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], Jan Patrick Pett[2], Quitz Jeng[4], Krzysztof Polanski[2], Ziqi Dong[4], Liam Bolt[2], Laura Richardson[2], Lira Mamanova[2], Monika Dabrowska[2], Anna Wilbrey-Clark[2], Elo Madissoon[2,3], Zewen Kelvin Tuong[2,5], Emma Dann[2], Chenqu Suo[2,10], Isaac Goh Kai'En[6], Xiaoling He[7], Roger Barker[7], Sarah A Teichmann[2,8], John C. Marioni[2,3,9], Kerstin B Meyer[2,*], Emma L Rawlins[4,*]

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, Wellcome Trust/MRC Stem Cell Institute, University of Cambridge, Cambridge, CB2 1QN, UK.
5. Molecular Immunity Unit, University of Cambridge Department of Medicine, Cambridge, UK
6. Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK
7. John van Geest Centre for Brain Repair, Department of Clinical Neurosciences and Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK
8. Dept Physics/Cavendish Laboratory, University of Cambridge, JJ Thomson Ave, Cambridge CB3 0HE, UK
9. Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK
10. Department of Paediatrics, Cambridge University Hospitals, Hills Road, Cambridge CB2 0 QQ, UK

* Lead contacts: km16@sanger.ac.uk, e.rawlins@gurdon.cam.ac.uk

Highlights

- Spatiotemporal atlas of human lung development from 5-22 post conception weeks identifies 147 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 that resemble cells in adult fibrotic lungs and a new subtype of neuroendocrine cell related to human small cell lung cancer; observations which strengthen the connections between development and disease/regeneration. Our datasets are available for the community to download and interact with through our web interface (https://fetal-lung.cellgeni.sanger.ac.uk). 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 et al., 2021; Delorey et al., 2021; Deprez et al., 2020; Homer et al., 2020; Sauler et al., 2020; Travaglini et al., 2020; Vieira Braga et al., 2019). These studies are providing a foundation for the comparison of normal lung cellular physiology with disease end-states and generating hypotheses about the contributions of signalling and transcriptional changes to disease. However, the adult lung is a very static organ with low rates of cell turnover (Blenkinsopp, 1967; Rawlins and Hogan, 2008). This makes it difficult to capture transition states and progenitor cell types, even in very large adult single cell RNA sequencing (scRNAseq) studies. 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 progenitor states, predict differentiation trajectories and potential gene regulatory networks. This will provide a baseline for studying adult homeostasis and disease, particularly the contribution of developmental mechanisms to disease processes. Moreover, the transcriptional core states recently observed for fibroblasts across different organs, species and perturbations (Buechler et al., 2021) show that the insights our atlas provides will be of wide relevance to human cell biology research.

The lung buds are specified in the developing human foregut endoderm at ~5 post conception weeks (pcw) (Burri, 1984; Nikolić et al., 2018). Subsequent morphogenesis is driven by branching of the distal-most bud tips. The bud tip epithelium comprises SOX9^{+}, ID2^{+} multipotent progenitor cells which self-renew at the tip during branching (Alanis et al., 2014; Miller et al., 2018; Nikolić et al., 2017; Rawlins et al., 2009). As the bud tip epithelium branches into the surrounding mesoderm, the epithelial cells that remain in the stalk region start to differentiate, initially into bronchiolar (airway) epithelium (~5-16 pcw) and at later stages (from ~16 pcw) into the alveolar epithelium (Nikolić et al., 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. This means that coupling single cell state identification to *in vivo* spatial visualization can provide high confidence in the identification of novel progenitor cell states in the developing lung. Moreover, detailed spatial analysis of the cell states allows cell identity designations to be compared to more traditional histological definitions.

A number of cellular models are now available for studying human lung development (Hawkins et al., 2021; Kanagaki et al., 2021; Miller et al., 2018; Nikolić et al., 2017; Sucre et al., 2018; Sun et al., 2021b), including new genetic tools for functional studies in primary lung embryonic organoid cells (Sun et al., 2021a). These models are difficult to bench-mark against the primary tissue, but could potentially be exploited for functional validation of single cell predictions as well as disease modelling.

We have generated a high-resolution single cell atlas of human lung development using a combination of scRNAseq, scATACseq, Visium Spatial Transcriptomics, and mRNA *in situ* hybridisation using bespoke probes and the hybridisation chain reaction (HCR) method (Trivedi et al., 2018). Combining these data sources has allowed us to identify 147 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 cell 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. The data sets we have generated are available for interactive analysis in a dedicated web interface.

Results

A single cell atlas of human lung development comprising 147 cell states

We obtained human embryonic and fetal lungs from 5-22 pcw for scRNAseq and scATACseq analysis. To focus on epithelial differentiation and region specialization, 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, with deeper sampling in a companion study (co-submitted paper). These cell samples (except for one at 6 pcw) were split and processed for both scRNAseq and scATACseq. We used a mixture of cell dissociation methods to obtain a balanced mixture of cell types from all lung compartments (Fig. 1A) and produced high-quality profiles of transcriptome (Fig. S1A) and DNA accessibility (Fig. S1O,P) at more than 2400 genes per cell and more than 18,000 fragments per nucleus on average. The RNA profiles of cells from different dissociation protocols and compartments were then iteratively parsed out by clustering (Fig. S1C) and subclustering (Fig.

S1D) without batch correction to maintain biological features, similar to a previous study (Travaglini et al., 2020). To enrich biological features and mitigate technical ones (see methods), 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 (Cao et al., 2020; He 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-enriched samples), showed that fibroblasts were the most prominent cell type seen, particularly in the 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 via sub-clustering (see methods) was carried out and naming of clusters was based on observed and published marker genes (Sup Table 1). These analyses resulted in a final assignment of 147 cell types/states (Fig. S2A). Sample age was a strong determinant of clustering (χ^2=163727, p ≈ 0), reflecting progressive cell type maturity in these developing lungs (Fig. S2B). Clusters mostly grouped into three distinct regions which we categorized as early (5, 6 pcw), mid (9, 11 pcw) and late (15-22 pcw) stages. In addition, 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, the latter is only prominent in a small number of proximally-located cell types (Fig. S2H), suggesting that most proximal to distal regions of the complex 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 in these samples (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 facilitating cell release and/or increased fragility of older neurons. PNS cells were clustered and assigned to cell types, but their scarcity and lack of available literature precluded further analysis (Fig. S2A,F,G). Logistic regression-based comparison (see methods) showed that the gene expression of our annotated cells corresponds well to those from adult lungs (Madissoon et al., 2021), particularly at the later stages of fetal development (Fig. S3A-E).

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

The epithelial cells separate into three major groups as a function of age (Fig 2A,B), with many basal cells, MUC16$^+$ ciliated cells and secretory cells highly enriched in the proximally-dissected tissue (Fig 2B; S2H). We detected a hierarchy of epithelial progenitor and differentiating cell states

in our single cell data (Fig. 2C) and visualized these cells along the distal to proximal axis of the developing human lungs (Fig. 2G; S4). 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). In our single cell data, tip cells are separated by developmental age into early (5,6 pcw), mid (9,11 pcw) and late (15-22 pcw) populations (Fig. 2A,B; discussed in Fig. 3). 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*$^{+}$, *SCGB3A2*$^{-/LO}$) (Fig. 2A). These three cell types form a differentiation trajectory in pseudotime from mid tip - mid stalk - mid airway progenitor which then branches into the neuroendocrine, or secretory, lineages (Fig. 2D,E). Visualizing gene expression along the inferred trajectory shows that the tip and stalk cells are transcriptionally similar (Fig. 2E,F). The stalk cells lose some tip markers, including functionally-important TFs *FOXP2* and *SOX9* (Li et al., 2016), and gain expression of a relatively small number of genes including PDPN and AGER. By contrast, the newly-defined airway progenitors upregulate a more specific set of marker genes, including *CYTL1*, *CLDN4, SCGB3A2* and the Notch-responsive TF *HES4,* which are associated with airway fates (Guha et al., 2012; Kaarteenaho et al., 2010) (Fig. 2E,F). The tip, stalk, airway progenitors and secretory progenitors can be clearly visualized in a distal-proximal sequence in the tip-stalk region of the tissue at all stages tested (10-17 pcw) (Fig. 2G; S4A,B). In summary, we detect a novel airway differentiation trajectory in the developing human lungs. Cells first exit the tip into a tip-adjacent stalk-state, activate airway progenitor markers and eventually commit to secretory or neuroendocrine lineages.

Consistent with previous data (Cutz et al., 1985), the earliest differentiated epithelial cells detected in our single cell atlas were neuroendocrine (NE) cells in 5 pcw lungs (Fig. 2A-C). We identified two types of NE cells secreting different neuropeptides: 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 (with 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 the classical *GRP*$^{+}$ NE cells were always observed closer to the budding tips, suggesting that they begin to differentiate prior to the *GHRL*$^{+}$ NE cells (Fig. 2H). However, 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 in the distal regions (Fig. S4C). We cannot identify the GHRL^{+} NE cells in mouse lungs by reanalysis of published data (Hurskainen et al., 2021; Negretti et al., 2021), however *Ghrl* is expressed in mouse ciliated cells (not shown). Moreover, extensive unsuccessful attempts have previously been made to spatially identify a second NE cell population in the mouse lungs (Borromeo et al., 2016). Together, these data suggest that the *GHRL*$^{+}$ NE cells are not conserved between humans and mice.

We annotated 5 sub-types of differentiating secretory cells and one more immature secretory progenitor. (i) The secretory progenitors (*SCGB3A2*$^{+}$, *SCGB1A1*$^{-}$, *SCGB3A1*$^{-}$, *CYLT1*$^{+}$) were

detected in the single cell atlas at 9 pcw, prominent at 11 pcw, but rarer in older lungs consistent with a transient progenitor state (Fig. 2A-C). They were located more proximally than airway progenitors *in vivo* and lie between the airway progenitors and other secretory cells in pseudotime trajectory analysis (Fig. 2D-G; S4B,B'). (ii) Club cells (S*CGB3A2*$^{+}$*, SCGB1A1*$^{+}$*, SCGB3A1*$^{-}$*, SPDEF*$^{-}$*, MUC16*$^{-}$) were detected from 15 pcw in the single cell data (Fig 2A-C), or 12 pcw in the tissue (Fig. S4D). They were localised in dispersed clusters throughout the non-cartilaginous airways, but as solitary cells in the more proximal cartilaginous airways (Fig. S4D). (iii) Secretory 1 (*SCGB1A1*^{LO}*, SCGB3A2*$^{+}$*, SCGB3A1*$^{+}$) and (iv) secretory 2 (*SCGB1A1*$^{+}$*, SCGB3A2*$^{+}$*, SCGB3A1*$^{+}$) appeared from 11 pcw (Fig. 2I; S4E). Both were *SPDEF*$^{+}$ (Chen et al., 2009), but *MUC5AC*$^{-}$ and *MUC5B*$^{-}$, suggesting that they differentiate into goblet or mucous cells which are prominent in the adult proximal airways. By contrast, (v) secretory 3 (*SCGB1A1*$^{+}$*, SCGB3A2*$^{LO/-}$*, SCGB3A1*$^{+}$) was detected from 15 pcw and was *SPDEF*$^{-}$. All three luminal secretory cell populations were located predominantly in the proximal cartilaginous airways and were *MUC16*$^{+}$ (Fig. 2C; S4E,F). (vi) The SMG secretory cells (*LTF*$^{+}$*, SCGB3A1*$^{+}$*, SPDEF*$^{+}$) were detected from 15 pcw in the single cell data and are likely to be a precursor of serous or mucous-secreting SMG cells (Fig. 2C,I; S4G). These secretory cell designations are concordant with a recent single cell study of human adult airway epithelium which identified club cells (*SCGB1A1*$^{+}$), mucous and goblet cells (*SPEDEF*$^{+}$) and an SMG-localised secretory cell (*LTF*$^{+}$) (Carraro et al., 2021). Logistic-regression prediction suggests that our club, secretory 1, secretory 2 and SMG-secretory cells are the developmental counterparts of these adult cells (Fig. S3F), while secretory3 is likely to be a progenitor, or a cell type specific to prenatal or paediatric lungs. Interestingly, the developing secretory cell populations we have identified are all *SCGB3A2*$^{+}$ or *SCGB3A2*$^{+}$*, SCGB3A1*$^{+}$. These phenotypes become prominent in secretory cells in end-stage pulmonary fibrosis lungs (Carraro et al., 2020; Habermann et al., 2020), suggesting that the disease cell states recapitulate developmental processes.

The airways also include ciliated and basal cells. Ciliated cells (*FOXJ1*$^{+}$*, ALOX15*$^{+}$) were detected in the single cell data from 11 pcw and were localised interspersed with secretory cells throughout the airways (Fig. 2A-C; S4H). Rare deuterosomal cells (*FOXJ1*$^{+}$*, CDC20B*$^{+}$) appeared at the same stages (Fig. 2A-C). MUC16^{+} ciliated cells (*FOXJ1*$^{+}$, *DNAH*$^{+}$*, MUC16*^{LO}) were also detected from 11 pcw, but confined to proximal dissected regions in the single cell analysis (Fig. 2A-C). They were located in patches in the most proximal cartilaginous airways (Fig. S4I), and are likely to represent *MUC16*$^{+}$ secretory cells generating ciliated cells, as has previously been suggested in the adult (Carraro et al., 2021; Deprez et al., 2020; Vieira Braga et al., 2019). Basal cells (*TP63*$^{+}$*, F3*$^{+}$) were present in the single cell atlas from 9 pcw (Fig. 2A-C; S4J) and were more frequent in proximal regions of the airway tree where specific subpopulations were seen (Fig. 2A-C). Rarer cell types (ionocytes, brush cells) that have been consistently identified in adult airways were not present in our single cell data, suggesting they are not differentiated by 22 pcw. However, 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 hypothesize that they are an unknown progenitor cell, consistent with their transient appearance in our data and the observation that Ascl1^{+} NE cells in adult mice can generate club, ciliated and mucous cells following injury (Ouadah et al., 2019; Yao et al., 2018).

In summary, we have identified multiple epithelial progenitor states (tip; stalk; airway progenitor; secretory progenitor) and differentiating airway cells which we have also localised to a spatial differentiation gradient along the proximal-distal axis of the epithelium (summarised in Fig. S4L). The developing secretory cell populations show expression profiles that are similar to cells recently described in diseased lungs, suggesting the recapitulation of developmental states. 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*$^{+}$) 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*, *STC1* and *SFTPA* could be detected readily in the late tip cells where they were co-expressed with lower levels of core tip markers (Fig. 3A,B). This is a unique tip cell transcriptional state which has not been detected in single cell analyses of developing mouse lungs (Blackwell et al., 2021; Zepp et al., 2021). This change in tip gene expression correlated with a pseudotime differentiation trajectory from mid to late tip cells to AT2 and AT1 cell fates, suggesting that late tip cells are primarily producing alveolar-fated descendants (Fig. 3C,D; S5A). Throughout this period, similar to earlier stages, the late tip cells remain SOX9^{+} and the late stalk cells PDPN^{+} (Fig. S5B). Consistent with a switch in developmental competence of the tips from the production of airway-fated to alveolar-fated descendants, a small number of AT2 cells appear in the single cell data from 15 pcw, but are much more prominent from 22 pcw (Fig. 2A). Similarly, at 16 pcw late tip cells (*SOX9*$^{+}$*, SFTPC*$^{+}$) were clearly visualized in the tissue, but differentiating AT2 cells (*SOX9*$^{LO/-}$*, SFTPC*$^{+}$) were rare suggesting that AT2 production is just beginning (Fig. 3E,F). Over the following weeks of development, the size of the tip regions decreased and increasing numbers of differentiating AT2 cells (*SOX9*$^{LO/-}$*, SFTPC*$^{+}$) were detected (Fig. 3E,F). 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. 3G, S5C-E). Consistent with the predicted change in tip fate potential (Fig. 3C-E), when late tip cells (16-20 pcw) were grown as organoids they retained a late tip phenotype *in vitro* (co-expression of tip and AT2 markers) and much more readily differentiated to mature AT2 fate than organoids derived from earlier developmental stages (9-12 pcw) (Lim et al., 2021).

In our single cell atlas, differentiating AT1 cells were first visible at 18 pcw, but much more prominent by 22 pcw (Fig. 2A-C). Similarly in tissue sections, AT1 cells were not detected at 17 pcw (Fig. S5F). However, by 20 pcw differentiating AT1 cells (*SPOCK2*^{LO}*, SFTPC*$^{-}$) were visible

and at 21 pcw AT1 cells (*SPOCK2*$^{+}$, *SFTPC*$^{-}$) lined the developing air sacs (Fig. 3I; S5G,H). Within the tissue sections, AT1 markers were only detected in cells which had undetectable, or extremely low levels of, *SFTPC* (Fig. S5F-H). Moreover, *SFTPC*-negative cells were always observed in the stalk regions from 16 pcw onwards (Fig. 3E). These 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, 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 the trajectory analysis (Fig. 2D), supporting our hypothesis that at all stages of lung development, cells enter a stalk-state after exiting the tip and prior to acquisition of fate identity. 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 (Frank et al., 2019; Zepp et al., 2021); confirming the importance of human cellular models to study aspects of human lung development that cannot be modelled in mice.

Lung endothelial cells exhibit early specialization into arterial and venous identities

In the 5-6 pcw lungs, the endothelial cells (ECs) comprised capillary ECs (early Cap: *THY1*$^{+}$, *CD24*$^{+}$), GRIA2^{+} arterial ECs (*GRIA2*$^{+}$, *GJA5*$^{+}$) and lymphatic ECs (*PROX1*$^{+}$, *STAB1*$^{+}$, *UCP2*^{LO}) (Fig. S6A-C); showing that capillaries and lymphatic vessels are already distinct from the earliest stages of lung development and that arterial specification begins prior to venous specification. With increasing developmental age, the capillary EC lineage on the UMAP moves from early (*THY1*$^{+}$, *CD24*$^{+}$, *EGLN1*$^{+}$) to mid (*CA4*$^{+}$, *KIT*$^{+}$, *EGLN1*$^{+}$) to late (*CA4*$^{+}$, *KIT*$^{+}$) sub-types (Fig. S6A-C), similar to the age transitions seen in other compartments. In addition, trajectory analysis suggests that both mid and late Cap cells generate arterial and venous ECs (Fig. S6G,H). Recently described capillary ECs (Gillich et al., 2020; Vila Ellis et al., 2020) specialized for gas exchange and leukocyte trafficking, aerocytes (*CA4*^{LO}, *S100A3*$^{+}$), were also observed at 20 pcw and more clearly at 22 pcw (Fig. S6A-C). Spatially the aerocytes are arranged around the developing air sacs (Fig. S6D). As reported in the mouse (Vila Ellis et al., 2020), 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 the tissue at 20 pcw (Fig. S6E). Overall, three distinct arterial ECs were detected. GRIA2^{+} ECs and arterial ECs (*DKK2*$^{+}$, *SSUH2*$^{+}$) form a continuous differentiation trajectory in pseudotime (Fig. S6G,H) and the GRIA2^{+} ECs are likely to be a more immature form of arterial endothelium. The OMD^{+} ECs (*GJA5*$^{+}$, *DKK2*$^{+}$, *PTGIS*$^{+}$, *OMD*$^{+}$), cluster with the arterial ECs on the UMAP and can be seen lining the larger arterial vessels on the Visium data (Fig. S9B). The venous ECs (*PVLAP*$^{+}$, *ACKR3*$^{+}$, *HDAC9*$^{+}$) and the pulmonary venous ECs (*PVLAP*$^{+}$, *PTGS1*$^{+}$, *PTGIS*$^{+}$) separate into two distinct clusters, with the pulmonary venous ECs represented by a small number of cells detected at 20/22 pcw (Fig. S6A-C). Systemic versus pulmonary circulation EC gene expression differences have previously

been detected in the veins of adult lungs (Schupp et al., 2021); here, we show that these are becoming established by 20 pcw while the fetal circulatory shunts are still present.

The lymphatic ECs were distinct from the other ECs. Two major lymphatic ECs were detected, lymphatic ECs (*PROX1*$^{+}$, *STAB1*$^{+}$, *UCP2*^{LO}) and SCG3^{+} lymphatic ECs (*PROX1*$^{+}$, *SCG3*$^{+}$), with an intermediate EC population connecting them (Fig. S6A-C,F). The SCG3^{+} lymphatic ECs resemble a lymphatic valve population which was recently described at a transcriptional level (Takeda et al., 2019).

Haematopoietic cell types in the developing lung

Alongside the development of the vascular system, erythrocytes and leukocytes were captured. At the early stages (5-6 pcw) when arterial, capillary and lymphatic endothelial cells 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 the relative numbers of lymphoid and myeloid cells grew rapidly, dominated by macrophages, dendritic cells, NK cells, ILCs, T cells and B cells (Fig. 1C, S2A,B, S7A-K). Immature T cells are largely absent from the atlas, consistent with the idea that T cell development is restricted to the thymus. In contrast, a range of early B cell precursors and the ILC precursor were detected. We further enriched TCR and BCR fragments from our scRNA-seq libraries which supported cell-type identities and subdivision (Fig. S7). 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 fibroblast cells, leukocytes in our atlas are transcriptionally highly similar to those of other organs with minimal evidence of lung-specificity (Fig. S7L).

Developmental trajectories of mesenchymal cells

Our scRNAseq atlas reveals new types of mesenchymal cells which we have localised to specific niches. The broad fibroblast cluster comprises fibroblasts, myofibroblasts, airway and vascular smooth muscle (ASM and vSMC), pericytes, mesothelium and chondrocytes (Fig. 4A,B). There is a distinct separation of cell clusters by age (Fig. 4C), but proximal-specific cell types were not detected except for airway fibroblasts (Fig. S2H). 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 scRNAseq time course: vSMC1 (*NTRK3*$^{+}$, *NTN4*$^{+}$, *PLN*$^{-}$) and vSMC2 (*NTRK3*$^{+}$, *NTN4*$^{+}$, *PLN*$^{+}$) (Fig. 4A,B). Both were localised to blood vessels on Visium analysis (Fig. 4D) and could be seen intermingled around the same vessels on tissue sections (Fig. S8A). 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. S8B). Since intermingling of vSMC subtypes with different levels of contractility proteins has been seen in adult lungs (Frid et al., 1997), our developmental observation suggests that these represent normal functional/ontological

differences, rather than pathology. Also associated with vessels, pericytes (*NDUF4LA2*$^{+}$) were visualized as a cell layer adjacent to the endothelium around arteries and veins (Fig. S8C-E).

The most common cells isolated from the 5-15 pcw lungs were fibroblasts (Fig. 1C). At 5-6 pcw, early fibroblasts (*SFRP2*$^{+}$, *WNT2*$^{+}$) predominated, although multiple smaller fibroblast populations were also detected (Fig. 4A,B) including an SEMA3E^{+} population that is transcriptomically similar to chondrocytes and vSMC2. A SEMA3E^{+} mid-fibroblast was also detected and their high integrin alpha 8 (*ITGA8*) expression coincides with the marker of an adult lung fibroblast population in elastic fiber-rich connective tissues (Matsushima et al., 2020); however, we did not detect these cells beyond 11 pcw. 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 (Reversade et al., 2021). In the most mature lungs sequenced, there were three distinct fibroblast populations: adventitial (*SFRP2*$^{+}$, *PI16*$^{+}$), airway (*S100A4*$^{+}$) and alveolar (*WNT2*$^{+}$, *FGFR4*$^{+}$) fibroblasts with distinct spatial locations (Fig. 4A,B,E-H). In addition, an intermediate fibroblast population connected the advential and alveolar fibroblasts on the UMAP and showed graded gene expression (Fig. 4A,B), possibly representing a transitional state. Pseudotime trajectory analysis suggested a distinct differentiation hierarchy from the early and mid fibroblasts to adventitial fibroblasts; with alveolar and airway fibroblasts forming separate differentiation trajectories (Fig. 4I-K). Alternatively, the presence of 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 cell types in 15-22 pcw lungs all expressed high levels of genes associated with ECM organisation, but also had distinct gene expression patterns and spatial localisations. The adventitial fibroblasts (*SFRP2*$^{+}$, *PI16*$^{+}$) were localised around the larger blood vessels in the Visium analysis (Fig 4D). This was confirmed by HCR where they formed diffusely arranged layers of cells surrounding the tightly packed concentric rings of endothelial cells, pericytes and smooth muscle (Fig 4E, S8C). Adventitial fibroblasts were highly enriched in gene expression associated with ECM organisation and signalling, including BMP, TGFβ, WNT (Fig. 4J,K; S8F,G) consistent with described roles providing structural support to the perivascular region of blood vessels (Dahlgren and Molofsky, 2019). By contrast, the 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; S8F,G). The advential and alveolar fibroblasts shared key markers such as collagens, but also expressed unique genes (adventitial: *SERPINF1, SFRP2, PI16*; alveolar: *FGFR4, VEGFD;* Fig. 4K), with similar populations described in the adult lung (Travaglini et al., 2020). By contrast, the airway fibroblasts (*S100A4*$^{+}$, note that *S100A4* is expressed in various immune and airway epithelial cells) formed a distinct cell layer adjacent to the airway smooth muscle and were highly enriched in signalling molecules associated with morphogenesis (Fig. 4G,J,K; S8F,G). Airway fibroblasts currently do

not have a mature counterpart in any of the published adult lung single cell atlases and may be a development-specific cell with a role in a local signalling niche. They are likely analogous to the injury-induced airway fibroblasts observed in mice which also express high levels of Tnc and Fgf ligands and promote airway regeneration following injury (Moiseenko et al., 2020), again suggesting that the airways reuse developmental mechanisms during regeneration. We did not detect lipofibroblasts, which have been described in the adult lung (Travaglini et al., 2020), meaning that they are either exceptionally rare or formed later than 22 pcw.

Myofibroblasts formed three distinct groups in our single cell data. Myofibroblast 1 (*CXCL14*$^+$, *KCNK17*$^+$, *CT45A3*$^+$) appeared at 9 pcw and persisted to 20 pcw. Myofibroblast 2 (*CXCL14*$^+$, *KCNK17*$^+$, *CT45A3*$^+$, *THBD*$^+$) and myofibroblast 3 (*CXCL14*$^+$, *KCNK17*$^+$, *CT45A3*$^-$, *THBD*$^-$) were predominantly identified at 22 pcw (Fig. 4A,B). Throughout development, myofibroblasts (*CXCL14*$^+$, *KCNK17*$^+$) were visualized surrounding the developing stalk region of the epithelium, suggesting a close signalling relationship (Fig. 4H, S8H). Although not detected in significant numbers in the scRNAseq data until 22 pcw, we see myofibroblast 2 (PDGFRA$^+$, THBD$^+$, *NOTUM*$^+$) around the stalk epithelium from 15 pcw (Fig. S8I,J,L), coincident with the acquisition of AT2 markers by the late tip cells, and suggesting that it is a more mature state of myofibroblast 1. Myofibroblast 2 was enriched in the expression of genes associated with cell contractility and focal adhesions, as well as WNT signalling (Fig. S8K,L). Contractility gene expression may reflect a role in branching morphogenesis, as suggested for mouse lungs (Goodwin et al., 2019). 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,* myofibroblast 2 modulates the WNT2 signal from the alveolar fibroblasts mediating spatial patterning of epithelial AT2 and AT1 fate in normal 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. S8K,L). Myofibroblast 3 was detected in lung sections from 21 pcw and always localised to the developing air sacs (Fig. S8J), rather than the stalk epithelium. These cells are likely to be precursors of the alveolar myofibroblasts which organise the alveolar niche during morphogenesis (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 in the developing lung 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 three major fibroblast populations (Fig. 4E-G), analysed signalling within three local niches, defined as follows - 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 (Sup. Table 2) which we curated by plotting the expression of ligand-receptor pairs representing major signalling pathways (Fig. 5A,B; S9A). We observed expected interactions, including high levels of Notch ligands and receptors, and RSPO3-LGR6 and CXCL12-CXCR4 signalling in the advential niche (Fig. S9A) (Herbert and Stainier, 2011). Similarly, expected signalling interactions predicted in the alveolar niche included aerocytes to late cap cells (ALPN-ALPNR) and alveolar epithelial cells to microvascular endothelial cells, particularly AT1 to aerocyte, (VEGFA-FLT1/FLT4/KDR) (Fig. S9B) (Gillich et al., 2020; Vila Ellis et al., 2020).

Airway fibroblasts were predicted to signal via TGFβ3 and BMP4 to the developing airway epithelium, consistent with observed roles in human basal cell specification and differentiation (Miller et al., 2020; Mou et al., 2016). The airway fibroblasts and smooth muscle 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 also predicted in the alveolar niche, interactions were based on lower levels of *FGF* expression, but higher levels of canonical *WNT2* and its cognate receptor expressed on epithelial cells (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 (Lim et al., 2021; Nikolić et al., 2017). Tissue staining showed that although *FGF7* is expressed fairly ubiquitously throughout the lung, the airway fibroblasts and smooth muscle cells form a distinct barrier between the airway epithelium and the canonical *WNT2* expression (Fig. 5C-E). Based on these data, we predicted that removing canonical WNT signalling, 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).

scATACseq analysis identifies putative cell fate regulators

Single cell ATACseq provides an independent method of assessing cell type based on open chromatin regions and also allows cell type-specific TFs to be predicted using motif analysis. After tissue dissociation, the single cell suspensions were split and half of the cells processed for nuclear isolation and scATACseq (Fig. 1A). Following quality control and doublet removal, 67 scATACseq clusters comprising a total of ~100K cells were obtained and a label transfer process was used to annotate scATACseq clusters based on our scRNAseq data (Fig. 6A; Methods). Not every cell state detected by scRNAseq was distinguishable by scATACseq, consistent with previous work (Cao et al., 2020; Domcke et al., 2020). For example, separate tip, stalk and airway progenitor clusters were discerned by scRNAseq (Fig. 2A), but a combined cluster with strong similarity to all three cell types was detected by scATACseq (Fig. 6A). Similarly, the resolution of scATACseq 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 scRNAseq and ATACseq data in terms of capturing cell types, including many of the novel/lesser-known cell types we identified by scRNAseq (mid and late tip, mid and late airway progenitors, GHRL$^+$ NE, MUC16$^+$ ciliated, dueterosomal, airway fibroblasts, aerocytes, SCG3$^+$ lymphatic endothelium).

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. S10). This revealed some expected TF signatures, for example TCF21 in the alveolar, adventitial and airway fibroblasts (Quaggin et al., 1999), GHRL and FOXA1/2 in epithelial cells (Gao et al., 2013; Wan et al., 2005), and SOX17 in arterial endothelium (Corada et al., 2013). Focusing in more detail on the epithelial cells and filtering on TF expression in the corresponding cell type in the scRNAseq data (Fig. 6B), we noticed that 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 et al., 2019), and TP63 in basal cells (Rock et al., 2009). Unexpected TF signatures included KLF factors in the secretory and AT1/AT2 cells, 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 the scRNAseq and scATACseq data (Fig. 2A, 6A). ASCL1 is required for mouse NE cell differentiation (Borges et al., 1997; Ito et al., 2000) 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, the two cell types have distinct, unique regions of open chromatin. especially in the neighborhood of cell-type specific genes such as GRP and GHRL (Fig. 6C, D).

In summary, we have produced a high-resolution scATACseq data set for the developing human lungs which is highly consistent with our scRNAseq data. Mining this data set provides hypotheses for lineage-determining TFs in lung development, many of which have been reported to be dysregulated in lung disease (Apostolopoulou et al., 2007; Koczulla et al., 2012; Shen et al., 2020). 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 cells and GHRL$^+$ NE cells share the expression of many TFs and open chromatin regions, but are transcriptionally distinct. In our scRNAseq data, they were both observed along a maturation trajectory (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 and was intermediate in 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 the intermediate NEs are an additional transition

population (Fig. S11A,B). The transition states between pulmonary NE and GHRL$^+$ NE were confirmed in fetal lung samples (Fig. S11C). 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. We confirmed our cell type designations by analysis of the tissue. *ASCL1* was co-expressed with *GRP*, but rarely with *GHRL*. We also observed *ASCL1* single positive cells, likely representing the 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. S11D).

Differential expression of *ASCL1* and *NEUROD1* defines A-type and N-type human small cell lung cancer (SCLC), which are considered to derive from NE cells (Gay et al., 2021). Interestingly, these two transcription factors coincide with the scRNA-seq marker genes and scATAC-seq TF motif enrichment of our fetal NE cells (Fig. 6B,7B). Therefore, 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). Given their relevance to human disease states, we sought to use our single cell atlas to identify the 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., 2021a) 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 the airway progenitor to GHRL$^+$ NE trajectory (Fig. S11A,B) to predict 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. S11E) and *NKX2.2*, 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 factors 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 scRNAseq. Individual TFs were overexpressed from a doxycycline-inducible construct for 3 days and organoids were maintained in the self-renewing (tip cell state-promoting) medium throughout to rigorously assay the lineage-determining competence of the TF (Fig. 7H; S12A), followed by scRNA-seq.

When mapped to epithelial cells of our fetal lung atlas, the majority of the mNeonGreen-3xNLS expressing organoid cells were projected towards mid tip progenitors or mid stalk cells as

expected, further validating the organoid model (Fig. 7I). Whereas overexpression of DeltaNTP63 resulted in basal cell-like lineages (S12B) consistent with a previous report (Warner et al., 2013), indicating that this simple *in vitro* assay could capture TF function accurately. Overexpression of *RFX6*, *TFAP2A* or *PAX9* did not result in the predicted lineage progression at a transcriptome level (S12B); possibly due to the weak effects of these TFs, or the low sensitivity of scRNAseq. However, *ASCL1*-overexpression organoids progressed into pulmonary NE precursors (Fig. 7I). By contrast, *NEUROD1* overexpression promoted tip progenitor cell differentiation into *GHRL*$^+$ NE precursors (Fig. 7I). Intriguingly, *NEUROG3* overexpression also led to *GHRL*$^+$ NE precursor formation (Fig. S12B). This indicated 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. S12C). By contrast, *NKX2-2* and *PROX1* were upregulated by other TFs, indicating they are relatively low in the hierarchy (Fig. S12C). *NKX2-2* and *PROX1* expression in the organoid assay matched their expression patterns in NE cells *in vivo* (Fig. S12C), showing that this assay recapitulated key features of the TF network. These experiments have allowed us to test gene regulatory network (GRN) predictions from the single cell atlas, confirm the predicted lineage trajectory and provide a foundation for studying human SCLC. This is particularly 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 147 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 states in the developing airway hierarchy with unprecedented resolution, including a new neuroendocrine cell subtype related to SCLC. We show 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 the 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.

Our data show that lung maturation occurs in concert across cell compartments, for example with epithelial AT1 cells and endothelial aerocytes differentiating in parallel at 20-22 pcw. Moreover, we have observed many aspects of cellular differentiation in utero for the first time showing that

they are controlled by prenatal factors, rather than the transition to air breathing with its associated mechanical and hormonal changes. For example, pulmonary venous endothelial cells, two distinct types of vSMCs and two types of lymphatic endothelial cells are established prior to major alterations in blood flow that occur postnatally.

We characterize in detail the proximal-distal differentiation trajectory in the airway epithelium: descendants of tip cells enter the stalk region and lose tip markers, before acquiring airway progenitor fate and entering into neuroendocrine or secretory lineages (Fig. 2). We define a new late tip cell state in which tip epithelial cells upregulate markers of AT2 cells, but retain canonical tip markers, from ~15 pcw (Fig. 3). This late tip cell has not been observed in mice and our further analyses have shown that it is a functionally distinct tip cell state with high propensity to produce alveolar epithelium (Lim et al., 2021). These data show that human distal tip cells are sequentially potent, rather than multipotent, with distinct tip cell states generating airway and alveolar descendants. In addition, we show that the spatial patterning of human alveolar epithelial lineages is analogous to the process for airway cells: cells exit the late tip and turn-off both tip and AT2 markers, prior to initiating differentiation into either AT1 or AT2 cell states (Fig. 3I). This is different to reports in the developing mouse lung of early commitment (even during the airway branching phase) to AT2 and AT1 lineages (Frank et al., 2019; Zepp et al., 2021). We speculate that the difference may originate from the compressed timing of mouse embryonic development.

The mesenchymal compartment contains multiple fibroblast and myofibroblast cell states and we focus on those present during the later stages of lung development. We show that airway, adventitial and alveolar fibroblasts are all localized 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 characterize a population of myofibroblasts which contacts the developing epithelial stalk region and expresses high levels of the secreted Wnt-inhibitor, *NOTUM*; whereas alveolar fibroblasts express high levels of the canonical *WNT2* ligand (Fig. 4). In a separate study, using surface markers identified from 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 signaling interaction between alveolar fibroblasts, myofibroblast 2 cells and late tip cells can control human AT2 spatial patterning.

Recent 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 (Jadhav et al., 2017; Larsen and Jensen, 2021). Our atlas confirms transcriptional similarities between human lung development and disease: developmental secretory progenitors express high levels of secretoglobins previously reported only in secretory cells from end-stage fibrosis lungs

(Carraro et al., 2020; Habermann et al., 2020); airway fibroblasts are highly reminiscent of a cell state observed in mouse lung regeneration (Moiseenko et al., 2020); GHRL$^+$ NE cells are transcriptionally similar to a NEUROD1$^+$ subtype of SCLC (Fig. 7). GHRL$^+$ NE cells have not yet been identified in adult lungs and our data may indicate either that they are a cancer cell of origin, or that NEUROD1$^+$ SCLCs reuse developmental transcriptional networks. Our functional analyses of NE cell differentiation in organoids will provide tools to test these hypotheses. Detailed ATACseq datasets are not yet available for human lung disease. Our ATACseq atlas has three times the median number of fragments per cell and more than seven times the number of lung cell types than published multi-organ atlases. It will provide a baseline for analyses when adult lung atlases are published. In summary, our multi-component atlas is a community resource for future analyses of human development, regeneration and disease.

Acknowledgements

We would like to acknowledge the Gurdon Institute Imaging Facility, and the Cellular Generation 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 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).

REFERENCES

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086.

Alanis, D.M., Chang, D.R., Akiyama, H., Krasnow, M.A., and Chen, J. (2014). Two nested developmental waves demarcate a compartment boundary in the mouse lung. Nat. Commun. *5*, 1–15.

Apostolopoulou, K., Pateras, I.S., Evangelou, K., Tsantoulis, P.K., Liontos, M., Kittas, C., Tiniakos, D.G., Kotsinas, A., Cordon-Cardo, C., and Gorgoulis, V.G. (2007). Gene amplification is a relatively frequent event leading to ZBTB7A (Pokemon) overexpression in non-small cell lung cancer. J. Pathol. *213*, 294–302.

Blackwell, T.S., Banovich, N.E., Kropski, J.A., and Sucre, J.M.S. (2021). A Single Cell Atlas of Lung Development. bioRxiv.

Blenkinsopp, W.K. (1967). Proliferation of respiratory tract epithelium in the rat. Exp. Cell Res. *46*, 144–154.

Borges, M., Linnoila, R.I., van de Velde, H.J., Chen, H., Nelkin, B.D., Mabry, M., Baylin, S.B., and Ball, D.W. (1997). An achaete-scute homologue essential for neuroendocrine differentiation in the lung. Nature *386*, 852–855.

Borromeo, M.D., Savage, T.K., Kollipara, R.K., He, M., Augustyn, A., Osborne, J.K., Girard, L., Minna, J.D., Gazdar, A.F., Cobb, M.H., et al. (2016). ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. Cell Rep. *16*, 1259–1272.

Buechler, M.B., Pradhan, R.N., Krishnamurty, A.T., Cox, C., Calviello, A.K., Wang, A.W., Yang, Y.A., Tam, L., Caothien, R., Roose-Girma, M., et al. (2021). Cross-tissue organization of the fibroblast lineage. Nature *593*, 575–579.

Burri, P.H. (1984). Fetal and postnatal development of the lung. Annu. Rev. Physiol. *46*, 617–628.

Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A., Aldinger, K.A., Blecher-Gonen, R., Zhang, F., et al. (2020). A human cell atlas of fetal gene expression. Science *370*.

Carraro, G., Mulay, A., Yao, C., Mizuno, T., Konda, B., Petrov, M., Lafkas, D., Arron, J.R., Hogaboam, C.M., Chen, P., et al. (2020). Single-Cell Reconstruction of Human Basal Cell Diversity in Normal and Idiopathic Pulmonary Fibrosis Lungs. Am. J. Respir. Crit. Care Med. *202*, 1540–1550.

Carraro, G., Langerman, J., Sabri, S., Lorenzana, Z., Purkayastha, A., Zhang, G., Konda, B., Aros, C.J., Calvert, B.A., Szymaniak, A., et al. (2021). Transcriptional analysis of cystic fibrosis

airways at single-cell resolution reveals altered epithelial cell states and composition. Nat. Med. *27*, 806–814.

Chen, G., Korfhagen, T.R., Xu, Y., Kitzmiller, J., Wert, S.E., Maeda, Y., Gregorieff, A., Clevers, H., and Whitsett, J.A. (2009). SPDEF is required for mouse pulmonary goblet cell differentiation and regulates a network of genes associated with mucus production. J. Clin. Invest. *119*, 2914–2924.

Corada, M., Orsenigo, F., Morini, M.F., Pitulescu, M.E., Bhat, G., Nyqvist, D., Breviario, F., Conti, V., Briot, A., Iruela-Arispe, M.L., et al. (2013). Sox17 is indispensable for acquisition and maintenance of arterial identity. Nat. Commun. *4*, 2609.

Cutz, E., Gillan, J.E., and Bryan, A.C. (1985). Neuroendocrine cells in the developing human lung: morphologic and functional considerations. Pediatr. Pulmonol. *1*, S21–S29.

Dahlgren, M.W., and Molofsky, A.B. (2019). Adventitial Cuffs: Regional Hubs for Tissue Immunity. Trends Immunol. *40*, 877–887.

Delorey, T.M., Ziegler, C.G.K., Heimberg, G., Normand, R., Yang, Y., Segerstolpe, Å., Abbondanza, D., Fleming, S.J., Subramanian, A., Montoro, D.T., et al. (2021). COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. Nature *595*, 107–113.

Deprez, M., Zaragosi, L.-E., Truchi, M., Becavin, C., Ruiz García, S., Arguel, M.-J., Plaisant, M., Magnone, V., Lebrigand, K., Abelanet, S., et al. (2020). A Single-Cell Atlas of the Human Healthy Airways. Am. J. Respir. Crit. Care Med. *202*, 1636–1645.

Domcke, S., Hill, A.J., Daza, R.M., Cao, J., O'Day, D.R., Pliner, H.A., Aldinger, K.A., Pokholok, D., Zhang, F., Milbank, J.H., et al. (2020). A human cell atlas of fetal chromatin accessibility. Science.

Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat. Protoc. *15*, 1484–1506.

Frank, D.B., Penkala, I.J., Zepp, J.A., Sivakumar, A., Linares-Saldana, R., Zacharias, W.J., Stolz, K.G., Pankin, J., Lu, M., Wang, Q., et al. (2019). Early lineage specification defines alveolar epithelial ontogeny in the murine lung. Proc. Natl. Acad. Sci. U. S. A. *116*, 4362–4371.

Frid, M.G., Dempsey, E.C., Durmowicz, A.G., and Stenmark, K.R. (1997). Smooth muscle cell heterogeneity in pulmonary and systemic vessels. Importance in vascular disease. Arterioscler. Thromb. Vasc. Biol. *17*, 1203–1209.

Gao, X., Vockley, C.M., Pauli, F., Newberry, K.M., Xue, Y., Randell, S.H., Reddy, T.E., and Hogan, B.L.M. (2013). Evidence for multiple roles for grainyhead-like 2 in the establishment and maintenance of human mucociliary airway epithelium.[corrected]. Proc. Natl. Acad. Sci. U. S. A. *110*, 9356–9361.

Gay, C.M., Stewart, C.A., Park, E.M., Diao, L., Groves, S.M., Heeke, S., Nabet, B.Y., Fujimoto,

J., Solis, L.M., Lu, W., et al. (2021). Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. Cancer Cell *39*, 346–360.e7.

Gillich, A., Zhang, F., Farmer, C.G., Travaglini, K.J., Tan, S.Y., Gu, M., Zhou, B., Feinstein, J.A., Krasnow, M.A., and Metzger, R.J. (2020). Capillary cell-type specialization in the alveolus. Nature *586*, 785–789.

Goodwin, K., Mao, S., Guyomar, T., Miller, E., Radisky, D.C., Košmrlj, A., and Nelson, C.M. (2019). Smooth muscle differentiation shapes domain branches during mouse lung development. Development *146*.

Guha, A., Vasconcelos, M., Cai, Y., Yoneda, M., Hinds, A., Qian, J., Li, G., Dickel, L., Johnson, J.E., Kimura, S., et al. (2012). Neuroepithelial body microenvironment is a niche for a distinct subset of Clara-like precursors in the developing airways. Proc. Natl. Acad. Sci. U. S. A. *109*, 12592–12597.

Habermann, A.C., Gutierrez, A.J., Bui, L.T., Yahn, S.L., Winters, N.I., Calvi, C.L., Peter, L., Chung, M.-I., Taylor, C.J., Jetter, C., et al. (2020). Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. Sci Adv *6*, eaba1972.

Hawkins, F.J., Suzuki, S., Beermann, M.L., Barillà, C., Wang, R., Villacorta-Martin, C., Berical, A., Jean, J.C., Le Suer, J., Matte, T., et al. (2021). Derivation of Airway Basal Stem Cells from Human Pluripotent Stem Cells. Cell Stem Cell *28*, 79–95.e8.

He, P., Williams, B.A., Trout, D., Marinov, G.K., Amrhein, H., Berghella, L., Goh, S.-T., Plajzer-Frick, I., Afzal, V., Pennacchio, L.A., et al. (2020). The changing mouse embryo transcriptome at whole tissue and single-cell resolution. Nature *583*, 760–767.

Herbert, S.P., and Stainier, D.Y.R. (2011). Molecular control of endothelial cell behaviour during blood vessel morphogenesis. Nat. Rev. Mol. Cell Biol. *12*, 551–564.

Homer, R., Yan, X., Rosas, I.O., and Kaminski, N. (2020). Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. Science.

Hurskainen, M., Mižíková, I., Cook, D.P., Andersson, N., Cyr-Depauw, C., Lesage, F., Helle, E., Renesme, L., Jankov, R.P., Heikinheimo, M., et al. (2021). Single cell transcriptomic analysis of murine lung development on hyperoxia-induced damage. Nat. Commun. *12*, 1–19.

Ireland, A.S., Micinski, A.M., Kastner, D.W., Guo, B., Wait, S.J., Spainhower, K.B., Conley, C.C., Chen, O.S., Guthrie, M.R., Soltero, D., et al. (2020). MYC Drives Temporal Evolution of Small Cell Lung Cancer Subtypes by Reprogramming Neuroendocrine Fate. Cancer Cell *38*, 60–78.e12.

Ito, T., Udaka, N., Yazawa, T., Okudela, K., Hayashi, H., Sudo, T., Guillemot, F., Kageyama, R., and Kitamura, H. (2000). Basic helix-loop-helix transcription factors regulate the neuroendocrine differentiation of fetal mouse pulmonary epithelium. Development *127*, 3913–3921.

Jadhav, U., Saxena, M., O'Neill, N.K., Saadatpour, A., Yuan, G.-C., Herbert, Z., Murata, K., and Shivdasani, R.A. (2017). Dynamic Reorganization of Chromatin Accessibility Signatures during Dedifferentiation of Secretory Precursors into Lgr5+ Intestinal Stem Cells. Cell Stem Cell *21*, 65–77.e5.

Kaarteenaho, R., Merikallio, H., Lehtonen, S., Harju, T., and Soini, Y. (2010). Divergent expression of claudin -1, -3, -4, -5 and -7 in developing human lung. Respir. Res. *11*, 59.

Kanagaki, S., Ikeo, S., Suezawa, T., Yamamoto, Y., Seki, M., Hirai, T., Hagiwara, M., Suzuki, Y., and Gotoh, S. (2021). Directed induction of alveolar type I cells derived from pluripotent stem cells via Wnt signaling inhibition. Stem Cells *39*, 156–169.

Kimura, S., Ostrin, E.J., and Chen, J. (2019). Transcriptional control of lung alveolar type 1 cell development and maintenance by NK homeobox 2-1. Proceedings of the.

Koczulla, A.-R., Jonigk, D., Wolf, T., Herr, C., Noeske, S., Klepetko, W., Vogelmeier, C., von Neuhoff, N., Rische, J., Wrenger, S., et al. (2012). Krüppel-like zinc finger proteins in end-stage COPD lungs with and without severe alpha1-antitrypsin deficiency. Orphanet J. Rare Dis. *7*, 29.

Kumar, M.E., Bogard, P.E., Espinoza, F.H., Menke, D.B., Kingsley, D.M., and Krasnow, M.A. (2014). Mesenchymal cells. Defining a mesenchymal progenitor niche at single-cell resolution. Science *346*, 1258810.

Larsen, H.L., and Jensen, K.B. (2021). Reprogramming cellular identity during intestinal regeneration. Curr. Opin. Genet. Dev. *70*, 40–47.

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

Li, S., Morley, M., Lu, M., Zhou, S., Stewart, K., French, C.A., Tucker, H.O., Fisher, S.E., and Morrisey, E.E. (2016). Foxp transcription factors suppress a non-pulmonary gene expression program to permit proper lung development. Dev. Biol. *416*, 338–346.

Lim, K., Tang, W., Sun, D., Peng, H., Teichmann, S., Marioni, J.C., Meyer, K.B., and Rawlins, E.L. (2021). Acquisition of alveolar fate and differentiation competence by human fetal lung epithelial progenitor cells. bioRxiv.

Madissoon, E., Oliver, A.J., Kleshchevnikov, V., Wilbrey-Clark, A., Polanski, K., Orsi, A.R., Mamanova, L., Bolt, L., Richoz, N., Elmentaite, R., et al. (2021). A spatial multi-omics atlas of the human lung reveals a novel immune cell survival niche.

Matsushima, S., Aoshima, Y., Akamatsu, T., Enomoto, Y., Meguro, S., Kosugi, I., Kawasaki, H., Fujisawa, T., Enomoto, N., Nakamura, Y., et al. (2020). CD248 and integrin alpha-8 are candidate markers for differentiating lung fibroblast subtypes. BMC Pulm. Med. *20*, 21.

Mellitzer, G., Beucher, A., Lobstein, V., Michel, P., Robine, S., Kedinger, M., and Gradwohl, G. (2010). Loss of enteroendocrine cells in mice alters lipid absorption and glucose homeostasis and impairs postnatal survival. J. Clin. Invest. *120*, 1708–1721.

Miller, A.J., Hill, D.R., Nagy, M.S., Aoki, Y., Dye, B.R., Chin, A.M., Huang, S., Zhu, F., White, E.S., Lama, V., 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*, 101–119.

Miller, A.J., Yu, Q., Czerwinski, M., Tsai, Y.-H., Conway, R.F., Wu, A., Holloway, E.M., Walker, T., Glass, I.A., Treutlein, B., et al. (2020). In Vitro and In Vivo Development of the Human Airway at Single-Cell Resolution. Dev. Cell *53*, 117–128.e6.

Moiseenko, A., Vazquez-Armendariz, A.I., Kheirollahi, V., Chu, X., Tata, A., Rivetti, S., Günther, S., Lebrigand, K., Herold, S., Braun, T., et al. (2020). Identification of a Repair-Supportive Mesenchymal Cell Population during Airway Epithelial Regeneration. Cell Rep. *33*, 108549.

Mou, H., Vinarsky, V., Tata, P.R., Brazauskas, K., Choi, S.H., Crooke, A.K., Zhang, B., Solomon, G.M., Turner, B., Bihler, H., et al. (2016). Dual SMAD Signaling Inhibition Enables Long-Term Expansion of Diverse Epithelial Basal Cells. Cell Stem Cell *19*, 217–231.

Naya, F.J., Huang, H.P., Qiu, Y., Mutoh, H., DeMayo, F.J., Leiter, A.B., and Tsai, M.J. (1997). Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. Genes Dev. *11*, 2323–2334.

Negretti, N.M., Plosa, E.J., Benjamin, J.T., Schuler, B.A., Christian Habermann, A., Jetter, C., Gulleman, P., Taylor, C.J., Nichols, D., Matlock, B.K., et al. (2021). A Single Cell Atlas of Lung Development.

Nikolić, M.Z., Caritg, O., Jeng, Q., Johnson, J.-A., Sun, D., Howell, K.J., Brady, J.L., Laresgoiti, U., Allen, G., Butler, R., 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*.

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

Ouadah, Y., Rojas, E.R., Riordan, D.P., Capostagno, S., Kuo, C.S., and Krasnow, M.A. (2019). Rare Pulmonary Neuroendocrine Cells Are Stem Cells Regulated by Rb, p53, and Notch. Cell *179*, 403–416.e23.

Quaggin, S.E., Schwartz, L., Cui, S., Igarashi, P., Deimling, J., Post, M., and Rossant, J. (1999). The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. Development *126*, 5771–5783.

Rawlins, E.L., and Hogan, B.L.M. (2008). Ciliated epithelial cell lifespan in the mouse trachea and lung. Am. J. Physiol. Lung Cell. Mol. Physiol. *295*, L231–L234.

Rawlins, E.L., Ostrowski, L.E., Randell, S.H., and Hogan, B.L.M. (2007). Lung development and repair: contribution of the ciliated lineage. Proc. Natl. Acad. Sci. U. S. A. *104*, 410–417.

Rawlins, E.L., Clark, C.P., Xue, Y., and Hogan, B.L.M. (2009). The Id2+ distal tip lung epithelium contains individual multipotent embryonic progenitor cells. Development *136*, 3741–

3745.

Reversade, B., Yan, K.S., Kuo, C.J., and Spence, J.R. (2021). R-SPONDIN2+ Mesenchymal Cells Form the Bud Tip Progenitor Niche During Human Lung Development. bioRxiv.

Rock, J.R., Onaitis, M.W., Rawlins, E.L., Lu, Y., Clark, C.P., Xue, Y., Randell, S.H., and Hogan, B.L.M. (2009). Basal cells as stem cells of the mouse trachea and human airway epithelium. Proc. Natl. Acad. Sci. U. S. A. *106*, 12771–12775.

Sauler, M., McDonough, J.E., Adams, T.S., Kothapalli, N., Schupp, J.C., Nouws, J., Chioccioli, M., Omote, N., Cosme, C., Jr, Poli, S., et al. (2020). Single-cell RNA sequencing identifies aberrant transcriptional profiles of cellular populations and altered alveolar niche signalling networks in Chronic Obstructive Pulmonary Disease (COPD) (medRxiv).

Schupp, J.C., Adams, T.S., Cosme, C., Jr, Raredon, M.S.B., Yuan, Y., Omote, N., Poli, S., Chioccioli, M., Rose, K.-A., Manning, E.P., et al. (2021). Integrated Single-Cell Atlas of Endothelial Cells of the Human Lung. Circulation *144*, 286–302.

Shen, W., Weng, Z., Fan, M., Wang, S., Wang, R., Zhang, Y., Tian, H., Wang, X., Wu, X., Yang, X., et al. (2020). Mechanisms by Which the MBD2/miR-301a-5p/CXCL12/CXCR4 Pathway Regulates Acute Exacerbations of Chronic Obstructive Pulmonary Disease. Int. J. Chron. Obstruct. Pulmon. Dis. *15*, 2561–2572.

van Soldt, B.J., Qian, J., Li, J., Tang, N., Lu, J., and Cardoso, W.V. (2019). Yap and its subcellular localization have distinct compartment-specific roles in the developing lung. Development *146*.

Sucre, J.M.S., Jetter, C.S., Loomans, H., Williams, J., Plosa, E.J., Benjamin, J.T., Young, L.R., Kropski, J.A., Calvi, C.L., Kook, S., et al. (2018). Successful Establishment of Primary Type II Alveolar Epithelium with 3D Organotypic Coculture. Am. J. Respir. Cell Mol. Biol. *59*, 158–166.

Sun, D., Evans, L., Perrone, F., Sokleva, V., Lim, K., Rezakhani, S., Lutolf, M., Zilbauer, M., and Rawlins, E.L. (2021a). A functional genetic toolbox for human tissue-derived organoids. Elife *10*.

Sun, Y.L., Hurley, K., Villacorta-Martin, C., Huang, J., Hinds, A., Gopalan, K., Caballero, I.S., Russo, S.J., Kitzmiller, J.A., Whitsett, J.A., et al. (2021b). Heterogeneity in Human iPSC-derived Alveolar Epithelial Type II Cells Revealed with ABCA3/SFTPC Reporters. Am. J. Respir. Cell Mol. Biol.

Takeda, A., Hollmén, M., Dermadi, D., Pan, J., Brulois, K.F., Kaukonen, R., Lönnberg, T., Boström, P., Koskivuo, I., Irjala, H., et al. (2019). Single-Cell Survey of Human Lymphatics Unveils Marked Endothelial Cell Heterogeneity and Mechanisms of Homing for Neutrophils. Immunity *51*, 561–572.e5.

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

RNA sequencing. Nature *587*, 619–625.

Trivedi, V., Choi, H.M.T., Fraser, S.E., and Pierce, N.A. (2018). Multidimensional quantitative analysis of mRNA expression within intact vertebrate embryos. Development *145*.

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

Vila Ellis, L., Cain, M.P., Hutchison, V., Flodby, P., Crandall, E.D., Borok, Z., Zhou, B., Ostrin, E.J., Wythe, J.D., and Chen, J. (2020). Epithelial Vegfa Specifies a Distinct Endothelial Population in the Mouse Lung. Dev. Cell *52*, 617–630.e6.

Wan, H., Dingle, S., Xu, Y., Besnard, V., Kaestner, K.H., Ang, S.-L., Wert, S., Stahlman, M.T., and Whitsett, J.A. (2005). Compensatory Roles of Foxa1 and Foxa2 during Lung Morphogenesis *. J. Biol. Chem. *280*, 13809–13816.

Warner, S.M.B., Hackett, T.-L., Shaheen, F., Hallstrand, T.S., Kicic, A., Stick, S.M., and Knight, D.A. (2013). Transcription factor p63 regulates key genes and wound repair in human airway epithelial basal cells. Am. J. Respir. Cell Mol. Biol. *49*, 978–988.

Yao, E., Lin, C., Wu, Q., Zhang, K., Song, H., and Chuang, P.-T. (2018). Notch Signaling Controls Transdifferentiation of Pulmonary Neuroendocrine Cells in Response to Lung Injury. Stem Cells *36*, 377–391.

Zepp, J.A., Morley, M.P., Loebel, C., Kremp, M.M., Chaudhry, F.N., Basil, M.C., Leach, J.P., Liberti, D.C., Niethamer, T.K., Ying, Y., et al. (2021). Genomic, epigenomic, and biophysical cues controlling the emergence of the lung alveolus. Science *371*.

FIGURE LEGENDS

Figure 1. Experimental overview.

(A) Overview of sample collection for scRNAseq (circles) and scATACseq (squares) experiments from whole lung (purple), distal (red) and proximal (blue) lung section, followed by cell processing and clustering, with cluster number referring to those used in the data portal (https://fetal-lung.cellgeni.sanger.ac.uk).
(B) UMAP representation of approximately 80,000 good quality cells obtained, indicating epithelial, endothelial, fibroblasts and leukocyte/erythroid compartments.
(C) Cell type proportions of the whole lung over developmental time.

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

(A, B) UMAP visualization of epithelial cells, coloured by cell types (A), stage (B, *left*), and dissection (B, *right*).
(C) Dot plot describing differential marker gene expression level for all identified cell types.
(D) UMAP visualising the predicted epithelial cell lineage trajectory using scvelo, developmental age is plotted in the inset.
(E, F) UMAP (E) and associated gene expression heatmap (F) delineating a lineage trajectory from Mid tip to Secretory cells using Monocle 3.
(G, H) *In situ* HCR analysis of distal fetal lungs at 11 (H) and 12 (G) pcw. (G) Airway progenitors and secretory progenitors are marked by *SOX9*$^{-}$/*CYTL1*$^{low/+}$/*SCGB3A2*$^{low/-}$ and *SOX9*$^{-}$/*CYTL1*$^{high/+}$/*SCGB3A2*$^{high/+}$, respectively. *SOX9* (tip epithelial cell, white), *CYTL1* (red), and *SCGB3A2* (green). (H) Two subtypes of neuroendocrine cells are marked as *GHRL*$^{+}$ (*GHRL*$^{+}$ neuroendocrine, red) or *GRP*+ (pulmonary neuroendocrine, green).
(I) *In situ* HCR analysis of proximal cartilaginous airway regions of fetal lungs at 19 pcw. Four subtypes of secretory cells were observed: Secretory 1, *SCGB3A2*$^{+}$/*SCGB3A1*$^{+}$/ *SCGB1A1*$^{low/-}$ (*arrowhead*); Secretory 2, *SCGB3A2*$^{+}$/*SCGB3A1*$^{+}$/*SCGB1A1*$^{+}$ (*asterisk**); Secretory 3, *SCGB3A2*$^{low/-}$/*SCGB3A1*$^{+}$/*SCGB1A1*$^{+}$ (*arrow*); Secretory 4, submucosal gland cells (SMGs), *SCGB3A1*$^{+}$.
DAPI, nuclei. Scale bars, 50 µm.

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

(A, B) *In situ* HCR analysis of distal lung regions at 10 (A, B), 15 (B), and 20 (A) pcw. (A) The early tip progenitor marker, *TPPP3* (red) at the pseudoglandular stage, 10 pcw, and at the canalicular stage, 20 pcw. *SOX9* (white, upper panel) at 10 pcw and *SFTPC* (*green*, lower panel) at 20 pcw indicate tip regions (arrows). (B) A late tip marker, *SFTPA1*, was rarely observed in 10

pcw fetal lungs (red, upper panel), clearly appeared in the tip regions of 15 pcw fetal lungs, with *STC1* (white), a Mid/Late tip progenitor marker. *SFTPC* (green).
(C) UMAP visualization of Mid/Late tip, AT1 and AT2 cells, coloured by stage.
(D) Lineage trajectory from Mid tip, Late tip, AT2, to AT1 (right) derived using Monocle 3.
(E) Heatmap showing gene expression along the lineage trajectory in D.
(F) *In situ* HCR analysis of distal fetal lung regions at 16, 19, and 21 pcw, *SFTPC* (green) and SOX9 (white). Arrows, *SFTPC*$^{+}$/*SOX9*$^{+}$ tip cells; arrowheads, differentiating AT2 cells in stalk/air sac regions, *SFTPC*$^{+}$/*SOX9*$^{low/-}$; asterisk (*), air sacs.
(G) Quantification of AT2 cells, *SFTPC*$^{+}$/*SOX9*$^{low/-}$, in stalk/air sac regions at 14, 15, 17, 19, and 21 pcw fetal lungs. 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 of 21 pcw fetal lung, visualizing the localization of AT2 (H) and AT1 (I) cells, each marked by *SFTPC*$^{+}$ and *NAPSA*$^{+}$ (*arrowheads;* H, I) and *SFTPC*$^{-}$/*MMP28*$^{+}$/*SPOCK2*$^{+}$ (*arrow*s; I).
(J) Diagram describing the differentiation process of late epithelial tip cells to acquire an alveolar progenitor identity, followed by alveolar epithelial differentiation to AT2 and AT1 cell lineages.
DAPI, nuclei. Scale bars, 50 µm.

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

(A) UMAP visualization of mesenchymal cells, coloured by cell types.
(B) Dot plot describing differential marker gene expression within a cluster by cell type.
(C) UMAP visualization of mesenchymal cells, coloured by stages.
(D) Visium spatial feature plots visualizing adventitial fibroblasts, vascular SMC 1 and 2 on a 17 pcw lung section.
(E-H) *In situ* HCR assay (E, F, H) and immunostaining (G) for fetal lung tissues at 19 (E), 18 (F), 15 (G), and 17 (H) pcw. (E) Adventitial fibroblasts (*SFRP2*$^{+}$,white/*PI16*$^{+}$,red; arrowheads). Endothelial cell (*PECAM1*, green). (F) Alveolar fibroblasts (*WNT2*$^{+}$, white/*FGFR4*$^{+}$,red), tip cells (*SFTPC*, green). Asterisks (*; myofibroblasts). (G) Airway fibroblasts (S100A4^{+},red; dashed lines), tip cells (CD44, green), smooth muscle (ACTA2, white, arrowheads). (H) Myofibroblasts (*KCNK17*$^{+}$,white/*CXCL14*$^{+}$,red; arrowheads), tip cells (*SFTPC*, green). DAPI, nuclei. Scale bars, 50 µm.
(I) UMAP visualization of fibroblast differentiation trajectories, coloured by cell types and stage.
(J, K) UMAPs (J) and the relevant gene expression heatmaps (K) displaying potential lineage trajectories from Mid tip to adventitial fibroblasts (top), alveolar fibroblasts (middle), or airway fibroblasts (bottom) derived using Monocle 3.

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

(A, B) Overview of predicted ligand–receptor interactions from CellPhoneDB in alveolar (A) and airway (B) niches. Dot plots visualize gene expression by cell type and dashed arrows indicate a predicted direction of signaling from ligands to receptors.

(C-E) Immunofluorescence of fetal lung tissues at 15 (C) and 19 (D, E) pcw. S100A4/*S100A4*, airway fibroblasts; ACTA2, airway SMCs; CD44, tip epithelium in alveolar region; PECAM1, endothelial cells. S100A4^{+} airway fibroblasts form a boundary (dashed lines) between alveolar and airway regions. Lines indicate a boundary between airway fibroblasts/airway SMCs and airway epithelial cells. DAPI, nuclei. Scale bars, 50 μm.

(F) Experimental outline for airway differentiation of lung tip organoids. Organoids were cultured in FGF7/10-containing medium, in the presence (self-renewal medium; SNM) or absence (differentiation medium; DM) of CHIR99021, a Wnt agonist, for 30 days.

(G) Changes in relative mRNA levels of maker genes of the tip progenitors (*SOX9*, *SOX2*), basal cells (*TP63*, *KRT5*), secretory cells (*SCGB3A2*, *SCGB1A1*), and ciliated cells (*FOXJ1*) by qRT-PCR. Data was normalized 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) for 30 days. DAPI, nuclei. Scale bar, 25 μm.

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

(A) Single-cell DNA accessibility profiles mapped onto 2D UMAP plane. Colors represent cell types/states.

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

(C-D) Read coverage tracks of in silico aggregated "pseudo-bulk" epithelial clusters over the GRP locus (C) and GHRL locus (D).

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

(A) Zoom-in UMAP plot of NE lineages: *GHRL*$^{+}$ NE, Pulmonary NE, precursor cells, and a transition population (intermediate NE).

(B) Dot plot showing selected gene expression in *GHRL*$^{+}$ NE, Pulmonary NE and *NEUROD1*$^{+}$ pulmonary NE. Intermediate NE presented a transition signature and some cells expressed *NEUROG3*.

(C) A schematic model of NE lineage formation based on trajectory analysis. Airway progenitor cells give rise to pulmonary NE and *GHRL*$^{+}$ NE, via *ASCL1* or *NEUROD1* expression respectively.

A second route to *GHRL*$^+$ NE is for differentiating pulmonary NE cells to turn on *NEUROG3* expression and transition to the *GHRL*$^+$ state.

(D) *ASCL1* enrichment in *GRP*$^+$ pulmonary NE cells. Left: Representative HCR images showing *ASCL1* enriched in *GRP*$^+$ pulmonary NE cells. *GRP* (green), *GHRL* (red), *ASCL1* (white). Yellow line indicates the *GRP*$^+$*ASCL*$^+$ pulmonary NE cells. Dashed yellow line marks *GRP*$^-$*ASCL1*$^+$ pulmonary NE precursor cells. White dashed line indicates the *GHRL*$^+$*ASCL1*$^-$ *GHRL*$^+$ NE cells. Right panel: Mean ± SEM of *ASCL1*$^+$ cell types, N = 3 human fetal lungs, n = 243 *ASCL1*$^+$ cells.

(E) *NEUROD1* enrichment in *GHRL*$^+$ NE cells. Left: Representative HCR images showing *NEUROD1* expression in *GHRL*$^+$ NE cells. *GRP* (green), *NEUROD1* (red), *GHRL* (white). Yellow line indicates the *NEUROD1*$^+$*GHRL*$^+$ pulmonary NE cells. Dashed yellow line marks *NEUROD1*$^+$*GHRL*^{low} *GHRL*$^+$ NE precursor cells. Right panel: 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 in all panels.

(F) Pulmonary NE and *GHRL*$^+$ NE ressemble A-type (ASCL1$^+$) and N-type (NEUROD1$^+$) SCLC, respectively. Gene signature scoring of A-type and N-type SCLC features in the epithelial compartment.

(G) Scenic analysis of transcription factor network governing Mid tip progenitor cells to pulmonary NE transition. Trajectory and colour coding match Fig. S9B.

(H) Schematic of TF overexpression experiments. 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 and given 10-13 days to recover before Dox treatment. Organoid cells were harvested 3 days post-Dox for scRNA-Seq. N=3 organoid lines used.

(I) scRNA-Seq analysis revealed *ASCL1* and *NEUROD1* overexpression differentiated tip organoids into pulmonary NE precursors and *GHRL*$^+$ NE precursors, respectively. Left: reference UMAP of primary human fetal lung epithelium. AT1, alveolar type I cells; AT2, alveolar type 2 cells; GNE Pre, *GHRL*$^+$ NE precursors; GNE, *GHRL*$^+$ NE cells; INE, intermediate transition NE cells; PNE Pre, pulmonary NE precursors; PNE, pulmonary NE cells. Mid and right panels: scRNA-Seq of organoids overexpressing *mNeonGreen-NLS*, *ASCL1* or *NEUROD1* projected onto the primary data.

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 147 cell types or cell states

(A-E) All of the curated 147 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) Proximal-distal biases of cell types. Cell types are shown as dots with x representing p values of Fisher's exact test and y values representing fold of proximal enrichment.

Figure S3. Comparing fetal lung scRNA-seq with adult scRNA-seq using logistic regression models.

(A-E) Fractions of fetal cells in each cluster in row labels predicted to be each adult cell cluster (Madissoon et al. 2021 biorxiv) in column labels visualized using heatmaps, grouped into epithelial (A), PNS (B), fibroblast and SMC (C), erythrocyte and endothelial (D), and leukocyte compartments (E).
(F) Fractions of fetal cell clusters in each cluster in row labels predicted to be each adult cell cluster (Carraro et al. 2021) in column labels visualized using heatmaps.

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 and secretory progenitor cells in distal fetal lungs at 10 (B) and 16 (B') pcw. The secretory progenitor cells marked by $SOX9^{-}/CYTL1^{high/+}/SCGB3A2^{high/+}$ are proximally

located down to the airway progenitors marked by *CYTL1*$^{low/+}$/*SCGB3A2*$^{low/-}$/*SCGB1A1*$^{-}$. *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) Secretory 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, secretory progenitor, and club cells are lined from the distal tip regions to the proximal non-cartilaginous airway regions. *SCGB3A2* (green), *SCGB1A1* (red).

(E) Secretory 1 (arrowhead) and Secretory 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) Secretory 2 (arrowhead) and Secretory 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) Secretory 4 (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).

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

DAPI, nuclei. Scale bars, 50 μm.

Figure S5. Late epithelial tip cells differentiate to AT2 and AT1 cells.

(A) Trajectory UMAP visualization of a lineage trajectory derived using scvelo from Mid tip to AT1 cells.

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

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

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

DAPI, nuclei. Scale bars, 50 µm.

Figure S6. Endothelial cell types in the developing human lung.

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

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

(C-E) *In situ* HCR analysis of distal lung regions at 20 (D), and 21 (C, E) pcw, visualizing aerocytes (*S100A3*$^+$ red/*CA4*$^+$ white; C), arterial endothelial cells (*GJA5*$^+$ red; D), venous endothelial cells (*ACKR3*$^+$ white; D), and lymphatic endothelial cells (*PROX1*$^+$ white; E). *PECAM1* (green) indicates all endothelial cell types. DAPI, nuclei. Scale bars, 50 µm.

(F) Trajectory UMAP visualizing potential endothelial cell lineage hierarchy from Mid/Late capillary endothelial cells to arterial endothelial cells, aerocytes, or pulmonary venous endothelial cells coloured by cell types and stages.

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

Figure S7. 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 S8. 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*$^{-/low}$ (VSM-1, arrows) and *NTN4*$^{+}$/*PLN*$^{+/high}$ (VSM-2, lines).
(B) Dot plot describing differential gene expression between vascular SMC-1 and -2.
(C) Pericytes and adventitial fibroblasts in 17 pcw fetal lung. *NDUF4AL2*$^{+}$, red/*NTRK3*$^{+}$, white pericytes (arrows) are surrounded by *NDUF4AL2*$^{-}$/*NTRK3*$^{+}$ adventitial fibroblasts (arrowheads). *PECAM1* (green) indicates an endothelial cell tube.
(D) Arterial endothelial cells and pericytes/vascular SMCs in 16 pcw fetal lung. *PECAM1*$^{+}$, green/*GJA5*$^{+}$, red arterial endothelial cells are surrounded by *NTRK3*$^{+}$, white pericytes/vascular SMCs.
(E) *NTRK3*$^{low/+}$ pericytes are surrounding *PECAM1*$^{+}$/*GJA5*$^{-}$ venous endothelial cells. Red dashed lines indicate boundaries between the pericytes and venous endothelial cells.
(F) Dot plot describing differential marker gene expression level between alveolar, adventitial and airway fibroblasts.
(G) Concept network visualization of gene ontology (GO) analysis using clusterProfiler for differentially expressed genes in alveolar, adventitial and airway fibroblasts.
(H-J) Immunostaining of fetal lung tissues at 11 (H), 15 (I), and 21 (J) pcw, to visualise myofibroblast populations, Myofibroblast-1 (H) and -2 (I) surrounding the developing stalk epithelial tubes, and Myofibroblast-3 (J) surrounding the developing air sacs. ACTA2^{+}/PDGFRA^{+} Myofibroblast-1 (THBD^{weak}; H) and -2 (THBD^{high}, arrows; I). PDGFRA^{+} Myofibroblast-3 at 21 pcw, does not express ACTA2 (arrows; J).
(K) Dot plot describing differential gene expression level between myofibroblast-2 and -3. The myofibroblast-2 population showed enriched expression of Wnt signaling associated genes, e.g. *NOTUM*, *LEF1*, and *DACH2*.
(L) *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 S9. Signalling ligand-receptor interactions in the adventitial niche.

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

(B) Visium spot transcriptome cluster map visualizing 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 S10. Global landscape of motif enrichment.

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

Figure S11. 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 S12. 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 foetal 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.

STAR METHODS

RESOURCE AVAILABILITY

Lead contact

Lead contact Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contacts, Emma L. Rawlins (e.rawlins@gurdon.cam.ac.uk) and Kerstin B. Meyer (km16@sanger.ac.uk).

Materials Availability

Human lung organoid lines used in this study are available from the lead contact, Emma L. Rawlins (e.rawlins@gurdon.cam.ac.uk), with a completed Materials Transfer Agreement.

Data and code availability

- Sequencing data have been deposited at ENA and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Microscopy data reported in this paper will be shared by the lead contacts upon request.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contacts upon request.
- ATAC-seq pseudobulk coverage profiles can be browsed at https://genome.ucsc.edu/s/brianpenghe/scATAC_fetal_lung20211206

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human embryonic and foetal lung tissue

Human embryonic and foetal lung tissues were provided from terminations of pregnancy from Cambridge University Hospitals NHS Foundation Trust under permission from NHS Research Ethical Committee (96/085) and the MRC/Wellcome Trust Human Developmental Biology Resource (London and Newcastle, University College London (UCL) site REC reference: 18/LO/0822; Newcastle site REC reference: 18/NE/0290; Project 200454; www.hdbr.org). Sample age ranged from 4 to 23 weeks of gestation (post-conception weeks; pcw). Stages of the samples were determined according to their external physical appearance and measurements. Sample names and gestational ages are listed in Sup Table 3. None of the samples used for the current study had any known genetic abnormalities.

METHOD DETAILS

Cell isolation for 10X single cell RNA and ATAC seq

Proximal and distal regions for human fetal lung samples ≥15 pcw were separated as indicated in Figure 1 and minced with scissors. Whole fetal lung samples <15pcw were directly minced with scissors. Minced tissues were transferred into a 15 ml Falcon tube and mixed with 5 ml of dissociation solution (collagenase, 0.125mg/ml, Sigma, C9891-100MG; dispase, 1U/ml, Merck, 4942078001; DNaseI, 0.1 mg/ml, Merck, D4527-10KU). The mixture was incubated in a shaker incubator at 37°C with horizontal shaking at 135 rpm for 30min (after 15min of incubation, the mixture was triturated with 10ml straight pipette). 5 ml of termination solution (2% fetal bovine serum in PBS) was added to terminate the digestion reaction. A brief spin at 100X g was performed to pellet large tissue pieces. The supernatant was passed through a 40 µm filter. Any large undigested pieces were further trypsinized with 3 ml of 5X trypsin (Trypsin EDTA X10, Thermo Fisher Scientific, 15400054) for 3-6 min in 37°C water baths to further expose epithelial cells. The reaction was stopped using 5ml of termination solution, filtered through a 40 µm cell strainer and collected. Cells were pelleted at 500X g for 5 min at 4°C. If the pellets were red, a red blood cell (RBC) removal step was performed by resuspending cells in 1X RBC lysis buffer (Thermo Fisher, 00-4300-54) for 3 min at room temperature. RBC lysis buffer was neutralised with 10 ml of termination solution. The cell suspension was passed through a 40 µm filter again. For the trypsinized cells, a CD326 (EpCAM) MACS enrichment (Miltenyi Biotec, 130-061-101) was performed to further enrich epithelial cells. Cells were counted, pelleted and resuspended in appropriate volume with PBS/0.04%BSA and single cell RNA and ATAC seq was carried out using 10X Chromium Single Cell V(D)J Kits (v1) and Chromium Next GEM Single Cell ATAC Kits (v1), respectively.

Human fetal lung organoid maintenance

Human fetal lung organoids were derived and maintained as previously described (Nikolic et al., 2017). In brief, human foetal lung tissues were treated with Dispase (8 U/ml Thermo Fisher Scientific, 17105041) at room temperature (RT) for 2 min to digest mesenchymal connections. Mesothelium and mesenchymal cells were carefully removed by needles. Branching epithelial tips were micro-dissected by needles, transferred into Matrigel (356231, Corning) and seeded in a 24 well low-attachment plate (M9312-100EA, Greiner) with 4-5 tips per 50 µl Matrigel dome per well. The plate was incubated at 37°C for 5-10 min until the Matrigel domes solidified. 600 µl of self-renewing medium containing: N2 (1: 100, Thermo Fisher Scientific, 17502001), B27 (1: 50, Thermo Fisher Scientific, 12587001), N-acetylcysteine (1.25 mM), EGF (50 ng/ml, PeproTech, AF-100-15), FGF10 (100 ng/ml, PeproTech, 100-26), FGF7 (100 ng/ml, PeproTech, 100-19), Noggin (100 ng/ml, PeproTech, 120-10C), R-spondin (5% v/v, Stem Cell Institute, University of Cambridge), CHIR99021 (3 µM, Stem Cell Institute, University of Cambridge) and SB 431542

(10 μM, bio-techne, 1614), was added. Organoids were cultured under standard tissue culture conditions (37°C, 5% CO2), maintained in self-renewing medium and passaged by mechanically breaking using P200 pipettes every 4-7 days.

Human foetal lung organoid bronchiolar differentiation

The progenitor organoids were expanded in self-renewal medium in BME (Basement Membrane Extract, R&D Systems, 3533-010-02). For airway differentiation, the organoids were dissociated by TrypLE and cultured in the differentiation medium (AdvDMEM+++, 1X B27, 1X N2, 1.25 mM N-acetylcysteine, 100 ng/mL FGF10, 100 ng/mL FGF7, 50 nM Dexamethasone, 0.1 mM cAMP, 0.1 mM IBMX, 10 μM Y-27632) for 15-30 days.

Human foetal lung organoid immunofluorescence

The differentiated organoids were released from the BME and fixed in 4% PFA at 4℃ for 30 min. Then the organoids were washed in PBS, incubated in 0.3% PBTX (0.3% Triton X-100 in PBS) at 4℃ for 1 hour, and blocked (1% bovine serum albumin, 5% normal donkey serum, 0.3% Triton X-100 in PBS) at 4℃ overnight. The organoids were incubated with primary antibodies: SCGB3A2 (1:800, Abcam, ab181853), KRT5 (1: 500, BioLegend, 905901), E-cadherin (1: 500; Thermo Fisher Scientific, 13-1900), SOX9 (1: 400; Merck, AB5535), SCGB1A1 (1: 800, Proteintech, 10490-1-AP), FOXJ1 (1: 300, eBioscience, 14-9965-80) at 4℃ overnight. The organoids were washed by PBS and further incubated with secondary antibodies (donkey anti-chicken 488, 1: 1000, Jackson Immunoresearch, 703-545-155; donkey anti-mouse 594, 1: 1000, Invitrogen, A-21203; donkey anti-rabbit 594, 1: 1000, Invitrogen, A-21207; donkey anti-rat 647, 1: 1000, Jackson Immunoresearch, 712-605-153; donkey anti-rabbit 647, 1: 1000, Invitrogen, A-31573). After DAPI staining (1 μg/mL) at 4℃ for 1 hour, the organoids were processed through a thiodiethanol series (25%, 50%, 75% and 97% v/v concentration in PBS) at 4℃ for imaging.

Transcription factor overexpression in human fetal lung organoids

Plasmid cloning

cDNAs for genes ASCL1, NEUROD1, NEUROG3, RFX6 and PAX9 were purchased from Genscript. cDNAs for gene TFAP2A and mNeonGreen-3XNLS were gifts from Azim Surani's Group. cDNA for DeltaNTP63 was purchased from IDT as a gBlock fragment. cDNA sequences were cloned into a Doxycycline inducible vector pLenti-tetON-KRAB-dCas9-DHFR-EF1a-TagRFP-2A-tet3G (Addgene: #167935) (Sun et al., 2021) using XhoI and BamHI sites by Infusion cloning (Takara, 638910). Plasmids used in this study will be deposited to Addgene.

Lentivirus packaging

We packaged the lentivirus as described previously (Sun et al., 2021). In brief, HEK293T cells were grown in 10-cm dishes to 70%-80% confluence. Lentiviral vector (10 µg) was co-transfected with packaging MD2.G (3 µg, Addgene plasmid # 12259), psPAX2 (6 µg, Addgene plasmid # 12260) and pAdVAntage (3 µg, E1711, Promega) using Lipofectamine 2000 Transfection Reagent (11668019, Thermo Fisher Scientific) according to manufacturer's protocol. Medium was refreshed the next morning. Lentivirus containing cell medium was harvested at 24 hrs and 48 hrs after medium refreshing and pooled together. Cell fragments were removed by 300X g centrifugation. Supernatant was then passed through a 0.45 µm filter. Lentivirus was concentrated using Lenti-X™ Concentrator (631232, Takara) according to the manufacture's instructions. Lentivirus pellets were dissolved in 400 µl PBS, aliquoted and frozen in -80°C.

Lentivirus transduction

Lentivirus transduction was performed as previously described (Sun et al., 2021). In brief, human foetal lung organoids derived from 3 independent donors were incubated with prewarmed TrypLE for 10min with trituration after 5 min. Organoid single cells and small fragments were collected, counted, pelleted and resuspended to around 100K cells/500 µl self-renewing medium with ROCKi (10 µM Y-27632). 0.5-2 µl of lentivirus was added and incubated overnight. Organoid cells were harvested the next morning, pelleted and re-seeded into Matrigel.

Overexpression of transcription factors and scRNA-Seq

After 3 days of lentivirus transduction, organoids were dissociated by incubation with prewarmed TrypLE for 10min with trituration after 5 min. TagRFP positive cells were sorted (20-40% of TagRFP positive rate), seeded back to Matrigel and allowed to recover for 10-12 days with self-renewing medium plus ROCKi (10 µM Y-27632). Organoids were treated with Doxycycline (2 µg/ml) for 3 days. Organoids were then fully dissociated into single cells by incubation with prewarmed TrypLE (Thermo Fisher Scientific, 12605028) for 15-20 min with trituration every 5 min. Organoid cells were counted, pelleted, resuspended in proper amounts of PBS/0.04%BSA and proceeded to scRNA-Seq according to 10X Chromium Single Cell V(D)J Kit manual.

In situ hybridization chain reaction and immunofluorescence

In situ HCR v3.0 was performed according to the manufacturer's protocol (Molecular Instruments (Choi et al., 2018)). Probes were designed according to the manual, and amplifiers with buffers were supplied by Molecular Instruments. All the sequence information of the probes is listed in Sup Table 4. In brief, the frozen human tissue sections fixed in 4% PFA/DEPC-treated PBS were cut into 20 µm slices and rinsed in nuclease-free ultrapure water, followed by 10 µg/mL proteinase K solution (Thermo Fisher Scientific, AM2546) for 2 min at 37°C. Then, the tissue slices were

incubated with 2 pmol of probes at 37°C overnight. After washing, the slices were treated with 6 pmol of the amplifiers at room temperature overnight. The amplifiers, consisting of a pair of hairpins conjugated to fluorophores, Alexa 488, 546, or 647, were used at final concentration of 0.03 µM. After rinsing excess hairpins in 5X SSC (sodium chloride sodium citrate) solution containing 0.1% Triton X-100, nuclei were counterstained with DAPI.

For immunostaining, the frozen human tissue sections at 20 µm thickness were permeabilized in 0.3% Triton-X/DEPC-treated PBS for 20 min at room temperature. Then the tissues were treated with a blocking solution containing 5% NDS, 1% BSA, 0.1% Triton-X in DEPC-treated PBS at 4°C for 3 hours. After rinsing with cold PBS, treated with primary antibodies against ACTA2 (1:500; Thermo Fisher Scientific, MA1-06110), THBD (1:100; PE-conjugated; BioLegend, 344104), PDGFRA (1:200; Cell Signaling Technology, 3174), S100A4 (1:200; Proteintech, 16105-1-AP), CD44 (1:200; Thermo Fisher Scientific, 17-0441-82), SOX9 (1: 200, Merck, AB5535), PDPN (1:200; R&D Systems, AF3670), or E-cadherin (1: 500; Thermo Fisher Scientific, 13-1900) for 24 hr. Secondary antibodies were treated for 3 hr at room temperature. The tissue was washed three times in PBS at room temperature and counterstained with DAPI. Images were collected under Leica SP8 confocal microscope.

Library Generation and Sequencing

Chromium Single Cell 5' V(D)J Reagent Kits (V1.0 chemistry) were used for scRNAseq library construction. Gene expression libraries (GEX) and V(D)J libraries were prepared according to the manufacturer's protocol (10X Genomics) using individual Chromium i7 Sample Indices. Libraries for gamma/delta TCR variable regions were amplified as previously described (Mimitou et al., 2019)(Conde et al.). GEX and V(D)J were pooled in 1:0.1 ratio respectively and sequenced on a NovaSeq 6000 S4 Flowcell (paired-end (PE), 150-bp reads) aiming for a minimum of 50,000 PE reads per cell for GEX libraries and 5,000 PE reads per cell for V(D)J libraries.

Visium Spatial Transcriptomics

Foetal lung samples at 12-20 post conception week (pcw) from the HDBR, up to 0.5cm^3 in size, were embedded in OCT and flash-frozen in dry-ice cooled isopentane. Twelve-micron cryosections were cut onto Visium slides, haematoxylin and eosin stained and imaged at 20X magnification on a Hamamatsu Nanozoomer 2.0 HT Brightfield. These were then further processed according to the 10X Genomics Visium protocol, using a permeabilization time of 18min for 12-17 pcw samples and 24 min for 19 pcw and older samples. Images were exported as tiled tiffs for analysis. Dual-indexed libraries were prepared as in the 10X Genomics protocol, pooled at 2.25 nM and sequenced 4 samples per Illumina Novaseq SP flow cell with read lengths 28bp R1, 10bp i7 index, 10bp i5 index, 90bp R2.

Reads mapping and quantification.

scRNA-seq data were mapped with STARsolo 2.7.3a (Kaminow et al., 2021) to the 10X distributed GRCh38 reference, version 3.0.0, derived from Ensembl 93. Cell calling was post-processed with an implementation of EmptyDrops (Lun et al., 2019) extracted from Cell Ranger 3.0.2 (distributed as empty drops on PyPi). For transduced organoid cells, exogenous genes were added to the reference as appropriate for organoids, with the transgene sequence truncated (length(R2)-1) bp after the end of the synthetic promoter to avoid reads from endogenous transcripts being mapped onto transgenes. For single-cell V(D)J data, reads were mapped with Cell Ranger 4.0.0 to the 10X distributed VDJ reference, version 4.0.0. Visium reads were mapped with Space Ranger 1.1.0 to the 10X distributed GRCh38 reference, version 3.0.0, derived from Ensembl 93 for consistency with the single cell data. scATAC reads were mapped with Cellranger-atac 1.2.0 to reference GRCh38-1.2.0.

VDJ analysis

Both TCR and BCR contigs contained in respective all_contigs.fasta and all_contig_annotations.csv files were re-annotated with igblastn (v1.17.1) using reference sequences curated from IMGT database (downloaded 01-Aug-2021) as per described with changeo (v1.0.0). For BCR contigs, heavy chain constant region calls were re-annotated using blastn (v2.12.0+) against curated sequences of CH1 regions corresponding to respective isotype classes from IMGT. BCR heavy chain V-gene alleles were corrected for individual genotypes using tigger (v1.0.0) (Gadala-Maria et al., 2015). Contigs were then filtered for basic quality control as described previously (Stephenson et al., 2021). Briefly, the following occurrences would lead to removal of contigs from further analysis: i) contigs were annotated with V, D, J or constant gene calls that are not from the same locus; ii) multiple long/heavy chain contigs present in the same cell; iii) there were only short/light chain contigs in a cell; and/or iv) there are multiple short/light chain contigs in a cell. Cells with multiple contigs were nevertheless retained if a) contigs were assessed to have identical V(D)J sequences but were assigned to a different contig by cellranger-vdj (presumably due to differences in non-V(D)J elements); b) UMI count differences were large in which case the contig with the highest UMI count is retained; and c) only IgM and IgD were both assigned to a cell. These checks were all performed using dandelion (Stephenson et al., 2021) singularity container (v0.1.10).

Single-cell RNA-seq processing and cell type annotation

Count matrices were loaded into Scanpy and concatenated. Cells expressing no more than 200 genes, and genes detected in no more than 5 cells, were removed. Cells having more than 20% of their reads mapped to mitochondria were also discarded. Counts were then divided by total counts and multiplied by a factor of 10000, followed by log transformation, all implemented in Scanpy's default setting (Wolf et al., 2018).

$Y_{ij} = ln(\frac{X_{ij}}{\sum_{i=1}^{n} X_{ij}} \cdot 10000 + 1)$, where X_{ij} is the raw count of ith gene in j^{th} cell

Feature genes were selected in three steps: For each sample, highly variable genes were calculated using Scanpy's default settings that extract genes with highest dispersion (variance divided by mean) values of log-transformed counts. Next, highly correlated genes for each sample were extracted using the DeepTree algorithm described in (He et al., 2020), reimplemented in our python-genomics toolkit. Genes extracted in at least two samples were merged as the final feature gene list. The log-transformed counts of these genes were then scaled and cell-cycle scores were regressed out using Scanpy's default scoring and regression functions. Using the top 50 PCs and 10 neighbors with resolution at 0.01, initial clustering was generated, yielding 10 major clusters (Figure S1C) corresponding to different compartments. These clusters were subsequently and recursively subclustered, curated and annotated manually (Figure S1D). Annotation was based on markers summarized in Table S1.

Artifact evaluation and removal for scRNA-seq data.

Doublets were evaluated using Scrublet in a batch-by-batch fashion (Figure S1G). To capture rare doublet clusters, we developed a method for Doublet Cluster Labeling (DouCLing, Figure S1E). Briefly, we calculated relative marker genes for each subcluster compared to other subclusters in the same parental large cluster. Then these marker genes were used to score all the cells in the atlas. If the top-scoring cells (above the mean score of the current subcluster) are mostly (>60%) from another large cluster, the clusters are flagged as doublet-like (Figure S1H). We then removed doublet-like clusters based on these two methods with manual curation (Figure S1K).

Maternally derived cells were evaluated based on SNP variations between the transcribed paternal genome in the fetuses and the maternal counterparts in the maternal cells. To do this, we indexed and pooled samples from the same donor into "Supersamples". Then we applied Souporcell to compare known common variants captured in scRNA-seq reads, setting the sample number to 2. Supersamples without maternal cells would split into two equal-sized groups while other supersamples would putatively capture maternal cells as a minor genotype group (Figure S1F). Based on this analysis, maternal-like cells do not contribute to scRNA-seq clusters (Figure S1M) and were thus kept for downstream analysis. For libraries with two multiplexed donors, we only used the Souporcell workflow to demultiplex the donors without maternal genetic detection.

Low-quality cells would usually have a relatively high percentage of mitochondria reads (Figure S1I) or a low number of genes detected (Figure S1J). Based on these we manually curated and removed low-quality clusters (Figure S1N).

An additional four clusters of contaminants coming from other organs were further removed. These were cardiomyocytes $ACTN2^{+}$ $MYH6^{+}$ (Litviňuková et al., 2020), esophagus epithelial cells

SOX2$^+$ TP63$^+$ TRH$^+$ (Madissoon et al., 2019), APOA1$^+$ APOA2$^+$ (Popescu et al., 2019) and cytotrophoblasts from the placenta PAGE4$^+$ GSTA3$^+$ (Vento-Tormo et al., 2018).

Visium Spatial Transcriptomics data analysis

Mapped Visium count matrices and scRNA-seq count matrices (after artifact removal) were both imported into Seurat 3 (Stuart et al., 2019) and transformed using SCTransform (Hafemeister and Satija, 2019), with mitochondria percentage of scRNA-seq data regressed out. Next, the scRNA-seq data were subsetted into a "pcw11,15,18" subgroup and a "pcw18,20,22" subgroup for cell-type prediction. The prediction was done for each Visium library using its corresponding scRNA-seq subgroup following the default label transfer pipeline of Seurat using the top 50 PCs.

Marker gene calculation

Ambrient RNA was removed with SoupX 1.4.5 with default parameters. Using the corrected count matrices, Scanpy.tl.rank_genes_groups was applied with default settings but keeping all the genes. These ranked genes were then filtered using Scanpy.tl.filter_rank_genes_groups with max_out_group_fraction=0.25 and min_fold_change=2. To compare specific cell types, Scanpy.tl.rank_genes_groups was applied for each cell type with only the other cell types of this subset as a reference. Over-representation analysis (hypergeometric test) with gene sets from GO BP, KEGG and MSigDB was performed using the clusterProfiler R package (Yu et al., 2012).

Statistical analysis for cell-type composition biases.

Chi-squared test of independence was performed for sample gestation age, cell-cycle stage and proximal/distal dissection regions against cell type categories. For proximal/distal biases, Fisher's exact test was used for each cell type and Benjamini-Hochberg correction was performed for multiple testing. The effect size for each cell type was defined as the odds ratio of its fraction of proximally sampled cells over that of distally sampled ones.

Differential gene expression along trajectories

The single cell transcriptomics data was preprocessed using Scanpy (Wolf et al., 2018) version 1.8.1. The cell cycle effect was regressed out using sc.pp.regress_out() and batch correction was performed using bbknn (Polański et al., 2020), before denoising the knn-graph using diffusion maps (Haghverdi et al., 2015) with sc.tl.diffmap() and applying PAGA (Wolf et al., 2019) with sc.tl.paga() to examine the connectivities between cell types. The final UMAPs were computed using the results of PAGA on Leiden (Traag et al., 2019) clusters as described in (Wolf et al., 2019). Data and UMAPs were exported into R, and monocle3 (Cao et al., 2019; Trapnell et al., 2014) was used to find a principal graph and define pseudotime. Differentially expressed genes were then computed along pseudotime using a graph-based test (morans' I) (Cao et al., 2019; Moran, 1950) and the principal graph in monocle3, which allows identification of genes

upregulated at any point in pseudotime. The results were visualized with heatmaps using the complexHeatmap (Gu et al., 2016) and seriation (Hahsler et al., 2008) packages, after smoothing gene expression with smoothing splines in R (smooth.spline(), df=12).

CellPhoneDB analysis

Filtered single-cell RNA-seq data were partitioned into early- (5-6pcw), middle- (9-11) and late-stage (15-22) subsets. These datasets were used as input for CellPhoneDB (Vento-Tormo et al., 2018) (command: cellphonedb method statistical_analysis --database v2.0.0 --threads 20 --counts-data gene_name --project-name FetalLungBroadLate --subsampling --subsampling-log False --subsampling-num-cells=$TotalCellNumber --iterations=10000 --result-precision=4). Interaction pairs were manually curated from the outputs.

Velocity analysis

Velocity analysis (La Manno et al., 2018) was performed using scvelo (Bergen et al., 2020) version 0.2.3. The preprocessed dataset was merged with spliced and unspliced read counts computed with velocyto, before using scv.pp.moments(), scv.tl.velocity() and scv.tl.velocity_graph() to compute velocities using the stochastic mode in scvelo.

Gene regulatory network analysis

The Scenic pipeline (Aibar et al., 2017; Van de Sande et al., 2020) was used (pySCENIC version 0.11.2) to predict transcription factors and putative target genes regulated throughout neuroendocrine cell differentiation. First, gene regulatory interactions were calculated based on co-expression across the single cell dataset with GRNBoost2 (Moerman et al., 2019), followed by pruning interactions using known TF binding motifs and the construction of dataset specific regulatory modules (regulons) (Imrichová et al., 2015). Regulons were then scored in each individual cell using AUCell. Cells of the neuroendocrine differentiation trajectory computed with monocle3 (as described above) were selected. The regulon target genes were filtered for differentially expressed genes along pseudotime for this trajectory. A network of TFs and target genes was then constructed by linking individual regulons.

Comparing fetal neuroendocrine transcriptome with SCLC

A-type and N-type signatures were selected from previous data ‘ASCL1High and NEUROD1High Gene Signatures and the Stratified Primary Tumor Samples’ (Borromeo et al., 2016). Top 10 genes with the highest fold enrichment were selected to score epithelial cells, using Scanpy’s tl.score_genes function.

Compare scRNA-seq datasets of the fetal lung and other studies.

Annotated scRNA-seq adult lung datasets were acquired from (Madissoon et al., 2021) and downloaded from the study by (Carraro et al., 2021). The multi-organ scRNA-seq dataset was downloaded from the study by (Cao et al., 2020). A logistic regression model was trained based on the fine-grained cell-types for each of these datasets, using sklearn.linear_model. LogisticRegression (Pedregosa et al., 2011). The trained models were then used to predict the cell types of single-cell transcriptomic profiles of the fetal lung (Figure S3, S7L).

Single-cell ATAC-seq processing and annotation

Cellranger-atac outputs were loaded into and processed by ArchR (Granja et al., 2021). The top 50 dimensions were used for LSI and no batch effect was carried out to preserve weak biological features. Doublets were removed using ArchR's default settings. Cells with TSSEnrichment score < 8 or ReadsInTSS < 1000 were discarded. Initial clustering was performed at resolution = 0.01 to be consistent with scRNA-seq, resulting in 7 large clusters corresponding to compartments. These clusters were further subclustered, similar to the workflow for scRNA-seq.

To annotate cell types and doublets, the annotated scRNA-seq h5ad object was loaded into Seurat3 by Seurat3-plus and integrated to scATAC-seq data using ArchR. The predicted cell type/state labels were used as a major reference for annotation. Clusters mapped to scRNA-seq doublet clusters were removed. Clusters with high fractions of blacklisted reads were also manually discarded.

Peaks were then called based on pseudo-bulk coverages by macs2. Marker peaks were calculated with default settings. Motifs from cis-bp database that are enriched in marker peaks were calculated and plotted.

Comparing organoid scRNA-seq with fetal lung scRNA-seq

Organoid scRNA-seq data were imported and filtered in the same way as described above. Organoid data were then projected onto fetal tissue data by Scanpy's tl.ingest function. Donors were demultiplexed using Souporcell with k=3 donors, based on common variants.

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods 14, 1083–1086.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. 38, 1408–1414.

Borromeo, M.D., Savage, T.K., Kollipara, R.K., He, M., Augustyn, A., Osborne, J.K., Girard, L.,

Minna, J.D., Gazdar, A.F., Cobb, M.H., et al. (2016). ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. Cell Rep. 16, 1259–1272.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502.

Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A., Aldinger, K.A., Blecher-Gonen, R., Zhang, F., et al. (2020). A human cell atlas of fetal gene expression. Science.

Carraro, G., Langerman, J., Sabri, S., Lorenzana, Z., Purkayastha, A., Zhang, G., Konda, B., Aros, C.J., Calvert, B.A., Szymaniak, A., et al. (2021). Transcriptional analysis of cystic fibrosis airways at single-cell resolution reveals altered epithelial cell states and composition. Nat. Med. 27, 806–814.

Conde, C.D., Domínguez Conde, C., Xu, C., Jarvis, L.B., Gomes, T., Howlett, S.K., Rainbow, D.B., Suchanek, O., King, H.W., Mamanova, L., et al. Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture in humans.

Gadala-Maria, D., Yaari, G., Uduman, M., and Kleinstein, S.H. (2015). Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. Proc. Natl. Acad. Sci. U. S. A. 112, E862–E870.

Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. 53, 403–411.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32, 2847–2849.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296.

Haghverdi, L., Buettner, F., and Theis, F.J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics 31, 2989–2998.

Hahsler, M., Hornik, K., and Buchta, C. (2008). Getting Things in Order: An Introduction to theRPackageseriation. Journal of Statistical Software 25.

He, P., Williams, B.A., Trout, D., Marinov, G.K., Amrhein, H., Berghella, L., Goh, S.-T., Plajzer-Frick, I., Afzal, V., Pennacchio, L.A., et al. (2020). The changing mouse embryo transcriptome at whole tissue and single-cell resolution. Nature 583, 760–767.

Imrichová, H., Hulselmans, G., Atak, Z.K., Potier, D., and Aerts, S. (2015). i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. Nucleic Acids Res. 43, W57–W64.

Kaminow, B., Yunusov, D., and Dobin, A. (2021). STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature 560, 494–498.

Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C.L., Lindberg, E.L., Kanda, M., Polanski, K., Heinig, M., Lee, M., et al. (2020). Cells of the adult human heart. Nature 588, 466–472.

Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., participants in the 1st Human Cell Atlas Jamboree, and Marioni, J.C. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. 20, 63.

Madissoon, E., Wilbrey-Clark, A., Miragaia, R.J., Saeb-Parsy, K., Mahbubani, K.T., Georgakopoulos, N., Harding, P., Polanski, K., Huang, N., Nowicki-Osuch, K., et al. (2019). scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. Genome Biol. 21, 1.

Madissoon, E., Oliver, A.J., Kleshchevnikov, V., Wilbrey-Clark, A., Polanski, K., Orsi, A.R., Mamanova, L., Bolt, L., Richoz, N., Elmentaite, R., et al. (2021). A spatial multi-omics atlas of the human lung reveals a novel immune cell survival niche.

Mimitou, E.P., Cheng, A., Montalbano, A., Hao, S., Stoeckius, M., Legut, M., Roush, T., Herrera, A., Papalexi, E., Ouyang, Z., et al. (2019). Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat. Methods 16, 409–412.

Moerman, T., Aibar Santos, S., Bravo González-Blas, C., Simm, J., Moreau, Y., Aerts, J., and Aerts, S. (2019). GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics 35, 2159–2161.

Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. Biometrika 37, 17–23.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Polański, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., and Park, J.-E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics 36, 964–965.

Popescu, D.-M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E.F., Polanski, K., Goh, I., Efremova, M., et al. (2019). Decoding human fetal liver haematopoiesis. Nature 574, 365–371.

Stephenson, E., Reynolds, G., Botting, R.A., Calero-Nieto, F.J., Morgan, M.D., Tuong, Z.K., Bach, K., Sungnak, W., Worlock, K.B., Yoshida, M., et al. (2021). Single-cell multi-omics analysis of the immune response in COVID-19. Nat. Med. 27, 904–916.

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

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Scientific Reports 9.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32, 381–386.

Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., et al. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat. Protoc. 15, 2247–2276.

Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.-E., Stephenson, E., Polański, K., Goncalves, A., et al. (2018). Single-cell reconstruction of the early maternal–fetal interface in humans. Nature 563, 347–353.

Wolf, F.A., Alexander Wolf, F., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biology 19.

Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 20, 59.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology 16, 284–287.

Figure 1

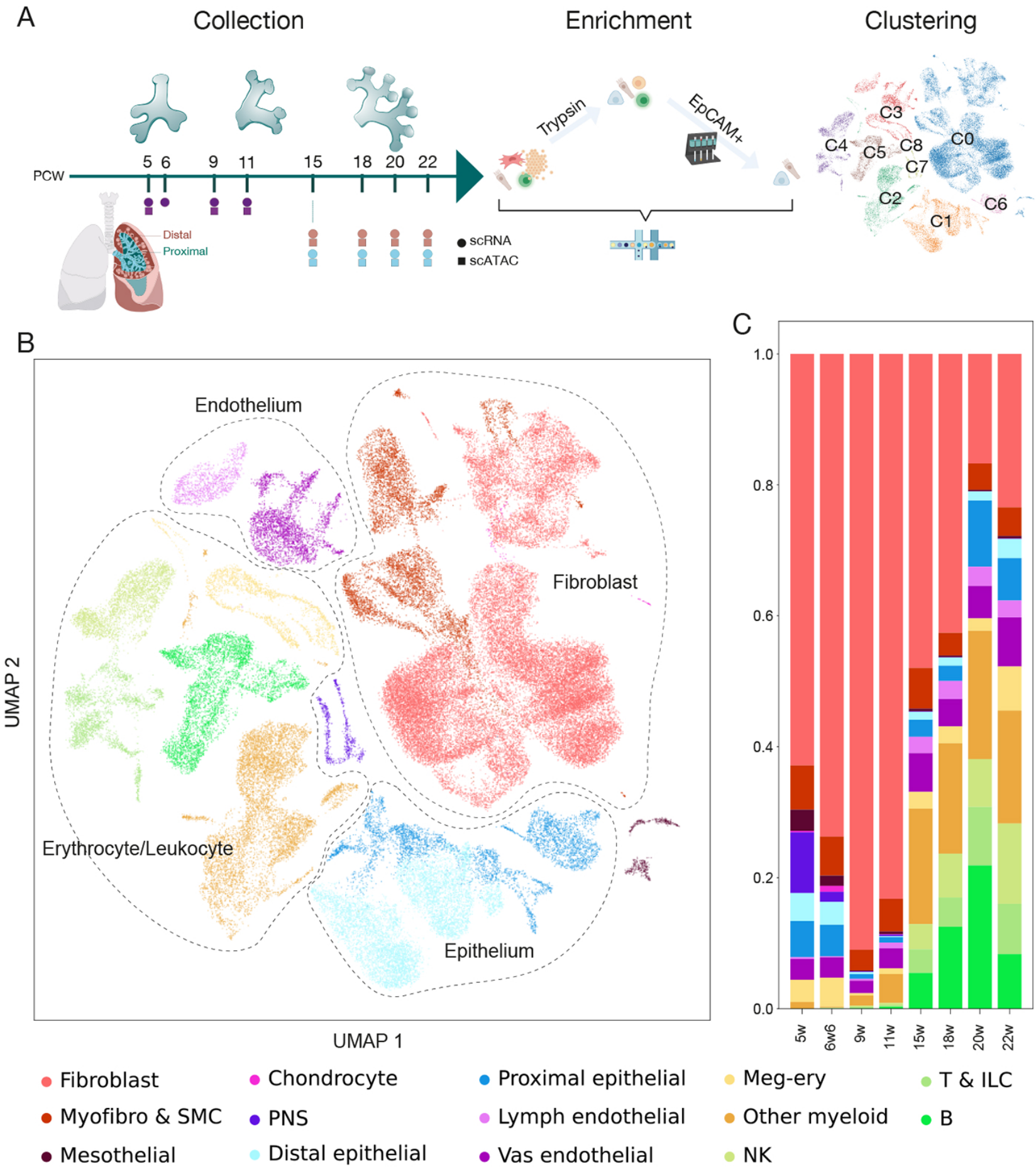

A
Collection
Enrichment
Clustering
PCW
5 6 9 11 15 18 20 22
Distal
Proximal
scRNA
scATAC
Trypsin
EpCAM+
C0
C1
C2
C3
C4
C5
C6
C7
C8
B
Endothelium
Fibroblast
Erythrocyte/Leukocyte
Epithelium
UMAP 2
UMAP 1
C
1.0
0.8
0.6
0.4
0.2
0.0
5w
6w6
9w
11w
15w
18w
20w
22w
Fibroblast
Myofibro & SMC
Mesothelial
Chondrocyte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

Figure 2

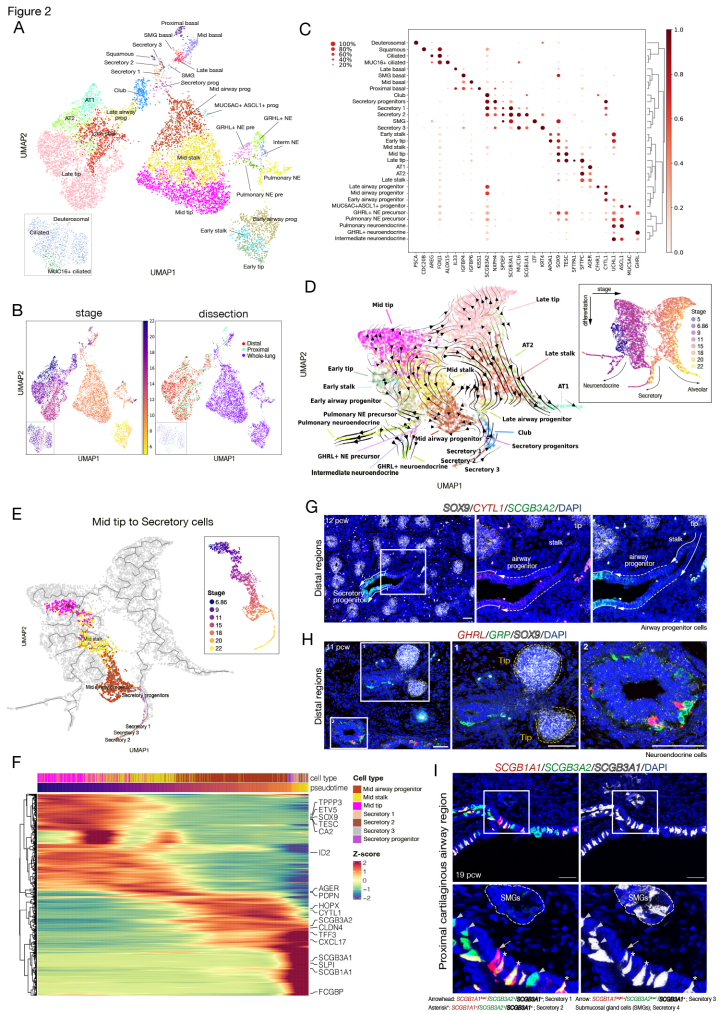
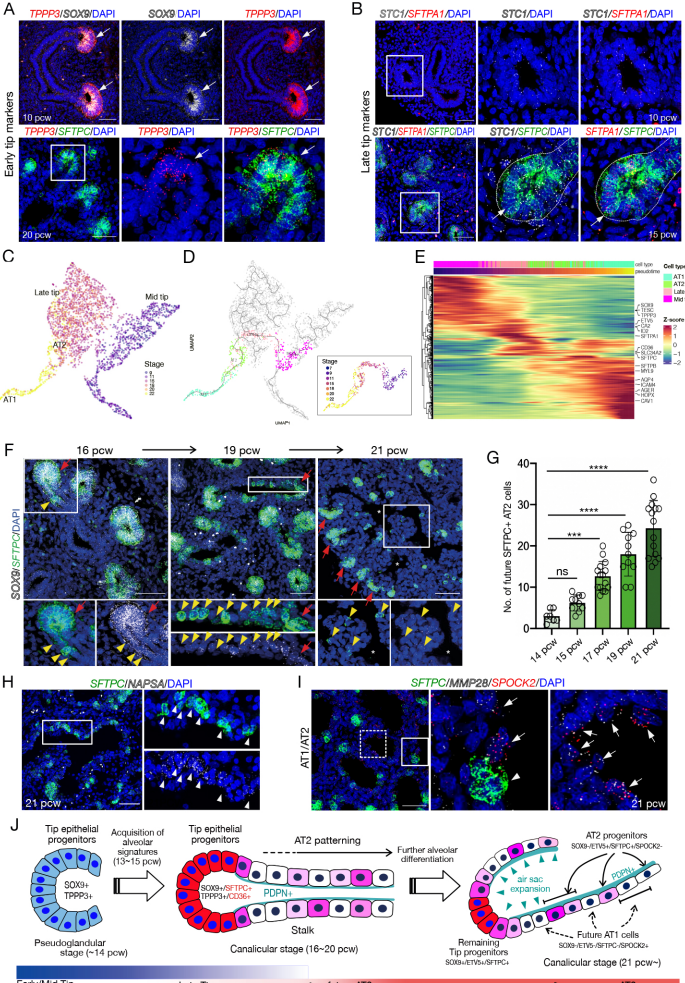

A
Proximal basal
SMG basal
Mid basal
Secretory 3
Squamous
Secretory 2
Secretory 1
Late basal
SMG
Secretory prog
Club
Mid airway prog
AT1
MUC5AC+ ASCL1+ prog
Late airway prog
AT2
GRHL+ NE
Ciliated
GRHL+ NE pre
Interm NE
Late tip
Mid stalk
Pulmonary NE
Pulmonary NE pre
Mid tip
Early airway prog
Early stalk
Early tip
Deuterosomal
Ciliated
Mid16+ ciliated
UMAP2
UMAP1
B
stage
dissection
Distal
Proximal
Whole-lung
UMAP2
UMAP1
C
100%
80%
60%
40%
20%
Deuterosomal
Squamous
Ciliated
MUC16+ ciliated
Late basal
SMG basal
Mid basal
Proximal basal
Club
Secretory progenitors
Secretory 3
Secretory 2
SMG
Secretory 1
Early stalk
Early tip
Mid stalk
Mid tip
Late tip
AT1
AT2
Late stalk
Late airway progenitor
Mid airway progenitor
Early airway progenitor
MUC5AC+ASCL1+ progenitor
GRHL+ NE precursor
Pulmonary NE precursor
Pulmonary neuroendocrine
GRHL+ neuroendocrine
Intermediate neuroendocrine
PIGA
CDH26
ARSG
FDX1
GUCA2A
L33
GPBP1
DGS1
KCNQ1
NPPA4
SCGB3A2
SCGB3A1
SCGB1A1
SCGB1A1
ITP
APOA1
CDO1
TEC
SYTL1
SYTL2
PAR1
CHGA
CYTL1
ASCL1
NKX2-1
GRPR
C
1.0
0.8
0.6
0.4
0.2
0.0
D
Mid tip
Late tip
AT2
Late stalk
Mid stalk
Early tip
Early stalk
AT1
Early airway progenitor
Pulmonary NE precursor
Pulmonary neuroendocrine
Late airway progenitor
Club
Mid airway progenitor
Secretory progenitors
Secretory 1
GRHL+ NE precursor
Secretory 2
GRHL+ neuroendocrine
Secretory 3
Intermediate neuroendocrine
UMAP2
UMAP1
stage
differentiation
Stage
5
6.86
9
11
15
18
20
22
Neuroendocrine
Secretory
Alveolar
E
Mid tip to Secretory cells
Stage
6.86
9
11
15
18
20
22
Mid stalk
Mid airway progenitor
Secretory progenitors
Secretory 1
Secretory 3
Secretory 2
UMAP2
UMAP1
G
SOX9/CYTL1/SCGB3A2/DAPI
12 pcw
Distal regions
Secretory progenitors
stalk
airway progenitor
Airway progenitor cells
H
GHRL/GRP/SOX9/DAPI
11 pcw
Distal regions
Tip
Neuroendocrine cells
F
cell type
pseudotime
Cell type
Mid airway progenitor
Mid stalk
Mid tip
Secretory 1
Secretory 2
Secretory 3
Secretory progenitor
Z-score
2
1
0
−1
−2
TPPP3
ETV5
SOX9
TESC
CA2
IO2
ASER
PDPN
HOPX
CYTL1
SCGB3A2
CLDN4
TFF3
CXCL17
SCGB3A1
SLPI
SCGB1A1
FCGBP
I
SCGB1A1/SCGB3A2/SCGB3A1/DAPI
Proximal cartilaginous airway region
19 pcw
SMGs
Arrowhead: SCGB1A1low/SCGB3A2+/SCGB3A1+; Secretory 1
Asterisk*: SCGB1A1+/SCGB3A2+/SCGB3A1+; Secretory 2
Arrow: SCGB1A1low/SCGB3A2low/SCGB3A1+; Secretory 3
Submucosal gland cells (SMGs), Secretory 4

Figure 3



A
TPPP3/SOX9/DAPI
SOX9 DAPI
TPPP3/DAPI
Early tip markers
10 pcw
TPPP3/SFTPC/DAPI
TPPP3/DAPI
TPPP3/SFTPC/DAPI
20 pcw
B
STC1/SFTPA1/DAPI
STC1/DAPI
STC1/SFTPA1/DAPI
Late tip markers
10 pcw
STC1/SFTPA1/SFTPC/DAPI
STC1/SFTPC/DAPI
SFTPA1/SFTPC/DAPI
15 pcw
C
Late tip
Mid tip
AT2
AT1
Stage
D
Stage
E
cell type
pseudotime
Cell type
AT1
AT2
Late tip
Mid tip
Z-score
F
16 pcw
19 pcw
21 pcw
SOX9/SFTPC/DAPI
G
No. of future SFTPC+ AT2 cells
****
****
***
ns
14 pcw
15 pcw
17 pcw
19 pcw
21 pcw
H
SFTPC/NAPSA/DAPI
21 pcw
I
SFTPC/MMP28/SPOCK2/DAPI
AT1/AT2
21 pcw
J
Tip epithelial progenitors
SOX9+
TPPP3+
Pseudoglandular stage (~14 pcw)
Acquisition of alveolar signatures (13~15 pcw)
Tip epithelial progenitors
AT2 patterning
SOX9+/SFTPC+
TPPP3+/CD36+
PDPN+
Stalk
Canalicular stage (16~20 pcw)
Further alveolar differentiation
air sac expansion
AT2 progenitors
SOX9+/ETV5+/SFTPC+/SPOCK2-
PDPN+
Remaining Tip progenitors
SOX9+/ETV5+/SFTPC+
Future AT1 cells
SOX9-/ETV5-/SFTPC-/SPOCK2+
Canalicular stage (21 pcw~)
Early/Mid Tip
Late Tip
future AT2
AT2
AT1

Figure 4

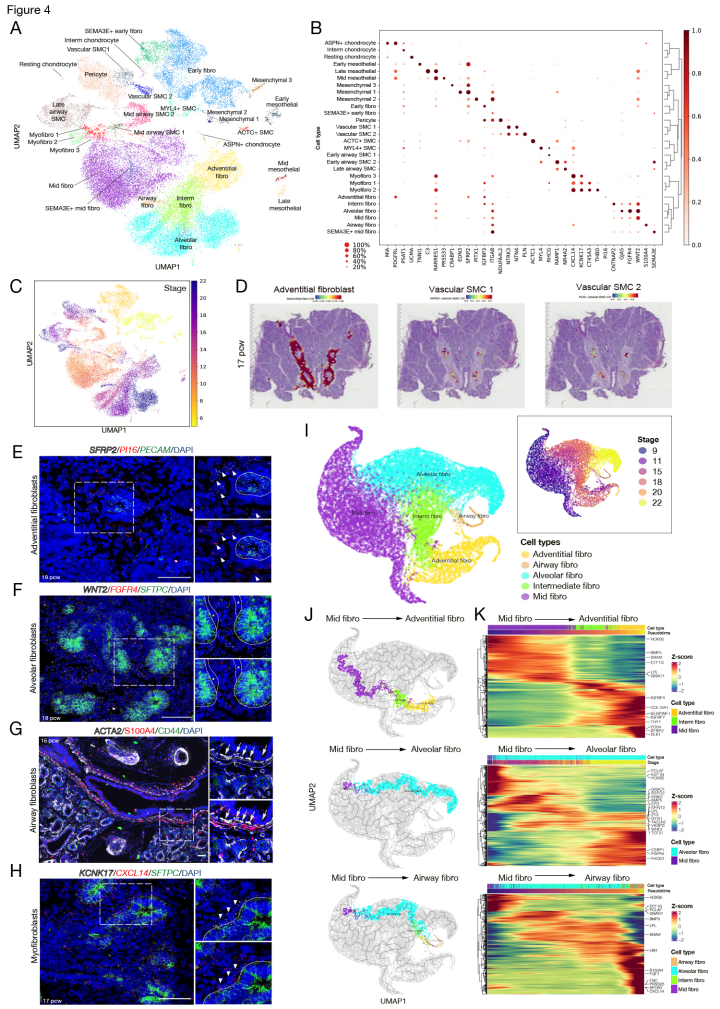

A
UMAP1
UMAP2
B
Cell type
C
Stage
D
Adventitial fibroblast
Vascular SMC 1
Vascular SMC 2
17 pcw
E
SFRP2/PI16/PECAM/DAPI
Adventitial fibroblasts
19 pcw
F
WNT2/FGFR4/SFTPC/DAPI
Alveolar fibroblasts
18 pcw
G
ACTA2/S100A4/CD44/DAPI
Airway fibroblasts
19 pcw
H
KCNK17/CXCL14/SFTPC/DAPI
Myofibroblasts
17 pcw
I
Stage
Cell types
Adventitial fibro
Airway fibro
Alveolar fibro
Intermediate fibro
Mid fibro
J
Mid fibro → Adventitial fibro
Mid fibro → Alveolar fibro
Mid fibro → Airway fibro
K
Z-score
Cell type
Pseudotime

Figure 5

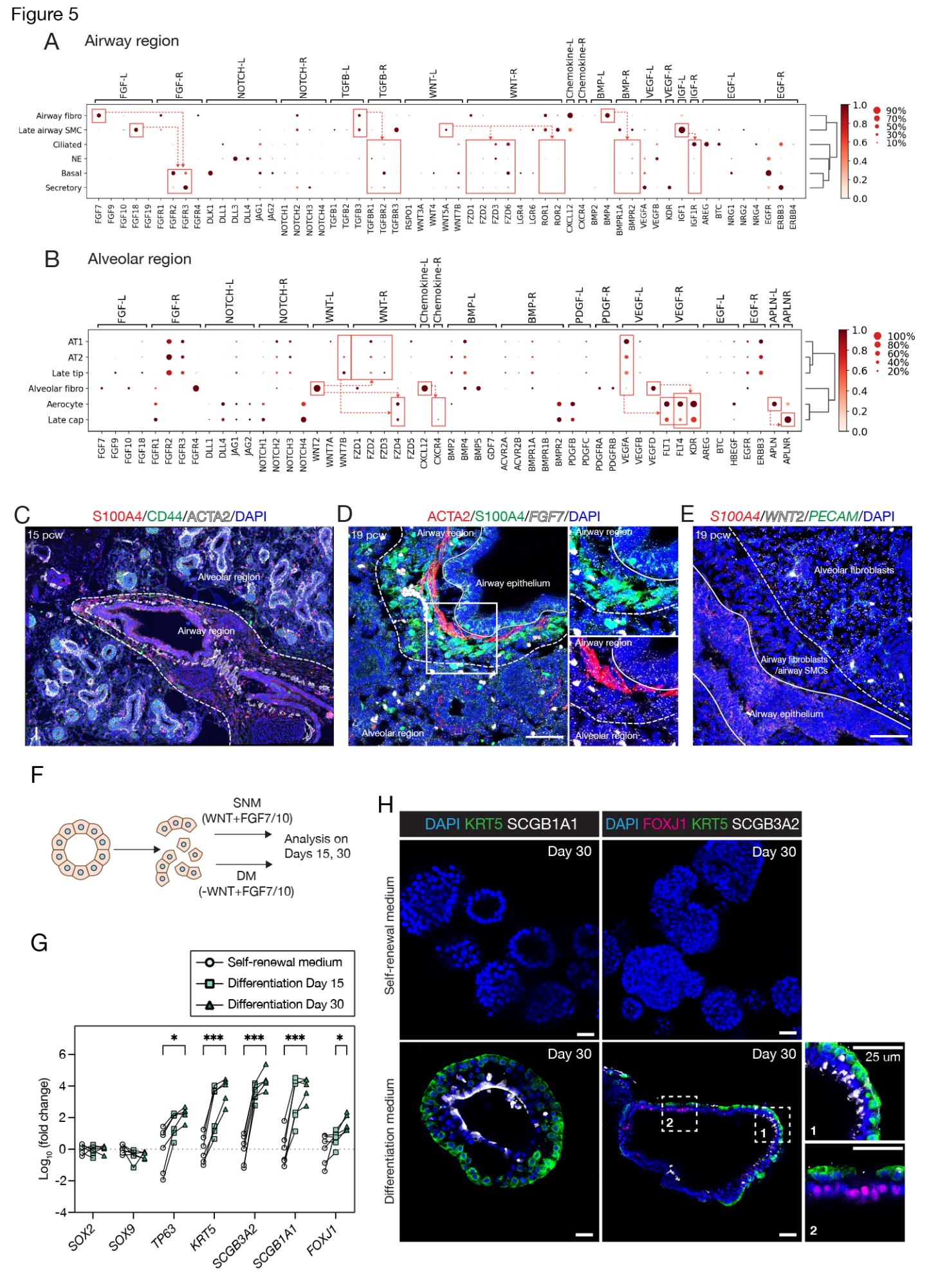

A
Airway region
FGF-L
FGF-R
NOTCH-L
NOTCH-R
TGFB-L
TGFB-R
WNT-L
WNT-R
Chemokine-L
Chemokine-R
BMP-L
BMP-R
VEGF-L
VEGF-R
IGF-L
IGF-R
EGF-L
EGF-R
Airway fibro
Late airway SMC
Ciliated
NE
Basal
Secretory
FGF7 FGF9 FGF10 FGF18 FGF19 FGFR1 FGFR2 FGFR3 FGFR4 DLK1 DLL1 DLL3 DLL4 JAG1 JAG2 NOTCH1 NOTCH2 NOTCH3 NOTCH4 TGFB1 TGFB2 TGFB3 TGFBR1 TGFBR2 TGFBR3 RSPO1 WNT3A WNT4 WNT5A WNT7B FZD1 FZD2 FZD3 FZD6 LGR4 LGR6 ROR1 ROR2 CXCL12 CXCR4 BMP2 BMP4 BMPR1A BMPR2 VEGFA VEGFB KDR IGF1 IGF1R AREG BTC NRG1 NRG2 NRG4 EGFR ERBB3 ERBB4
1.0 0.8 0.6 0.4 0.2 0.0
90% 70% 50% 30% 10%
B
Alveolar region
FGF-L
FGF-R
NOTCH-L
NOTCH-R
WNT-L
WNT-R
Chemokine-L
Chemokine-R
BMP-L
BMP-R
PDGF-L
PDGF-R
VEGF-L
VEGF-R
EGF-L
EGF-R
APLN-L
APLNR
AT1
AT2
Late tip
Alveolar fibro
Aerocyte
Late cap
FGF7 FGF9 FGF10 FGF18 FGFR1 FGFR2 FGFR3 FGFR4 DLL1 DLL4 JAG1 JAG2 NOTCH1 NOTCH2 NOTCH3 NOTCH4 WNT2 WNT7A WNT7B FZD1 FZD2 FZD3 FZD4 FZD5 CXCL12 CXCR4 BMP2 BMP4 BMP5 GDF7 ACVR2A ACVR2B BMPR1A BMPR1B BMPR2 PDGFB PDGFC PDGFRA PDGFRB VEGFA VEGFB VEGFD FLT1 FLT4 KDR AREG BTC HBEGF EGFR ERBB3 APLN APLNR
1.0 0.8 0.6 0.4 0.2 0.0
100% 80% 60% 40% 20%
C
S100A4/CD44/ACTA2/DAPI
15 pcw
Alveolar region
Airway region
D
ACTA2/S100A4/FGF7/DAPI
19 pcw
Airway region
Airway epithelium
Alveolar region
Airway region
Airway region
Alveolar region
E
S100A4/WNT2/PECAM/DAPI
19 pcw
Alveolar fibroblasts
Airway fibroblasts
/airway SMCs
Airway epithelium
F
SNM
(WNT+FGF7/10)
Analysis on
Days 15, 30
DM
(-WNT+FGF7/10)
G
Self-renewal medium
Differentiation Day 15
Differentiation Day 30
Log10 (fold change)
6
4
2
0
-2
-4
*
***
***
***
*
SOX2
SOX9
TP63
KRT5
SCGB3A2
SCGB1A1
FOXJ1
H
DAPI KRT5 SCGB1A1
DAPI FOXJ1 KRT5 SCGB3A2
Day 30
Day 30
Self-renewal medium
Day 30
Day 30
Differentiation medium
2
1
25 um
1
2

Figure 6

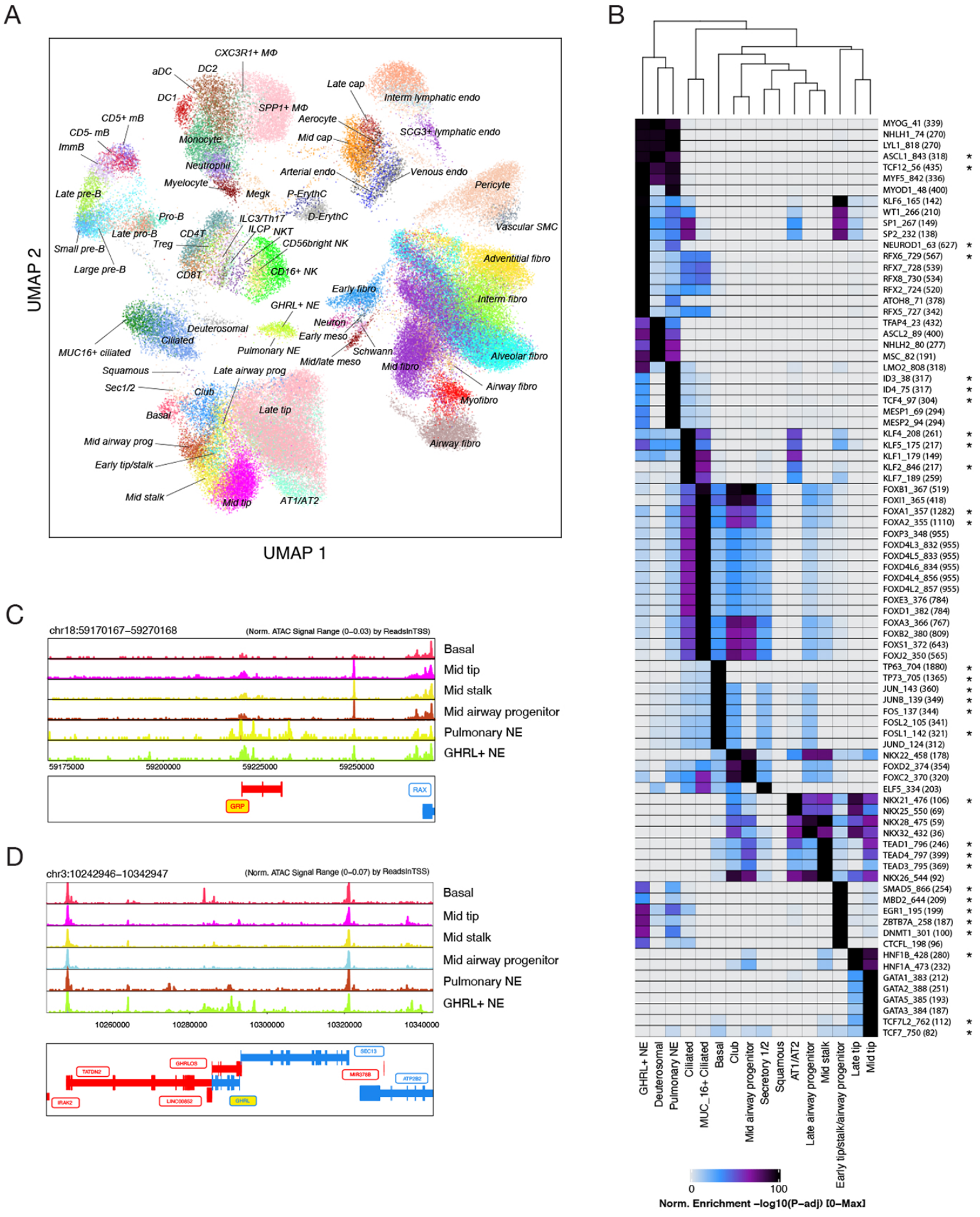

A
CXC3R1+ MΦ
aDC
DC2
DC1
SPP1+ MΦ
Late cap
Interm lymphatic endo
CD5+ mB
CD5- mB
ImmB
Monocyte
Aerocyte
Mid cap
SCG3+ lymphatic endo
Neutrophil
Arterial endo
Venous endo
Pericyte
Myelocyte
Megk
P-ErythC
Late pre-B
Pro-B
ILC3/Th17
D-ErythC
Vascular SMC
Late pro-B
CD4T
ILCP
NKT
Small pre-B
Treg
CD56bright NK
Adventitial fibro
Large pre-B
CD8T
CD16+ NK
Early fibro
Interm fibro
GHRL+ NE
Neuton
Deuterosomal
Ciliated
Early meso
MUC16+ ciliated
Pulmonary NE
Schwann
Mid/late meso
Alveolar fibro
Mid fibro
Squamous
Late airway prog
Airway fibro
Sec1/2
Club
Myofibro
Basal
Late tip
Mid airway prog
Airway fibro
Early tip/stalk
Mid stalk
Mid tip
AT1/AT2
UMAP 2
UMAP 1
B
MYOG_41 (339)
NHLH1_74 (270)
LYL1_818 (270)
ASCL1_843 (318) *
TCF12_56 (435) *
MYF5_842 (336)
MYOD1_48 (400)
KLF6_165 (142)
WT1_266 (210)
SP1_267 (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)
FOXB2_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_370 (320)
ELF5_334 (203)
NKX21_476 (106) *
NKX25_550 (69)
NKX28_475 (59)
NKX32_432 (36)
TEAD1_796 (246) *
TEAD4_797 (399) *
TEAD3_795 (369) *
NKX26_544 (92)
SMAD5_866 (254) *
MBD2_644 (209) *
EGR1_195 (199) *
ZBTB7A_258 (187) *
DNMT1_301 (100) *
CTCFL_198 (96)
HNF1B_428 (280) *
HNF1A_473 (232)
GATA1_383 (212)
GATA2_388 (251)
GATA5_385 (193)
GATA3_384 (187)
TCF7L2_762 (112) *
TCF7_750 (82) *
GHRL+ NE
Deuterosomal
Pulmonary NE
Ciliated
MUC_16+ Ciliated
Basal
Club
Mid airway progenitor
Secretory 1/2
Squamous
AT1/AT2
Late airway progenitor
Mid stalk
Early tip/stalk/airway progenitor
Late tip
Mid tip
0
100
Norm. Enrichment −log10(P−adj) [0−Max]
C
chr18:59170167–59270168
(Norm. ATAC Signal Range (0–0.03) by ReadsInTSS)
Basal
Mid tip
Mid stalk
Mid airway progenitor
Pulmonary NE
GHRL+ NE
59175000
59200000
59225000
59250000
GRP
RAX
D
chr3:10242946–10342947
(Norm. ATAC Signal Range (0–0.07) by ReadsInTSS)
Basal
Mid tip
Mid stalk
Mid airway progenitor
Pulmonary NE
GHRL+ NE
10260000
10280000
10300000
10320000
10340000
SEC13
GHRLOS
TATDN2
MIR378B
ATP2B2
IRAK2
LINC00852
GHRL

Figure 7

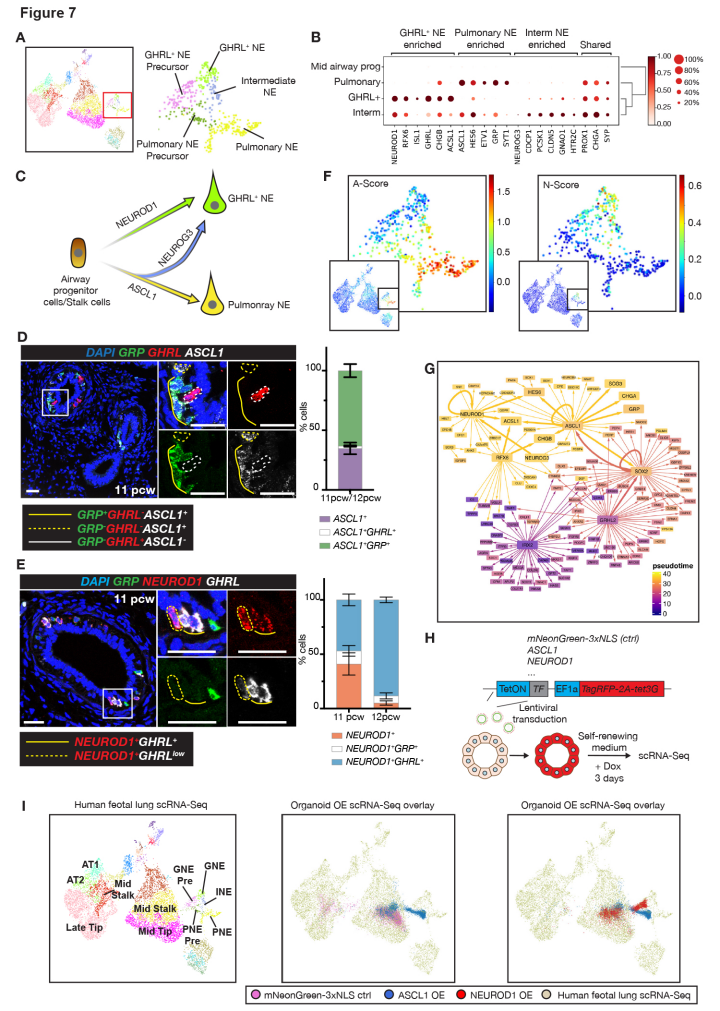

A
GHRL+ NE Precursor
GHRL+ NE
Intermediate NE
Pulmonary NE Precursor
Pulmonary NE
B
GHRL+ NE enriched
Pulmonary NE enriched
Interm NE enriched
Shared
Mid airway prog
Pulmonary
GHRL+
Interm
NEUROD1
NEUROG3
ISL1
GHRL
CHGB
ASCL1
ASCL1
HES6
ETV1
GRP
SYT1
NEUROD2
CDX2
PCSK1
CLDN5
SMAD1
GATA6
PROX1
CHGA
SYP
100%
80%
60%
40%
20%
1.00
0.75
0.50
0.25
0.00
C
NEUROD1
NEUROG3
ASCL1
GHRL+ NE
Airway progenitor cells/Stalk cells
Pulmonray NE
F
A-Score
1.5
1.0
0.5
0.0
N-Score
0.6
0.4
0.2
0.0
D
DAPI GRP GHRL ASCL1
11 pcw
GRP+GHRL-ASCL1+
GRP-GHRL-ASCL1+
GRP-GHRL+ASCL1-
100
50
0
% cells
11pcw
12pcw
ASCL1+
ASCL1+GHRL+
ASCL1+GRP+
E
DAPI GRP NEUROD1 GHRL
11 pcw
NEUROD1+GHRL+
NEUROD1+GHRLlow
100
50
0
% cells
11 pcw
12pcw
NEUROD1+
NEUROD1+GRP+
NEUROD1+GHRL+
G
NEUROD1
ASCL1
CHGB
RFX6
NEUROD2
HES6
SOX9
CHGA
GRP
SOX2
SYT1
pseudotime
40
30
20
10
0
H
mNeonGreen-3xNLS (ctrl)
ASCL1
NEUROD1
...
TetON
TF
EF1a
TagRFP-2A-tet3G
Lentiviral transduction
Self-renewing medium
+ Dox
3 days
scRNA-Seq
I
Human feotal lung scRNA-Seq
AT1
AT2
Mid Stalk
GNE Pre
GNE
INE
Late Tip
Mid Stalk
Mid Tip
PNE Pre
PNE
Organoid OE scRNA-Seq overlay
Organoid OE scRNA-Seq overlay
mNeonGreen-3xNLS ctrl
ASCL1 OE
NEUROD1 OE
Human feotal lung scRNA-Seq

Figure S1

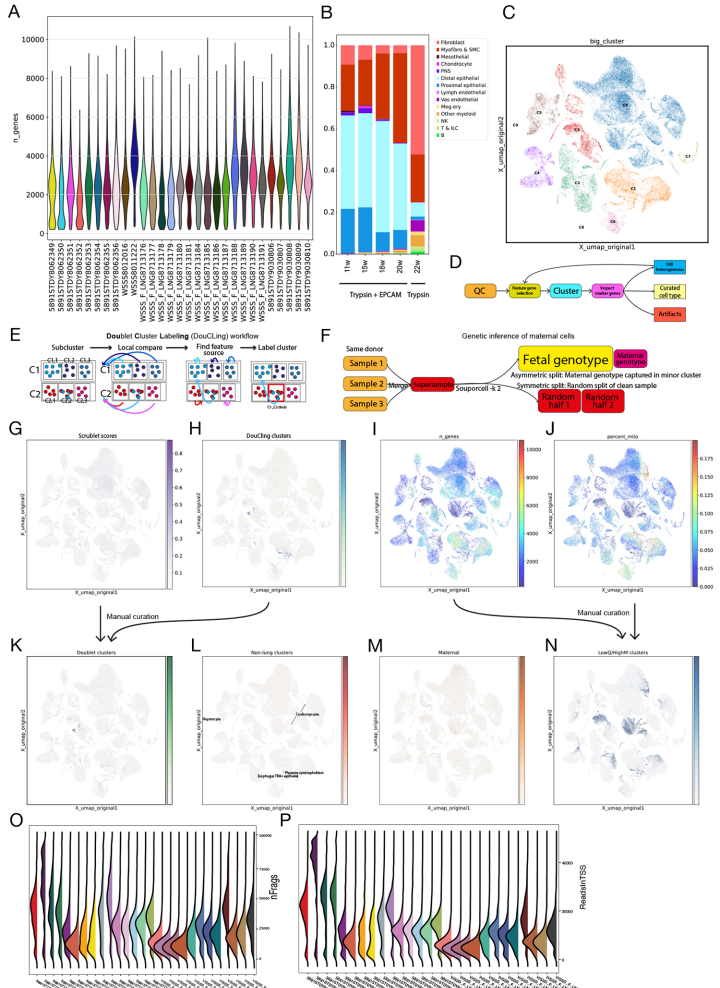

A
n_genes
B
Fibroblast
Myofibro & SMC
Mesothelial
Chondrocyte
PNS
Distal epithelial
Proximal epithelial
Lymph endothelial
Vas endothelial
Meg-ery
Other myeloid
NK
T & ILC
B
Trypsin + EPCAM
Trypsin
C
big_cluster
X_umap_original1
X_umap_original2
D
QC
Feature gene selection
Cluster
Inspect marker genes
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
C1.2 doub
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
Myoscyte
Cardiomyocyte
Esophagus Non-epithelial
Plasma cytomyelodium
M
Maternal
N
LowQ/HighM clusters
O
nFrags
P
ReadsInTSS

Figure S2

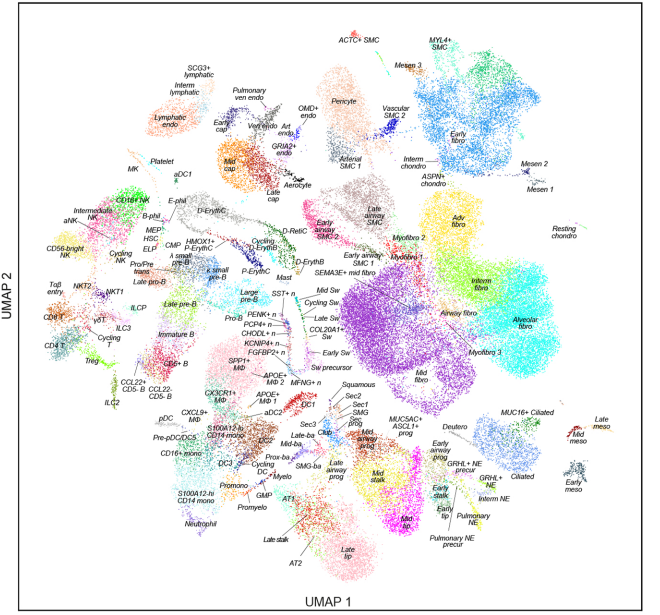

A
UMAP 1
UMAP 2
ACTC+ SMC
MYL4+ SMC
Mesen 3
SCG3+ lymphatic
Interm lymphatic
Pulmonary ven endo
Pericyte
OMD+ endo
Vascular SMC 2
Lymphatic endo
Early cap
Ven endo
Art endo
Early fibro
GRIA2+ endo
Mid cap
Arterial SMC 1
Interm chondro
Mesen 2
Platelet
MK
aDC1
Aerocyte
Late cap
ASPN+ chondro
Mesen 1
CD16+ NK
E-phal
Late airway SMC
Adv fibro
Intermediate aNK
B-phal
DI-ErythC
Early airway SMC 2
Resting chondro
MEP
HSC
ELP
D-ReloC
Myofibro 2
CD56-bright NK
CMP
HMOX1+ P-ErythC
Cycling Erythro
Cycling NK
Pro/Pre trans
x small pre-B
Early airway SMC 1
Myofibro 1
D-ErythB
SEMA3E+ mid fibro
Late pro-B
P-ErythC
Mast
Interm fibro
Tcd8 entry
NKT2
NKT1
Late pre-B
Large pre-B
SST+ n
Mid Sw
Cycling Sw
ILCP
Airway fibro
CD8 T
γδT
ILC3
Pro-B
PENK+ n
PCP4+ n
Late Sw
COL20A1+ Sw
Alveolar fibro
Cycling T
CD4 T
Immature B
CHODL+ n
KCNIP4+ n
FGFBP2+ n
Early Sw
Myofibro 3
Treg
GPR+ B
SPP1+ MØ
Sw precursor
APOE+ MØ 2
MFNG+ n
Squamous
Mid fibro
ILC2
CCL22+ CD5- B
CCL22- CD5- B
CX3CR1+ MØ
APOE+ MØ 1
DC1
Sec2
Sec1
SMG prog
MUC16+ Ciliated
Late meso
pDC
CXCL9+ MØ
aDC2
Sec3
MUC5AC+ ASCL1+ prog
Deutero
Mid meso
Pre-pDC/DC5
S100A12-hi CD14-mono
Clip
Late-ba
Mid airway prog
Early airway prog
CD16+ mono
Mid-ba
GRNL+ NE precur
Proc-ba
Cycling DC
SMG-ba
Late airway prog
Ciliated
Early meso
DC3
Myelo
Mid stalk
GRHL+ NE
S100A12-hi CD14 mono
Promono
Early stalk
Interm NE
AT1
Promyelo
Early tip
Pulmonary NE
Neutrophil
Mid tip
Late stalk
Pulmonary NE precur
Late tip
AT2
B
stage
C
phase
G1
G2M
S
D
dissection
Distal
Proximal
Whole-lung
E
chemistry
Normal
Trypsin+
Trypsin+CD326
F
SST+ n
PCP4+ n
TM4SF4+ PENK+ n
KCNIP4+ n
FGFBP2+ n
TM4SF4+ CHODL+ n
Mid airway
Late airway
COL20A1+ Sw
MFNG+ CD200+
Early Schwann
Schwann precursor
Proliferating Schwann
UMAP1
UMAP2
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
NKX6-1
SST
GRP
PCP4
PYDC
PENK
DLK1
TMEM176A
CXCL12
GPM6B
MPZL1
APOA1
TYMS
COL20A1
CDH19
ALDH1A1
MFAP5
HMGN2
CTRB1
1.0
0.8
0.6
0.4
0.2
0.0
H
Proximal basal
Secretory 3
Mid basal
Pulmonary venous endo
Late Schwann
Late basal
MUC16+ ciliated
Airway fibro
SMG basal
Effect size (Proximal enrichment)
Benjamini-Hochberg corrected p-values
0.05
3
1.0
1000
100
10
1
0.1
1E-33
1E-30
1E-27
1E-24
1E-21
1E-18
1E-15
1E-12
1E-9
1E-6
1E-3
1

Figure S3

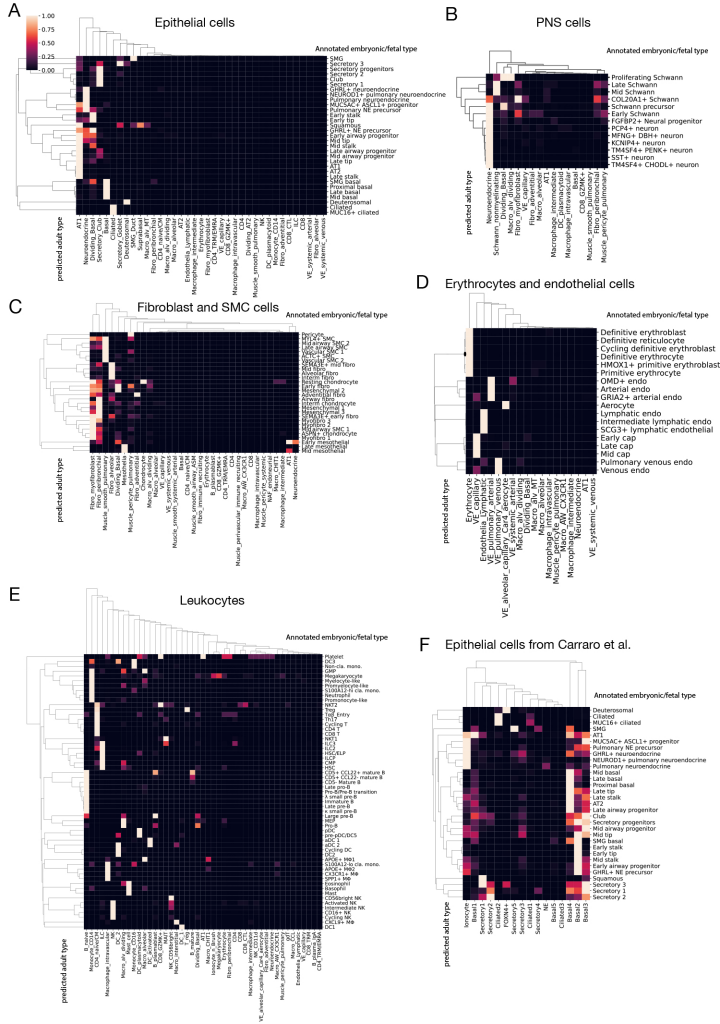
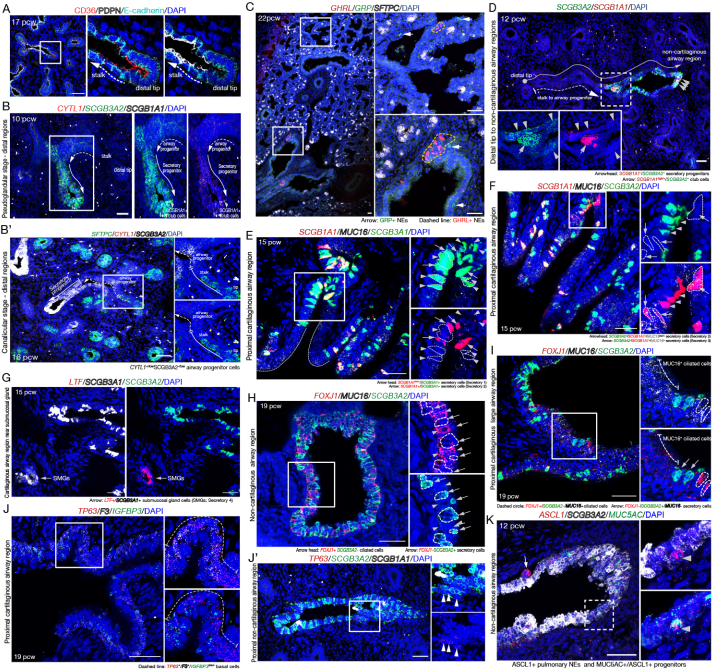
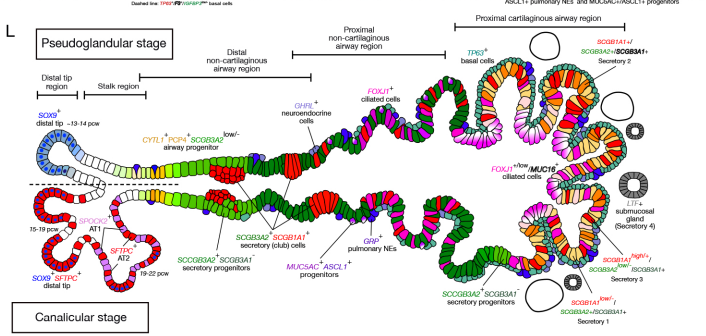

A
Epithelial cells
Annotated embryonic/fetal type
predicted adult type
B
PNS cells
Annotated embryonic/fetal type
Proliferating Schwann
Late Schwann
Mid Schwann
COL20A1+ Schwann
Schwann precursor
Early Schwann
FGFBP2+ Neural progenitor
PCP4+ neuron
NPNT+ DBH+ neuron
KCNIP4+ neuron
TMM5F4+ PENK+ neuron
SST+ neuron
TMM5F4+ CHODL+ neuron
predicted adult type
C
Fibroblast and SMC cells
Annotated embryonic/fetal type
predicted adult type
D
Erythrocytes and endothelial cells
Annotated embryonic/fetal type
Definitive erythroblast
Definitive reticulocyte
Cycling definitive erythroblast
Definitive erythrocyte
HMOX1+ primitive erythroblast
Primitive erythrocyte
OMD+ endo
Arterial endo
GRIA2+ arterial endo
Aerocyte
Lymphatic endo
Intermediate lymphatic endo
SCG3+ lymphatic endothelial
Early cap
Late cap
Mid cap
Pulmonary venous endo
Venous endo
predicted adult type
E
Leukocytes
Annotated embryonic/fetal type
predicted adult type
F
Epithelial cells from Carraro et al.
Annotated embryonic/fetal type
Deuterosomal
Ciliated
MUC16+ ciliated
SMG
AT2
MUC5AC+ ASCL1+ progenitor
Pulmonary NE precursor
GHRL+ neuroendocrine
NEUROD1+ pulmonary neuroendocrine
Pulmonary neuroendocrine
Mid basal
Late basal
Proximal basal
Late tip
Late stalk
AT1
Late airway progenitor
Club
Secretory progenitors
Mid airway progenitor
Mid tip
SMG basal
Early stalk
Early tip
Mid stalk
Early airway progenitor
GHRL+ NE precursor
Squamous
Secretory 3
Secretory 1
Secretory 2
predicted adult type

Figure S4



A
CD36/PDPN/E-cadherin/DAPI
17 pcw
stalk
distal tip
B
CYTL1/SCGB3A2/SCGB1A1/DAPI
10 pcw
Pseudoglandular stage - distal regions
Stalk
distal tip
Airway progenitor
Secretory progenitor
SCGB3A2+ Nestin-Ki-Ki-Ki-Ki
SCGB1A1+ Nestin-Ki-Ki-Ki-Ki
B'
SFTPC/CYTL1/SCGB3A2/DAPI
18 pcw
Canalicular stage - distal regions
Stalk
distal tip
Airway progenitor
Secretory progenitor
CYTL1+/SCGB3A2+/Ki-Ki-Ki-Ki-Ki airway progenitor cells
C
GHRL/GRP/SFTPC/DAPI
22pcw
Arrow: GRP+ NEs
Dashed line: GHRL+ NEs
D
SCGB3A2/SCGB1A1/DAPI
12 pcw
Distal tip to non-cartilaginous airway regions
non-cartilaginous airway region
distal tip
stalk to airway progenitor
Arrowhead: SCGB1A1+/SCGB3A2+ secretory progenitors
Arrow: SCGB1A1+/SCGB3A2+ club cells
E
SCGB1A1/MUC16/SCGB3A1/DAPI
15 pcw
Proximal cartilaginous airway region
Arrow head: SCGB1A1+/SCGB3A1+ secretory cells (Secretory 2)
Arrow: SCGB1A1+/SCGB3A1+ secretory cells (Secretory 2)
F
SCGB1A1/MUC16/SCGB3A2/DAPI
15 pcw
Proximal cartilaginous airway region
Arrowhead: SCGB1A1+/SCGB3A2+ secretory cells (Secretory 3)
Arrow: SCGB1A1+/SCGB3A2+ secretory cells (Secretory 3)
G
LTF/SCGB3A1/SCGB3A2/DAPI
16 pcw
Cartilaginous airway region near submucosal gland
SMGs
Arrow: LTF+/SCGB3A1+ submucosal gland cells (SMGs, Secretory 4)
H
FOXJ1/MUC16/SCGB3A2/DAPI
19 pcw
Non-cartilaginous airway region
Arrow head: FOXJ1+/SCGB3A2+ ciliated cells
Arrow: FOXJ1+/SCGB3A2+ secretory cells
I
FOXJ1/MUC16/SCGB3A2/DAPI
19 pcw
Proximal cartilaginous large airway region
MUC16+ ciliated cells
MUC16+ ciliated cells
Dashed circle: FOXJ1+/SCGB3A2+/MUC16+ ciliated cells
Arrow: FOXJ1+/SCGB3A2+/MUC16- secretory cells
J
TP63/F3/IGFBP3/DAPI
19 pcw
Proximal cartilaginous airway region
Dashed line: TP63+/F3+/IGFBP3+/Ki-Ki-Ki-Ki-Ki basal cells
J'
TP63/SCGB3A2/SCGB1A1/DAPI
Proximal non-cartilaginous airway region
K
ASCL1/SCGB3A2/MUC5AC/DAPI
12 pcw
Non-cartilaginous airway regions
ASCL1+ pulmonary NEs and MUC5AC+/ASCL1+ progenitors



L
Pseudoglandular stage
Proximal cartilaginous airway region
Proximal non-cartilaginous airway region
Distal non-cartilaginous airway region
Distal tip region
Stalk region
SOX9+ distal tip ~13-14 pcw
CYTL1+ PCP4+ SCGB3A2low/- airway progenitor
GHRL+ neuroendocrine cells
FOXJ1+ ciliated cells
TP63+ basal cells
SCGB1A1+/SCGB3A2+/SCGB3A1+ Secretory 2
FOXJ1+/low/MUC16+ ciliated cells
LTF+ submucosal gland (Secretory 4)
SCGB3A2+ SCGB1A1+ secretory (club) cells
GRP+ pulmonary NEs
SCCGB3A2+ SCGB3A1- secretory progenitors
MUC5AC+ ASCL1+ progenitors
SCCGB3A2+ SCGB3A1- secretory progenitors
SCGB1A1low/-/SCGB3A2+/SCGB3A1+ Secretory 3
SCGB1A1low/-/SCGB3A2+/SCGB3A1+ Secretory 1
SPOCK2+ AT1
15-19 pcw
SFTPC+ AT2
19-22 pcw
SOX9+ SFTPC+ distal tip
Canalicular stage

Figure S5

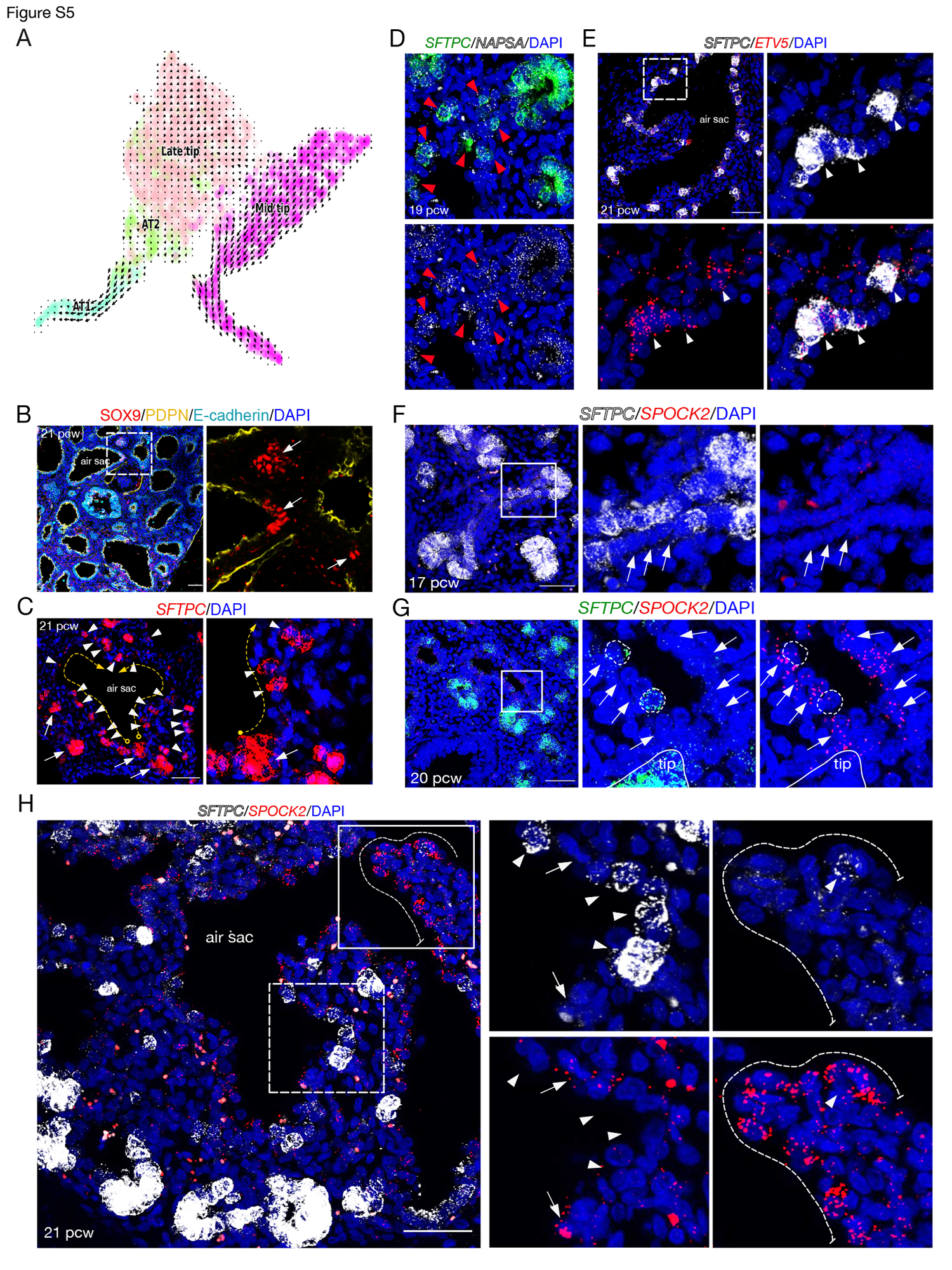

A
Late tip
Mid tip
AT2
AT1
D
SFTPC/NAPSA/DAPI
19 pcw
E
SFTPC/ETV5/DAPI
air sac
21 pcw
B
SOX9/PDPN/E-cadherin/DAPI
21 pcw
air sac
F
SFTPC/SPOCK2/DAPI
17 pcw
C
SFTPC/DAPI
21 pcw
air sac
G
SFTPC/SPOCK2/DAPI
20 pcw
tip
tip
H
SFTPC/SPOCK2/DAPI
air sac
21 pcw

Figure S6

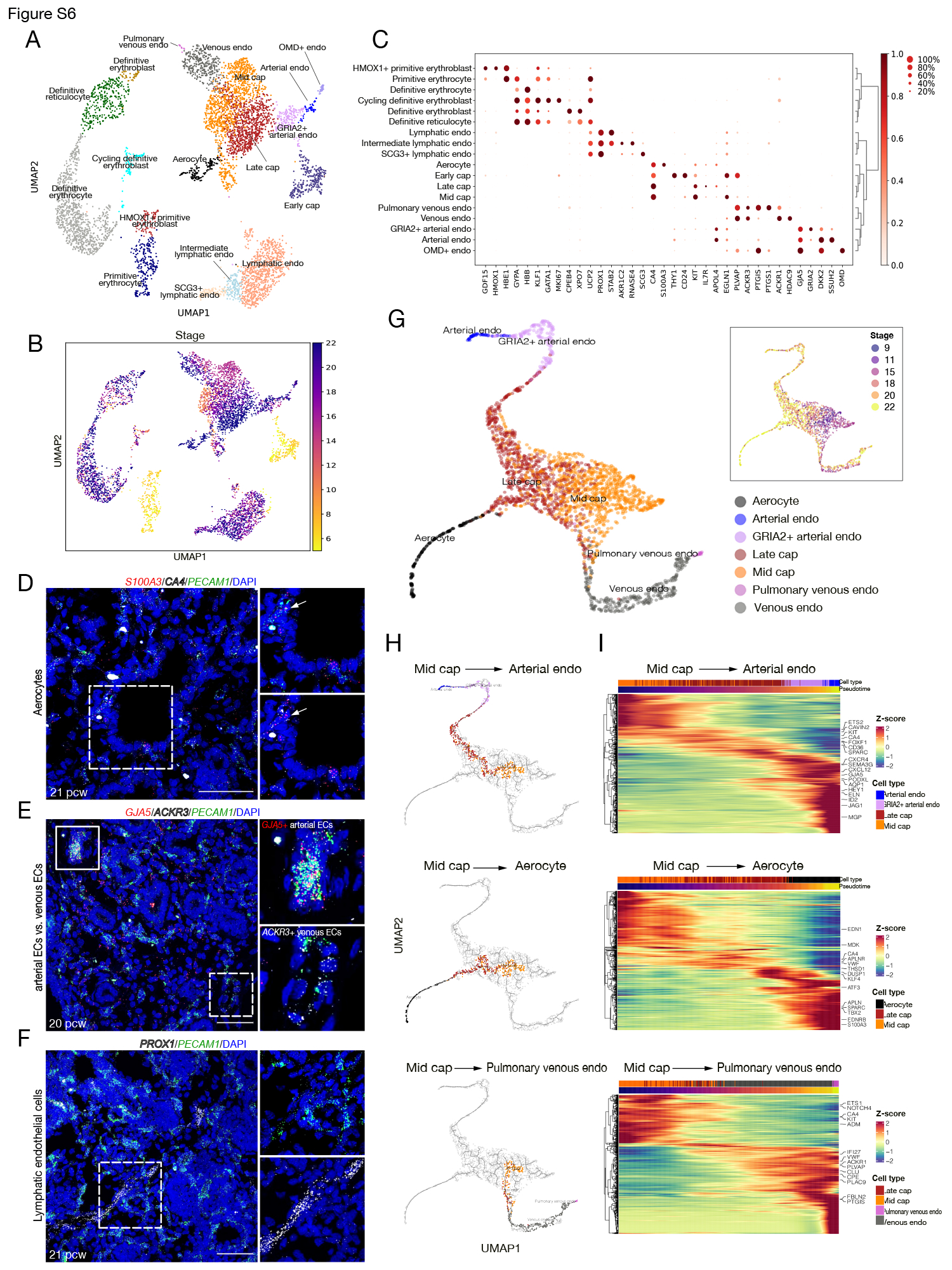

A
Pulmonary venous endo
Venous endo
OMD+ endo
Definitive erythroblast
Arterial endo
Mid cap
Definitive reticulocyte
GRIA2+ arterial endo
Cycling definitive erythroblast
Aerocyte
Late cap
UMAP2
Definitive erythrocyte
Early cap
HMOX1+ primitive erythroblast
Intermediate lymphatic endo
Lymphatic endo
Primitive erythrocyte
SCG3+ lymphatic endo
UMAP1
B
Stage
UMAP2
UMAP1
C
HMOX1+ primitive erythroblast
Primitive erythrocyte
Definitive erythrocyte
Cycling definitive erythroblast
Definitive erythroblast
Definitive reticulocyte
Lymphatic endo
Intermediate lymphatic endo
SCG3+ lymphatic endo
Aerocyte
Early cap
Late cap
Mid cap
Pulmonary venous endo
Venous endo
GRIA2+ arterial endo
Arterial endo
OMD+ endo
GDF15 HMOX1 HBE1 GYPA HBB KLF1 GATA1 MKI67 CPEB4 XPO7 UCP2 PROX1 STAB2 AKR1C2 RNASE4 SCG3 CA4 S100A3 THY1 CD24 KIT IL7R APOL4 EGLN1 PLVAP ACKR3 PTGIS PTGS1 ACKR1 HDAC9 GJA5 GRIA2 DKK2 SSUH2 OMD
100%
80%
60%
40%
20%
D
S100A3/CA4/PECAM1/DAPI
Aerocytes
21 pcw
E
GJA5/ACKR3/PECAM1/DAPI
GJA5+ arterial ECs
ACKR3+ venous ECs
arterial ECs vs. venous ECs
20 pcw
F
PROX1/PECAM1/DAPI
Lymphatic endothelial cells
21 pcw
G
Arterial endo
GRIA2+ arterial endo
Late cap
Mid cap
Aerocyte
Pulmonary venous endo
Venous endo
Stage
9
11
15
18
20
22
Aerocyte
Arterial endo
GRIA2+ arterial endo
Late cap
Mid cap
Pulmonary venous endo
Venous endo
H
Mid cap → Arterial endo
Mid cap → Aerocyte
Mid cap → Pulmonary venous endo
UMAP2
UMAP1
I
Mid cap → Arterial endo
Cell type
Pseudotime
Z-score
ETS2
CAVIN2
KIT
CA4
FOXF1
CD36
SPARC
CXCR4
SEMA3G
CXCL12
GJA5
PODXL
AQP1
HEY1
ELN
ID2
JAG1
MGP
Cell type
Arterial endo
GRIA2+ arterial endo
Late cap
Mid cap
Mid cap → Aerocyte
EDN1
MDK
CA4
APLNR
VWF
THSD1
DUSP1
KLF4
ATF3
APLN
SPARC
TBX2
EDNRB
S100A3
Cell type
Aerocyte
Late cap
Mid cap
Mid cap → Pulmonary venous endo
ETS1
NOTCH4
CA4
KIT
ADM
IFI27
VWF
ACKR1
PLVAP
CLU
CPE
PLAC9
FBLN2
PTGIS
Cell type
Late cap
Mid cap
Pulmonary venous endo
Venous endo

Figure S7

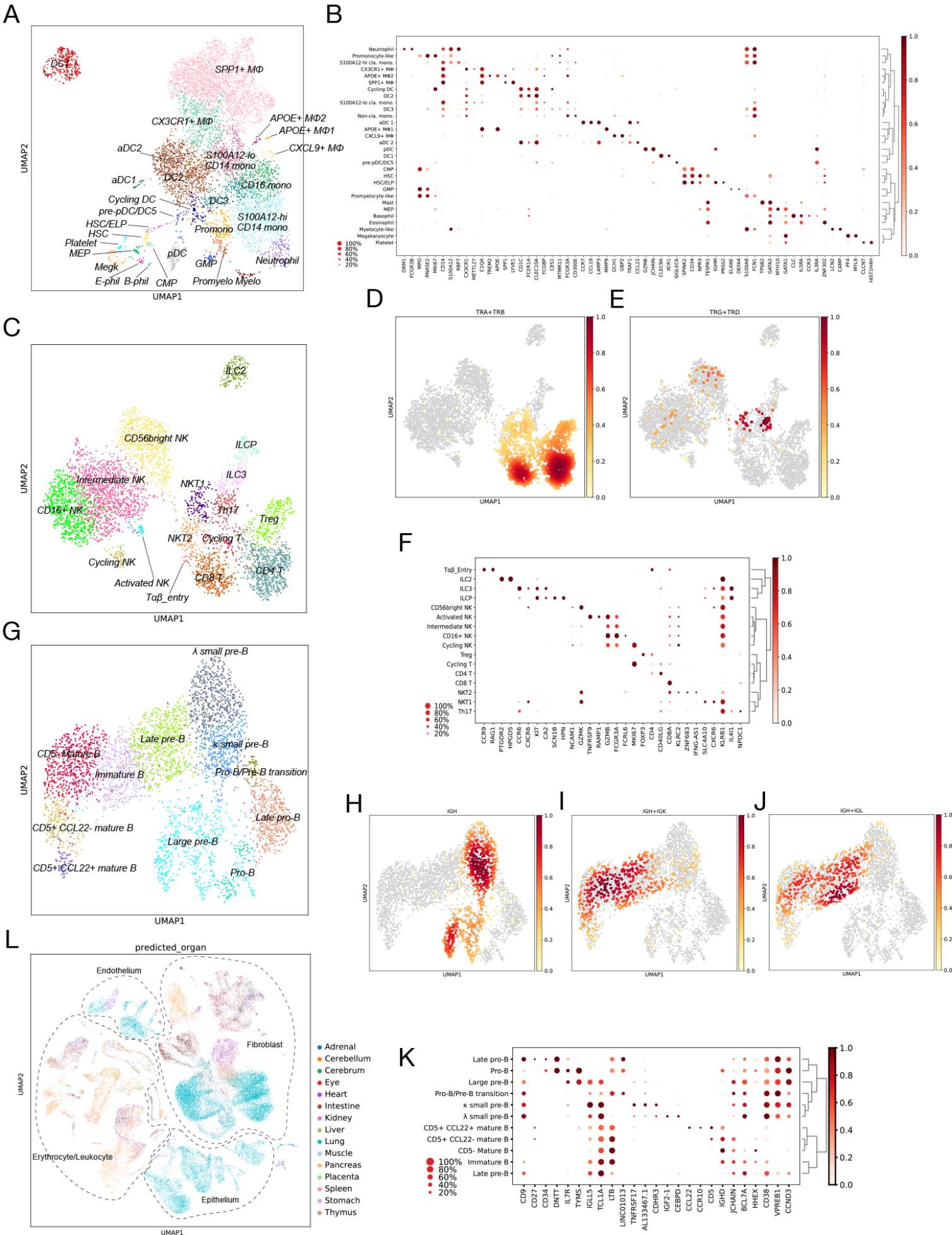
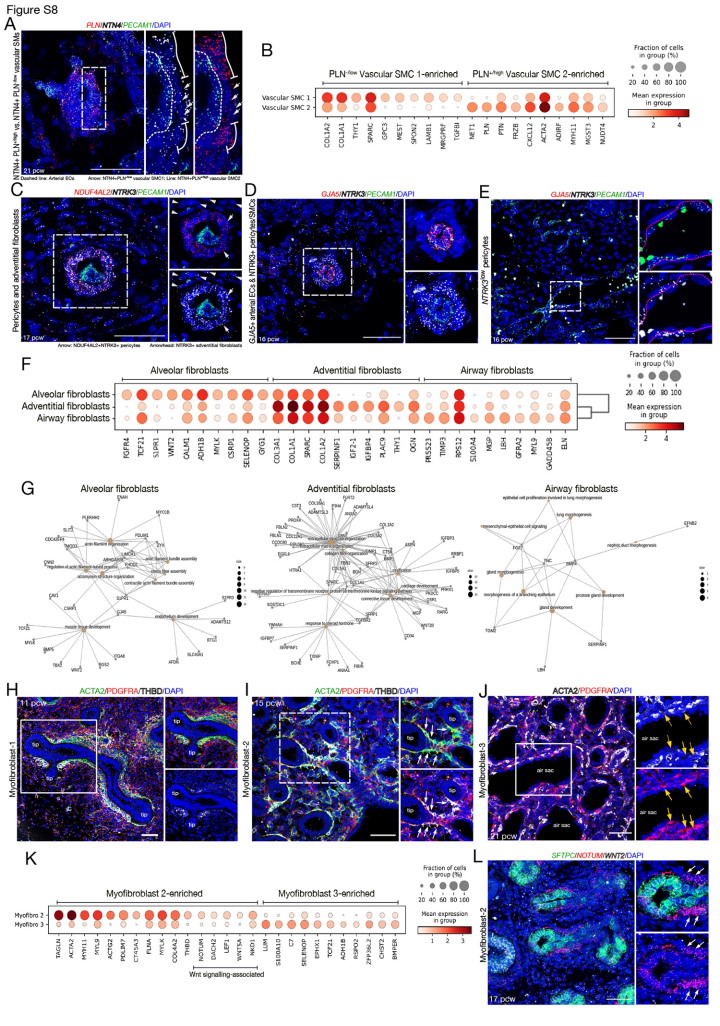

A
DC1
SPP1+ MΦ
CX3CR1+ MΦ
APOE+ MΦ2
APOE+ MΦ1
CXCL9+ MΦ
aDC2
S100A12-lo CD14 mono
aDC1
DC2
CD16 mono
Cycling DC
DC3
pre-pDC/DC5
S100A12-hi CD14 mono
HSC/ELP
Promono
HSC
Platelet
MEP
pDC
GMP
Neutrophil
Megk
E-phil B-phil
CMP
Promyelo Myelo
UMAP2
UMAP1
B
Neutrophil
Promonocyte-like
S100A12-hi cla. mono.
CX3CR1+ MΦ
APOE+ MΦ2
SPP1+ MΦ
Cycling DC
DC2
S100A12-lo cla. mono.
DC3
Non-cla. mono.
aDC 1
APOE+ MΦ1
CXCL9+ MΦ
aDC 2
pDC
DC1
pre-pDC/DC5
CMP
HSC
HSC/ELP
GMP
Promyelocyte-like
Mast
MEP
Basophil
Eosinophil
Myelocyte-like
Megakaryocyte
Platelet
100%
80%
60%
40%
20%
C
ILC2
CD56bright NK
ILCP
Intermediate NK
ILC3
NKT1
CD16+ NK
Th17
Treg
NKT2
Cycling T
Cycling NK
CD4 T
Activated NK
CD8 T
Tαβ_entry
UMAP2
UMAP1
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
UMAP2
UMAP1
H
IGH
I
IGH+IGK
J
IGH+IGL
L
predicted_organ
Endothelium
Fibroblast
Erythrocyte/Leukocyte
Epithelium
UMAP2
UMAP1
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
100%
80%
60%
40%
20%

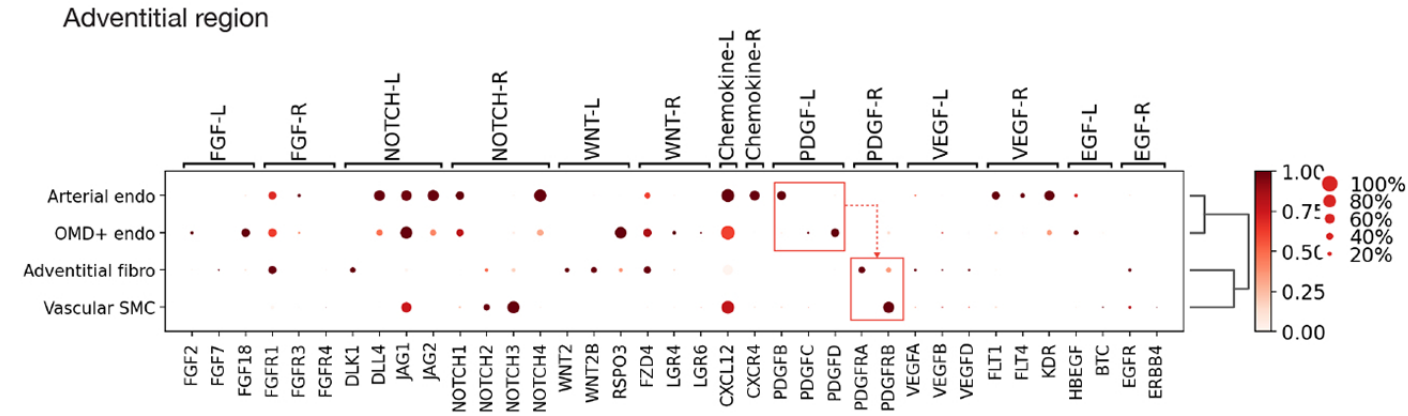
Figure S8



A
PLN/NTN4/PECAM1/DAPI
NTN4+ PLN^high vs. NTN4+ PLN^low 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
THY1
SPARC
GPC3
MEST
SPON2
LAMB1
MRGPRF
TGFBI
NET1
PLN
PTN
FRZB
CXCL12
ACTA2
ADIRF
MYH11
MGST3
NUDT4
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
2 4
C
NDUF4/AL2/NTRK3/PECAM1/DAPI
Pericytes and adventitial fibroblasts
17 pcw
Arrow: NDUF4/AL2/NTRK3+ pericytes
Arrowhead: NTRK3+ adventitial fibroblasts
D
GJA5/NTRK3/PECAM1/DAPI
GJA5+ arterial ECs & NTRK3+ pericytes/SMCs
16 pcw
E
GJA5/NTRK3/PECAM1/DAPI
NTRK3^low pericytes
16 pcw
F
Alveolar fibroblasts
Adventitial fibroblasts
Airway fibroblasts
Alveolar fibroblasts
Adventitial fibroblasts
Airway fibroblasts
FGFR4
TCF21
S1PR1
WNT2
CALM1
ADH1B
MYLK
CSRP1
SELENOP
GYG1
COL3A1
COL1A1
SPARC
COL1A2
SERPINF1
EGF2-1
IGFBP4
PLAC9
THY1
OGN
PRSS23
TIMP3
SPOX2
S100A4
MGP
LBH
GFRA2
MYL9
GAD0455
ELN
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
2 4
G
Alveolar fibroblasts
Adventitial fibroblasts
Airway fibroblasts
H
ACTA2/PDGFRA/THBD/DAPI
Myofibroblast-1
11 pcw
tip
I
ACTA2/PDGFRA/THBD/DAPI
Myofibroblast-2
15 pcw
J
ACTA2/PDGFRA/DAPI
Myofibroblast-3
21 pcw
air sac
K
Myofibroblast 2-enriched
Myofibroblast 3-enriched
Myofibro 2
Myofibro 3
TAGLN
ACTA2
MYH11
MYL9
ACTG2
PDLIM7
CTKSA3
FLNA
MYLK
COL4A2
THBD
NOTUM
DACH2
LEF1
WNT5A
NKD1
LUM
S100A10
C7
SELENOP
EPHX1
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
L
SFTPC/NOTUM/WNT2/DAPI
Myofibroblast-2
17 pcw

Figure S9

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
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
1.00
0.75
0.50
0.25
0.00
100%
80%
60%
40%
20%

B

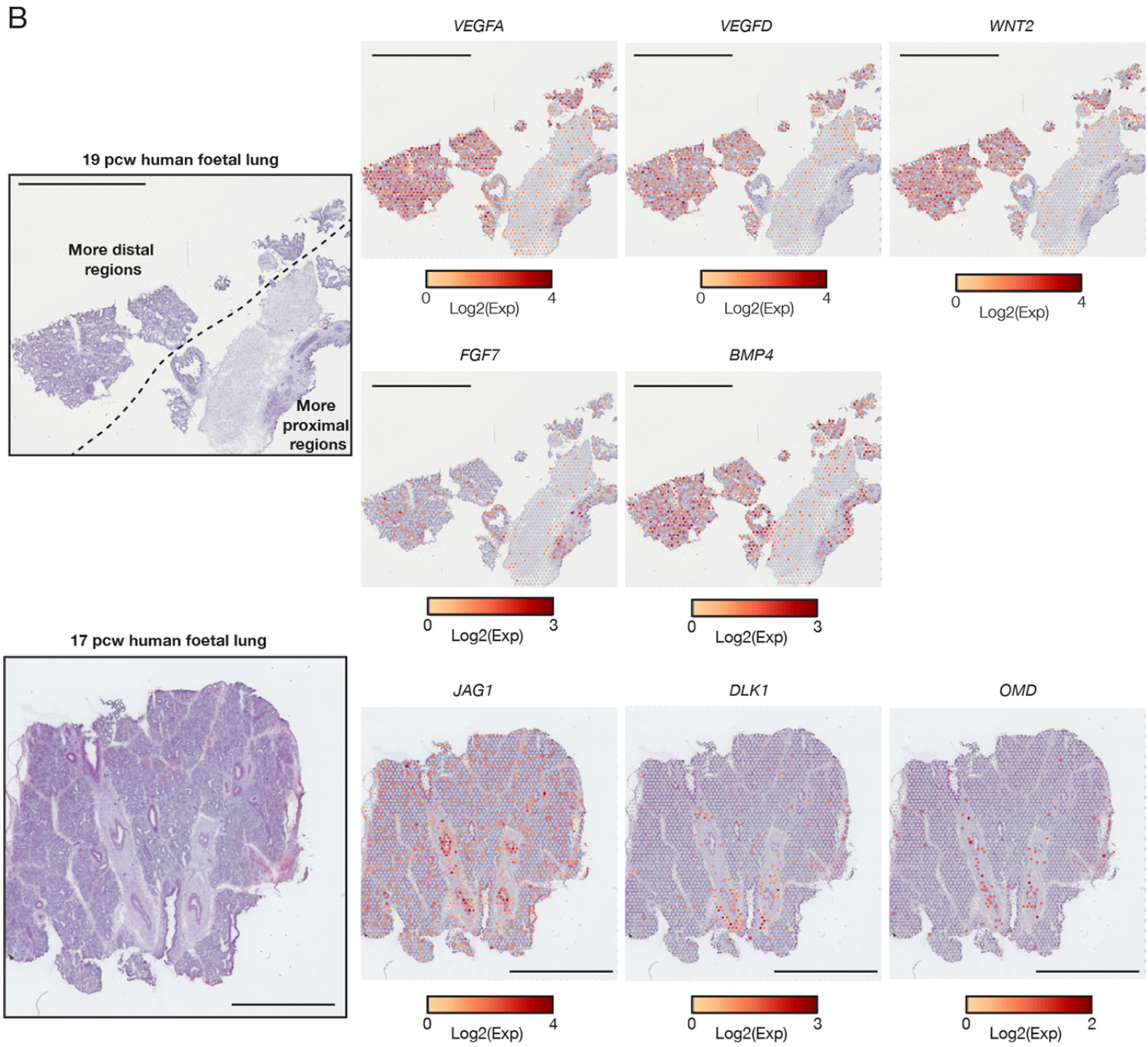

19 pcw human foetal lung
More distal regions
More proximal regions
VEGFA
VEGFD
WNT2
FGF7
BMP4
JAG1
DLK1
OMD
Log2(Exp)
0
4
3
2
17 pcw human foetal lung

Figure S10

A

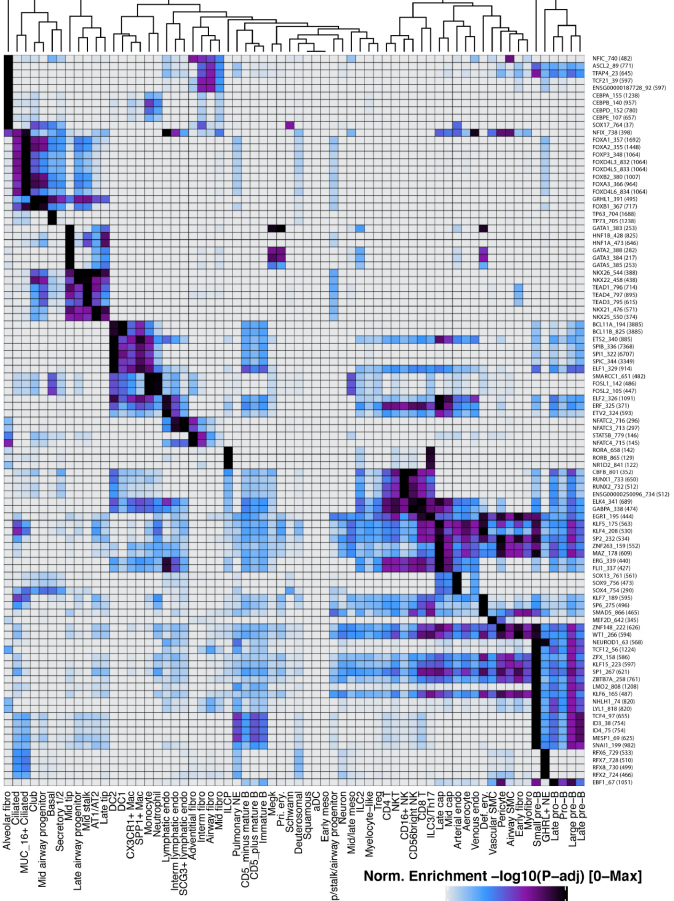
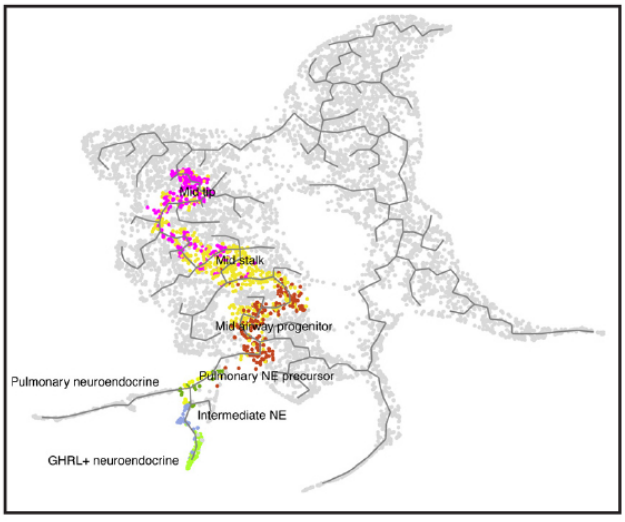
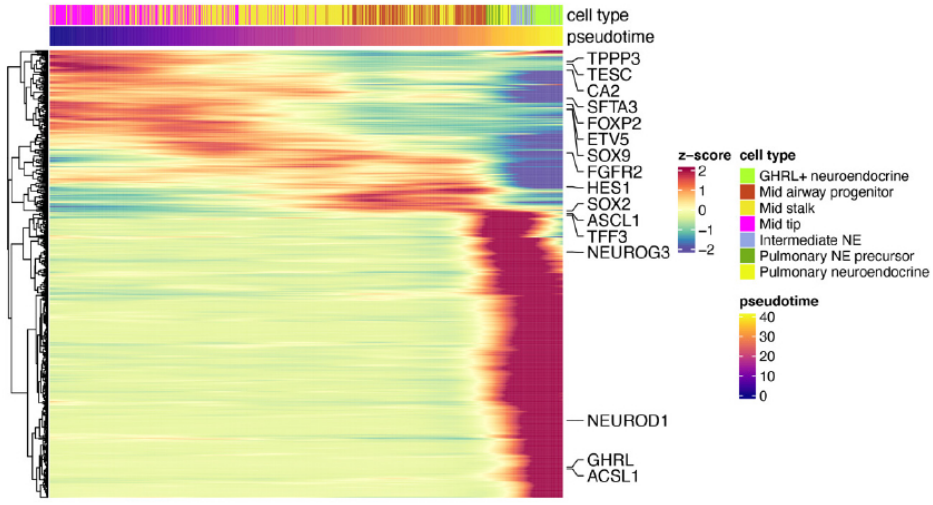

NFIC_790 (482)
ASCL2_89 (771)
TBX21_23 (645)
TCF21_39 (597)
ENSG00000187728_92 (597)
CEBPA_155 (1238)
CEBPB_140 (937)
CEBPD_152 (780)
CEBPE_107 (657)
SOX17_764 (37)
NFIX_738 (788)
FOXA1_357 (1692)
FOXA2_355 (1448)
FOXP3_348 (1064)
FOXD4L1_832 (1064)
FOXD4L5_833 (1064)
FOXB2_380 (1037)
FOXA3_366 (904)
FOXD4L6_834 (1064)
GRHL1_391 (495)
FOXB1_367 (717)
TP63_764 (1688)
TP73_795 (1338)
GATA1_383 (253)
HNF1B_428 (825)
HNF1A_473 (646)
GATA2_388 (282)
GATA3_384 (217)
GATA5_385 (253)
NKX28_544 (588)
NKX22_458 (438)
TEAD1_796 (714)
TEAD4_797 (893)
TEAD3_795 (613)
NKX21_476 (571)
NKX25_550 (374)
BCL11A_194 (3885)
BCL11B_825 (3885)
ETS2_340 (885)
SPIB_336 (7368)
SPI1_322 (6707)
SPIC_344 (3349)
ELF1_329 (914)
SMARCC1_651 (482)
FOSL1_142 (986)
FOSL2_165 (447)
ELF3_326 (1091)
ERF_325 (371)
ETV2_324 (993)
NFATC2_716 (296)
NFATC3_719 (297)
STAT5B_779 (146)
NFATC4_715 (145)
RORA_658 (142)
RORB_865 (129)
NR1D2_841 (122)
CBFB_901 (352)
RUNX1_723 (650)
RUNX2_732 (512)
ENSG00000250096_734 (512)
ELK4_341 (989)
GABPB_338 (474)
EGR1_195 (444)
KLF5_175 (963)
KLF4_208 (530)
SP2_232 (534)
ZNF263_159 (552)
MAZ_178 (609)
ERG_339 (440)
FLI1_337 (421)
SOX13_761 (561)
SOX9_756 (473)
SOX4_754 (390)
KLF7_189 (595)
SP6_275 (484)
SMAD5_866 (465)
MEF2D_642 (345)
ZNF148_222 (626)
WT1_266 (594)
NEUROD1_63 (560)
TCF12_56 (1224)
ZFX_158 (586)
KLF15_223 (597)
SP1_267 (621)
ZBTB7A_258 (761)
LMO2_808 (1208)
KLF6_165 (467)
NHLH1_74 (820)
LYL1_818 (820)
TCF4_97 (655)
ID3_38 (754)
ID4_75 (754)
MESP1_86 (625)
SNAI1_599 (982)
RFX6_729 (533)
RFX7_728 (510)
RFX8_730 (499)
RFX2_724 (466)
EBF1_67 (1051)
Alveolar fibro
Ciliated
Club
Mid airway progenitor
Secretory 1/2
Late airway progenitor
Mid stalk
AT1/AT2
Late tip
DC1
DC2
CX3CR1+ Mac
SPP1+ Mac
Mono-Mac
Neutrophil
Lymphatic endo
Interm lymphatic endo
SCG3+ Adrenal endo
Interm fibro
Airway fibro
Alveolar fibro
ILCP
Pulmonary NE
CD5_minus mature B
CD5_plus mature B
Immature B
Megk
Im. ery.
Erythroblast
Deuterosomal
Squamous DC
Early meso
Early tip/stalk/airway progenitor
Neuron
Mid late meso
Myelocyte-like
Treg
CD4T
NKT
CD16+ NK
CD56bright NK
ILC3/CTB T
ILC2/CTB T
Late cap
Mid cap
Artery peri
Aerocyte
Venous endo
Capillary
Vascular SMC
Airway SMC
Early fibro
Mid fibro
Small pre-B
GRHL+ PJE
Late fibro
Pro-B
Late pre-B
Late pre-B
Norm. Enrichment −log10(P−adj) [0−Max]
0
100

Figure S11

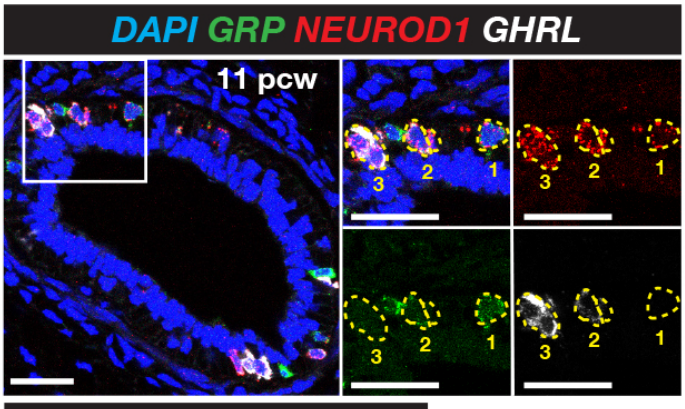
A



B

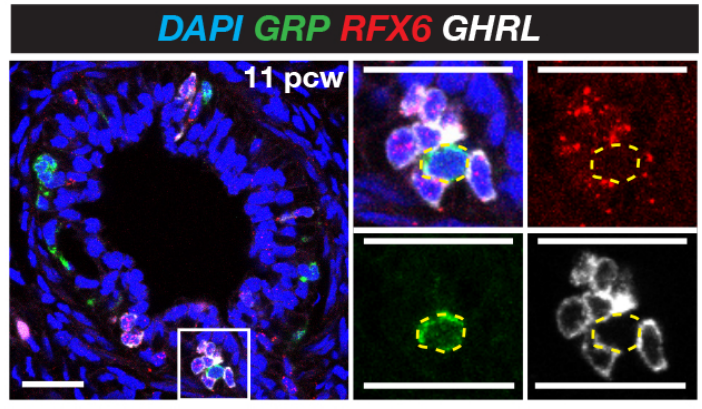


C

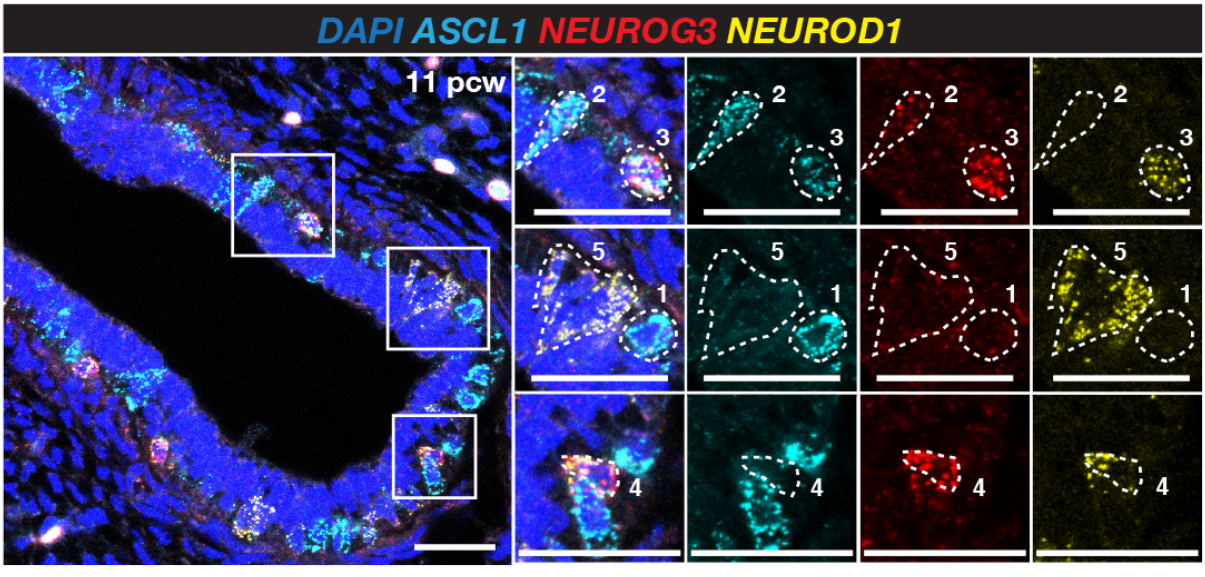


1: $GRP^{+}NEUROD1^{low}GHRL^{-}$
2: $GRP^{low}NEUROD1^{+}GHRL^{low}$
3: $GRP^{-}NEUROD1^{+}GHRL^{+}$

E



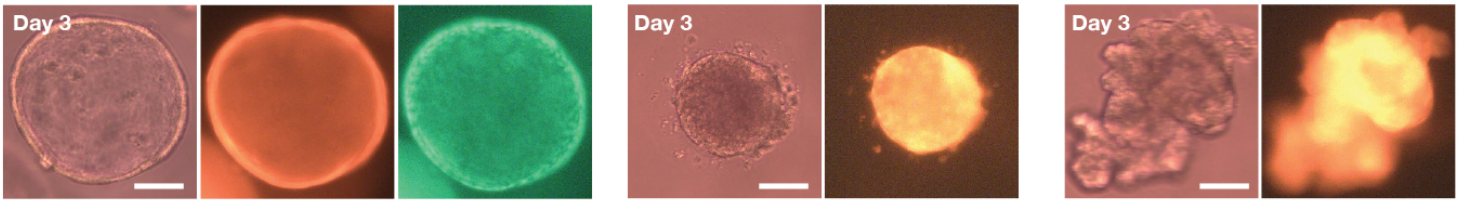
D



1: $ASCL1^{+}NEUROG3^{-}NEUROD1^{-}$
2: $ASCL1^{+}NEUROG3^{+}NEUROD1^{-}$
3: $ASCL1^{+}NEUROG3^{+}NEUROD1^{+}$
4: $ASCL1^{-}NEUROG3^{+}NEUROD1^{+}$
5: $ASCL1^{-}NEUROG3^{-}NEUROD1^{+}$

Figure S12

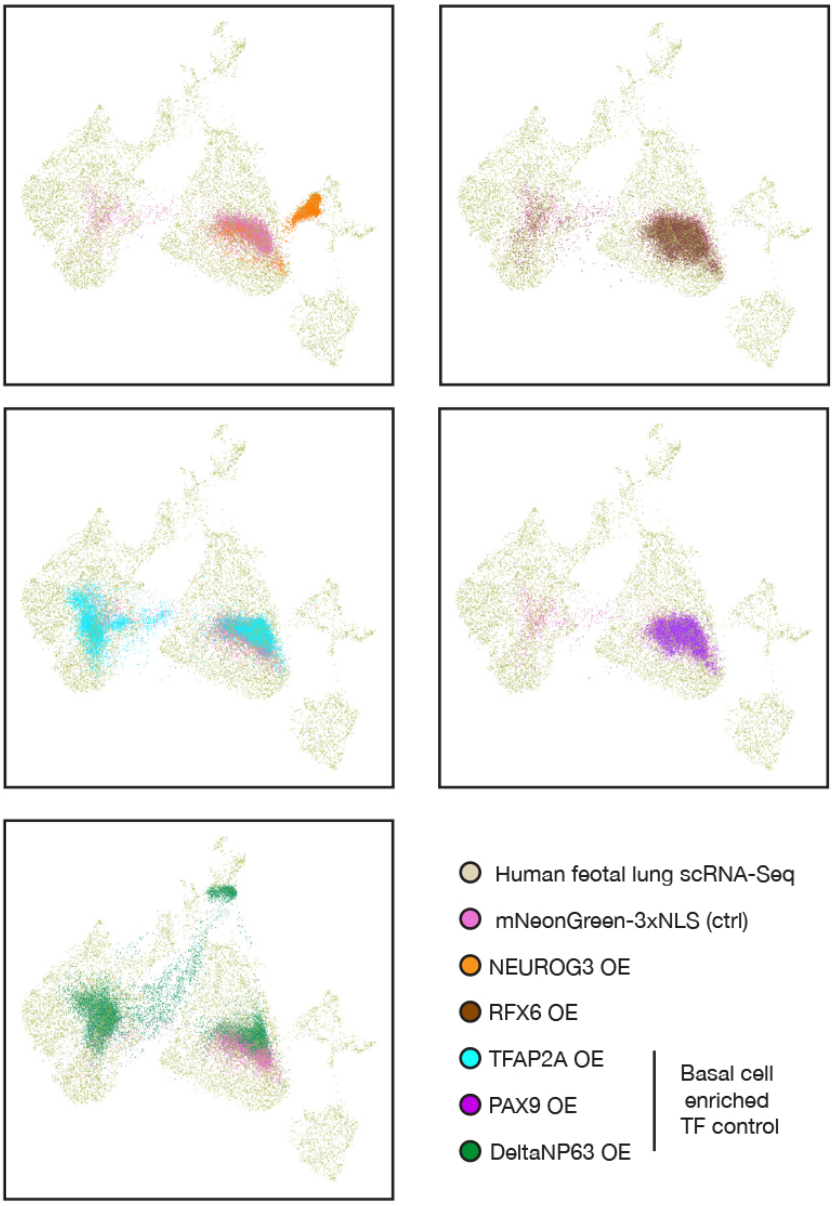
A



mNeonGreen-3xNLS Ctrl | TagRFP | mNG-3xNLS | ASCL1 OE | TagRFP | NEUROD1 OE | TagRFP

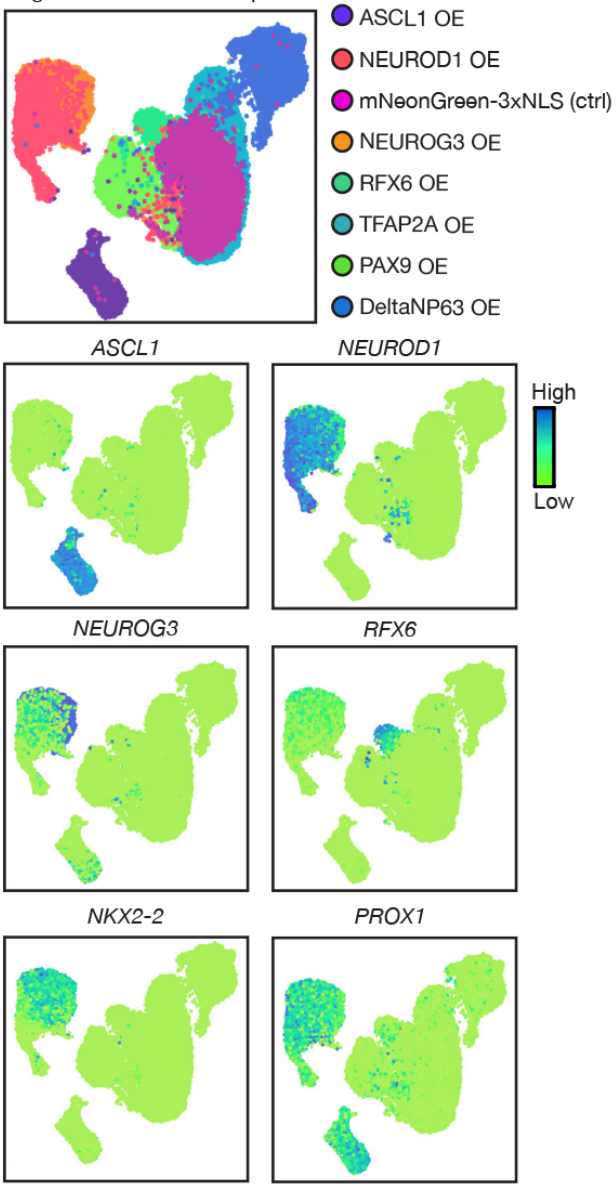
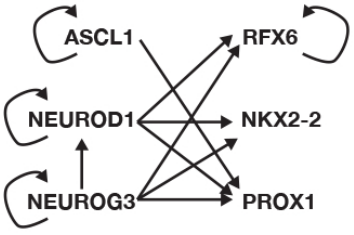
B

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