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

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

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

15. Lead contact

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

Highlights

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

Abstract

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

Introduction

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

63 the embryonic and fetal human lung will identify developmental precursors and progenitors,
64 predict differentiation trajectories and potential gene regulatory networks. This will provide a
65 baseline for studying adult homeostasis and disease.
66
67 The lung buds are specified in the human foregut endoderm at ~5 post conception weeks (pcw)
68 (Burri 1984; Nikolić, Sun, and Rawlins 2018). Subsequent morphogenesis is driven by branching
69 of the distal-most bud tips. The bud tip epithelium comprises $SOX9^+$, $ID2^+$ multipotent progenitors
70 which self-renew during branching (Rawlins, Clark, et al. 2009; Alanis et al. 2014; Nikolić et al.
71 2017; Miller et al. 2018). As the bud tip epithelium branches into the surrounding mesoderm, the
72 epithelial cells that remain in the stalk region start to differentiate into bronchiolar (airway)
73 epithelium (~5-16 pcw) and then (from ~16 pcw) into alveolar epithelium (Nikolić, Sun, and
74 Rawlins 2018). The pattern of growth from undifferentiated multipotent epithelial progenitors at
75 the distal tips means that the position of a cell along the proximal-distal axis of the lung epithelial
76 tree is a strong predictor of its maturity. The more mature cells, which exited the tip first, are found
77 more proximally, whereas the most immature cell states, which exited the tip recently, are found
78 in the tip-adjacent (stalk) regions (Rawlins et al. 2007). In other words, space reflects time in lung
79 development. Therefore, coupling single cell state identification to *in vivo* spatial visualisation can
80 provide high confidence in the identification of novel progenitor cells in the developing lung.
81 Moreover, detailed spatial analysis of cell states allows cell identity designations to be compared
82 to more traditional histological definitions.
83
84 We have generated a high-resolution single cell atlas of human lung development using a
85 combination of scRNA-seq, scATAC-seq, Visium Spatial Transcriptomics, and mRNA *in situ*
86 hybridisation using hybridisation chain reaction (HCR) (Trivedi et al. 2018). Combining these data
87 sources has allowed us to identify 144 cell states/types in 5-22 pcw lung samples. These include
88 novel progenitor cell states, transition populations and a new subtype of neuroendocrine cell which
89 is related to a subtype of human small cell lung cancer. We observe increasing cell maturation over
90 time with many cell states identified in the adult human lungs already present at 22 pcw. Moreover,
91 we have used our atlas to make predictions about progenitor cell states, signalling interactions and
92 lineage-defining transcription factors, and we demonstrate how these can be efficiently tested

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

Results

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

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

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

124 late (15-22 pcw) stages. Cell cycle phase (Fig. S2C, χ^2=25361, p ≈ 0) and dissected region (Fig.
125 S2D, χ^2=968, p = 8.9E-131) were also associated with clustering. However, region was only
126 prominent for a small number of proximally-located cell types (Fig. S2H), suggesting that most
127 proximal to distal regions of the airway branch structure were still represented in both dissected
128 regions of the lung. Epithelial cells were mostly derived from the trypsin-treated and CD326-
129 enriched samples, although airway smooth muscle, myofibroblasts and alveolar fibroblasts were
130 also enriched here (Fig. S2E). Peripheral Nervous System (PNS) cells and chondrocytes were only
131 obtained from 5-6 pcw lungs, likely correlating with lower extracellular matrix (ECM) complexity
132 and/or increased fragility of older neurons. PNS cells were clustered and assigned to cell types,
133 but their scarcity precluded further analysis (Fig. S2A,F,G). Data integration and logistic
134 regression-based comparison showed that gene expression of our annotated cells corresponds well
135 to those from adult lungs (Madissoon et al. 2021) (Fig. S3A-C).
136
137 **A differentiation trajectory of airway progenitor states lies along the distal to proximal axis**
138 **of the developing lungs**
139 The epithelial cells separate by age (Fig. 2A,B), with many basal cells, MUC16$^+$ ciliated cells and
140 secretory cells enriched in the proximally-dissected tissue (Fig. 2B; S2H). The most immature
141 epithelial progenitors are tip cells: SOX9$^+$ multipotent progenitors located at the distal branching
142 tips of the respiratory tree (Nikolić et al. 2017). Tip cells were separated by developmental age
143 into early (5,6 pcw), mid (9,11 pcw) and late (15-22 pcw) populations (Fig. 2A,B) with some
144 shared and some stage specific markers (Fig. 2C). On the epithelial UMAP, each tip population
145 clusters closely with adjacent stalk cells (*SOX9*$^{LO/-}$, *PDPN*^{LO}, *HOPX*^{LO}) and airway progenitors
146 (*CYTL1*$^{LO/+}$, *PCP4*$^+$, *SCGB3A*$^{+/LO}$) (Fig. 2A). The tip, stalk and airway progenitors can be
147 visualised in a distal-proximal sequence in the tissue at all stages tested (10-16 pcw) (Fig. 2E;
148 S4A,B, Supplemental video 1), consistent with the most proximal cells being the most mature.
149 These three cell types form a predicted differentiation trajectory from mid-tip to mid-stalk to mid-
150 airway progenitor which branches into the neuroendocrine, or secretory, lineages (Fig. 2D).
151
152 **Two subtypes of neuroendocrine cells are present in the developing airways**

153 Consistent with previous data (Cutz, Gillan, and Bryan 1985), the earliest differentiated epithelial
154 cells detected were neuroendocrine (NE) cells in 5 pcw lungs (Fig. 2A-C). We identified two types
155 of NE cells: classical pulmonary NE cells (*GRP*$^{+}$) and GHRL^{+} NE cells (*TTR*$^{+}$, *GHRL*$^{+}$) in
156 agreement with a recent human fetal cell atlas (Cao et al. 2020). We observed increasing maturity
157 of NE cells over time (specific populations denoted as precursors on the UMAP). In addition, an
158 intermediate NE population, a putative transition state, connected the two NE cell types (Fig. 2A).
159 At 11 pcw, *GRP*$^{+}$ NE cells were always observed closer to the budding tips, suggesting that they
160 begin to differentiate prior to the *GHRL*$^{+}$ NE cells (Fig. 2F). This spatial difference was not
161 apparent in the oldest samples where both *GRP*$^{+}$ and *GHRL*$^{+}$ cells were observed at all airway
162 levels, although less abundant distally (Fig. S4C). Mouse GHRL^{+} NE cells were not detected in
163 re-analysis of published mouse data (Negretti, Plosa, Benjamin, Schuler, Christian Habermann, et
164 al. 2021; Zepp et al. 2021), or spatially (Borromeo et al. 2016). However, *Ghrl* is expressed in
165 mouse ciliated cells which cluster with human fetal GHRL^{+} NE cells following scVI integration
166 (Lopez et al. 2018) and clustering analysis (Fig. S3D-F) (Zepp et al. 2021).
167

168 **Multiple secretory cell subtypes in the proximal cartilaginous airways**

169 We annotated 5 sub-types of differentiating secretory cells and one more immature proximal
170 secretory progenitor. (i) The proximal secretory progenitors (*SCGB3A2*$^{+}$, *SCGB1A1*$^{-}$, *SCGB3A1*$^{-}$
171 $^{/LO}$, *CYTL1*$^{+}$) were detected in the single cell atlas at 9 pcw, prominent at 11 pcw, but rarer in older
172 lungs consistent with a progenitor state (Fig. 2A-C,G). (ii) Club cells (S*CGB3A2*$^{+}$, *SCGB1A1*$^{+}$,
173 *SCGB3A1*$^{-}$, *SPDEF*$^{-}$, *MUC16*$^{-}$) were detected from 15 pcw in the single cell data (Fig 2A-C,G), or
174 12 pcw in the tissue localised in clusters more distally, but largely dispersed in the more proximal
175 non-cartilaginous regions (Fig. S4D). (iii) Submucosal gland (SMG) secretory cells (*LTF*$^{+}$,
176 *SCGB3A1*$^{+}$, *SPDEF*$^{+}$) were detected from 15 pcw in the single cell data, located in SMG ducts
177 and likely to be a precursor of serous and/or mucous-secreting SMG cells (Fig. 2A-C,G; S4G).
178 (iv) Proximal secretory 1 (*SCGB1A1*^{LO}, *SCGB3A2*$^{+}$, *SCGB3A1*$^{+}$) and (v) proximal secretory 2
179 (*SCGB1A1*$^{+}$, *SCGB3A2*$^{+}$, *SCGB3A1*$^{+}$) appeared from 11 pcw (Fig. 2A-C,G-H; S4E). Both were
180 *SPDEF*$^{+}$, *MUC5B*$^{+}$, *SERPINA1*$^{+}$ (Fig. 2I), suggesting they differentiate into goblet or mucous
181 cells. By contrast, (vi) proximal secretory 3 (*SCGB1A1*$^{+}$, *SCGB3A2*$^{LO/-}$, *SCGB3A1*$^{+}$) was detected
182 from 15 pcw and was *SPDEF*$^{-}$ (Fig. 2A-C), but *CYP2F1*$^{+}$, *MUC4*$^{+}$ and *KRT4*$^{+}$ (Fig. 2G,J). All
183 three luminal proximal secretory cell populations were located in the proximal cartilaginous

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

Other airway cells

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

Predicted airway epithelial differentiation trajectories

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

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

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

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

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

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

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

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

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

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

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

306 Integration of our fetal atlas with adult data revealed high correlation between expected groups:
307 fetal airway progenitors with adult secretory club cells; fetal and adult ciliated and deuterosomal
308 cells; proximal secretory fetal cells with adult goblet cells (Fig. S3A). The AT2 and AT1 cells we
309 detect in the fetal lungs cluster closely with the adult (Fig. S3A; Pearson correlation coefficients:
310 fetal-adult AT2 0.66; AT1 0.80). However, the fetal cells are immature and differ in gene
311 expression to their adult counterparts (for example, for AT2/1, Fig. S3G).
312
313 **Lung endothelial cells exhibit early specialisation into arterial and venous identities**
314 At 5-6 pcw, the endothelial cells (ECs) comprised capillary (early Cap: *THY1*$^{+}$*, CD24*$^{+}$), GRIA2^{+}
315 arterial (*GRIA2*$^{+}$*, GJA5*$^{+}$) and lymphatic ECs (*PROX1*$^{+}$*, STAB1*$^{+}$*, UCP2*^{LO}) (Fig. S8A-C); showing
316 that capillaries and lymphatic vessels are distinct from the earliest stages of lung development and
317 that arterial specification begins prior to venous. With increasing age, the capillary EC lineage
318 moves from early (*THY1*$^{+}$*, CD24*$^{+}$*, EGLN1*$^{+}$) to mid (*CA4*$^{+}$*, KIT*$^{+}$*, EGLN1*$^{+}$) to late (*CA4*$^{+}$*, KIT*$^{+}$)
319 sub-types (Fig. S8A-C), similar to the age transitions in other compartments. Trajectory analysis
320 predicts that both mid- and late-Cap cells generate arterial and venous ECs (Fig. S8G,H).
321 Aerocytes (*CA4*^{LO}*, S100A3*$^{+}$), capillary ECs specialised for gas exchange and leukocyte trafficking
322 (Gillich et al. 2020; Vila Ellis et al. 2020), were observed at 20-22 pcw (Fig. S8A-C) arranged
323 around the developing air sacs (Fig. S8D). Microvasculature specification therefore occurs
324 relatively late in human fetal life coincident with the development of AT1 cells.
325
326 Broad markers of arterial and venous specification were clear in sections at 20 pcw (Fig. S8E).
327 Three distinct arterial ECs were detected. GRIA2^{+} and arterial ECs (*DKK2*$^{+}$*, SSUH2*$^{+}$) form a
328 continuous differentiation trajectory in pseudotime (Fig. S8G,H) with GRIA2^{+} ECs likely to be a
329 more immature form. The OMD^{+} ECs (*GJA5*$^{+}$*, DKK2*$^{+}$*, PTGIS*$^{+}$*, OMD*$^{+}$) cluster with arterial ECs,
330 are more proximal (Fig. S2H) and line the larger arterial vessels (Fig. S11B). By contrast, venous
331 ECs (*PVLAP*$^{+}$*, ACKR3*$^{+}$*, HDAC9*$^{+}$) do not have clear subclusters (Fig. S8A-C). Systemic and
332 pulmonary circulation ECs have been found in adult lungs (Schupp et al. 2021); we cannot detect
333 these in fetal lungs.
334

335 Two major lymphatic ECs were detected, lymphatic ECs (*PROX1*$^{+}$, *STAB1*$^{+}$, *UCP2*^{LO}) and SCG3^{+}
336 lymphatic ECs (*PROX1*$^{+}$, *SCG3*$^{+}$), with an intermediate population connecting them (Fig. S8A-
337 C,F). SCG3^{+} lymphatic ECs resemble a lymphatic valve population (Takeda et al. 2019).
338
339 **Haematopoietic cell types in the developing lung**
340 At the early stages (5-6 pcw) when arterial, capillary and lymphatic ECs were present, embryonic
341 erythrocyte, HMOX1^{+} erythroblast and a small number of macrophages and ILC progenitors were
342 detected, representing the early progenitors of haematopoiesis. After 11 pcw relative numbers of
343 lymphoid and myeloid cells increased, dominated by macrophages, ILCs, dendritic, NK, T and B
344 cells (Fig. 1C, S2A,B, S9A-C,F,G,K). Immature T cells are largely absent from the atlas, consistent
345 with the restriction of T cell development to the thymus. In contrast, a range of early B cell
346 precursors and the ILC precursor were detected. We enriched TCR and BCR fragments from our
347 scRNA-seq libraries which supported cell-type identities and subdivision (Fig. S9D-E,H-J). To
348 look for lung-specific features of the leukocyte cells, we compared our atlas with a pan-fetal human
349 atlas (Cao et al. 2020). Unlike epithelial, endothelial and fibroblasts, leukocytes in our atlas are
350 transcriptionally highly similar to those of other organs with minimal evidence of lung-specificity
351 (Fig. S9L).
352
353 **Developmental trajectories of mesenchymal cells**
354 The broad fibroblast cluster comprises fibroblasts, myofibroblasts, airway and vascular smooth
355 muscle (ASM and vSMC), pericytes, mesothelium and chondrocytes (Fig. 4A,B). Airway
356 fibroblasts and chondrocytes were proximally-enriched; mesothelium distally-enriched (Fig. S2H;
357 4D). There is a distinct separation of cell clusters by age (Fig. 4C). Airway SM cells were observed
358 from 9 pcw, consistent with previous immunostaining (Nikolić et al. 2017), and showed increasing
359 maturity over time (Fig. 4A-C). Two distinct populations of vSMC were observed throughout the
360 time course: vSMC1 (*NTRK3*$^{+}$, *NTN4*$^{+}$, *PLN*$^{-}$) and vSMC2 (*NTRK3*$^{+}$, *NTN4*$^{+}$, *PLN*$^{+}$) (Fig. 4A,B)
361 and were intermingled around the same vessels on tissue sections (Fig. S10A,C). vSMC1 was
362 enriched in genes relating to ECM organisation and cell adhesion, whereas vSMC2 was enriched
363 for transcripts encoding contractility proteins and signalling molecules (Fig. S10B). Intermingling
364 of vSMC subtypes with different levels of contractility proteins is seen in adult lungs (Frid et al.
365 1997); our developmental observation suggests that these represent normal functional/ontological

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

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

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

397 (Fig. 4D,G,J,K; S10E,F). We did not detect lipofibroblasts (Travaglini et al. 2020), meaning that
398 they are either exceptionally rare, or form later than 22 pcw, or do not form distinct clusters in all
399 lung data sets (Madissoon et al. 2021). Endothelial and fibroblast populations align well between
400 fetal and adult data (Fig. S3B,C), but with some unique developmental states, such as fetal
401 early/mid-fibroblasts and myofibroblasts.
402
403 Myofibroblasts formed three distinct groups in our single cell data. Myofibroblast 1 (*CXCL14*$^{+}$*,*
404 *KCNK17*$^{+}$*, CT45A3*$^{+}$*, THBD*^{LO}) appeared at 9 pcw and persisted to 20 pcw. Myofibroblast 2
405 (*CXCL14*$^{+}$*, KCNK17*$^{+}$*, CT45A3*$^{+}$*, THBD*^{HI}) and myofibroblast 3 (*CXCL14*$^{+}$*, KCNK17*$^{+}$*, CT45A3*$^{-}$
406 *, THBD*$^{-}$*)* were predominantly identified at 22 pcw (Fig. 4A,B). Throughout development,
407 myofibroblasts (*CXCL14*$^{+}$*, KCNK17*$^{+}$) were visualised surrounding the developing stalk region of
408 the epithelium, suggesting a close signalling relationship (Fig. 4D,H, S10G,H). Although not
409 detected in significant numbers in the scRNA-seq data until 22 pcw, we see myofibroblast 2
410 (PDGFRA^{+}, THBD^{HI}, *NOTUM*$^{+}$) around the stalk epithelium from 15 pcw (Fig. 4D; S10H,J,K),
411 the same position as myofibroblast 1. The appearance of myofibroblast 2 is coincident with the
412 acquisition of AT2 markers by the late tip cells, and it may be a more mature state of myofibroblast
413 1. Myofibroblast 2 was enriched in gene expression associated with cell contractility and focal
414 adhesions, as well as WNT signalling (Fig. S10J,K). Co-expression of the Wnt-responsive genes
415 *LEF1*, *NOTUM* and *NKD1* suggests that myofibroblast 2 is responding to local Wnt expression
416 (*WNT2* is high in alveolar fibroblasts) and producing the secreted Wnt inhibitor NOTUM;
417 potentially to regulate local cell patterning. We tested this hypothesis using co-culture experiments
418 where myofibroblast 2 was shown to both respond to WNT and to modulate the Wnt-response of
419 co-cultured epithelial cells (Lim et al. 2021). *In vivo,* it is likely that myofibroblast 2 modulates
420 the WNT2 signal from the alveolar fibroblasts mediating spatial patterning of epithelial AT2 and
421 AT1 fate in human lung development. By contrast, myofibroblast 3 has higher expression of genes
422 associated with ECM organisation and a variety of signalling molecules, including *C7*, *RSPO2* and
423 *BMPER* (Fig. S10J). Myofibroblast 3 was always localised to the developing air sacs (Fig. S10I),
424 rather than the stalk epithelium, and are likely to be precursors of the alveolar myofibroblasts (Li
425 et al. 2020; Zepp et al. 2021).
426
427 **Signalling niches in lung development**

428 We used CellPhoneDB (Efremova et al. 2020) to analyse cell-cell communication with the aim of
429 predicting signalling interactions controlling cell fate allocation. We focused on 15-22 pcw cells
430 and, based on the spatial localisation of the major fibroblast populations (Fig. 4E-G), analysed
431 signalling within three niches, defined as - Airway niche: airway fibroblasts; late airway SMCs;
432 airway epithelial cells. Alveolar niche: alveolar fibroblasts, aerocytes, late Cap cells, late tip cells,
433 AT1, AT2. Adventitial niche: adventitial fibroblasts, arterial endothelium, OMD^+ endothelium,
434 vascular smooth muscle cells. CellPhoneDB predicts numerous signalling interactions
435 (Supplemental Table 2) which we curated by plotting the expression of ligand-receptor pairs
436 representing major signalling pathways (Fig. 5A,B; S11A). We observed expected interactions,
437 including high levels of Notch ligands and receptors and CXCL12-CXCR4 signalling in the
438 advential niche (Fig. S11A,B) (Herbert and Stainier 2011). Similarly, expected signalling
439 predicted in the alveolar niche included aerocytes to late cap cells (ALPN-ALPNR) and alveolar
440 epithelial cells to microvascular ECs (VEGFA-FLT1/FLT4/KDR) (Fig. 5B; S11B) (Gillich et al.
441 2020; Vila Ellis et al. 2020).
442
443 Airway fibroblasts were predicted to signal via TGFβ3 and BMP4 to the airway epithelium,
444 consistent with roles for these signals in human basal cell specification and differentiation (Miller
445 et al. 2020; Mou et al. 2016). Airway fibroblasts and ASM were also predicted to signal to the
446 epithelium via FGF7/18 to FGFR2/3 and non-canonical WNT5A to FZD/ROR (Fig. 5A). By
447 contrast, although FGF and WNT signalling interactions were predicted in the alveolar niche,
448 interactions were based on lower levels of *FGF*, but higher levels of canonical *WNT2* and its
449 receptor (Fig. 5B). The predicted FGF and WNT signalling interactions in the alveolar niche/late
450 tip cells are consistent with the requirement of these factors for long-term self-renewal of human
451 distal tip organoids (Nikolić et al. 2017; Lim et al. 2021). Tissue staining showed that although
452 *FGF7* is expressed fairly ubiquitously, the airway fibroblasts and ASM form a distinct barrier
453 between the airway epithelium and the *WNT2* expression (Fig. 5C-E). Based on these data, we
454 predicted that removing canonical WNT, but retaining FGF signalling would promote airway
455 differentiation in the human distal tip organoids (Fig. 5F). Indeed, we observed robust basal,
456 secretory and ciliated cell differentiation in response to FGF-containing medium (Fig. 5G,H).
457
458 **scATAC-seq analysis identifies putative cell fate regulators**

459 Single cell ATAC-seq provides an independent method of assessing cell type based on open
460 chromatin regions and allows cell type-specific TFs to be predicted. After tissue dissociation, the
461 single cell suspensions were split and half of the cells processed for nuclear isolation and scATAC-
462 seq (Fig. 1A). Following quality control and doublet removal, 67 scATAC-seq clusters comprising
463 ~100K cells were obtained and a label transfer process was used to annotate scATAC-seq clusters
464 based on our scRNA-seq data (Fig. 6A). Not every cell state detected by scRNA-seq was
465 distinguishable by scATAC-seq, consistent with previous work (Domcke et al. 2020; Cao et al.
466 2020). For example, separate early-tip, stalk and airway progenitor clusters were discerned by
467 scRNA-seq (Fig. 2A), but a combined cluster with strong similarity to all three cell types was
468 detected by scATAC-seq (Fig. 6A). Similarly, the resolution of scATAC-seq allowed us to identify
469 a combined AT1/AT2 cluster and single arterial endothelial, vascular smooth muscle,
470 myofibroblast and basal cell clusters (Fig. 6A). Nevertheless, there was broad agreement between
471 the scRNA-seq and ATAC-seq data in terms of capturing cell types, including many of the
472 novel/lesser-known cell types we identified by scRNA-seq (mid and late tip, mid and late airway
473 progenitors, GHRL^{+} NE, MUC16^{+} ciliated, dueterosomal, airway fibroblasts, aerocytes, SCG3^{+}
474 lymphatic endothelial cells).
475
476 We analysed TF binding motifs in the unique/enriched open chromatin regions in each cluster and
477 plotted the top 5 TF motifs per cell type (Fig. S12). As expected, TFs belonging to the same family
478 are frequently enriched in the same cell type due to similarities in their binding motifs. This
479 analysis revealed some expected TF signatures, for example TCF21 in the alveolar, adventitial and
480 airway fibroblasts (Quaggin et al. 1999), GRHL and FOXA1/2 in epithelium (Gao et al. 2013;
481 Wan et al. 2005), and SOX17 in arterial endothelium (Corada et al. 2013). Examining epithelial
482 cells and focussing on TFs expressed in the corresponding cell type in the scRNA-seq data (Fig.
483 6B,C, marked by asterisk in B), TEAD motifs were enriched in mid-stalk cells consistent with a
484 key role for Yap (van Soldt et al. 2019), NKX2.1 in AT1/AT2 cells (Kimura, Ostrin, and Chen
485 2019), KLF factors in secretory and AT1/AT2 (Liberti et al. 2022) and TP63 in basal cells (Rock
486 et al. 2009). Unexpected TF signatures included HNF1B in late tip cells, and ZBTB7A in early
487 tip/stalk/airway progenitors. We focused on the pulmonary and GHRL^{+} NE cells which cluster
488 closely in both data sets (Fig. 2A, 6A). ASCL1 is required for mouse NE cell differentiation (Ito
489 et al. 2000; Borges et al. 1997) and this motif is strongly associated with both pulmonary and

490 GHRL$^+$ NE cells (Fig. 6B). However, both cell types also have specific TF motifs including
491 NEUROD1 and RFX6 in the GHRL$^+$ NEs, and TCF4 and ID in the pulmonary NEs (Fig. 6B).
492 Consistent with this, there are distinct, unique regions of open chromatin, especially in the
493 neighbourhood of cell-type specific genes such as GRP and GHRL (Fig. 6D).
494
495 We have produced a high-resolution scATAC-seq data set for the developing human lungs which
496 is highly consistent with our scRNA-seq data. Mining this data provides hypotheses for lineage-
497 determining TFs in lung development. As further scATAC datasets become available for control
498 and diseased adult lungs, our data will provide a resource for comparing normal and aberrant TFs
499 and chromatin regulation.
500
501 **Transcriptional control of neuroendocrine cell subtype formation**
502 Pulmonary NE and GHRL$^+$ NE cells share the expression of many TFs and open chromatin
503 regions, but are transcriptionally distinct. In our scRNA-seq data, they were both observed along
504 a maturation trajectory (from what are labelled as precursors), shared classical NE markers
505 (*CHGA, SYP*), but differed in TF and hormone expression (Fig. 7A,B). A third NE population
506 (intermediate NE) clustered between pulmonary and GHRL$^+$ NE cells with intermediate gene
507 expression (Fig. 7A,B), although it did contain a small number of cells expressing the unique
508 marker *NEUROG3.* Pseudotime trajectory analysis suggested that pulmonary NE and GHRL$^+$ NE
509 cells were derived from airway progenitors/stalk cells and that intermediate NEs are an additional
510 transition population (Fig. S13A,B). Transition states between pulmonary NE and GHRL$^+$ NE
511 were observed in sections (Fig. S13C). We therefore postulated that Pulmonary NE precursors
512 could acquire *NEUROG3* and convert to GHRL$^+$ NE fate (Fig. 7C), or vice-versa - GHRL$^+$
513 precursors converting to pulmonary NE fate. In sections, *ASCL1* was co-expressed with *GRP*, but
514 rarely with *GHRL*. We also observed *ASCL1* single-positive cells, likely representing pulmonary
515 NE precursors (Fig. 7D). *NEUROD1* was co-expressed with *GHRL*, but also observed with *GRP*
516 (Fig. 7E). Whereas *NEUROG3* was co-expressed with *ASCL1* and/or *NEUROD1*, supporting a
517 role in a transition population (Fig. S13D).
518
519 Differential expression of *ASCL1* and *NEUROD1* defines A- and N-type human small cell lung
520 cancer (SCLC), which likely derive from NE cells (Gay et al. 2021). Interestingly, these two TFs

521 coincide with the scRNA-seq marker genes and scATAC-seq TF motif enrichment of our fetal NE
522 cells (Fig. 6B,7B). We generated SCLC feature gene lists (Borromeo et al. 2016) and performed
523 gene signature scoring, showing that the A-type signature resembles pulmonary NEs, whereas the
524 N-type resembles GHRL$^+$ NEs (Fig. 7F). These data suggest that either there are two different NE
525 cells of origin for human SCLCs, or that SCLCs reuse developmental mechanisms as suggested
526 by some mouse models (Ireland et al. 2020). We have been unable to detect GHRL$^+$ NEs in the
527 adult airways using HCR (5 biological replicates). However, a small number of GHRL$^+$ cells are
528 present within a tuft cell cluster in an integrated adult lung cell atlas containing 2.2 million cells
529 (Sikkema et al. 2022), suggesting that GHRL$^+$ NEs could be a rare cell state in the adult airways.
530 Given their relevance to human disease states, we sought to use our single cell atlas to identify
531 lineage-defining TFs controlling NE cell differentiation and test these predictions using our
532 organoid system. We reasoned that overexpression of lineage-defining transcription factors in lung
533 tip organoids (Nikolić et al. 2017; Sun et al. 2021) would promote cell type-specific differentiation.
534
535 Multiple TFs were differentially expressed between pulmonary NE and *GHRL*$^+$ NE cells (Fig. 7B).
536 We used SCENIC analysis of gene regulatory networks (GRNs) (Aibar et al. 2017) along a
537 predicted airway progenitor to GHRL$^+$ NE trajectory (Fig. S13A,B) to identify putative lineage-
538 defining TFs (Fig. 7G). ASCL1, NEUROD1 and NEUROG3 all emerged as potential key nodes
539 and are required for endocrine cell differentiation in various organs (Borges et al. 1997; Ito et al.
540 2000; Mellitzer et al. 2010; Naya et al. 1997). We also selected the *GHRL*$^+$ NE-specific *RFX6*
541 (Fig. S13E) and *NKX2.2* (Fig. 7B), the pan-NE *PROX1* (Fig. 7B) and, as controls, the basal cell-
542 specific TFs *DeltaNTP63*, *TFAP2A*, *PAX9,* and *mNeonGreen-3xNLS*. Overexpression of *PROX1*
543 or *NKX2-2* did not result in NE gene upregulation based on qRT-PCR (not shown) and these TFs
544 were not followed up. The other factors resulted in increased expression of basal or NE markers
545 compared to *mNeonGreen-3xNLS* controls and the experiments were repeated using scRNA-seq.
546 Individual TFs were overexpressed from a doxycycline-inducible construct for 3 days and
547 organoids were maintained in the self-renewing (tip cell-promoting) medium throughout to
548 rigorously assay the lineage-determining competence of the TF (Fig. 7H; S14A), followed by
549 scRNA-seq.
550

551 When mapped to epithelial cells of our fetal lung atlas, the majority of the *mNeonGreen-3xNLS*
552 expressing organoid cells projected to mid-tip or -stalk cells as expected (Fig. 7I). Whereas,
553 overexpression of *DeltaNTP63* resulted in basal cell-like lineages (S14B) consistent with a
554 previous report (Warner et al. 2013), indicating that this simple assay can report TF function.
555 Overexpression of *RFX6*, *TFAP2A* or *PAX9* did not result in the predicted lineage progression at
556 a transcriptome level (S14B). However, *ASCL1*-overexpressing organoids progressed into
557 pulmonary NE precursors (Fig. 7I) and *NEUROD1* overexpression promoted differentiation into
558 *GHRL*$^+$ NE precursors (Fig. 7I). *NEUROG3* overexpression also led to *GHRL*$^+$ NE precursor
559 formation (Fig. S14B), suggesting that the *GHRL*$^+$ NE lineage is the destination of the intermediate
560 NE population (Fig. 7C).
561
562 The 5' differences between the transgenes and endogenous TFs allowed us to distinguish these
563 transcripts and infer gene regulation hierarchy. We observed autoregulation of *ASCL1*,
564 *NEUROD1*, *NEUROG3* and *RFX6* (Fig. S14C). By contrast, *NKX2-2* and *PROX1* were
565 upregulated by other TFs, indicating they are relatively low in the hierarchy (Fig. S14C). *NKX2-2*
566 and *PROX1* expression in the organoid assay matched their expression in NE cells *in vivo* (Fig.
567 7B, S14C), showing that this assay recapitulated key features of the TF network. These
568 experiments have allowed us to test gene GRN predictions from the single cell atlas, confirm the
569 predicted lineage trajectory and provide a foundation for studying human SCLC. This is significant
570 given that there is no evidence that GHRL$^+$ NE cells are present in mice (Borromeo et al. 2016),
571 making the use of mouse models difficult.
572
573 **Discussion**
574 Using a combination of single cell and spatial approaches we have identified 144 cell types, or
575 states, in the developing human lungs across the 5-22 pcw period. We take advantage of a known
576 proximal-distal gradient in epithelial differentiation to identify progenitor and differentiating states
577 in the developing airway, including a new neuroendocrine cell subtype related to SCLC. We
578 suggest that human alveolar epithelial differentiation follows a tip-stalk-AT2/1 fate decision
579 pattern that is different to the prevailing cellular models of mouse alveolar development.
580 Moreover, analysis of the mesenchymal compartment identified three niche regions with distinct
581 signalling interactions, allowing us to identify signalling conditions that are sufficient for airway

582 differentiation of human embryonic lung organoids. We tested GRN predictions for NE cell
583 differentiation in an organoid system, allowing us to identify lineage-defining TFs and provide
584 directionality to the inferred differentiation trajectory. This study provides a paradigm for
585 combining single cell datasets with spatial analysis of the tissue and functional analyses in a human
586 organoid system to provide mechanistic insights into human development.
587
588 We show that lung maturation occurs in concert across cell compartments, for example epithelial
589 AT1 cells and endothelial aerocytes differentiating in parallel at 20-22 pcw. Moreover, we have
590 observed many aspects of cellular differentiation *in utero* showing that they are controlled by
591 prenatal factors, rather than the transition to air breathing with its associated mechanical and
592 hormonal changes. For example, two distinct types of vSMCs and lymphatic endothelial cells are
593 established prior to major alterations in blood flow that occur postnatally.
594
595 The mesenchymal compartment contains multiple fibroblast and myofibroblast cell states and we
596 focus on those present during the later stages of lung development. Airway, adventitial and
597 alveolar fibroblasts are all localised in distinct niche regions and participate in different signalling
598 interactions. Airway and adventitial fibroblasts both express unique combinations of signalling
599 molecules and also form physical barriers between the neighbouring airway epithelium, or vascular
600 endothelium, and the widespread alveolar fibroblasts (Fig. 4,5). Similarly, we characterise a
601 population of myofibroblasts which contacts the developing epithelial stalk region and expresses
602 high levels of the secreted Wnt-inhibitor, *NOTUM* (Fig. S10K); whereas alveolar fibroblasts
603 express high levels of the canonical *WNT2* ligand (Fig. 4). In a separate study, using surface
604 markers identified in this single cell atlas, we were able to specifically isolate alveolar fibroblasts
605 and myofibroblast 2 cells and perform co-culture experiments with late tip organoids (Lim et al.
606 2021). Those experiments confirmed that a three-way signalling interaction between alveolar
607 fibroblasts, myofibroblast 2 cells and late tip cells can control human AT2 spatial patterning.
608
609 We find that GHRL^{+} NE cells are transcriptionally similar to the NEUROD1^{+} N-subtype of SCLC
610 (Fig. 7). Our functional analyses of NE cell differentiation in organoids will provide tools to test
611 these hypotheses. Mouse studies show that fetal transcriptional and chromatin cell states are
612 accessed during the normal process of tissue regeneration and may contribute to neoplasm in

613 chronic inflammation (Larsen and Jensen 2021; Jadhav et al. 2017). Detailed ATAC-seq datasets
614 are not yet available for human lung disease. Our high quality ATAC-seq atlas will provide a
615 baseline for further analyses when adult chromatin accessibility lung atlases are published. In
616 summary, our multi-component atlas is a community resource for future analyses of human
617 development, regeneration and disease.
618
619 **Limitations of the study**
620 We provide a carefully annotated, descriptive cell atlas resource. Many conclusions are derived
621 from trajectory inference or TF binding site analyses and require future validation. The trajectory
622 inference analyses predict lineage relationships and transitions between cell types. However, these
623 transitions reflect similarities in gene expression, for which direct differentiation processes are
624 only one possible explanation. Further, they do not imply directionality. RNA velocity (used in
625 Fig. 2D) provides directionality to trajectory inference predictions using the ratio of spliced versus
626 unspliced mRNA and assuming steady-state transcription and degradation rates. However, over
627 developmental time these assumptions are unlikely to hold across all developmental transitions,
628 which can lead to the inference of incorrect directionality (Bergen et al. 2021). Similarly, when
629 applying RNA velocity algorithms to scRNA-seq data of a known differentiation trajectory,
630 reversed velocities have been reported (Gorin et al. 2022). For these reasons, we performed most
631 of our trajectory inference analysis using Monocle3 (Trapnell et al. 2014). Monocle3 requires user-
632 defined starting and end points and calculates the most likely routes between these points (shown
633 as grey lines on the plots), guided by known biological features of the data (age and spatial
634 arrangement of cells). A further confounding factor for trajectory inference in development is that
635 upon maturation, some cell types are likely to act as progenitors themselves (Rawlins, Okubo, et
636 al. 2009; Rock et al. 2009; Yang et al. 2018). Adding to the difficulty of reconstructing complete
637 lineages, our single-cell dissociation protocols are expected to under-sample certain cells, such as
638 mature neurons and chondrocytes. Furthermore, validation assays for lineage analysis in human
639 systems rely on *in vitro* experiments, including organoid and iPSC differentiation. It is important
640 to acknowledge that these usually define differentiation competence and do not necessarily mean
641 that a specific differentiation route occurs *in vivo*.
642

643 We have performed data integration and regression analysis to compare the identity of our fetal
644 human lung cells with adult human lung cells. There are approximately three decades between the
645 oldest fetal and youngest adult human lung samples sequenced, including a rapid period of
646 postnatal growth and morphogenesis, puberty and an unknown number of
647 infections/environmental insults. Despite this, many fetal-adult similarities can be seen.
648 Nevertheless, it will be important to sequence additional lungs and, when possible, to fill the age
649 gap. Similarly, mouse-human fetal lung cell comparisons have discerned similarities and
650 differences. However, the differences in experimental protocols and annotation granularity
651 between the mouse and human data might have also contributed to the differences we see.
652 Moreover, mice were selected as lab animals partly due to their small size and rapid gestation. It
653 will be informative in the future to make comparisons with a range of fetal lungs, including larger,
654 long-developing species such as pig and sheep, to distinguish between differences due to species,
655 size and gestation period.
656

657 **Acknowledgements**

658 We would like to acknowledge the Gurdon Institute Imaging Facility, and the Cellular Genetics
659 IT and Phenotyping group, New Pipeline Group and DNA pipelines of Sanger Institute; Menna
660 Clatworthy and Muzz Haniffa and RE, CS, ED, IG, MH, CD, WS for discussions on cell-type
661 annotations; MP, AP, SL and CT for informatics support. KL is supported by the Basic Science
662 Research Program through the National Research Foundation of Korea (NRF) funded by the
663 Ministry of Education (2018R1A6A3A03012122). DS is supported by a Wellcome Trust PhD
664 studentship (109146/Z/15/Z) and the Department of Pathology, University of Cambridge. PH
665 holds a non-stipendiary research fellowship at St Edmund's College, University of Cambridge.
666 EM is supported by ESPOD fellowship of EMBL-EBI and Sanger Institute. JPP is supported by
667 the MSCA Postdoctoral Fellowship. ELR is supported by the MRC (MR/P009581/1;
668 MR/S035907/1). ZD is supported by a Wellcome Trust PhD studentship (222275/Z/20/Z). RAB
669 is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) and was
670 an NIHR senior investigator. Gurdon Institute Core support from the Wellcome Trust
671 (203144/Z/16/Z) and Cancer Research UK (C6946/A24843). KBM, JCM and SAT acknowledge
672 funding from the MRC (MR/S035907/1) and from Wellcome (WT211276/Z/18/Z and Sanger core
673 grant WT206194). MZN acknowledges funding from a MRC Clinician Scientist Fellowship

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

Author Contributions

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

Declaration of Interests

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

809 Lim, K., W. Tang, D. Sun, H. Peng, S. Teichmann, J. C. Marioni, K. B. Meyer, and E. L. Rawlins.
810 2021. "Acquisition of Alveolar Fate and Differentiation Competence by Human Fetal Lung
811 Epithelial Progenitor Cells." *bioRxiv*.
812 https://www.biorxiv.org/content/10.1101/2021.06.30.450501.abstract.

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

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

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

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

826 Miller, Alyssa J., David R. Hill, Melinda S. Nagy, Yoshiro Aoki, Briana R. Dye, Alana M. Chin,
827 Sha Huang, et al. 2018. "In Vitro Induction and In Vivo Engraftment of Lung Bud Tip
828 Progenitor Cells Derived from Human Pluripotent Stem Cells." *Stem Cell Reports* 10 (1):
829 101–19.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

877 Sikkema, L., D. Strobl, L. Zappia, E. Madissoon, N. S. Markov, L. Zaragosi, M. Ansari, et al.
878 2022. "An Integrated Cell Atlas of the Human Lung in Health and Disease." *bioRxiv*.
879 https://doi.org/10.1101/2022.03.10.483747.

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

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

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

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

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

898 Trivedi, Vikas, Harry M. T. Choi, Scott E. Fraser, and Niles A. Pierce. 2018. "Multidimensional
899 Quantitative Analysis of mRNA Expression within Intact Vertebrate Embryos." *Development*
900 145 (1). https://doi.org/10.1242/dev.156869.

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

904 Vila Ellis, Lisandra, Margo P. Cain, Vera Hutchison, Per Flodby, Edward D. Crandall, Zea Borok,
905 Bin Zhou, Edwin J. Ostrin, Joshua D. Wythe, and Jichao Chen. 2020. "Epithelial Vegfa
906 Specifies a Distinct Endothelial Population in the Mouse Lung." *Developmental Cell* 52 (5):
907 617–30.e6.

908 Wan, Huajing, Sharon Dingle, Yan Xu, Valérie Besnard, Klaus H. Kaestner, Siew-Lan Ang, Susan
909 Wert, Mildred T. Stahlman, and Jeffrey A. Whitsett. 2005. "Compensatory Roles of Foxa1
910 and Foxa2 during Lung Morphogenesis *." *The Journal of Biological Chemistry* 280 (14):
911 13809–16.

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

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

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

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

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

929 **FIGURE LEGENDS**

930

931 **Figure 1. Experimental overview.**

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

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

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

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

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

958 (C,D) UMAP visualisation of early to late tip, late stalk, fetal AT1 and AT2 cells, coloured by cell
959 types (C) and stages (C'); PAGA analysis (C''); Monocle3 trajectories (D).
960 (E) Gene expression heatmap of trajectory coloured in D.
961 (F) *In situ* HCR at 16, 19, and 21 pcw, *SFTPC* (green), *TPPP3* (red), and *SOX9* (white). White
962 lines/red arrows: columnar tip prognitors, $SFTPC^{+}/SOX9^{+/high}/TPPP3^{+}$. Arrowheads/dashed lines
963 in stalk/air sac regions: cuboidal differentiating fetal AT2 cells, $SFTPC^{+}/SOX9^{low/-}/TPPP3^{low/-}$.
964 Asterisk (*) primitive air sacs.
965 (G) Quantification of cuboidal $SFTPC^{+}/SOX9^{low/-}$ fetal AT2 cells in stalk/air sac regions in F. The
966 $SFTPC^{+}$ tip epithelial cells were excluded by their columnar morphology and marker expression
967 ($SOX9^{low/-}$). Mean ± SD, n >7. Significance evaluated by 1-way ANOVA with Tukey multiple
968 comparison post-test; ns: not significant, *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.
969 (H, I) *In situ* HCR analysis at 21 pcw. Fetal AT2 $SFTPC^{+}$ and $NAPSA^{+}$ (*arrowheads;* H, I) and
970 fetal AT1 $SFTPC^{-}/MMP28^{+}/SPOCK2^{+}$ (*arrow*s; I).
971 (J) Diagram of the acquisition of alveolar progenitor identity by late epithelial tips, followed by
972 differentiation to fetal AT2 and AT1 lineages.
973 DAPI, nuclei. Scale bars, 50 μm.
974 See also Figure S7 and S8.

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

977 (A) UMAP visualisation of mesenchymal cells.
978 (B) Dot plot describing mesenchymal differential marker gene expression.
979 (C) UMAP visualisation of mesenchymal cells coloured by stage.
980 (D) Visium spatial feature plots visualising adventitial fibroblasts, airway fibroblasts, ASPN+
981 chondrocytes, and myofibroblasts-2 on 17 and 20 pcw lung sections. Scores are conservative
982 estimates of cell-type abundance per voxel.
983 (E-H) *In situ* HCR assay (E-H) and immunostaining (G). (E) Adventitial fibroblasts
984 (*SFRP2,*white/*PI16,*red; arrowheads), ECs (*PECAM1*, green). (F) Alveolar fibroblasts (*WNT2*
985 white; *FGFR4* red), tip cells (*SFTPC* green). Asterisks (* myofibroblasts). (G) Airway fibroblasts
986 (*S100A4* red; *AGTR2* white), smooth muscle (ACTA2 green, dashed line). (H) Myofibroblasts

987 (*KCNK17* white, *CXCL14* red; arrowheads), tip cells (*SFTPC*, green). DAPI, nuclei. Scale bars,
988 50 µm.
989 (I) UMAP visualisation of cell types (I), stage (I') and PAGA analysis (I'') of fibroblast
990 differentiation trajectories, .
991 (J, K) UMAPs with Monocle3 trajectories (J) and selected trajectory gene expression heatmaps
992 (K) for mid tip to adventitial fibroblasts (top), alveolar fibroblasts (middle), or airway fibroblasts
993 (bottom).
994 See also Figure S10.

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

996 (A, B) Curated ligand–receptor interaction predictions from CellPhoneDB in alveolar (A) and
997 airway (B) niches. Dot plots visualise gene expression by cell type; dashed arrows indicate the
998 predicted direction of signalling.
999 (C-E) Immunofluorescence/HCR. S100A4/*S100A4*, airway fibroblasts; ACTA2, ASM; CD44, tip
1000 epithelium; PECAM1, ECs. Airway fibroblasts/ASM form a boundary (dashed lines) between
1001 alveolar and airway regions. Lines are between airway fibroblasts/SMCs and airway epithelium.
1002 DAPI, nuclei. Scale bars, 50 µm.
1003 (F) Organoids were cultured in FGF7/10-containing medium, in the presence (self-renewal
1004 medium; SNM) or absence (differentiation medium; DM) of CHIR99021, for 30 days.
1005 (G) qRT-PCR quantifications normalised to the organoids cultured in SNM. Significance
1006 evaluated by 2-way ANOVA with Tukey multiple comparison post-test; **P*<0.05, ***P*<0.01,
1007 ****P*<0.001; *n*=6 organoid lines.
1008 (H) Whole mount immunofluorescence of lung organoids cultured in self-renewal medium (upper)
1009 and differentiation medium (lower). DAPI, nuclei. Scale bar, 25 µm.
1010 See also Figure S11.

1011 **Figure 6. DNA accessibility and motif enrichment revealed by scATAC-seq.**

1012 (A) Single-cell DNA accessibility profiles mapped onto 2D UMAP plane. Coloured for cell states.
1013 (B) A union set of top 10 enriched motifs in the marker peaks among epithelial cell types/states.
1014 Statistical significance is visualised as a heatmap according to the colour bar below. Transcription
1015 factors concordantly expressed based on scRNA-seq data are marked with asterisks.

1016 (C) Expression dotplot of the concordant transcription factors from (B) in epithelial cell types.
1017 (D-E) Read coverage tracks of in silico aggregated "pseudo-bulk" epithelial clusters over the GRP
1018 locus (D) and GHRL locus (E).
1019 See also Figure S12.

1020 **Figure 7. *ASCL1* and *NEUROD1* regulate the formation of two subtypes of neuroendocrine**
1021 **cells.**

1022 (A) Zoom-in UMAP plot of NE lineages.
1023 (B) Dot plot showing selected gene expression in NE lineages.
1024 (C) Schematic model of NE lineage formation.
1025 (D) Left: HCR, *GRP* (green), *GHRL* (red), *ASCL1* (white). Right: Mean ± SEM of *ASCL1*$^+$ cell
1026 types, N = 3 human fetal lungs, n = 243 *ASCL1*$^+$ cells.
1027 (E) Left: HCR, *GRP* (green), *NEUROD1* (red), *GHRL* (white). Right: Mean ± SEM of *NEUROD1*$^+$
1028 cell types: N = 2, 11 pcw human fetal lungs, n = 129; N = 3, 12 pcw human fetal lungs, n = 132.
1029 Scale bars 25 μm.
1030 (F) Gene signature scoring of A-type and N-type SCLC features in the epithelial UMAP.
1031 (G) Scenic analysis of predicted TF network governing mid tip progenitor cells to pulmonary NE.
1032 Trajectory and colour coding match Fig. S13A,B.
1033 (H) Organoids from 8 pcw human fetal lungs were transduced with Doxycline (Dox) inducible TF,
1034 or mNeonGreen-NLS, lentivirus. Transduced organoids were isolated by flow cytometry based on
1035 TagRFP expression, seeded in Matrigel for 10-13 days prior to Dox treatment. Organoid cells were
1036 harvested 3 days post-Dox for scRNA-Seq. N=3 organoid lines.
1037 (I) Left: reference UMAP of primary human fetal lung epithelium. Mid and right: scRNA-Seq of
1038 organoids overexpressing *mNeonGreen-NLS*, *ASCL1* or *NEUROD1* projected onto the primary
1039 data.
1040 See also Figures S13 and S14.

1041 **Figure S1. Quality control for scRNA-seq and scATAC-seq data**

1042 (A) Distributions of the number of genes detected per cell, grouped by 10X libraries.
1043 (B) Proportions of broad cell types in samples treated with Trypsin, and Trypsin plus EPCAM
1044 enrichment following colour codes in Figure 1C.

1045 (C) Initial clusters of data separating compartments, before subclustering.
1046 (D-F) Workflows of the recursive subclustering method (D), the Doublet Cluster Labeling
1047 (DouCLing) method to identify doublet-driven clusters (E), and inference of maternal cells using
1048 Souporcell (F).
1049 (G) Doublet scores calculated by Scrublet.
1050 (H) Inferred doublet clusters using DouCLing.
1051 (I) Number of genes detected projected on UMAP.
1052 (J) Percentage of mitochondrial reads.
1053 (K) Cells in curated doublet clusters.
1054 (L) Cells in clusters of cells coming from other organs. Marker genes in parentheses.
1055 (M) Inferred maternal cells.
1056 (N) Cells in curated low-quality cell clusters.
1057 (O,P) scATAC-seq quality metrics of fragment detection per cell (O) and reads mapped in
1058 transcription-start sites (P).
1059

1060 **Figure S2. Overview of 144 cell types or cell states**

1061 (A-E) All of the curated 144 clusters of single cells projected on UMAP space of transcriptomes,
1062 colored by cell type/state (A), developmental stage (B), inferred cell-cycle phase (C), dissection
1063 region (D) and dissociation/enrichment strategy (E).
1064 (F) Cells from the initial PNS cluster (C7) projected on UMAP space of transcriptomes, colored
1065 by cell type/state.
1066 (G) Selected feature genes of cell types/states in the initial PNS cluster.
1067 (H) Spatiotemporal biases of cell types. Cell types are shown as dots with x representing the
1068 weighted average of developmental stages, y representing the score of proximal enrichment and
1069 the size corresponding to the cluster size.
1070

1071 **Figure S3. Comparing fetal lung scRNA-seq with adult human and mouse lung scRNA-seq.**

1072 (A-F) Correlations of scVI latent variables between human fetal lung cell clusters and those of
1073 previously annotated adult cell clusters (Madissoon et al. 2021, A-C) and mouse lung cell clusters

1074 (Zepp et al. 2021, D-F), focusing on epithelial (A, D), fibroblast (B, E) and endothelial (C, F)
1075 compartments.
1076 (G) Expression dotplot of genes shared or unique to fetal/adult lung AT1/AT2 cell clusters..
1077

1078 **Figure S4. Spatial analysis of airway epithelial cells in the developing human lungs by *in situ***
1079 **HCR.**

1080 (A) Tip and stalk epithelial cells in distal regions of fetal lungs at 17 pcw, immunostained using
1081 antibodies against CD36 (tip epithelial cells, red), PDPN (stalk epithelial cells, white), and E-
1082 cadherin (epithelium, cyan).
1083 (B, B') Airway progenitor cells in distal fetal lungs at 10 (B) and 16 (B') pcw. The airway
1084 progenitor cells marked by *SOX9*$^{-}$*/CYTL1*$^{+}$*/SCGB3A2*$^{+}$ are located proximally to the *CYTL1*$^{-}$
1085 */SCGB3A2*$^{-}$ stalk. *SCGB1A1* indicates club cells (B, white). *SFTPC* is mainly expressed in the tip
1086 and partly located in stalk regions (B', green).
1087 (C) GHRL^{+} neuroendocrine (dashed line, red) and GRP^{+} pulmonary neuroendocrine cells (arrow,
1088 green) in fetal lungs at 22 pcw. *SFTPC* indicates tip epithelial cells (white).
1089 (D) Airway progenitor (arrowhead) and club cells (arrow) in non-cartilaginous airway regions of
1090 fetal lungs at 12 pcw are marked by *SCGB3A2*$^{+}$*/SCGB1A1*$^{-}$ and *SCGB3A2*$^{+}$*/SCGB1A1*$^{+}$,
1091 respectively. Tip, stalk, airway progenitor, and club cells are localised progressively more
1092 proximally from the distal tip regions to the proximal non-cartilaginous airway regions. *SCGB3A2*
1093 (green), *SCGB1A1* (red).
1094 (E) Proximal secretory 1 (arrowhead) and 2 (arrow) are distinguishable by the presence or absence
1095 of *SCGB1A1* expression, each marked by *SCGB3A1*$^{+}$/*SCGB1A1*$^{low/-}$/MUC16$^{low/-}$ and
1096 *SCGB3A1*$^{+}$/*SCGB1A1*$^{+}$/MUC16$^{low/+}$, respectively, in the proximal cartilaginous airway in 15 pcw
1097 fetal lungs. *MUC16*$^{+}$ only cells are MUC16^{+} ciliated cells. *SCGB3A2* (green), *SCGB1A1* (red),
1098 *MUC16* (white).
1099 (F) Proximal secretory 2 (arrowhead) and 3 (arrow) are distinguishable by the presence or absence
1100 of *SCGB3A2* and *MUC16* expression, marked by *SCGB3A2*$^{+}$/*SCGB1A1*$^{+}$/MUC16$^{low/+}$ and
1101 *SCGB3A2*$^{low/-}$/*SCGB1A1*$^{+}$/MUC16^{+}, respectively, in the proximal cartilaginous airway of fetal
1102 lungs at 15 pcw. *SCGB3A2* (green), *SCGB1A1* (red), *MUC16* (white).

1103 (G) Submucosal gland cells (arrow) located in SMGs are marked by strong *LTF* expression with
1104 *SCGB3A1*$^{+}$/*SCGB3A2*$^{-}$ in the proximal cartilaginous airway regions of fetal lungs at 15 pcw.
1105 *SCGB3A2* (green), *LTF* (red), *SCGB3A1* (white).
1106 (H) Ciliated cells and secretory cells are distinguishable by expression of *FOXJ1* (red) or
1107 *SCGB3A2* (green) in the non-cartilaginous airway regions at 19 pcw lungs. Ciliated cells
1108 (arrowhead), *FOXJ1*$^{+}$/*SCGB3A2*$^{-}$; secretory cells (arrow), *FOXJ1*$^{-}$/*SCGB3A2*$^{+}$.
1109 (I) MUC16^{+} ciliated cells (dashed line), ciliated cells (dashed circle), and secretory cells (arrow)
1110 located in the proximal cartilaginous airway regions of fetal lungs at 19 pcw. The MUC16^{+} ciliated
1111 cells express *MUC16* (white) with a weak level of *FOXJ1* (red), whereas the ciliated cells only
1112 express strong *FOXJ1* without *MUC16* expression. *SCGB3A2* (green)
1113 (J, J') Proximal basal cells (J, dashed line) line the basal layer of the proximal cartilaginous
1114 pseudostratified airway in fetal lungs at 19 pcw and are marked by *TP63* (red), *F3* (white), and
1115 *IGFBP3* (green). In contrast, only a few *TP63*$^{+}$ basal cells (J', red, arrowheads) are observed in
1116 the non-cartilaginous, non-pseudostratified airway regions.
1117 (K) *ASCL1*$^{+}$ pulmonary neuroendocrine (arrow) and MUC5AC^{+}/ASCL1^{+} progenitors (arrowhead)
1118 in the non-cartilaginous airway regions of fetal lung at 12 pcw. MUC5AC (green), *ASCL1* (red),
1119 *SCGB3A2* (white).
1120 DAPI, nuclei. Scale bars, 50 μm.
1121 (L) Diagram describing spatial location of epithelial cell types observed in the developing human
1122 lungs.
1123

1124 **Figure S5. Spatiotemporal location, distribution, and quantification of major epithelial cell**
1125 **types along the distal to proximal axis of the developing lungs.**

1126 (A) *In situ* HCR analysis of fetal human lung tissues at mid (10-12 pcw) and late (15-21 pcw)
1127 stages, showing spatiotemporal location and distribution of major epithelial cell types along the
1128 distal to proximal axis of the developing lungs. The lung regions were divided for imaging into
1129 tip, stalk to terminal airway; distal to proximal non-cartilaginous airway; and proximal
1130 cartilaginous airway.
1131 (B) Quantification of cells expressing marker genes of airway lineages along the airway regions at
1132 mid (10-12 pcw, upper) and late (15-21 pcw, lower) stages. *SCGB3A2*, airway progenitors/all

1133 secretory lineage cells; *CYTL1*, airway progenitor cells; *NDUFA4L2*, club/proximal secretory
1134 cells; *FMO2*, club/proximal secretory cells; *FOXJ1*, ciliated cells; *TP63*, basal cells; *SCGB1A1*,
1135 club/proximal secretory cells; *SCGB3A1*, proximal secretory cell subtypes 1-3. Significance was
1136 evaluated by 1-way ANOVA with Tukey multiple comparison post-test; n=3 biological replicates;
1137 ns: not significant *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.
1138 (C) Proportion of proximal secretory progenitor cells, proximal secretory cell subtypes 1-3 within
1139 the proximal cartilaginous airway regions by ages, at 10-12, 15-16, and 19-21 pcw. The secretory
1140 cells in the proximal cartilaginous airway regions were counted: Prox Secretory Prog,
1141 *SCGB3A2*$^{+}$*SCGB3A1*$^{-}$*SCGB1A1*$^{-}$; Prox Secretory 1, *SCGB3A2*$^{+}$*SCGB3A1*$^{+}$*SCGB1A1*$^{-}$; Prox
1142 Secretory 2, *SCGB3A2*$^{+}$*SCGB3A1*$^{+}$*SCGB1A1*$^{+}$; Prox Secretory 3, *SCGB3A2*$^{-}$
1143 *SCGB3A1*$^{+}$*SCGB1A1*$^{+}$. Club cells located in the non-cartilaginous airway regions were excluded.
1144 Significance was evaluated by 2-way ANOVA with Tukey multiple comparison post-test; n=4
1145 biological replicates; ns: not significant *P<0.05, **P<0.01, ***P<0.001.
1146 (D) Diagram describing spatiotemporal distribution of major cell type markers along the distal to
1147 proximal axis of the developing lungs, at mid and late stages. Mid stage only, blue; Late stage
1148 only, red; Mid-to-late stages, green. Arrows indicate narrowed (*CYTL1*) or expanded (*NDUFA4L2*,
1149 *FOXJ1*) distribution after mid to late stage transition.
1150 (E) *In situ* HCR analysis of rare cell type markers of putative ionocytes (*FOXI1* yellow) and tuft
1151 cells (*POU2F3*, red). *E-cadherin*, green.
1152 DAPI, nuclei. Scale bar, 50 μm.

1153 **Figure S6. Trajectory analysis of airway lineage differentiation via airway progenitor cells**
1154 **in the developing human lung.**

1155 (A, A') UMAP visualization (A) and PAGA analysis (A') of a lineage trajectory from mid tip to
1156 proximal secretory lineage cells, including proximal secretory progenitor and proximal secretory
1157 cell subtypes 1 to 3. Mid and late basal cells were shown to be disconnected from other proximal
1158 secretory cell types in the PAGA analysis (A').
1159 (B-C) Trajectory UMAPs, by cell type (B) and stages (B'), and the relevant gene expression
1160 heatmap (C) displaying the selected lineage trajectory from mid tip to proximal secretory cell
1161 subtypes 1 and 2, analysed by Monocle 3. (Note that the grey lines in UMAP indicate all of the
1162 predicted differentiation paths from a user-defined starting point.)

1163 (D, D') UMAP visualization (D) and PAGA analysis (D') of a lineage trajectory from late tip, late
1164 stalk, late airway progenitor to club cells. Basal cells, including late basal, proximal basal, and
1165 SMG basal cells were shown to be left out of the trajectory as they do not connect clearly to the
1166 other cell types in this analysis (D').
1167 (E-F) Trajectory UMAPs, by cell types (E) and stages (E'), and the relevant gene expression
1168 heatmap (C) showing the selected lineage trajectory from late tip to club cells, analysed by
1169 Monocle 3.
1170 (G) Purification of distal *SCGB3A2*-GFP^{+} airway cells from human fetal lung tissues at 8-11 pcw.
1171 The epithelial cells were isolated using EPCAM magnetic microbeads (MACS) from the dissected
1172 distal and proximal airway tissues, followed by infection with lentivirus habouring *SCGB3A2*
1173 promoter-driven GFP. The *SCGB3A2*-GFP positive cell fractions were sorted and analysed by
1174 FACS after 48 hrs and *in vitro* cultured for 28 and 45 days in the airway differentiation medium.
1175 (H) Gene expression profile of the freshly purified *SCGB3A2*-GFP positive cells derived from
1176 distal and proximal airway tissues were investigated by qRT-PCR and compared with dissected
1177 tip cells. *SOX9*, distal tip progenitor marker. *CYTL1*, airway progenitor marker. *SCGB1A1* and
1178 *SCGB3A2*, airway/secretory cell lineage markers. *SCGB3A1*, proximal secretory cell marker. Data
1179 was normalised to *SCGB3A2*-GFP negative cells derived from distal tip/stalk tissues; mean ± SD
1180 of 3 biological replicates. Significance was evaluated by 1-way ANOVA with Tukey multiple
1181 comparison post-test; *P<0.05, **P<0.01, ***P<0.001.
1182 (I) Gene expression analysis of the *in vitro* cultured *SCGB3A2*-GFP positive cells (airway
1183 progenitors) derived from distal airway tissues by qRT-PCR. Airway organoids were formed from
1184 the *SCGB3A2*-GFP positive cells and collected at Day 0, 14, and 28 days after culture for the
1185 analysis. Data were normalised to *SCGB3A2*-GFP negative cells derived from distal tip/stalk
1186 tissues; mean ± SD of 4 biological replicates. Significance was evaluated by 1-way ANOVA with
1187 Tukey multiple comparison post-test; ns: not significant *P<0.05, **P<0.01.
1188 (J-M) Immunofluorescence analysis of two biologically independent, SCGB3A2-GFP+ cell-
1189 derived airway organoids cultured in the airway differentiation medium for 28 (J-L) and 45 (M)
1190 days. SCGB1A1 (J, red), airway progenitor/secretory cell marker. TP63 (K, red), basal cell marker.
1191 FOXJ1 (L, magenta), ciliated cell marker. SCGB3A1 (M, red), proximal secretory cell marker.
1192 DAPI, nuclei. Scale bar, 50 μm.

1193 **Figure S7. Late epithelial tip cells differentiate to AT2 and AT1 cells.**

1194 (A, B) UMAP visualisation (A) of a lineage trajectory from early/mid/late tip to fetal AT2 and
1195 AT1 cells and the relevant gene expression heatmap (B) showing the selected lineage trajectory
1196 analysed by Monocle 3.

1197 (C) *In situ* HCR (*TPPP3* and *SFTPC*) and immunostaining (SOX9) analysis of 15 pcw fetal lung,
1198 describing SOX9^{+}*TPPP3*$^{+}$*SFTPC*$^{+}$ tip epithelial progenitors (lines) and SOX9^{-}*TPPP3*$^{-}$*SFTPC*$^{+}$
1199 fetal AT2 cell population (dashed circles) lining the stalk.

1200 (D) Immunostaining of 21 pcw fetal lung using antibodies against SOX9 (red), PDPN (white), and
1201 E-cadherin (green). Arrows indicate the late tip cell population, which does not co-express the
1202 stalk marker, PDPN.

1203 (E-G) *In situ* HCR analysis of 19 (F) and 21 pcw (E, G) fetal lungs, showing the *SFTPC*$^{+}$ fetal
1204 AT2 cell population (arrowheads) lining the developing air sacs. Arrows indicate *SFTPC*$^{+}$ late tip
1205 cells. (E) *SFTPC* (red). (F, G) *NAPSA* (white; F) and *ETV5* (red; G) overlap with *SFTPC* in the
1206 fetal AT2 cells.

1207 (H-J) *In situ* HCR analysis of distal lung regions at 17 (H), 20 (I), and 21 (J) pcw, visualising
1208 *SFTPC*$^{-}$ fetal AT1 cells (arrows). *SFTPC*$^{-}$/*SPOCK2*$^{-}$ stalk cells at 17 pcw (H) began to express
1209 *SPOCK2* (red) at 20 pcw (I) and further developed to future AT1 cells (*SFTPC*$^{-}$/*SPOCK2*$^{+}$) at 21
1210 pcw (J). Dashed circles (I) and arrowheads (J) indicate fetal AT2 cells. Dashed line (J) shows fetal
1211 AT1 cells lining the developing air sacs.

1212 DAPI, nuclei. Scale bars, 50 μm.

1213 **Figure S8. Endothelial cell types in the developing human lung.**

1214 (A,B) UMAP visualisation of endothelial cells, coloured by cell types (A) and stages (B).

1215 (C) Dot plot describing differential marker gene expression level by cell type.

1216 (D-F) *In situ* HCR analysis of distal lung regions at 20 (E), and 21 (D, F) pcw. (D) Aerocytes
1217 (*S100A3*$^{+}$ red/*CA4*$^{+}$ white), capillary endothelium (*CA4*$^{+}$ white), and all endothelial cells
1218 (*PECAM*$^{+}$, green). (E) Arterial endothelial cells (*GJA5*$^{+}$ red), venous endothelial cells (*ACKR3*$^{+}$
1219 white), and all endothelial cells (*PECAM*$^{+}$, green). (F) Lymphatic endothelial cells (*PROX1*$^{+}$
1220 white) and all endothelial cells (*PECAM*$^{+}$, green). DAPI, nuclei. Scale bars, 50 μm.

1221 (G) Trajectory UMAP and PAGA plot (G'') visualising potential endothelial cell lineage hierarchy
1222 from Mid/Late capillary endothelial cells to arterial endothelial cells, aerocytes, or venous
1223 endothelial cells coloured by cell types (G) and stages (G').
1224 (H, I) Individual trajectory UMAPs (H) and the relevant gene expression heatmaps (I) displaying
1225 potential lineage trajectories derived by Monocle 3 from Mid/Late capillary endothelial cells to
1226 arterial endothelial cells (*top*), aerocytes (*middle*), or venous endothelial cells (*bottom*).
1227

1228 **Figure S9. Clustering and cell type markers for immune cell types and comparison to other**
1229 **fetal data sets.**

1230 (A,C,G) UMAP embeddings of different immune compartments showing myeloid cell types/states
1231 (A), T, NK and ILC lymphoid cells (C) and B lymphoid cells (G).
1232 (B,F,K) Dot plots showing expression of selected marker genes of cell types/states in the three
1233 immune compartments.
1234 (D, E, H, I, J) Enrichment of each class of immune receptors based on abTCR, gdTCR and BCR-
1235 enriched scRNA-seq.
1236 (L) Predicted organ-of-source with highest scores for cells shown in Figure 1, based on the
1237 reference atlas in (Cao et al. 2020).
1238

1239 **Figure S10. Spatial analysis of mesenchymal cell types in the developing human lungs by *in***
1240 ***situ* HCR assay and immunostaining.**

1241 (A) Vascular SMC 1 and 2 are surrounding arterial endothelial cells (*PECAM1*$^{+}$, dashed line), each
1242 marked by *NTN4*$^{+}$/*PLN*$^{-/low}$ (vSMC 1, arrows) and *NTN4*$^{+}$/*PLN*$^{+/high}$ (vSMC 2, lines).
1243 (B) Dot plot describing differential gene expression between vascular SMC 1 and 2.
1244 (C) Vascular SMCs and adventitial fibroblasts in 17 pcw fetal lung. *NDUF4AL2*$^{+}$ red/*NTRK3*$^{+}$
1245 vSMCs (arrows) are surrounded by *NDUF4AL2*$^{-}$/*NTRK3*$^{+}$ adventitial fibroblasts (arrowheads).
1246 *PECAM1* (green) indicates an endothelial cell tube.
1247 (D) *FAM162B*$^{+}$ pericytes (red) are surrounding *PECAM1*$^{+}$ endothelial cells (green) in the
1248 microvascular regions.

1249 (E) Dot plot describing differential marker gene expression level between alveolar, adventitial and
1250 airway fibroblasts.
1251 (F) Concept network visualisation of gene ontology (GO) analysis using clusterProfiler for
1252 differentially expressed genes in alveolar, adventitial and airway fibroblasts.
1253 (G-I) Immunostaining of fetal lung tissues at 11 (G), 15 (H), and 21 (I) pcw, to visualise
1254 myofibroblast populations: Myofibroblast-1 (G) and -2 (H) surrounding the developing stalk
1255 epithelial tubes, and Myofibroblast-3 (I) surrounding the developing air sacs. $ACTA2^{+}$/$PDGFRA^{+}$
1256 Myofibroblast-1 ($THBD^{weak}$; G) and -2 ($THBD^{high}$, arrows; H). $PDGFRA^{+}$ Myofibroblast-3 at 21
1257 pcw, does not express ACTA2 (arrows; I).
1258 (J) Dot plot describing differential gene expression level between myofibroblast-2 and -3. The
1259 myofibroblast-2 population showed enriched expression of Wnt signalling associated genes, e.g.
1260 *NOTUM*, *LEF1*, and *DACH2*.
1261 (K) *In situ* HCR assay of 17 pcw fetal lung tissues. Myofibroblast-2 expresses *NOTUM* (red), a
1262 Wnt antagonist, to block local Wnt signals from alveolar fibroblasts (white, *WNT2*) to the stalk
1263 epithelium.
1264 DAPI, nuclei. Scale bars, 50 μm.
1265

1266 **Figure S11. Signalling ligand-receptor interactions in the adventitial niche.**

1267 (A) Overview of predicted ligand–receptor interactions using CellPhoneDB in the adventitial
1268 niche. Dot plots visualise gene expression by cell type and dashed arrows indicate a predicted
1269 direction of signalling from ligands to receptors.
1270 (B) Visium spot transcriptome cluster map visualising signalling ligands expressed in the fetal
1271 lung tissues at 19 (upper) and 17 (lower) pcw. Scale bars denote 2.5 mm (upper) and 2 mm (lower),
1272 respectively.
1273

1274 **Figure S12. Global landscape of motif enrichment.**

1275 Top 5 enriched motifs in the marker peaks among all the cell types/states. Statistical significance
1276 is visualised as a heatmap according to the colour bar below.
1277

1278 **Figure S13. Transcription factor regulatory network controlling NE subtypes.**

1279 (A) Selected trajectory from Mid tip cells to *GHRL*$^{+}$ NE cells via Intermediate NEs, a transition
1280 cell population.

1281 (B) Heatmap of genes differentially expressed along the trajectory.

1282 (C) Representative HCR images showing the transition between two types of NE cells. *GRP*
1283 (green), *NEUROD1* (red), *GHRL* (white). #1 labelled *GRP*$^{+}$*NEUROD1*^{low}*GHRL*$^{-}$ cells, which have
1284 just started the transition from GRP^{+} pulmonary NE/precursor cells. #2 labelled
1285 *GRP*^{low}*NEUROD1*$^{+}$*GHRL*^{low} cells, in transition to *GHRL*$^{+}$ NE cells. #3 labelled *GRP*$^{-}$
1286 *NEUROD1*$^{+}$*GHRL*$^{+}$, *GHRL*$^{+}$ NE cells. Right: Mean ± SEM of *NEUROD1*$^{+}$ cell types. 11 pcw:
1287 N=2 fetal lungs, n = 129 *NEUROD1*$^{+}$ cells; 12 pcw N=3 fetal lungs, n=132 *NEUROD1*$^{+}$ cells.
1288 Scale bars = 25 μm in all panels.

1289 (D) Representative HCR images showing *NEUROG3* co-expression with *ASCL1* and *NEUROD1*.
1290 Dashed white lines label representative cells showing different combinations of the three
1291 transcription factors, further indicated by #1-#5 labelling. *ASCL1* (cyan), *NEUROG3* (red),
1292 *NEUROD1* (yellow).

1293 (E) Representative HCR images showing *RFX6* expression in *GHRL*$^{+}$ NE cells. Dash yellow line
1294 labelled *GRP*$^{+}$*RFX6*$^{-}$ pulmonary NE cells. Scale bars = 25 μm in all panels.

1295

1296 **Figure S14. Validation of NE transcription factors using human fetal lung organoid system.**

1297 (A) Representative epifluorescent microscopic images showing organoid morphology after 3 days
1298 of mNeonGreen-3xNLS (control), ASCL1, or NEUROD1 overexpression.

1299 (B) ScRNA-seq results of organoid transcription factor overexpression overlay on human fetal
1300 lung scRNA-seq as a reference.

1301 (C) scRNA-seq results of transcription factor overexpression; organoid data only in the UMAP.
1302 Selected transcription factor expression was shown in the middle panel. A regulatory network of
1303 the selected transcription factors were drawn based on the organoid OE data at the bottom of the
1304 panel. (Note that the arrows do not necessarily denote direct interactions).

Figure 1

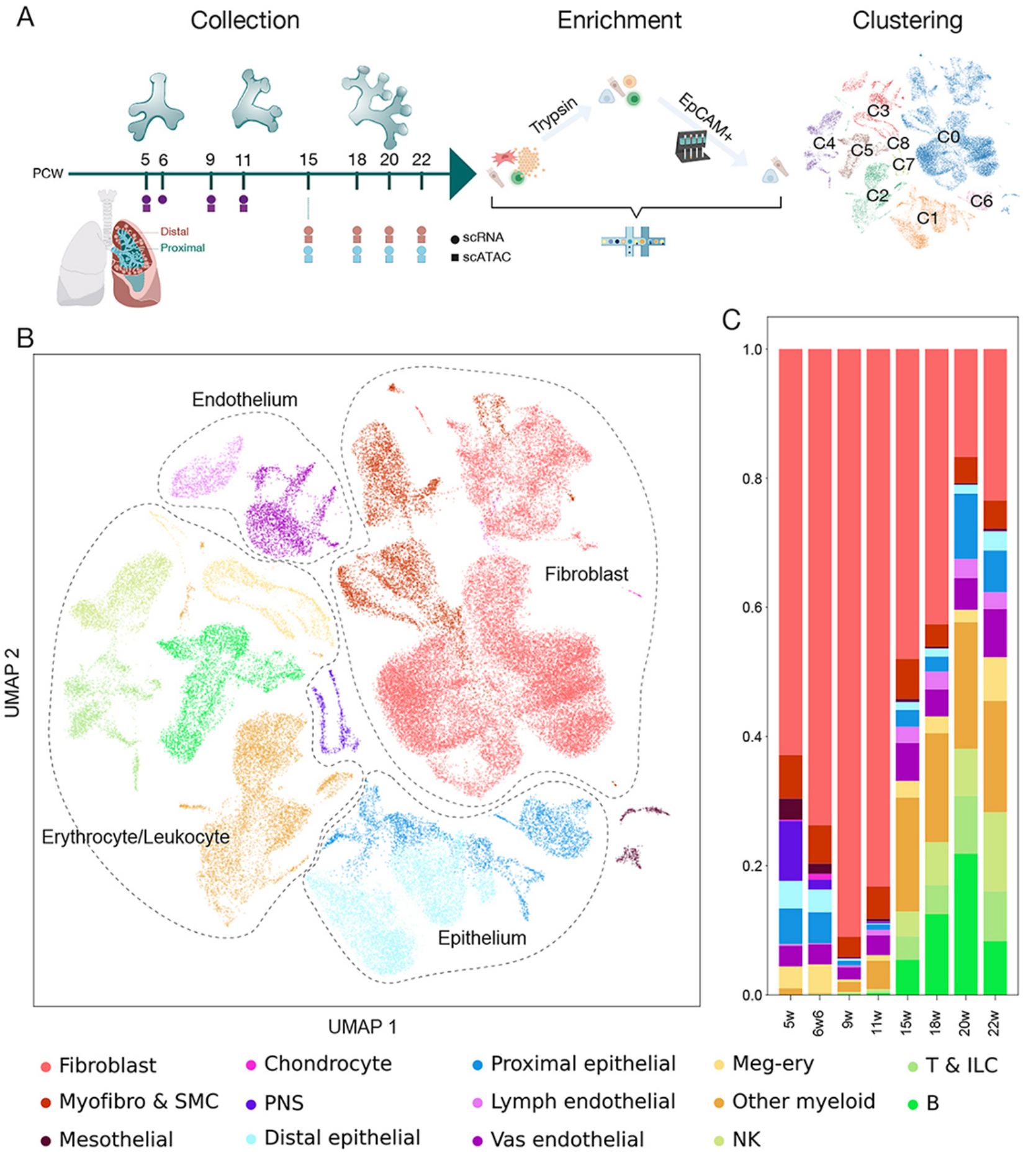

A
Collection
Enrichment
Clustering
PCW
5 6 9 11 15 18 20 22
Distal
Proximal
scRNA
scATAC
Trypsin
EpCAM+
C0
C1
C2
C3
C4
C5
C6
C7
C8
B
Endothelium
Fibroblast
Erythrocyte/Leukocyte
Epithelium
UMAP 2
UMAP 1
C
1.0
0.8
0.6
0.4
0.2
0.0
5w
6w6
9w
11w
15w
18w
20w
22w
Fibroblast
Myofibro & SMC
Mesothelial
Chondrocyte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

Figure 2

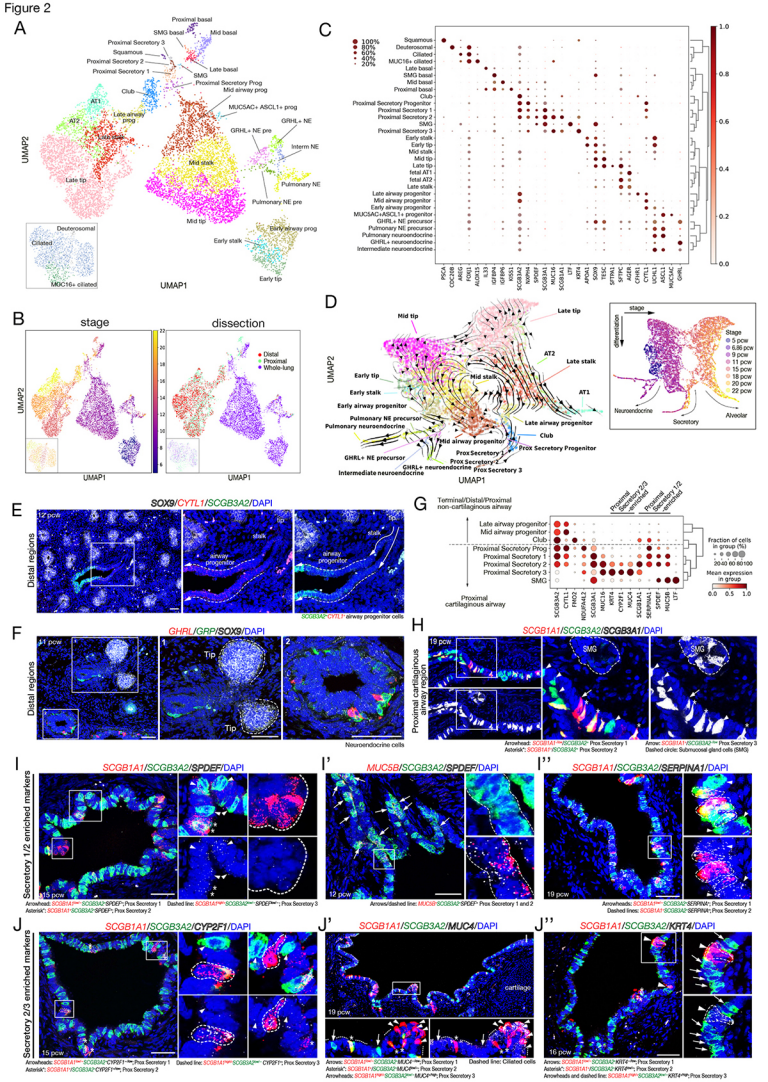
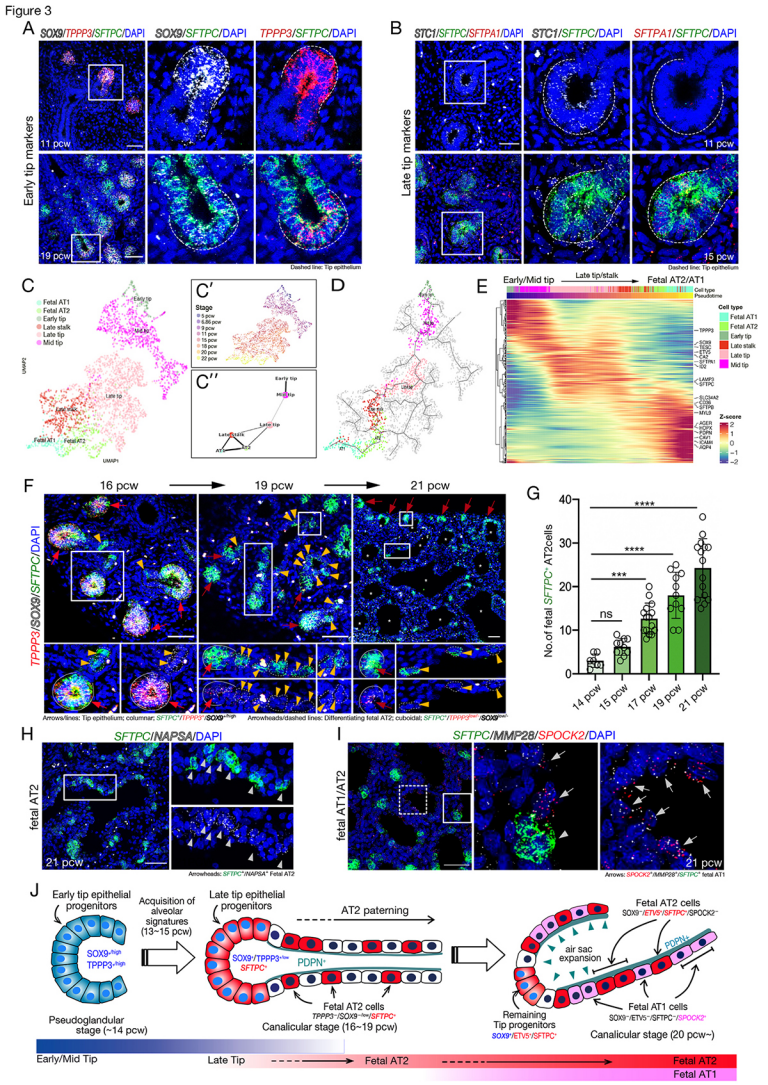
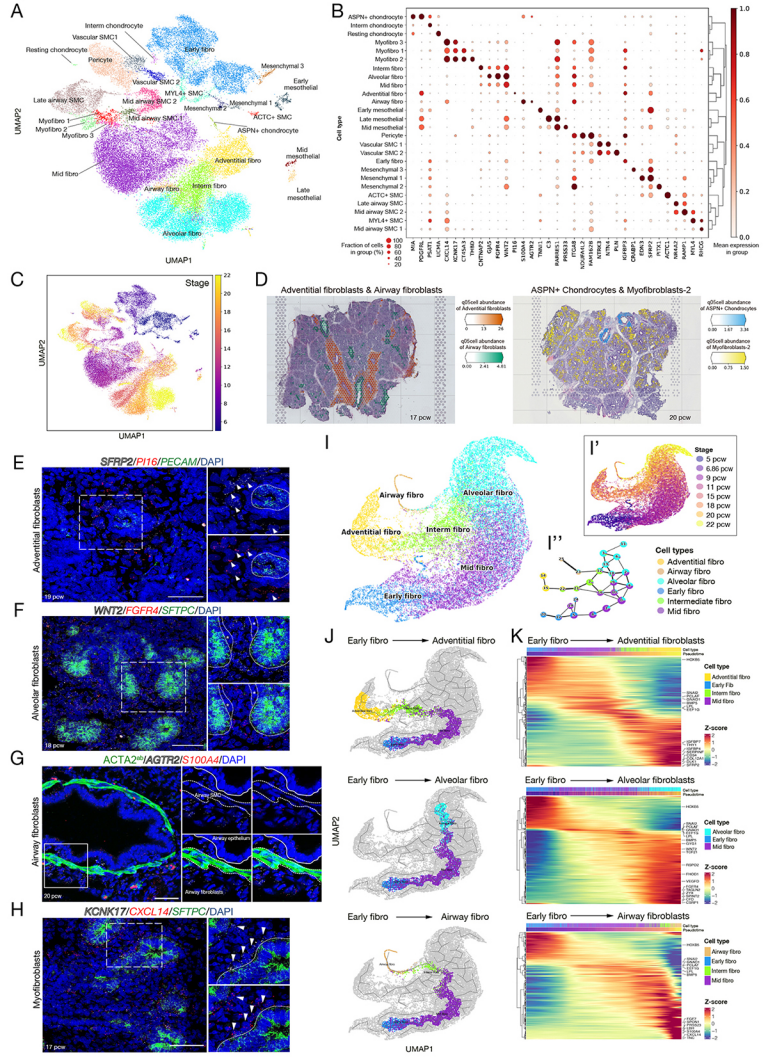

A
Proximal basal
SMG basal
Mid basal
Proximal Secretory 3
Squamous
Proximal Secretory 2
Proximal Secretory 1
Late basal
SMG
Proximal Secretory Prog
Mid airway prog
Club
AT1
AT2
Late airway progenitor
Ciliated
MUC5AC+ ASCL1+ prog
GRHL+ NE
GRHL+ NE pre
Interm NE
Late tip
Mid stalk
Pulmonary NE
Pulmonary NE pre
Mid tip
Early airway prog
Early stalk
Early tip
Deuterosomal
Ciliated
MDC16+ ciliated
UMAP2
UMAP1
B
stage
dissection
Distal
Proximal
Whole-lung
UMAP2
UMAP1
UMAP1
C
100%
80%
60%
40%
20%
Squamous
Deuterosomal
Ciliated
MUC16+ ciliated
Late basal
SMG basal
Mid basal
Proximal basal
Club
Proximal Secretory Progenitor
Proximal Secretory 1
Proximal Secretory 2
SMG
Proximal Secretory 3
Early stalk
Early tip
Mid stalk
Mid tip
Late tip
fetal AT1
fetal AT2
Late stalk
Late airway progenitor
Mid airway progenitor
Early airway progenitor
MUC5AC+ASCL1+ progenitor
GRHL+ NE precursor
Pulmonary NE precursor
Pulmonary neuroendocrine
GRHL+ neuroendocrine
Intermediate neuroendocrine
PSCA
CCK188
CEACAM5
FOXI1
AGR2
L33
GFRPA
MUC16
KRT5
IGSS1
SCGB3A2
SCGB1A1
SPDEF
S100A14
SCGB3A1
MUC1
MUC5B
SFTPC
SFTPB
CYP2F1
CYP4B1
CYTL1
ASCL1
NEUROD1
GHRL
1.0
0.8
0.6
0.4
0.2
0.0
D
Mid tip
Late tip
ATZ
Late stalk
Early tip
Mid stalk
AT1
Early stalk
Early airway progenitor
Pulmonary NE precursor
Pulmonary neuroendocrine
Late airway progenitor
Club
Mid airway progenitor
Prox Secretory 1
Prox Secretory 2
Prox Secretory Progenitor
GRHL+ neuroendocrine
Intermediate neuroendocrine
Prox Secretory 3
UMAP2
UMAP1
differentiation
stage
Stage
5 pcw
6.5/6 pcw
9 pcw
11 pcw
15 pcw
18 pcw
20 pcw
22 pcw
Neuroendocrine
Secretory
Alveolar
E
SOX9/CYTL1/SCGB3A2/DAPI
12 pcw
Distal regions
tip
stalk
airway progenitor
airway progenitor
SCGB3A2+CYTL1+ airway progenitor cells
F
GHRL/GRP/SOX9/DAPI
11 pcw
Distal regions
Tip
Tip
Neuroendocrine cells
G
Terminal/Distal/Proximal non-cartilaginous airway
Late airway progenitor
Mid airway progenitor
Club
Proximal Secretory Prog
Proximal Secretory 1
Proximal Secretory 2
Proximal Secretory 3
SMG
Proximal cartilaginous airway
Proximal Secretory 2/3 enriched
Proximal Secretory 1/2 enriched
SCGB3A2
CYTL1
SCGB1A1
KRT5
SCGB3A1
KRT4
CYP2F1
SCGB1A2
SERPINA1
MUC5B
LTF
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
H
SCGB1A1/SCGB3A2/SCGB3A1/DAPI
19 pcw
Proximal cartilaginous airway region
SMG
SMG
Arrowhead: SCGB1A1+ SCGB3A2+ Prox Secretory 1
Asterisk*: SCGB1A1+ SCGB3A2+ Prox Secretory 2
Arrow: SCGB1A1+SCGB3A2+ Prox Secretory 3
Dashed circle: Submucosal gland cells (SMG)
I
SCGB1A1/SCGB3A2/SPDEF/DAPI
Secretory 1/2 enriched markers
15 pcw
Arrowhead: SCGB1A1+ SCGB3A2+ SPDEF+ Prox Secretory 1
Asterisk*: SCGB1A1+SCGB3A2+ SPDEF+ Prox Secretory 2
Dashed line: SCGB1A1+ SCGB3A2+ SPDEF+ Prox Secretory 3
I'
MUC5B/SCGB3A2/SPDEF/DAPI
12 pcw
Arrows/dashed line: MUC5B+ SCGB3A2+ SPDEF+ Prox Secretory 1 and 2
I''
SCGB1A1/SCGB3A2/SERPINA1/DAPI
19 pcw
Arrowheads: SCGB1A1+ SCGB3A2+ SERPINA1+ Prox Secretory 1
Dashed lines: SCGB1A1+ SCGB3A2+ SERPINA1+ Prox Secretory 2
J
SCGB1A1/SCGB3A2/CYP2F1/DAPI
Secretory 2/3 enriched markers
15 pcw
Arrowheads: SCGB1A1+ SCGB3A2+ CYP2F1+ Prox Secretory 1
Asterisk*: SCGB1A1+SCGB3A2+ CYP2F1+ Prox Secretory 2
Dashed line: SCGB1A1+ SCGB3A2+ CYP2F1+ Prox Secretory 3
J'
SCGB1A1/SCGB3A2/MUC4/DAPI
cartilage
19 pcw
Arrows: SCGB1A1+ SCGB3A2+ MUC4+ Prox Secretory 1
Asterisk*: SCGB1A1+SCGB3A2+ MUC4+ Prox Secretory 2
Arrowheads: SCGB1A1+ SCGB3A2+ MUC4+ Prox Secretory 3
Dashed line: Ciliated cells
J''
SCGB1A1/SCGB3A2/KRT4/DAPI
16 pcw
Arrows: SCGB1A1+ SCGB3A2+ KRT4+ Prox Secretory 1
Asterisk*: SCGB1A1+SCGB3A2+ KRT4+ Prox Secretory 2
Arrowheads and dashed line: SCGB1A1+ SCGB3A2+ KRT4+ Prox Secretory 3

Figure 3



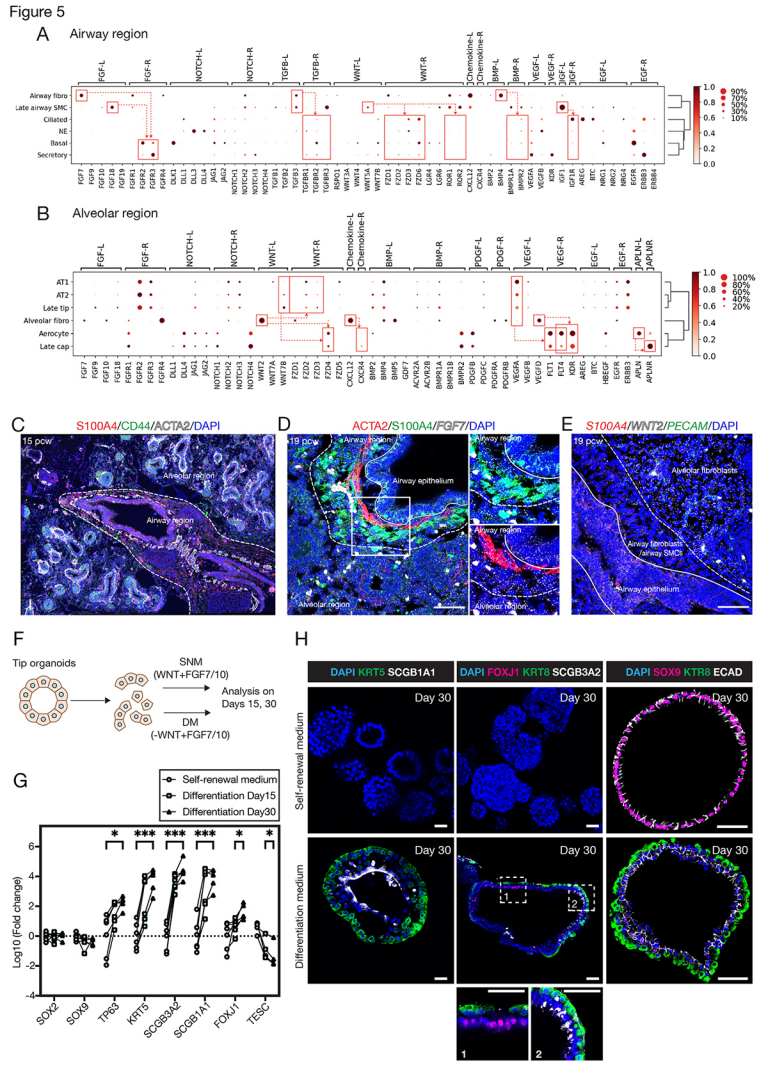
A
SOX9/TPPP3/SFTPC/DAPI
SOX9/SFTPC/DAPI
TPPP3/SFTPC/DAPI
Early tip markers
11 pcw
19 pcw
Dashed line: Tip epithelium
B
STC1/SFTPC/SFTPA1/DAPI
STC1/SFTPC/DAPI
SFTPA1/SFTPC/DAPI
Late tip markers
11 pcw
15 pcw
Dashed line: Tip epithelium
C
Fetal AT1
Fetal AT2
Early tip
Late stalk
Late tip
Mid tip
UMAP1
UMAP2
C′
Stage
5 pcw
6.50 pcw
9 pcw
11 pcw
15 pcw
19 pcw
20 pcw
22 pcw
C″
Early tip
Mid tip
Late tip
Late stalk
AT1
AT2
D
E
Early/Mid tip
Late tip/stalk
Fetal AT2/AT1
Cell type
Pseudotime
Cell type
Fetal AT1
Fetal AT2
Early tip
Late stalk
Late tip
Mid tip
Z-score
F
16 pcw
19 pcw
21 pcw
TPPP3/SOX9/SFTPC/DAPI
Arrows/lines: Tip epithelium; columnar; SFTPC+/TPPP3+/SOX9+high
Arrowheads/dashed lines: Differentiating fetal AT2; cuboidal; SFTPC+/TPPP3+low/SOX9+low
G
No.of fetal SFTPC+ AT2cells
****
****
***
ns
14 pcw
15 pcw
17 pcw
19 pcw
21 pcw
H
SFTPC/NAPSA/DAPI
fetal AT2
21 pcw
Arrowheads: SFTPC+/NAPSA+ Fetal AT2
I
SFTPC/MMP28/SPOCK2/DAPI
fetal AT1/AT2
21 pcw
Arrows: SPOCK2+/MMP28+/SFTPC+ fetal AT1
J
Early tip epithelial progenitors
SOX9+high
TPPP3+high
Pseudoglandular stage (~14 pcw)
Acquisition of alveolar signatures (13~15 pcw)
Late tip epithelial progenitors
AT2 paterning
SOX9+/TPPP3+low
SFTPC+
PDPN+
Fetal AT2 cells
TPPP3+/SOX9+low/SFTPC+
Canalicular stage (16~19 pcw)
Fetal AT2 cells
SOX9+/ETV5+/SFTPC+/SPOCK2+
air sac expansion
PDPN+
Remaining Tip progenitors
SOX9+/ETV5+/SFTPC+
Fetal AT1 cells
SOX9+/ETV5+/SFTPC+/SPOCK2+
Canalicular stage (20 pcw~)
Early/Mid Tip
Late Tip
Fetal AT2
Fetal AT2
Fetal AT1

Figure 4



A
Interm chondrocyte
Vascular SMC1
Resting chondrocyte
Pericyte
Early fibro
Mesenchymal 3
Early mesothelial
Vascular SMC 2
MYL4+ SMC
Late airway SMC
Mid airway SMC 2
Mesenchymal 1
Mesenchymal 2
ACTC+ SMC
Myofibro 1
Myofibro 2
Myofibro 3
Mid airway SMC 1
ASPN+ chondrocyte
Mid mesothelial
Adventitial fibro
Mid fibro
Airway fibro
Interm fibro
Late mesothelial
Alveolar fibro
UMAP1
UMAP2
B
ASPN+ chondrocyte
Interm chondrocyte
Resting chondrocyte
Myofibro 3
Myofibro 1
Myofibro 2
Interm fibro
Alveolar fibro
Mid fibro
Adventitial fibro
Airway fibro
Early mesothelial
Late mesothelial
Mid mesothelial
Pericyte
Vascular SMC 1
Vascular SMC 2
Early fibro
Mesenchymal 3
Mesenchymal 1
Mesenchymal 2
ACTC+ SMC
Late airway SMC
Mid airway SMC 2
MYL4+ SMC
Mid airway SMC 1
Cell type
Fraction of cells in group (%)
Mean expression in group
C
Stage
UMAP1
UMAP2
D
Adventitial fibroblasts & Airway fibroblasts
ASPN+ Chondrocytes & Myofibroblasts-2
17 pcw
20 pcw
E
SFRP2/PI16/PECAM/DAPI
Adventitial fibroblasts
19 pcw
F
WNT2/FGFR4/SFTPC/DAPI
Alveolar fibroblasts
18 pcw
G
ACTA2ab/AGTR2/S100A4/DAPI
Airway fibroblasts
20 pcw
H
KCNK17/CXCL14/SFTPC/DAPI
Myofibroblasts
17 pcw
I
Airway fibro
Alveolar fibro
Adventitial fibro
Interm fibro
Mid fibro
Early fibro
I'
Stage
5 pcw
6.86 pcw
9 pcw
11 pcw
15 pcw
18 pcw
20 pcw
22 pcw
I''
Cell types
Adventitial fibro
Airway fibro
Alveolar fibro
Early fibro
Intermediate fibro
Mid fibro
J
Early fibro → Adventitial fibro
Early fibro → Alveolar fibro
Early fibro → Airway fibro
UMAP1
UMAP2
K
Early fibro → Adventitial fibroblasts
Early fibro → Alveolar fibroblasts
Early fibro → Airway fibroblasts
Cell type
Z-score

Figure 5


A
Airway region
Airway fibro
Late airway SMC
Ciliated
NE
Basal
Secretory
B
Alveolar region
AT1
AT2
Late tip
Alveolar fibro
Aerocyte
Late cap
C
S100A4/CD44/ACTA2/DAPI
15 pcw
Alveolar region
Airway region
D
ACTA2/S100A4/FGF7/DAPI
19 pcw
Airway region
Airway epithelium
Alveolar region
E
S100A4/WNT2/PECAM/DAPI
19 pcw
Alveolar fibroblasts
Airway fibroblasts
Airway SMCs
Airway epithelium
F
Tip organoids
SNM
(WNT+FGF7/10)
DM
(-WNT+FGF7/10)
Analysis on
Days 15, 30
G
Self-renewal medium
Differentiation Day15
Differentiation Day30
Log10 (Fold change)
SOX2
SOX9
TP63
KRT5
SCGB3A2
SCGB1A1
FOXJ1
TESC
H
DAPI KRT5 SCGB1A1
DAPI FOXJ1 KRT8 SCGB3A2
DAPI SOX9 KTR8 ECAD
Day 30
Self-renewal medium
Differentiation medium

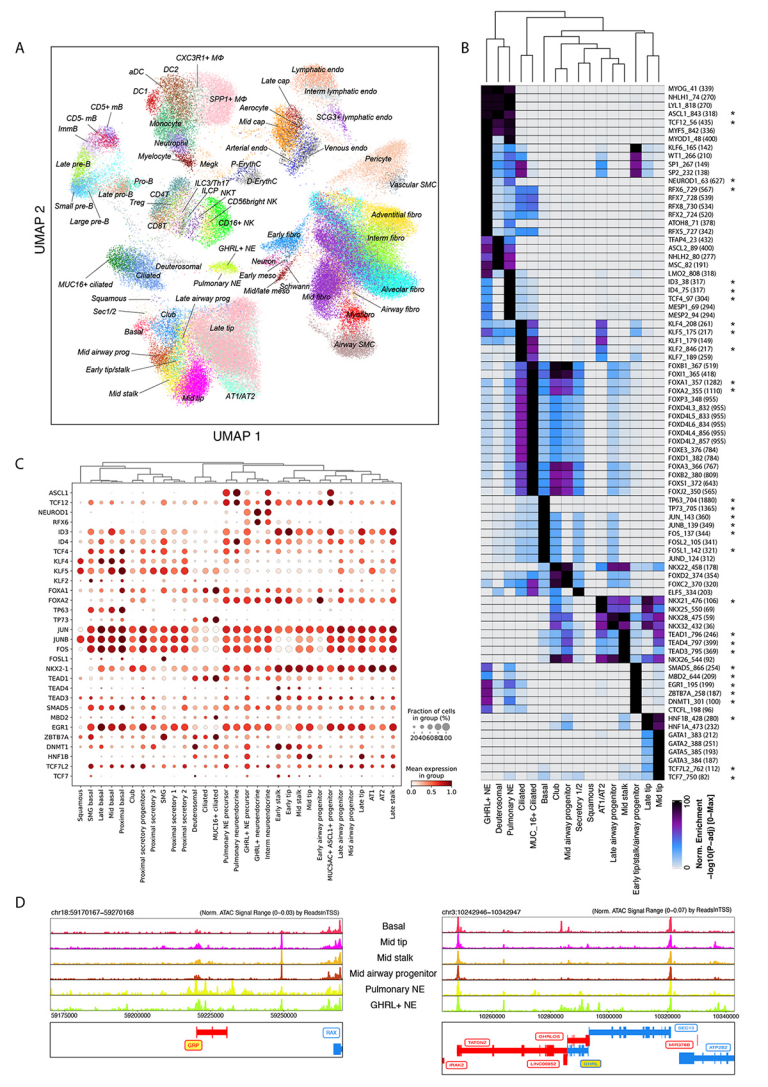
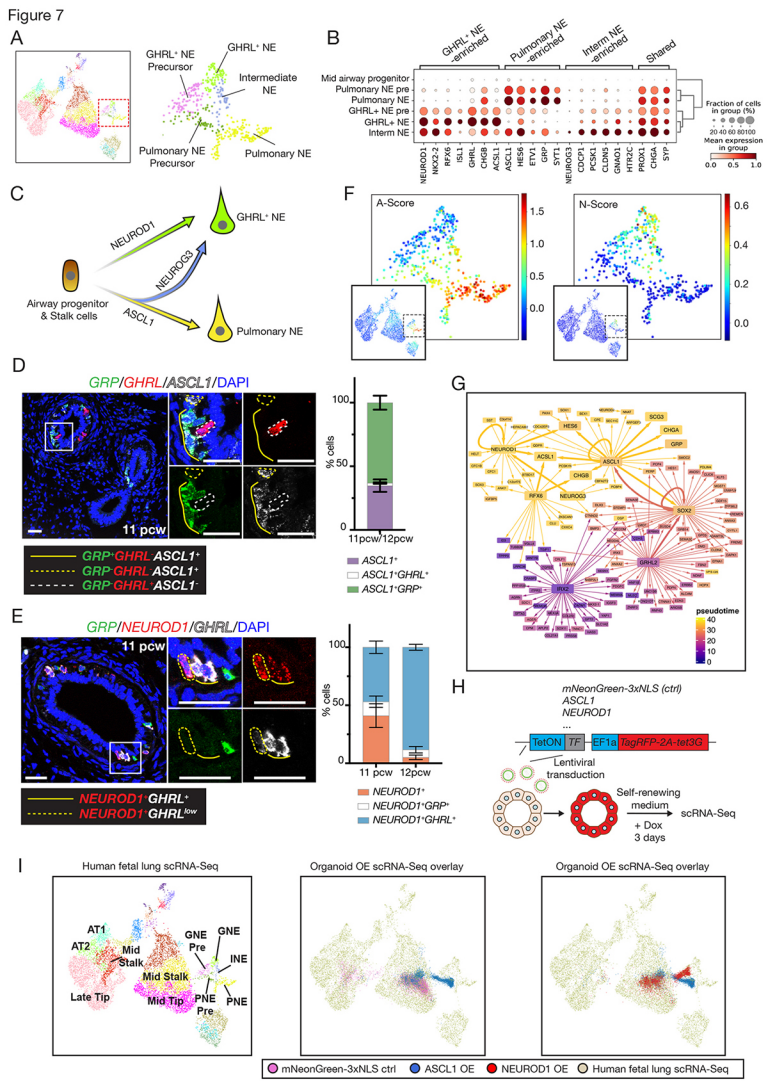
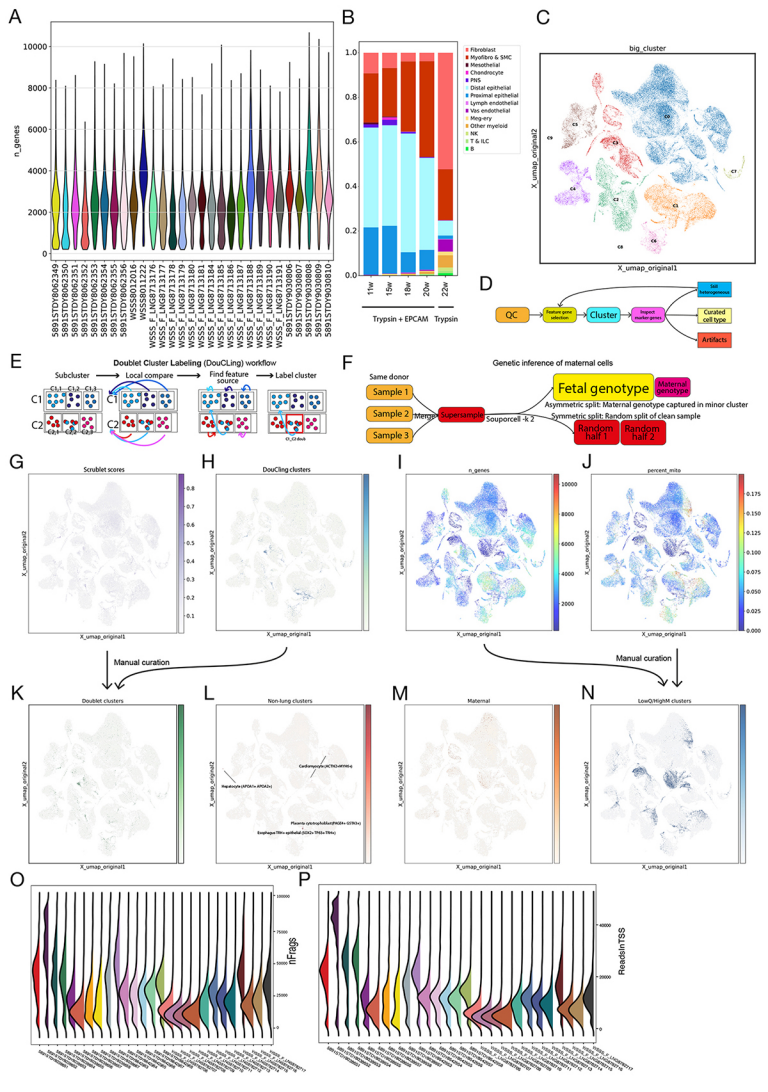
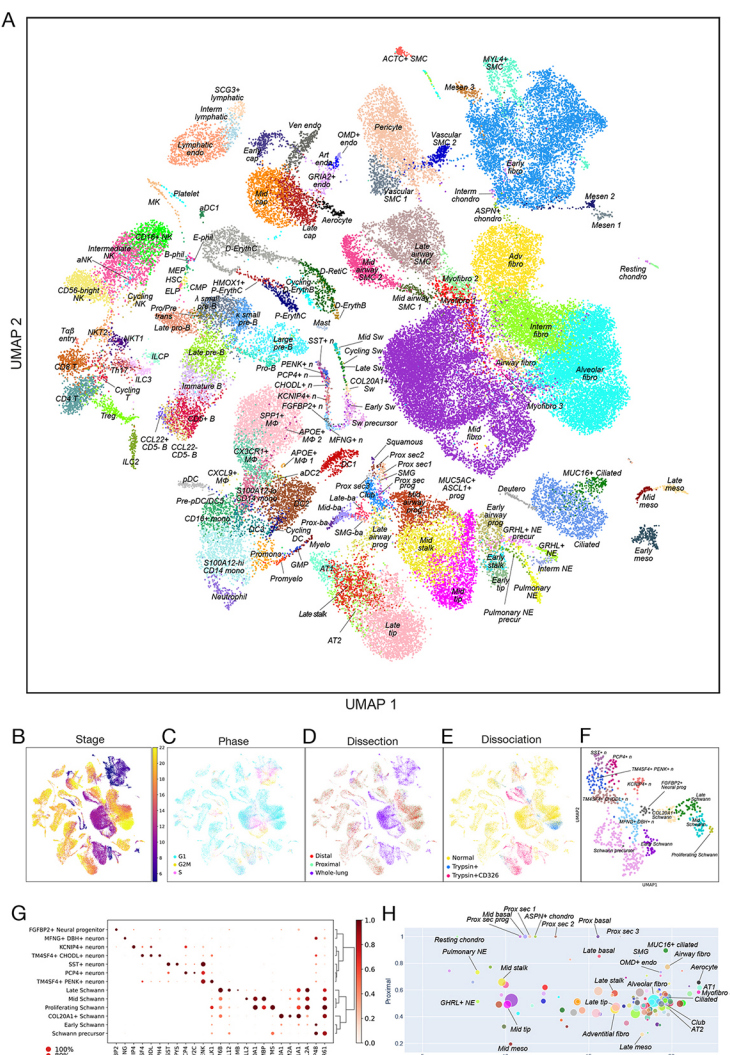

A
UMAP 1
UMAP 2
CXCR3R1+ Mφ
aDC
DC2
DC1
SPP1+ Mφ
Lymphatic endo
Late cap
Interm lymphatic endo
Aerocyte
Mid cap
SCG3+ lymphatic endo
CD5+ mB
CD5- mB
ImmB
Monocyte
Neutrophil
Arterial endo
Venous endo
Pericyte
Myelocyte
Megk
P-ErythC
D-ErythC
Late pre-B
Pro-B
ILC3/Th17
ILCP
NKT
CD56bright NK
Vascular SMC
Small pre-B
Late pro-B
CD4T
Treg
CD16+ NK
Large pre-B
CD8T
Adventitial fibro
Early fibro
Interm fibro
GHRL+ NE
Deuterosomal
Ciliated
Neuron
Early meso
MUC16+ ciliated
Pulmonary NE
Schwann
Alveolar fibro
Squamous
Late airway prog
Mid/late meso
Mid fibro
Sec1/2
Club
Myofibro
Airway fibro
Basal
Late tip
Airway SMC
Mid airway prog
Early tip/stalk
Mid stalk
Mid tip
AT1/AT2
B
MYOG_41 (339)
MHLH1_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)
FOXDL3_832 (955)
FOXDL5_833 (955)
FOXDL6_834 (955)
FOXDL4_856 (955)
FOXDL2_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 (369) *
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)
GATA3_385 (193)
GATA3_384 (187)
TCF7L2_762 (12) *
TCF7_750 (82) *
GHRL+ NE
Deuterosomal
SMG basal
Ciliated
MUC_16+ Ciliated
Basal
Club
Mid airway progenitor
Secretory 1/2
Squamous
Proximal secretory 2
Late airway progenitor
Mid stalk
Early tip/stalk/airway progenitor
Late tip
Mid tip
Norm. Enrichment (0-100)
−log10(P-adj) [0-Max]
C
ASCL1
TCF12
NEUROD1
RFX6
ID3
ID4
TCF4
KLF4
KLF5
KLF2
FOXA1
FOXA2
TP63
TP73
JUN
JUNB
FOS
FOSL1
NKX2-1
TEAD1
TEAD4
TEAD3
SMAD5
MBD2
EGR1
ZBTB7A
DNMT1
HNF1B
TCF7L2
TCF7
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
Squamous
SMG basal
Late basal
Ionocyte/tuft
Proximal basal
Club
Proximal secretory progenitors
Proximal secretory 2
SMG
Proximal secretory 1
Proximal secretory 2
Deuterosomal
Ciliated
MUC16+ ciliated
Pulmonary NE precursor
Pulmonary NE
GHRL+ NE precursor
GHRL+ neuroendocrine
Interm neuroendocrine
Early meso
Early tip
Mid stalk
Mid tip
Early airway progenitor
MUC5AC+ ASCL1+ progenitor
Late airway progenitor
Mid airway progenitor
Late tip
AT1
AT2
Late stalk
D
chr18:59170167–59270168
(Norm. ATAC Signal Range (0–0.03) by ReadsInTSS)
59175000
59200000
59225000
59250000
GRP
RAX
Basal
Mid tip
Mid stalk
Mid airway progenitor
Pulmonary NE
GHRL+ NE
chr3:10242946–10342947
(Norm. ATAC Signal Range (0–0.07) by ReadsInTSS)
10260000
10280000
10300000
10320000
10340000
TMTC1N2
EYHA1CS1
SEC11A
LINC00862
ANAK2
TMTC1N3
ATP2B2

Figure 7



A
GHRL+ NE Precursor
GHRL+ NE
Intermediate NE
Pulmonary NE Precursor
Pulmonary NE
B
GHRL+ NE -enriched
Pulmonary NE -enriched
Interm NE -enriched
Shared
Mid airway progenitor
Pulmonary NE pre
Pulmonary NE
GHRL+ NE pre
GHRL+ NE
Interm NE
NEUROD1
NKX2-2
RFX6
ISL1
GHRL
CHGB
ACSL1
ASCL1
HES6
ETV1
GRP
SYT1
NEUROG3
CDCP1
PCSK1
CLDN5
GNAO1
HTR2C
PROX1
CHGA
SYP
Fraction of cells in group (%)
20 40 60 80100
Mean expression in group
0.0 0.5 1.0
C
NEUROD1
NEUROG3
ASCL1
GHRL+ NE
Airway progenitor & Stalk cells
Pulmonary NE
F
A-Score
1.5
1.0
0.5
0.0
N-Score
0.6
0.4
0.2
0.0
D
GRP/GHRL/ASCL1/DAPI
11 pcw
GRP+GHRL-ASCL1+
GRP-GHRL-ASCL1+
GRP-GHRL+ASCL1-
% cells
100
50
0
11pcw/12pcw
ASCL1+
ASCL1+GHRL+
ASCL1+GRP+
G
pseudotime
40
30
20
10
0
E
GRP/NEUROD1/GHRL/DAPI
11 pcw
NEUROD1+GHRL+
NEUROD1+GHRLlow
% cells
100
50
0
11 pcw
12pcw
NEUROD1+
NEUROD1+GRP+
NEUROD1+GHRL+
H
mNeonGreen-3xNLS (ctrl)
ASCL1
NEUROD1
...
TetON
TF
EF1a
TagRFP-2A-tet3G
Lentiviral transduction
Self-renewing medium
+ Dox
3 days
scRNA-Seq
I
Human fetal lung scRNA-Seq
AT1
AT2
Mid Stalk
Mid Stalk
Late Tip
Mid Tip
GNE Pre
GNE
INE
PNE Pre
PNE
Organoid OE scRNA-Seq overlay
Organoid OE scRNA-Seq overlay
mNeonGreen-3xNLS ctrl
ASCL1 OE
NEUROD1 OE
Human fetal lung scRNA-Seq


A
n_genes
B
Fibroblast
Myofibro & SMC
Mesothelial
Chondrocyte
PNS
Distal epithelial
Proximal epithelial
Lymph endothelial
Vas endothelial
Meg-ery
Other myeloid
NK
T & ILC
B
1Tw
15w
18w
20w
22w
Trypsin + EPCAM
Trypsin
C
big_cluster
X_umap_original1
X_umap_original2
D
QC
Feature gene selection
Cluster
Impact cluster genes
Still heterogeneous
Curated cell type
Artifacts
E
Doublet Cluster Labeling (DouCLing) workflow
Subcluster
Local compare
Find feature source
Label cluster
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
K
Doublet clusters
L
Non-lung clusters
M
Maternal
N
LowQ/HighMt clusters
O
nFrags
P
ReadsInTSS


A
UMAP 2
UMAP 1
ACTC+ SMC
MYL4+ SMC
Mesen 3
SCG3+ lymphatic
Interm lymphatic
Lymphatic endo
Ven endo
OMD+ endo
Pericyte
Vascular SMC 2
Early fibro
Art endo
GRIA2+ endo
Mid cap
Vascular SMC 1
Interm chondro
Mesen 2
Mesen 1
ASPN+ chondro
Platelet
MK
aDC1
Aerocyte
Late cap
CD16+ NK
E-phil
D-ErythC
Intermediate NK
aNK
B-phil
MEP
HSC
ELP
CMP
HMOX1+ P-ErythC
D-RetiC
Late airway SMC
Mid airway SMC 2
Adv fibro
Resting chondro
CD56-bright NK
Cycling NK
Pro/Pre trans
A small pre-B
Cycling ErythB
D-ErythB
Mid airway SMC 1
Myofibro 2
Mast
Late pro-B
P-ErythC
Interm fibro
Taβ entry
NKT2
NKT1
Large pre-B
SST+ n
Mid Sw
Cycling Sw
Late pre-B
ILCP
Pro-B
PENK+ n
PCP4+ n
Late Sw
COL20A1+ Sw
Airway fibro
Alveolar fibro
CD4 T
Th17
ILC3
Cycling
Immature B
CHODL+ n
KCNIP4+ n
FGFBP2+ n
Early Sw
Myofibro 3
CD4 T
Treg
CD5+ B
SPP1+ MΦ
Sw precursor
Mid fibro
CCL22+ CD5- B
CCL22- CD5- B
APOE+ MΦ 2
MFNG+ n
Squamous
CX3CR1+ MΦ
APOE+ MΦ 1
Prox sec2
Prox sec1
ILC2
aDC2
DC1
SMG
Prox sec prog
MUC5AC+ ASCL1+ prog
MUC16+ Ciliated
Late meso
CXCL9+ MΦ
pDC
S100A12-hi CD14 mono
Prox sec
Late-ba
Clip
Mid-ba
Mid airway prog
Deutero
Mid meso
Pre-pDC/DC3
CD16+ mono
Prox-ba
Cycling DC
SMG-ba
Late airway prog
Early airway prog
GRHL+ NE precur
Mid stalk
Ciliated
Early meso
Promono
Myelo
S100A12-hi CD14 mono
GMP
AT1
Early stalk
Interm NE
Promyelo
Early tip
Pulmonary NE
Neutrophil
Mid tip
Pulmonary NE precur
Late stalk
Late tip
AT2
B
Stage
C
Phase
G1
G2M
S
D
Dissection
Distal
Proximal
Whole-lung
E
Dissociation
Normal
Trypsin+
Trypsin+CD326
F
UMAP2
UMAP1
G
FGFBP2+ Neural progenitor
MFNG+ DBH+ neuron
KCNIP4+ neuron
TM4SF4+ CHODL+ neuron
SST+ neuron
PCP4+ neuron
TM4SF4+ PENK+ neuron
Late Schwann
Mid Schwann
Proliferating Schwann
COL20A1+ Schwann
Early Schwann
Schwann precursor
100%
80%
60%
40%
20%
H
Proximal
Stage
Prox sec 1
Mid basal
Prox sec prog
ASPN+ chondro
Prox sec 2
Prox basal
Prox sec 3
Resting chondro
Pulmonary NE
Late basal
SMG
MUC16+ ciliated
Airway fibro
OMD+ endo
Aerocyte
Mid stalk
Late stalk
Alveolar fibro
AT1
Myofibro 3
Ciliated
GRHL+ NE
Late tip
Club
AT2
Mid tip
Adventitial fibro
Mid meso
Late meso

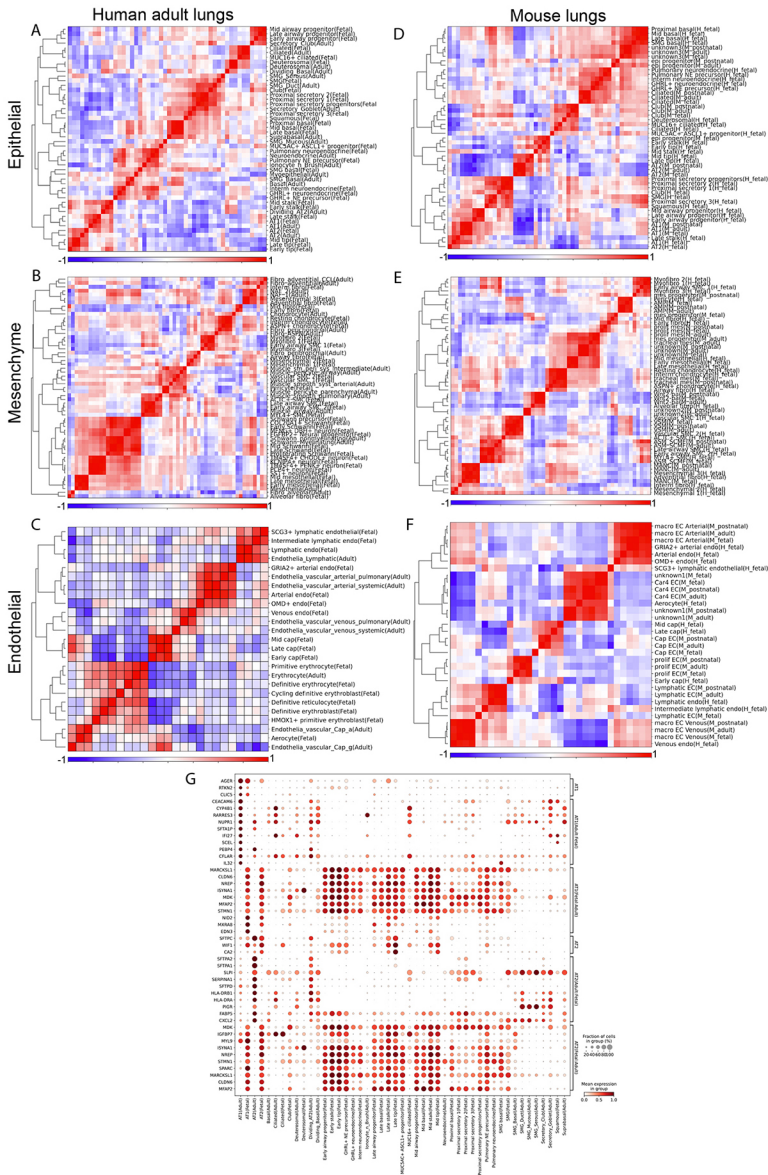
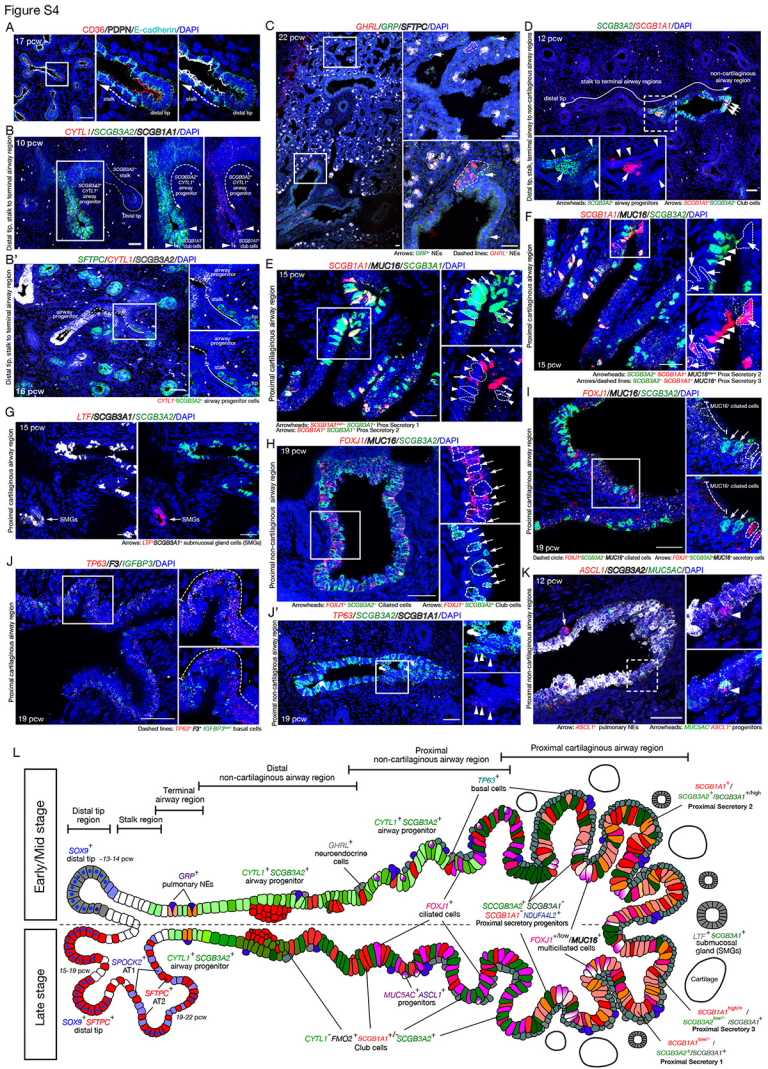

Human adult lungs
Mouse lungs
Epithelial
Mesenchyme
Endothelial
A
B
C
D
E
F
G
-1
1
SCG3+ lymphatic endothelial(Fetal)
Intermediate lymphatic endo(Fetal)
Lymphatic endo(Fetal)
Endothelia_lymphatic(Adult)
GIRIA2+ arterial endo(Fetal)
Endothelia_vascular_arterial_pulmonary(Adult)
Endothelia_vascular_arterial_systemic(Adult)
Arterial endo(Fetal)
OMD+ endo(Fetal)
Venous endo(Fetal)
Endothelia_vascular_venous_pulmonary(Adult)
Endothelia_vascular_venous_systemic(Adult)
Mid cap(Fetal)
Late cap(Fetal)
Early cap(Fetal)
Primitive erythrocyte(Fetal)
Erythrocyte(Adult)
Definitive erythrocyte(Fetal)
Cycling definitive erythroblast(Fetal)
Definitive reticulocyte(Fetal)
Definitive erythroblast(Fetal)
HMOX1+ primitive erythroblast(Fetal)
Endothelia_vascular_Cap_a(Adult)
Aerocyte(Fetal)
Endothelia_vascular_Cap_g(Adult)
macro EC Arterial(M_postnatal)
macro EC Arterial(M_adult)
macro EC Arterial(M_fetal)
GRIA2+ arterial endo(H_fetal)
Arterial endo(H_fetal)
OMD+ endo(H_fetal)
SCG3+ lymphatic endothelial(H_fetal)
unknown1(M_fetal)
Car4 ECIM_fetal)
Car4 ECIM_postnatal)
Car4 ECIM_adult)
Aerocyte(H_fetal)
unknown1(M_postnatal)
unknown1(M_adult)
Mid cap(H_fetal)
Late cap(H_fetal)
Cap ECIM_postnatal)
Cap ECIM_adult)
Cap ECIM_fetal)
proliF ECIM_postnatal)
proliF ECIM_adult)
proliF ECIM_fetal)
Early cap(H_fetal)
Lymphatic ECIM_postnatal)
Lymphatic ECIM_adult)
Lymphatic endo(H_fetal)
Intermediate lymphatic endo(H_fetal)
Lymphatic ECIM_fetal)
macro EC Venous(M_postnatal)
macro EC Venous(M_adult)
macro EC Venous(M_fetal)
Venous endo(H_fetal)
Fraction of cells in group (%)
Mean expression in group

Figure S4



A
CD36/PDPN/E-cadherin/DAPI
17 pcw
distal tip
stalk
distal tip
B
CYTL1/SCGB3A2/SCGB1A1/DAPI
10 pcw
SCGB3A2+ stalk
SCGB3A2+ CYTL1+ airway progenitor
Distal tip
SCGB3A2+ CYTL1+ airway progenitor
SCGB3A2+ CYTL1+ airway progenitor
SCGB3A2+ stalk cells
SCGB3A2+ CYTL1+ airway progenitor
SCGB3A2+ stalk cells
Distal tip, stalk to terminal airway region
B'
SFTPC/CYTL1/SCGB3A2/DAPI
airway progenitor
stalk
airway progenitor
airway progenitor
stalk
16 pcw
CYTL1+ SCGB3A2+ airway progenitor cells
Distal tip, stalk to terminal airway region
C
GHRL/GRP/SFTPC/DAPI
22 pcw
Arrows: GRP+ NEs
Dashed lines: GHRL+ NEs
D
SCGB3A2/SCGB1A1/DAPI
12 pcw
stalk to terminal airway regions
non-cartilaginous airway region
distal tip
Distal tip, stalk, terminal airway to non-cartilaginous airway regions
Arrowheads: SCGB3A2+ airway progenitors
Arrows: SCGB1A1+ SCGB3A2+ Club cells
E
SCGB1A1/MUC16/SCGB3A1/DAPI
15 pcw
Proximal cartilaginous airway region
Arrowheads: SCGB1A1+ SCGB3A1+ Prox Secretory 1
Arrows: SCGB1A1+ SCGB3A1+ Prox Secretory 2
F
SCGB1A1/MUC16/SCGB3A2/DAPI
15 pcw
Proximal cartilaginous airway region
Arrowheads: SCGB3A2+ SCGB1A1+ MUC16 Prox Secretory 2
Arrows/dashed lines: SCGB3A2+ SCGB1A1+ MUC16+ Prox Secretory 3
G
LTF/SCGB3A1/SCGB3A2/DAPI
15 pcw
SMGs
SMGs
Proximal cartilaginous airway region
Arrows: LTF+ SCGB3A1+ submucosal gland cells (SMGs)
H
FOXJ1/MUC16/SCGB3A2/DAPI
19 pcw
Proximal non-cartilaginous airway region
Arrowheads: FOXJ1+ SCGB3A2+ Ciliated cells
Arrows: FOXJ1+ SCGB3A2+ Club cells
I
FOXJ1/MUC16/SCGB3A2/DAPI
MUC16+ ciliated cells
MUC16+ ciliated cells
19 pcw
Proximal cartilaginous airway region
Dashed circle: FOXJ1+ SCGB3A2+ MUC16+ ciliated cells
Arrows: FOXJ1+ SCGB3A2+ MUC16+ secretory cells
J
TP63/F3/IGFBP3/DAPI
19 pcw
Proximal cartilaginous airway region
Dashed lines: TP63+ F3+ IGFBP3 basal cells
J'
TP63/SCGB3A2/SCGB1A1/DAPI
19 pcw
Proximal non-cartilaginous airway region
K
ASCL1/SCGB3A2/MUC5AC/DAPI
12 pcw
Proximal non-cartilaginous airway region
Arrow: ASCL1+ pulmonary NEs
Arrowheads: MUC5AC+ ASCL1+ progenitors
L
Early/Mid stage
Late stage
Proximal non-cartilaginous airway region
Proximal cartilaginous airway region
Distal non-cartilaginous airway region
Terminal airway region
Distal tip region
Stalk region
SOX9+ distal tip ~13-14 pcw
GRP+ pulmonary NEs
CYTL1+ SCGB3A2+ airway progenitor
GHRL+ neuroendocrine cells
CYTL1+ SCGB3A2+ airway progenitor
TP63+ basal cells
SCGB1A1+/ SCGB3A2+/SCGB3A1+/high Proximal Secretory 2
FOXJ1+ ciliated cells
SCGB3A2+ SCGB3A1+ SCGB1A1+ NDUFA4L2+ Proximal secretory progenitors
FOXJ1+/low/MUC16+ multiciliated cells
LTF+ SCGB3A1+ submucosal gland (SMGs)
Cartilage
15-19 pcw
SPOCK2+ AT1
CYTL1+ SCGB3A2+ airway progenitor
SFTPC+ AT2
19-22 pcw
SOX9+ SFTPC+ distal tip
MUC5AC+ ASCL1+ progenitors
CYTL1+ FMO2+ SCGB1A1+/SCGB3A2+ Club cells
SCGB1A1high/+ / SCGB3A2low/- /SCGB3A1+ Proximal Secretory 3
SCGB1A1low/- / SCGB3A2+/SCGB3A1+ Proximal Secretory 1

Figure S5

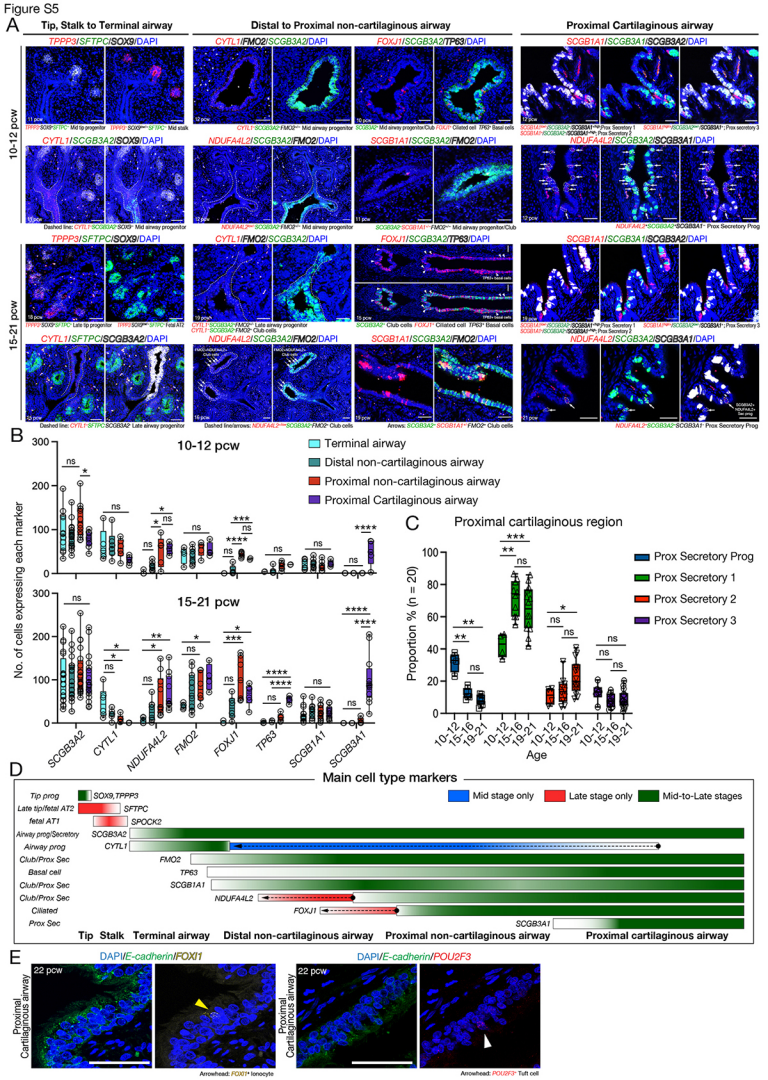

A
Tip, Stalk to Terminal airway
Distal to Proximal non-cartilaginous airway
Proximal Cartilaginous airway
10-12 pcw
15-21 pcw
B
10-12 pcw
15-21 pcw
Terminal airway
Distal non-cartilaginous airway
Proximal non-cartilaginous airway
Proximal Cartilaginous airway
No. of cells expressing each marker
SCGB3A2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

Figure S6

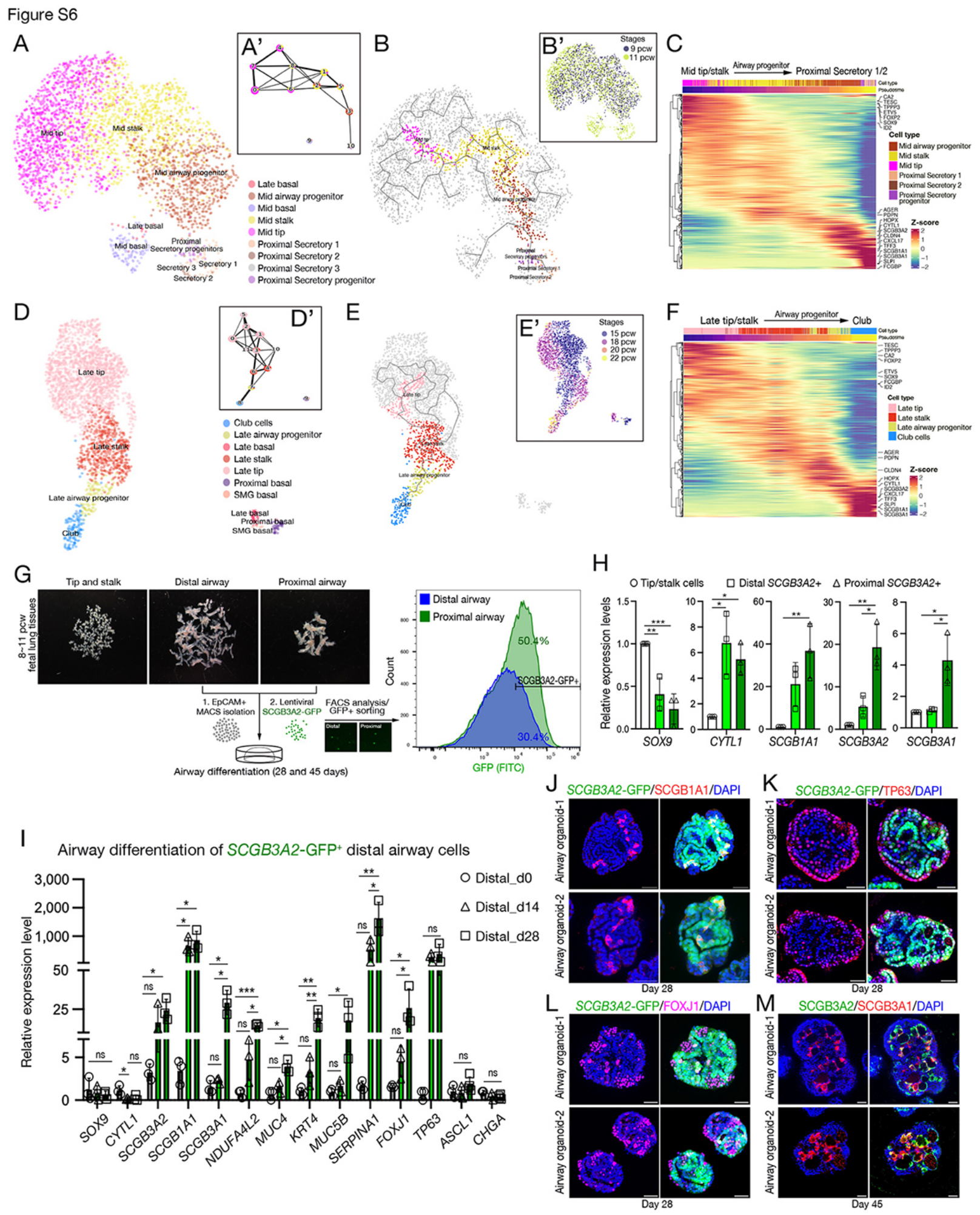

A
A'
B
B'
Stages
9 pcw
11 pcw
C
Mid tip/stalk
Airway progenitor
Proximal Secretory 1/2
Cell type
Pseudotime
Cell type
Mid airway progenitor
Mid stalk
Mid tip
Proximal Secretory 1
Proximal Secretory 2
Proximal Secretory progenitor
Z-score
Mid tip
Mid stalk
Mid airway progenitor
Late basal
Mid basal
Proximal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

Figure S7

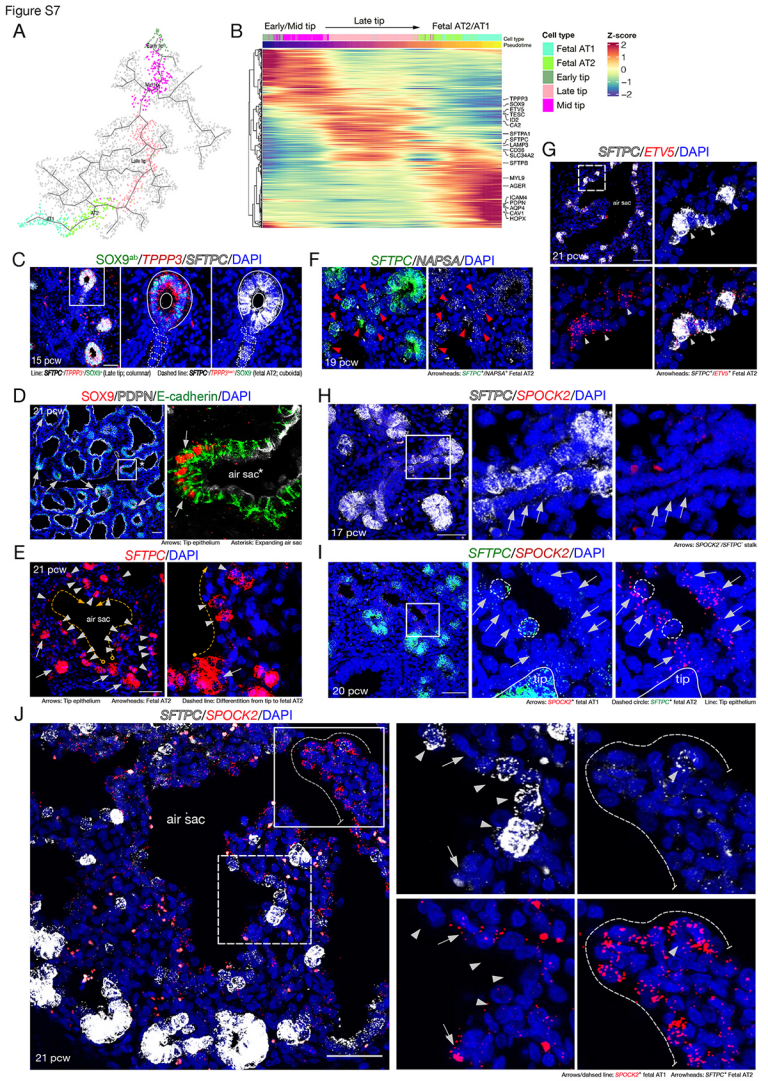

A
Early tip
Late tip
AT1
AT2
B
Early/Mid tip
Late tip
Fetal AT2/AT1
Cell type
Pseudotime
TPPP3
SOX9
ETV5
TESC
IGS
CA2
SFTPA1
SFTPC
LAMP3
CO2
SLC8A2
SFTPB
MYL9
AGER
ICAM4
PDPN
AQP4
CAV1
HOPX
Cell type
Fetal AT1
Fetal AT2
Early tip
Late tip
Mid tip
Z-score
2
1
0
−1
−2
C
SOX9ab/TPPP3/SFTPC/DAPI
15 pcw
Line: SFTPC+/TPPP3+/SOX9+ (Late tip; columnar)
Dashed line: SFTPC+/TPPP3−/SOX9− (fetal AT2; cuboidal)
F
SFTPC/NAPSA/DAPI
19 pcw
Arrowheads: SFTPC+/NAPSA+ Fetal AT2
G
SFTPC/ETV5/DAPI
air sac
21 pcw
Arrowheads: SFTPC+/ETV5+ Fetal AT2
D
SOX9/PDPN/E-cadherin/DAPI
21 pcw
air sac*
Arrows: Tip epithelium
Asterisk: Expanding air sac
H
SFTPC/SPOCK2/DAPI
17 pcw
Arrows: SPOCK2+/SFTPC+ stalk
E
SFTPC/DAPI
21 pcw
air sac
Arrows: Tip epithelium
Arrowheads: Fetal AT2
Dashed line: Differentition from tip to fetal AT2
I
SFTPC/SPOCK2/DAPI
20 pcw
tip
tip
Arrows: SPOCK2+ fetal AT1
Dashed circle: SFTPC+ fetal AT2
Line: Tip epithelium
J
SFTPC/SPOCK2/DAPI
air sac
21 pcw
Arrows/dahsed line: SPOCK2+ fetal AT1
Arrowheads: SFTPC+ Fetal AT2

Figure S8

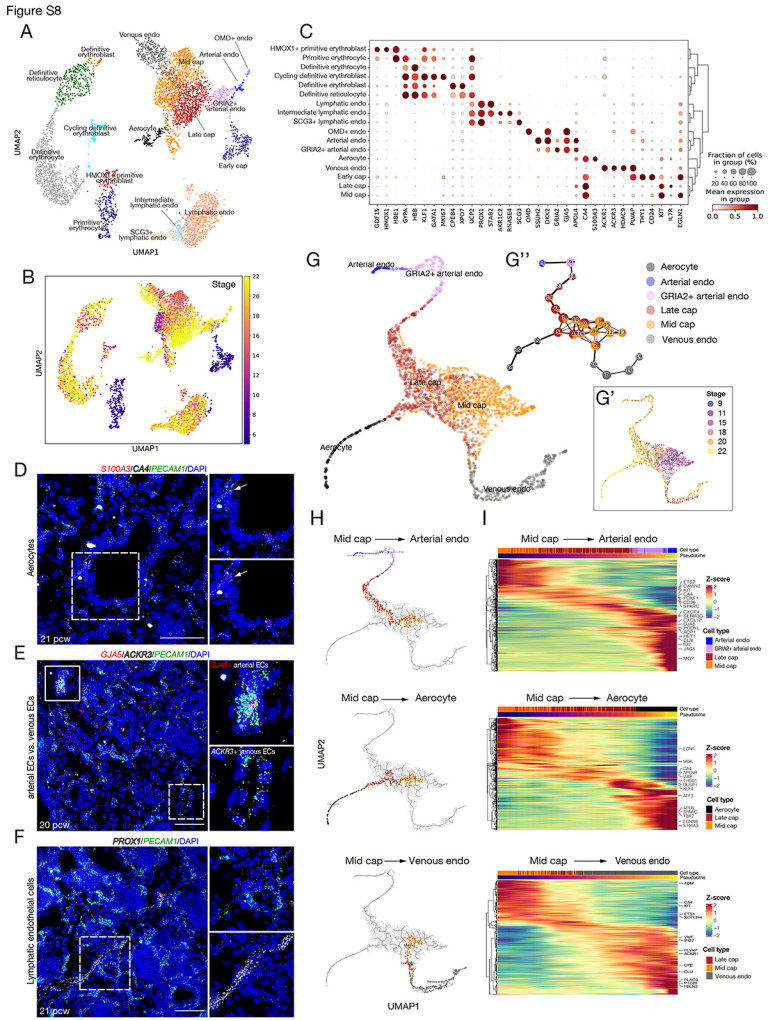

A
Venous endo
OMD+ endo
Definitive erythroblast
Definitive reticulocyte
Arterial endo
Mid cap
GRIA2+ arterial endo
Cycling definitive erythroblast
Aerocyte
Late cap
Early cap
Primitive erythrocyte
HMOX1+ primitive erythroblast
Intermediate lymphatic endo
Lymphatic endo
Primitive erythrocyte
SCG3+ lymphatic endo
UMAP1
UMAP2
B
Stage
UMAP1
UMAP2
C
HMOX1+ primitive erythroblast
Primitive erythrocyte
Definitive erythrocyte
Cycling definitive erythroblast
Definitive erythroblast
Definitive reticulocyte
Lymphatic endo
Intermediate lymphatic endo
SCG3+ lymphatic endo
OMD+ endo
Arterial endo
GRIA2+ arterial endo
Aerocyte
Venous endo
Early cap
Late cap
Mid cap
GDF15
HMOX1
HBE1
HBZ
HBB
KLF1
GATA1
MKI67
CPB84
SLC4A1
UCP2
PROX1
STMN2
ARX
ARL4D
RNASE4
CHGB
CHGA
SSUH2
GRIA2
GRIA1
GJA5
APOA1
CA4
S100A3
ACKR3
SLC9A9
ACKR1
HOMER2
HOMER3
PLVAP
VWF
CD24
KIT
IL32
EGFL7
Fraction of cells in group (%)
20 40 60 80100
Mean expression in group
0.0
0.5
1.0
G
Arterial endo
GRIA2+ arterial endo
Late cap
Mid cap
Aerocyte
Venous endo
G''
Aerocyte
Arterial endo
GRIA2+ arterial endo
Late cap
Mid cap
Venous endo
G'
Stage
9
11
15
18
20
22
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
H
Mid cap → Arterial endo
Mid cap → Aerocyte
Mid cap → Venous endo
UMAP2
UMAP1
I
Mid cap → Arterial endo
Cell type
Pseudotime
Z-score
Cell type
Arterial endo
GRIA2+ arterial endo
Late cap
Mid cap
Mid cap → Aerocyte
Aerocyte
Mid cap → Venous endo
Late cap
Mid cap
Venous endo

Figure S9

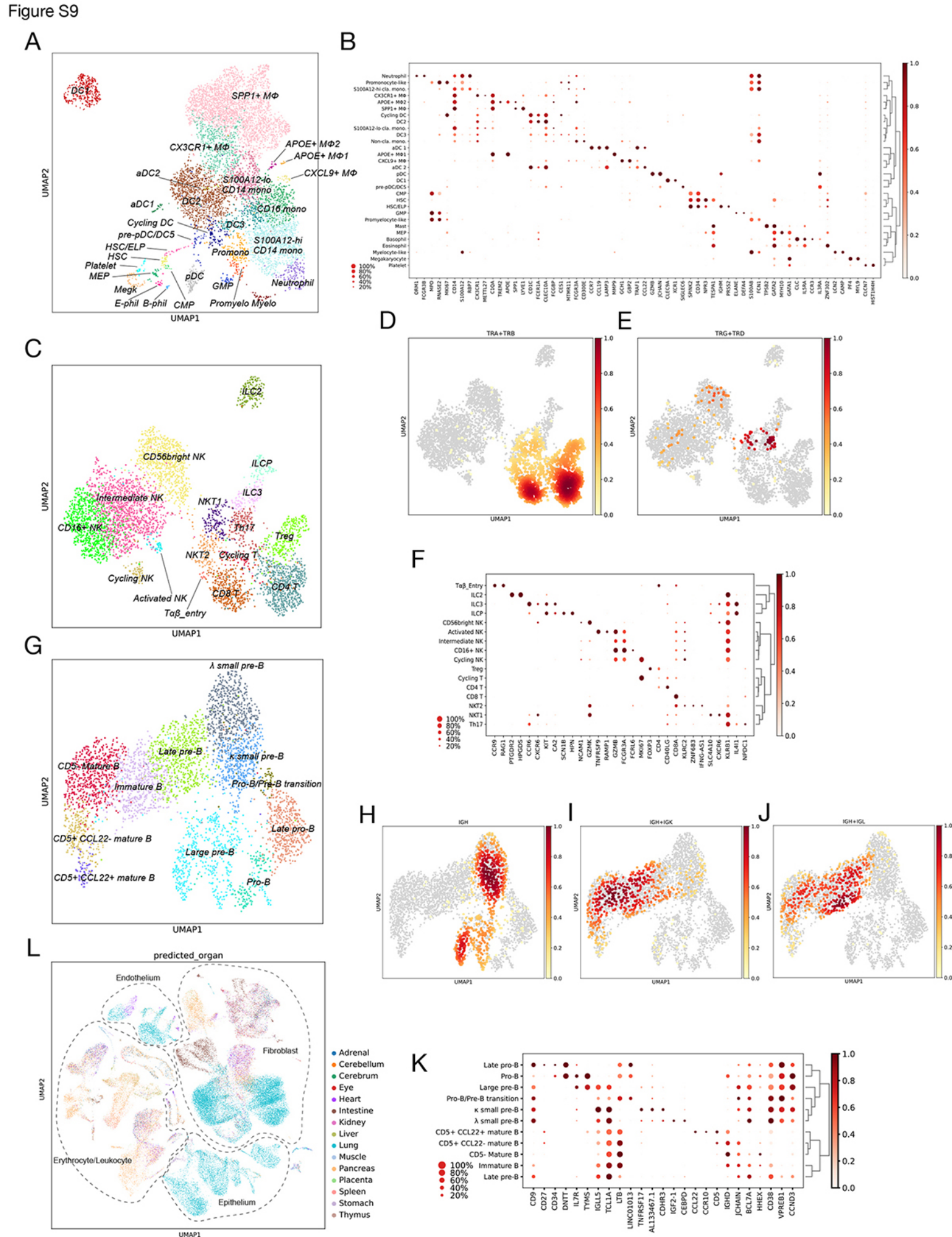
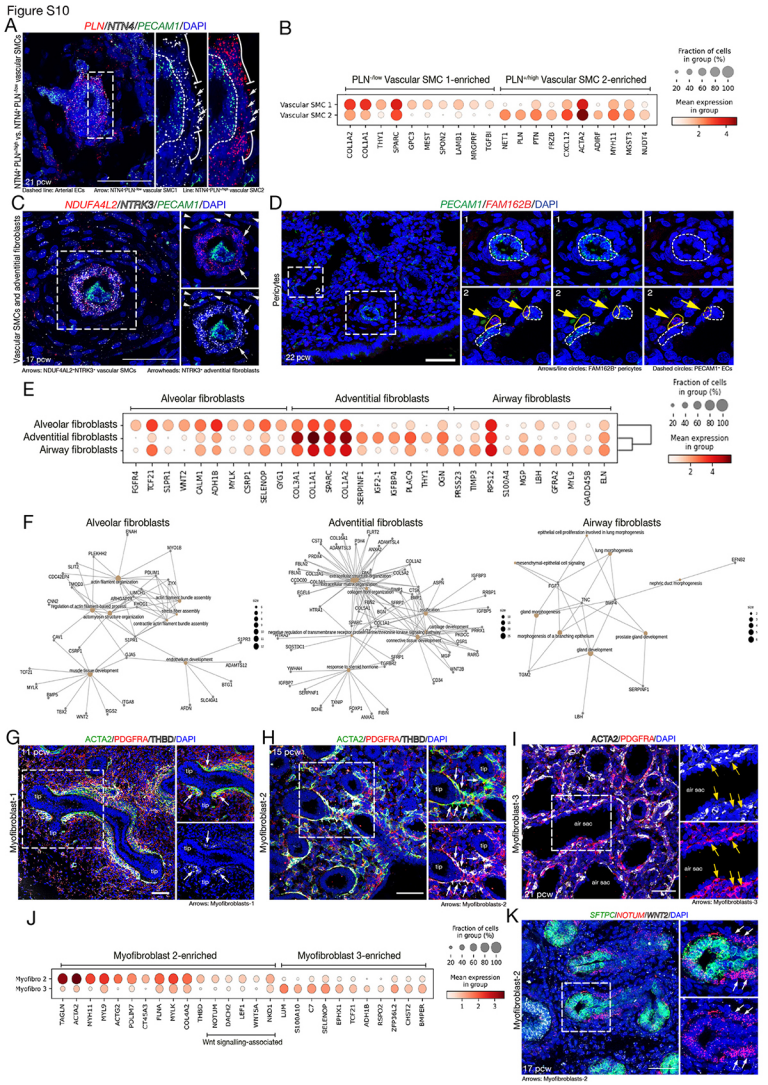

A
DC1
SPP1+ MΦ
CX3CR1+ MΦ
APOE+ MΦ2
APOE+ MΦ1
CXCL9+ MΦ
aDC2
S100A12-lo CD14 mono
DC2
aDC1
CD16 mono
Cycling DC
DC3
pre-pDC/DC5
S100A12-hi CD14 mono
HSC/ELP
Promono
HSC
Platelet
MEP
pDC
GMP
Neutrophil
Megk
E-phil
B-phil
CMP
Promyelo
Myelo
UMAP2
UMAP1
B
Neutrophil
Promonocyte-like
S100A12-hi cla. mono.
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
UMAP2
UMAP1
E
TRG+TRD
UMAP2
UMAP1
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%
CCR9
RAG1
PTGDR2
HPGDS
CCR6
CXCR6
KIT
CA2
SCN1B
HPN
NCAM1
GZMK
TNFRSF9
RAMP1
GZMB
FCGR3A
FCRL6
MKI67
FOXP3
CD4
CD40LG
CD8A
KLRC2
ZNF683
IFNG-AS1
SLC4A10
CXCR6
KLRB1
IL4I1
NPDC1
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
UMAP2
UMAP1
I
IGH+IGK
UMAP2
UMAP1
J
IGH+IGL
UMAP2
UMAP1
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%
CD9
CD27
CD34
DNTT
IL7R
TYMS
IGLL5
TCL1A
LTB
LINC01013
TNFRSF17
AL133467.1
CDHR3
IGF2-1
CEBPD
CCL22
CCR10
CD5
IGHD
JCHAIN
BCL7A
HHEX
CD38
VPREB1
CCND3

Figure S10



A
PLN/NTN4/PECAM1/DAPI
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
ADRF
MYH11
MGST3
NUDT4
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
2 4
C
NDUFA4L2/NTRK3/PECAM1/DAPI
Vascular SMCs and adventitial fibroblasts
17 pcw
Arrows: NDUFA4L2-NTRK3- vascular SMCs
Arrowheads: NTRK3- adventitial fibroblasts
D
PECAM1/FAM162B/DAPI
Pericytes
22 pcw
Arrows/line circles: FAM162B- pericytes
Dashed circles: PECAM1- ECs
E
Alveolar fibroblasts
Adventitial fibroblasts
Airway fibroblasts
FGFR4
TCF21
SLPR1
WNT2
CALM1
ADH1B
MYLK
CSRP1
SELENOP
GYG1
COL3A1
COL1A1
SPARC
COL1A2
SERPINF1
IGF2-1
IGFBP4
PLAC9
THY1
OGN
PRSS23
TIMP3
RPS12
S100A4
MGP
LBH
GFRA2
MYL9
GADD45B
ELN
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
2 4
F
Alveolar fibroblasts
Adventitial fibroblasts
Airway fibroblasts
G
ACTA2/PDGFRA/THBD/DAPI
Myofibroblast-1
11 pcw
tip
Arrows: Myofibroblasts-1
H
ACTA2/PDGFRA/THBD/DAPI
Myofibroblast-2
15 pcw
tip
Arrows: Myofibroblasts-2
I
ACTA2/PDGFRA/DAPI
Myofibroblast-3
air sac
21 pcw
Arrows: Myofibroblasts-3
J
Myofibroblast 2-enriched
Myofibroblast 3-enriched
Myofibro 2
Myofibro 3
TAGLN
ACTA2
MYH11
MYL9
ACTG2
POLR2F
CT45A3
FLNA
MYLK
COL4A2
THBD
NOTUM
DACH2
LEF1
WNT5A
NND1
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
K
SFTPC/NOTUM/WNT2/DAPI
Myofibroblast-2
17 pcw
Arrows: Myofibroblasts-2

Figure S11

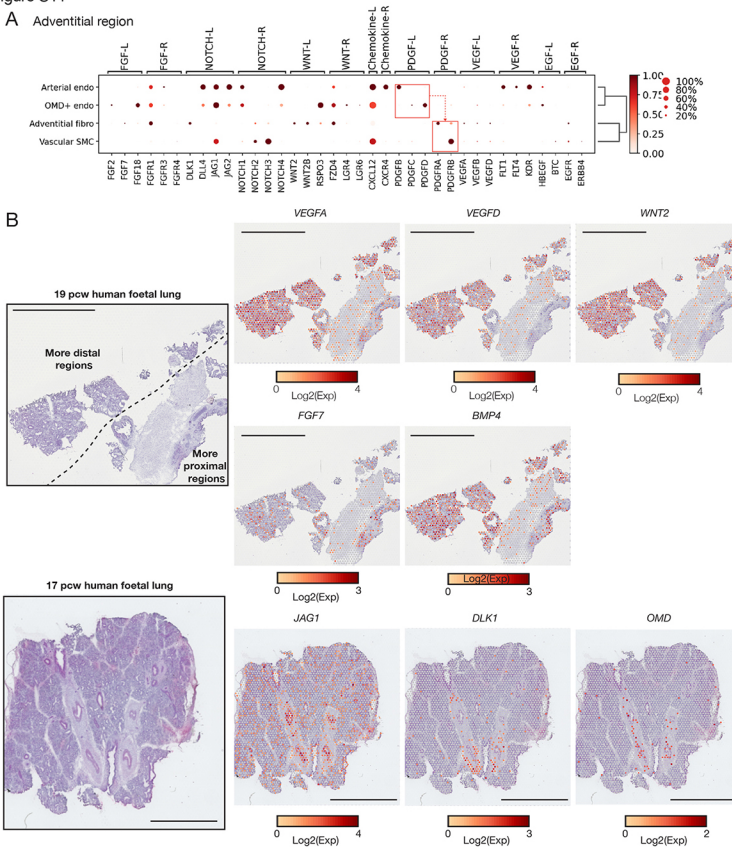

A
Adventitial region
FGF-L
FGF-R
NOTCH-L
NOTCH-R
WNT-L
WNT-R
Chemokine-L
Chemokine-R
PDGF-L
PDGF-R
VEGF-L
VEGF-R
EGF-L
EGF-R
Arterial endo
OMD+ endo
Adventitial fibro
Vascular SMC
FGF2
FGF7
FGF18
FGFR1
FGFR3
FGFR4
DLK1
DLL4
JAG1
JAG2
NOTCH1
NOTCH2
NOTCH3
NOTCH4
WNT2
WNT2B
RSPO3
FZD4
LGR4
LGR6
CXCL12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

Figure S12

A

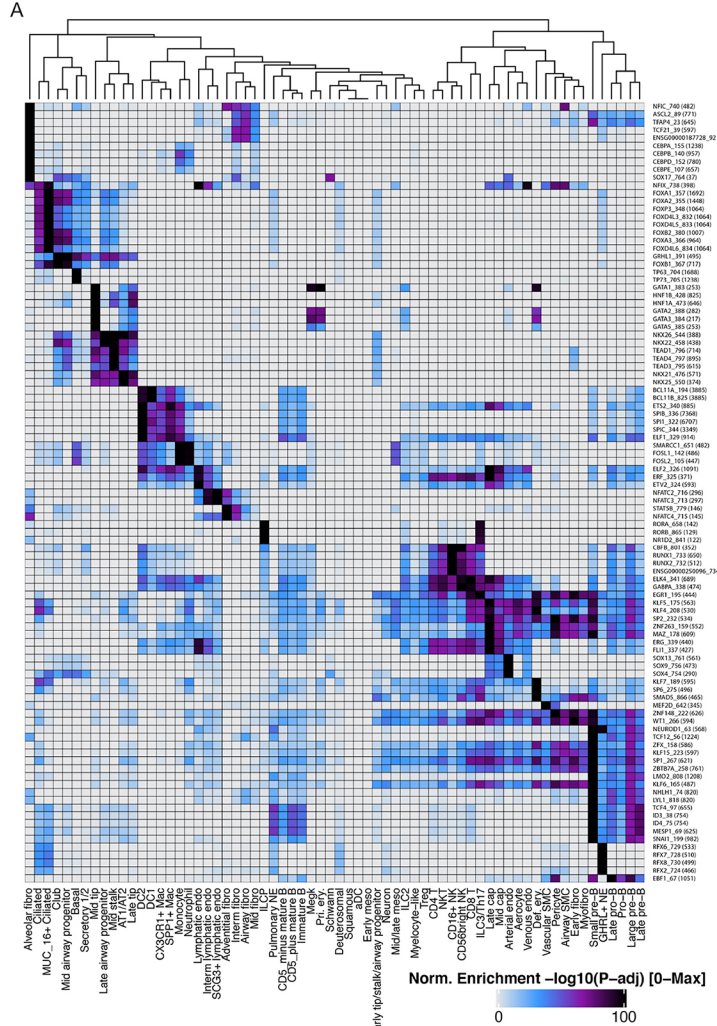

NFIC_740 (482)
ASCL2_89 (771)
TFAP4_23 (645)
TCF21_39 (597)
ENSG00000187728_92 (597)
CEBPA_155 (1238)
CEBPB_140 (957)
CEBPD_152 (780)
CEBPE_107 (657)
SOX17_764 (37)
NFIX_718 (398)
FOXA1_357 (1692)
FOXA2_355 (1448)
FOXP3_348 (1064)
FOXD6L3_832 (1064)
FOXD4L5_839 (1064)
FOXB2_380 (1007)
FOXA3_366 (964)
FOXD4L6_834 (1064)
GRHL1_391 (695)
FOXB1_367 (717)
TP63_784 (1688)
TP73_705 (1238)
GATA1_383 (253)
HNF1B_428 (825)
HNF1A_473 (646)
GATA2_388 (282)
GATA3_384 (217)
GATA5_385 (253)
NKX26_544 (388)
NKX22_458 (438)
TEAD1_796 (714)
TEAD4_797 (895)
TEAD3_795 (615)
NKX21_476 (571)
NKX25_550 (374)
BCL11A_194 (3885)
BCL11B_825 (3885)
ETS2_340 (885)
SPIB_336 (7368)
SPI1_322 (6707)
SPC_344 (3349)
ELF1_329 (914)
SMARCC1_651 (462)
FOSL1_142 (486)
FOSL2_105 (447)
ELF3_326 (1691)
ERE_325 (371)
ETV2_324 (593)
NFATC2_716 (296)
NFATC3_713 (297)
STAT5B_779 (146)
NFATC4_715 (145)
RORA_658 (142)
RORB_865 (129)
NR1D2_841 (122)
CBFB_891 (352)
RUNX1_733 (650)
RUNX2_732 (512)
ENSG00000250996_734 (512)
ELK4_341 (689)
GABPA_338 (474)
EGR3_195 (444)
KLF5_175 (563)
KLF4_208 (530)
SP2_232 (534)
ZNF263_159 (552)
MAZ_178 (669)
ERG_339 (440)
FLI1_337 (427)
SOX13_761 (561)
SOX9_756 (473)
SOX4_754 (290)
KLF7_189 (595)
SP6_275 (496)
SMAD5_866 (465)
MEF2D_642 (345)
ZNF148_222 (626)
WT1_266 (594)
NEUROD1_63 (568)
TCF12_56 (1224)
ZFX_158 (586)
KLF15_223 (597)
SP1_267 (621)
ZBTB7A_258 (761)
LMO2_808 (1208)
KLF6_165 (487)
NHLH1_74 (820)
LVL1_818 (820)
TCF4_97 (655)
ID3_38 (754)
ID4_75 (754)
MESP1_69 (625)
SNAI1_199 (982)
RFX6_729 (533)
RFX7_728 (510)
RFX8_730 (499)
RFX2_724 (466)
EBF1_67 (1051)
Alveolar fibro
Ciliated
MUC_16+ Ciliated
Ciliated
Mid airway progenitor
Basal
Secretory 22
Mid tip
Late airway progenitor
Mid alveolar
AT1/AT2
Late tip
DC2
DC1
CX3CR1+ Mac
SPP1+ Mac
Monocyte
Neutrophil
Lymphatic endo
Interm lymphatic endo
SCG3+ lymphatic endo
Adventitial fibro
Interm fibro
Airway fibro
Mid ILCP
ILCP
Pulmonary NE
CD5_immature B
CD5_mature B
Immature B
Plasma
Pre-Myelin
Schwann
Deuterosomal
Suprabasal
Early meso
Early tip/stalk/airway progenitor
Neuron
Mid/late meso
Myeloid
Myelocyte–like
CD4+
CD4T
NKT
CD16+ NK
CD56bright NK
CD8 T
ILC3/Th17
Mid cap
Arterial endo
Mid vein
Venous endo
Del-sev
Vascular SMC
Pericyte
Airway SMC
Myofibro
Small pre-B
Pro-B
Late pro-B
Large pre-B
Late pre-B
Norm. Enrichment −log10(P−adj) [0–Max]
0
100

Figure S13

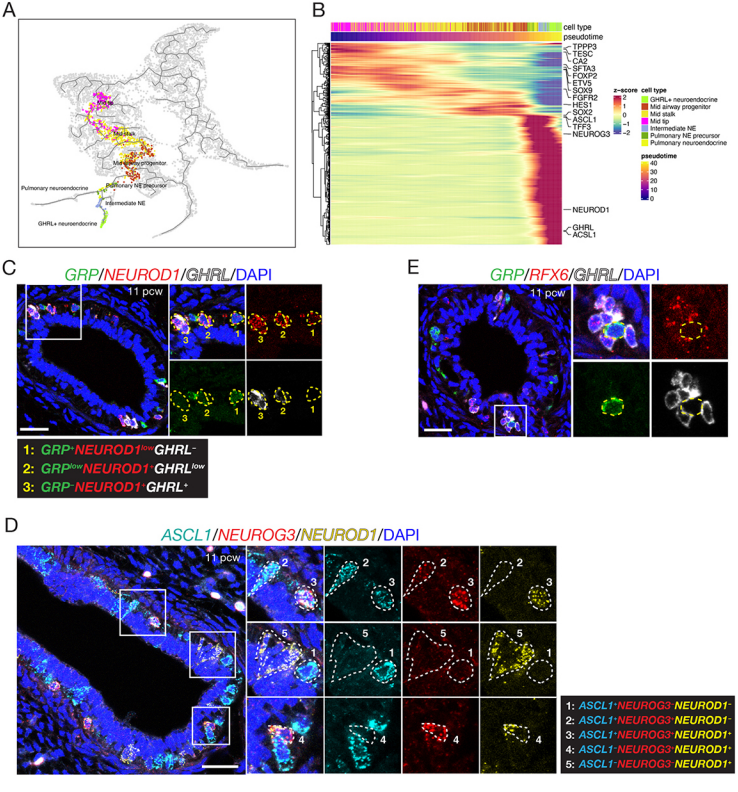
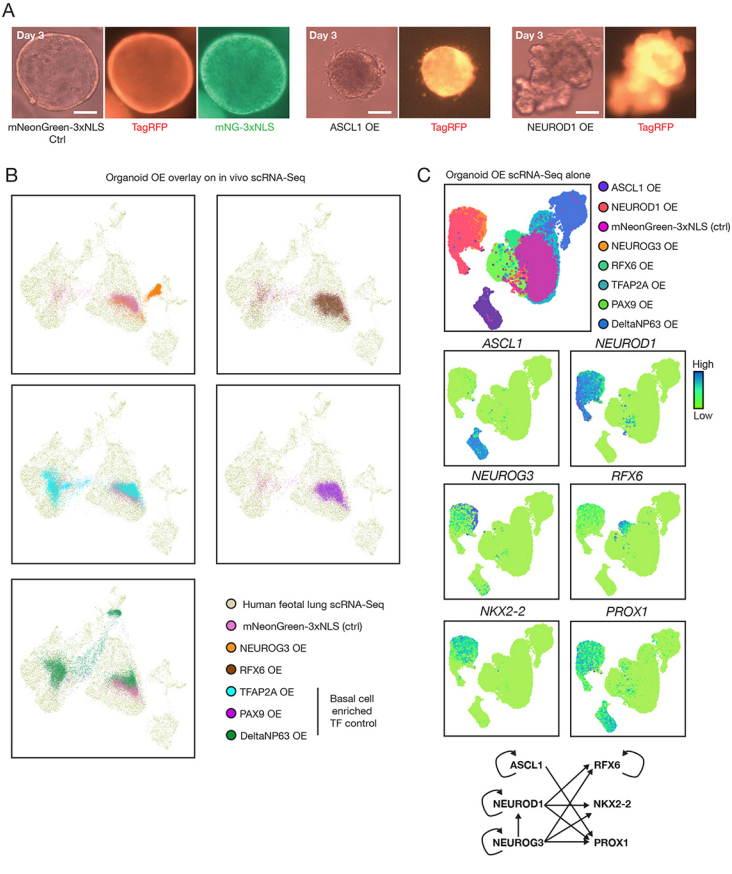

A
Mid stem
Mid airway progenitor
Pulmonary neuroendocrine
Pulmonary NE precursor
Intermediate NE
GHRL+ neuroendocrine
B
cell type
pseudotime
TPPP3
TESC
CA2
SFTA3
FOXP2
ETV5
SOX9
FGFR2
HES1
SOX2
ASCL1
TFF3
NEUROG3
NEUROD1
GHRL
ACSL1
z-score
2
1
0
−1
−2
cell type
GHRL+ neuroendocrine
Mid airway progenitor
Mid stalk
Mid tip
Intermediate NE
Pulmonary NE precursor
Pulmonary neuroendocrine
pseudotime
40
30
20
10
0
C
GRP/NEUROD1/GHRL/DAPI
11 pcw
1: GRP+NEUROD1low GHRL-
2: GRPlow NEUROD1- GHRLlow
3: GRP- NEUROD1+ GHRL+
E
GRP/RFX6/GHRL/DAPI
11 pcw
D
ASCL1/NEUROG3/NEUROD1/DAPI
11 pcw
1: ASCL1+NEUROG3- NEUROD1-
2: ASCL1+NEUROG3- NEUROD1-
3: ASCL1+NEUROG3- NEUROD1+
4: ASCL1+NEUROG3- NEUROD1+
5: ASCL1- NEUROG3- NEUROD1+

Figure S14


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