A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates

Peng He[1,2,3], Kyungtae Lim[1,4], Dawei Sun[1,4, 14], Jan Patrick Pett[2], Quitz Jeng[4], Krzysztof Polanski[2], Ziqi Dong[4], Liam Bolt[2,12], Laura Richardson[2], Lira Mamanova[2,13], Monika Dabrowska[2], Anna Wilbrey-Clark[2], Elo Madissoon[2,3], Zewen Kelvin Tuong[2,5], Emma Dann[2], Chenqu Suo[2,6], Isaac Goh[7], Masahiro Yoshida[8], Marko Z Nikolić[8], Sam M Janes[8], Xiaoling He[9], Roger A Barker[9], Sarah A Teichmann[2,10,*], John C. Marioni[2,3,11,*], Kerstin B Meyer[2,*], Emma L Rawlins[4,15,*]

1. Joint first authors
2. Wellcome Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK
3. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, UK.
4. Wellcome Trust/CRUK Gurdon Institute, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, CB2 1QN, UK.
5. Molecular Immunity Unit, University of Cambridge Department of Medicine, Cambridge, UK
6. Department of Paediatrics, Cambridge University Hospitals, Hills Road, Cambridge CB2 0 QQ, UK
7. Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK
8. Lungs for Living Research Centre, UCL Respiratory, University College London, London, UK
9. John van Geest Centre for Brain Repair, Department of Clinical Neurosciences and Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK
10. Department of Physics, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK
11. Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK
12. Current affiliation: Genomics England, Hinxton, CB10 1DR, UK
13. Current affiliation: CS Genomics, Cambridge CB1 2JH, UK
14. Current affiliation: Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA

15. Lead contact

*Corresponding authors: sarah.teichmann@sanger.ac.uk, marioni@ebi.ac.uk, km16@sanger.ac.uk, elr21@cam.ac.uk

Highlights

- Spatiotemporal atlas of human lung development from 5-22 post conception weeks identifies 144 cell types/states.
- Tracking the developmental origins of multiple cell compartments, including new progenitor states.
- Functional diversity of fibroblasts in distinct anatomical signalling niches.
- Resource applied to interrogate and experimentally test the transcription factor code controlling neuroendocrine cell heterogeneity and the origins of small cell lung cancer.

Abstract

We present a multiomic cell atlas of human lung development that combines single cell RNA and ATAC sequencing, high throughput spatial transcriptomics and single cell imaging. Coupling single cell methods with spatial analysis has allowed a comprehensive cellular survey of the epithelial, mesenchymal, endothelial and erythrocyte/leukocyte compartments from 5-22 post conception weeks. We identify new cell states in all compartments. These include developmental-specific secretory progenitors and a new subtype of neuroendocrine cell related to human small cell lung cancer. Our datasets are available through our web interface (https://lungcellatlas.org). Finally, to illustrate its general utility, we use our cell atlas to generate predictions about cell-cell signalling and transcription factor hierarchies which we test using organoid models.

Introduction

Single cell mapping of cell states in the adult human lung in health and disease is being performed at increasing resolution (Carraro and Stripp 2022) providing a foundation for understanding lung cellular physiology. The adult lung has low rates of cell turnover (Blenkinsopp 1967; Rawlins and Hogan 2008) making it difficult to capture transition states and progenitor cells. Moreover, there are developmental-specific cell states that do not exist in the adult. A high-resolution cell atlas of

the embryonic and fetal human lung will identify developmental precursors and progenitors, predict differentiation trajectories and potential gene regulatory networks. This will provide a baseline for studying adult homeostasis and disease.

The lung buds are specified in the human foregut endoderm at ~5 post conception weeks (pcw) (Burri 1984; Nikolić, Sun, and Rawlins 2018). Subsequent morphogenesis is driven by branching of the distal-most bud tips. The bud tip epithelium comprises $SOX9^{+}$, $ID2^{+}$ multipotent progenitors which self-renew during branching (Rawlins, Clark, et al. 2009; Alanis et al. 2014; Nikolić et al. 2017; Miller et al. 2018). As the bud tip epithelium branches into the surrounding mesoderm, the epithelial cells that remain in the stalk region start to differentiate into bronchiolar (airway) epithelium (~5-16 pcw) and then (from ~16 pcw) into alveolar epithelium (Nikolić, Sun, and Rawlins 2018). The pattern of growth from undifferentiated multipotent epithelial progenitors at the distal tips means that the position of a cell along the proximal-distal axis of the lung epithelial tree is a strong predictor of its maturity. The more mature cells, which exited the tip first, are found more proximally, whereas the most immature cell states, which exited the tip recently, are found in the tip-adjacent (stalk) regions (Rawlins et al. 2007). In other words, space reflects time in lung development. Therefore, coupling single cell state identification to *in vivo* spatial visualisation can provide high confidence in the identification of novel progenitor cells in the developing lung. Moreover, detailed spatial analysis of cell states allows cell identity designations to be compared to more traditional histological definitions.

We have generated a high-resolution single cell atlas of human lung development using a combination of scRNA-seq, scATAC-seq, Visium Spatial Transcriptomics, and mRNA *in situ* hybridisation using hybridisation chain reaction (HCR) (Trivedi et al. 2018). Combining these data sources has allowed us to identify 144 cell states/types in 5-22 pcw lung samples. These include novel progenitor cell states, transition populations and a new subtype of neuroendocrine cell which is related to a subtype of human small cell lung cancer. We observe increasing cell maturation over time with many cell states identified in the adult human lungs already present at 22 pcw. Moreover, we have used our atlas to make predictions about progenitor cell states, signalling interactions and lineage-defining transcription factors, and we demonstrate how these can be efficiently tested

using a genetically-tractable human fetal lung organoid model. These data sets are available for interactive analysis at https://lungcellatlas.org.

Results

A single cell atlas of human lung development comprising 144 cell states

We obtained human embryonic and fetal lungs from 5-22 pcw for scRNA-seq and scATAC-seq. To focus on differentiation we deeply sampled 15, 18, 20 and 22 pcw lungs and separated proximal and distal regions, while leaving lungs at 5, 6, 9 and 11 pcw intact. We used a mixture of cell dissociation methods to obtain a balanced mixture of cell types (Fig. 1A) and produced high-quality transcriptome (Fig. S1A; average >2400 genes/cell) and DNA accessibility (Fig. S1O,P; average >18,000 fragments/nucleus) data. The RNA profiles of cells from different dissociation protocols and compartments were iteratively parsed out by clustering (Fig. S1C) and subclustering (Fig. S1D) without batch correction to maintain biological features. To enrich biological features and mitigate technical ones, we removed doublet-driven clusters (Fig. S1E,G,H,K), stressed or low-quality clusters (except those expressing known markers, such as erythroid) (Fig. S1I,J,N), clusters composed of cells from only one sample when replicates are available, and clusters of cells from other organs due to contamination during dissection (Fig. S1L) (He et al. 2020; Cao et al. 2020). We identified maternal cells, which are a minimal fraction in our atlas (Fig. S1M), based on genetic background (Fig. S1F). The result of this clustering is shown as a Uniform Manifold Approximation and Projection (UMAP) (Fig. 1A), on which we manually annotated fibroblast, epithelial, endothelial and erythrocyte/leukocyte lineages (Fig. 1B). Plotting the cell type distribution against time (excluding trypsin/CD326-treated samples, shown in Fig. S1B), showed that fibroblasts were the most prominent cell, particularly in younger lungs (Fig. 1C). Leukocytes and erythrocytes were observed in all lungs sampled, with B, T and NK cells becoming prominent from 15 pcw (Fig. 1C).

Further cell type annotation was performed via sub-clustering (see methods) and cluster naming was based on observed and published marker genes (Supplemental Table 1), resulting in assignment of 144 cell types/states (Fig. S2A). Sample age was a strong determinant of clustering (χ^2=163727, p ≈ 0), reflecting progressive cell type maturity over time (Fig. S2B). Clusters mostly grouped into three distinct regions which we categorised as early (5, 6 pcw), mid (9, 11 pcw) and

late (15-22 pcw) stages. Cell cycle phase (Fig. S2C, χ^2=25361, p ≈ 0) and dissected region (Fig. S2D, χ^2=968, p = 8.9E-131) were also associated with clustering. However, region was only prominent for a small number of proximally-located cell types (Fig. S2H), suggesting that most proximal to distal regions of the airway branch structure were still represented in both dissected regions of the lung. Epithelial cells were mostly derived from the trypsin-treated and CD326-enriched samples, although airway smooth muscle, myofibroblasts and alveolar fibroblasts were also enriched here (Fig. S2E). Peripheral Nervous System (PNS) cells and chondrocytes were only obtained from 5-6 pcw lungs, likely correlating with lower extracellular matrix (ECM) complexity and/or increased fragility of older neurons. PNS cells were clustered and assigned to cell types, but their scarcity precluded further analysis (Fig. S2A,F,G). Data integration and logistic regression-based comparison showed that gene expression of our annotated cells corresponds well to those from adult lungs (Madissoon et al. 2021) (Fig. S3A-C).

A differentiation trajectory of airway progenitor states lies along the distal to proximal axis of the developing lungs

The epithelial cells separate by age (Fig. 2A,B), with many basal cells, MUC16$^+$ ciliated cells and secretory cells enriched in the proximally-dissected tissue (Fig. 2B; S2H). The most immature epithelial progenitors are tip cells: SOX9$^+$ multipotent progenitors located at the distal branching tips of the respiratory tree (Nikolić et al. 2017). Tip cells were separated by developmental age into early (5,6 pcw), mid (9,11 pcw) and late (15-22 pcw) populations (Fig. 2A,B) with some shared and some stage specific markers (Fig. 2C). On the epithelial UMAP, each tip population clusters closely with adjacent stalk cells (*SOX9*$^{LO/-}$, *PDPN*^{LO}, *HOPX*^{LO}) and airway progenitors (*CYTL1*$^{LO/+}$, *PCP4*$^+$, *SCGB3A*$^{+/LO}$) (Fig. 2A). The tip, stalk and airway progenitors can be visualised in a distal-proximal sequence in the tissue at all stages tested (10-16 pcw) (Fig. 2E; S4A,B, Supplemental video 1), consistent with the most proximal cells being the most mature. These three cell types form a predicted differentiation trajectory from mid-tip to mid-stalk to mid-airway progenitor which branches into the neuroendocrine, or secretory, lineages (Fig. 2D).

Two subtypes of neuroendocrine cells are present in the developing airways

Consistent with previous data (Cutz, Gillan, and Bryan 1985), the earliest differentiated epithelial cells detected were neuroendocrine (NE) cells in 5 pcw lungs (Fig. 2A-C). We identified two types of NE cells: classical pulmonary NE cells (*GRP*$^{+}$) and GHRL^{+} NE cells (*TTR*$^{+}$, *GHRL*$^{+}$) in agreement with a recent human fetal cell atlas (Cao et al. 2020). We observed increasing maturity of NE cells over time (specific populations denoted as precursors on the UMAP). In addition, an intermediate NE population, a putative transition state, connected the two NE cell types (Fig. 2A). At 11 pcw, *GRP*$^{+}$ NE cells were always observed closer to the budding tips, suggesting that they begin to differentiate prior to the *GHRL*$^{+}$ NE cells (Fig. 2F). This spatial difference was not apparent in the oldest samples where both *GRP*$^{+}$ and *GHRL*$^{+}$ cells were observed at all airway levels, although less abundant distally (Fig. S4C). Mouse GHRL^{+} NE cells were not detected in re-analysis of published mouse data (Negretti, Plosa, Benjamin, Schuler, Christian Habermann, et al. 2021; Zepp et al. 2021), or spatially (Borromeo et al. 2016). However, *Ghrl* is expressed in mouse ciliated cells which cluster with human fetal GHRL^{+} NE cells following scVI integration (Lopez et al. 2018) and clustering analysis (Fig. S3D-F) (Zepp et al. 2021).

Multiple secretory cell subtypes in the proximal cartilaginous airways

We annotated 5 sub-types of differentiating secretory cells and one more immature proximal secretory progenitor. (i) The proximal secretory progenitors (*SCGB3A2*$^{+}$, *SCGB1A1*$^{-}$, *SCGB3A1*$^{-/LO}$, *CYTL1*$^{+}$) were detected in the single cell atlas at 9 pcw, prominent at 11 pcw, but rarer in older lungs consistent with a progenitor state (Fig. 2A-C,G). (ii) Club cells (S*CGB3A2*$^{+}$, *SCGB1A1*$^{+}$, *SCGB3A1*$^{-}$, *SPDEF*$^{-}$, *MUC16*$^{-}$) were detected from 15 pcw in the single cell data (Fig 2A-C,G), or 12 pcw in the tissue localised in clusters more distally, but largely dispersed in the more proximal non-cartilaginous regions (Fig. S4D). (iii) Submucosal gland (SMG) secretory cells (*LTF*$^{+}$, *SCGB3A1*$^{+}$, *SPDEF*$^{+}$) were detected from 15 pcw in the single cell data, located in SMG ducts and likely to be a precursor of serous and/or mucous-secreting SMG cells (Fig. 2A-C,G; S4G). (iv) Proximal secretory 1 (*SCGB1A1*^{LO}, *SCGB3A2*$^{+}$, *SCGB3A1*$^{+}$) and (v) proximal secretory 2 (*SCGB1A1*$^{+}$, *SCGB3A2*$^{+}$, *SCGB3A1*$^{+}$) appeared from 11 pcw (Fig. 2A-C,G-H; S4E). Both were *SPDEF*$^{+}$, *MUC5B*$^{+}$, *SERPINA1*$^{+}$ (Fig. 2I), suggesting they differentiate into goblet or mucous cells. By contrast, (vi) proximal secretory 3 (*SCGB1A1*$^{+}$, *SCGB3A2*$^{LO/-}$, *SCGB3A1*$^{+}$) was detected from 15 pcw and was *SPDEF*$^{-}$ (Fig. 2A-C), but *CYP2F1*$^{+}$, *MUC4*$^{+}$ and *KRT4*$^{+}$ (Fig. 2G,J). All three luminal proximal secretory cell populations were located in the proximal cartilaginous

airways and were *MUC16*$^{+}$ (Fig. 2C,G,H; S4E,F). Detailed spatial-temporal analysis of 10-21 pcw airways revealed that the proportion of proximal secretory progenitors decreased with developmental age, whilst proximal secretory cells 1 and 2 increased (Fig. S5A-C); consistent with a progenitor function for proximal secretory progenitors.

Other airway cells

We detected ciliated cells (*FOXJ1*$^{+}$*, ALOX15*$^{+}$) from 11 pcw, interspersed with secretory/club cells throughout the airways (Fig. 2A-C; S4H; S5A,B). Rarer deuterosomal cells (*FOXJ1*$^{+}$*, CDC20B*$^{+}$) appeared at the same time (Fig. 2A-C). MUC16^{+} ciliated cells (*FOXJ1*$^{+}$, *DNAH*$^{+}$*, MUC16*^{LO}) were also detected from 11 pcw, but confined to proximal dissected regions (Fig. 2A-C; S2H). They were located in patches in the most proximal cartilaginous airways (Fig. S4I), and likely represent *MUC16*$^{+}$ secretory cells generating ciliated cells, as suggested in the adult (Deprez et al. 2020; Carraro et al. 2021; Vieira Braga et al. 2019). Basal cells (*TP63*$^{+}$*, F3*$^{+}$) were present from 9 pcw (Fig. 2A-C; S4J) and more frequent in proximal regions (Fig. 2C; S2H; S5A,B). Rarer cells (ionocytes, tuft) that have been consistently identified in adult airways were not present in our single cell data. However, we found putative ionocytes (*FOXI1*$^{+}$; 4/4 lungs) and putative tuft cells (*POU2F3*$^{+}$; 2/4 lungs) in the most proximal cartilaginous airways of 21-22 pcw lung sections (Fig. S5E)**,** suggesting they begin to differentiate mid-gestation. Moreover, we reproducibly detected a small population of *MUC5AC*$^{+}$, *ASCL1*$^{+}$ cells in 9,11 pcw lungs (Fig. 2A-C). These were localised to the proximal non-cartilaginous airways where they appeared as solitary, somewhat basal, non-columnar cells (Fig. S4K). We hypothesise that they are an unknown progenitor, consistent with their transient appearance and the observation that Ascl1^{+} NE cells in adult mice can generate club, ciliated and mucous cells following injury (Yao et al. 2018; Ouadah et al. 2019).

Predicted airway epithelial differentiation trajectories

A detailed spatio-temporal analysis of the major airway epithelial cell types from 10-21 pcw confirms that cell maturation begins in the more proximal regions. An example is lack of ciliated and club cells in the distal non-cartilaginous airways at 10-12 pcw, but presence at 15-21 pcw (Fig. S5A-D). Conversely, airway progenitors are found throughout the non-cartilaginous airways at 10-12 pcw, but restricted to the terminal airways by 15-21 pcw (Fig. S5A-D). In addition, proximal

secretory cells are spatially restricted to the cartilaginous airways, whilst club cells are found in the non-cartilaginous regions (Fig. S5A-D).

This spatial separation means that predicted differentiation trajectories which combine proximal secretory cells and club cells (Fig. 2D) can reveal general trends, but are likely to be over-simplified. We therefore predicted mid- (Fig. S6A-C) and late-stage (Fig. S6D-F) airway lineage trajectories separately. In both cases, basal cells formed discrete clusters on the UMAPs (Fig. S6A',D'). Trajectory inference analysis suggests a differentiation route from mid-tip to stalk to airway progenitors to proximal secretory progenitors and proximal secretory cells (Fig. S6B), consistent with sample age (Fig. S6B'). Visualising gene expression along the inferred trajectory shows mid-tip and stalk cells are similar (Fig. S6C). The stalk cells lose some tip markers, including *FOXP2* and *SOX9*, and gain a relatively small number of genes including *PDPN* and *AGER*. By contrast, the newly-defined airway progenitors upregulate marker genes associated with airway fates, including *CYTL1*, *CLDN4* and *SCGB3A2* (Kaarteenaho et al. 2010; Guha et al. 2012) (Fig. S6C). A similar differentiation trajectory was predicted from late-tip to late-stalk to late-airway progenitor to club cells (Fig. S6E), although the oldest tip and stalk cells included in this analysis may produce alveolar lineages (Fig. S6E'; 3C-E; S7A,B). Visualising gene expression along the inferred late-airway trajectory shows that the late-tip and stalk cells are transcriptionally similar and undergo gene expression changes analogous to mid-tip and stalk (loss of *SOX9*, *FOXP2*; gain of *PDPN*, *AQP5*; Fig. S6F).

These analyses predict that cells first exit the tip into a tip-adjacent stalk-state, followed by gain of airway progenitor identity before commitment to a specific differentiation state that likely depends on local signalling cues. Although we cannot predict the origin of the basal cells using trajectory inference methods, we hypothesise that they are derived from a columnar progenitor (possibly the airway progenitor), but will themselves act as progenitor/stem cells following differentiation analogous to previous observations in mice (Yang et al. 2018).

Our trajectory inference (Fig. S6A-F) predicts that airway progenitors will differentiate readily to airway cell types. At 9-10 pcw, *CYLT1*$^{+}$, *SCGB3A2*$^{+}$ airway progenitors are found throughout the airway tree (Fig. 2E; S4B,D,L; S5A,B,D). We devised a strategy to isolate airway progenitors

using a combination of distal non-cartilaginous airway micro-dissection and transduction with a lentiviral *SCGB3A2* transcriptional reporter (*SCGB3A2-GFP*, Fig. S6G). Freshly-isolated distal *SCGB3A2*-GFP$^+$ cells were *SOX9*^{LO}, *CYTL1*^{HI}, *SCGB1A1*^{LO}, *SCGB3A2*^{LO}, and *SCGB3A1*^{LO} compared to tip/stalk cells and more proximal *SCGB3A2*-GFP$^+$ cells from the same lungs (Fig. S6H), confirming airway progenitor identity. When single cells were placed into an FGF-containing differentiation medium (Hawkins et al. 2021), distal *SCGB3A2*-GFP$^+$ cells produced basal, ciliated and mature secretory cells (Fig. S6I-K). This demonstrates that, consistent with the trajectory analysis, the airway progenitors are competent to differentiate into airway lineages.

In summary, we have identified multiple epithelial progenitor states (tip; stalk; airway progenitor; proximal secretory progenitor) and differentiating airway cells which localise to a spatial differentiation gradient along the proximal-distal axis of the epithelium (summarised in Fig. S4L; S5D). Moreover, we identify *GHRL*$^+$ neuroendocrine cells which do not exist in the mouse.

Late epithelial tip cells acquire alveolar identity prior to alveolar epithelial differentiation

The tip cells expressed a core set of tip-specific markers (*SOX9*$^+$, *ETV5*$^+$, *TESC*$^+$, *TPPP3*$^+$, *STC1*$^+$) at all stages sampled (Fig. 2A-C). We observed a gradual decrease in tip marker expression and an increase in alveolar type 2 (AT2) cell gene expression in tip cells with developmental age (Fig. 2C). Indeed, by 15 pcw the AT2 markers *SFTPC* and *SFTPA* were detected readily in the late-tip cells where they were co-expressed with lower levels of core tip markers (Fig. 3A,B). This late-tip is a unique tip cell transcriptional state which has not been detected in developing mouse lungs (Negretti, Plosa, Benjamin, Schuler, Habermann, et al. 2021; Zepp et al. 2021). The change in tip gene expression correlates with predicted differentiation trajectories from early to mid to late-tip cells to (late-stalk to) fetal AT2 and AT1 cell fates (Fig. 3C,D; without late-stalk in S7A). However, trajectory inference analysis at this transitional developmental stage is challenging. It is likely that some of the late-tip cells are producing the terminal branches of the conducting airways (Fig. S6D-F). Moreover, the inferred connections between mid-tip and late-tip cells are weak (Fig. 3C) and we cannot exclude a novel origin for late-tip cells perhaps emerging as new buds from a stalk position, although this hypothesis is not strongly supported by our analysis (Fig. 2D). Nevertheless throughout this period, similar to earlier stages, the late-tip cells remain SOX9$^+$ and the late-stalk cells turn off tip markers and acquire *PDPN/AGER* (Fig. S4A; S7D).

A small number of AT2 cells appear in the single cell data from 15 pcw, but are more prominent from 22 pcw (Fig. 2A). Similarly, at 16 pcw late tip cells (*SOX9*$^{+}$*,TPPP3*$^{+}$*,SFTPC*$^{+}$) were clearly visualised in the tissue, but differentiating AT2 cells (*SOX9*$^{LO/-}$*,TPPP3*$^{LO/-}$*,SFTPC*$^{+}$) were rare suggesting AT2 production is just beginning (Fig. 3F,G; S7C). Over the following weeks, the size of the tip regions decreased and more differentiating AT2 cells (*SOX9*$^{LO/-}$*,TPPP3*$^{LO/-}$*,SFTPC*$^{+}$) were detected (Fig. 3F,G). At 21 pcw smaller numbers of late tip cells persist and AT2 cells (*SOX9*$^{-}$*, SFTPC*$^{+}$*, NASPA*$^{+}$*, ETV5*$^{+}$*)* were found scattered throughout the developing air sacs (Fig. 3H, S7E-J). Consistent with the predicted change in tip fate potential (Fig. 3C-E), late-tip cells (16-20 pcw) grown as organoids retained a late-tip phenotype *in vitro* and were much more readily differentiated to mature AT2s than organoids derived from earlier developmental stages (Lim et al. 2021).

In our single cell atlas, differentiating AT1 cells were first visible at 18 pcw, but more prominent by 22 pcw (Fig. 2A-C). Similarly in tissue sections, AT1 cells were not detected at 17 pcw (Fig. S7H). However, by 20 pcw differentiating AT1 cells (*SPOCK2*^{LO}*, SFTPC*$^{-}$) were visible and at 21 pcw AT1 cells (*SPOCK2*$^{+}$*, SFTPC*$^{-}$) were interspersed with AT2 cells lining the developing air sacs (Fig. 3I; S7I,J). In sections, AT1 markers were only detected in cells which had undetectable, or extremely low levels of, *SFTPC* (Fig. S7H-J). Moreover, *SFTPC*-negative cells were always observed in the stalk regions from 16 pcw onwards (Fig. 3F). These spatial expression data are consistent with an alveolar epithelial differentiation model in which from ~16 pcw the late-tip progenitors first exit the tip state, turning off co-expressed AT2 cell markers, and enter the late-stalk cell state, prior to initiating AT1 or AT2 cell differentiation in response to local signalling cues (Fig. 3J). Furthermore, the late-stalk cells are connected to AT2, AT1 and late airway progenitors in trajectory inference analysis (Fig. 2D; 3C,D), supporting our hypothesis that at all stages of lung development, cells exit the tip and enter a stalk-state prior to differentiation. This model for spatial patterning of human alveolar development is different from the current prevailing mouse model in which AT2 and AT1 cells are thought to be specified early in development (Zepp et al. 2021; Frank et al. 2019).

Integration of our fetal atlas with adult data revealed high correlation between expected groups: fetal airway progenitors with adult secretory club cells; fetal and adult ciliated and deuterosomal cells; proximal secretory fetal cells with adult goblet cells (Fig. S3A). The AT2 and AT1 cells we detect in the fetal lungs cluster closely with the adult (Fig. S3A; Pearson correlation coefficients: fetal-adult AT2 0.66; AT1 0.80). However, the fetal cells are immature and differ in gene expression to their adult counterparts (for example, for AT2/1, Fig. S3G).

Lung endothelial cells exhibit early specialisation into arterial and venous identities

At 5-6 pcw, the endothelial cells (ECs) comprised capillary (early Cap: *THY1*$^{+}$*, CD24*$^{+}$), GRIA2^{+} arterial (*GRIA2*$^{+}$*, GJA5*$^{+}$) and lymphatic ECs (*PROX1*$^{+}$*, STAB1*$^{+}$*, UCP2*^{LO}) (Fig. S8A-C); showing that capillaries and lymphatic vessels are distinct from the earliest stages of lung development and that arterial specification begins prior to venous. With increasing age, the capillary EC lineage moves from early (*THY1*$^{+}$*, CD24*$^{+}$*, EGLN1*$^{+}$) to mid (*CA4*$^{+}$*, KIT*$^{+}$*, EGLN1*$^{+}$) to late (*CA4*$^{+}$*, KIT*$^{+}$) sub-types (Fig. S8A-C), similar to the age transitions in other compartments. Trajectory analysis predicts that both mid- and late-Cap cells generate arterial and venous ECs (Fig. S8G,H). Aerocytes (*CA4*^{LO}*, S100A3*$^{+}$), capillary ECs specialised for gas exchange and leukocyte trafficking (Gillich et al. 2020; Vila Ellis et al. 2020), were observed at 20-22 pcw (Fig. S8A-C) arranged around the developing air sacs (Fig. S8D). Microvasculature specification therefore occurs relatively late in human fetal life coincident with the development of AT1 cells.

Broad markers of arterial and venous specification were clear in sections at 20 pcw (Fig. S8E). Three distinct arterial ECs were detected. GRIA2^{+} and arterial ECs (*DKK2*$^{+}$*, SSUH2*$^{+}$) form a continuous differentiation trajectory in pseudotime (Fig. S8G,H) with GRIA2^{+} ECs likely to be a more immature form. The OMD^{+} ECs (*GJA5*$^{+}$*, DKK2*$^{+}$*, PTGIS*$^{+}$*, OMD*$^{+}$) cluster with arterial ECs, are more proximal (Fig. S2H) and line the larger arterial vessels (Fig. S11B). By contrast, venous ECs (*PVLAP*$^{+}$*, ACKR3*$^{+}$*, HDAC9*$^{+}$) do not have clear subclusters (Fig. S8A-C). Systemic and pulmonary circulation ECs have been found in adult lungs (Schupp et al. 2021); we cannot detect these in fetal lungs.

Two major lymphatic ECs were detected, lymphatic ECs (*PROX1*$^{+}$, *STAB1*$^{+}$, *UCP2*^{LO}) and SCG3^{+} lymphatic ECs (*PROX1*$^{+}$, *SCG3*$^{+}$), with an intermediate population connecting them (Fig. S8A-C,F). SCG3^{+} lymphatic ECs resemble a lymphatic valve population (Takeda et al. 2019).

Haematopoietic cell types in the developing lung

At the early stages (5-6 pcw) when arterial, capillary and lymphatic ECs were present, embryonic erythrocyte, HMOX1^{+} erythroblast and a small number of macrophages and ILC progenitors were detected, representing the early progenitors of haematopoiesis. After 11 pcw relative numbers of lymphoid and myeloid cells increased, dominated by macrophages, ILCs, dendritic, NK, T and B cells (Fig. 1C, S2A,B, S9A-C,F,G,K). Immature T cells are largely absent from the atlas, consistent with the restriction of T cell development to the thymus. In contrast, a range of early B cell precursors and the ILC precursor were detected. We enriched TCR and BCR fragments from our scRNA-seq libraries which supported cell-type identities and subdivision (Fig. S9D-E,H-J). To look for lung-specific features of the leukocyte cells, we compared our atlas with a pan-fetal human atlas (Cao et al. 2020). Unlike epithelial, endothelial and fibroblasts, leukocytes in our atlas are transcriptionally highly similar to those of other organs with minimal evidence of lung-specificity (Fig. S9L).

Developmental trajectories of mesenchymal cells

The broad fibroblast cluster comprises fibroblasts, myofibroblasts, airway and vascular smooth muscle (ASM and vSMC), pericytes, mesothelium and chondrocytes (Fig. 4A,B). Airway fibroblasts and chondrocytes were proximally-enriched; mesothelium distally-enriched (Fig. S2H; 4D). There is a distinct separation of cell clusters by age (Fig. 4C). Airway SM cells were observed from 9 pcw, consistent with previous immunostaining (Nikolić et al. 2017), and showed increasing maturity over time (Fig. 4A-C). Two distinct populations of vSMC were observed throughout the time course: vSMC1 (*NTRK3*$^{+}$, *NTN4*$^{+}$, *PLN*$^{-}$) and vSMC2 (*NTRK3*$^{+}$, *NTN4*$^{+}$, *PLN*$^{+}$) (Fig. 4A,B) and were intermingled around the same vessels on tissue sections (Fig. S10A,C). vSMC1 was enriched in genes relating to ECM organisation and cell adhesion, whereas vSMC2 was enriched for transcripts encoding contractility proteins and signalling molecules (Fig. S10B). Intermingling of vSMC subtypes with different levels of contractility proteins is seen in adult lungs (Frid et al. 1997); our developmental observation suggests that these represent normal functional/ontological

differences, rather than pathology. Pericytes (*FAM162B*$^{+}$) were visualised adjacent to the microvascular endothelium (Fig. S10D).

The most common cells isolated from 5-15 pcw lungs were fibroblasts (Fig. 1C). At 5-6 pcw, early-fibroblasts (*SFRP2*$^{+}$,*WNT2*$^{+}$) predominated, although multiple populations were detected (Fig. 4A,B). In 9,11 pcw lungs, early-fibroblasts had matured into mid-fibroblasts (*WNT2*$^{+}$, *FGFR4*^{LO}) which have recently been shown to promote epithelial tip cell fate (Hein et al. 2022). In the most mature lungs sequenced, there were three distinct fibroblast populations: adventitial (*SFRP2*$^{+}$, *PI16*$^{+}$), airway (*AGTR2*$^{+}$, *S100A4*$^{+}$) and alveolar (*WNT2*$^{+}$, *FGFR4*$^{+}$) with distinct spatial locations (Fig. 4A,B,D-H). In addition, an intermediate fibroblast population connected the more mature fibroblasts on the UMAP (Fig. 4A,B), possibly representing a transitional state. Pseudotime analysis predicted a differentiation hierarchy from the early and mid fibroblasts to adventitial fibroblasts; with alveolar and airway fibroblasts forming separate branches (Fig. 4I-K). Alternatively, the intermediate fibroblast population may indicate plasticity in the fibroblast lineage in normal development as previously suggested (Kumar et al. 2014).

The three major fibroblast types in 15-22 pcw lungs expressed high levels of genes associated with ECM organisation, but had distinct gene expression patterns and spatial localisation. Adventitial fibroblasts (*SFRP2*$^{+}$, *PI16*$^{+}$) surrounded the larger blood vessels (Fig 4D). They formed diffusely arranged layers of cells surrounding the tightly packed concentric rings of ECs, pericytes and smooth muscle (Fig 4E, S10C). Adventitial fibroblasts were enriched in gene expression associated with ECM organisation and signalling, including BMP, TGFβ, WNT (Fig. 4J,K; S10E,F) consistent with described roles providing structural support to the perivascular region (Dahlgren and Molofsky 2019). Alveolar fibroblasts (*WNT2*$^{+}$, *FGFR4*$^{+}$) were observed throughout the lung, particularly surrounding the tip cells and close to the microvasculature (Fig 4F). They were enriched in genes associated with actin organisation, focal adhesions and morphogenesis, as well as signalling molecules (Fig. 4J,K; S10E,F). Adventitial and alveolar fibroblasts shared key markers such as collagens, but also expressed unique genes (adventitial: *SERPINF1, SFRP2, PI16*; alveolar: *FGFR4, VEGFD;* Fig. 4K). By contrast, the airway fibroblasts (*AGTR2*$^{+}$, *S100A4*$^{+}$, note that *S100A4* is expressed in various immune and airway epithelial cells) were adjacent to the airway smooth muscle and highly enriched in signalling molecules associated with morphogenesis

(Fig. 4D,G,J,K; S10E,F). We did not detect lipofibroblasts (Travaglini et al. 2020), meaning that they are either exceptionally rare, or form later than 22 pcw, or do not form distinct clusters in all lung data sets (Madissoon et al. 2021). Endothelial and fibroblast populations align well between fetal and adult data (Fig. S3B,C), but with some unique developmental states, such as fetal early/mid-fibroblasts and myofibroblasts.

Myofibroblasts formed three distinct groups in our single cell data. Myofibroblast 1 (*CXCL14*$^{+}$, *KCNK17*$^{+}$, *CT45A3*$^{+}$, *THBD*^{LO}) appeared at 9 pcw and persisted to 20 pcw. Myofibroblast 2 (*CXCL14*$^{+}$, *KCNK17*$^{+}$, *CT45A3*$^{+}$, *THBD*^{HI}) and myofibroblast 3 (*CXCL14*$^{+}$, *KCNK17*$^{+}$, *CT45A3*$^{-}$, *THBD*$^{-}$) were predominantly identified at 22 pcw (Fig. 4A,B). Throughout development, myofibroblasts (*CXCL14*$^{+}$, *KCNK17*$^{+}$) were visualised surrounding the developing stalk region of the epithelium, suggesting a close signalling relationship (Fig. 4D,H, S10G,H). Although not detected in significant numbers in the scRNA-seq data until 22 pcw, we see myofibroblast 2 (PDGFRA^{+}, THBD^{HI}, *NOTUM*$^{+}$) around the stalk epithelium from 15 pcw (Fig. 4D; S10H,J,K), the same position as myofibroblast 1. The appearance of myofibroblast 2 is coincident with the acquisition of AT2 markers by the late tip cells, and it may be a more mature state of myofibroblast 1. Myofibroblast 2 was enriched in gene expression associated with cell contractility and focal adhesions, as well as WNT signalling (Fig. S10J,K). Co-expression of the Wnt-responsive genes *LEF1*, *NOTUM* and *NKD1* suggests that myofibroblast 2 is responding to local Wnt expression (*WNT2* is high in alveolar fibroblasts) and producing the secreted Wnt inhibitor NOTUM; potentially to regulate local cell patterning. We tested this hypothesis using co-culture experiments where myofibroblast 2 was shown to both respond to WNT and to modulate the Wnt-response of co-cultured epithelial cells (Lim et al. 2021). *In vivo,* it is likely that myofibroblast 2 modulates the WNT2 signal from the alveolar fibroblasts mediating spatial patterning of epithelial AT2 and AT1 fate in human lung development. By contrast, myofibroblast 3 has higher expression of genes associated with ECM organisation and a variety of signalling molecules, including *C7*, *RSPO2* and *BMPER* (Fig. S10J). Myofibroblast 3 was always localised to the developing air sacs (Fig. S10I), rather than the stalk epithelium, and are likely to be precursors of the alveolar myofibroblasts (Li et al. 2020; Zepp et al. 2021).

Signalling niches in lung development

We used CellPhoneDB (Efremova et al. 2020) to analyse cell-cell communication with the aim of predicting signalling interactions controlling cell fate allocation. We focused on 15-22 pcw cells and, based on the spatial localisation of the major fibroblast populations (Fig. 4E-G), analysed signalling within three niches, defined as - Airway niche: airway fibroblasts; late airway SMCs; airway epithelial cells. Alveolar niche: alveolar fibroblasts, aerocytes, late Cap cells, late tip cells, AT1, AT2. Adventitial niche: adventitial fibroblasts, arterial endothelium, OMD^{+} endothelium, vascular smooth muscle cells. CellPhoneDB predicts numerous signalling interactions (Supplemental Table 2) which we curated by plotting the expression of ligand-receptor pairs representing major signalling pathways (Fig. 5A,B; S11A). We observed expected interactions, including high levels of Notch ligands and receptors and CXCL12-CXCR4 signalling in the advential niche (Fig. S11A,B) (Herbert and Stainier 2011). Similarly, expected signalling predicted in the alveolar niche included aerocytes to late cap cells (ALPN-ALPNR) and alveolar epithelial cells to microvascular ECs (VEGFA-FLT1/FLT4/KDR) (Fig. 5B; S11B) (Gillich et al. 2020; Vila Ellis et al. 2020).

Airway fibroblasts were predicted to signal via TGFβ3 and BMP4 to the airway epithelium, consistent with roles for these signals in human basal cell specification and differentiation (Miller et al. 2020; Mou et al. 2016). Airway fibroblasts and ASM were also predicted to signal to the epithelium via FGF7/18 to FGFR2/3 and non-canonical WNT5A to FZD/ROR (Fig. 5A). By contrast, although FGF and WNT signalling interactions were predicted in the alveolar niche, interactions were based on lower levels of *FGF*, but higher levels of canonical *WNT2* and its receptor (Fig. 5B). The predicted FGF and WNT signalling interactions in the alveolar niche/late tip cells are consistent with the requirement of these factors for long-term self-renewal of human distal tip organoids (Nikolić et al. 2017; Lim et al. 2021). Tissue staining showed that although *FGF7* is expressed fairly ubiquitously, the airway fibroblasts and ASM form a distinct barrier between the airway epithelium and the *WNT2* expression (Fig. 5C-E). Based on these data, we predicted that removing canonical WNT, but retaining FGF signalling would promote airway differentiation in the human distal tip organoids (Fig. 5F). Indeed, we observed robust basal, secretory and ciliated cell differentiation in response to FGF-containing medium (Fig. 5G,H).

scATAC-seq analysis identifies putative cell fate regulators

Single cell ATAC-seq provides an independent method of assessing cell type based on open chromatin regions and allows cell type-specific TFs to be predicted. After tissue dissociation, the single cell suspensions were split and half of the cells processed for nuclear isolation and scATAC-seq (Fig. 1A). Following quality control and doublet removal, 67 scATAC-seq clusters comprising ~100K cells were obtained and a label transfer process was used to annotate scATAC-seq clusters based on our scRNA-seq data (Fig. 6A). Not every cell state detected by scRNA-seq was distinguishable by scATAC-seq, consistent with previous work (Domcke et al. 2020; Cao et al. 2020). For example, separate early-tip, stalk and airway progenitor clusters were discerned by scRNA-seq (Fig. 2A), but a combined cluster with strong similarity to all three cell types was detected by scATAC-seq (Fig. 6A). Similarly, the resolution of scATAC-seq allowed us to identify a combined AT1/AT2 cluster and single arterial endothelial, vascular smooth muscle, myofibroblast and basal cell clusters (Fig. 6A). Nevertheless, there was broad agreement between the scRNA-seq and ATAC-seq data in terms of capturing cell types, including many of the novel/lesser-known cell types we identified by scRNA-seq (mid and late tip, mid and late airway progenitors, GHRL$^+$ NE, MUC16$^+$ ciliated, dueterosomal, airway fibroblasts, aerocytes, SCG3$^+$ lymphatic endothelial cells).

We analysed TF binding motifs in the unique/enriched open chromatin regions in each cluster and plotted the top 5 TF motifs per cell type (Fig. S12). As expected, TFs belonging to the same family are frequently enriched in the same cell type due to similarities in their binding motifs. This analysis revealed some expected TF signatures, for example TCF21 in the alveolar, adventitial and airway fibroblasts (Quaggin et al. 1999), GRHL and FOXA1/2 in epithelium (Gao et al. 2013; Wan et al. 2005), and SOX17 in arterial endothelium (Corada et al. 2013). Examining epithelial cells and focussing on TFs expressed in the corresponding cell type in the scRNA-seq data (Fig. 6B,C, marked by asterisk in B), TEAD motifs were enriched in mid-stalk cells consistent with a key role for Yap (van Soldt et al. 2019), NKX2.1 in AT1/AT2 cells (Kimura, Ostrin, and Chen 2019), KLF factors in secretory and AT1/AT2 (Liberti et al. 2022) and TP63 in basal cells (Rock et al. 2009). Unexpected TF signatures included HNF1B in late tip cells, and ZBTB7A in early tip/stalk/airway progenitors. We focused on the pulmonary and GHRL$^+$ NE cells which cluster closely in both data sets (Fig. 2A, 6A). ASCL1 is required for mouse NE cell differentiation (Ito et al. 2000; Borges et al. 1997) and this motif is strongly associated with both pulmonary and

GHRL$^+$ NE cells (Fig. 6B). However, both cell types also have specific TF motifs including NEUROD1 and RFX6 in the GHRL$^+$ NEs, and TCF4 and ID in the pulmonary NEs (Fig. 6B). Consistent with this, there are distinct, unique regions of open chromatin, especially in the neighbourhood of cell-type specific genes such as GRP and GHRL (Fig. 6D).

We have produced a high-resolution scATAC-seq data set for the developing human lungs which is highly consistent with our scRNA-seq data. Mining this data provides hypotheses for lineage-determining TFs in lung development. As further scATAC datasets become available for control and diseased adult lungs, our data will provide a resource for comparing normal and aberrant TFs and chromatin regulation.

Transcriptional control of neuroendocrine cell subtype formation

Pulmonary NE and GHRL$^+$ NE cells share the expression of many TFs and open chromatin regions, but are transcriptionally distinct. In our scRNA-seq data, they were both observed along a maturation trajectory (from what are labelled as precursors), shared classical NE markers (*CHGA, SYP*), but differed in TF and hormone expression (Fig. 7A,B). A third NE population (intermediate NE) clustered between pulmonary and GHRL$^+$ NE cells with intermediate gene expression (Fig. 7A,B), although it did contain a small number of cells expressing the unique marker *NEUROG3.* Pseudotime trajectory analysis suggested that pulmonary NE and GHRL$^+$ NE cells were derived from airway progenitors/stalk cells and that intermediate NEs are an additional transition population (Fig. S13A,B). Transition states between pulmonary NE and GHRL$^+$ NE were observed in sections (Fig. S13C). We therefore postulated that Pulmonary NE precursors could acquire *NEUROG3* and convert to GHRL$^+$ NE fate (Fig. 7C), or vice-versa - GHRL$^+$ precursors converting to pulmonary NE fate. In sections, *ASCL1* was co-expressed with *GRP*, but rarely with *GHRL*. We also observed *ASCL1* single-positive cells, likely representing pulmonary NE precursors (Fig. 7D). *NEUROD1* was co-expressed with *GHRL*, but also observed with *GRP* (Fig. 7E). Whereas *NEUROG3* was co-expressed with *ASCL1* and/or *NEUROD1*, supporting a role in a transition population (Fig. S13D).

Differential expression of *ASCL1* and *NEUROD1* defines A- and N-type human small cell lung cancer (SCLC), which likely derive from NE cells (Gay et al. 2021). Interestingly, these two TFs

coincide with the scRNA-seq marker genes and scATAC-seq TF motif enrichment of our fetal NE cells (Fig. 6B,7B). We generated SCLC feature gene lists (Borromeo et al. 2016) and performed gene signature scoring, showing that the A-type signature resembles pulmonary NEs, whereas the N-type resembles GHRL$^+$ NEs (Fig. 7F). These data suggest that either there are two different NE cells of origin for human SCLCs, or that SCLCs reuse developmental mechanisms as suggested by some mouse models (Ireland et al. 2020). We have been unable to detect GHRL$^+$ NEs in the adult airways using HCR (5 biological replicates). However, a small number of GHRL$^+$ cells are present within a tuft cell cluster in an integrated adult lung cell atlas containing 2.2 million cells (Sikkema et al. 2022), suggesting that GHRL$^+$ NEs could be a rare cell state in the adult airways. Given their relevance to human disease states, we sought to use our single cell atlas to identify lineage-defining TFs controlling NE cell differentiation and test these predictions using our organoid system. We reasoned that overexpression of lineage-defining transcription factors in lung tip organoids (Nikolić et al. 2017; Sun et al. 2021) would promote cell type-specific differentiation.

Multiple TFs were differentially expressed between pulmonary NE and *GHRL*$^+$ NE cells (Fig. 7B). We used SCENIC analysis of gene regulatory networks (GRNs) (Aibar et al. 2017) along a predicted airway progenitor to GHRL$^+$ NE trajectory (Fig. S13A,B) to identify putative lineage-defining TFs (Fig. 7G). ASCL1, NEUROD1 and NEUROG3 all emerged as potential key nodes and are required for endocrine cell differentiation in various organs (Borges et al. 1997; Ito et al. 2000; Mellitzer et al. 2010; Naya et al. 1997). We also selected the *GHRL*$^+$ NE-specific *RFX6* (Fig. S13E) and *NKX2.2* (Fig. 7B), the pan-NE *PROX1* (Fig. 7B) and, as controls, the basal cell-specific TFs *DeltaNTP63*, *TFAP2A*, *PAX9,* and *mNeonGreen-3xNLS*. Overexpression of *PROX1* or *NKX2-2* did not result in NE gene upregulation based on qRT-PCR (not shown) and these TFs were not followed up. The other factors resulted in increased expression of basal or NE markers compared to *mNeonGreen-3xNLS* controls and the experiments were repeated using scRNA-seq. Individual TFs were overexpressed from a doxycycline-inducible construct for 3 days and organoids were maintained in the self-renewing (tip cell-promoting) medium throughout to rigorously assay the lineage-determining competence of the TF (Fig. 7H; S14A), followed by scRNA-seq.

When mapped to epithelial cells of our fetal lung atlas, the majority of the *mNeonGreen-3xNLS* expressing organoid cells projected to mid-tip or -stalk cells as expected (Fig. 7I). Whereas, overexpression of *DeltaNTP63* resulted in basal cell-like lineages (S14B) consistent with a previous report (Warner et al. 2013), indicating that this simple assay can report TF function. Overexpression of *RFX6*, *TFAP2A* or *PAX9* did not result in the predicted lineage progression at a transcriptome level (S14B). However, *ASCL1*-overexpressing organoids progressed into pulmonary NE precursors (Fig. 7I) and *NEUROD1* overexpression promoted differentiation into *GHRL*$^+$ NE precursors (Fig. 7I). *NEUROG3* overexpression also led to *GHRL*$^+$ NE precursor formation (Fig. S14B), suggesting that the *GHRL*$^+$ NE lineage is the destination of the intermediate NE population (Fig. 7C).

The 5' differences between the transgenes and endogenous TFs allowed us to distinguish these transcripts and infer gene regulation hierarchy. We observed autoregulation of *ASCL1*, *NEUROD1*, *NEUROG3* and *RFX6* (Fig. S14C). By contrast, *NKX2-2* and *PROX1* were upregulated by other TFs, indicating they are relatively low in the hierarchy (Fig. S14C). *NKX2-2* and *PROX1* expression in the organoid assay matched their expression in NE cells *in vivo* (Fig. 7B, S14C), showing that this assay recapitulated key features of the TF network. These experiments have allowed us to test gene GRN predictions from the single cell atlas, confirm the predicted lineage trajectory and provide a foundation for studying human SCLC. This is significant given that there is no evidence that GHRL$^+$ NE cells are present in mice (Borromeo et al. 2016), making the use of mouse models difficult.

Discussion

Using a combination of single cell and spatial approaches we have identified 144 cell types, or states, in the developing human lungs across the 5-22 pcw period. We take advantage of a known proximal-distal gradient in epithelial differentiation to identify progenitor and differentiating states in the developing airway, including a new neuroendocrine cell subtype related to SCLC. We suggest that human alveolar epithelial differentiation follows a tip-stalk-AT2/1 fate decision pattern that is different to the prevailing cellular models of mouse alveolar development. Moreover, analysis of the mesenchymal compartment identified three niche regions with distinct signalling interactions, allowing us to identify signalling conditions that are sufficient for airway

differentiation of human embryonic lung organoids. We tested GRN predictions for NE cell differentiation in an organoid system, allowing us to identify lineage-defining TFs and provide directionality to the inferred differentiation trajectory. This study provides a paradigm for combining single cell datasets with spatial analysis of the tissue and functional analyses in a human organoid system to provide mechanistic insights into human development.

We show that lung maturation occurs in concert across cell compartments, for example epithelial AT1 cells and endothelial aerocytes differentiating in parallel at 20-22 pcw. Moreover, we have observed many aspects of cellular differentiation *in utero* showing that they are controlled by prenatal factors, rather than the transition to air breathing with its associated mechanical and hormonal changes. For example, two distinct types of vSMCs and lymphatic endothelial cells are established prior to major alterations in blood flow that occur postnatally.

The mesenchymal compartment contains multiple fibroblast and myofibroblast cell states and we focus on those present during the later stages of lung development. Airway, adventitial and alveolar fibroblasts are all localised in distinct niche regions and participate in different signalling interactions. Airway and adventitial fibroblasts both express unique combinations of signalling molecules and also form physical barriers between the neighbouring airway epithelium, or vascular endothelium, and the widespread alveolar fibroblasts (Fig. 4,5). Similarly, we characterise a population of myofibroblasts which contacts the developing epithelial stalk region and expresses high levels of the secreted Wnt-inhibitor, *NOTUM* (Fig. S10K); whereas alveolar fibroblasts express high levels of the canonical *WNT2* ligand (Fig. 4). In a separate study, using surface markers identified in this single cell atlas, we were able to specifically isolate alveolar fibroblasts and myofibroblast 2 cells and perform co-culture experiments with late tip organoids (Lim et al. 2021). Those experiments confirmed that a three-way signalling interaction between alveolar fibroblasts, myofibroblast 2 cells and late tip cells can control human AT2 spatial patterning.

We find that GHRL^{+} NE cells are transcriptionally similar to the NEUROD1^{+} N-subtype of SCLC (Fig. 7). Our functional analyses of NE cell differentiation in organoids will provide tools to test these hypotheses. Mouse studies show that fetal transcriptional and chromatin cell states are accessed during the normal process of tissue regeneration and may contribute to neoplasm in

chronic inflammation (Larsen and Jensen 2021; Jadhav et al. 2017). Detailed ATAC-seq datasets are not yet available for human lung disease. Our high quality ATAC-seq atlas will provide a baseline for further analyses when adult chromatin accessibility lung atlases are published. In summary, our multi-component atlas is a community resource for future analyses of human development, regeneration and disease.

Limitations of the study

We provide a carefully annotated, descriptive cell atlas resource. Many conclusions are derived from trajectory inference or TF binding site analyses and require future validation. The trajectory inference analyses predict lineage relationships and transitions between cell types. However, these transitions reflect similarities in gene expression, for which direct differentiation processes are only one possible explanation. Further, they do not imply directionality. RNA velocity (used in Fig. 2D) provides directionality to trajectory inference predictions using the ratio of spliced versus unspliced mRNA and assuming steady-state transcription and degradation rates. However, over developmental time these assumptions are unlikely to hold across all developmental transitions, which can lead to the inference of incorrect directionality (Bergen et al. 2021). Similarly, when applying RNA velocity algorithms to scRNA-seq data of a known differentiation trajectory, reversed velocities have been reported (Gorin et al. 2022). For these reasons, we performed most of our trajectory inference analysis using Monocle3 (Trapnell et al. 2014). Monocle3 requires user-defined starting and end points and calculates the most likely routes between these points (shown as grey lines on the plots), guided by known biological features of the data (age and spatial arrangement of cells). A further confounding factor for trajectory inference in development is that upon maturation, some cell types are likely to act as progenitors themselves (Rawlins, Okubo, et al. 2009; Rock et al. 2009; Yang et al. 2018). Adding to the difficulty of reconstructing complete lineages, our single-cell dissociation protocols are expected to under-sample certain cells, such as mature neurons and chondrocytes. Furthermore, validation assays for lineage analysis in human systems rely on *in vitro* experiments, including organoid and iPSC differentiation. It is important to acknowledge that these usually define differentiation competence and do not necessarily mean that a specific differentiation route occurs *in vivo*.

We have performed data integration and regression analysis to compare the identity of our fetal human lung cells with adult human lung cells. There are approximately three decades between the oldest fetal and youngest adult human lung samples sequenced, including a rapid period of postnatal growth and morphogenesis, puberty and an unknown number of infections/environmental insults. Despite this, many fetal-adult similarities can be seen. Nevertheless, it will be important to sequence additional lungs and, when possible, to fill the age gap. Similarly, mouse-human fetal lung cell comparisons have discerned similarities and differences. However, the differences in experimental protocols and annotation granularity between the mouse and human data might have also contributed to the differences we see. Moreover, mice were selected as lab animals partly due to their small size and rapid gestation. It will be informative in the future to make comparisons with a range of fetal lungs, including larger, long-developing species such as pig and sheep, to distinguish between differences due to species, size and gestation period.

Acknowledgements

We would like to acknowledge the Gurdon Institute Imaging Facility, and the Cellular Genetics IT and Phenotyping group, New Pipeline Group and DNA pipelines of Sanger Institute; Menna Clatworthy and Muzz Haniffa and RE, CS, ED, IG, MH, CD, WS for discussions on cell-type annotations; MP, AP, SL and CT for informatics support. KL is supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1A6A3A03012122). DS is supported by a Wellcome Trust PhD studentship (109146/Z/15/Z) and the Department of Pathology, University of Cambridge. PH holds a non-stipendiary research fellowship at St Edmund's College, University of Cambridge. EM is supported by ESPOD fellowship of EMBL-EBI and Sanger Institute. JPP is supported by the MSCA Postdoctoral Fellowship. ELR is supported by the MRC (MR/P009581/1; MR/S035907/1). ZD is supported by a Wellcome Trust PhD studentship (222275/Z/20/Z). RAB is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) and was an NIHR senior investigator. Gurdon Institute Core support from the Wellcome Trust (203144/Z/16/Z) and Cancer Research UK (C6946/A24843). KBM, JCM and SAT acknowledge funding from the MRC (MR/S035907/1) and from Wellcome (WT211276/Z/18/Z and Sanger core grant WT206194). MZN acknowledges funding from a MRC Clinician Scientist Fellowship

(MR/W00111X/1), the Rosetrees Trust (M899) and Action Medical Research (GN2911). This work was partly undertaken at UCLH/UCL who received a proportion of funding from the Department of Health's NIHR Biomedical Research Centre's funding scheme.

Author Contributions

Conceptualization: PH, KL, DS, KBM and ELR. Methodology: PH, KL, DS. Software: PH. Formal Analysis: PH, JPP, KP, ZKT. Investigation: KL, DS, QJ, ZD, LB, LR, LM, MD, AW and MY. Resources: EM, XH, RAB, SMJ. Data Curation: PH, KL, DS, EM, ZKT, ED, CS, IG. Writing - original draft: PH, KL, DS, JPP, KBM and ELR. Writing - review and editing: PH, KL, DS, KBM, ELR, SAT, JBM. Supervision: MZN, RAB, SAT, JBM, KBM and ELR. Funding Acquisition: KL, EM, JPP, RAB, MZN, SAT, JBM, KBM and ELR.

Declaration of Interests

SAT is a member of the Scientific Advisory Board for the following companies: Biogen, Foresite Labs, GSK, Qiagen, CRG Barcelona, Jax Labs, SciLife Lab, Allen Institute. She is a consultant for Genentech and Roche. She is co-founder of Transition Bio and a member of the Board. ZKT has received consulting fees from Synteny Biotechnologies for activities unrelated to this work.

REFERENCES

Aibar, Sara, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, et al. 2017. "SCENIC: Single-Cell Regulatory Network Inference and Clustering." *Nature Methods* 14 (11): 1083–86.

Alanis, Denise Martinez, Daniel R. Chang, Haruhiko Akiyama, Mark A. Krasnow, and Jichao Chen. 2014. "Two Nested Developmental Waves Demarcate a Compartment Boundary in the Mouse Lung." *Nature Communications* 5 (1): 1–15.

Bergen, Volker, Ruslan A. Soldatov, Peter V. Kharchenko, and Fabian J. Theis. 2021. "RNA Velocity-Current Challenges and Future Perspectives." *Molecular Systems Biology* 17 (8): e10282.

Blenkinsopp, W. K. 1967. "Proliferation of Respiratory Tract Epithelium in the Rat." *Experimental Cell Research* 46 (1): 144–54.

Borges, M., R. I. Linnoila, H. J. van de Velde, H. Chen, B. D. Nelkin, M. Mabry, S. B. Baylin, and D. W. Ball. 1997. "An Achaete-Scute Homologue Essential for Neuroendocrine Differentiation in the Lung." *Nature* 386 (6627): 852–55.

Borromeo, Mark D., Trisha K. Savage, Rahul K. Kollipara, Min He, Alexander Augustyn, Jihan K. Osborne, Luc Girard, et al. 2016. "ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs." *Cell Reports* 16 (5): 1259–72.

Burri, P. H. 1984. "Fetal and Postnatal Development of the Lung." *Annual Review of Physiology* 46: 617–28.

Cao, Junyue, Diana R. O'Day, Hannah A. Pliner, Paul D. Kingsley, Mei Deng, Riza M. Daza, Michael A. Zager, et al. 2020. "A Human Cell Atlas of Fetal Gene Expression." *Science* 370 (6518). https://doi.org/10.1126/science.aba7721.

Carraro, Gianni, Justin Langerman, Shan Sabri, Zareeb Lorenzana, Arunima Purkayastha, Guangzhu Zhang, Bindu Konda, et al. 2021. "Transcriptional Analysis of Cystic Fibrosis Airways at Single-Cell Resolution Reveals Altered Epithelial Cell States and Composition." *Nature Medicine* 27 (5): 806–14.

Carraro, Gianni, and Barry R. Stripp. 2022. "Insights Gained in the Pathology of Lung Disease through Single-Cell Transcriptomics." *The Journal of Pathology* 257 (4): 494–500.

Corada, Monica, Fabrizio Orsenigo, Marco Francesco Morini, Mara Elena Pitulescu, Ganesh Bhat, Daniel Nyqvist, Ferruccio Breviario, et al. 2013. "Sox17 Is Indispensable for Acquisition and Maintenance of Arterial Identity." *Nature Communications* 4: 2609.

Cutz, E., J. E. Gillan, and A. C. Bryan. 1985. "Neuroendocrine Cells in the Developing Human Lung: Morphologic and Functional Considerations." *Pediatric Pulmonology* 1 (3 Suppl): S21–29.

731 Dahlgren, Madelene W., and Ari B. Molofsky. 2019. "Adventitial Cuffs: Regional Hubs for Tissue
732 Immunity." *Trends in Immunology* 40 (10): 877–87.

733 Deprez, Marie, Laure-Emmanuelle Zaragosi, Marin Truchi, Christophe Becavin, Sandra Ruiz
734 García, Marie-Jeanne Arguel, Magali Plaisant, et al. 2020. "A Single-Cell Atlas of the Human
735 Healthy Airways." *American Journal of Respiratory and Critical Care Medicine* 202 (12):
736 1636–45.

737 Domcke, Silvia, Andrew J. Hill, Riza M. Daza, Junyue Cao, Diana R. O'Day, Hannah A. Pliner,
738 Kimberly A. Aldinger, et al. 2020. "A Human Cell Atlas of Fetal Chromatin Accessibility."
739 *Science*, November. https://doi.org/10.1126/science.aba7612.

740 Efremova, Mirjana, Miquel Vento-Tormo, Sarah A. Teichmann, and Roser Vento-Tormo. 2020.
741 "CellPhoneDB: Inferring Cell–cell Communication from Combined Expression of Multi-
742 Subunit Ligand–receptor Complexes." *Nature Protocols* 15 (4): 1484–1506.

743 Frank, David B., Ian J. Penkala, Jarod A. Zepp, Aravind Sivakumar, Ricardo Linares-Saldana,
744 William J. Zacharias, Katharine G. Stolz, et al. 2019. "Early Lineage Specification Defines
745 Alveolar Epithelial Ontogeny in the Murine Lung." *Proceedings of the National Academy of*
746 *Sciences of the United States of America* 116 (10): 4362–71.

747 Frid, M. G., E. C. Dempsey, A. G. Durmowicz, and K. R. Stenmark. 1997. "Smooth Muscle Cell
748 Heterogeneity in Pulmonary and Systemic Vessels. Importance in Vascular Disease."
749 *Arteriosclerosis, Thrombosis, and Vascular Biology* 17 (7): 1203–9.

750 Gao, Xia, Christopher M. Vockley, Florencia Pauli, Kimberly M. Newberry, Yan Xue, Scott H.
751 Randell, Timothy E. Reddy, and Brigid L. M. Hogan. 2013. "Evidence for Multiple Roles for
752 Grainyhead-like 2 in the Establishment and Maintenance of Human Mucociliary Airway
753 Epithelium.[corrected]." *Proceedings of the National Academy of Sciences of the United*
754 *States of America* 110 (23): 9356–61.

755 Gay, Carl M., C. Allison Stewart, Elizabeth M. Park, Lixia Diao, Sarah M. Groves, Simon Heeke,
756 Barzin Y. Nabet, et al. 2021. "Patterns of Transcription Factor Programs and Immune
757 Pathway Activation Define Four Major Subtypes of SCLC with Distinct Therapeutic
758 Vulnerabilities." *Cancer Cell* 39 (3): 346–60.e7.

759 Gillich, Astrid, Fan Zhang, Colleen G. Farmer, Kyle J. Travaglini, Serena Y. Tan, Mingxia Gu,
760 Bin Zhou, Jeffrey A. Feinstein, Mark A. Krasnow, and Ross J. Metzger. 2020. "Capillary
761 Cell-Type Specialization in the Alveolus." *Nature* 586 (7831): 785–89.

762 Gorin, Gennady, Meichen Fang, Tara Chari, and Lior Pachter. 2022. "RNA Velocity Unraveled."
763 *bioRxiv*. https://doi.org/10.1101/2022.02.12.480214.

764 Guha, Arjun, Michelle Vasconcelos, Yan Cai, Mitsuhiro Yoneda, Anne Hinds, Jun Qian, Guihua
765 Li, et al. 2012. "Neuroepithelial Body Microenvironment Is a Niche for a Distinct Subset of
766 Clara-like Precursors in the Developing Airways." *Proceedings of the National Academy of*
767 *Sciences of the United States of America* 109 (31): 12592–97.

768 Hawkins, Finn J., Shingo Suzuki, Mary Lou Beermann, Cristina Barillà, Ruobing Wang, Carlos

769 Villacorta-Martin, Andrew Berical, et al. 2021. "Derivation of Airway Basal Stem Cells from
770 Human Pluripotent Stem Cells." *Cell Stem Cell* 28 (1): 79–95.e8.

771 Hein, Renee F. C., Joshua H. Wu, Emily M. Holloway, Tristan Frum, Ansley S. Conchola, Yu-
772 Hwai Tsai, Angeline Wu, et al. 2022. "R-SPONDIN2+ Mesenchymal Cells Form the Bud
773 Tip Progenitor Niche during Human Lung Development." *Developmental Cell*, June.
774 https://doi.org/10.1016/j.devcel.2022.05.010.

775 He, Peng, Brian A. Williams, Diane Trout, Georgi K. Marinov, Henry Amrhein, Libera Berghella,
776 Say-Tar Goh, et al. 2020. "The Changing Mouse Embryo Transcriptome at Whole Tissue and
777 Single-Cell Resolution." *Nature* 583 (7818): 760–67.

778 Herbert, Shane P., and Didier Y. R. Stainier. 2011. "Molecular Control of Endothelial Cell
779 Behaviour during Blood Vessel Morphogenesis." *Nature Reviews. Molecular Cell Biology*
780 12 (9): 551–64.

781 Ireland, Abbie S., Alexi M. Micinski, David W. Kastner, Bingqian Guo, Sarah J. Wait, Kyle B.
782 Spainhower, Christopher C. Conley, et al. 2020. "MYC Drives Temporal Evolution of Small
783 Cell Lung Cancer Subtypes by Reprogramming Neuroendocrine Fate." *Cancer Cell* 38 (1):
784 60–78.e12.

785 Ito, T., N. Udaka, T. Yazawa, K. Okudela, H. Hayashi, T. Sudo, F. Guillemot, R. Kageyama, and
786 H. Kitamura. 2000. "Basic Helix-Loop-Helix Transcription Factors Regulate the
787 Neuroendocrine Differentiation of Fetal Mouse Pulmonary Epithelium." *Development* 127
788 (18): 3913–21.

789 Jadhav, Unmesh, Madhurima Saxena, Nicholas K. O'Neill, Assieh Saadatpour, Guo-Cheng Yuan,
790 Zachary Herbert, Kazutaka Murata, and Ramesh A. Shivdasani. 2017. "Dynamic
791 Reorganization of Chromatin Accessibility Signatures during Dedifferentiation of Secretory
792 Precursors into Lgr5+ Intestinal Stem Cells." *Cell Stem Cell* 21 (1): 65–77.e5.

793 Kaarteenaho, Riitta, Heta Merikallio, Siri Lehtonen, Terttu Harju, and Ylermi Soini. 2010.
794 "Divergent Expression of Claudin -1, -3, -4, -5 and -7 in Developing Human Lung."
795 *Respiratory Research* 11 (May): 59.

796 Kimura, S., E. J. Ostrin, and J. Chen. 2019. "Transcriptional Control of Lung Alveolar Type 1 Cell
797 Development and Maintenance by NK Homeobox 2-1." *Proceedings of the*.
798 https://www.pnas.org/content/116/41/20545.short.

799 Kumar, Maya E., Patrick E. Bogard, F. Hernán Espinoza, Douglas B. Menke, David M. Kingsley,
800 and Mark A. Krasnow. 2014. "Mesenchymal Cells. Defining a Mesenchymal Progenitor
801 Niche at Single-Cell Resolution." *Science* 346 (6211): 1258810.

802 Larsen, Hjalte L., and Kim B. Jensen. 2021. "Reprogramming Cellular Identity during Intestinal
803 Regeneration." *Current Opinion in Genetics & Development* 70 (October): 40–47.

804 Liberti, Derek C., William A. Liberti Iii, Madison M. Kremp, Ian J. Penkala, Fabian L. Cardenas-
805 Diaz, Michael P. Morley, Apoorva Babu, Su Zhou, Rafael J. Fernandez Iii, and Edward E.
806 Morrisey. 2022. "Klf5 Defines Alveolar Epithelial Type 1 Cell Lineage Commitment during

Lung Development and Regeneration." *Developmental Cell*, July. https://doi.org/10.1016/j.devcel.2022.06.007.

Lim, K., W. Tang, D. Sun, H. Peng, S. Teichmann, J. C. Marioni, K. B. Meyer, and E. L. Rawlins. 2021. "Acquisition of Alveolar Fate and Differentiation Competence by Human Fetal Lung Epithelial Progenitor Cells." *bioRxiv*. https://www.biorxiv.org/content/10.1101/2021.06.30.450501.abstract.

Li, Rongbo, Xiaoping Li, James Hagood, Min-Sheng Zhu, and Xin Sun. 2020. "Myofibroblast Contraction Is Essential for Generating and Regenerating the Gas-Exchange Surface." *The Journal of Clinical Investigation* 130 (6): 2859–71.

Lopez, Romain, Jeffrey Regier, Michael B. Cole, Michael I. Jordan, and Nir Yosef. 2018. "Deep Generative Modeling for Single-Cell Transcriptomics." *Nature Methods* 15 (12): 1053–58.

Madissoon, Elo, Amanda J. Oliver, Vitalii Kleshchevnikov, Anna Wilbrey-Clark, Krzysztof Polanski, Ana Ribeiro Orsi, Lira Mamanova, et al. 2021. "A Spatial Multi-Omics Atlas of the Human Lung Reveals a Novel Immune Cell Survival Niche." *bioRxiv*. https://doi.org/10.1101/2021.11.26.470108.

Mellitzer, Georg, Anthony Beucher, Viviane Lobstein, Pascal Michel, Sylvie Robine, Michèle Kedinger, and Gérard Gradwohl. 2010. "Loss of Enteroendocrine Cells in Mice Alters Lipid Absorption and Glucose Homeostasis and Impairs Postnatal Survival." *The Journal of Clinical Investigation* 120 (5): 1708–21.

Miller, Alyssa J., David R. Hill, Melinda S. Nagy, Yoshiro Aoki, Briana R. Dye, Alana M. Chin, Sha Huang, et al. 2018. "In Vitro Induction and In Vivo Engraftment of Lung Bud Tip Progenitor Cells Derived from Human Pluripotent Stem Cells." *Stem Cell Reports* 10 (1): 101–19.

Miller, Alyssa J., Qianhui Yu, Michael Czerwinski, Yu-Hwai Tsai, Renee F. Conway, Angeline Wu, Emily M. Holloway, et al. 2020. "In Vitro and In Vivo Development of the Human Airway at Single-Cell Resolution." *Developmental Cell* 53 (1): 117–28.e6.

Mou, Hongmei, Vladimir Vinarsky, Purushothama Rao Tata, Karissa Brazauskas, Soon H. Choi, Adrianne K. Crooke, Bing Zhang, et al. 2016. "Dual SMAD Signaling Inhibition Enables Long-Term Expansion of Diverse Epithelial Basal Cells." *Cell Stem Cell* 19 (2): 217–31.

Naya, F. J., H. P. Huang, Y. Qiu, H. Mutoh, F. J. DeMayo, A. B. Leiter, and M. J. Tsai. 1997. "Diabetes, Defective Pancreatic Morphogenesis, and Abnormal Enteroendocrine Differentiation in BETA2/neuroD-Deficient Mice." *Genes & Development* 11 (18): 2323–34.

Negretti, Nicholas M., Erin J. Plosa, John T. Benjamin, Bryce A. Schuler, A. Christian Habermann, Christopher Jetter, Peter Gulleman, et al. 2021. "A Single Cell Atlas of Lung Development." *bioRxiv*. https://doi.org/10.1101/2021.01.21.427641.

Negretti, Nicholas M., Erin J. Plosa, John T. Benjamin, Bryce A. Schuler, A. Christian Habermann, Christopher S. Jetter, Peter Gulleman, et al. 2021. "A Single-Cell Atlas of Mouse Lung Development." *Development* 148 (24). https://doi.org/10.1242/dev.199512.

Nikolić, Marko Z., Oriol Caritg, Quitz Jeng, Jo-Anne Johnson, Dawei Sun, Kate J. Howell, Jane L. Brady, et al. 2017. "Human Embryonic Lung Epithelial Tips Are Multipotent Progenitors That Can Be Expanded in Vitro as Long-Term Self-Renewing Organoids." *eLife* 6 (June). https://doi.org/10.7554/eLife.26575.

Nikolić, Marko Z., Dawei Sun, and Emma L. Rawlins. 2018. "Human Lung Development: Recent Progress and New Challenges." *Development* 145 (16). https://doi.org/10.1242/dev.163485.

Ouadah, Youcef, Enrique R. Rojas, Daniel P. Riordan, Sarah Capostagno, Christin S. Kuo, and Mark A. Krasnow. 2019. "Rare Pulmonary Neuroendocrine Cells Are Stem Cells Regulated by Rb, p53, and Notch." *Cell* 179 (2): 403–16.e23.

Quaggin, S. E., L. Schwartz, S. Cui, P. Igarashi, J. Deimling, M. Post, and J. Rossant. 1999. "The Basic-Helix-Loop-Helix Protein pod1 Is Critically Important for Kidney and Lung Organogenesis." *Development* 126 (24): 5771–83.

Rawlins, Emma L., Cheryl P. Clark, Yan Xue, and Brigid L. M. Hogan. 2009. "The Id2+ Distal Tip Lung Epithelium Contains Individual Multipotent Embryonic Progenitor Cells." *Development* 136 (22): 3741–45.

Rawlins, Emma L., and Brigid L. M. Hogan. 2008. "Ciliated Epithelial Cell Lifespan in the Mouse Trachea and Lung." *American Journal of Physiology. Lung Cellular and Molecular Physiology* 295 (1): L231–34.

Rawlins, Emma L., Tadashi Okubo, Yan Xue, David M. Brass, Richard L. Auten, Hiroshi Hasegawa, Fan Wang, and Brigid L. M. Hogan. 2009. "The Role of Scgb1a1+ Clara Cells in the Long-Term Maintenance and Repair of Lung Airway, but Not Alveolar, Epithelium." *Cell Stem Cell* 4 (6): 525–34.

Rawlins, Emma L., Lawrence E. Ostrowski, Scott H. Randell, and Brigid L. M. Hogan. 2007. "Lung Development and Repair: Contribution of the Ciliated Lineage." *Proceedings of the National Academy of Sciences of the United States of America* 104 (2): 410–17.

Rock, Jason R., Mark W. Onaitis, Emma L. Rawlins, Yun Lu, Cheryl P. Clark, Yan Xue, Scott H. Randell, and Brigid L. M. Hogan. 2009. "Basal Cells as Stem Cells of the Mouse Trachea and Human Airway Epithelium." *Proceedings of the National Academy of Sciences of the United States of America* 106 (31): 12771–75.

Schupp, Jonas C., Taylor S. Adams, Carlos Cosme Jr, Micha Sam Brickman Raredon, Yifan Yuan, Norihito Omote, Sergio Poli, et al. 2021. "Integrated Single-Cell Atlas of Endothelial Cells of the Human Lung." *Circulation* 144 (4): 286–302.

Sikkema, L., D. Strobl, L. Zappia, E. Madissoon, N. S. Markov, L. Zaragosi, M. Ansari, et al. 2022. "An Integrated Cell Atlas of the Human Lung in Health and Disease." *bioRxiv*. https://doi.org/10.1101/2022.03.10.483747.

Soldt, Benjamin J. van, Jun Qian, Jiao Li, Nan Tang, Jining Lu, and Wellington V. Cardoso. 2019. "Yap and Its Subcellular Localization Have Distinct Compartment-Specific Roles in the Developing Lung." *Development* 146 (9). https://doi.org/10.1242/dev.175810.

Sun, Dawei, Lewis Evans, Francesca Perrone, Vanesa Sokleva, Kyungtae Lim, Saba Rezakhani, Matthias Lutolf, Matthias Zilbauer, and Emma L. Rawlins. 2021. "A Functional Genetic Toolbox for Human Tissue-Derived Organoids." *eLife* 10 (October). https://doi.org/10.7554/eLife.67886.

Takeda, Akira, Maija Hollmén, Denis Dermadi, Junliang Pan, Kevin Francis Brulois, Riina Kaukonen, Tapio Lönnberg, et al. 2019. "Single-Cell Survey of Human Lymphatics Unveils Marked Endothelial Cell Heterogeneity and Mechanisms of Homing for Neutrophils." *Immunity* 51 (3): 561–72.e5.

Trapnell, Cole, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J. Lennon, Kenneth J. Livak, Tarjei S. Mikkelsen, and John L. Rinn. 2014. "The Dynamics and Regulators of Cell Fate Decisions Are Revealed by Pseudotemporal Ordering of Single Cells." *Nature Biotechnology* 32 (4): 381–86.

Travaglini, Kyle J., Ahmad N. Nabhan, Lolita Penland, Rahul Sinha, Astrid Gillich, Rene V. Sit, Stephen Chang, et al. 2020. "A Molecular Cell Atlas of the Human Lung from Single-Cell RNA Sequencing." *Nature* 587 (7835): 619–25.

Trivedi, Vikas, Harry M. T. Choi, Scott E. Fraser, and Niles A. Pierce. 2018. "Multidimensional Quantitative Analysis of mRNA Expression within Intact Vertebrate Embryos." *Development* 145 (1). https://doi.org/10.1242/dev.156869.

Vieira Braga, Felipe A., Gozde Kar, Marijn Berg, Orestes A. Carpaij, Krzysztof Polanski, Lukas M. Simon, Sharon Brouwer, et al. 2019. "A Cellular Census of Human Lungs Identifies Novel Cell States in Health and in Asthma." *Nature Medicine* 25 (7): 1153–63.

Vila Ellis, Lisandra, Margo P. Cain, Vera Hutchison, Per Flodby, Edward D. Crandall, Zea Borok, Bin Zhou, Edwin J. Ostrin, Joshua D. Wythe, and Jichao Chen. 2020. "Epithelial Vegfa Specifies a Distinct Endothelial Population in the Mouse Lung." *Developmental Cell* 52 (5): 617–30.e6.

Wan, Huajing, Sharon Dingle, Yan Xu, Valérie Besnard, Klaus H. Kaestner, Siew-Lan Ang, Susan Wert, Mildred T. Stahlman, and Jeffrey A. Whitsett. 2005. "Compensatory Roles of Foxa1 and Foxa2 during Lung Morphogenesis *." *The Journal of Biological Chemistry* 280 (14): 13809–16.

Warner, Stephanie M. B., Tillie-Louise Hackett, Furquan Shaheen, Teal S. Hallstrand, Anthony Kicic, Stephen M. Stick, and Darryl A. Knight. 2013. "Transcription Factor p63 Regulates Key Genes and Wound Repair in Human Airway Epithelial Basal Cells." *American Journal of Respiratory Cell and Molecular Biology* 49 (6): 978–88.

Yang, Ying, Paul Riccio, Michael Schotsaert, Munemasa Mori, Jining Lu, Dong-Kee Lee, Adolfo García-Sastre, Jianming Xu, and Wellington V. Cardoso. 2018. "Spatial-Temporal Lineage Restrictions of Embryonic p63+ Progenitors Establish Distinct Stem Cell Pools in Adult Airways." *Developmental Cell* 44 (6): 752–61.e4.

Yao, Erica, Chuwen Lin, Qingzhe Wu, Kuan Zhang, Hai Song, and Pao-Tien Chuang. 2018. "Notch Signaling Controls Transdifferentiation of Pulmonary Neuroendocrine Cells in

922 Response to Lung Injury." *Stem Cells* 36 (3): 377–91.

923 Zepp, Jarod A., Michael P. Morley, Claudia Loebel, Madison M. Kremp, Fatima N. Chaudhry,
924 Maria C. Basil, John P. Leach, et al. 2021. "Genomic, Epigenomic, and Biophysical Cues
925 Controlling the Emergence of the Lung Alveolus." *Science* 371 (6534).
926 https://doi.org/10.1126/science.abc3172.

927

928

FIGURE LEGENDS

Figure 1. Experimental overview.

(A) Overview of sample collection for scRNA-seq (circles) and scATAC-seq (squares) experiments from whole lung (purple), distal (red) and proximal (blue) regions, cell processing and broad clustering; cluster number refers to the data portal (https://lungcellatlas.org).
(B) UMAP representation of ~80,000 good quality cells, indicating epithelial, endothelial, fibroblasts and leukocyte/erythroid compartments.
(C) Cell type proportions of the whole lung over developmental time.
See also Figures S1, S2, S3 and S9.

Figure 2. Epithelial cell types, states and location over developmental time

(A, B) UMAP visualisation of epithelial cells, coloured by cell types (A), stage (B, *left*), and region (B, *right*).
(C) Dot plot describing differential marker gene expression level for epithelial cells.
(D) UMAP visualising the predicted epithelial cell lineage trajectory using scvelo, inset: developmental age.
(E, F) *In situ* HCR at 11 (F) and 12 (E) pcw. (E) *SOX9* (tip epithelium, white), *CYTL1* (red), *SCGB3A2* (green). (F) *GHRL*$^{+}$ (*GHRL*$^{+}$ neuroendocrine, red), *GRP*+ (pulmonary neuroendocrine, green).
(G) Dot plot showing differential marker genes across secretory cell subtypes.
(H) *In situ* HCR at 19 pcw using *SCGB1A1* (red), *SCGB3A2* (green), and *SCGB3A1* (white).
(I-J) Differentially enriched genes in the proximal secretory cell subtypes. *SPDEF* (I, I'), *MUC5B* (I'), and *SERPINA1* (I''), *CYP2F1* (J), *MUC4* (J'), *KRT4* (J'') all white; *SCGB1A1* (red) and *SCGB3A2* (green).
DAPI, nuclei. Scale bars, 50 μm.
See also Figures S4, S5 and S6.

Figure 3. Late epithelial tip cells acquire an alveolar progenitor identity

(A, B) *In situ* HCR at 11 (A, B), 15 (B), and 19 (A) pcw. (A) *SFTPC* (green), *TPPP3* (red), *SOX9* (white). (B) *SFTPC* (green), *SFTPA1* (red), *STC1* (white). Dashed lines: tip epithelium.

(C,D) UMAP visualisation of early to late tip, late stalk, fetal AT1 and AT2 cells, coloured by cell types (C) and stages (C'); PAGA analysis (C''); Monocle3 trajectories (D).

(E) Gene expression heatmap of trajectory coloured in D.

(F) *In situ* HCR at 16, 19, and 21 pcw, *SFTPC* (green), *TPPP3* (red), and *SOX9* (white). White lines/red arrows: columnar tip prognitors, *SFTPC*$^{+}$/*SOX9*$^{+/high}$/*TPPP3*$^{+}$. Arrowheads/dashed lines in stalk/air sac regions: cuboidal differentiating fetal AT2 cells, *SFTPC*$^{+}$/*SOX9*$^{low/-}$/*TPPP3*$^{low/-}$. Asterisk (*) primitive air sacs.

(G) Quantification of cuboidal *SFTPC*$^{+}$/*SOX9*$^{low/-}$ fetal AT2 cells in stalk/air sac regions in F. The *SFTPC*$^{+}$ tip epithelial cells were excluded by their columnar morphology and marker expression (*SOX9*$^{low/-}$). Mean ± SD, n >7. Significance evaluated by 1-way ANOVA with Tukey multiple comparison post-test; ns: not significant, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

(H, I) *In situ* HCR analysis at 21 pcw. Fetal AT2 *SFTPC*$^{+}$ and *NAPSA*$^{+}$ (*arrowheads;* H, I) and fetal AT1 *SFTPC*$^{-}$/*MMP28*$^{+}$/*SPOCK2*$^{+}$ (*arrow*s; I).

(J) Diagram of the acquisition of alveolar progenitor identity by late epithelial tips, followed by differentiation to fetal AT2 and AT1 lineages.

DAPI, nuclei. Scale bars, 50 μm.

See also Figure S7 and S8.

Figure 4. Diverse mesenchymal cell types localise to distinct niches in the developing human lung.

(A) UMAP visualisation of mesenchymal cells.

(B) Dot plot describing mesenchymal differential marker gene expression.

(C) UMAP visualisation of mesenchymal cells coloured by stage.

(D) Visium spatial feature plots visualising adventitial fibroblasts, airway fibroblasts, ASPN+ chondrocytes, and myofibroblasts-2 on 17 and 20 pcw lung sections. Scores are conservative estimates of cell-type abundance per voxel.

(E-H) *In situ* HCR assay (E-H) and immunostaining (G). (E) Adventitial fibroblasts (*SFRP2,*white/*PI16,*red; arrowheads), ECs (*PECAM1*, green). (F) Alveolar fibroblasts (*WNT2* white; *FGFR4* red), tip cells (*SFTPC* green). Asterisks (* myofibroblasts). (G) Airway fibroblasts (*S100A4* red; *AGTR2* white), smooth muscle (ACTA2 green, dashed line). (H) Myofibroblasts

(*KCNK17* white, *CXCL14* red; arrowheads), tip cells (*SFTPC*, green). DAPI, nuclei. Scale bars, 50 μm.

(I) UMAP visualisation of cell types (I), stage (I') and PAGA analysis (I'') of fibroblast differentiation trajectories, .

(J, K) UMAPs with Monocle3 trajectories (J) and selected trajectory gene expression heatmaps (K) for mid tip to adventitial fibroblasts (top), alveolar fibroblasts (middle), or airway fibroblasts (bottom).

See also Figure S10.

Figure 5. Signalling ligand-receptor interactions in alveolar and airway niches.

(A, B) Curated ligand–receptor interaction predictions from CellPhoneDB in alveolar (A) and airway (B) niches. Dot plots visualise gene expression by cell type; dashed arrows indicate the predicted direction of signalling.

(C-E) Immunofluorescence/HCR. S100A4/*S100A4*, airway fibroblasts; ACTA2, ASM; CD44, tip epithelium; PECAM1, ECs. Airway fibroblasts/ASM form a boundary (dashed lines) between alveolar and airway regions. Lines are between airway fibroblasts/SMCs and airway epithelium. DAPI, nuclei. Scale bars, 50 μm.

(F) Organoids were cultured in FGF7/10-containing medium, in the presence (self-renewal medium; SNM) or absence (differentiation medium; DM) of CHIR99021, for 30 days.

(G) qRT-PCR quantifications normalised to the organoids cultured in SNM. Significance evaluated by 2-way ANOVA with Tukey multiple comparison post-test; **P*<0.05, ***P*<0.01, ****P*<0.001; *n*=6 organoid lines.

(H) Whole mount immunofluorescence of lung organoids cultured in self-renewal medium (upper) and differentiation medium (lower). DAPI, nuclei. Scale bar, 25 μm.

See also Figure S11.

Figure 6. DNA accessibility and motif enrichment revealed by scATAC-seq.

(A) Single-cell DNA accessibility profiles mapped onto 2D UMAP plane. Coloured for cell states.

(B) A union set of top 10 enriched motifs in the marker peaks among epithelial cell types/states. Statistical significance is visualised as a heatmap according to the colour bar below. Transcription factors concordantly expressed based on scRNA-seq data are marked with asterisks.

(C) Expression dotplot of the concordant transcription factors from (B) in epithelial cell types.
(D-E) Read coverage tracks of in silico aggregated "pseudo-bulk" epithelial clusters over the GRP locus (D) and GHRL locus (E).
See also Figure S12.

Figure 7. *ASCL1* and *NEUROD1* regulate the formation of two subtypes of neuroendocrine cells.

(A) Zoom-in UMAP plot of NE lineages.
(B) Dot plot showing selected gene expression in NE lineages.
(C) Schematic model of NE lineage formation.
(D) Left: HCR, *GRP* (green), *GHRL* (red), *ASCL1* (white). Right: Mean ± SEM of *ASCL1*$^+$ cell types, N = 3 human fetal lungs, n = 243 *ASCL1*$^+$ cells.
(E) Left: HCR, *GRP* (green), *NEUROD1* (red), *GHRL* (white). Right: Mean ± SEM of *NEUROD1*$^+$ cell types: N = 2, 11 pcw human fetal lungs, n = 129; N = 3, 12 pcw human fetal lungs, n = 132. Scale bars 25 μm.
(F) Gene signature scoring of A-type and N-type SCLC features in the epithelial UMAP.
(G) Scenic analysis of predicted TF network governing mid tip progenitor cells to pulmonary NE. Trajectory and colour coding match Fig. S13A,B.
(H) Organoids from 8 pcw human fetal lungs were transduced with Doxycline (Dox) inducible TF, or mNeonGreen-NLS, lentivirus. Transduced organoids were isolated by flow cytometry based on TagRFP expression, seeded in Matrigel for 10-13 days prior to Dox treatment. Organoid cells were harvested 3 days post-Dox for scRNA-Seq. N=3 organoid lines.
(I) Left: reference UMAP of primary human fetal lung epithelium. Mid and right: scRNA-Seq of organoids overexpressing *mNeonGreen-NLS*, *ASCL1* or *NEUROD1* projected onto the primary data.
See also Figures S13 and S14.

Figure S1. Quality control for scRNA-seq and scATAC-seq data

(A) Distributions of the number of genes detected per cell, grouped by 10X libraries.
(B) Proportions of broad cell types in samples treated with Trypsin, and Trypsin plus EPCAM enrichment following colour codes in Figure 1C.

(C) Initial clusters of data separating compartments, before subclustering.

(D-F) Workflows of the recursive subclustering method (D), the Doublet Cluster Labeling (DouCLing) method to identify doublet-driven clusters (E), and inference of maternal cells using Souporcell (F).

(G) Doublet scores calculated by Scrublet.

(H) Inferred doublet clusters using DouCLing.

(I) Number of genes detected projected on UMAP.

(J) Percentage of mitochondrial reads.

(K) Cells in curated doublet clusters.

(L) Cells in clusters of cells coming from other organs. Marker genes in parentheses.

(M) Inferred maternal cells.

(N) Cells in curated low-quality cell clusters.

(O,P) scATAC-seq quality metrics of fragment detection per cell (O) and reads mapped in transcription-start sites (P).

Figure S2. Overview of 144 cell types or cell states

(A-E) All of the curated 144 clusters of single cells projected on UMAP space of transcriptomes, colored by cell type/state (A), developmental stage (B), inferred cell-cycle phase (C), dissection region (D) and dissociation/enrichment strategy (E).

(F) Cells from the initial PNS cluster (C7) projected on UMAP space of transcriptomes, colored by cell type/state.

(G) Selected feature genes of cell types/states in the initial PNS cluster.

(H) Spatiotemporal biases of cell types. Cell types are shown as dots with x representing the weighted average of developmental stages, y representing the score of proximal enrichment and the size corresponding to the cluster size.

Figure S3. Comparing fetal lung scRNA-seq with adult human and mouse lung scRNA-seq.

(A-F) Correlations of scVI latent variables between human fetal lung cell clusters and those of previously annotated adult cell clusters (Madissoon et al. 2021, A-C) and mouse lung cell clusters

(Zepp et al. 2021, D-F), focusing on epithelial (A, D), fibroblast (B, E) and endothelial (C, F) compartments.

(G) Expression dotplot of genes shared or unique to fetal/adult lung AT1/AT2 cell clusters..

Figure S4. Spatial analysis of airway epithelial cells in the developing human lungs by *in situ* HCR.

(A) Tip and stalk epithelial cells in distal regions of fetal lungs at 17 pcw, immunostained using antibodies against CD36 (tip epithelial cells, red), PDPN (stalk epithelial cells, white), and E-cadherin (epithelium, cyan).

(B, B') Airway progenitor cells in distal fetal lungs at 10 (B) and 16 (B') pcw. The airway progenitor cells marked by *SOX9*$^{-}$*/CYTL1*$^{+}$*/SCGB3A2*$^{+}$ are located proximally to the *CYTL1*$^{-}$*/SCGB3A2*$^{-}$ stalk. *SCGB1A1* indicates club cells (B, white). *SFTPC* is mainly expressed in the tip and partly located in stalk regions (B', green).

(C) GHRL^{+} neuroendocrine (dashed line, red) and GRP^{+} pulmonary neuroendocrine cells (arrow, green) in fetal lungs at 22 pcw. *SFTPC* indicates tip epithelial cells (white).

(D) Airway progenitor (arrowhead) and club cells (arrow) in non-cartilaginous airway regions of fetal lungs at 12 pcw are marked by *SCGB3A2*$^{+}$*/SCGB1A1*$^{-}$ and *SCGB3A2*$^{+}$*/SCGB1A1*$^{+}$, respectively. Tip, stalk, airway progenitor, and club cells are localised progressively more proximally from the distal tip regions to the proximal non-cartilaginous airway regions. *SCGB3A2* (green), *SCGB1A1* (red).

(E) Proximal secretory 1 (arrowhead) and 2 (arrow) are distinguishable by the presence or absence of *SCGB1A1* expression, each marked by *SCGB3A1*$^{+}$*/SCGB1A1*$^{low/-}$/MUC16$^{low/-}$ and *SCGB3A1*$^{+}$*/SCGB1A1*$^{+}$/MUC16$^{low/+}$, respectively, in the proximal cartilaginous airway in 15 pcw fetal lungs. *MUC16*$^{+}$ only cells are MUC16^{+} ciliated cells. *SCGB3A2* (green), *SCGB1A1* (red), *MUC16* (white).

(F) Proximal secretory 2 (arrowhead) and 3 (arrow) are distinguishable by the presence or absence of *SCGB3A2* and *MUC16* expression, marked by *SCGB3A2*$^{+}$*/SCGB1A1*$^{+}$/MUC16$^{low/+}$ and *SCGB3A2*$^{low/-}$*/SCGB1A1*$^{+}$/MUC16^{+}, respectively, in the proximal cartilaginous airway of fetal lungs at 15 pcw. *SCGB3A2* (green), *SCGB1A1* (red), *MUC16* (white).

(G) Submucosal gland cells (arrow) located in SMGs are marked by strong *LTF* expression with *SCGB3A1*$^{+}$/*SCGB3A2*$^{-}$ in the proximal cartilaginous airway regions of fetal lungs at 15 pcw. *SCGB3A2* (green), *LTF* (red), *SCGB3A1* (white).

(H) Ciliated cells and secretory cells are distinguishable by expression of *FOXJ1* (red) or *SCGB3A2* (green) in the non-cartilaginous airway regions at 19 pcw lungs. Ciliated cells (arrowhead), *FOXJ1*$^{+}$/*SCGB3A2*$^{-}$; secretory cells (arrow), *FOXJ1*$^{-}$/*SCGB3A2*$^{+}$.

(I) MUC16^{+} ciliated cells (dashed line), ciliated cells (dashed circle), and secretory cells (arrow) located in the proximal cartilaginous airway regions of fetal lungs at 19 pcw. The MUC16^{+} ciliated cells express *MUC16* (white) with a weak level of *FOXJ1* (red), whereas the ciliated cells only express strong *FOXJ1* without *MUC16* expression. *SCGB3A2* (green)

(J, J') Proximal basal cells (J, dashed line) line the basal layer of the proximal cartilaginous pseudostratified airway in fetal lungs at 19 pcw and are marked by *TP63* (red), *F3* (white), and *IGFBP3* (green). In contrast, only a few *TP63*$^{+}$ basal cells (J', red, arrowheads) are observed in the non-cartilaginous, non-pseudostratified airway regions.

(K) *ASCL1*$^{+}$ pulmonary neuroendocrine (arrow) and MUC5AC^{+}/ASCL1^{+} progenitors (arrowhead) in the non-cartilaginous airway regions of fetal lung at 12 pcw. MUC5AC (green), *ASCL1* (red), *SCGB3A2* (white).

DAPI, nuclei. Scale bars, 50 μm.

(L) Diagram describing spatial location of epithelial cell types observed in the developing human lungs.

Figure S5. Spatiotemporal location, distribution, and quantification of major epithelial cell types along the distal to proximal axis of the developing lungs.

(A) *In situ* HCR analysis of fetal human lung tissues at mid (10-12 pcw) and late (15-21 pcw) stages, showing spatiotemporal location and distribution of major epithelial cell types along the distal to proximal axis of the developing lungs. The lung regions were divided for imaging into tip, stalk to terminal airway; distal to proximal non-cartilaginous airway; and proximal cartilaginous airway.

(B) Quantification of cells expressing marker genes of airway lineages along the airway regions at mid (10-12 pcw, upper) and late (15-21 pcw, lower) stages. *SCGB3A2*, airway progenitors/all

secretory lineage cells; *CYTL1*, airway progenitor cells; *NDUFA4L2*, club/proximal secretory cells; *FMO2*, club/proximal secretory cells; *FOXJ1*, ciliated cells; *TP63*, basal cells; *SCGB1A1*, club/proximal secretory cells; *SCGB3A1*, proximal secretory cell subtypes 1-3. Significance was evaluated by 1-way ANOVA with Tukey multiple comparison post-test; n=3 biological replicates; ns: not significant **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001.

(C) Proportion of proximal secretory progenitor cells, proximal secretory cell subtypes 1-3 within the proximal cartilaginous airway regions by ages, at 10-12, 15-16, and 19-21 pcw. The secretory cells in the proximal cartilaginous airway regions were counted: Prox Secretory Prog, *SCGB3A2*$^{+}$*SCGB3A1*$^{-}$*SCGB1A1*$^{-}$; Prox Secretory 1, *SCGB3A2*$^{+}$*SCGB3A1*$^{+}$*SCGB1A1*$^{-}$; Prox Secretory 2, *SCGB3A2*$^{+}$*SCGB3A1*$^{+}$*SCGB1A1*$^{+}$; Prox Secretory 3, *SCGB3A2*$^{-}$ *SCGB3A1*$^{+}$*SCGB1A1*$^{+}$. Club cells located in the non-cartilaginous airway regions were excluded. Significance was evaluated by 2-way ANOVA with Tukey multiple comparison post-test; n=4 biological replicates; ns: not significant **P*<0.05, ***P*<0.01, ****P*<0.001.

(D) Diagram describing spatiotemporal distribution of major cell type markers along the distal to proximal axis of the developing lungs, at mid and late stages. Mid stage only, blue; Late stage only, red; Mid-to-late stages, green. Arrows indicate narrowed (*CYTL1*) or expanded (*NDUFA4L2*, *FOXJ1*) distribution after mid to late stage transition.

(E) *In situ* HCR analysis of rare cell type markers of putative ionocytes (*FOXI1* yellow) and tuft cells (*POU2F3*, red). *E-cadherin*, green.

DAPI, nuclei. Scale bar, 50 μm.

Figure S6. Trajectory analysis of airway lineage differentiation via airway progenitor cells in the developing human lung.

(A, A') UMAP visualization (A) and PAGA analysis (A') of a lineage trajectory from mid tip to proximal secretory lineage cells, including proximal secretory progenitor and proximal secretory cell subtypes 1 to 3. Mid and late basal cells were shown to be disconnected from other proximal secretory cell types in the PAGA analysis (A').

(B-C) Trajectory UMAPs, by cell type (B) and stages (B'), and the relevant gene expression heatmap (C) displaying the selected lineage trajectory from mid tip to proximal secretory cell subtypes 1 and 2, analysed by Monocle 3. (Note that the grey lines in UMAP indicate all of the predicted differentiation paths from a user-defined starting point.)

(D, D') UMAP visualization (D) and PAGA analysis (D') of a lineage trajectory from late tip, late stalk, late airway progenitor to club cells. Basal cells, including late basal, proximal basal, and SMG basal cells were shown to be left out of the trajectory as they do not connect clearly to the other cell types in this analysis (D').

(E-F) Trajectory UMAPs, by cell types (E) and stages (E'), and the relevant gene expression heatmap (C) showing the selected lineage trajectory from late tip to club cells, analysed by Monocle 3.

(G) Purification of distal *SCGB3A2*-GFP^{+} airway cells from human fetal lung tissues at 8-11 pcw. The epithelial cells were isolated using EPCAM magnetic microbeads (MACS) from the dissected distal and proximal airway tissues, followed by infection with lentivirus habouring *SCGB3A2* promoter-driven GFP. The *SCGB3A2*-GFP positive cell fractions were sorted and analysed by FACS after 48 hrs and *in vitro* cultured for 28 and 45 days in the airway differentiation medium.

(H) Gene expression profile of the freshly purified *SCGB3A2*-GFP positive cells derived from distal and proximal airway tissues were investigated by qRT-PCR and compared with dissected tip cells. *SOX9*, distal tip progenitor marker. *CYTL1*, airway progenitor marker. *SCGB1A1* and *SCGB3A2*, airway/secretory cell lineage markers. *SCGB3A1*, proximal secretory cell marker. Data was normalised to *SCGB3A2*-GFP negative cells derived from distal tip/stalk tissues; mean ± SD of 3 biological replicates. Significance was evaluated by 1-way ANOVA with Tukey multiple comparison post-test; **P*<0.05, ***P*<0.01, ****P*<0.001.

(I) Gene expression analysis of the *in vitro* cultured *SCGB3A2*-GFP positive cells (airway progenitors) derived from distal airway tissues by qRT-PCR. Airway organoids were formed from the *SCGB3A2*-GFP positive cells and collected at Day 0, 14, and 28 days after culture for the analysis. Data were normalised to *SCGB3A2*-GFP negative cells derived from distal tip/stalk tissues; mean ± SD of 4 biological replicates. Significance was evaluated by 1-way ANOVA with Tukey multiple comparison post-test; ns: not significant **P*<0.05, ***P*<0.01.

(J-M) Immunofluorescence analysis of two biologically independent, SCGB3A2-GFP+ cell-derived airway organoids cultured in the airway differentiation medium for 28 (J-L) and 45 (M) days. SCGB1A1 (J, red), airway progenitor/secretory cell marker. TP63 (K, red), basal cell marker. FOXJ1 (L, magenta), ciliated cell marker. SCGB3A1 (M, red), proximal secretory cell marker. DAPI, nuclei. Scale bar, 50 μm.

Figure S7. Late epithelial tip cells differentiate to AT2 and AT1 cells.

(A, B) UMAP visualisation (A) of a lineage trajectory from early/mid/late tip to fetal AT2 and AT1 cells and the relevant gene expression heatmap (B) showing the selected lineage trajectory analysed by Monocle 3.

(C) *In situ* HCR (*TPPP3* and *SFTPC*) and immunostaining (SOX9) analysis of 15 pcw fetal lung, describing SOX9^{+}*TPPP3*$^{+}$*SFTPC*$^{+}$ tip epithelial progenitors (lines) and SOX9^{-}*TPPP3*$^{-}$*SFTPC*$^{+}$ fetal AT2 cell population (dashed circles) lining the stalk.

(D) Immunostaining of 21 pcw fetal lung using antibodies against SOX9 (red), PDPN (white), and E-cadherin (green). Arrows indicate the late tip cell population, which does not co-express the stalk marker, PDPN.

(E-G) *In situ* HCR analysis of 19 (F) and 21 pcw (E, G) fetal lungs, showing the *SFTPC*$^{+}$ fetal AT2 cell population (arrowheads) lining the developing air sacs. Arrows indicate *SFTPC*$^{+}$ late tip cells. (E) *SFTPC* (red). (F, G) *NAPSA* (white; F) and *ETV5* (red; G) overlap with *SFTPC* in the fetal AT2 cells.

(H-J) *In situ* HCR analysis of distal lung regions at 17 (H), 20 (I), and 21 (J) pcw, visualising *SFTPC*$^{-}$ fetal AT1 cells (arrows). *SFTPC*$^{-}$*/SPOCK2*$^{-}$ stalk cells at 17 pcw (H) began to express *SPOCK2* (red) at 20 pcw (I) and further developed to future AT1 cells (*SFTPC*$^{-}$*/SPOCK2*$^{+}$) at 21 pcw (J). Dashed circles (I) and arrowheads (J) indicate fetal AT2 cells. Dashed line (J) shows fetal AT1 cells lining the developing air sacs.

DAPI, nuclei. Scale bars, 50 μm.

Figure S8. Endothelial cell types in the developing human lung.

(A,B) UMAP visualisation of endothelial cells, coloured by cell types (A) and stages (B).

(C) Dot plot describing differential marker gene expression level by cell type.

(D-F) *In situ* HCR analysis of distal lung regions at 20 (E), and 21 (D, F) pcw. (D) Aerocytes (*S100A3*$^{+}$ red/*CA4*$^{+}$ white), capillary endothelium (*CA4*$^{+}$ white), and all endothelial cells (*PECAM*$^{+}$, green). (E) Arterial endothelial cells (*GJA5*$^{+}$ red), venous endothelial cells (*ACKR3*$^{+}$ white), and all endothelial cells (*PECAM*$^{+}$, green). (F) Lymphatic endothelial cells (*PROX1*$^{+}$ white) and all endothelial cells (*PECAM*$^{+}$, green). DAPI, nuclei. Scale bars, 50 μm.

(G) Trajectory UMAP and PAGA plot (G'') visualising potential endothelial cell lineage hierarchy from Mid/Late capillary endothelial cells to arterial endothelial cells, aerocytes, or venous endothelial cells coloured by cell types (G) and stages (G').

(H, I) Individual trajectory UMAPs (H) and the relevant gene expression heatmaps (I) displaying potential lineage trajectories derived by Monocle 3 from Mid/Late capillary endothelial cells to arterial endothelial cells (*top*), aerocytes (*middle*), or venous endothelial cells (*bottom*).

Figure S9. Clustering and cell type markers for immune cell types and comparison to other fetal data sets.

(A,C,G) UMAP embeddings of different immune compartments showing myeloid cell types/states (A), T, NK and ILC lymphoid cells (C) and B lymphoid cells (G).

(B,F,K) Dot plots showing expression of selected marker genes of cell types/states in the three immune compartments.

(D, E, H, I, J) Enrichment of each class of immune receptors based on abTCR, gdTCR and BCR-enriched scRNA-seq.

(L) Predicted organ-of-source with highest scores for cells shown in Figure 1, based on the reference atlas in (Cao et al. 2020).

Figure S10. Spatial analysis of mesenchymal cell types in the developing human lungs by *in situ* HCR assay and immunostaining.

(A) Vascular SMC 1 and 2 are surrounding arterial endothelial cells (*PECAM1*$^{+}$, dashed line), each marked by *NTN4*$^{+}$/*PLN*$^{-/low}$ (vSMC 1, arrows) and *NTN4*$^{+}$/*PLN*$^{+/high}$ (vSMC 2, lines).

(B) Dot plot describing differential gene expression between vascular SMC 1 and 2.

(C) Vascular SMCs and adventitial fibroblasts in 17 pcw fetal lung. *NDUF4AL2*$^{+}$ red/*NTRK3*$^{+}$ vSMCs (arrows) are surrounded by *NDUF4AL2*$^{-}$/*NTRK3*$^{+}$ adventitial fibroblasts (arrowheads). *PECAM1* (green) indicates an endothelial cell tube.

(D) *FAM162B*$^{+}$ pericytes (red) are surrounding *PECAM1*$^{+}$ endothelial cells (green) in the microvascular regions.

(E) Dot plot describing differential marker gene expression level between alveolar, adventitial and airway fibroblasts.

(F) Concept network visualisation of gene ontology (GO) analysis using clusterProfiler for differentially expressed genes in alveolar, adventitial and airway fibroblasts.

(G-I) Immunostaining of fetal lung tissues at 11 (G), 15 (H), and 21 (I) pcw, to visualise myofibroblast populations: Myofibroblast-1 (G) and -2 (H) surrounding the developing stalk epithelial tubes, and Myofibroblast-3 (I) surrounding the developing air sacs. $ACTA2^{+}$/$PDGFRA^{+}$ Myofibroblast-1 ($THBD^{weak}$; G) and -2 ($THBD^{high}$, arrows; H). $PDGFRA^{+}$ Myofibroblast-3 at 21 pcw, does not express ACTA2 (arrows; I).

(J) Dot plot describing differential gene expression level between myofibroblast-2 and -3. The myofibroblast-2 population showed enriched expression of Wnt signalling associated genes, e.g. *NOTUM*, *LEF1*, and *DACH2*.

(K) *In situ* HCR assay of 17 pcw fetal lung tissues. Myofibroblast-2 expresses *NOTUM* (red), a Wnt antagonist, to block local Wnt signals from alveolar fibroblasts (white, *WNT2*) to the stalk epithelium.

DAPI, nuclei. Scale bars, 50 μm.

Figure S11. Signalling ligand-receptor interactions in the adventitial niche.

(A) Overview of predicted ligand–receptor interactions using CellPhoneDB in the adventitial niche. Dot plots visualise gene expression by cell type and dashed arrows indicate a predicted direction of signalling from ligands to receptors.

(B) Visium spot transcriptome cluster map visualising signalling ligands expressed in the fetal lung tissues at 19 (upper) and 17 (lower) pcw. Scale bars denote 2.5 mm (upper) and 2 mm (lower), respectively.

Figure S12. Global landscape of motif enrichment.

Top 5 enriched motifs in the marker peaks among all the cell types/states. Statistical significance is visualised as a heatmap according to the colour bar below.

Figure S13. Transcription factor regulatory network controlling NE subtypes.

(A) Selected trajectory from Mid tip cells to *GHRL*$^{+}$ NE cells via Intermediate NEs, a transition cell population.

(B) Heatmap of genes differentially expressed along the trajectory.

(C) Representative HCR images showing the transition between two types of NE cells. *GRP* (green), *NEUROD1* (red), *GHRL* (white). #1 labelled *GRP*$^{+}$*NEUROD1*^{low}*GHRL*$^{-}$ cells, which have just started the transition from GRP^{+} pulmonary NE/precursor cells. #2 labelled *GRP*^{low}*NEUROD1*$^{+}$*GHRL*^{low} cells, in transition to *GHRL*$^{+}$ NE cells. #3 labelled *GRP*$^{-}$ *NEUROD1*$^{+}$*GHRL*$^{+}$, *GHRL*$^{+}$ NE cells. Right: Mean ± SEM of *NEUROD1*$^{+}$ cell types. 11 pcw: N=2 fetal lungs, n = 129 *NEUROD1*$^{+}$ cells; 12 pcw N=3 fetal lungs, n=132 *NEUROD1*$^{+}$ cells. Scale bars = 25 μm in all panels.

(D) Representative HCR images showing *NEUROG3* co-expression with *ASCL1* and *NEUROD1*. Dashed white lines label representative cells showing different combinations of the three transcription factors, further indicated by #1-#5 labelling. *ASCL1* (cyan), *NEUROG3* (red), *NEUROD1* (yellow).

(E) Representative HCR images showing *RFX6* expression in *GHRL*$^{+}$ NE cells. Dash yellow line labelled *GRP*$^{+}$*RFX6*$^{-}$ pulmonary NE cells. Scale bars = 25 μm in all panels.

Figure S14. Validation of NE transcription factors using human fetal lung organoid system.

(A) Representative epifluorescent microscopic images showing organoid morphology after 3 days of mNeonGreen-3xNLS (control), ASCL1, or NEUROD1 overexpression.

(B) ScRNA-seq results of organoid transcription factor overexpression overlay on human fetal lung scRNA-seq as a reference.

(C) scRNA-seq results of transcription factor overexpression; organoid data only in the UMAP. Selected transcription factor expression was shown in the middle panel. A regulatory network of the selected transcription factors were drawn based on the organoid OE data at the bottom of the panel. (Note that the arrows do not necessarily denote direct interactions).

Figure 1

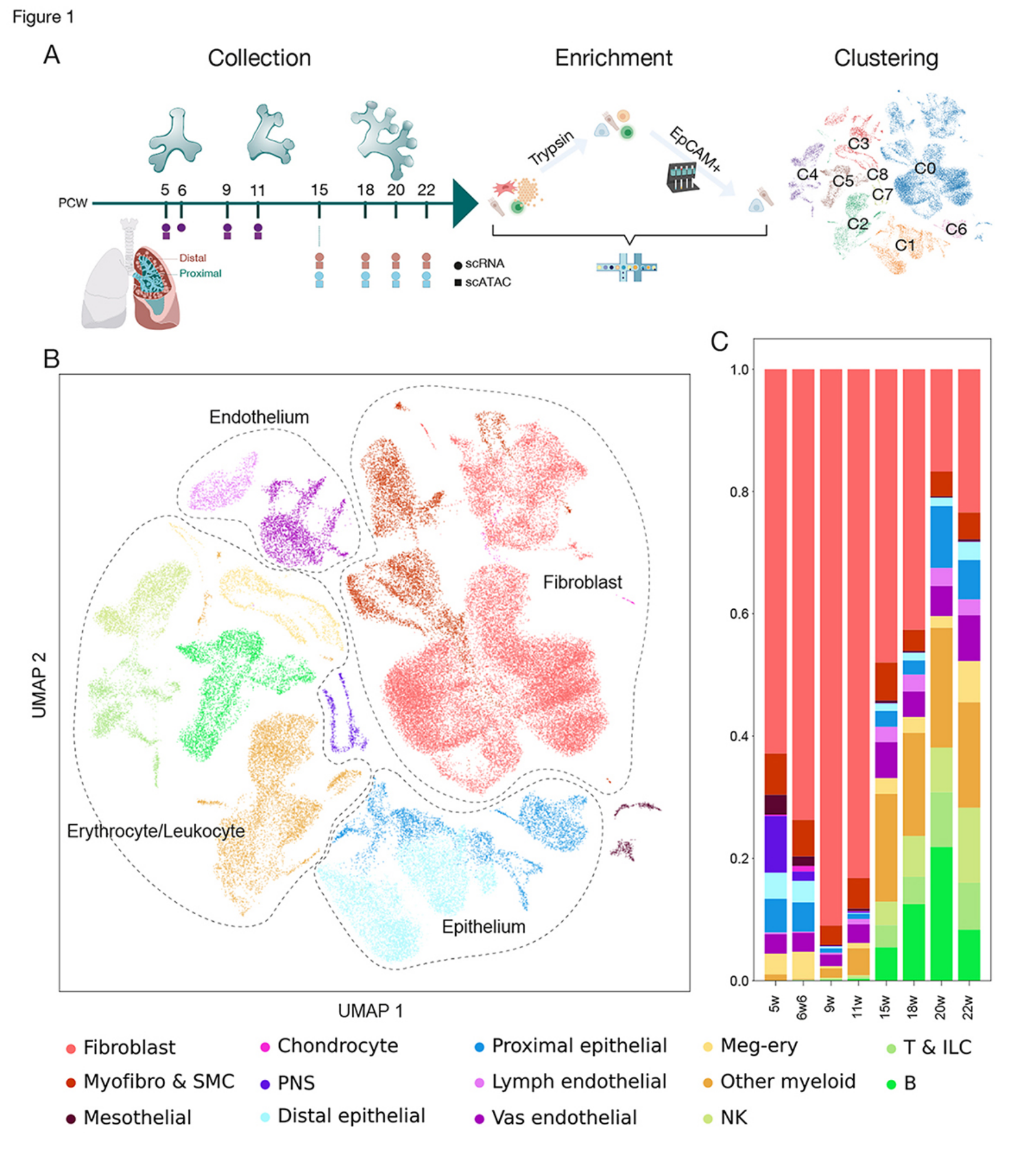

A
Collection
Enrichment
Clustering
PCW
5 6 9 11 15 18 20 22
Distal
Proximal
scRNA
scATAC
Trypsin
EpCAM+
C0
C1
C2
C3
C4
C5
C6
C7
C8
B
Endothelium
Fibroblast
Erythrocyte/Leukocyte
Epithelium
UMAP 1
UMAP 2
C
1.0
0.8
0.6
0.4
0.2
0.0
5w
6w6
9w
11w
15w
18w
20w
22w
Fibroblast
Myofibro & SMC
Mesothelial
Chondrocyte
PNS
Distal epithelial
Proximal epithelial
Lymph endothelial
Vas endothelial
Meg-ery
Other myeloid
NK
T & ILC
B

Figure 2

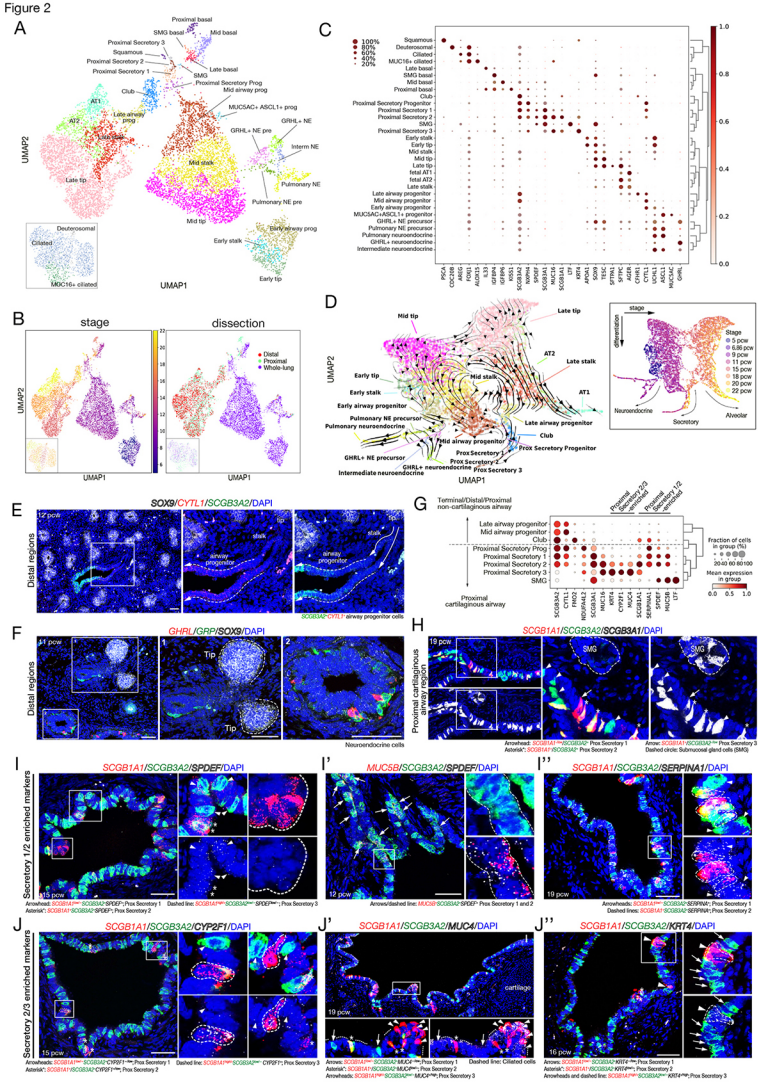

A
Proximal basal
SMG basal
Mid basal
Proximal Secretory 3
Squamous
Proximal Secretory 2
Proximal Secretory 1
Late basal
SMG
Proximal Secretory Prog
Mid airway prog
Club
AT1
AT2
Late airway prog
MUC5AC+ ASCL1+ prog
GRHL+ NE pre
GRHL+ NE
Interm NE
Mid stalk
Pulmonary NE
Late tip
Pulmonary NE pre
Mid tip
Early airway prog
Early stalk
Early tip
Deuterosomal
Ciliated
MUC16+ ciliated
UMAP2
UMAP1
B
stage
dissection
Distal
Proximal
Whole-lung
UMAP2
UMAP1
C
100%
80%
60%
40%
20%
Squamous
Deuterosomal
Ciliated
MUC16+ ciliated
Late basal
SMG basal
Mid basal
Proximal basal
Club
Proximal Secretory Progenitor
Proximal Secretory 1
Proximal Secretory 2
SMG
Proximal Secretory 3
Early stalk
Early tip
Mid stalk
Mid tip
Late tip
fetal AT1
fetal AT2
Late stalk
Late airway progenitor
Mid airway progenitor
Early airway progenitor
MUC5AC+ASCL1+ progenitor
GHRL+ NE precursor
Pulmonary NE precursor
Pulmonary neuroendocrine
GHRL+ neuroendocrine
Intermediate neuroendocrine
PSCA
CDC20B
FOXI1
ASCL3
IL33
GFRA1
SCGB1A1
HSD11
SCGB3A2
SFXBP
SCGB3A1
MUC16
SCGB3A1
KRT4
AP000783
SOX9
TSSC1
SFTPC
SFTPB
CYP4X1
CYTL1
ASCL1
NKX2-1
GHRL
1.0
0.8
0.6
0.4
0.2
0.0
D
Mid tip
Late tip
AT2
Late stalk
Early tip
Mid stalk
Early stalk
AT1
Early airway progenitor
Pulmonary NE precursor
Late airway progenitor
Pulmonary neuroendocrine
Club
Mid airway progenitor
Prox Secretory Progenitor
Prox Secretory 2
GHRL+ neuroendocrine
Prox Secretory 3
Intermediate neuroendocrine
UMAP2
UMAP1
stage
differentiation
Stage
5 pcw
6.66 pcw
9 pcw
11 pcw
15 pcw
18 pcw
20 pcw
22 pcw
Neuroendocrine
Secretory
Alveolar
E
SOX9/CYTL1/SCGB3A2/DAPI
12 pcw
Distal regions
tip
stalk
airway progenitor
SCGB3A2+/CYTL1+ airway progenitor cells
G
Terminal/Distal/Proximal non-cartilaginous airway
Late airway progenitor
Mid airway progenitor
Club
Proximal Secretory Prog
Proximal Secretory 1
Proximal Secretory 2
SMG
Proximal cartilaginous airway
Proximal Secretory 2/3 enriched
Proximal Secretory 1/2 enriched
SCGB3A2
CYTL1
SOX9
SCGB1A1
KRT4
CYP2F1
SFTPC
SCGB3A1
SFTPB
MUC5B
LTF
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
F
GHRL/GRP/SOX9/DAPI
11 pcw
Distal regions
Tip
Neuroendocrine cells
H
SCGB1A1/SCGB3A2/SCGB3A1/DAPI
19 pcw
Proximal cartilaginous airway region
SMG
Arrowhead: SCGB1A1low/SCGB3A2+ Prox Secretory 1
Asterisk*: SCGB1A1+/SCGB3A2+ Prox Secretory 2
Arrow: SCGB1A1+/SCGB3A2low Prox Secretory 3
Dashed circle: Submucosal gland cells (SMG)
I
SCGB1A1/SCGB3A2/SPDEF/DAPI
Secretory 1/2 enriched markers
15 pcw
Arrowhead: SCGB1A1low/SCGB3A2+/SPDEF+ Prox Secretory 1
Asterisk*: SCGB1A1+/SCGB3A2+/SPDEF+ Prox Secretory 2
Dashed line: SCGB1A1low/SCGB3A2low/SPDEFlow Prox Secretory 3
I'
MUC5B/SCGB3A2/SPDEF/DAPI
12 pcw
Arrows/dashed line: MUC5B+/SCGB3A2+/SPDEF+ Prox Secretory 1 and 2
I''
SCGB1A1/SCGB3A2/SERPINA1/DAPI
19 pcw
Arrowheads: SCGB1A1low/SCGB3A2+/SERPINA1+ Prox Secretory 1
Dashed lines: SCGB1A1+/SCGB3A2+/SERPINA1+ Prox Secretory 2
J
SCGB1A1/SCGB3A2/CYP2F1/DAPI
Secretory 2/3 enriched markers
15 pcw
Arrowheads: SCGB1A1low/SCGB3A2+/CYP2F1+low Prox Secretory 1
Asterisk*: SCGB1A1+/SCGB3A2+/CYP2F1+ Prox Secretory 2
Dashed line: SCGB1A1low/SCGB3A2low/CYP2F1low Prox Secretory 3
J'
SCGB1A1/SCGB3A2/MUC4/DAPI
19 pcw
cartilage
Arrows: SCGB1A1+/SCGB3A2+/MUC4+ Prox Secretory 1
Asterisk*: SCGB1A1+/SCGB3A2+/MUC4low Prox Secretory 2
Arrowheads: SCGB1A1low/SCGB3A2low/MUC4low Prox Secretory 3
Dashed line: Ciliated cells
J''
SCGB1A1/SCGB3A2/KRT4/DAPI
16 pcw
Arrows: SCGB1A1low/SCGB3A2+/KRT4low Prox Secretory 1
Asterisk*: SCGB1A1+/SCGB3A2+/KRT4+ Prox Secretory 2
Arrowheads and dashed line: SCGB1A1low/SCGB3A2low/KRT4low Prox Secretory 3

Figure 3

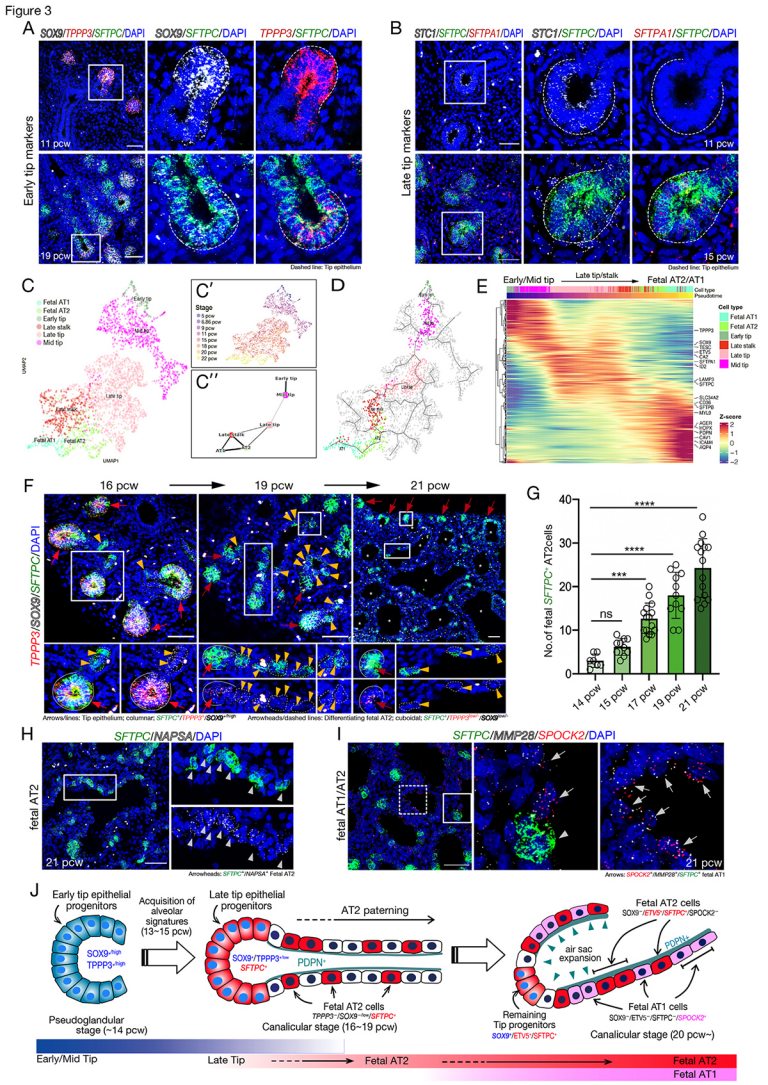

A
SOX9/TPPP3/SFTPC/DAPI
SOX9/SFTPC/DAPI
TPPP3/SFTPC/DAPI
Early tip markers
11 pcw
19 pcw
Dashed line: Tip epithelium
B
STC1/SFTPC/SFTPA1/DAPI
STC1/SFTPC/DAPI
SFTPA1/SFTPC/DAPI
Late tip markers
11 pcw
15 pcw
Dashed line: Tip epithelium
C
Fetal AT1
Fetal AT2
Early tip
Late stalk
Late tip
Mid tip
Early tip
Mid tip
Late tip
Late stalk
Fetal AT1
Fetal AT2
UMAP1
UMAP2
C′
Stage
9 pcw
10.5 pcw
9 pcw
11 pcw
15 pcw
19 pcw
20 pcw
22 pcw
C″
Early tip
Mid tip
Late tip
Late stalk
Late tip
AT1
AT2
D
Early tip
Mid tip
Late tip
Late stalk
AT1
AT2
E
Early/Mid tip
Late tip/stalk
Fetal AT2/AT1
Cell type Pseudotime
Cell type
Fetal AT1
Fetal AT2
Early tip
Late stalk
Late tip
Mid tip
TPPP3
SOX9
SOX2
ETV5
SFTPC
SOX4
LAMP3
SFTPD
ELOVL2
CLDN4
SFTPB
MYL9
AQER
PDPN
PSPN
CAV1
CABIN
HOPX
Z-score
2
1
0
−1
−2
F
16 pcw
19 pcw
21 pcw
TPPP3/SOX9/SFTPC/DAPI
Arrows/lines: Tip epithelium; columnar; SFTPC+/TPPP3+/SOX9+high
Arrowheads/dashed lines: Differentiating fetal AT2; cuboidal; SFTPC+/TPPP3+low/SOX9+low
G
No.of fetal SFTPC+ AT2cells
40
30
20
10
0
****
****
***
ns
14 pcw
15 pcw
17 pcw
19 pcw
21 pcw
H
SFTPC/NAPSA/DAPI
fetal AT2
21 pcw
Arrowheads: SFTPC+/NAPSA+ Fetal AT2
I
SFTPC/MMP28/SPOCK2/DAPI
fetal AT1/AT2
21 pcw
Arrows: SPOCK2+/MMP28+/SFTPC+ fetal AT1
J
Early tip epithelial progenitors
SOX9-high
TPPP3-high
Pseudoglandular stage (~14 pcw)
Acquisition of alveolar signatures (13~15 pcw)
Late tip epithelial progenitors
SOX9/TPPP3-low
SFTPC+
PDPN+
Fetal AT2 cells
TPPP3-low/SOX9-low/SFTPC+
Canalicular stage (16~19 pcw)
AT2 paterning
Fetal AT2 cells
SOX9+/ETV5+/SFTPC+/SPOCK2−
air sac expansion
PDPN+
Remaining Tip progenitors
SOX9/ETv5/SFTPC−
Fetal AT1 cells
SOX9−/ETV5−/SFTPC−/SPOCK2+
Canalicular stage (20 pcw~)
Early/Mid Tip
Late Tip
Fetal AT2
Fetal AT2
Fetal AT1

Figure 4

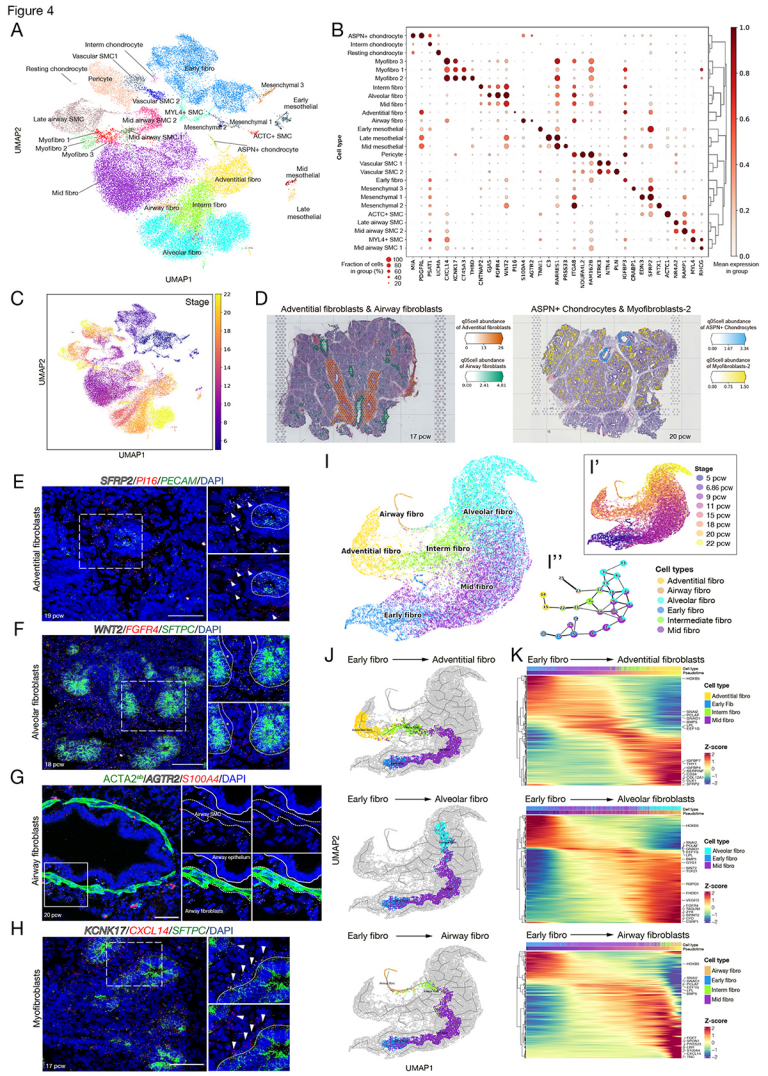
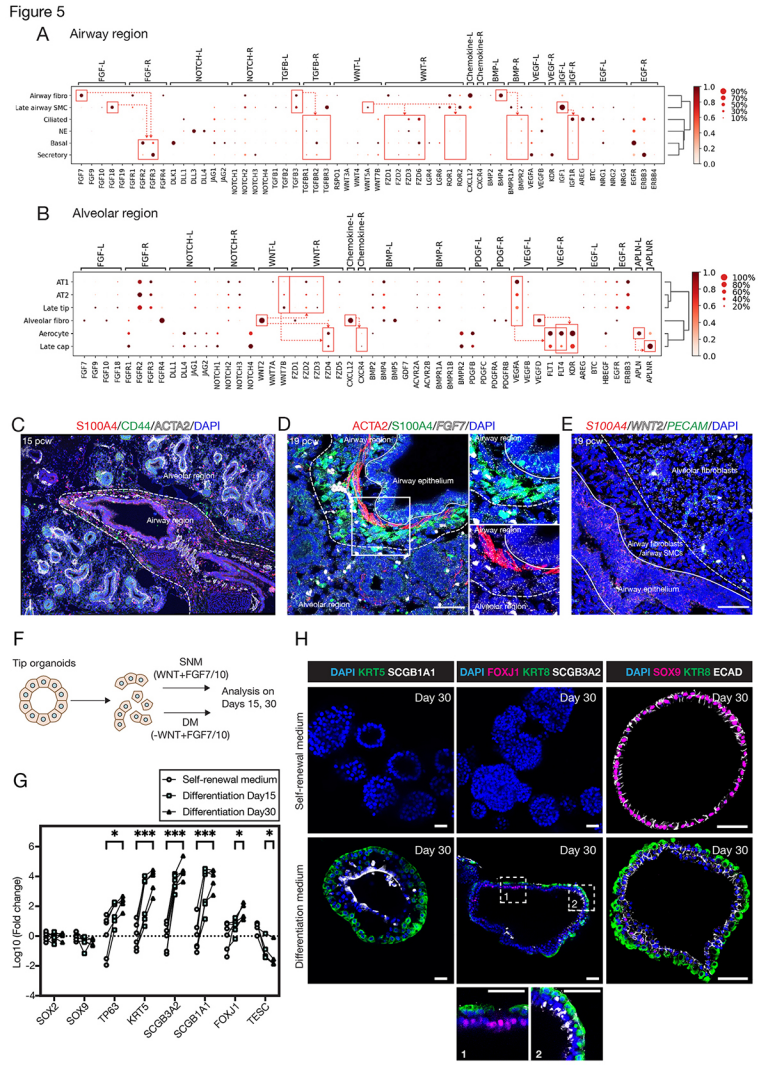

A
Interm chondrocyte
Vascular SMC1
Resting chondrocyte
Pericyte
Early fibro
Mesenchymal 3
Early mesothelial
Vascular SMC 2
MYL4+ SMC
Mid airway SMC 2
Mesenchymal 1
Mesenchymal 2
Late airway SMC
ACTC+ SMC
Myofibro 1
Myofibro 2
Myofibro 3
Mid airway SMC 1
ASPN+ chondrocyte
Mid mesothelial
Adventitial fibro
Mid fibro
Airway fibro
Interm fibro
Late mesothelial
Alveolar fibro
UMAP2
UMAP1
B
ASPN+ chondrocyte
Interm chondrocyte
Resting chondrocyte
Myofibro 3
Myofibro 1
Myofibro 2
Alveolar fibro
Interm fibro
Adventitial fibro
Airway fibro
Early mesothelial
Late mesothelial
Mid mesothelial
Pericyte
Vascular SMC 1
Vascular SMC 2
Early fibro
Mesenchymal 3
Mesenchymal 1
Mesenchymal 2
ACTC+ SMC
Late airway SMC
Mid airway SMC 2
MYL4+ SMC
Mid airway SMC 1
Cell type
Fraction of cells in group (%)
Mean expression in group
C
Stage
UMAP2
UMAP1
D
Adventitial fibroblasts & Airway fibroblasts
q05cell abundance of Adventitial fibroblasts
q05cell abundance of Airway fibroblasts
17 pcw
ASPN+ Chondrocytes & Myofibroblasts-2
q05cell abundance of ASPN+ Chondrocytes
q05cell abundance of Myofibroblasts-2
20 pcw
E
SFRP2/PI16/PECAM/DAPI
Adventitial fibroblasts
19 pcw
F
WNT2/FGFR4/SFTPC/DAPI
Alveolar fibroblasts
18 pcw
G
ACTA2alpha/AGTR2/S100A4/DAPI
Airway fibroblasts
Airway SMC
Airway epithelium
Airway fibroblasts
20 pcw
H
KCNK17/CXCL14/SFTPC/DAPI
Myofibroblasts
17 pcw
I
Airway fibro
Alveolar fibro
Adventitial fibro
Interm fibro
Mid fibro
Early fibro
I'
Stage
5 pcw
6.86 pcw
9 pcw
11 pcw
15 pcw
18 pcw
20 pcw
22 pcw
I''
Cell types
Adventitial fibro
Airway fibro
Alveolar fibro
Early fibro
Intermediate fibro
Mid fibro
J
Early fibro → Adventitial fibro
Early fibro → Alveolar fibro
Early fibro → Airway fibro
UMAP2
UMAP1
K
Early fibro → Adventitial fibroblasts
Early fibro → Alveolar fibroblasts
Early fibro → Airway fibroblasts
Cell type
Adventitial fibro
Alveolar fibro
Airway fibro
Early Fib
Early fibro
Interm fibro
Mid fibro
Z-score

Figure 5


A
Airway region
Airway fibro
Late airway SMC
Ciliated
NE
Basal
Secretory
B
Alveolar region
AT1
AT2
Late tip
Alveolar fibro
Aerocyte
Late cap
C
S100A4/CD44/ACTA2/DAPI
15 pcw
Alveolar region
Airway region
D
ACTA2/S100A4/FGF7/DAPI
19 pcw
Airway region
Airway epithelium
Alveolar region
E
S100A4/WNT2/PECAM/DAPI
19 pcw
Alveolar fibroblasts
Airway fibroblasts/ Airway SMCs
Airway epithelium
F
Tip organoids
SNM
(WNT+FGF7/10)
Analysis on
Days 15, 30
DM
(-WNT+FGF7/10)
G
Self-renewal medium
Differentiation Day15
Differentiation Day30
Log10 (Fold change)
SOX2
SOX9
TP63
KRT5
SCGB3A2
SCGB1A1
FOXJ1
TESC
H
DAPI KRT5 SCGB1A1
DAPI FOXJ1 KRT8 SCGB3A2
DAPI SOX9 KRT8 ECAD
Day 30
Self-renewal medium
Differentiation medium
1
2

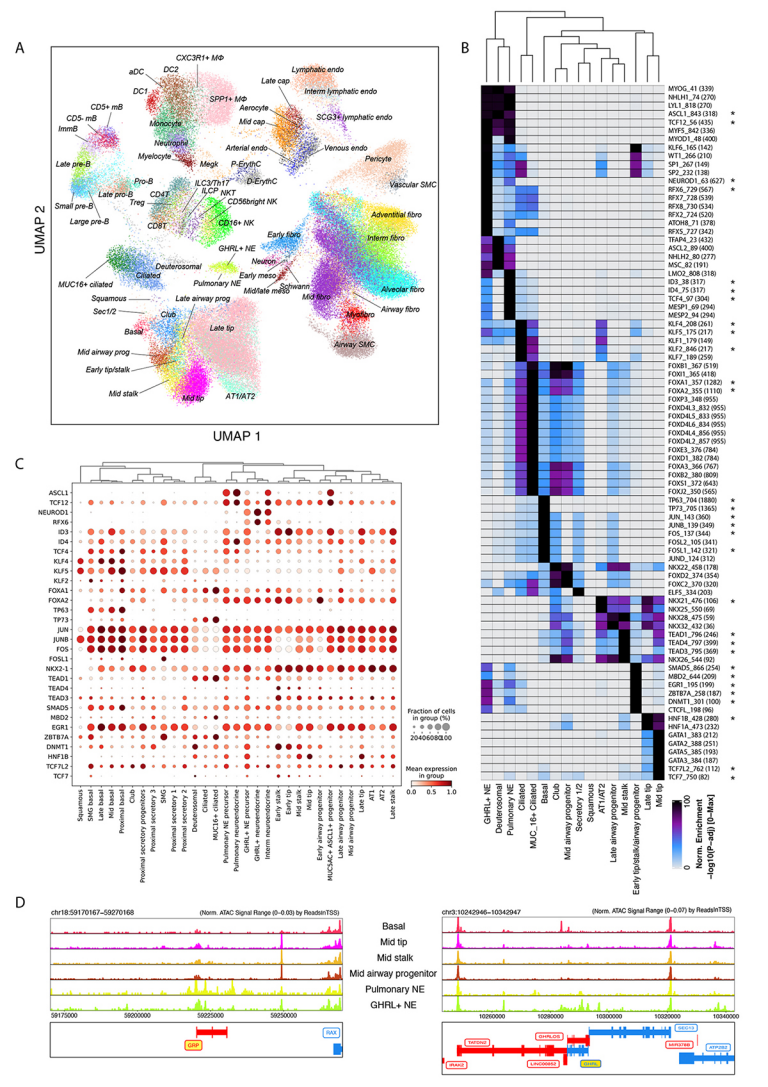
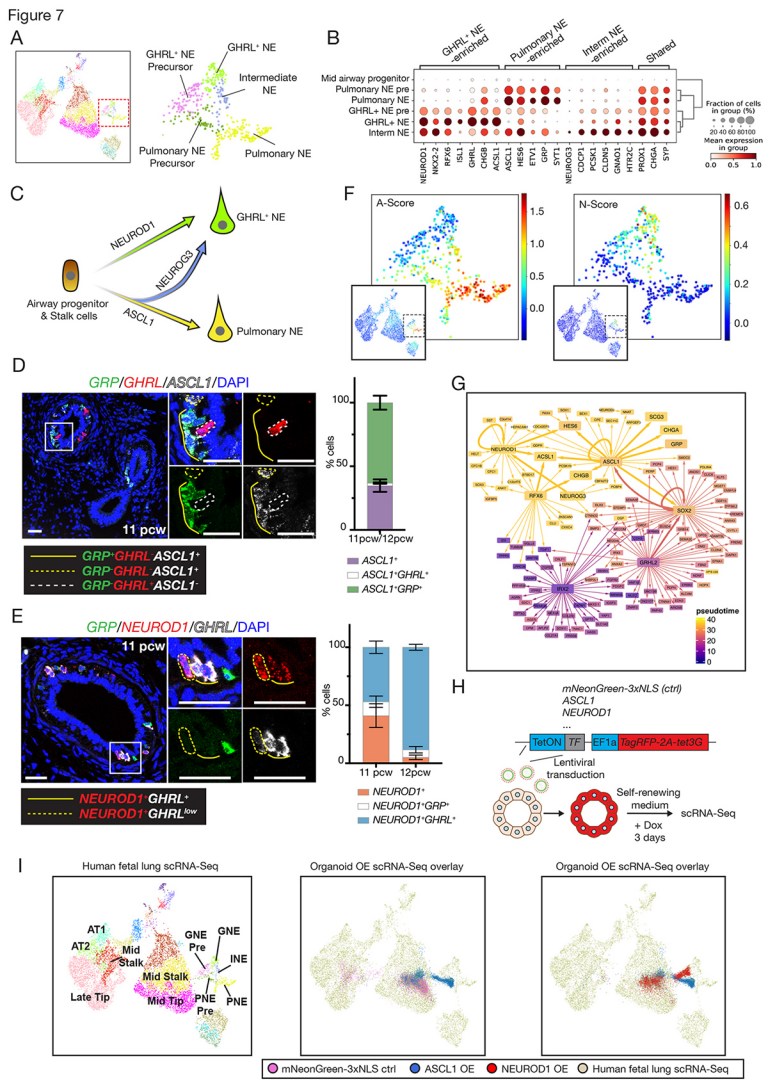
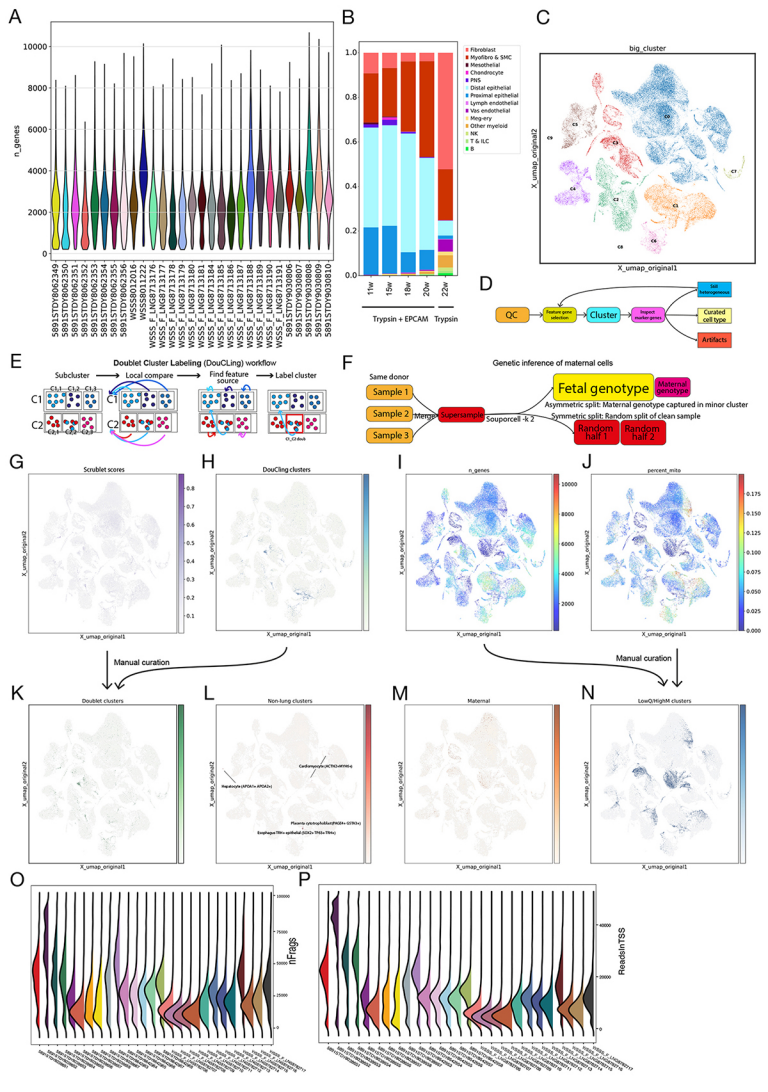
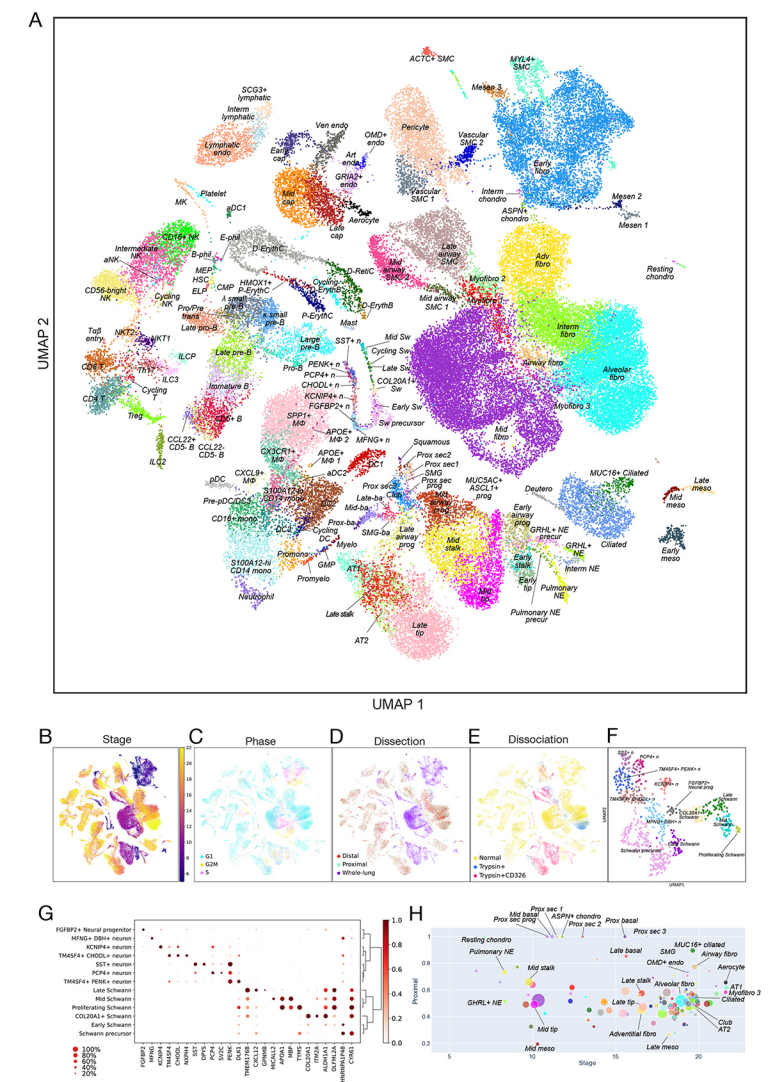

A
UMAP 2
UMAP 1
CXCR3R1+ Mϕ
aDC
DC2
DC1
Lymphatic endo
Late cap
Interm lymphatic endo
SPP1+ Mϕ
Aerocyte
CD5+ mB
CD5- mB
ImmB
Monocyte
Mid cap
SCGB+ lymphatic endo
Neutrophil
Arterial endo
Venous endo
Late pre-B
Myelocyte
Megk
P-ErythC
Pericyte
Pro-B
D-ErythC
Vascular SMC
ILC3/Th17
ILC1/ NKT
Late pro-B
CD4T
CD56bright NK
Treg
Small pre-B
Large pre-B
CD8T
CD16+ NK
Adventitial fibro
Early fibro
Interm fibro
GHRL+ NE
Deuterosomal
Neuron
Ciliated
Early meso
Schwann
MUC16+ ciliated
Pulmonary NE
Midflate meso
Alveolar fibro
Squamous
Late airway prog
Mid fibro
Sec1/2
Myofibro
Airway fibro
Club
Basal
Late tip
Airway SMC
Mid airway prog
Early tip/stalk
Mid stalk
Mid tip
AT1/AT2
B
MYOG_41 (339)
MKI.HI_74 (270)
LYL1_818 (270)
ASCL1_843 (318)
TCF12_56 (435)
MYF5_842 (336)
MYOD1_48 (400)
KLF6_165 (142)
WT1_266 (210)
SP1_367 (149)
SP2_232 (138)
NEUROD1_63 (627)
RFX6_729 (567)
RFX7_728 (539)
RFX8_730 (534)
RFX2_724 (520)
ATOH8_71 (378)
RFX5_727 (342)
TFAP4_23 (432)
ASCL2_89 (400)
NHLH2_80 (277)
MSC_82 (191)
LMO2_808 (318)
ID3_38 (317)
ID4_75 (317)
TCF4_97 (304)
MESP1_69 (294)
MESP2_94 (294)
KLF4_208 (261)
KLF5_175 (217)
KLF1_179 (149)
KLF2_846 (217)
KLF7_189 (259)
FOXB1_367 (519)
FOXI1_365 (418)
FOXA1_357 (1282)
FOXA2_355 (1110)
FOXP3_348 (955)
FOXD4L3_832 (955)
FOXD4L5_833 (955)
FOXD4L6_834 (955)
FOXD4L4_856 (955)
FOXD4L2_857 (955)
FOXE3_376 (784)
FOXD1_382 (784)
FOXA3_366 (767)
FOXD2_380 (809)
FOXS1_372 (643)
FOXJ2_350 (565)
TP63_704 (1880)
TP73_705 (1365)
JUN_143 (360)
JUNB_139 (349)
FOS_137 (344)
FOSL2_105 (341)
FOSL1_142 (321)
JUND_124 (312)
NKX22_458 (178)
FOXD2_374 (354)
FOXC2_379 (320)
ELF5_334 (203)
NKX21_476 (106)
NKX25_550 (69)
NKX28_475 (59)
NKX32_432 (36)
TEAD1_796 (246)
TEAD4_797 (309)
TEAD3_795 (369)
NKX26_544 (92)
SMAD5_866 (254)
MBD2_644 (209)
EGR1_595 (199)
ZBTB7A_258 (187)
DNMT1_301 (100)
CTCFL_196 (96)
HNF1B_428 (280)
HNF1A_473 (232)
GATA1_383 (212)
GATA2_388 (251)
GATA3_385 (193)
GATA3_384 (187)
TCF7L1_762 (112)
TCF7_750 (82)
GHRL+ NE
Deuterosomal
Pulmonary NE
Ciliated
MUC_16+ Ciliated
Club
Mid airway progenitor
Secretory 1/2
Squamous
SMG basal
Late airway progenitor
Mid stalk
Early tip/stalk/airway progenitor
Late tip
Mid tip
Norm. Enrichment (0-100)
−log10(P-adj) [0-Max]
C
ASCL1
TCF12
NEUROD1
RFX6
ID3
ID4
TCF4
KLF4
KLF5
KLF2
FOXA1
FOXA2
TP63
TP73
JUN
JUNB
FOS
FOSL1
NKX2-1
TEAD1
TEAD4
TEAD3
SMAD5
MBD2
EGR1
ZBTB7A
DNMT1
HNF1B
TCF7L2
TCF7
Fraction of cells in group (%)
Mean expression in group
Squamous
SMG basal
Late basal
Mid basal
Proximal basal
Club
Proximal secretory progenitors
Proximal secretory 2
SMG
Proximal secretory 1
Proximal secretory 2
Deuterosomal
Ciliated
MUC16+ ciliated
Pulmonary NE
Pulmonary NE precursor
GHRL+ NE precursor
GHRL+ neuroendocrine
Interm neuroendocrine
Early tip/stalk
Early tip
Mid stalk
Mid tip
Early airway progenitor
MUC5AC+ ASCL1+ progenitor
Late airway progenitor
Mid airway progenitor
Secretory 1/2
AT1
AT2
Late stalk
Basal
Mid tip
Mid stalk
Mid airway progenitor
Pulmonary NE
GHRL+ NE
D
chr18:59170167–59270168
(Norm. ATAC Signal Range (0–0.03) by ReadsInTSS)
59175000
59200000
59225000
59250000
GRP
RAX
chr3:10242946–10342947
(Norm. ATAC Signal Range (0–0.07) by ReadsInTSS)
10260000
10280000
10300000
10320000
10340000
TMTOP2
EHHADH
MECP2
TMTOP3B
ATP2B2
LYNX1SBP2
PPARG
IRAK2

Figure 7


A
GHRL+ NE Precursor
GHRL+ NE
Intermediate NE
Pulmonary NE Precursor
Pulmonary NE
B
GHRL+ NE -enriched
Pulmonary NE -enriched
Interm NE -enriched
Shared
Mid airway progenitor
Pulmonary NE pre
Pulmonary NE
GHRL+ NE pre
GHRL+ NE
Interm NE
NEUROD1
NKX2-2
PCSK1
ISL1
GHRL
CHGB
ACSL1
ASCL1
HES6
GRP
SYT1
NEUROG3
CDCP1
PCSK1
CLDN5
GHRL
KIF5C
PRDX1
CHGA
SYP
Fraction of cells in group (%)
20 40 60 80100
Mean expression in group
0.0 0.5 1.0
C
NEUROD1
NEUROG3
ASCL1
Airway progenitor & Stalk cells
GHRL+ NE
Pulmonary NE
F
A-Score
N-Score
D
GRP/GHRL/ASCL1/DAPI
11 pcw
GRP+GHRL-ASCL1+
GRP-GHRL+ASCL1+
GRP-GHRL-ASCL1-
% cells
11pcw 12pcw
ASCL1+
ASCL1+GHRL+
ASCL1+GRP+
G
pseudotime
E
GRP/NEUROD1/GHRL/DAPI
11 pcw
NEUROD1+GHRL+
NEUROD1-GHRLlow
11 pcw 12pcw
NEUROD1+
NEUROD1+GRP+
NEUROD1+GHRL+
H
mNeonGreen-3xNLS (ctrl)
ASCL1
NEUROD1
TetON
TF
EF1a
TagRFP-2A-tet3G
Lentiviral transduction
Self-renewing medium
+ Dox 3 days
scRNA-Seq
I
Human fetal lung scRNA-Seq
AT1
AT2
Mid Stalk
GNE Pre
GNE
INE
Mid Stalk
Late Tip
Mid Tip
PNE Pre
PNE
Organoid OE scRNA-Seq overlay
Organoid OE scRNA-Seq overlay
mNeonGreen-3xNLS ctrl
ASCL1 OE
NEUROD1 OE
Human fetal lung scRNA-Seq


A
n_genes
S891STDY8062349
S891STDY8062350
S891STDY8062351
S891STDY8062352
S891STDY8062353
S891STDY8062354
S891STDY8062355
S891STDY8062356
WSS8012016
WSS5_FUAG8713177
WSS5_FUAG8713178
WSS5_FUAG8713179
WSS5_FUAG8713180
WSS5_FUAG8713181
WSS5_FUAG8713184
WSS5_FUAG8713185
WSS5_FUAG8713186
WSS5_FUAG8713187
WSS5_FUAG8713188
WSS5_FUAG8713189
WSS5_FUAG8713190
S891STDY9030805
S891STDY9030806
S891STDY9030807
S891STDY9030808
S891STDY9030809
S891STDY9030810
B
Fibroblast
Myofibr & SMC
Mesothelial
Chondrocyte
PNS
Distal epithelial
Proximal epithelial
Lymph endothelial
Vas endothelial
Meg-ery
Other myeloid
NK
T & ILC
B
11w
15w
18w
20w
22w
Trypsin + EPCAM
Trypsin
C
big_cluster
X_umap_original1
X_umap_original2
D
QC
Feature gene selection
Cluster
Inspect cluster genes
Cell heterogeneous
Curated cell type
Artifacts
E
Doublet Cluster Labeling (DouCLing) workflow
Subcluster
Local compare
Find feature source
Label cluster
C1
C2
C1_1
C1_2
C1_3
C2_1
C2_2
C2_3
O_DJ-Ihrb
F
Genetic inference of maternal cells
Same donor
Sample 1
Sample 2
Sample 3
Merge
Supersample
Souporcell -k 2
Fetal genotype
Maternal genotype
Asymmetric split: Maternal genotype captured in minor cluster
Symmetric split: Random split of clean sample
Random half 1
Random half 2
G
Scrublet scores
H
DouCling clusters
I
n_genes
J
percent_mito
Manual curation
Manual curation
K
Doublet clusters
L
Non-lung clusters
Cardiomyocyte(CTNT+MYH6+)
Hepatocyte (AFOA+ APOA2+)
Pleuris cytokeratin&Vim(PKG8+ CITR3+)
Esophagus TP6+ epithelial (CK5+ TP53+ TP6+)
M
Maternal
N
LowQ/HighM clusters
O
nFrags
P
ReadsInTSS


A
UMAP 2
UMAP 1
ACTC+ SMC
MYL4+ SMC
Mesen 3
SCG3+ lymphatic
Interm lymphatic
Lymphatic endo
Ven endo
OMD+ endo
Pericyte
Vascular SMC 2
Early cap
Art endo
Early fibro
GRIA2+ endo
Mid cap
Vascular SMC 1
Interm chondro
Mesen 2
Platelet
MK
aDC1
Aerocyte
Late cap
ASPN+ chondro
Mesen 1
CD16+ NK
E-phil
Late airway SMC
Intermediate NK
B-phil
D-Erythc
Adv fibro
MEP
HSC
D-Ret6C
Resting chondro
aNK
CMP
Mid airway SMC 2
Myofibro 2
CD56-bright NK
ELP
HbMOX1+ P-Erythc
Cycling Erythro
Cycling NK
Mid airway SMC 1
Hyalin chondro
λ small pre-B
D-ErythB
ProFre fibro
κ small pre-B
P-ErythC
Interm fibro
Late pro-B
Mast
Tαβ entry
NKT2
SST+ n
Mid Sw
CD8 T
CD4NKT1
Cycling Sw
Airway fibro
ILCP
Late pre-B
Late Sw
Alveolar fibro
Th17
ILC3
PENK+ n
Cycling
PCP4+ n
Myofibro 3
CD4 T
Immature B
COL20A1+ Sw
Cycling
CHODL+ n
Treg
KCNIP4+ n
Early Sw
SPP1+ Mφ
FGFBP2+ n
Sw precursor
CCL22+ CD5- B
CCL22- CD5- B
APOE+ Mφ 2
MFNG+ n
Mid fibro
Squamous
CX3CR1+ Mφ
Prox sec2
APOE+ Mφ 1
Prox sec1
DC1
ILC2
SMG
aDC2
Late sec prog
CXCL9+ MΦ
Prox sec
MUC5AC+ ASCL1+ prog
MUC16+ Ciliated
Late meso
pDC
S100A12-hi CD14+ mono
Club
Deutero
Mid meso
Pre-pDC
Late-ba
Early airway prog
CD16+ mono
Mid-ba
GRHL+ NE precur
Prox-ba
Mid stalk
Cycling DC
Myelo
Late airway prog
GRHL+ NE
Ciliated
SMG-ba
Early meso
Early stalk
Interm NE
S100A12-hi CD14 mono
Promono
GMP
AT1
Early tip
Pulmonary NE
Promyelo
Mid tip
Neutrophil
Pulmonary NE precur
Late stalk
Late tip
AT2
B
Stage
C
Phase
G1
G2M
S
D
Dissection
Distal
Proximal
Whole-lung
E
Dissociation
Normal
Trypsin+
Trypsin+CD326
F
UMAP2
UMAP1
G
FGFBP2+ Neural progenitor
MFNG+ DBH+ neuron
KCNIP4+ neuron
TM4SF4+ CHODL+ neuron
SST+ neuron
PCP4+ neuron
TM4SF4+ PENK+ neuron
Late Schwann
Mid Schwann
Proliferating Schwann
COL20A1+ Schwann
Early Schwann
Schwann precursor
100%
80%
60%
40%
20%
FGFBP2
MFNG
KCNIP4
TM4SF4
CHODL
NXPH1
SST
DPP6
PCP4
SYT1
PENK
NPY
TM4SF4
CHODL2
MCAM
ALCAM
MPZ
SOX10
TYRO3
COL20A1
ITM2A
ALDH1A1
FABP7
HNRNPA1
CYR61
1.0
0.8
0.6
0.4
0.2
H
Proximal
Stage
Prox sec 1
Mid basal
Prox sec prog
ASPN+ chondro
Prox sec 2
Prox basal
Prox sec 3
Resting chondro
MUC16+ ciliated
Pulmonary NE
Late basal
SMG
Airway fibro
OMD+ endo
Aerocyte
Mid stalk
Late stalk
Alveolar fibro
AT1
Myofibro 3
GRHL+ NE
Late tip
Ciliated
Club
Mid tip
Adventitial fibro
AT2
Mid meso
Late meso

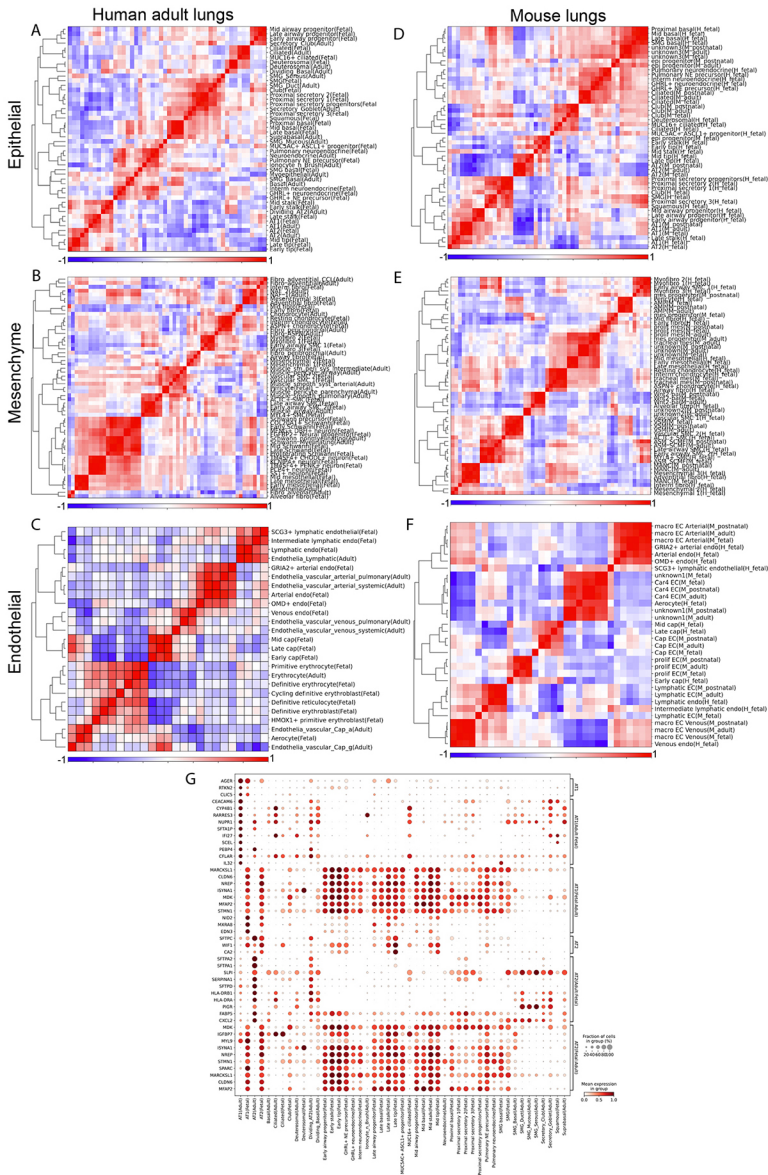

Human adult lungs
Mouse lungs
Epithelial
Mesenchyme
Endothelial
A
Mid airway progenitor(Fetal)
Late airway progenitor(Fetal)
Early airway progenitor(Fetal)
Secretory_Club(Adult)
Ciliated(Fetal)
Ciliated(Adult)
MUC16+ ciliated(Fetal)
Deuterosomal(Fetal)
Deuterosomal(Adult)
Dividing_Basal(Adult)
SMG_Serous(Adult)
SMG(Fetal)
SMG_Duct(Adult)
Club(Fetal)
Proximal secretory 2(Fetal)
Proximal secretory 1(Fetal)
Proximal secretory progenitors(Fetal
Secretory_Goblet(Adult)
Proximal secretory 3(Fetal)
Squamous(Fetal)
Proximal basal(Fetal)
Mid basal(Fetal)
Late basal(Fetal)
Suprabasal(Adult)
SMG_Mucous(Adult)
MUC5AC+ ASCL1+ progenitor(Fetal)
Pulmonary neuroendocrine(Fetal)
Neuroendocrine(Adult)
Pulmonary NE precursor(Fetal)
Ionocyte_n_Brush(Adult)
SMG basal(Fetal)
Myoepithelial(Adult)
SMG_Basal(Adult)
Basal(Adult)
Interm neuroendocrine(Fetal)
GHRL+ neuroendocrine(Fetal)
GHRL+ NE precursor(Fetal)
Mid stalk(Fetal)
Early stalk(Fetal)
Dividing_AT2(Adult)
Late stalk(Fetal)
AT1(Fetal)
AT1(Adult)
AT2(Fetal)
AT2(Adult)
Mid tip(Fetal)
Late tip(Fetal)
Early tip(Fetal)
-1
1
B
Fibro_adventitial_CCL(Adult)
Fibro-adventitial(Adult)
Interm fibro(Fetal)
NAF-2(Adult)
NAF-1(Adult)
Mesenchymal 3(Fetal)
Adventitial fibro(Fetal)
Mid fibro(Fetal)
Early fibro(Fetal)
Chondrocyte(Adult)
Resting chondrocyte(Fetal)
Interm chondrocyte(Fetal)
ASPN+ chondrocyte(Fetal)
Fibro_perichondrial(Adult)
Fibro-ASPN(Adult)
Myofibro 2(Fetal)
Myofibro 1(Fetal)
Early airway SMC 1(Fetal)
Myofibro 3(Fetal)
Fibro_peribronchial(Adult)
Airway fibro(Fetal)
Mesenchymal 2(Fetal)
Mesenchymal 1(Fetal)
Muscle_sm_peri_sys_intermediate(Adult)
Muscle-pericyte-airway(Adult)
Vascular SMC 2(Fetal)
Vascular SMC 1(Fetal)
Muscle_smooth_syst_arterial(Adult)
Pericyte(Fetal)
Muscle_pericyte_parenchyma(Adult)
Muscle-smooth_pulmonary(Adult)
ACTC+ SMC(Fetal)
Late airway SMC(Fetal)
Early airway SMC 2(Fetal)
Muscle_airway(Adult)
MYL4+ SMC(Fetal)
Schwann precursor(Fetal)
COL20A1+ Schwann(Fetal)
Early Schwann(Fetal)
MFNG+ DBH+ neuron(Fetal)
EGFBP2+ Neural progenitor(Fetal)
Schwann_nonmyelinating(Adult)
Schwann-Myelinating(Adult)
Mid Schwann(Fetal)
Late Schwann(Fetal)
Proliferating Schwann(Fetal)
TM4SF4+ CHODL+ neuron(Fetal)
KCNIP4+ neuron(Fetal)
TM4SF4+ PENK+ neuron(Fetal)
PCP4+ neuron(Fetal)
SST+ neuron(Fetal)
Mid mesothelial(Fetal)
Late mesothelial(Fetal)
Early mesothelial(Fetal)
Mesothelial(Adult)
Fibro_alveolar(Adult)
Alveolar fibro(Fetal)
-1
1
C
SCG3+ lymphatic endothelial(Fetal)
Intermediate lymphatic endo(Fetal)
Lymphatic endo(Fetal)
Endothelia_Lymphatic(Adult)
GRIA2+ arterial endo(Fetal)
Endothelia_vascular_arterial_pulmonary(Adult)
Endothelia_vascular_arterial_systemic(Adult)
Arterial endo(Fetal)
OMD+ endo(Fetal)
Venous endo(Fetal)
Endothelia_vascular_venous_pulmonary(Adult)
Endothelia_vascular_venous_systemic(Adult)
Mid cap(Fetal)
Late cap(Fetal)
Early cap(Fetal)
Primitive erythrocyte(Fetal)
Erythrocyte(Adult)
Definitive erythrocyte(Fetal)
Cycling definitive erythroblast(Fetal)
Definitive reticulocyte(Fetal)
Definitive erythroblast(Fetal)
HMOX1+ primitive erythroblast(Fetal)
Endothelia_vascular_Cap_a(Adult)
Aerocyte(Fetal)
Endothelia_vascular_Cap_g(Adult)
-1
1
D
Proximal basal(H_fetal)
Mid basal(H_fetal)
Late basal(H_fetal)
SMG basal(H_fetal)
unknown3(M_postnatal)
unknown3(M_adult)
unknown3(M_fetal)
epi progenitor(M_postnatal)
epi progenitor(M_adult)
Pulmonary neuroendocrine(H_fetal)
Pulmonary NE precursor(H_fetal)
Interm neuroendocrine(H_fetal)
GHRL+ neuroendocrine(H_fetal)
GHRL+ NE precursor(H_fetal)
Ciliated(M_postnatal)
Ciliated(M_adult)
Ciliated(M_fetal)
Club(M_postnatal)
Club(M_adult)
Club(M_fetal)
Deuterosomal(H_fetal)
MUC16+ ciliated(H_fetal)
Ciliated(H_fetal)
MUC5AC+ ASCL1+ progenitor(H_fetal)
epi progenitor(M_fetal)
Early stalk(H_fetal)
Early tip(H_fetal)
Mid stalk(H_fetal)
Mid tip(H_fetal)
Late tip(H_fetal)
AT2(M_postnatal)
AT2(M_adult)
AT2(M_fetal)
Proximal secretory progenitors(H_fetal)
Proximal secretory 2(H_fetal)
Proximal secretory 1(H_fetal)
Club(H_fetal)
SMG(H_fetal)
Proximal secretory 3(H_fetal)
Squamous(H_fetal)
Mid airway progenitor(H_fetal)
Late airway progenitor(H_fetal)
Early airway progenitor(H_fetal)
AT1(M_postnatal)
AT1(M_adult)
AT1(M_fetal)
Late stalk(H_fetal)
AT1(H_fetal)
AT2(H_fetal)
-1
1
E
Myofibro 2(H_fetal)
Myofibro 1(H_fetal)
Early airway SMC 1(H_fetal)
Myofibro 3(H_fetal)
mes progenitor(M_postnatal)
Pericyte(H_fetal)
AMP(M_fetal)
AMP(M_postnatal)
AMP(M_adult)
mes progenitor(M_fetal)
Mid fibro(H_fetal)
Early fibro(H_fetal)
prolif mes(M_postnatal)
prolif mes(M_fetal)
prolif mes(M_adult)
mes progenitor(M_adult)
tracheal mes(M_adult)
unknown(M_postnatal)
unknown(M_adult)
unknown(M_fetal)
Mid mesothelial(H_fetal)
Early mesothelial(H_fetal)
Late mesothelial(H_fetal)
Resting chondrocyte(H_fetal)
Interm chondrocyte(H_fetal)
tracheal mes(M_fetal)
tracheal mes(M_postnatal)
ASPN+ chondrocyte(H_fetal)
Airway fibro(H_fetal)
Wnt2-ba(M_postnatal)
Wnt2-ba(M_fetal)
Wnt2-ba(M_adult)
Alveolar fibro(H_fetal)
unknown2(M_postnatal)
unknown2(M_adult)
Vascular SMC 1(H_fetal)
VSM(M_fetal)
VSM(M_postnatal)
VSM(M_adult)
Vascular SMC 2(H_fetal)
ACTC+ SMC(H_fetal)
ASM_SCMF(M_postnatal)
ASM_SCMF(M_adult)
Late airway SMC(H_fetal)
Early airway SMC 2(H_fetal)
MYL4+ SMC(H_fetal)
ASM_SCMF(M_fetal)
MANC(M_postnatal)
MANC(M_adult)
Mesenchymal 3(H_fetal)
Adventitial fibro(H_fetal)
MANC(M_fetal)
Interm fibro(H_fetal)
Mesenchymal 2(H_fetal)
Mesenchymal 1(H_fetal)
-1
1
F
macro EC Arterial(M_postnatal)
macro EC Arterial(M_adult)
macro EC Arterial(M_fetal)
GRIA2+ arterial endo(H_fetal)
Arterial endo(H_fetal)
OMD+ endo(H_fetal)
SCG3+ lymphatic endothelial(H_fetal)
unknown1(M_fetal)
Car4 EC(M_fetal)
Car4 EC(M_postnatal)
Car4 EC(M_adult)
Aerocyte(H_fetal)
unknown1(M_postnatal)
unknown1(M_adult)
Mid cap(H_fetal)
Late cap(H_fetal)
Cap EC(M_postnatal)
Cap EC(M_adult)
Cap EC(M_fetal)
prolif EC(M_postnatal)
prolif EC(M_adult)
prolif EC(M_fetal)
Early cap(H_fetal)
Lymphatic EC(M_postnatal)
Lymphatic EC(M_adult)
Lymphatic endo(H_fetal)
Intermediate lymphatic endo(H_fetal)
Lymphatic EC(M_fetal)
macro EC Venous(M_postnatal)
macro EC Venous(M_adult)
macro EC Venous(M_fetal)
Venous endo(H_fetal)
-1
1
G
AGER
RTKN2
CLIC5
CEACAM6
CYP4B1
RARRES3
NUPR1
SFTA1P
IFI27
SCEL
PEBP4
CFLAR
IL32
MARCKSL1
CLDN6
NREP
ISYNA1
MDK
MFAP2
STMN1
NID2
MXRA8
EDN3
SFTPC
WIF1
CA2
SFTPA2
SFTPA1
SLPI
SERPINA1
SFTPD
HLA-DRB1
HLA-DRA
PIGR
FABP5
CXCL2
MDK
IGFBP7
MYL9
ISYNA1
NREP
STMN1
SPARC
MARCKSL1
CLDN6
MFAP2
AT1
AT1(Adult:Fetal)
AT1(Fetal:Adult)
AT2
AT2(Adult:Fetal)
AT2(Fetal:Adult)
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
AT1(Adult)
AT1(Fetal)
AT2(Adult)
AT2(Fetal)
Basal(Adult)
Ciliated(Adult)
Ciliated(Fetal)
Club(Fetal)
Deuterosomal(Adult)
Deuterosomal(Fetal)
Dividing_AT2(Adult)
Dividing_Basal(Adult)
Early airway progenitor(Fetal)
Early stalk(Fetal)
Early tip(Fetal)
GHRL+ NE precursor(Fetal)
GHRL+ neuroendocrine(Fetal)
Interm neuroendocrine(Fetal)
Ionocyte_n_Brush(Adult)
Late airway progenitor(Fetal)
Late basal(Fetal)
Late stalk(Fetal)
Late tip(Fetal)
MUC5AC+ ASCL1+ progenitor(Fetal)
MUC16+ ciliated(Fetal)
Mid airway progenitor(Fetal)
Mid basal(Fetal)
Mid stalk(Fetal)
Mid tip(Fetal)
Neuroendocrine(Adult)
Proximal basal(Fetal)
Proximal secretory 1(Fetal)
Proximal secretory 2(Fetal)
Proximal secretory 3(Fetal)
Proximal secretory progenitors(Fetal)
Pulmonary NE precursor(Fetal)
Pulmonary neuroendocrine(Fetal)
SMG basal(Fetal)
SMG(Fetal)
SMG_Basal(Adult)
SMG_Duct(Adult)
SMG_Mucous(Adult)
SMG_Serous(Adult)
Secretory_Club(Adult)
Secretory_Goblet(Adult)
Squamous(Fetal)
Suprabasal(Adult)

Figure S4

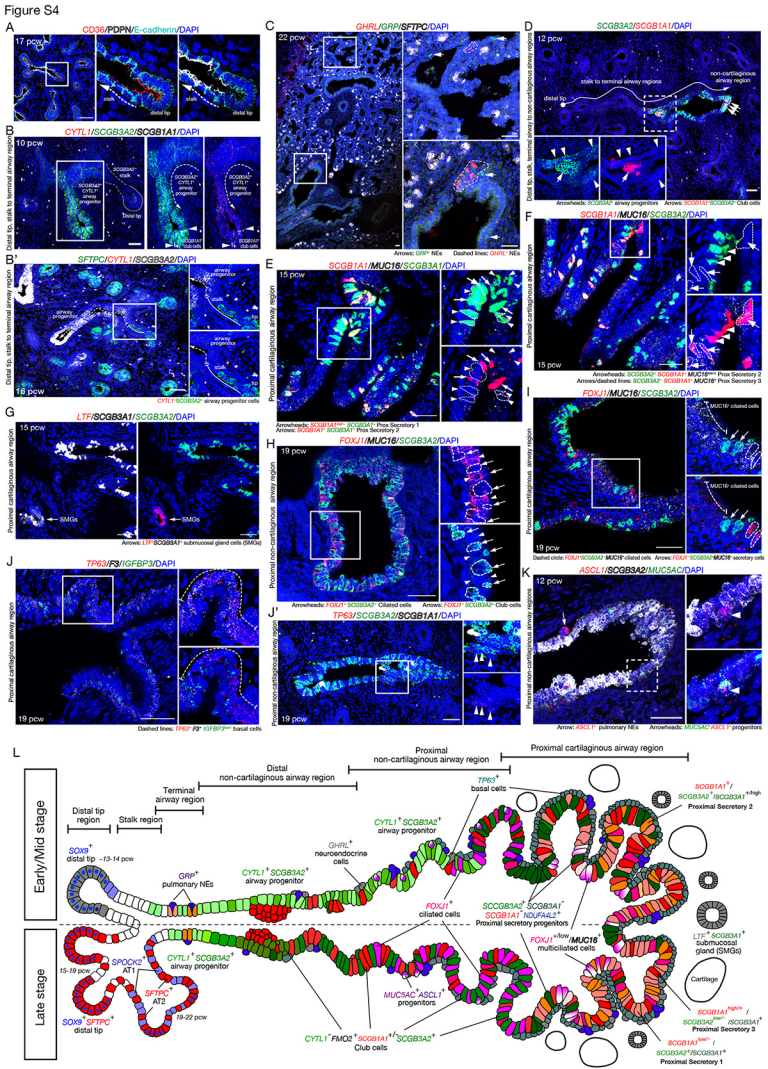

A
17 pcw
CD36/PDPN/E-cadherin/DAPI
distal tip
stalk
B
10 pcw
CYTL1/SCGB3A2/SCGB1A1/DAPI
Distal tip, stalk to terminal airway region
SCGB3A2+ stalk
SCGB3A2+ CYTL1+ airway progenitor
Distal tip
SCGB3A2+ tip cells
SCGB3A2+ club cells
B'
SFTPC/CYTL1/SCGB3A2/DAPI
Distal tip, stalk to terminal airway region
airway progenitor
stalk
airway progenitor
stalk
16 pcw
CYTL1+ SCGB3A2+ airway progenitor cells
C
22 pcw
GHRL/GRP/SFTPC/DAPI
Arrows: GRP+ NEs
Dashed lines: GHRL+ NEs
D
12 pcw
SCGB3A2/SCGB1A1/DAPI
distal tip
stalk to terminal airway regions
non-cartilaginous airway region
Distal tip, stalk, terminal airway to non-cartilaginous airway regions
Arrowheads: SCGB3A2+ airway progenitors
Arrows: SCGB1A1+ SCGB3A2+ Club cells
E
15 pcw
SCGB1A1/MUC16/SCGB3A1/DAPI
Proximal cartilaginous airway region
Arrowheads: SCGB1A1low SCGB3A1+ Prox Secretory 1
Arrows: SCGB1A1+ SCGB3A1+ Prox Secretory 2
F
SCGB1A1/MUC16/SCGB3A2/DAPI
Proximal cartilaginous airway region
15 pcw
Arrowheads: SCGB3A2+ SCGB1A1low MUC16low Prox Secretory 2
Arrows/dashed lines: SCGB3A2+ SCGB1A1+ MUC16+ Prox Secretory 3
G
15 pcw
LTF/SCGB3A1/SCGB3A2/DAPI
Proximal cartilaginous airway region
SMGs
SMGs
Arrows: LTF+ SCGB3A1+ submucosal gland cells (SMGs)
H
19 pcw
FOXJ1/MUC16/SCGB3A2/DAPI
Proximal non-cartilaginous airway region
Arrowheads: FOXJ1+ SCGB3A2+ Ciliated cells
Arrows: FOXJ1+ SCGB3A2+ Club cells
I
FOXJ1/MUC16/SCGB3A2/DAPI
Proximal cartilaginous airway region
MUC16+ ciliated cells
MUC16+ ciliated cells
19 pcw
Dashed circle: FOXJ1+SCGB3A2+ MUC16+ ciliated cells
Arrows: FOXJ1+ SCGB3A2+MUC16+ secretory cells
J
TP63/F3/IGFBP3/DAPI
Proximal cartilaginous airway region
19 pcw
Dashed lines: TP63+ F3+ IGFBP3low basal cells
J'
TP63/SCGB3A2/SCGB1A1/DAPI
Proximal non-cartilaginous airway region
19 pcw
K
12 pcw
ASCL1/SCGB3A2/MUC5AC/DAPI
Proximal non-cartilaginous airway regions
Arrow: ASCL1+ pulmonary NEs
Arrowheads: MUC5AC+ASCL1+ progenitors
L
Early/Mid stage
Late stage
Distal tip region
Stalk region
Terminal airway region
Distal non-cartilaginous airway region
Proximal non-cartilaginous airway region
Proximal cartilaginous airway region
SOX9+ distal tip ~13-14 pcw
GRP+ pulmonary NEs
CYTL1+ SCGB3A2+ airway progenitor
GHRL+ neuroendocrine cells
TP63+ basal cells
SCGB1A1+ SCGB3A2+ SCGB3A1+/+high Proximal Secretory 2
FOXJ1+ ciliated cells
SCGB3A2+ SCGB3A1+ SCGB1A1- NDUFAHL2+ Proximal secretory progenitors
LTF+ SCGB3A1+ submucosal gland (SMGs)
Cartilage
FOXJ1low/MUC16+ multiciliated cells
SCGB1A1high/+ / SCGB3A2low/- SCGB3A1+ Proximal Secretory 3
SCGB1A1low/- / SCGB3A2+ SCGB3A1+ Proximal Secretory 1
15-19 pcw
SPOCK2+ AT1
SFTPC+ AT2
SOX9+ SFTPC+ distal tip
19-22 pcw
CYTL1+ SCGB3A2+ airway progenitor
MUC5AC+ ASCL1+ progenitors
CYTL1+ FMO2+ SCGB3A1+/- SCGB3A2+ Club cells

Figure S5

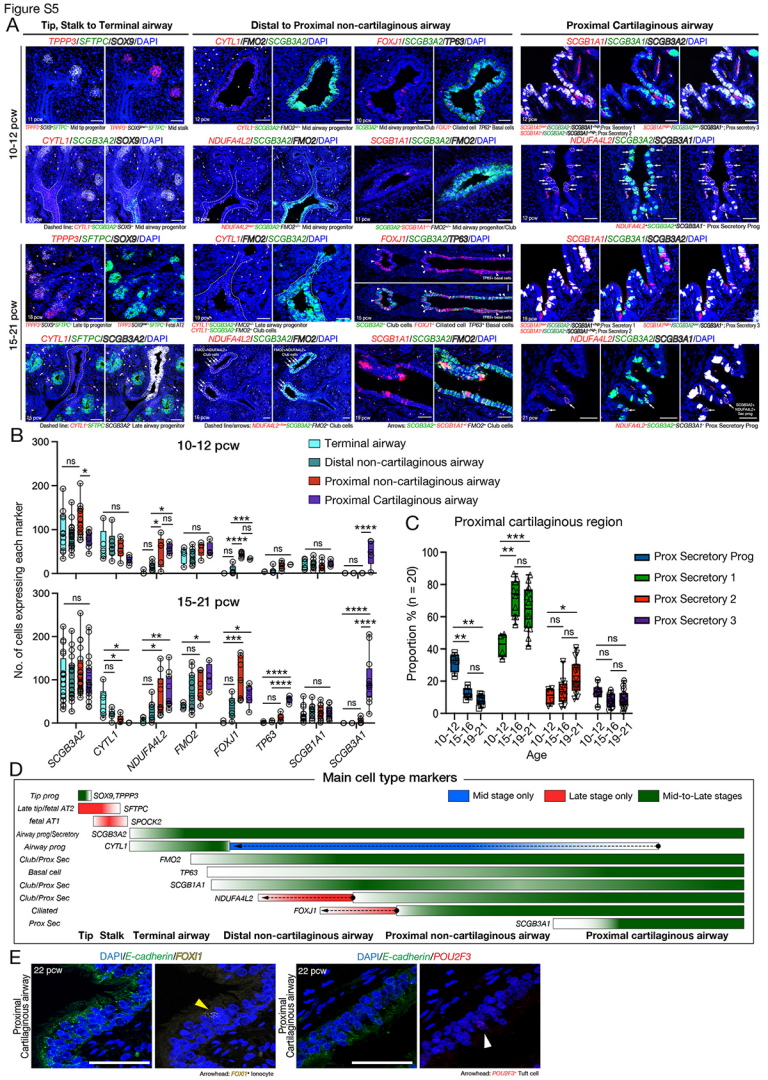

A
Tip, Stalk to Terminal airway
Distal to Proximal non-cartilaginous airway
Proximal Cartilaginous airway
10-12 pcw
15-21 pcw
TPPP3/SFTPC/SOX9/DAPI
CYTL1/FMO2/SCGB3A2/DAPI
FOXJ1/SCGB3A2/TP63/DAPI
SCGB1A1/SCGB3A1/SCGB3A2/DAPI
CYTL1/SCGB3A2/SOX9/DAPI
NDUFA4L2/SCGB3A2/FMO2/DAPI
SCGB1A1/SCGB3A2/FMO2/DAPI
NDUFA4L2/SCGB3A2/SCGB3A1/DAPI
B
10-12 pcw
15-21 pcw
Terminal airway
Distal non-cartilaginous airway
Proximal non-cartilaginous airway
Proximal Cartilaginous airway
No. of cells expressing each marker
SCGB3A2
CYTL1
NDUFA4L2
FMO2
FOXJ1
TP63
SCGB1A1
SCGB3A1
C
Proximal cartilaginous region
Proportion % (n = 20)
Age
Prox Secretory Prog
Prox Secretory 1
Prox Secretory 2
Prox Secretory 3
D
Main cell type markers
Mid stage only
Late stage only
Mid-to-Late stages
Tip prog
Late tip/fetal AT2
fetal AT1
Airway prog/Secretory
Airway prog
Club/Prox Sec
Basal cell
Club/Prox Sec
Club/Prox Sec
Ciliated
Prox Sec
SOX9,TPPP3
SFTPC
SPOCK2
SCGB3A2
CYTL1
FMO2
TP63
SCGB1A1
NDUFA4L2
FOXJ1
SCGB3A1
Tip
Stalk
Terminal airway
Distal non-cartilaginous airway
Proximal non-cartilaginous airway
Proximal cartilaginous airway
E
DAPI/E-cadherin/FOXI1
DAPI/E-cadherin/POU2F3
22 pcw
Proximal Cartilaginous airway
Arrowhead: FOXI1+ Ionocyte
Arrowhead: POU2F3+ Tuft cell

Figure S6

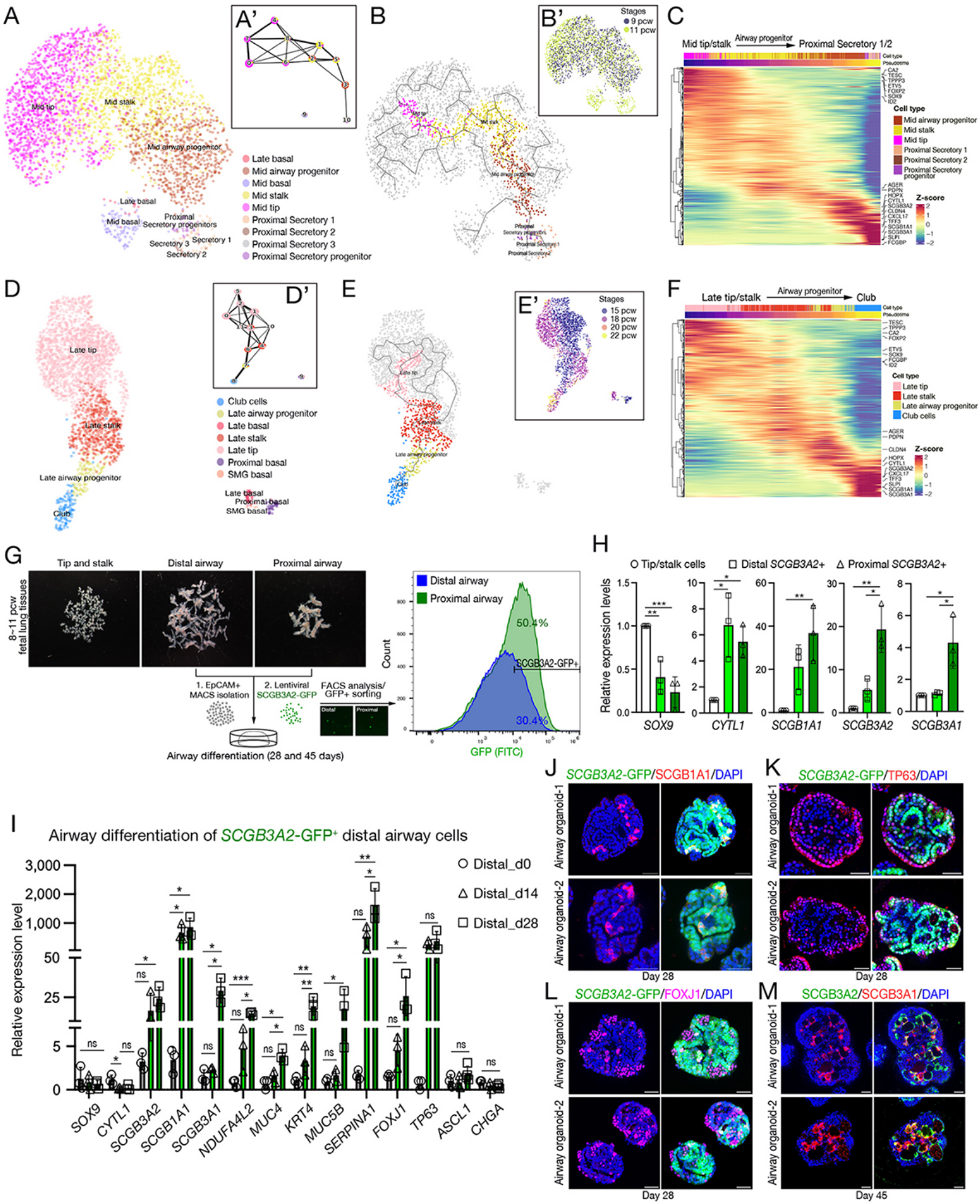

A
A'
B
B'
Stages
9 pcw
11 pcw
C
Mid tip/stalk
Airway progenitor
Proximal Secretory 1/2
Cell type
Mid airway progenitor
Mid stalk
Mid tip
Proximal Secretory 1
Proximal Secretory 2
Proximal Secretory progenitor
Z-score
Mid tip
Mid stalk
Mid airway progenitor
Late basal
Mid basal
Proximal Secretory progenitors
Secretory 3
Secretory 1
Secretory 2
Late basal
Mid airway progenitor
Mid basal
Mid stalk
Mid tip
Proximal Secretory 1
Proximal Secretory 2
Proximal Secretory 3
Proximal Secretory progenitor
D
D'
E
E'
Stages
15 pcw
18 pcw
20 pcw
22 pcw
F
Late tip/stalk
Airway progenitor
Club
Cell type
Late tip
Late stalk
Late airway progenitor
Club cells
Z-score
Late tip
Late stalk
Late airway progenitor
Club
Club cells
Late airway progenitor
Late basal
Late stalk
Late tip
Proximal basal
SMG basal
Late basal
Proximal basal
SMG basal
G
Tip and stalk
Distal airway
Proximal airway
8~11 pcw fetal lung tissues
1. EpCAM+ MACS isolation
2. Lentiviral SCGB3A2-GFP
FACS analysis/ GFP+ sorting
Distal
Proximal
Airway differentiation (28 and 45 days)
Distal airway
Proximal airway
50.4%
30.4%
SCGB3A2-GFP+
Count
GFP (FITC)
H
Tip/stalk cells
Distal SCGB3A2+
Proximal SCGB3A2+
Relative expression levels
SOX9
CYTL1
SCGB1A1
SCGB3A2
SCGB3A1
I
Airway differentiation of SCGB3A2-GFP+ distal airway cells
Distal_d0
Distal_d14
Distal_d28
Relative expression level
SOX9
CYTL1
SCGB3A2
SCGB1A1
SCGB3A1
NDUFA4L2
MUC4
KRT4
MUC5B
SERPINA1
FOXJ1
TP63
ASCL1
CHGA
J
SCGB3A2-GFP/SCGB1A1/DAPI
Airway organoid-1
Airway organoid-2
Day 28
K
SCGB3A2-GFP/TP63/DAPI
Airway organoid-1
Airway organoid-2
Day 28
L
SCGB3A2-GFP/FOXJ1/DAPI
Airway organoid-1
Airway organoid-2
Day 28
M
SCGB3A2/SCGB3A1/DAPI
Airway organoid-1
Airway organoid-2
Day 45

Figure S7

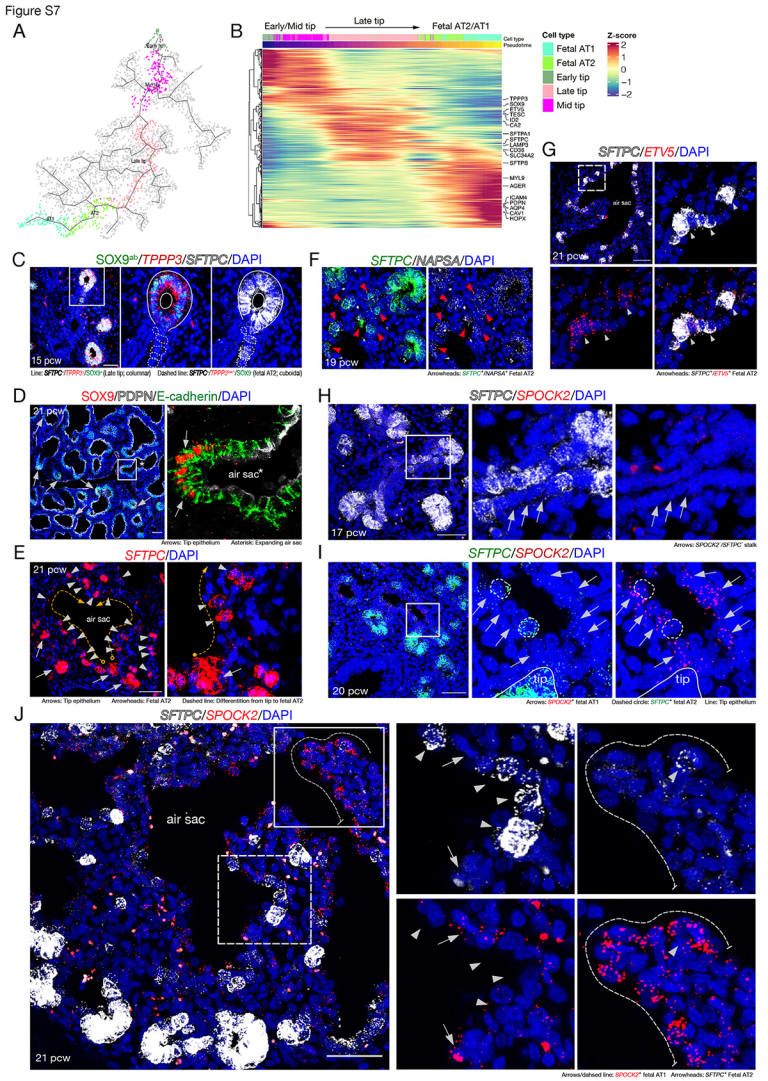

A
Early tip
Mid tip
Late tip
AT1
AT2
B
Early/Mid tip
Late tip
Fetal AT2/AT1
Cell type
Pseudotime
Cell type
Fetal AT1
Fetal AT2
Early tip
Late tip
Mid tip
Z-score
2
1
0
−1
−2
TPPP3
SOX9
ETV5
TESC
CA2
CA4
SFTPA1
SFTPC
LAMP3
CD68
SLC34A2
SFTPB
MYL9
AGER
ICAM1
PECAM1
AQP4
CAV1
HOPX
C
SOX9ab/TPPP3/SFTPC/DAPI
15 pcw
Line: SFTPC+/TPPP3+/SOX9+ (Late tip; columnar)
Dashed line: SFTPC+/TPPP3low/SOX9+ (fetal AT2; cuboidal)
F
SFTPC/NAPSA/DAPI
19 pcw
Arrowheads: SFTPC+/NAPSA+ Fetal AT2
G
SFTPC/ETV5/DAPI
air sac
21 pcw
Arrowheads: SFTPC+/ETV5+ Fetal AT2
D
SOX9/PDPN/E-cadherin/DAPI
21 pcw
air sac*
Arrows: Tip epithelium
Asterisk: Expanding air sac
H
SFTPC/SPOCK2/DAPI
17 pcw
Arrows: SPOCK2+/SFTPC− stalk
E
SFTPC/DAPI
21 pcw
air sac
Arrows: Tip epithelium
Arrowheads: Fetal AT2
Dashed line: Differentiation from tip to fetal AT2
I
SFTPC/SPOCK2/DAPI
20 pcw
tip
tip
Arrows: SPOCK2+ fetal AT1
Dashed circle: SFTPC+ fetal AT2
Line: Tip epithelium
J
SFTPC/SPOCK2/DAPI
air sac
21 pcw
Arrows/dashed line: SPOCK2+ fetal AT1
Arrowheads: SFTPC+ Fetal AT2

Figure S8

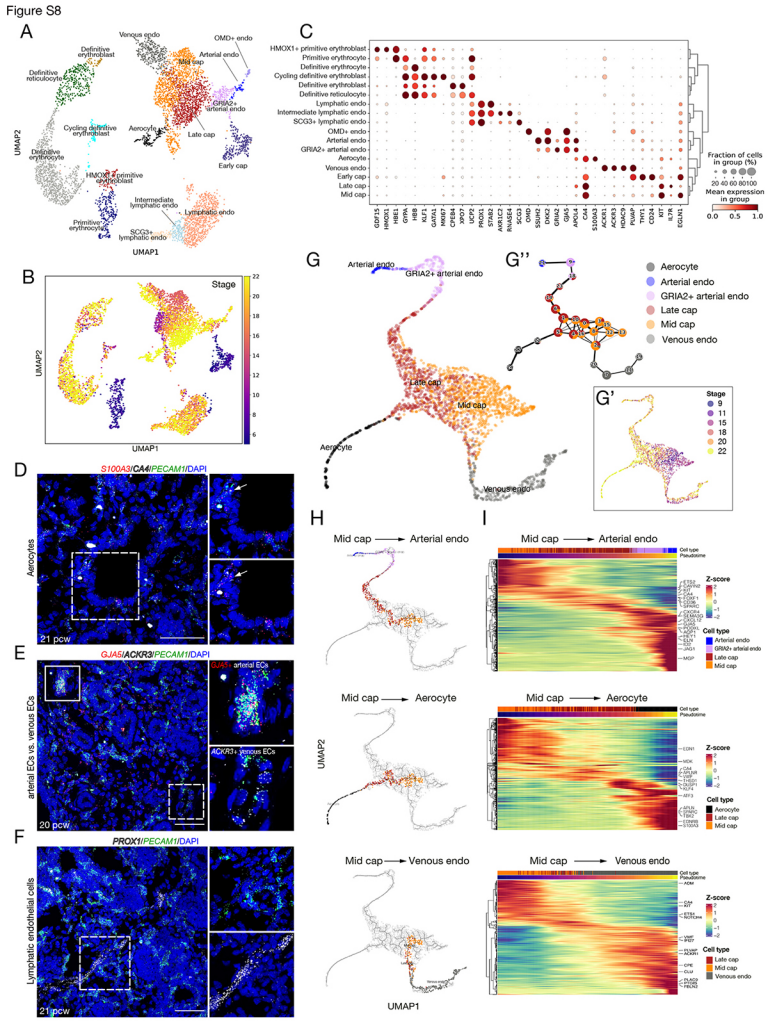

A
Venous endo
OMD+ endo
Definitive erythroblast
Arterial endo
Mid cap
Definitive reticulocyte
GRIA2+ arterial endo
Cycling definitive erythroblast
Aerocyte
Late cap
UMAP2
Early cap
Definitive erythrocyte
HMOX1+ primitive erythroblast
Intermediate lymphatic endo
Lymphatic endo
Primitive erythrocyte
SCG5+ lymphatic endo
UMAP1
C
HMOX1+ primitive erythroblast
Primitive erythrocyte
Definitive erythrocyte
Cycling definitive erythroblast
Definitive erythroblast
Definitive reticulocyte
Lymphatic endo
Intermediate lymphatic endo
SCG5+ lymphatic endo
OMD+ endo
Arterial endo
GRIA2+ arterial endo
Aerocyte
Venous endo
Early cap
Late cap
Mid cap
GATA1
HMOX1
HBE1
HBZ
HBB
KLF1
GATA2
MKI67
CPEB4
TOP2A
UCP2
PROX1
SCG5
CCL21
ANXA1/2
RNASE4
OMD
SOX18
BMX
GRIA2
GJA5
APOLD1
S100A3
ACKR3
HEY1
HDAC9
PLVAP
PECAM1
CD34
KIT
IL7R
EGFL7
Fraction of cells in group (%)
20 40 60 80100
Mean expression in group
0.0 0.5 1.0
B
Stage
UMAP2
UMAP1
G
Arterial endo
GRIA2+ arterial endo
Late cap
Mid cap
Aerocyte
Venous endo
G''
G'
Aerocyte
Arterial endo
GRIA2+ arterial endo
Late cap
Mid cap
Venous endo
Stage
D
S100A3/CA4/PECAM1/DAPI
Aerocytes
21 pcw
E
GJA5/ACKR3/PECAM1/DAPI
arterial ECs vs. venous ECs
GJA5+ arterial ECs
ACKR3+ venous ECs
20 pcw
F
PROX1/PECAM1/DAPI
Lymphatic endothelial cells
21 pcw
H
Mid cap → Arterial endo
Mid cap → Aerocyte
Mid cap → Venous endo
UMAP2
UMAP1
I
Mid cap → Arterial endo
Mid cap → Aerocyte
Mid cap → Venous endo
Cell type Pseudotime
Z-score
Cell type
Arterial endo
GRIA2+ arterial endo
Late cap
Mid cap
Aerocyte
Venous endo

Figure S9

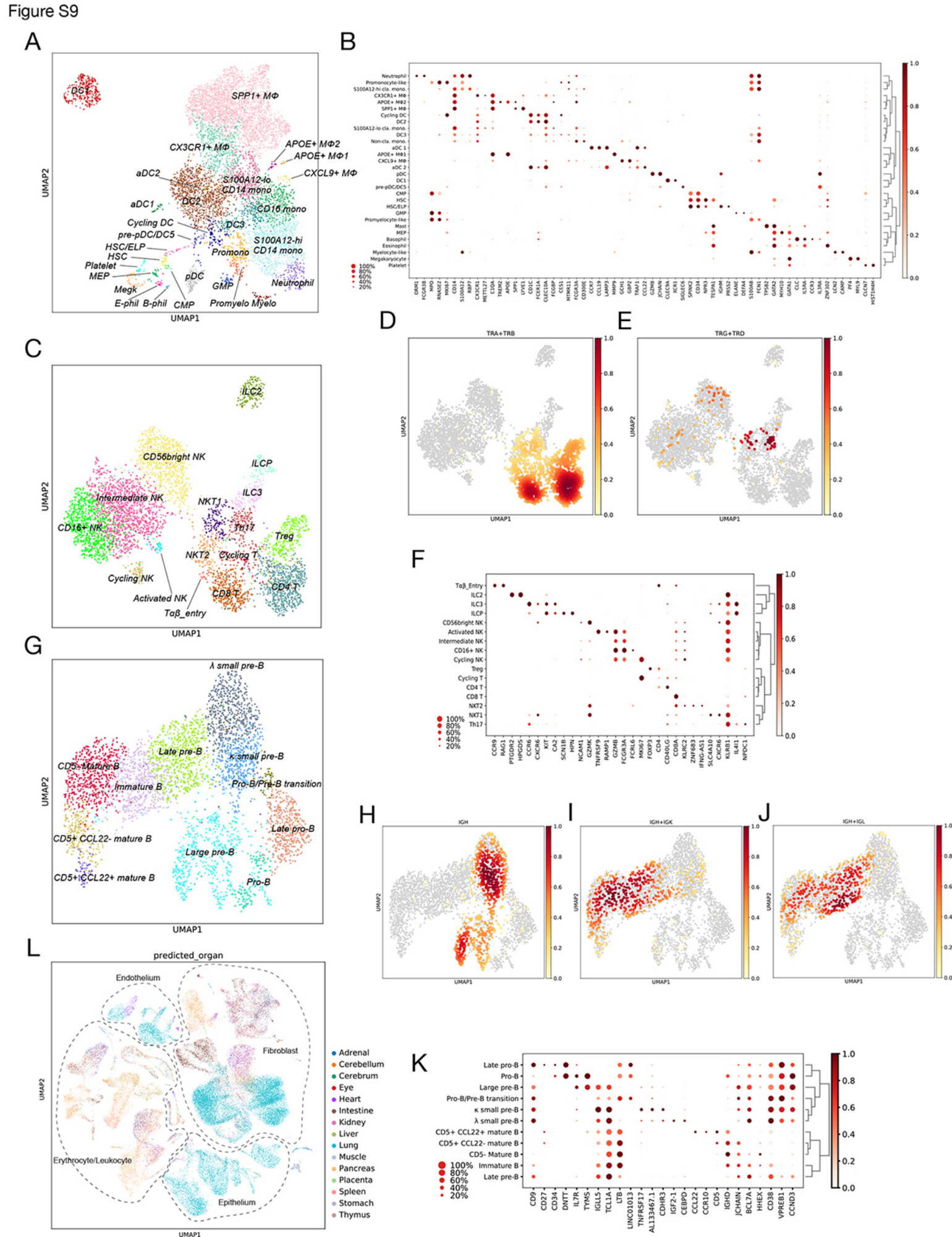

A
DC1
SPP1+ MΦ
CX3CR1+ MΦ
APOE+ MΦ2
APOE+ MΦ1
CXCL9+ MΦ
aDC2
S100A12-lo CD14 mono
aDC1
DC2
CD16 mono
Cycling DC
DC3
pre-pDC/DC5
S100A12-hi CD14 mono
HSC/ELP
HSC
Promono
Platelet
MEP
pDC
GMP
Neutrophil
Megk
E-phil
B-phil
CMP
Promyelo
Myelo
UMAP2
UMAP1
B
C
ILC2
CD56bright NK
ILCP
Intermediate NK
NKT1
ILC3
CD16+ NK
Th17
Treg
NKT2
Cycling T
Cycling NK
CD4 T
Activated NK
CD8 T
Tαβ_entry
D
TRA+TRB
E
TRG+TRD
F
Tαβ_Entry
ILC2
ILC3
ILCP
CD56bright NK
Activated NK
Intermediate NK
CD16+ NK
Cycling NK
Treg
Cycling T
CD4 T
CD8 T
NKT2
NKT1
Th17
100%
80%
60%
40%
20%
G
λ small pre-B
Late pre-B
κ small pre-B
CD5- Mature B
Immature B
Pro-B/Pre-B transition
Late pro-B
CD5+ CCL22- mature B
Large pre-B
CD5+ CCL22+ mature B
Pro-B
H
IGH
I
IGH+IGK
J
IGH+IGL
L
predicted_organ
Endothelium
Fibroblast
Erythrocyte/Leukocyte
Epithelium
Adrenal
Cerebellum
Cerebrum
Eye
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Placenta
Spleen
Stomach
Thymus
K
Late pro-B
Pro-B
Large pre-B
Pro-B/Pre-B transition
κ small pre-B
λ small pre-B
CD5+ CCL22+ mature B
CD5+ CCL22- mature B
CD5- Mature B
Immature B
Late pre-B

Figure S10

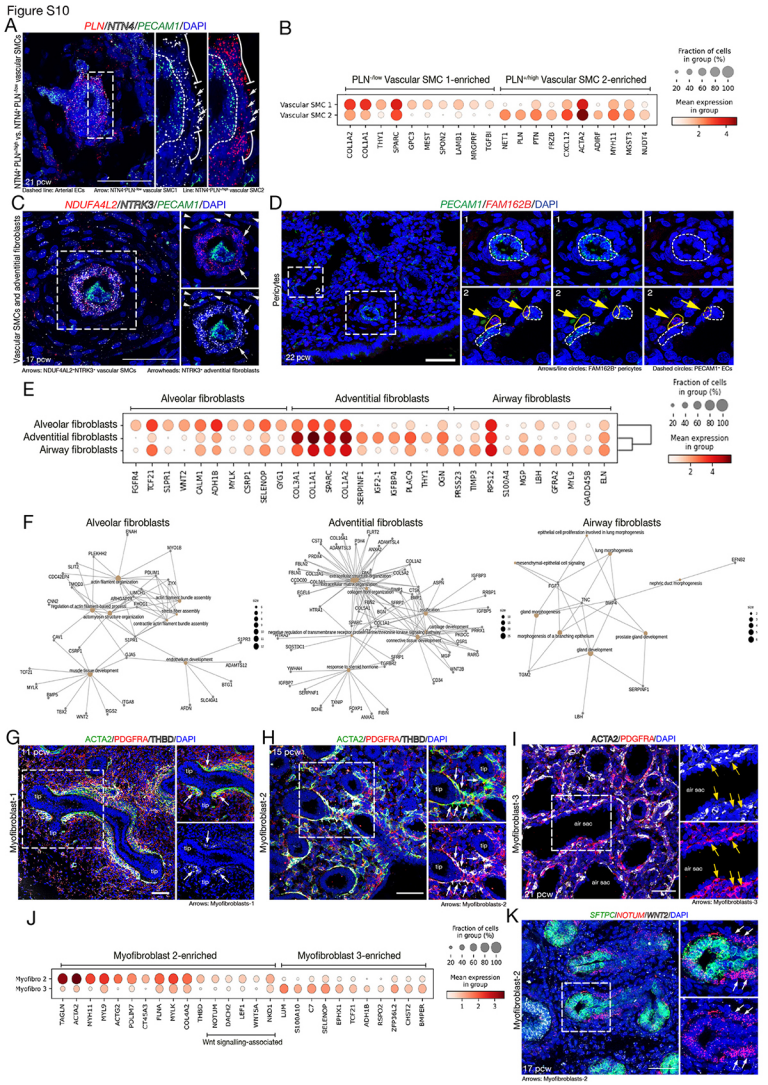

A
PLN/NTN4/PECAM1/DAPI
NTN4+ PLN-low vs. NTN4+ PLN+high vascular SMCs
21 pcw
Dashed line: Arterial ECs
Arrow: NTN4+PLN-low vascular SMC1
Line: NTN4+PLN+high vascular SMC2
B
PLN-low Vascular SMC 1-enriched
PLN+high Vascular SMC 2-enriched
Vascular SMC 1
Vascular SMC 2
COL1A2 COL1A1 TIMP1 SPARC GPC3 MEST SPON2 LAMB1 MRGPRF TGFBI MET1 PLN PTN FRZB CXCL12 ACTA2 ADRF MYH11 MOST3 NUDT4
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
2 4
C
NDUFA4L2/NTRK3/PECAM1/DAPI
Vascular SMCs and adventitial fibroblasts
17 pcw
Arrows: NDUFA4L2+NTRK3+ vascular SMCs
Arrowheads: NTRK3+ adventitial fibroblasts
D
PECAM1/FAM162B/DAPI
Pericytes
22 pcw
Arrows/line circles: FAM162B+ pericytes
Dashed circles: PECAM1+ ECs
E
Alveolar fibroblasts
Adventitial fibroblasts
Airway fibroblasts
FGFR4 TCF21 SLPR1 WNT2 CALB1 ADH1B MYLK CSRP1 SELENOP GYG1 COL3A1 COL1A1 SPARC COL1A2 SERPINF1 KCF2-1 KIFBP4 PLAC9 TH7-1 OCN PRSS23 TIMP3 RPS12 S100A4 MGP LBH GFRA2 MYL9 GADD45B ELN
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
2 4
F
Alveolar fibroblasts
Adventitial fibroblasts
Airway fibroblasts
epithelial cell proliferation involved in lung morphogenesis
lung morphogenesis
mesenchymal-epithelial cell signaling
FGF10
nephric duct morphogenesis
gland morphogenesis
prostate gland development
gland development
morphogenesis of a branching epithelium
TGM2
SERPINF1
LBH
G
ACTA2/PDGFRA/THBD/DAPI
Myofibroblast-1
11 pcw
tip
sp
Arrows: Myofibroblasts-1
H
ACTA2/PDGFRA/THBD/DAPI
Myofibroblast-2
15 pcw
tip
sp
Arrows: Myofibroblasts-2
I
ACTA2/PDGFRA/DAPI
Myofibroblast-3
air sac
17 pcw
Arrows: Myofibroblasts-3
J
Myofibroblast 2-enriched
Myofibroblast 3-enriched
Myofibro 2
Myofibro 3
TAGLN ACTA2 MYH11 MYL9 ACTG2 PDLIM7 CTXS43 FLNA MYLK COL4A2 THBD NDTUH DACH2 LEF1 WNT5A NKD1 LUM S100A10 C7 SELENOP BHD61 TCF21 ADH1B RSPO2 ZFP36L2 CHST2 BMPER
Wnt signalling-associated
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
1 2 3
K
SFTPC/NOTUM/WNT2/DAPI
Myofibroblast-2
17 pcw
Arrows: Myofibroblasts-2

Figure S11

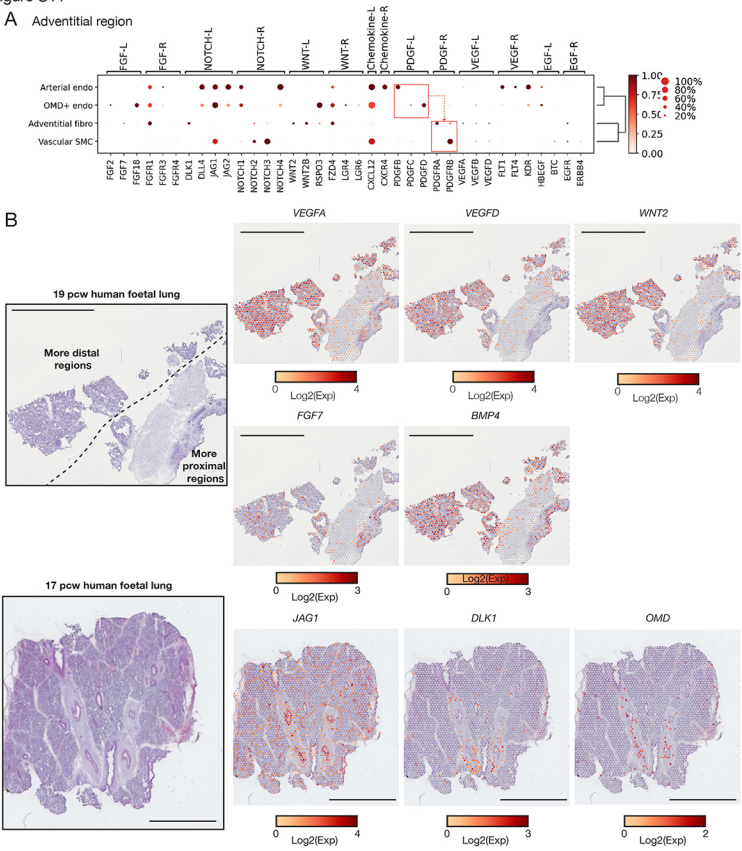

A
Adventitial region
FGF-L
FGF-R
NOTCH-L
NOTCH-R
WNT-L
WNT-R
Chemokine-L
Chemokine-R
PDGF-L
PDGF-R
VEGF-L
VEGF-R
EGF-L
EGF-R
Arterial endo
OMD+ endo
Adventitial fibro
Vascular SMC
100%
80%
60%
40%
20%
1.00
0.75
0.50
0.25
0.00
FGF2
FGF7
FGF18
FGFR1
FGFR3
FGFR4
DLK1
DLL4
JAG1
JAG2
NOTCH1
NOTCH2
NOTCH3
NOTCH4
WNT2
WNT2B
RSPO3
FZD4
LGR4
LGR6
CXCL12
CXCR4
PDGFB
PDGFC
PDGFD
PDGFRA
PDGFRB
VEGFA
VEGFB
VEGFD
FLT1
FLT4
KDR
HBEGF
BTC
EGFR
ERBB4
B
19 pcw human foetal lung
More distal regions
More proximal regions
17 pcw human foetal lung
VEGFA
VEGFD
WNT2
FGF7
BMP4
JAG1
DLK1
OMD
0
4
3
2
Log2(Exp)

Figure S12

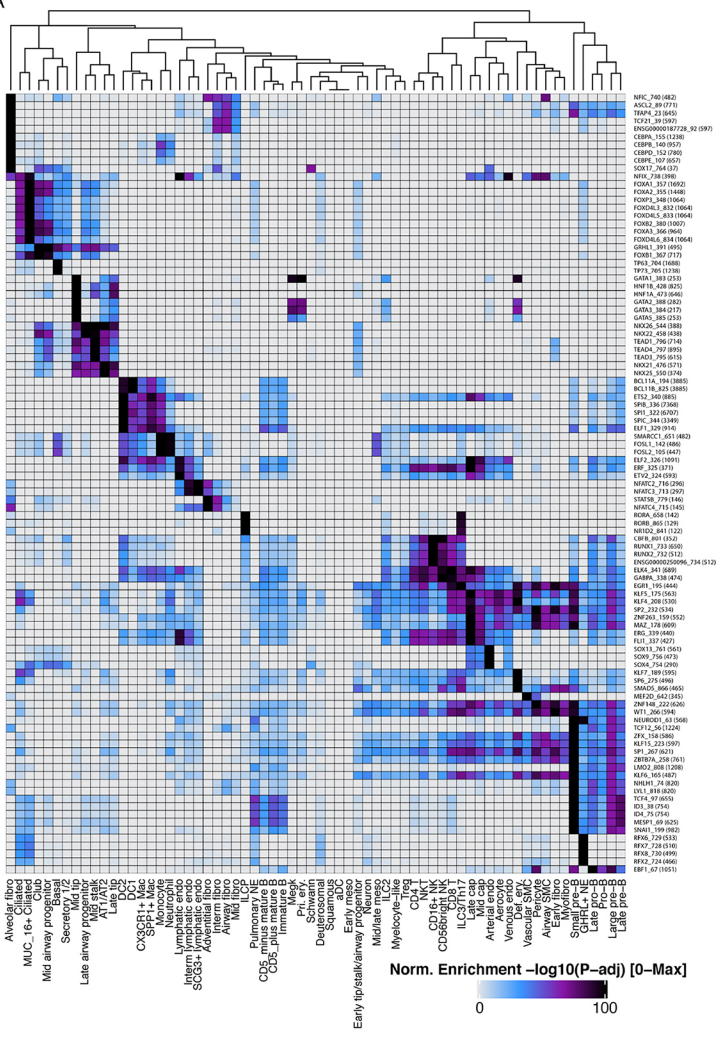

A
NFIC_740 (482)
ASCL2_89 (771)
TFAP4_23 (645)
TCF21_39 (597)
ENSG00000187728_92 (597)
CEBPA_155 (1238)
CEBPB_140 (957)
CEBPD_152 (780)
CEBPE_107 (657)
SOX17_764 (37)
NFIX_738 (398)
FOXA1_357 (1692)
FOXA2_355 (1448)
FOXP3_348 (1064)
FOXD4L3_852 (1064)
FOXD4L5_859 (1064)
FOXB2_380 (1007)
FOXA3_366 (964)
FOXD4L6_834 (1064)
GRHL1_391 (695)
FOXB1_367 (717)
TP63_784 (1688)
TP73_765 (1238)
GATA1_383 (253)
HNF1B_428 (825)
HNF1A_473 (646)
GATA2_388 (282)
GATA3_384 (217)
GATA5_385 (253)
NKX26_544 (388)
NKX22_458 (438)
TEAD1_796 (714)
TEAD4_797 (895)
TEAD3_795 (615)
NKX21_476 (571)
NKX25_550 (374)
BCL11A_194 (3885)
BCL11B_829 (3885)
ETS2_340 (885)
SPIB_336 (7368)
SPI1_322 (6707)
SPIC_344 (3349)
ELF1_329 (914)
SMARCC1_651 (482)
FOSL1_142 (486)
FOSL2_105 (447)
ELF3_326 (1691)
EBF_325 (371)
ETV2_324 (593)
NFATC2_716 (296)
NFATC3_713 (297)
STAT5B_779 (146)
NFATC4_715 (145)
RORA_658 (142)
RORB_865 (129)
NR1D2_841 (122)
CBFB_869 (352)
RUNX1_733 (650)
RUNX2_732 (512)
ENSG00000250996_734 (512)
ELK4_341 (689)
GABPA_338 (474)
EGR3_195 (444)
KLF5_175 (563)
KLF4_208 (530)
SP2_232 (534)
ZNF263_159 (552)
MAZ_178 (669)
ERG_339 (440)
FLI1_337 (427)
SOX13_761 (561)
SOX9_756 (473)
SOX4_754 (290)
KLF7_189 (595)
SP6_275 (496)
SMAD5_866 (465)
MEF2D_642 (345)
ZNF148_222 (626)
WT1_266 (594)
NEUROD1_63 (568)
TCF12_56 (1224)
ZFX_158 (586)
KLF15_223 (597)
SP1_267 (621)
ZBTB7A_258 (761)
LMO2_808 (1208)
KLF6_165 (487)
NHLH1_74 (820)
LYL1_818 (820)
TCF4_97 (655)
ID3_38 (754)
ID4_75 (754)
MESP1_69 (625)
SNAI1_599 (982)
RFX6_729 (553)
RFX7_728 (510)
RFX8_730 (499)
RFX2_724 (466)
EBF1_67 (1051)
Alveolar fibro
Ciliated
MUC_16+ Ciliated
Club
Mid airway progenitor
Basal
Secretory 22
Mid tip
Late airway progenitor
Mid AT1
AT1/AT2
Late tip
DC2
DC1
CX3CR1+ Mac
SPP1+ Mac
Monocyte
Neutrophil
Lymphatic endo
Interm lymphatic endo
SCG3+ Lymphatic endo
Adventitial fibro
Interm fibro
Airway fibro
Mid ILC2
ILC1
Pulmonary NE
CD5+ mature B
CD5- mature B
Immature B
Pre-B
Plasma
Schwann
Deep sensorimal
Sensorimotor
Early meso
Paraxial meso
Neuron
Mid/late meso
Myelocyte–like
CD4 T
NK/T
CD16+ NK
CD56+ NK
CD8 T
ILC3/Th17
Mid cap
Arterial endo
Aerocyte
Venous endo
Del. cap
Vascular SMC
Pericyte
Airway SMC
Myofibro
Small Pre-B
Late pro-B
Pro-B
Large Pre-B
Late pre-B
Norm. Enrichment –log10(P–adj) [0–Max]
0
100

Figure S13

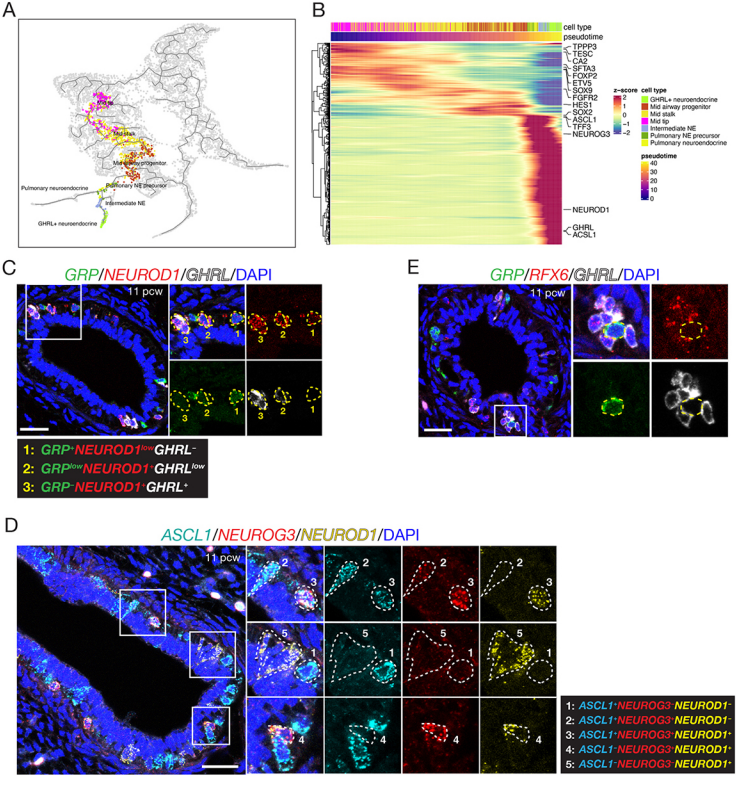
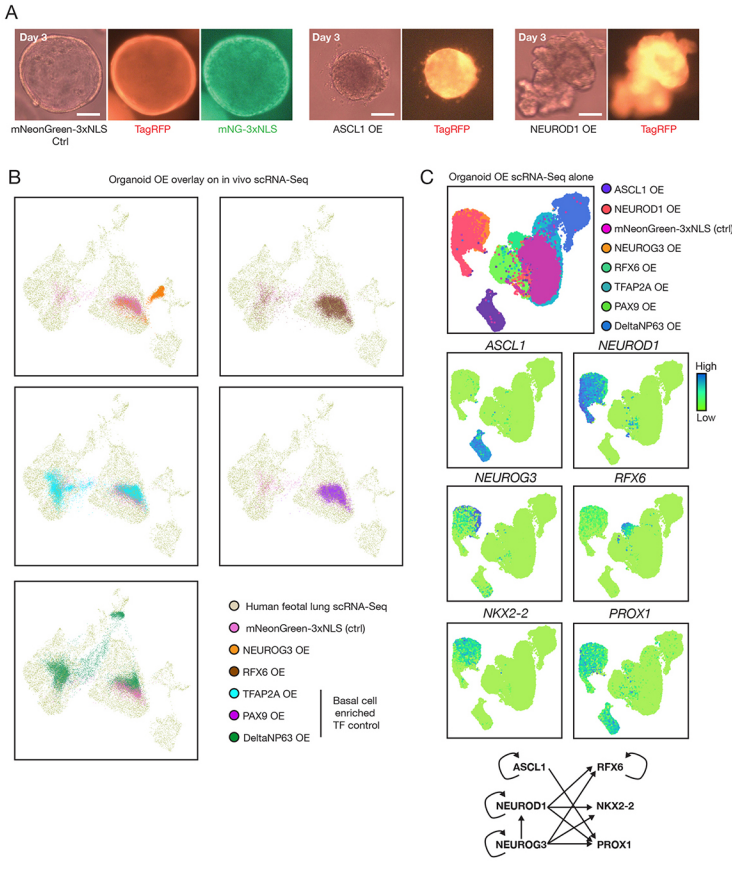

A
Mid stalk
Mid airway progenitor
Pulmonary neuroendocrine
Pulmonary NE precursor
Intermediate NE
GHRL+ neuroendocrine
B
cell type
pseudotime
TPPP3
TESC
CA2
SFTA3
FOXP2
ETV5
SOX9
FGFR2
HES1
SOX2
ASCL1
TFF3
NEUROG3
NEUROD1
GHRL
ACSL1
z-score
cell type
GHRL+ neuroendocrine
Mid airway progenitor
Mid stalk
Mid tip
Intermediate NE
Pulmonary NE precursor
Pulmonary neuroendocrine
pseudotime
C
GRP/NEUROD1/GHRL/DAPI
11 pcw
1: GRP+NEUROD1low GHRL-
2: GRPlow NEUROD1- GHRLlow
3: GRP- NEUROD1+ GHRL+
D
ASCL1/NEUROG3/NEUROD1/DAPI
11 pcw
1: ASCL1+NEUROG3-NEUROD1-
2: ASCL1+NEUROG3-NEUROD1-
3: ASCL1+NEUROG3-NEUROD1+
4: ASCL1+NEUROG3-NEUROD1+
5: ASCL1-NEUROG3-NEUROD1+
E
GRP/RFX6/GHRL/DAPI
11 pcw

Figure S14


A
Day 3
mNeonGreen-3xNLS Ctrl
TagRFP
mNG-3xNLS
Day 3
ASCL1 OE
TagRFP
Day 3
NEUROD1 OE
TagRFP
B
Organoid OE overlay on in vivo scRNA-Seq
Human feotal lung scRNA-Seq
mNeonGreen-3xNLS (ctrl)
NEUROG3 OE
RFX6 OE
TFAP2A OE
PAX9 OE
DeltaNP63 OE
Basal cell enriched TF control
C
Organoid OE scRNA-Seq alone
ASCL1 OE
NEUROD1 OE
mNeonGreen-3xNLS (ctrl)
NEUROG3 OE
RFX6 OE
TFAP2A OE
PAX9 OE
DeltaNP63 OE
ASCL1
NEUROD1
High
Low
NEUROG3
RFX6
NKX2-2
PROX1
ASCL1
RFX6
NEUROD1
NKX2-2
NEUROG3
PROX1