R., Lauria, M., Nam, H.S., Zhuang, Y., Benezra, R., Di Bernardo, D., et al. (2012). Id proteins synchronize stemness and anchorage to the niche of neural stem cells. Nat. Cell Biol. *14*, 477–487.

79. Jen, Y., Manova, K., and Benezra, R. (1996). Expression patterns of Id1, Id2, and Id3 are highly related but distinct from that of Id4 during mouse embryogenesis. Dev. Dyn. *207*, 235–252.

80. Liu, X., Rowan, S., Liang, J., Yao, C., Huang, G., Deng, N., Xie, T., Wu, D., Wang, Y., Burman, A., et al. (2020). Definition and Signatures of Lung Fibroblast Populations in Development and Fibrosis in Mice and Men. https://doi.org/10.1101/2020.07.15.203141.

81. Strunz, M., Simon, L.M., Ansari, M., Kathiriya, J.J., Angelidis, I., Mayr, C.H., Tsidiridis, G., Lange, M., Mattner, L.F., Yee, M., et al. (2020). Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. Nat. Commun. *11*, 3559.

82. Kobayashi, Y., Tata, A., Konkimalla, A., Katsura, H., Lee, R.F., Ou, J., Banovich, N.E., Kropski, J.A., and Tata, P.R. (2020). Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. Nat. Cell Biol. *22*, 934–946.

83. Choi, J., Park, J.E., Tsagkogeorga, G., Yanagita, M., Koo, B.K., Han, N., and Lee, J.H. (2020). Inflammatory Signals Induce AT2 Cell-Derived Damage-Associated Transient Progenitors that Mediate Alveolar Regeneration. Cell Stem Cell *27*, 366–382.e7.

84. Warburton, D. (2020). YAP and TAZ in lung development: The timing is important. Am. J. Respir. Cell Mol. Biol. *62*, 141–142.

85. Nantie, L.B., Young, R.E., Paltzer, W.G., Zhang, Y., Johnson, R.L., Verheyden, J.M., and Sun, X. (2018). *Lats1/2* inactivation reveals Hippo function in alveolar type I cell differentiation during lung transition to air breathing. Development *145*, dev163105.

86. Penkala, I., Liberti, D., Pankin, J., Sivakumar, A., Kremp, M., Jayachandran, S., et al. (2021). Age-dependent alveolar epithelial plasticity orchestrates lung homeostasis and regeneration. Cell Stem Cell *28*, 1775–1789.e5.

87. Gokey, J., Snowball, J., Sridharan, A., Sudha, P., Kitzmiller, J., Xu, Y., et al. (2021). YAP regulates alveolar epithelial cell differentiation and AGER via NFIB/KLF5/NKX2-1. iScience *24*, 102967.

88. Zhou, B., Stueve, T., Mihalakakos, E., Miao, L., Mullen, D., Wang, Y., et al. (2021). Comprehensive epigenomic profiling of human alveolar epithelial differentiation identifies key epigenetic states and transcription factor co-regulatory networks for maintenance of distal lung identity. BMC Genomics *22*, 906.

89. Isago, H., Mitani, A., Mikami, Y., Horie, M., Urushiyama, H., Hamamoto, R., Terasaki, Y., and Nagase, T. (2020). Epithelial expression of YAP and TAZ is sequentially required in lung development. Am. J. Respir. Cell Mol. Biol. *62*, 256–266.

90. Jean, J.C., George, E., Kaestner, K.H., Brown, L.A.S., Spira, A., and Joyce-Brady, M. (2013). Transcription factor Klf4, induced in the lung by oxygen at birth, regulates perinatal fibroblast and myofibroblast differentiation. PLoS One *8*, e54806.

91. Tsao, P.-N., Matsuoka, C., Wei, S.-C., Sato, A., Sato, S., Hasegawa, K., Chen, H.-K., Ling, T.-Y., Mori, M., Cardoso, W.V., and Morimoto, M. (2016). Epithelial Notch signaling regulates lung alveolar morphogenesis and airway epithelial integrity. Proc. Natl. Acad. Sci. USA *113*, 8242–8247.

92. McGowan, S.E., and McCoy, D.M. (2017). Platelet-derived growth factor receptor-α and Ras-related C3 botulinum toxin substrate-1 regulate mechano-responsiveness of lung fibroblasts. Am. J. Physiol. Lung Cell. Mol. Physiol. *313*, L1174–L1187.

93. Chesler, A.T., Szczot, M., Bharucha-Goebel, D., Čeko, M., Donkervoort, S., Laubacher, C., Hayes, L.H., Alter, K., Zampieri, C., Stanley, C., et al. (2016). The role of PIEZO2 in human mechanosensation. N. Engl. J. Med. *375*, 1355–1364.

94. Delle Vedove, A., Storbeck, M., Heller, R., Hölker, I., Hebbar, M., Shukla, A., Magnusson, O., Cirak, S., Girisha, K.M., O'Driscoll, M., et al. (2016). Biallelic Loss of Proprioception-Related PIEZO2 Causes Muscular Atrophy with Perinatal Respiratory Distress, Arthrogryposis, and Scoliosis. Am. J. Hum. Genet. *99*, 1206–1216.

95. Lake, B.B., Chen, S., Hoshi, M., Plongthongkum, N., Salamon, D., Knoten, A., Vijayan, A., Venkatesh, R., Kim, E.H., Gao, D., et al. (2019). A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. Nat. Commun. *10*, 2832.

96. Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.-L., Chen, S., et al. (2016). Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science *352*, 1586–1590.

97. Rai, V., Quang, D.X., Erdos, M.R., Cusanovich, D.A., Daza, R.M., Narisu, N., Zou, L.S., Didion, J.P., Guan, Y., Shendure, J., et al. (2020). Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. Mol. Metab. *32*, 109–121.

98. Kia, A., Gloeckner, C., Osothprarop, T., Gormley, N., Bomati, E., Stephenson, M., Goryshin, I., and He, M.M. (2017). Improved genome sequencing using an engineered transposase. BMC Biotechnol. *17*, 6.

99. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. Gigascience *10*, 1–4.

100. Fang, R., Preissl, S., Li, Y., Hou, X., Lucero, J., Wang, X., Motamedi, A., Shiau, A.K., Zhou, X., Xie, F., et al. (2021). Comprehensive analysis of single cell ATAC-seq data with SnapATAC. Nat. Commun. *12*, 1337.

101. Bravo González-Blas, C., Minnoye, L., Papasokrati, D., Aibar, S., Hulselmans, G., Christiaens, V., Davie, K., Wouters, J., and Aerts, S. (2019). cis-Topic: cis-regulatory topic modeling on single-cell ATAC-seq data. Nat. Methods *16*, 397–400.

102. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44 (W1)*, W160-5.

103. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

104. Cusanovich, D.A., Reddington, J.P., Garfield, D.A., Daza, R.M., Aghamirzaie, D., Marco-Ferreres, R., Pliner, H.A., Christiansen, L., Qiu, X., Steemers, F.J., et al. (2018). The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature 555, 538–542.

105. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat. Methods 18, 1333–1341.

106. Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383.

107. Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and bioconductor. Methods Mol. Biol. 1418, 335–351.

108. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
PIEZO2 Antibody (Rabbit polyclonal)	Invitrogen	Cat#PA5-56894; RRID: AB_2645539
Alpha smooth muscle actin Antibody (Mouse/IgG2a, kappa)	Invitrogen	Cat#14-9760-82; RRID: AB_2572996
Human LIMD1 Antibody (Mouse monoclonal IgG)	R&D Systems	Cat#MAB8494; Clone#919330; Accession#Q9UGP4
Wang et al., 2020 snRNA-seq	Wang et al., 2020	GEO: GSE161382
Human molecular cell atlas	Travaglini et al., 2020	EGA accession: EGAS00001004344
Alexa Fluor Plus 555 (Donkey anti-Goat IgG (H+L) Highly Cross-Adsorbed Secondary Antibody)	Invitrogen	Cat#A32816; RRID: AB_2762839
DyLight 755 (Donkey anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody)	Invitrogen	Cat#SA5-10043; RRID: AB_2556623
Alexa Fluor 647 (Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody)	Invitrogen	Cat#A32787; RRID: AB_2762830
Biological Samples		
Frozen human lung biopsies and fresh frozen paraffin-embedded sections	Biorepository for Investigation of Neonatal Diseases of the Lung (BRINDL)	https://brindl.urmc.rochester.edu/
Fresh frozen paraffin-embedded unused adult deceased donor lung tissue	University of California San Diego (UCSD)	N/A
Chemicals, Peptides, and Recombinant Proteins		
100x Citrate Buffer pH 6.0	Abcam	Cat#ab64236
VECTASHIELD Vibrance Antifade Mounting Medium with DAPI	Vector Laboratories	Cat#H-1800
Deposited Data		
Raw data	This paper	GEO: GSE154027
Cellular census of human lungs	Vieira Braga et al., 2019	GEO: GSE130148
Cell atlas of mouse lung development	Cohen et al., 2018	GEO: GSE119228
scRNA-seq mouse lung at birth	Guo et al., 2019	GEO: GSE122332
LGEA scRNA-seq	Du et al., 2015; Du et al., 2017; Du et al., 2021	www.lungmap.net
Experimental Models: Organisms/Strains		
C57BL/6 mice frozen whole lungs	Perl Lab	N/A
Oligonucleotides		
(See Lake et al., 2018 Table S12)	Lake et al., 2018	N/A
Software and Algorithms		
Code for processing scTHS-seq data	Lake et al., 2018	https://github.com/yanwu2014/scTHS-Seq-processing
Code for analyzing lung scTHS-seq data	This paper	https://github.com/ThuElizabethD/Lung-scTHS-seq
bcl2fastq	Github	https://github.com/brwnj/bcl2fastq
deindexer	Github	https://github.com/ws6/deindexer
BWA	Github	https://github.com/lh3/bwa
pysam	Github	https://github.com/pysam-developers/pysam
SAMtools	Danecek et al., 2021	https://www.htslib.org/
SnapATAC	Fang, Priessl et al., 2021	https://github.com/r3fang/SnapATAC
MACS	Github	https://github.com/macs3-project/MACS

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
cisTopic	Bravo González-Blas et al., 2019	https://github.com/aertslab/cisTopic
Seurat	Stuart, Butler et al., 2019	https://satijalab.org/seurat/
Cicero	Pliner et al., 2018	https://cole-trapnell-lab.github.io/cicero-release/
deepTools	Ramirez et al., 2016	https://github.com/deeptools/deepTools
ChromVAR	Schep et al., 2017	https://greenleaflab.github.io/chromVAR/index.html
Signac	Stuart et al., 2021	https://satijalab.org/signac/
ChromVARmotifs	Github	https://github.com/GreenleafLab/chromVARmotifs
ChIPseeker	Guangchuang et al., 2015	http://bioconductor.org/packages/release/bioc/html/ChIPseeker.html
GViz	Hahne and Ivanek, 2016	https://bioconductor.org/packages/release/bioc/html/Gviz.html
TxDb.Hsapiens.UCSC.hg38.knownGene R package v3.4.6	Bioconductor	https://doi.org/10.18129/B9.bioc.TxDb.Hsapiens.UCSC.hg38.knownGene
TxDb.Mmusculus.UCSC.mm10.knownGene R package v3.4.7	Bioconductor	https://doi.org/10.18129/B9.bioc.TxDb.Mmusculus.UCSC.mm10.knownGene
org.Hs.eg.db R package v3.8.2	Bioconductor	https://doi.org/10.18129/B9.bioc.org.Hs.eg.db
org.Mm.eg.db R package v3.8.2	Bioconductor	https://doi.org/10.18129/B9.bioc.org.Mm.eg.db
g-chromVAR	Ulirsch et al., 2019	https://github.com/caleblareau/gchromVAR
CausalDB fine-mapping pipeline	Github	https://github.com/mulinlab/CAUSALdb-finemapping-pip
R/Rstudio	R/Rstudio	https://rstudio.com/
Other		
GentleMACS Dissociator	Miltenyi Biotec	Cat#130-093-235
GentleMACS C Tubes	Miltenyi Biotec	Cat#130-093-237
Sterile single-pack CellTrics filters - 30 um	CellTrics	Cat#04-004-2326
S2002 Dako Pen	Agilent Technologies	Cat#S200232-3
scTHS-seq Assay	Lake et al., 2018	https://doi.org/10.17504/protocols.io.xf8fjrw
TC20 Automated Cell Counter	Bio-Rad	Cat#1450102
HiScribe T7 High Yield RNA Synthesis Kit	New England BioLabs	Cat#E2040S
Taq 5X Master Mix	New England BioLabs	Cat# M0285L
Taq DNA Polymerase with Standard Taq Buffer	New England BioLabs	Cat#M0273E
AMPure XP	Beckman	Cat#A63882
Clontech SMART MMLV reverse transcriptase kit	Takara	Cat#639523
RNase H	Enzymatics	Cat#Y9220L

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kun Zhang (kzhang@bioeng.ucsd.edu).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- Single-cell THS-seq data have been deposited at GEO and are publicly available. Accession numbers are listed in the key resources table. Microscopy data reported in this paper will be shared by the lead contact upon request.
- All code for analysis is publicly available as of the date of publication at this github repository: github.com/ThuElizabethD/Lung-scTHS-seq.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

All post-mortem human pediatric donor lung samples were obtained from the Biorepository for Investigation of Neonatal Diseases of the Lung (BRINDL) supported by the NHLBI LungMAP Human Tissue Core housed at the University of Rochester. Consent, tissue acquisition, and storage protocols can be found on the repository's website (brindl.urmc.rochester.edu/). Frozen human lung tissue biopsies weighing approximately 100 mg were shipped in cryovials overnight from BRINDL on dry ice and processed upon arrival simultaneously with whole mouse lungs. Normal adult human lung from unused deceased donor tissue consented for research use was obtained from LifeSharing, San Diego, CA and processed for histology by formalin fixation and paraffin embedding. Sample size and donor metadata are listed in Table S1.

Animals

Archived healthy C57BL/6 fresh-frozen whole mouse lungs used in this study were shipped from the University of Cincinnati on dry ice and immediately stored at −80C. All mice were not involved in previous procedures. Sample size, sex and developmental stage are provided in Table S1.

METHOD DETAILS

Nuclei isolation

Human and mouse lung samples were thawed in 1-1.5 mL 10% dimethyl sulfoxide (DMSO; Fisher Scientific) in fetal bovine serum (FBS; GIBCO) on ice for 5-10 min. Contents were transferred into GentleMACS C tubes and 4 mL ice-cold nuclei extraction buffer was added (0.32M sucrose, 5 mM CaCl, 3mM MgAc2, 0.1 mM EDTA, 10 mM Tris-HCl, 0.1% Triton X-100). GentleMACS dissociator preset programs A through E was used, placing tubes on ice in between programs. Samples and 1 mL of NEB buffer used to rinse C tubes were passed through 30-um filters into 15 mL tubes. Total volume was brought up to 10 mL with PBS (GIBCO) + 2 mM EGTA (PBSE, pH ~7.4), centrifuged for 10 min x 900 g at 4C, and pellets resuspended in 1 mL PBSE + 2% fatty-acid-free BSA (Gemini) containing 4',6-diamidino-2-phenylindole (DAPI).

scTHS-seq library generation

Subsequent DAPI+ flow cytometry sorting, nuclei processing, scTHS-seq library generation, and sequencing was completed as described previously.[14] Briefly, after flow cytometry, nuclei were prepared for counting and species-species mixing by centrifuging at 500 g for 5 min at 4°C and pellet resuspension in 1x lysis buffer, followed by 1.5x tagmentation buffer. Human and mouse nuclei were counted on a Bio-Rad TC20 cell counter. Optimal cell concentrations were obtained for each sample (2- 5 × 10^5 nuclei/mL, target ?1 million total nuclei for each sample). Equal volumes of human and mouse samples were combined for transposition and combinatorial indexing. Generation of barcoded transposome complex was done using annealed transposons and Tn5059.[14,98] For nuclei tagmentation and barcoding, human-mouse mixed cell samples were added to the 384-well plate of freshly generated uniquely barcoded Tn5059 r5 transposome complexes. EDTA 50 mM was added to each well to stop the reaction. Next, tagmented nuclei were pooled and prepared for flow cytometry sorting into 96-well plates (100 nuclei per well). Propidium iodide (eBioscience) staining was used to select nuclei and doublets were removed using forward and side scatterplots. For library preparation, 96-well plates were processed individually. Guanidine hydrochloride was added to each well. Reactions were purified using 1.8x AMPure SPRI beads, bead-pelleting, and 80% ethanol washes. For end fill-in, NEB Taq polymerase 1x was added. NEB HiScribe T7 high-yield synthesis kits were used for *in vitro* transcription amplification. Reactions were purified using 2x SPRI binding buffer, 80% ethanol washes, and SPRI beads resuspended in nuclease-free water. Clontech SMART MMLV reverse transcriptase kits were used for reverse transcription. To degrade RNA in cDNA-RNA hybrids, we added Enzymatics RNase H. For second-strand synthesis, we used 20 μM sss_scnXTv2 followed by NEB taq5X. Reactions were purified again using 2x SPRI binding buffer, 80% ethanol washes, and SPRI beads resuspended in nuclease-free water. Then, double-stranded cDNA fragments were simultaneously fragmented and 3′ adaptors added with a custom nXTv2_i7 Tn5059 transposome. Reactions were purified and resuspended in nuclease-free water. Finally, samples were eluted off SPRI beads using a magnetic plate and qPCR was performed using standard Illumina Nextera XT v2 barcoding in an 8x12 (i5xi7) format using custom scTHS-seq i5 indexes and standard Illumina i7 indexes. Validated size-selected libraries were sequenced on high-throughput Illumina HiSeq 2500 (50+8+32 single-end reads).

Immunofluorescent staining and imaging

10 μM thick sections were cut from flash frozen paraffin embedded BRINDL and UCSD blocks and deparaffinized/rehydrated in xylene 3 × 15 min, 2 × 100%, 2 × 95%, 1 × 80% alcohol. A container containing 1x citrate buffer was placed in a glass dish filled XXX with diH20 and prewarmed in a microwave for 5 min, 100% power. Slides underwent antigen retrieval for 1 × 20 min, 50% power, 1 × 10 min, 50% power, making sure that the slides were submerged in citrate buffer at the end of the 1st 20 min. Keeping the slides in citrate buffer, they were cooled in 4°C for 45 – 60 min and then washed with running diH20. Tissue sections were encircled with PAP pen, followed by 10 min 03.% tween/PBS for permeabilization. Slides were then washed for 5 min with PBS and placed in a humidity chamber to be blocked with 5% bovine serum albumin (BSA). Without washing, BSA was tapped off each slide and primary

antibody, made in PBS, pipetted onto each slide (see chart for antibody dilutions). Slides, placed in a humidity chamber, were incubated at 4°C overnight. The next day, slides were washed with PBS using a squirt bottle then left for 5 min in PBS. Secondary was pipetted directly onto the slides and incubated for 1 h at room temperature. Slides were washed with PBS using a squirt bottle plus a 5-min PBS wash and coverslipped with Vectashield + DAPI. Slides were visualized using a Leica Stellaris 5.

Antigen or Fluorophore	Dilution
PIEZO2	1/100
ACTA2	1/50
LIMD1	1/100
RAGE	1/200
SPC	1/100
Alexa-555	1/200 (RAGE)
Dylight-750	1/100 (PIEZO2, SPC)
Alexa-647	1/200 (LIMD1), 1/300 (ACTA2)

QUANTIFICATION AND STATISTICAL ANALYSIS

scTHS-seq data processing

Alignment and Sorting - BAM files for uniquely barcoded nuclei were generated from raw BCL files as previously described[14] including a tutorial (https://github.com/yanwu2014/scTHS-Seq-processing). Briefly, using bcl2fastq (v2.17.1.14), we produced Fastq files for Read1, Index1, and Index2 and identified reads that map to each unique barcode combination using deindexer with a zero mismatch stringency resulting in a single fastq file per cell barcode combination. After deindexing, cell barcode combinations for each read were appended to the read name, and all fastq files re-merged for alignment. Merged fastq files were aligned to hg38 (GCA_000001405.15_GRCh38_no_alt_plus_hs38d1_analysis_set) and mm10 (GCA_000001635.5_ GRCm38.p3_no_alt_ analysis_set) reference genomes using BWA (0.7.12-r1039). Individual bam files were generated with assigned barcode, plate id, and sample id using pysam. Using samtools[99], bam files were merged and sorted into a single file for each species.

Quality filtering and peak calling

SnapATAC (v2)[100] snaptools snap-pre command (with –keep-single = TRUE) and snap-add-bmat (5 kb bin size) were ran to create a bins by cells matrix and high-quality cell barcodes were selected based on a total number of fragments per barcode ('fragment.num') greater than 1500 and a uniquely mapped ratio ('umap ratio') greater than 0.6. Next, the matrix was binarized and additional filtering removed any bins overlapping with ENCODE hg38 and mm10 blacklists (http://mitra.stanford.edu/kundaje/akundaje/release/ blacklists/) and the top 5% of bins with highest coverage.

Dimensionality reduction with PCA and construction of a KNN graph using the top 16 PCs (k = 20) was completed. R igraph clustering on the bins by cells matrix produced raw clusters and peak calling using combined reads from each cluster was run using MACS2 (v2.1.1.20160309) with the command "–nomodel–shift 37–ext 73–qvalue 5e-2 -B–SPMR–call-summits." Peaks from all clusters were merged to generate a consensus list of peaks, and a binary chromosomal peak locations by cell barcodes matrix was generated using 'createPmat' function.[100] Peak locations were filtered to keep only chromosomes 1-22, X for human and 1-19, X for mouse, and only peaks belonging to a minimum of 25 cells and cells with at least 200 accessible sites were kept.

Dimensionality reduction and clustering

Using cisTopic,[101] Latent Dirichlet Allocation (LDA) was completed on the peaks by cells matrix with the optimal number of topics selected by cisTopic's model selection functionality (human topics = 45, mouse topics = 40). LDA topics were utilized to generate UMAP coordinates for 2D visualization and both LDA topics and UMAP coordinates were embedded into Seurat (v3) for clustering using 'FindNeighbors' with k = 30 and FindClusters with resolution = 0.8.[28]

Gene activities and cell-type annotation

A gene activity matrix was created using Cicero with default parameters except for a co-accessibility cutoff of 0.1.[16] Peaks were annotated by assigning them to genes whose promoters were closer than 5kb upstream or downstream. Co-accessibility (the proportion of cells in which both peaks were simultaneously accessible) was calculated between these annotated peaks and peaks further than 5kb away from promoters (distal peaks). Each distal peak was assigned to the gene of the annotated peak to which it was most co-accessible.

To annotate clusters present in the entire human and mouse dataset, we systematically classified cell types first based on lung compartment (epithelial, mesenchymal, endothelial and immune) followed by more specific cell types based on gene activity and peaks in promoter regions using canonical marker genes. To visualize peaks, bam files containing cell barcodes from clusters were merged and sorted using SAMtools (v1.3.1), and converted to bigwig files using 'bamCoverage' from the deepTools

(v2.1.0).[102] Bigwigs were uploaded to UCSC Genome Browser (http://genome.ucsc.edu/).[103] To confirm assignment of cell type labels, average gene activities of each cluster were correlated with published lung reference sc/snRNA-seq datasets[17–24] from corresponding species. An iterative clustering strategy was used to further classify subpopulations by subsetting target cells or clusters, re-selecting topics for UMAP visualization, and clustering as described above.

Differential accessible peak and gene analysis
Differentially accessible peaks were identified for each cell type and sample using a modified Fisher's exact test as previously described.[14] To get a sense of the effect size of each differentially accessible peak, log-fold change was computed for each peak in each group, defined here as the log2 transform of the fraction of cells in a cluster that are accessible at that peak divided by the fraction of all other cells that are accessible at that peak. A specificity score was also computed as previously described to assess how specific the accessibility of a peak is to a cluster.[104] We defined a peak as differentially accessible for a cell type or sample if it had an adjusted p value of less than 0.05, a log fold-change of greater than 0.5, and a specificity score of greater than 0.0001. Using default parameters, top differential peaks for each cluster were analyzed using GREAT to link genomic regions to genes and associate significant biological processes.[43] Differential genes for each cluster were identified using the Cicero gene activity matrix and Seurat's "FindAllMarkers" which utilizes a logistic regression and likelihood-ratio test to identify differential genes.[28]

Transcription factor analysis
Differential TF activity between clusters and motif enrichment in differential peaks linked to cell-type marker genes were accomplished using ChromVAR,[27] Signac,[105] and Seurat function.[28] First, a filtered collection of human and mouse position weight matrices ("human_pwms_v2" and "mouse_pwms_v2") from ChromVARmotifs was utilized to look for the presence of motifs in each peak to construct a motif occurrence matrix. Then, ChromVar was used to compute a deviations matrix (motif accessibility for each TF for each cell). Next, Seurat was utilized to identify differential TF activities for each cluster.

To perform motif enrichment in a set of differentially accessible peaks linked to marker genes, top differential peaks for each cluster (qval < 0.01, logfc > 0.5) were identified and linked to promoter regions of genes using 'annotatePeak' (+/− 5kb from TSS) from ChIPseeker (v1.20.0).[106] Annotated peaks were then filtered by mapped genes. Only peaks which were mapped to a top differential gene for annotated clusters (adjusted p.value < 0.05, ave_logFC > 0.1) were kept. Motif enrichment was performed on peaks of marker genes with more than one associated differential peak using Signac 'FindMotifs'.

Connectivity spider plots and bigwig track visualizations
The connectivity spider plots consisted of three major tracks: peak connection, bigwig for each cell type, and gene annotation. Tracks were processed individually and added to a plot list for final plotting. Peak connection tracks only contained peaks related to gene-of-interest. To filter for target peaks, the top 20,000 differential peaks per cell type cluster were first identified (q-value cutoff = 0.01, specificity cutoff = 1e-4, log fold-change cutoff = 0.5). Then, a peaks-by-genes co-accessibility matrix was generated using all accessible peaks linked to promoter regions of genes and a co-accessibility score above 0.1 or 0.25. The peaks by genes co-accessibility matrix was then subsetted to a vector of one gene by non-zero value peaks. The Cicero peaks by peaks conns object was also subsetted to contain the same peaks. Peak connections were plotted using Cicero plotConnections.[16] Using GViz (v3.11),[107] cluster-specific bigwigs were visualized using DataTrack and gene annotation tracks were built using GeneRegionTrack with the corresponding genomes (TxDb.Hsapiens.UCSC.hg38.knownGene R package v3.4.6 and TxDb.Mmusculus.UCSC.mm10.knownGene R package v3.4.7). Gene annotations were converted to gene symbols using ENTREZID (org.Hs.eg.db and org.Mm.eg.db R packages v3.8.2). For final visualization, GViz 'plotTracks' function was utilized.

GWAS analysis
We used g-chromVAR[108] to identify cell types that are enriched in certain lung diseases or phenotypes. Briefly, g-chromVAR requires a bulk or single-cell chromatin accessibility matrix and fine-mapped GWAS summary statistics, which describe the probability that a SNP is the causal SNP for a given phenotype. We used pre-computed fine-mapping results computed using CaviarBF from CausalDB (http://mulinlab.tmu.edu.cn/causaldb/index.html) for most phenotypes. For the remaining phenotypes, we ran the CausalDB fine-mapping pipeline with default parameters on the UK Biobank Results published by the Neale group (http://www.nealelab.is/uk-biobank/) for Idiopathic Pulmonary Fibrosis (IPF) and a 2017 meta-analysis of lung function for FEV1, FVC, and FEV1/FVC (UK Biobank, 2018).[56] To run g-chromVAR, we first binarized the single-cell accessibility matrix and then summed the accessibilities of all cells belonging to a given cell type to create a pseudobulk accessibility matrix with one column per cell type. We then ran g-chromVAR with default parameters. Because the results of g-chromVAR can vary depending on the choice of background peaks, we ran g-chromVAR 100 times and used the average of those 100 runs as our final result.

LGEA scRNA-seq analysis
Raw expression matrices were downloaded from www.lungmap.net. Human 1 day old neonate (LungMAP ID LMEX0000001624) was analyzed alone using Seurat's v4.0.1 standard workflow[108] and integrated with 3 additional donors (LMEX0000001625, LMEX0000001626, and LMEX0000003690) as described by Stuart, Butler et al., 2019.[28] QC parameters included nFeature_RNA > 200 and < 2500. For UMAP visualization and clustering, we chose dims = 1:20, npcs = 30, and resolution = 1 and 1.5. Top marker

genes were generated for both resolutions using p_val_adj < 0.04, pct.2 < 0.1, pct.1 > 0.3, ave_log2FC > 1. Cell types were annotated by aligning signature expression gene lists available on the LGEA web portal (https://research.cchmc.org/pbge/lunggens/mainportal.html).

scTHS-seq pseudobulk and bulk LGEA RNA-seq comparison

Gene activity levels for each major cell type in Figure S1 were summed and divided by cell count proportion to create scTHS-seq pseudobulk levels. Human and mouse sorted epithelial, mesenchymal, endothelial, and immune bulk RNA-seq counts per million taken from the LGEA web portal (https://research.cchmc.org/pbge/lunggens/mainportal.html) for each gene were added across age groups (human - neonate, infant and child; mouse – PND1, PND7, PND14, PND28). See Table S7.

Comparison of scTHS-seq with published data

We obtained gene count matrices from GEO and www.lungmap.net. Reference datasets were normalized and variable genes identified using default settings for Seurat NormalizeData and FindVariableFeatures functions. We used the union of all marker genes for scTHS-seq and reference dataset to compute the Pearson correlation between the average scTHS-seq gene activity and average gene expression of cell types in each dataset. Dot plots using LGEA scRNA-seq[13,23,24] and Wang et al., 2020[21] snRNA-seq were generated using the Seurat DotPlot function.

Supplemental information

A single-cell regulatory map of postnatal lung

alveologenesis in humans and mice

Thu Elizabeth Duong, Yan Wu, Brandon Chin Sos, Weixiu Dong, Siddharth Limaye, Lauraine H. Rivier, Greg Myers, James S. Hagood, and Kun Zhang
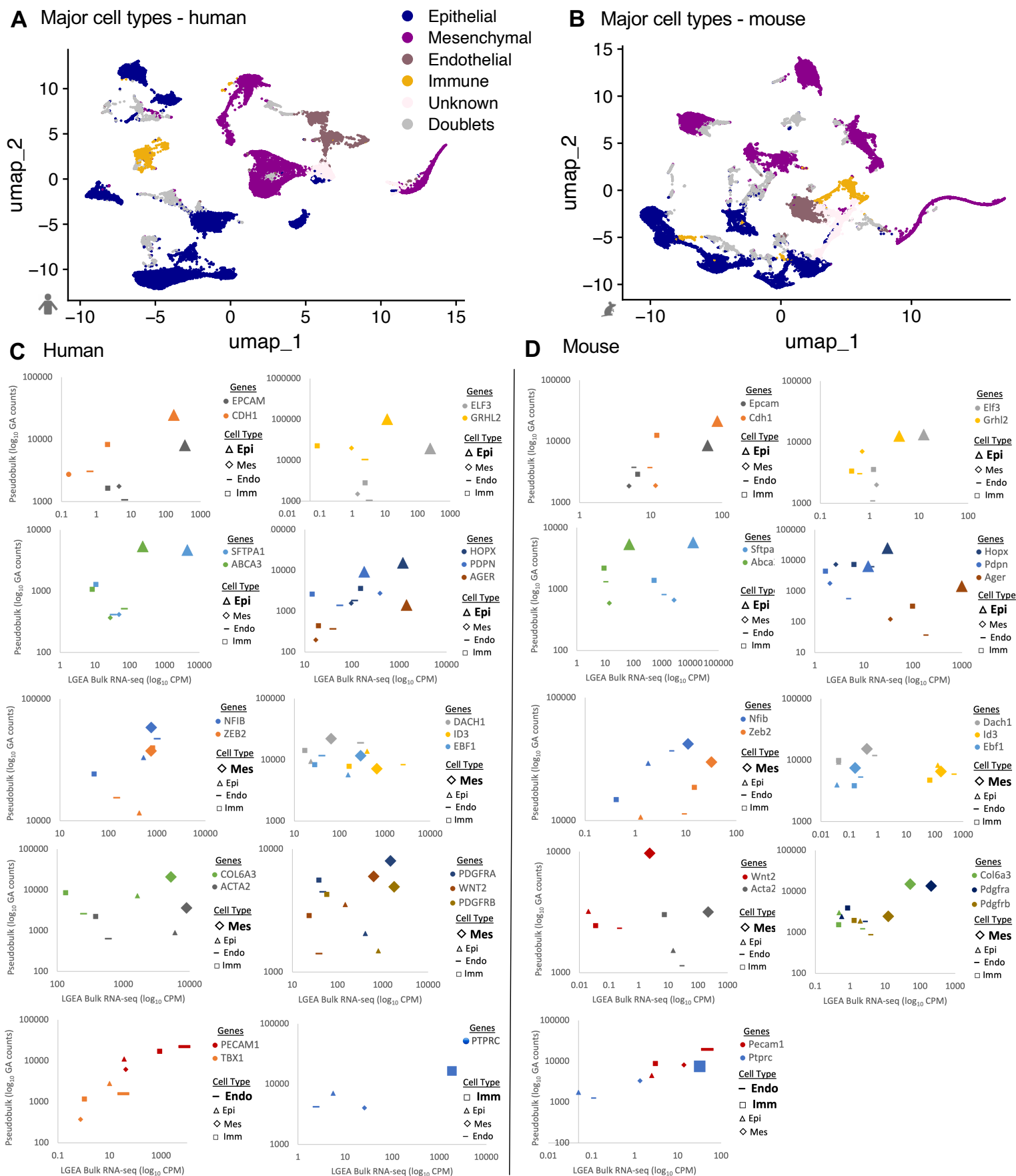
Figure S1. Correlation of marker genes between scTHS-seq and LGEA bulk RNA-seq, Related to Figure 1.
(A-B) UMAP of human and mouse scTHS-seq in Figure 1B-C colored by major cell type. **(C-D)** Scatter plots of pseudobulk scTHS-seq gene activity counts for marker genes used to classify clusters in major cell types with corresponding expression (log10 cells counts per million) of LGEA bulk RNA-seq of sorted epithelial, mesenchymal, endothelial, and immune cells [S1-3].
See also Table S7.
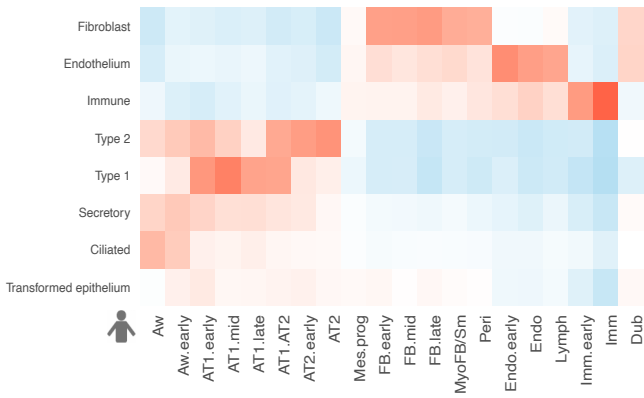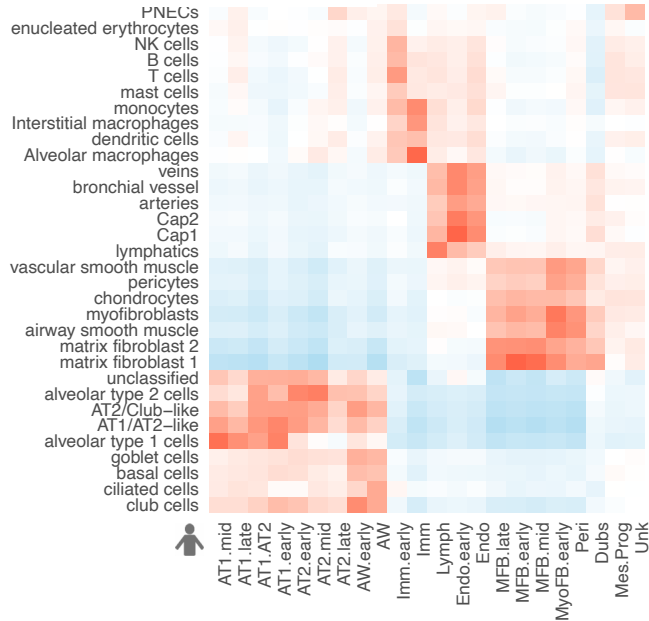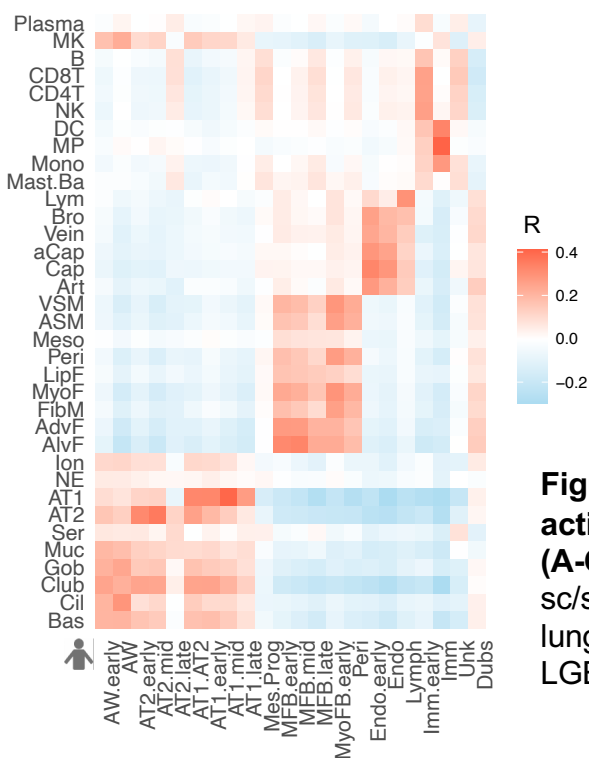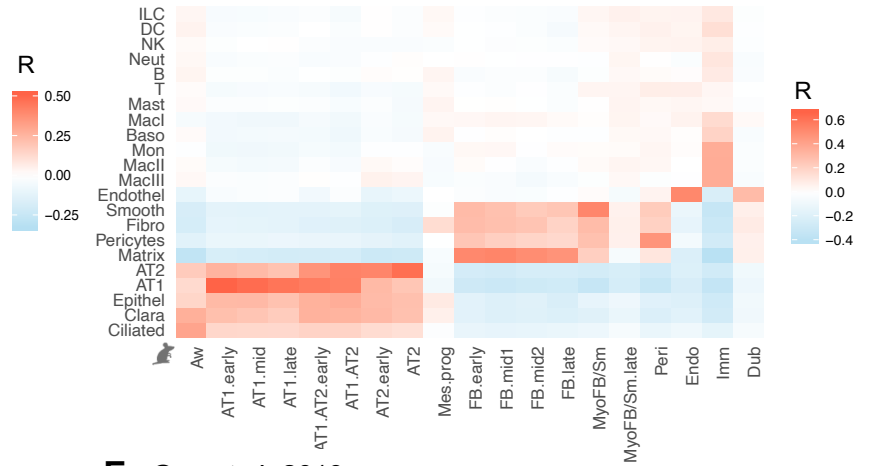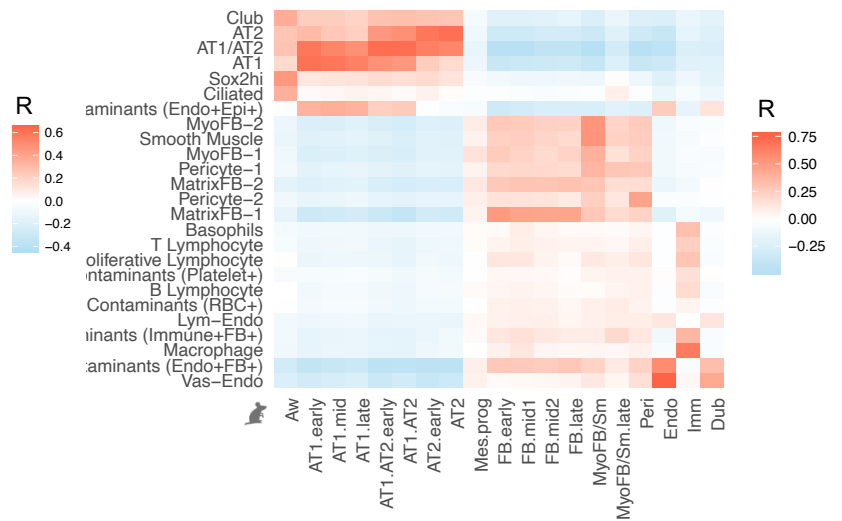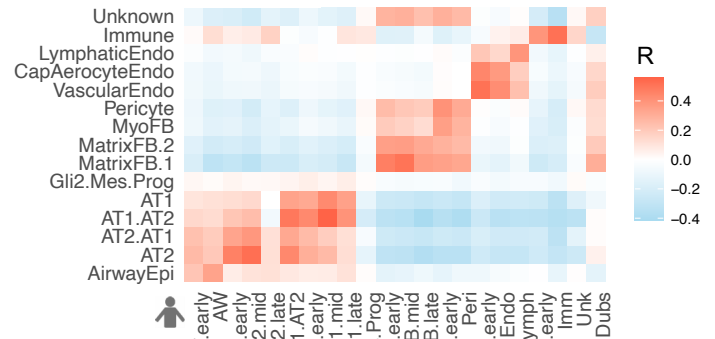
Figure S2. Correlation of human and mouse scTHS-seq gene activity with scRNA-seq lung datasets, Related to Figure 1. **(A-C)** Correlation with published control adult human lung sc/snRNA-seq datasets [S4-6]. **(D and E)** Correlation with mouse lung scRNA-seq datasets [S7-8]. **(F)** Correlation with integrated LGEA human lung scRNA-seq datasets from 4 donors [S1-3].
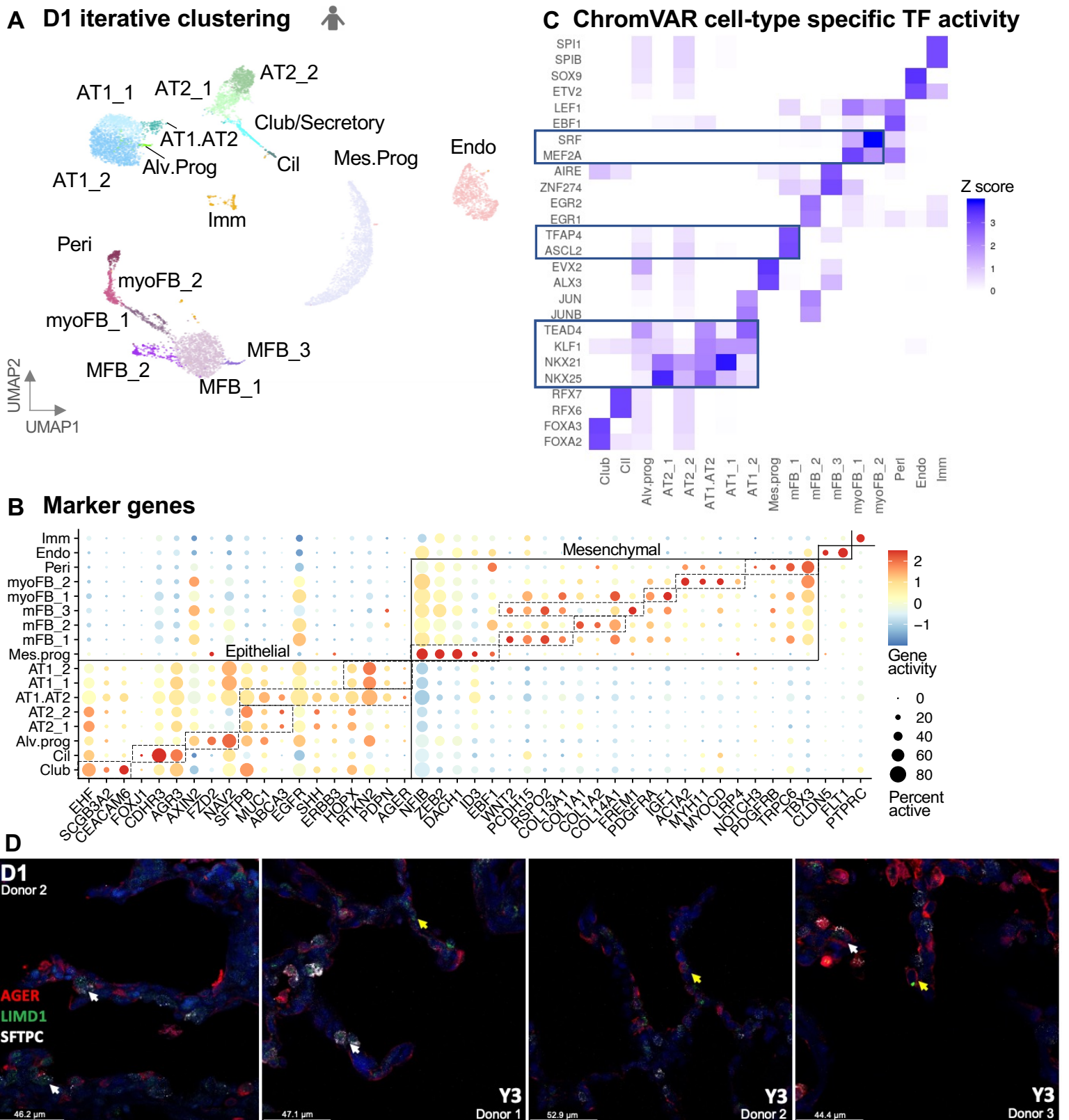
Figure S3. Iterative clustering of human day 1 lung cells and immunofluorescent staining for AT1.AT2 cells, Related to Figure 2-3, 5-6. (A) UMAP of annotated D1 cell types. **(B)** Dot plot of log-normalized gene activity scores for marker genes in each cell state. **(C)** Heatmap of top 2 differential TFs for each cluster using ChromVAR TF activity Z scores. Rectangles highlight key TFs explored in downstream construction of regulatory modules. **(D)** *LIMD1, AGER*, and *SFTPC* staining in alveolar region of human tissue in additional D1 and Y3 donors. White arrows indicate triple positive AT1.AT2 cells. Yellow arrows indicate *AGER* and *LIMD1* in AT1 cells. scTHS-seq data collected from donor 1 in each age group. See also Table S4.
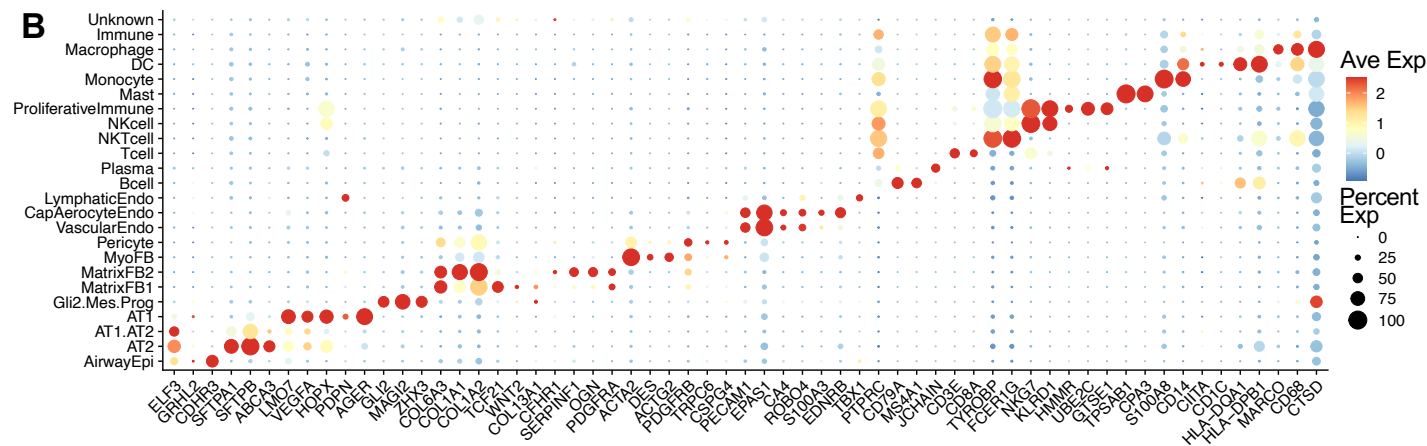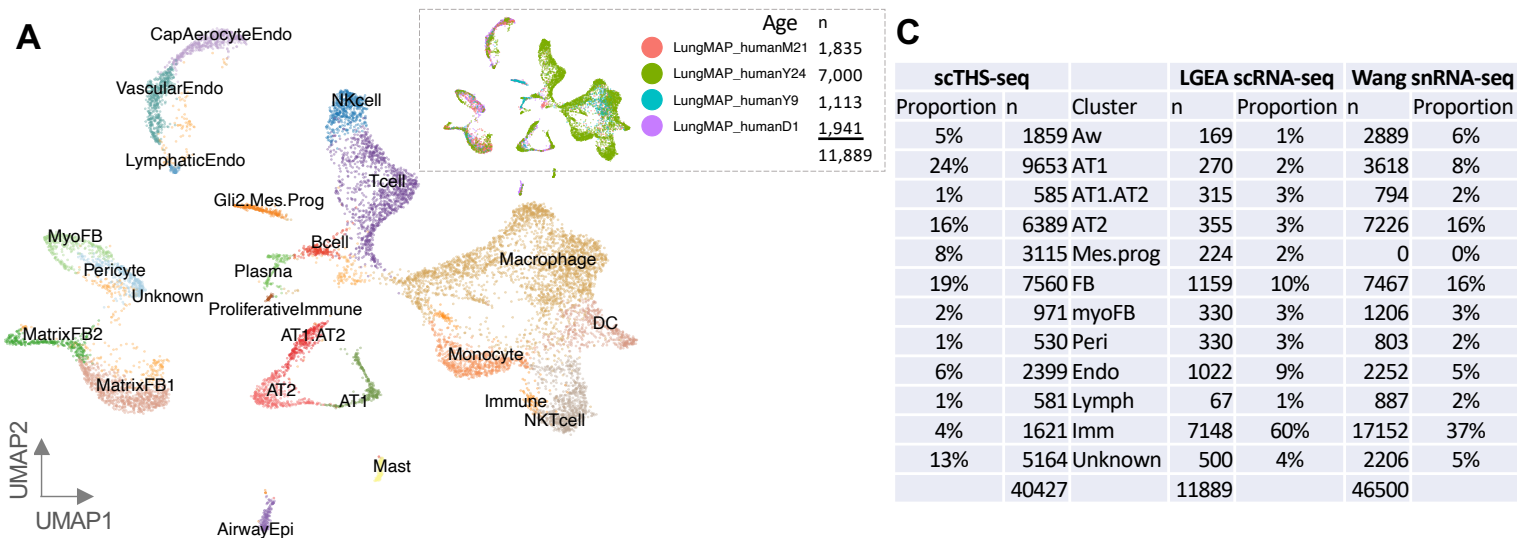
Figure S4. Human LGEA scRNA-seq analysis, Related to Figure 3 and 6.
(A) UMAPs of integrated LGEA human scRNA-seq data from D1, M24, Y9, and Y24 donors [S1-3] with annotated cell types based on signature expression of genes. Inset recolors UMAP by donor. **(B)** Dot plot of DE genes used to annotate cell types in A. **(C)** Table of approximate annotated cell type proportions in scTHS-seq, LGEA scRNA-seq, and Wang et al. 2020 [S5] snRNA-seq datasets. Annotated cell types were grouped by best fit into cluster labels listed. **(D)** UMAP of annotated cell types in LGEA D1 human scRNA-seq dataset. **(E)** Dot plot of DE genes used to annotate cell types in D.

Table C:

scTHS-seq			LGEA scRNA-seq		Wang snRNA-seq	
Proportion	n	Cluster	n	Proportion	n	Proportion
5%	1859	Aw	169	1%	2889	6%
24%	9653	AT1	270	2%	3618	8%
1%	585	AT1.AT2	315	3%	794	2%
16%	6389	AT2	355	3%	7226	16%
8%	3115	Mes.prog	224	2%	0	0%
19%	7560	FB	1159	10%	7467	16%
2%	971	myoFB	330	3%	1206	3%
1%	530	Peri	330	3%	803	2%
6%	2399	Endo	1022	9%	2252	5%
1%	581	Lymph	67	1%	887	2%
4%	1621	Imm	7148	60%	17152	37%
13%	5164	Unknown	500	4%	2206	5%
	40427		11889		46500	

A P1 iterative clustering

B Marker genes

C ChromVAR cell-type specific TF activities
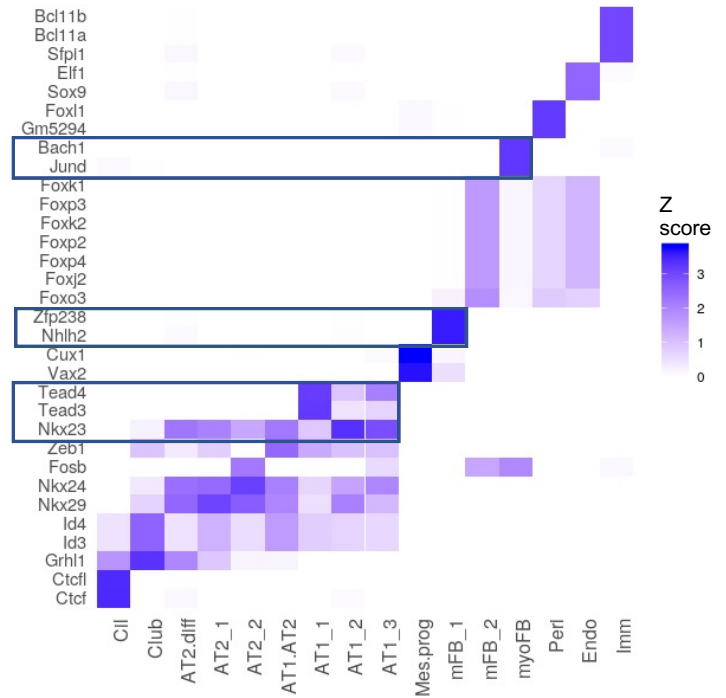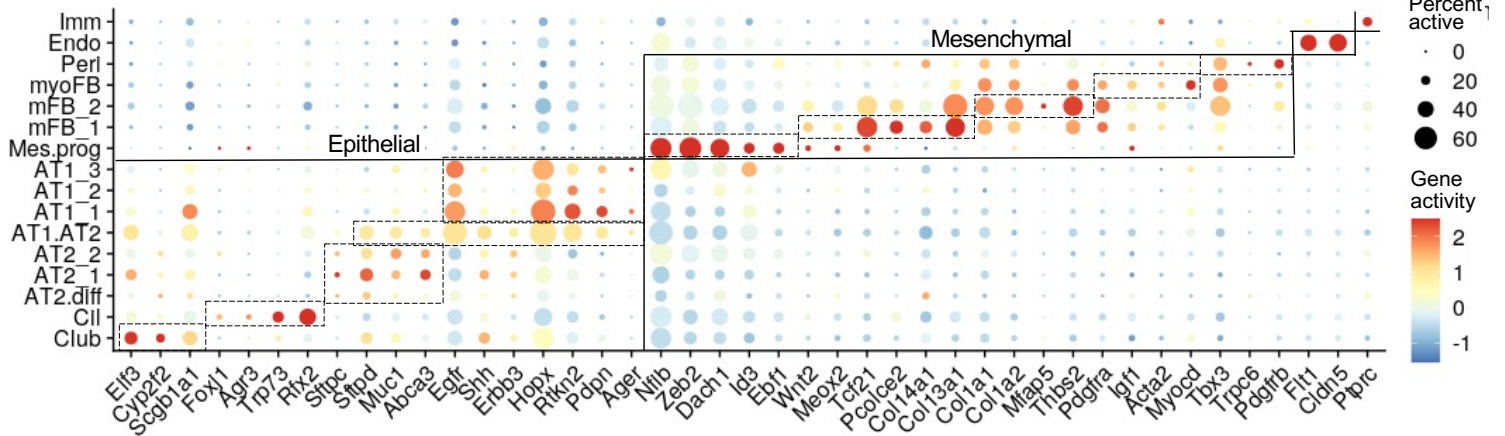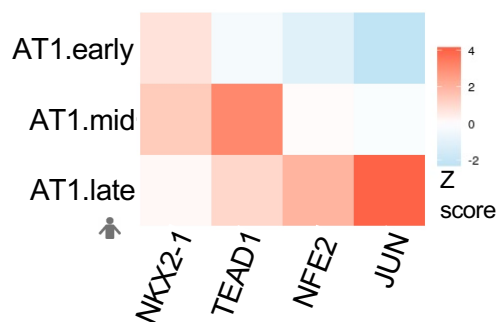
D AT1.AT2 genes
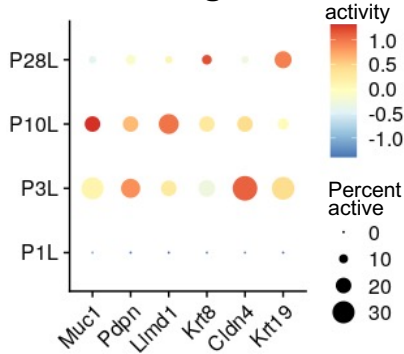
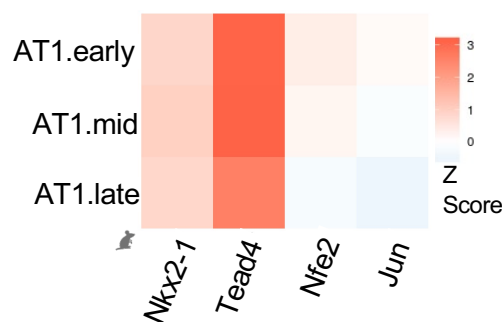**Figure S5. Iterative clustering of mouse postnatal day 1 lung cells, Related to Figure 2, 5, and 6.**
(A) UMAP annotated P1 cells. **(B)** Dot plot of gene activity scores for P1 annotated clusters. **(C)** Top 2 TFs for each cluster using ChromVAR differential TF activity Z scores. **(D)** Dot plot of DA genes in AT1.AT2 cluster by donor.
See also Table S4.
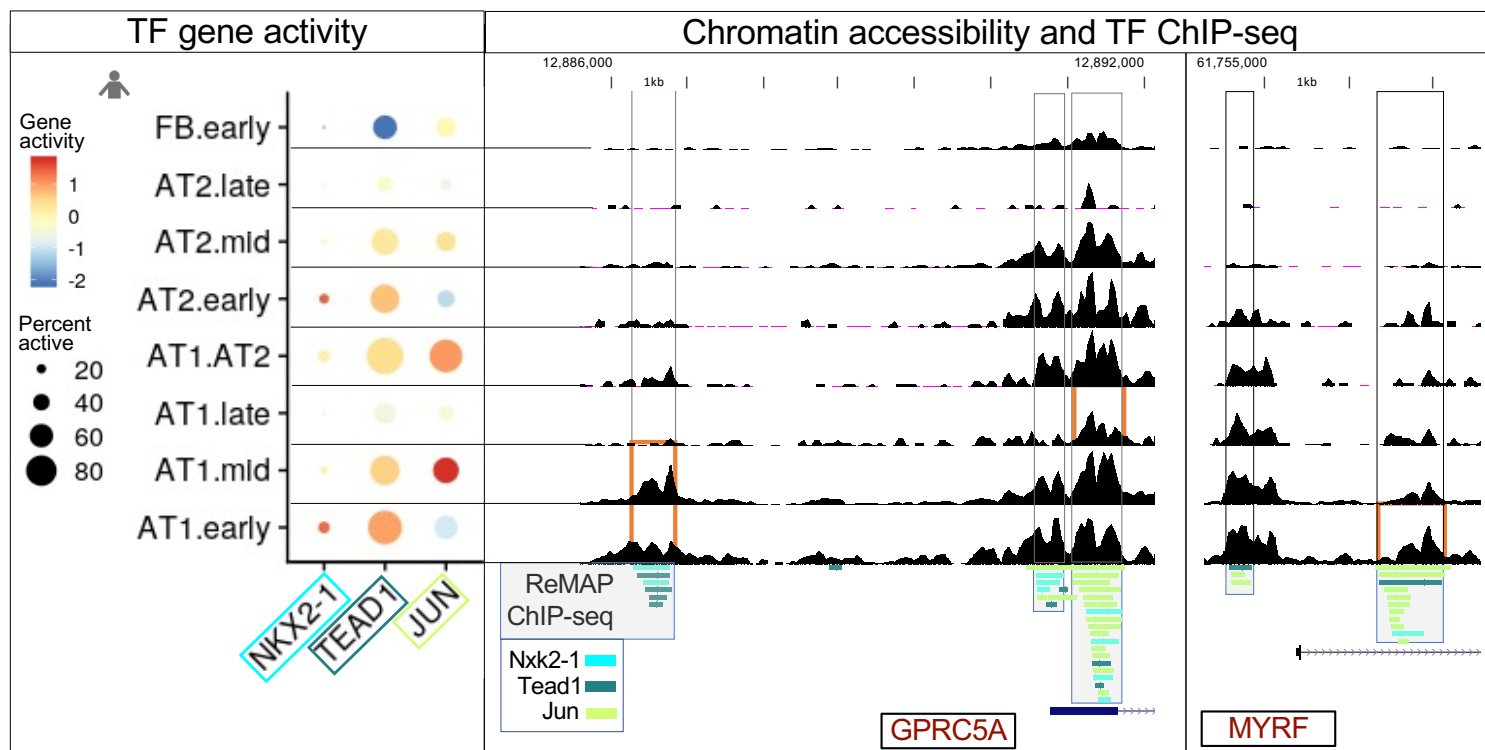
Figure S6. Dynamic AT1 cell regulation with increasing age and across species, Related to Figure 1.
(A and B) TF activity Z scores for human and mouse AT1 subsets from Figure 1B and 1C. **(C)** Aligning temporal TF availability with corresponding TF binding to accessible peaks linked to AT1 marker genes. Gene activity scores for *NKX2-1, TEAD1,* and *JUN* for age-defined clusters from Figure 1B (left panel). Genome browser tracks for corresponding cell types with highlighted overlap between accessible peaks and ChIP-seq peaks for *NKX2-1, TEAD1,* and *JUN* from ReMAP 2018 [S9] near *GPRC5A* and *MYRF* TSS. Bars for ChIP-seq peaks are colored by TF to which they belong (right panel).
See also Table S2.
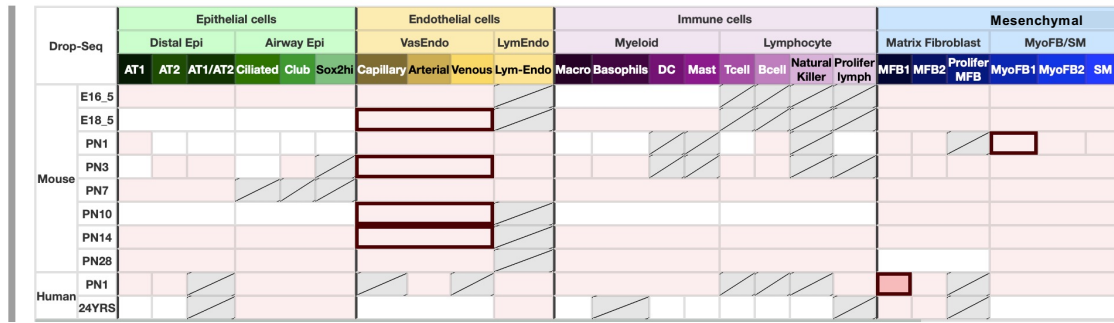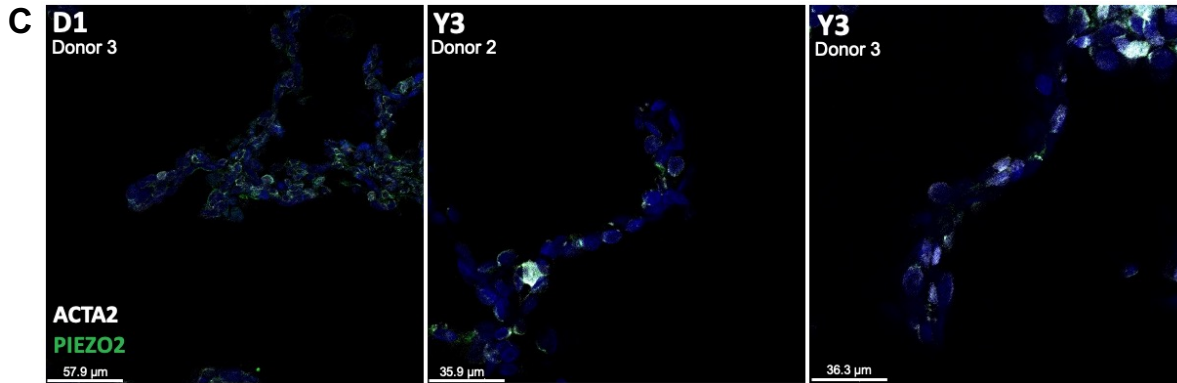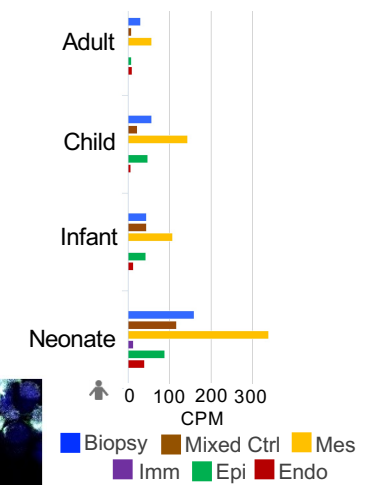
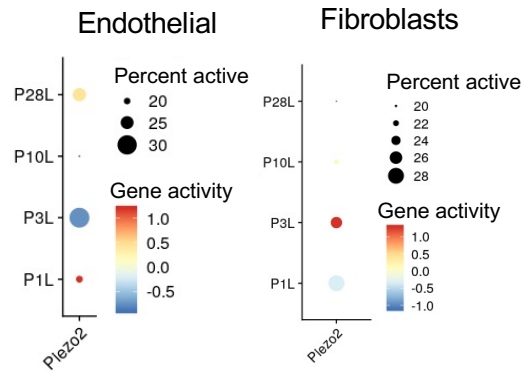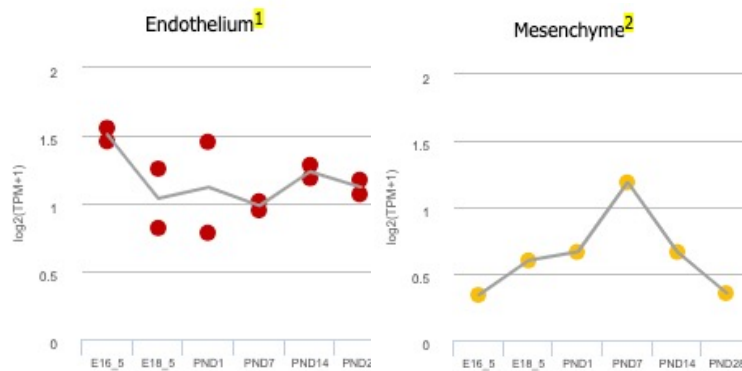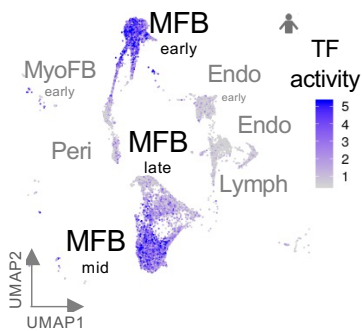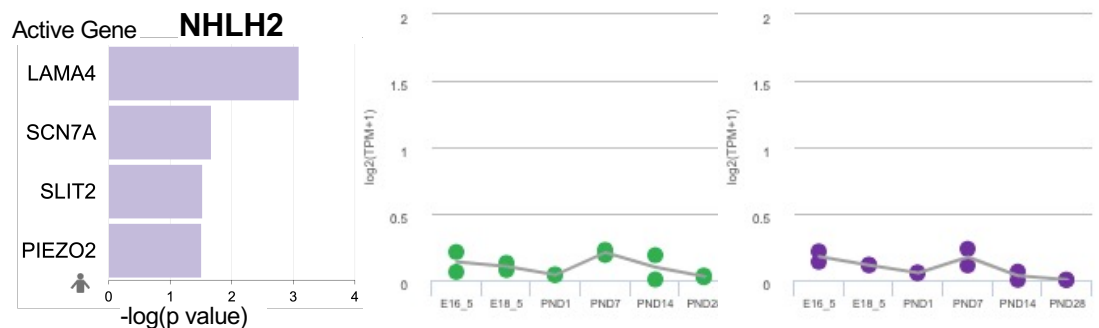**Figure S7. Accessibility, expression, and immunostaining of *PIEZO2*, Related to Figure 6.**
(A) Table from LGEA with *PIEZO2* SE (red boxes) from scDrop-seq experiments in varying human and mouse cell types at different developmental ages. **(B)** Average counts per million (CPM) of *PIEZO2* expression from LGEA bulk sorted epithelial, mesenchymal, endothelial, immune, biopsy, and mixed control populations from neonatal, infant, child and adult human lung cells. **(C)** *PIEZO2* and *ACTA2* staining in alveolar region of human tissue in additional D1 and Y3 donors. **(D)** *Piezo2* gene activity for all mouse endothelial and fibroblasts grouped by age. **(E)** *Piezo2* bulk gene expression (log2(TPM+1)) in LGEA endothelium, mesenchyme, epithelium, and immune sorted lung populations from different murine developmental time points. **(F)** UMAP of stromal subset from Figure 1B highlighting *NHLH2* TF activity across development. **(G)** Barplot of -log10(p values) for active genes in D1 MFB_1 cells with significant *NHLH2* motif enrichment.

Table S6. Summary of key findings and supporting evidence, Related to Figure 1-7.

Finding	Evidence
Human and mouse mesenchymal progenitor (Mes.Prog) population -Signature genes: *NFIB, ZEB2, DACH1, ID3, EBF1*	Table S3 – Overlapping differential genes in similar cell populations in the following manuscripts and LGEA: 1. Human and mouse Ebf1+ fibroblasts (Liu et al., iScience 2021 PMID 34151224)[S10] 2. Mouse mesenchymal progenitor (MP) (Xie et al., Cell Rep. 2018 PMID 29590628)[S11] 3. Mouse proliferative mesenchymal progenitors (PMP) (LGEA E16.5)[S1-3]
AT1.AT2 cells -Signature genes: *LIMD1, KRT8, KRT19*, and *CLDN*	Corresponding LGEA [S1-3] and Wang et al., Elife 2020 [S5] sc/snRNA-seq gene expression in human lung. Figure 2C and S3D *LIMD1* protein immunostaining in human lung at D1, Y3, and adult. Three donors each for D1 and Y3. Similar cell populations: 1. Human and mouse Pre-alveolar type 1 transitional cell state (PATS) (Kobayashi et al., Nat Cell Bio 2020 PMID 32661339)[S12] 2. Mouse Krt8+ alveolar differentiation intermediate (ADI) (Strunz/Simon et al., Nat Comm 2020 PMID 32678092)[S13] 3. Mouse damage-associated transient progenitors (DATPs) (Choi et al., Cell Stem Cell 2020 PMID 32750316)[S14]
KLF and *TEAD* TF involvement in AT1 cell differentiation AT1 *PDGFA* ligand to myofibroblast *PDGFRA* receptor signaling drives postnatal alveolar septation	Corresponding TF and gene expression in LGEA [S1-3] and Wang et al., Elife 2020 [S5] sc/snRNA-seq. 1. Mouse scRNA-seq and scATAC-seq data (Zepp et al., Science 2021 PMID 33707239)[S15] – AT1 cells are a signaling hub to receptive, force-exerting myofibroblasts required for postnatal respiration.
Increase in number of myofibroblasts during human and mouse alveolar septation	LGEA ACTA2 protein immunostaining in human lung (Figure 4A). Figure 6H and S7C *ACTA2* protein immunostaining (co-staining with PIEZO2) in human lung. Three donors each for D1 and Y3.
Matrix fibroblast type 1 Signature gene: *PIEZO2*	Corresponding human and mouse single-cell expression and LGEA bulk expression of PIEZO2 (Figure 6 and S7). Figure 6H and S7C *PIEZO2* protein immunostaining in human lung. Three donors each for D1 and Y3.

Supplemental Information References:

1. Du, Y., Guo, M., Whitsett, J. A. & Xu, Y. (2015). 'LungGENS': A web-based tool for mapping single-cell gene expression in the developing lung. *Thorax* **70**, 1092–1094.
2. Du, Y., Kitzmiller, J. A., Sridharan, A., Perl, A. K., Bridges, J. P., Misra, R. S., Pryhuber, G. S., Mariani, T. J., Bhattacharya, S., Guo, M., *et al.* (2017). Lung Gene Expression Analysis (LGEA): An integrative web portal for comprehensive gene expression data analysis in lung development. *Thorax* vol. 72 481–484.
3. Du, Y., Ouyang, W., Kitzmiller, J. A., Guo, M., Zhao, S., Whitsett, J. A. & Xu, Y. (2021). Lung gene expression analysis web portal version 3: Lung-at-a-glance. *American Journal of Respiratory Cell and Molecular Biology* vol. 64 146–149.
4. Vieira Braga, F. A., Kar, G., Berg, M., Carpaij, O. A., Polanski, K., Simon, L. M., Brouwer, S., Gomes, T., Hesse, L., Jiang, J., *et al.* (2019). A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* 1 doi:10.1038/s41591-019-0468-5.
5. Wang, A., Chiou, J., Poirion, O. B., Buchanan, J., Valdez, M. J., Verheyden, J. M., Hou, X., Kudtarkar, P., Narendra, S., Newsome, J. M., *et al.* (2020). Single-cell multiomic profiling of human lungs reveals cell-type-specific and age-dynamic control of sars-cov2 host genes. *Elife* **9**, 1–28.
6. Travaglini, K. J., Nabhan, A. N., Penland, L., Sinha, R., Gillich, A., Sit, R. V., Chang, S., Conley, S. D., Mori, Y., Seita, J., *et al.* (2020). A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625.
7. Cohen, M., Giladi, A., Gorki, A. D., Solodkin, D. G., Zada, M., Hladik, A., Miklosi, A., Salame, T. M., Halpern, K. B., David, E., *et al.* (2018). Lung Single-Cell Signaling Interaction Map Reveals Basophil Role in Macrophage Imprinting. *Cell* **175**, 1031-1044.e18.
8. Guo, M., Du, Y., Gokey, J. J., Ray, S., Bell, S. M., Adam, M., Sudha, P., Perl, A. K., Deshmukh, H., Potter, S. S., *et al.* (2019). Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. *Nat. Commun.* **10**, 37.
9. Chèneby, J., Gheorghe, M., Artufel, M., Mathelier, A. & Ballester, B. (2018). ReMap 2018: An updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. *Nucleic Acids Res.* **46**, D267–D275.