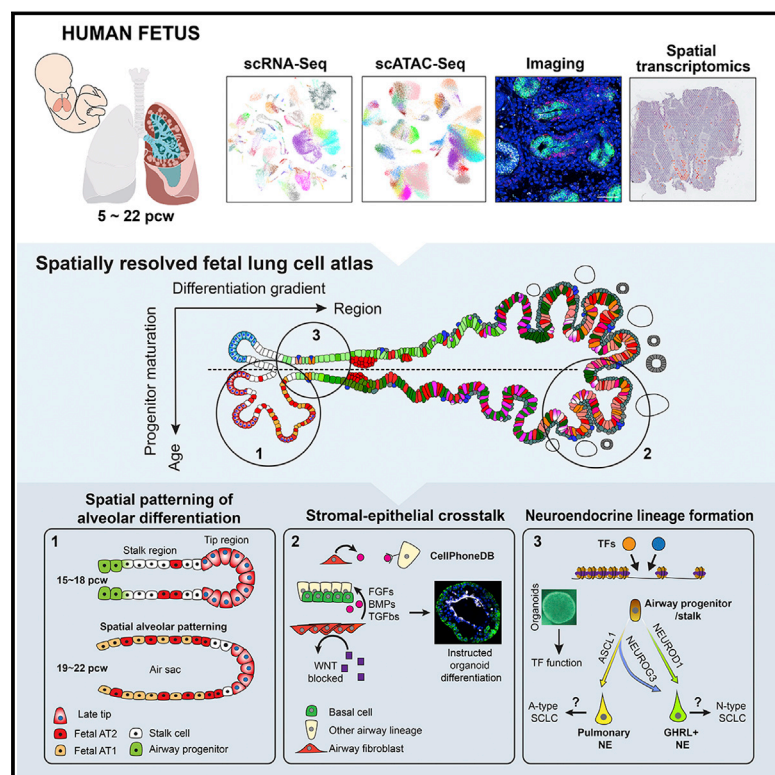


# A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates

## Graphical abstract



## Authors

Peng He, Kyungtae Lim, Dawei Sun, ..., John C. Marioni, Kerstin B. Meyer, Emma L. Rawlins

## Correspondence

e.rawlins@gurdon.cam.ac.uk

## In brief

Multiomic analysis of human fetal lungs from 5–22 post-conception weeks unveils cell-lineage trajectories across different cell types during development and will provide fresh insights into lung disease progression in adults.

## Highlights

- Spatiotemporal atlas of human lung development identifies 144 cell types/states
- Tracking the developmental origins of multiple cell types, including new progenitors
- Functional diversity of fibroblasts in distinct anatomical signaling niches
- Experimental validation of TFs controlling neuroendocrine cell heterogeneity



## Resource

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

Peng He,<sup>1,2,15</sup> Kyungtae Lim,<sup>3,15</sup> Dawei Sun,<sup>3,13,15</sup> Jan Patrick Pett,<sup>1</sup> Quitz Jeng,<sup>3</sup> Krzysztof Polanski,<sup>1</sup> Ziqi Dong,<sup>3</sup> Liam Bolt,<sup>1,11</sup> Laura Richardson,<sup>1</sup> Lira Mamanova,<sup>1,12</sup> Monika Dabrowska,<sup>1</sup> Anna Wilbrey-Clark,<sup>1</sup> Elo Madisson,<sup>1,2</sup> Zewen Kelvin Tuong,<sup>1,4</sup> Emma Dann,<sup>1</sup> Chenqu Suo,<sup>1,5</sup> Isaac Goh,<sup>6</sup> Masahiro Yoshida,<sup>7</sup> Marko Z. Nikolić,<sup>7</sup> Sam M. Janes,<sup>7</sup> Xiaoling He,<sup>8</sup> Roger A. Barker,<sup>8</sup> Sarah A. Teichmann,<sup>1,9,14</sup> John C. Marioni,<sup>1,2,10,14</sup> Kerstin B. Meyer,<sup>1,14</sup> and Emma L. Rawlins<sup>3,14,16,\*</sup>

<sup>1</sup>Wellcome Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

<sup>2</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, UK

<sup>3</sup>Wellcome Trust/CRUK Gurdon Institute, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 1QN, UK

<sup>4</sup>Molecular Immunity Unit, University of Cambridge Department of Medicine, Cambridge, UK

<sup>5</sup>Department of Paediatrics, Cambridge University Hospitals, Hills Road, Cambridge CB2 0 QQ, UK

<sup>6</sup>Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK

<sup>7</sup>Lungs for Living Research Centre, UCL Respiratory, University College London, London, UK

<sup>8</sup>John van Geest Centre for Brain Repair, Department of Clinical Neurosciences and Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

<sup>9</sup>Department of Physics, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK

<sup>10</sup>Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK

<sup>11</sup>Present address: Genomics England, Hinxton CB10 1DR, UK

<sup>12</sup>Present address: CS Genomics, Cambridge CB1 2JH, UK

<sup>13</sup>Present address: Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA

<sup>14</sup>Senior author

<sup>15</sup>These authors contributed equally

<sup>16</sup>Lead contact

\*Correspondence: [e.rawlins@gurdon.cam.ac.uk](mailto:e.rawlins@gurdon.cam.ac.uk)  
<https://doi.org/10.1016/j.cell.2022.11.005>

## SUMMARY

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 previously uncharacterized cell states in all compartments. These include developmental-specific secretory progenitors and a subtype of neuroendocrine cell related to human small cell lung cancer. Our datasets are available through our web interface (<https://lungcellatlas.org>). To illustrate its general utility, we use our cell atlas to generate predictions about cell-cell signaling and transcription factor hierarchies which we rigorously test using organoid models.

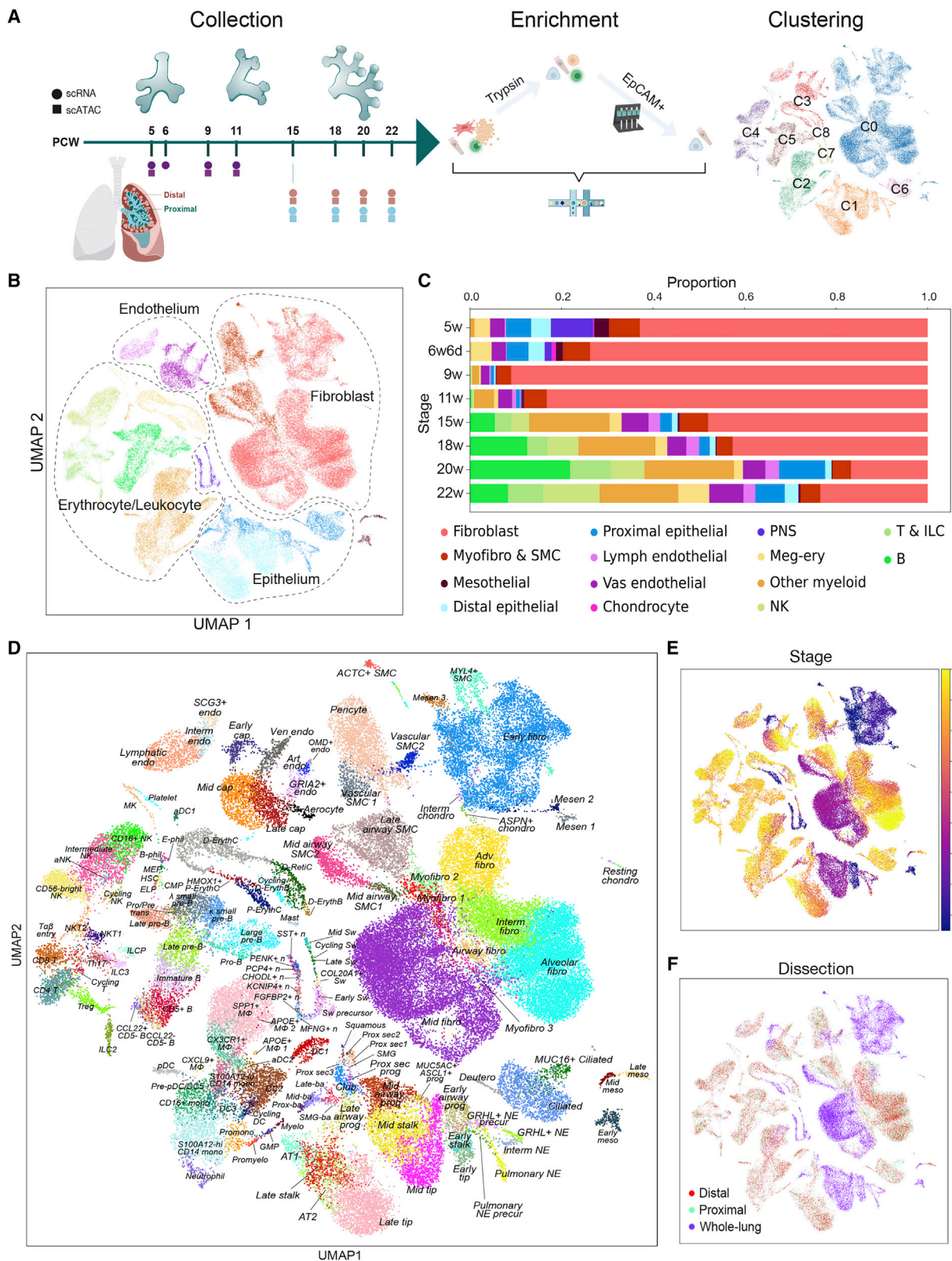
## INTRODUCTION

Single-cell mapping of cell states in the adult human lung in health and disease is being performed at increasing resolution,<sup>1</sup> providing a foundation for understanding lung cellular physiology. The adult lung has low rates of cell turnover,<sup>2,3</sup> making it difficult to capture transition states and progenitors. Moreover, there are developmental-specific cell states that do not exist in the adult. A high-resolution cell atlas of the embryonic and fetal human lung will identify developmental precursors and progeni-

tors and predict differentiation trajectories and potential gene regulatory networks. This will provide a baseline for studying adult homeostasis and disease.

The lung buds are specified in the human foregut endoderm at ~5 post-conception weeks (pcw).<sup>4,5</sup> Subsequent morphogenesis is driven by branching of the distal-most bud tips. The bud tip epithelium comprises SOX9<sup>+</sup>, ID2<sup>+</sup> multipotent progenitors that self-renew during branching.<sup>6–9</sup> As the bud tip epithelium branches into the surrounding mesoderm, the epithelial cells that remain in the stalk region start to differentiate into





(legend on next page)



bronchiolar (airway, ~5–16 pcw) and later (from ~16 pcw) into alveolar epithelium.<sup>5</sup> The pattern of growth from multipotent epithelial progenitors at the distal tips means that the position of a cell along the proximal-distal axis of the lung epithelial tree is a strong predictor of its maturity. The more mature cells, which exited the tip first, are more proximal, whereas the most immature cell states, which exited the tip recently, are found in the tip-adjacent (stalk) regions.<sup>10</sup> In other words, space reflects time in lung development. Therefore, coupling single-cell state analysis to *in vivo* spatial visualization can provide high confidence in the identification of novel progenitor cells in the developing lung. Moreover, detailed spatial analysis of cell states allows cell identity designations to be compared to more traditional histological definitions.

We have generated a high-resolution single-cell atlas of human lung development using a combination of scRNA-seq, scATAC-seq, Visium Spatial Transcriptomics, and mRNA *in situ* hybridization using hybridization chain reaction (HCR).<sup>11</sup> Combining these data sources has allowed us to identify 144 cell states/types in 5–22 pcw lung samples. These include previously uncharacterized progenitor cell states, transition populations, and a subtype of neuroendocrine cell related to a subtype of human small cell lung cancer (SCLC). We observe increasing cell maturation over time, with many cell states identified in adult lungs already present at 22 pcw. We have used our atlas to make predictions about progenitor cell states, signaling interactions, and lineage-defining transcription factors, and we demonstrate how these can be efficiently tested using a genetically tractable human fetal lung organoid model. The datasets 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 (Figure 1A) and produced high-quality transcriptome (Figure S1A; average > 2,400 genes/cell) and DNA accessibility (Figures S1K and S1L; average > 18,000 fragments/nucleus) data. After iterative clustering (Figures S1C and S1D), removal of doublet-driven clusters (Figures S1E, S1G, S1G', and S1G''), stressed or low-quality clusters (except those expressing known markers, such as erythroid) (Figures S1I, S1I', and S1I''), clusters composed of cells from only one sample when replicates are available, and clusters of cells from other organs (Figure S1H)<sup>12,13</sup> and maternal cell evaluation (Figures S1F and S1J), we present 71,752 cells shown as a uniform manifold

approximation and projection (UMAP) (Figure 1A), on which we manually annotated fibroblast, epithelial, endothelial, and erythrocyte/leukocyte lineages (Figure 1B). Plotting the cell-type distribution against time (excluding trypsin/CD326-treated samples, shown in Figure S1B) showed that fibroblasts were the most prominent cell, particularly in younger lungs (Figure 1C). Leukocytes and erythrocytes were observed in all lungs sampled, with B, T, and NK cells becoming prominent from 15 pcw (Figure 1C).

Further cell-type annotation was performed based on marker genes (Table S1), resulting in assignment of 144 cell types/states (Figure 1D). Sample age was a strong determinant of clustering ( $\chi^2 = 163,727$ ,  $p \approx 0$ ), reflecting progressive cell maturity over time (Figure 1E). Clusters mostly grouped into three distinct regions which we categorized as early (5, 6 pcw), mid (9–11 pcw), and late (15–22 pcw) stages. Cell cycle phase (Figure S1M,  $\chi^2 = 25,361$ ,  $p \approx 0$ ) and dissected region (Figure 1F,  $\chi^2 = 968$ ,  $p = 8.9E-131$ ) were also associated with clustering. However, the dissection region was only prominent for a small number of proximally located cell types (Figure 1F), suggesting that most proximal-to-distal regions of the airway structure were still represented in both dissected regions of the lung. Epithelial cells were mostly derived from the trypsin-treated and CD326-enriched samples, although airway smooth muscle, myofibroblasts, and alveolar fibroblasts were also enriched here (Figure S1M'). Peripheral nervous system (PNS) cells and chondrocytes were only obtained from 5–6 pcw lungs, likely correlating with lower extracellular matrix (ECM) complexity in younger lungs and/or increased fragility of older neurons. PNS cells were clustered and assigned to cell types, but scarcity precluded further analysis (Figure 1D, S1N, and S1N'). Data integration and logistic regression-based comparison showed that gene expression of our annotated cells corresponds well to those of adult lungs<sup>14</sup> (Figures S2A–S2C).

### A differentiation trajectory of airway progenitor states lies along the developing lung distal-to-proximal axis

The epithelial cells separate by age (Figures 2A and 2B), with many basal cells, MUC16<sup>+</sup> ciliated cells, and secretory cells enriched in the proximally dissected tissue (Figures 2B and S1O). The most immature epithelial progenitors are tip cells: SOX9<sup>+</sup> multipotent progenitors located at the distal branching tips of the respiratory tree.<sup>8</sup> Tip cells were separated into early (5, 6 pcw), mid (9–11 pcw) and late (15–22 pcw) populations (Figures 2A and 2B) with both shared and stage-specific markers (Figure 2C). On the epithelial UMAP, each tip population clusters closely with adjacent stalk cells (SOX9<sup>LO/-</sup>, PDPN<sup>LO</sup>, HOPX<sup>LO</sup>) and airway progenitors (CYTL1<sup>LO/+</sup>, PCP4<sup>+</sup>, SCGB3A<sup>+LO</sup>) (Figure 2A). The tip, stalk, and airway progenitors can be visualized in a distal-proximal sequence in the tissue at all stages tested (10–16 pcw) (Figures 2E, S3A, and S1B, Video S1), consistent with the most proximal cells being the most mature. These three cell types form a predicted differentiation trajectory from mid-tip

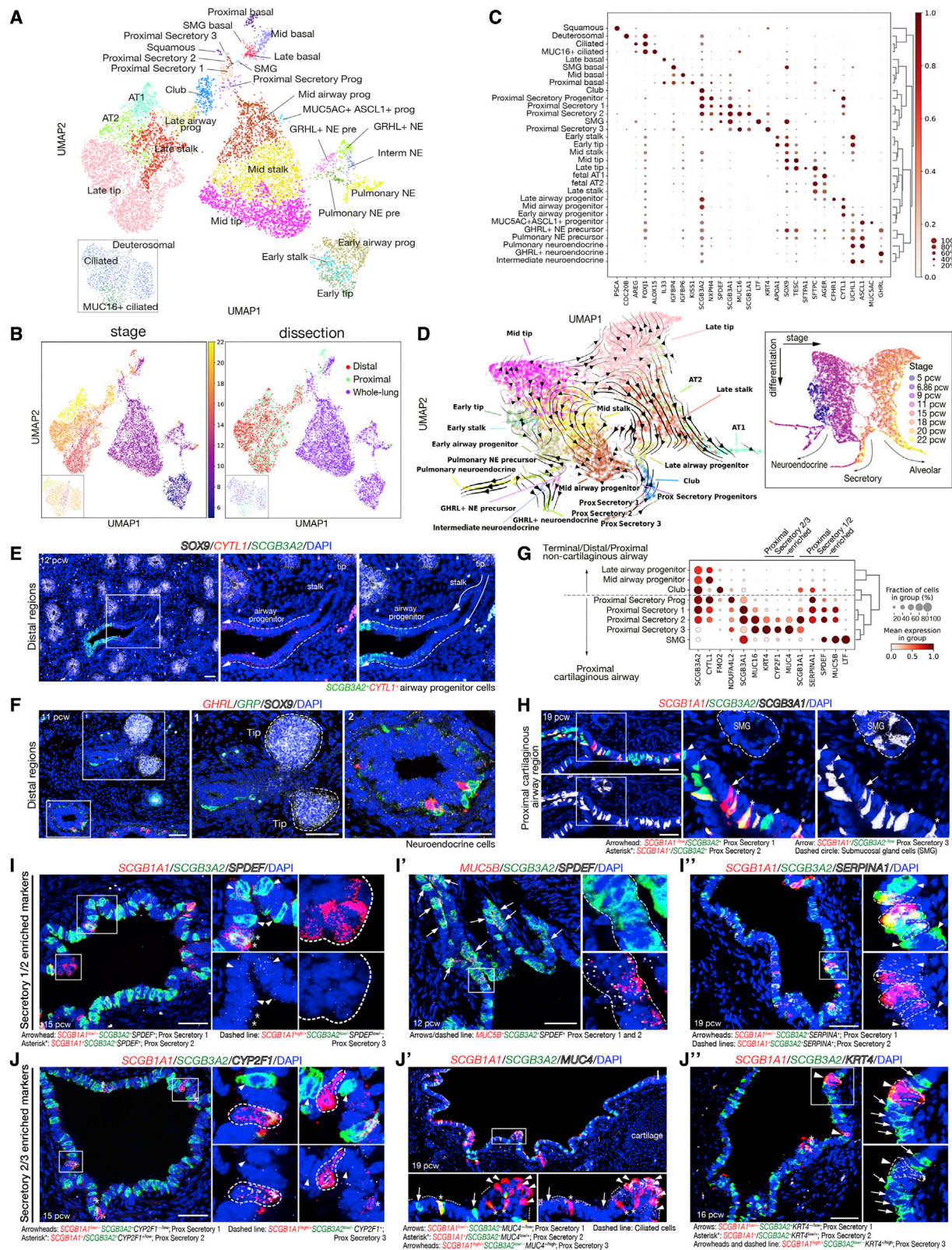
#### Figure 1. Data and experimental overview

(A) Overview of sample collection for scRNA-seq (circles) and scATAC-seq (squares) experiments from whole lung (purple), distal (red), and proximal (blue) regions, cell processing and broad clustering; cluster number refers to the data portal (<https://lungcellatlas.org>).

(B and C) UMAP representation (B) and cell-type proportion (C) of 71,752 good-quality cells, indicating epithelial, endothelial, fibroblast, and leukocyte/erythroid compartments.

(D–F) UMAP visualization by cell type/state (D), developmental stage (E), and dissection region (F). See also Figures S1 and S2.





(legend on next page)

to mid-stalk to mid-airway progenitor that branches into the neuroendocrine, or secretory, lineages (Figure 2D).

### Two subtypes of neuroendocrine cells are present in the developing airways

Consistent with previous data,<sup>15</sup> the earliest differentiated epithelial cells detected were neuroendocrine (NE) cells in 5 pcw lungs (Figures 2A–2C). We identified two types of NE cells: classical pulmonary NE cells (*GRP*<sup>+</sup>) and GHRL<sup>+</sup> NE cells (*TTR*<sup>+</sup>, *GHRL*<sup>+</sup>) in agreement with a recent human fetal cell atlas.<sup>13</sup> We observed increasing maturity of NE cells over time (specific populations denoted as precursors on the UMAP). In addition, an intermediate NE population, a putative transition state, connected the two NE cells (Figure 2A). At 11 pcw, *GRP*<sup>+</sup> pulmonary NE cells were observed closer to the budding tips, suggesting that they begin to form prior to the *GHRL*<sup>+</sup> NE cells (Figure 2F). This spatial difference was not apparent in the oldest samples where both *GRP*<sup>+</sup> and *GHRL*<sup>+</sup> cells were observed at all airway levels, although less abundant distally (Figure S3C). Mouse *Ghrh*<sup>+</sup> NE cells were not detected in re-analysis of published mouse data,<sup>16,17</sup> or spatially.<sup>18</sup> However, *Ghrh* is expressed in mouse ciliated cells that cluster with human fetal GHRL<sup>+</sup> NE cells (Figure S2D).<sup>17</sup>

### Multiple secretory cell subtypes in the proximal cartilaginous airways

We annotated 5 sub-types of differentiating secretory cells and one proximal secretory progenitor. (1) The proximal secretory progenitors (*SCGB3A2*<sup>+</sup>, *SCGB1A1*<sup>+</sup>, *SCGB3A1*<sup>L/O</sup>, *CYTL1*<sup>+</sup>) were detected in the single-cell atlas at 9 pcw, prominent at 11 pcw, but rarer in older lungs consistent with a progenitor state (Figures 2A–2C, and 2G). (2) Club cells (*SCGB3A2*<sup>+</sup>, *SCGB1A1*<sup>+</sup>, *SCGB3A1*<sup>+</sup>, *SPDEF*<sup>+</sup>, *MUC16*<sup>+</sup>) were detected from 15 pcw in the single-cell data (Figures 2A–2C, and 2G), or 12 pcw in the tissue localized in clusters more distally, but dispersed in the more proximal non-cartilaginous regions (Figure S3D). (3) Submucosal gland (SMG) secretory cells (*LTF*<sup>+</sup>, *SCGB3A1*<sup>+</sup>, *SPDEF*<sup>+</sup>) were detected from 15 pcw in the single-cell data, located in SMG ducts and likely to be a precursor of serous and/or mucous-secreting SMG cells (Figures 2A–2C, 2G, and S3G). (4) Proximal secretory 1 (*SCGB1A1*<sup>L/O</sup>, *SCGB3A2*<sup>+</sup>, *SCGB3A1*<sup>+</sup>) and (5) proximal secretory 2 (*SCGB1A1*<sup>+</sup>, *SCGB3A2*<sup>+</sup>, *SCGB3A1*<sup>+</sup>) appeared from 11 pcw (Figure 2A–2C, 2G–2H, and S3E). Both were *SPDEF*<sup>+</sup>, *MUC5B*<sup>+</sup>, *SERPINA1*<sup>+</sup> (Figure 2I), suggesting they differentiate into goblet or mucous cells. By contrast, (6) proximal secretory 3 (*SCGB1A1*<sup>+</sup>, *SCGB3A2*<sup>L/O</sup>, *SCGB3A1*<sup>+</sup>) was detected from 15 pcw and was *SPDEF*<sup>+</sup> (Figures 2A–2C), but *CYP2F1*<sup>+</sup>, *MUC4*<sup>+</sup>, and *KRT4*<sup>+</sup> (Figures 2G and 2J). All three

luminal proximal secretory cell populations were located in the proximal cartilaginous airways and were *MUC16*<sup>+</sup> (Figures 2C, 2G, 2H, S3E, and S3F). Detailed spatial-temporal analysis of 10–21 pcw airways revealed that the proportion of proximal secretory progenitors decreased with developmental age, while proximal secretory cells 1 and 2 increased (Figures S4A–S4C), consistent with a progenitor function for proximal secretory progenitors.

### Other airway cells

We detected ciliated cells (*FOXJ1*<sup>+</sup>, *ALOX15*<sup>+</sup>) from 11 pcw, interspersed with secretory/club cells (Figures 2A–2C, S3H, S4A, and S4B). Rarer deuterosomal cells (*FOXJ1*<sup>+</sup>, *CDC20B*<sup>+</sup>) appeared at the same time (Figures 2A–2C). *MUC16*<sup>+</sup> ciliated cells (*FOXJ1*<sup>+</sup>, *DNAH*<sup>+</sup>, *MUC16*<sup>L/O</sup>) were also detected from 11 pcw but confined to proximal dissected regions (Figures 2A–2C and S3H). They were located in patches in the most proximal cartilaginous airways (Figure S3I) and likely represent *MUC16*<sup>+</sup> secretory cells generating ciliated cells, as suggested in the adult.<sup>19–21</sup> Basal cells (*TP63*<sup>+</sup>, *F3*<sup>+</sup>) were present from 9 pcw (Figures 2A–2C and S3J) and more frequent in proximal regions (Figures 2C, S4A, and S4B). Rarer cells (ionocytes, tuft) that have been identified in adult airways were not present in our single-cell data. However, we found putative ionocytes (*FOXI1*<sup>+</sup>; 4/4 lungs) and putative tuft cells (*POU2F3*<sup>+</sup>; 2/4 lungs) in the most proximal cartilaginous airways of 21–22 pcw lung sections (Figure S4E), suggesting they begin to differentiate mid-gestation. Moreover, we reproducibly detected a small population of *MUC5AC*<sup>+</sup>, *ASCL1*<sup>+</sup> cells in 9–11 pcw lungs (Figures 2A–2C). These were localized to the proximal non-cartilaginous airways where they appeared as solitary, somewhat basal, non-columnar cells (Figure S3K). We hypothesize that they are an unknown progenitor, consistent with their transient appearance and the observation that *Ascl1*<sup>+</sup> NE cells in adult mice can generate club, ciliated, and mucous cells following injury.<sup>22,23</sup>

### Predicted airway epithelial differentiation trajectories

A detailed spatiotemporal analysis of major airway epithelial cell types from 10–21 pcw confirms that cell maturation begins more proximally. 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 (Figures S4A–S4D). Conversely, airway progenitors are found throughout the non-cartilaginous airways at 10–12 pcw but restricted to terminal airways by 15–21 pcw (Figures S4A–S4D). In addition, proximal secretory cells are spatially restricted to the cartilaginous airways, while club cells are found in the non-cartilaginous regions (Figures S4A–S4D).

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

(A and B) UMAP visualization of epithelial cells, colored by cell types (A), stage (B, left), and region (B, right).

(C) Dot plot describing differential marker gene expression level for epithelial cells.

(D) UMAP visualizing the predicted epithelial cell lineage trajectory using scvelo; inset: developmental age.

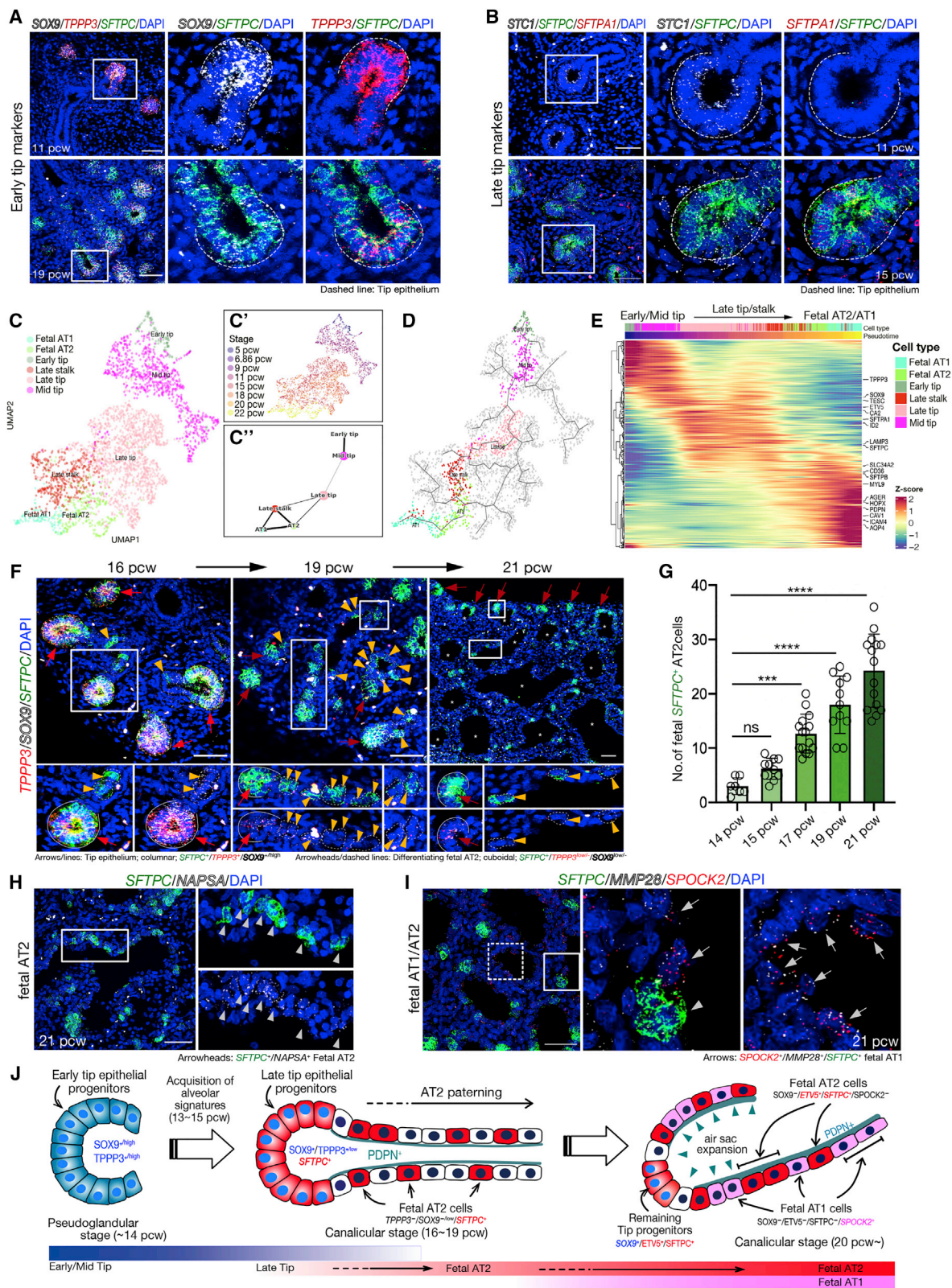
(E and F) *In situ* HCR at 11 (F) and 12 (E) pcw. (E) *SOX9* (tip epithelium, white), *CYTL1* (red), *SCGB3A2* (green). (F) *GHRL*<sup>+</sup> (*GHRL*<sup>+</sup> neuroendocrine, red), *GRP*<sup>+</sup> (pulmonary neuroendocrine, green).

(G) Dot plot showing differential marker genes across secretory cell subtypes.

(H) *In situ* HCR at 19 pcw using *SCGB1A1* (red), *SCGB3A2* (green), and *SCGB3A1* (white).

(I and J) Differentially enriched genes in the proximal secretory cell subtypes. *SPDEF* (I, I'), *SERPINA1* (I''), *CYP2F1* (J), *MUC4* (J'), and *KRT4* (J'') all white; *MUC5B* (I') and *SCGB1A1* all red, and *SCGB3A2* green. DAPI, nuclei. Scale bars, 50  $\mu$ m. See also Figures S3, S4, and S5.





(legend on next page)



This spatial separation means that predicted differentiation trajectories that combine proximal secretory cells and club cells (Figure 2D) can reveal general trends but are likely to be oversimplified. We therefore predicted mid- (Figures S5A–S5C) and late-stage (Figures S5D–S5F) airway lineage trajectories separately. In both cases, basal cells formed discrete clusters on the UMAPs (Figures S5A' and S5D'). Trajectory inference analysis suggests a differentiation route from mid-tip to stalk to airway progenitors to proximal secretory progenitors and proximal secretory cells (Figure S5B), consistent with sample age (Figure S5B'). Visualizing gene expression along the inferred trajectory shows mid-tip and stalk cells are similar (Figure S5C). Stalk cells lose some tip markers, including *FOXP2* and *SOX9*, and gain a 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*<sup>24,25</sup> (Figure S5C). A similar differentiation trajectory was predicted from late-tip to late-stalk to late-airway progenitor to club cells (Figure S5E), although the oldest tip and stalk cells included in this analysis may produce alveolar lineages (Figures S5E', 3C, 3E, S6A, and S6B). Visualizing 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*; Figure S5F).

These analyses predict that cells exit the tip to the stalk state, followed by gain of airway progenitor identity before commitment to a specific differentiation state that likely depends on local signaling cues. Although we cannot predict the origin of the basal cells using trajectory inference methods, we hypothesize 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.<sup>26</sup>

Our trajectory inference (Figures S5A–S5F) predicts that airway progenitors will differentiate readily to airway cell types. At 9–10 pcw, *CYTL1*<sup>+</sup> and *SCGB3A2*<sup>+</sup> airway progenitors are found throughout the airway tree (Figures 2E, S3B, S3D, S3L, S4A, S4B, and S4D). We isolated airway progenitors using a combination of distal non-cartilaginous airway micro-dissection and transduction with a lentiviral *SCGB3A2* transcriptional reporter (*SCGB3A2-GFP*, Figure S5G). Freshly isolated distal *SCGB3A2-GFP*<sup>+</sup> cells were *SOX9*<sup>LO</sup>, *CYTL1*<sup>HI</sup>, *SCGB1A1*<sup>LO</sup>, *SCGB3A2*<sup>LO</sup>, and *SCGB3A1*<sup>LO</sup> compared to tip/stalk cells and more proximal

*SCGB3A2-GFP*<sup>+</sup> cells from the same lungs (Figure S5H). When single cells were placed into an FGF-containing differentiation medium,<sup>27</sup> distal *SCGB3A2-GFP*<sup>+</sup> cells produced basal, ciliated, and mature secretory cells (Figures S5I–S5M). 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, and proximal secretory progenitor) and differentiating airway cells that localize to a spatial differentiation gradient along the proximal-distal axis of the epithelium (summarized in Figures S3L and S4D). Moreover, we identify *GHRL*<sup>+</sup> neuroendocrine cells that do not exist in the mouse.

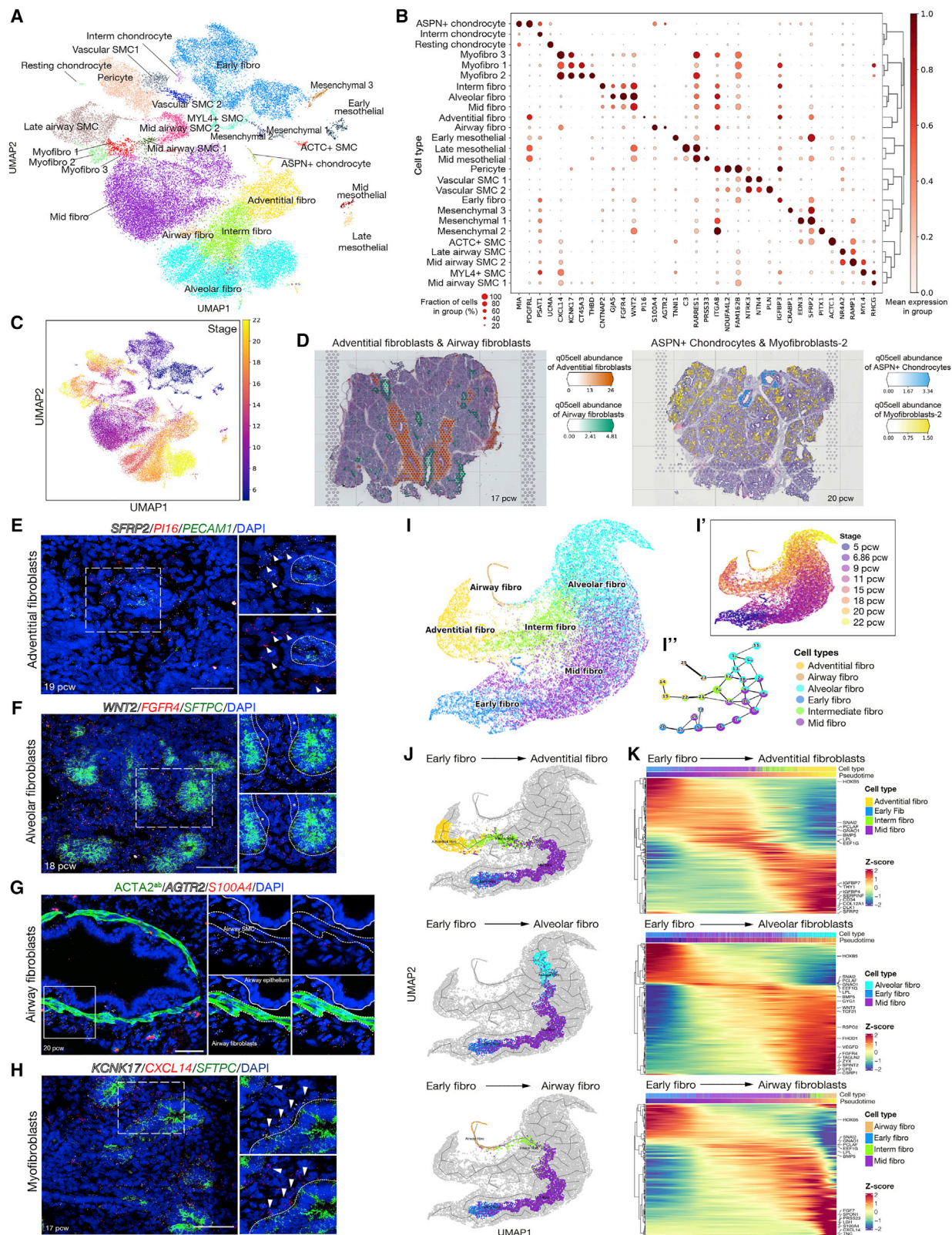
### Late epithelial tip cells acquire alveolar identity prior to alveolar differentiation

Tip cells express a core set of tip-specific markers (*SOX9*<sup>+</sup>, *ETV5*<sup>+</sup>, *TESC*<sup>+</sup>, *TPPP3*<sup>+</sup>, and *STC1*<sup>+</sup>) at all stages sampled (Figures 2A–2C). 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 (Figure 2C). By 15 pcw the AT2 markers *SFTPC* and *SFTPA* were detected readily in late-tip cells where they were co-expressed with lower levels of core tip markers (Figures 3A and 3B). The late tip is a transcriptional state that has not been detected in developing mouse lungs.<sup>17,28</sup> The change in expression profile that is observed upon the transition to late tips correlates with a change in the predicted differentiation trajectory from late-tip cells to late-stalk to fetal AT2 and AT1 cells (Figures 3C and 3D; without late stalk in S6A). However, trajectory inference analysis at this transitional stage is challenging. It is likely that some of the late-tip cells produce the terminal branches of the conducting airways (Figures S5D–S5F). Moreover, the inferred connections between mid-tip and late-tip cells are weak (Figure 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 (Figure 2D). Nevertheless, throughout this period, similar to earlier stages, late-tip cells remain *SOX9*<sup>+</sup> and late-stalk cells turn off tip markers and acquire *PDPN*/*AGER* (Figures S3A and S6D).

A small number of AT2 cells appear in the single-cell data from 15 pcw but are more prominent from 22 pcw (Figure 2A). Similarly, at around 16 pcw, late-tip cells (*SOX9*<sup>+</sup>, *TPPP3*<sup>+</sup>, *SFTPC*<sup>+</sup>) were clearly visualized in the tissue, but differentiating AT2 cells (*SOX9*<sup>LO/-</sup>, *TPPP3*<sup>LO/-</sup>, *SFTPC*<sup>+</sup>) were rare (Figures 3F and 3G, and

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

(A and B) *In situ* HCR at 11 (A and B), 15 (B), and 19 (A) pcw. (A) *SFTPC* (green), *TPPP3* (red), *SOX9* (white). (B) *SFTPC* (green), *SFTPA1* (red), *STC1* (white). Dashed lines represent tip epithelium.  
(C and D) UMAP visualization of early to late tip, late stalk, fetal AT1 and AT2 cells, colored by cell types (C) and stages (C'); PAGA analysis (C''); Monocle3 trajectories (D).  
(E) Gene expression heatmap of trajectory colored in (D).  
(F) *In situ* HCR at 16, 19, and 21 pcw, *SFTPC* (green), *TPPP3* (red), and *SOX9* (white). White lines/red arrows: columnar tip progenitors, *SFTPC*<sup>+</sup>/*SOX9*<sup>high</sup>/*TPPP3*<sup>+</sup>. Arrowheads/dashed lines in stalk/air sac regions: cuboidal differentiating fetal AT2 cells, *SFTPC*<sup>+</sup>/*SOX9*<sup>low</sup>/*TPPP3*<sup>low</sup>. Asterisks (\*) represent primitive air sacs.  
(G) Quantification of cuboidal *SFTPC*<sup>+</sup>/*SOX9*<sup>low</sup> fetal AT2 cells in stalk/air sac regions in (F). The *SFTPC*<sup>+</sup> tip epithelial cells were excluded by their columnar morphology and marker expression (*SOX9*<sup>low</sup>). Mean ± SD, n > 7. Significance evaluated by one-way ANOVA with Tukey multiple comparison post-test; ns: not significant, \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001.  
(H and I) *In situ* HCR analysis at 21 pcw. Fetal AT2 *SFTPC*<sup>+</sup> and *NAPSA*<sup>+</sup> (arrowheads; H and I) and fetal AT1 *SFTPC*<sup>+</sup>/*MMP28*<sup>+</sup>/*SPOCK2*<sup>+</sup> (arrows; I).  
(J) Diagram of the acquisition of alveolar progenitor identity by late epithelial tips, followed by differentiation to fetal AT2 and AT1 lineages. DAPI, nuclei. Scale bars, 50 μm. See also Figure S6.



(legend on next page)



S6C). Over the following weeks, the size of the tip regions decreased and more differentiating AT2 cells were detected (Figures 3F and 3G). At 21 pcw, smaller numbers of late-tip cells persist, and AT2 cells (*SOX9*<sup>+</sup>, *SFTPC*<sup>+</sup>, *NASPA*<sup>+</sup>, *ETV5*<sup>+</sup>) were found scattered throughout the developing air sacs (Figures 3H and S6E–S6J). Consistent with the predicted change in tip fate potential (Figures 3C–3E), late-tip cells (16–20 pcw) grown as organoids retained a late-tip phenotype *in vitro* and more readily differentiated to mature AT2 cells than organoids derived from earlier developmental stages.<sup>29</sup>

In our single-cell atlas, differentiating AT1 cells were first visible at 18 pcw but more prominent by 22 pcw (Figures 2A–2C). Similarly in tissue sections, AT1 cells were not detected at 17 pcw (Figure S6H). However, by 20 pcw, differentiating AT1 cells (*SPOCK2*<sup>LO</sup>, *SFTPC*<sup>+</sup>) were visible, and at 21 pcw, AT1 cells (*SPOCK2*<sup>+</sup>, *SFTPC*<sup>+</sup>) were interspersed with AT2 cells lining the developing air sacs (Figures 3I, S6I, and S6J). In sections, AT1 markers were only detected in cells which had no or extremely low levels of *SFTPC* (Figures S6H–S6J). Moreover, *SFTPC*-negative cells were always observed in the stalk regions from 16 pcw onwards (Figure 3F). These 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 AT2 cell markers, and enter the late-stalk cell state, prior to initiating AT1 or AT2 cell differentiation in response to local signaling cues (Figure 3J). Furthermore, the late-stalk cells are connected to AT2, AT1, and late airway progenitors in trajectory inference analysis (Figures 2D, 3C, and 3D), supporting our hypothesis that at all stages of lung development, cells exit the tip and enter a stalk state prior to differentiation.

Integration of our fetal cell atlas with adult data revealed high correlation between expected groups: fetal airway progenitors with adult secretory club cells, fetal and adult ciliated and deuterosomal cells, and proximal secretory fetal cells with adult goblet cells (Figure S2A). The AT2 and AT1 cells we detect in the fetal lungs cluster closely with the adult (Pearson correlation coefficients: fetal-adult AT2 0.66; AT1 0.80). However, the fetal cells are immature and differ in gene expression to their adult counterparts (Figure S2G).

### Lung endothelial cells exhibit early specialization into arterial and venous identities

At 5–6 pcw, the endothelial cells (ECs) comprised capillary (early Cap: *THY1*<sup>+</sup>, *CD24*<sup>+</sup>), *GRIA2*<sup>+</sup> arterial (*GRIA2*<sup>+</sup>, *GJA5*<sup>+</sup>), and

lymphatic ECs (*PROX1*<sup>+</sup>, *STAB1*<sup>+</sup>, and *UCP2*<sup>LO</sup>), showing that capillaries and lymphatic vessels are distinct from the earliest stages of lung development and that arterial specification begins prior to venous (Figure S1T). At later stages, trajectory analysis predicts that both mid- and late-Cap cells generate arterial and venous ECs (Figure S7A). Aerocytes (*CA4*<sup>LO</sup>, *S100A3*<sup>+</sup>), capillary ECs specialized for gas exchange and leukocyte trafficking,<sup>30,31</sup> were observed at 20–22 pcw around the developing air sacs (Figure S7B). Microvasculature specification therefore occurs relatively late in human fetal life coincident with the development of AT1 cells.

Broad markers of arterial and venous specification were clear in sections at 20 pcw (Figures S7C and S7D). Three distinct arterial ECs were detected. *GRIA2*<sup>+</sup> and arterial ECs (*DKK2*<sup>+</sup>, *SSUH2*<sup>+</sup>) form a continuous differentiation trajectory in pseudotime (Figure S7A) with *GRIA2*<sup>+</sup> ECs likely to be a more immature form. The *OMD*<sup>+</sup> ECs (*GJA5*<sup>+</sup>, *DKK2*<sup>+</sup>, *PTGIS*<sup>+</sup>, and *OMD*<sup>+</sup>) cluster with arterial ECs and are more proximal (Figure S1O). By contrast, venous ECs (*PVLAP*<sup>+</sup>, *ACKR3*<sup>+</sup>, and *HDAC9*<sup>+</sup>) do not have clear subclusters. Systemic and pulmonary circulation ECs have been found in adult lungs<sup>32</sup>; we cannot detect these in fetal lungs. Two major lymphatic ECs were detected: lymphatic ECs (*PROX1*<sup>+</sup>, *STAB1*<sup>+</sup>, and *UCP2*<sup>LO</sup>) and SCG3<sup>+</sup> lymphatic ECs (*PROX1*<sup>+</sup>, *SCG3*<sup>+</sup>) (Figure S7E). SCG3<sup>+</sup> lymphatic ECs resemble a lymphatic valve population.<sup>33</sup>

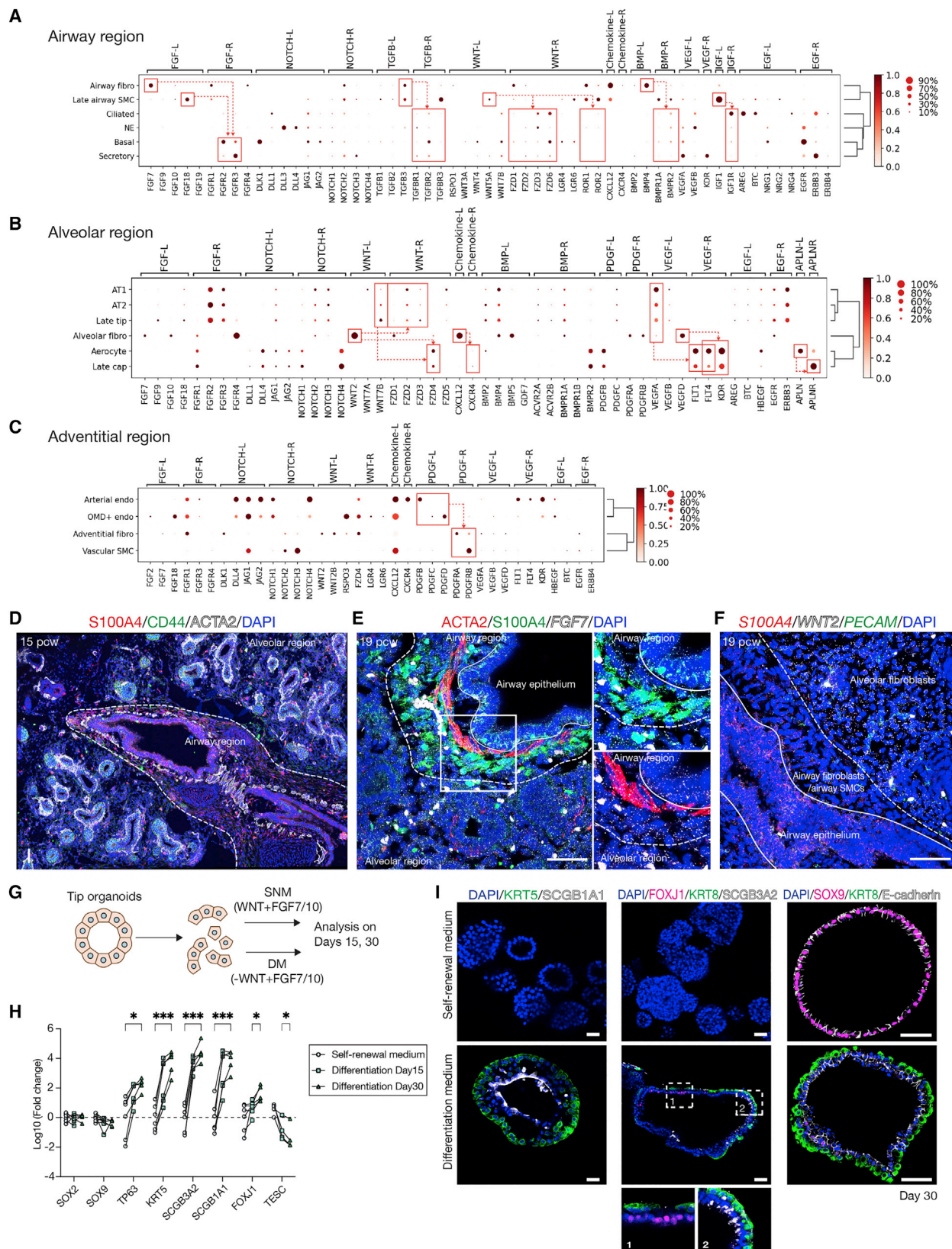
### Hematopoietic cell types in the developing lung

At the early stages (5–6 pcw) when arterial, capillary, and lymphatic ECs were present, embryonic erythrocytes, *HMOX1*<sup>+</sup> erythroblasts, and a small number of macrophages and ILC progenitors were detected, representing the early progenitors of hematopoiesis. After 11 pcw, relative numbers of lymphoid and myeloid cells increased, dominated by macrophages; ILCs; and dendritic, NK, T, and B cells (Figures 1C–1E, S1P, S1R, and S1T). Immature T cells are largely absent from the atlas, consistent with the restriction of T cell development to the thymus. In contrast, a range of early B cell precursors and the ILC precursor were detected. TCR and BCR scRNA-seq supported cell-type identities and subdivision (Figures S1Q'' and S1R''). We compared our atlas with a pan-fetal human atlas<sup>13</sup> and found that leukocytes were transcriptionally highly similar to those of other organs (Figure S1S).

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

- (A) UMAP visualization of mesenchymal cells.
- (B) Dot plot of mesenchymal differential marker gene expression.
- (C) UMAP visualization of mesenchymal cells colored by stage.
- (D) Visium spatial feature plots visualizing adventitial fibroblasts, airway fibroblasts, *ASPN*<sup>+</sup> chondrocytes, and myofibroblast-2 on 17 and 20 pcw lung sections. Scores are conservative estimates of cell-type abundance per voxel.
- (E–H) *In situ* HCR assay (E–H) and immunostaining (G).
- (E) Adventitial fibroblasts (*SFRP2*, white/*PI16*, red; arrowheads), ECs (*PECAM1*, green).
- (F) Alveolar fibroblasts (*WNT2* white; *FGFR4* red), tip cells (*SFTPC* green). Asterisks (\*myofibroblasts).
- (G) Airway fibroblasts (*S100A4* red; *AGTR2* white), smooth muscle (*ACTA2* green, dashed line).
- (H) Myofibroblasts (*KCNK17* white, *CXCL14* red; arrowheads), tip cells (*SFTPC*, green). DAPI, nuclei. Scale bars, 50  $\mu$ m.
- (I) UMAP visualization of cell types (I) and stage (I') and PAGA analysis (I'') of fibroblast differentiation trajectories.
- (J and K) UMAPs with Monocle3 trajectories (J) and selected trajectory gene expression heatmaps (K) for mid tip to adventitial fibroblasts (top), alveolar fibroblasts (middle), or airway fibroblasts (bottom). See also Figure S7.





(legend on next page)

### Developmental trajectories of mesenchymal cells

The broad fibroblast cluster comprises fibroblasts, myofibroblasts, airway and vascular smooth muscle (ASM and vSMC), pericytes, mesothelium, and chondrocytes (Figures 4A and 4B). Airway fibroblasts and chondrocytes were proximally enriched and mesothelium distally enriched (Figures 4D and S1O). Cell clusters separated by age (Figure 4C). ASM cells were observed from 9 pcw, consistent with previous immunostaining,<sup>8</sup> and showed increasing maturity over time (Figures 4A–4C). Two distinct populations of vSMC were observed throughout the time course, vSMC1 (*NTRK3*<sup>+</sup>, *NTN4*<sup>+</sup>, and *PLN*<sup>+</sup>) and vSMC2 (*NTRK3*<sup>+</sup>, *NTN4*<sup>+</sup>, and *PLN*<sup>+</sup>) (Figures 4A and 4B), and were intermingled around the same vessels on tissue sections (Figures S7F and S7H). vSMC1 was enriched in genes relating to ECM organization and cell adhesion, and vSMC2 for transcripts encoding contractility proteins and signaling molecules (Figure S7G). Intermingling of vSMC subtypes with different levels of contractility proteins is seen in adult lungs<sup>34</sup>; our developmental observation suggests that these represent normal functional/ontological differences, rather than pathology. Pericytes (*FAM162B*<sup>+</sup>) were visualized adjacent to the microvascular endothelium (Figure S7I).

The most common cells isolated from 5–15 pcw lungs were fibroblasts (Figure 1C). At 5–6 pcw, early fibroblasts (*SFRP2*<sup>+</sup>, *WNT2*<sup>+</sup>) predominated, although multiple populations were detected (Figures 4A and 4B). In 9–11 pcw lungs, early fibroblasts had matured into mid fibroblasts (*WNT2*<sup>+</sup>, *FGFR4*<sup>L/O</sup>) which can promote epithelial tip cell fate.<sup>35</sup> In the oldest lungs sequenced, there were three distinct fibroblasts: adventitial (*SFRP2*<sup>+</sup>, *PI16*<sup>+</sup>), airway (*AGTR2*<sup>+</sup>, *S100A4*<sup>+</sup>), and alveolar (*WNT2*<sup>+</sup>, *FGFR4*<sup>+</sup>) with distinct locations (Figures 4A, 4B, and 4D–4G). In addition, an intermediate fibroblast connected the more mature fibroblasts on the UMAP (Figures 4A and 4B), 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 (Figures 4I–4K). Alternatively, the intermediate fibroblast population may indicate lineage plasticity as previously suggested.<sup>36</sup>

The three major fibroblast types in 15–22 pcw lungs expressed high levels of genes associated with ECM organization but had distinct gene expression patterns and spatial localization. Adventitial fibroblasts (*SFRP2*<sup>+</sup>, *PI16*<sup>+</sup>) surrounded the larger blood vessels (Figure 4D). They formed diffuse layers of cells surrounding the tightly packed concentric rings of ECs, pericytes, and vSMCs (Figures 4E and S7H). Adventitial fibroblasts were enriched in genes associated with ECM organization and signaling, including BMP, TGFβ, and WNT (Figures 4J, 4K, and S7J), consistent with

described roles providing structural support to the perivascular region.<sup>37</sup> Alveolar fibroblasts (*WNT2*<sup>+</sup>, *FGFR4*<sup>+</sup>) were observed throughout the lung, particularly surrounding tip cells and microvasculature (Figure 4F). They were enriched in genes associated with actin organization, focal adhesions, and morphogenesis, as well as signaling molecules (Figures 4J, 4K, and S7J). Adventitial and alveolar fibroblasts expressed shared and unique genes (adventitial: *SERPINF1*, *SFRP2*, and *PI16*; alveolar: *FGFR4*, *VEGFD*; Figure 4K). By contrast, the airway fibroblasts (*AGTR2*<sup>+</sup>, *S100A4*<sup>+</sup>; note *S100A4* is expressed in various immune and airway epithelial cells) were adjacent to the ASM and highly enriched in signaling molecules associated with morphogenesis (Figures 4D, 4G, 4J, 4K, and S7J). We did not detect lipofibroblasts,<sup>38</sup> meaning that they are either rare, form later than 22 pcw, or do not form distinct clusters in all lung datasets.<sup>14</sup> Endothelial and fibroblast populations align well between fetal and adult data (Figures S2B and S2C), but with some unique developmental states, such as fetal early/mid-fibroblasts and myofibroblasts.

Myofibroblasts formed three distinct groups in our single-cell data. Myofibroblast 1 (*CXCL14*<sup>+</sup>, *KCNK17*<sup>+</sup>, *CT45A3*<sup>+</sup>, and *THBD*<sup>L/O</sup>) appeared at 9 pcw and persisted to 20 pcw. Myofibroblast 2 (*CXCL14*<sup>+</sup>, *KCNK17*<sup>+</sup>, *CT45A3*<sup>+</sup>, and *THBD*<sup>H</sup>) and myofibroblast 3 (*CXCL14*<sup>+</sup>, *KCNK17*<sup>+</sup>, *CT45A3*<sup>+</sup>, and *THBD*<sup>+</sup>) were predominantly identified at 22 pcw (Figures 4A and 4B). Throughout development, myofibroblasts (*CXCL14*<sup>+</sup>, *KCNK17*<sup>+</sup>) were visualized surrounding the developing stalk region of the epithelium, suggesting a close signaling relationship (Figures 4D, 4H, S7K, and S7L). Although not detected in significant numbers in the scRNA-seq data until 22 pcw, we see myofibroblast 2 (*PDGFRA*<sup>+</sup>, *THBD*<sup>H</sup>, and *NOTUM*<sup>+</sup>) around the stalk epithelium from 15 pcw (Figures 4D, S7L, S7N, and S7O), the same position as myofibroblast 1. The appearance of myofibroblast 2 is coincident with the acquisition of AT2 markers by the late-tip cells and may be a more mature state of myofibroblast 1. Myofibroblast 2 was enriched in gene expression associated with cell contractility and focal adhesions, as well as WNT signaling (Figures S7N and S7O). Co-expression of the Wnt-responsive genes *LEF1*, *NOTUM*, and *NKD1* suggests that myofibroblast 2 is responding to local Wnt expression (*WNT2* is high in alveolar fibroblasts) and producing the secreted Wnt inhibitor *NOTUM*, potentially to regulate local cell patterning. By contrast, myofibroblast 3 has higher expression of genes associated with ECM organization and a variety of signaling molecules, including *C7*, *RSPO2*, and *BMPER* (Figure S7N). Myofibroblast 3 was localized to the developing air sacs (Figure S7M), rather than the stalk epithelium, and is likely to be a precursor of the alveolar myofibroblasts.<sup>17,39</sup>

### Figure 5. Signaling ligand-receptor interactions in specific niches

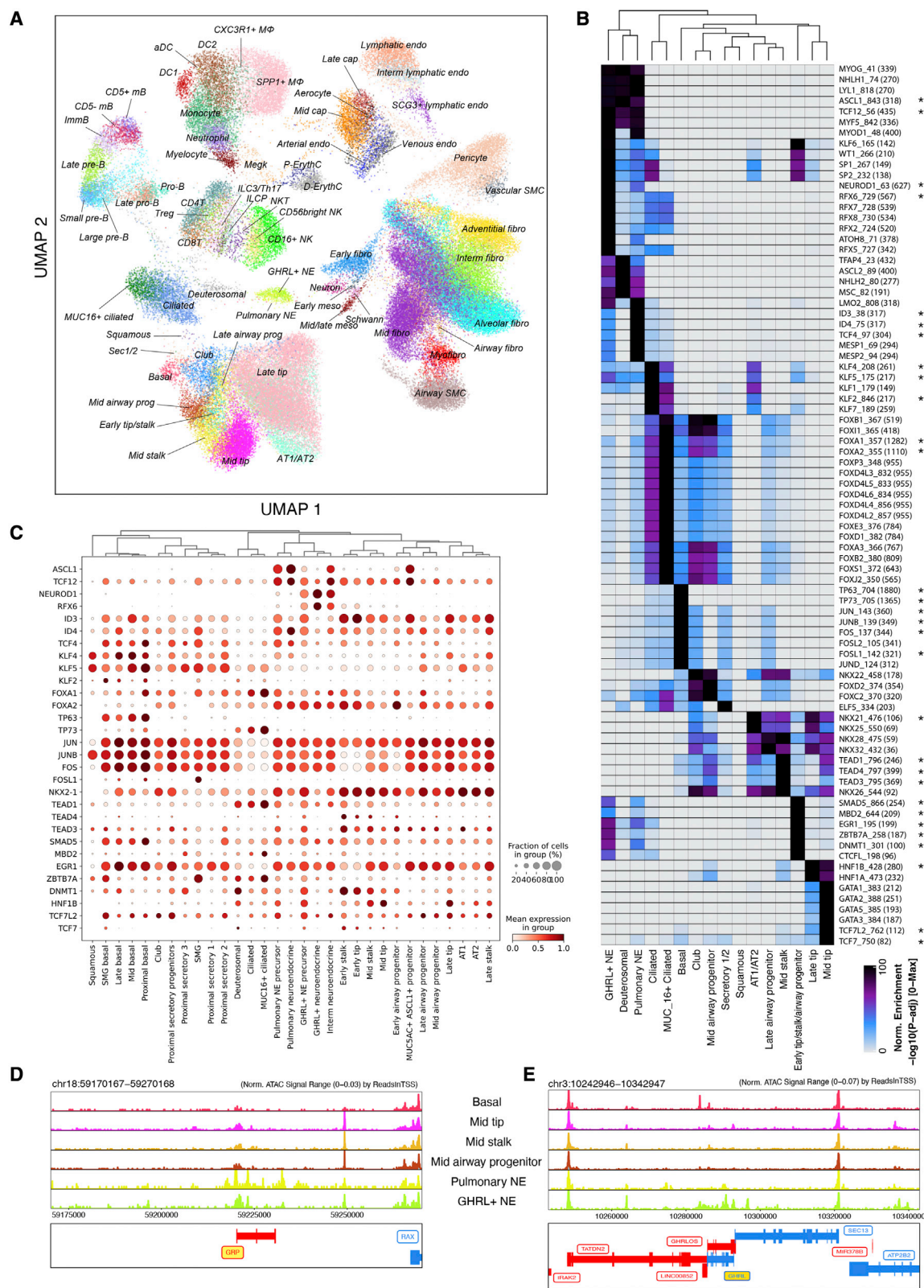
(A–C) Curated ligand-receptor interaction predictions from CellPhoneDB in airway (A), alveolar (B), and adventitial (C) niches. Dot plots visualize gene expression by cell type; dashed arrows indicate the predicted direction of signaling from ligands to receptors.

(D–F) Immunofluorescence/HCR. *S100A4/S100A4*, airway fibroblasts; *ACTA2*, ASM; *CD44*, tip epithelium; *PECAM1*, ECs. Airway fibroblasts/ASM form a boundary (dashed lines) between alveolar and airway regions. Lines are between airway fibroblasts/SMCs and airway epithelium. DAPI, nuclei. Scale bars, 50 μm. (G) Organoids were cultured in FGF7/10-containing medium, in the presence (self-renewal medium; SNM) or absence (differentiation medium; DM) of CHIR99021, for 30 days.

(H) qRT-PCR quantification normalized to organoids cultured in SNM. Significance evaluated by two-way ANOVA with Tukey multiple comparison post-test; \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001; n = 6 organoid lines.

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





(legend on next page)



### Signaling niches in lung development

We used CellPhoneDB<sup>40</sup> to predict signaling interactions controlling cell fate. We focused on 15–22 pcw cells and, based on the localization of the major fibroblast populations (Figures 4E–4G), analyzed signaling within three niches. The airway niche includes airway fibroblasts, late airway SMCs, and airway epithelial cells. The alveolar niche includes alveolar fibroblasts, aerocytes, late Cap cells, late-tip cells, AT1, and AT2. Finally, the adventitial niche includes adventitial fibroblasts, arterial endothelium, OMD<sup>+</sup> endothelium, and vascular smooth muscle cells. CellPhoneDB predicts numerous signaling interactions (Table S2) that we curated by plotting the expression of ligand-receptor pairs representing major signaling pathways (Figures 5A–5C). We observed expected interactions, including high levels of Notch ligands and receptors and CXCL12-CXCR4 signaling in the adventitial niche (Figure 5C).<sup>41</sup> Similarly, expected signaling predicted in the alveolar niche included aerocytes to late cap cells (ALPN-ALPNR) and alveolar epithelial cells to microvascular ECs (VEGFA-FLT1/FLT4/KDR) (Figure 5B).<sup>30,31</sup>

Airway fibroblasts were predicted to signal via TGFβ3 and BMP4 to the airway epithelium, consistent with roles for these signals in human basal cell specification and differentiation.<sup>42,43</sup> Airway fibroblasts and ASM were also predicted to signal to the epithelium via FGF7/18 to FGFR2/3 and non-canonical WNT5A to FZD/ROR (Figure 5A). By contrast, although FGF and WNT signaling interactions were also predicted in the alveolar niche, interactions were based on lower levels of FGF but higher levels of canonical WNT2 and its receptor (Figure 5B). The predicted FGF and WNT signaling interactions in the alveolar niche and late-tip cells are consistent with the requirement of these factors for long-term self-renewal of human distal tip organoids.<sup>8,29</sup> Tissue staining showed that although FGF7 is expressed fairly ubiquitously, the airway fibroblasts and ASM form a distinct barrier between the airway epithelium and the WNT2 expression (Figures 5D–5F). Based on these data, we predicted that removing canonical WNT but retaining FGF signaling would promote airway differentiation in the human distal tip organoids (Figure 5G). Indeed, we observed robust basal, secretory, and ciliated cell differentiation in response to FGF-containing medium (Figures 5H and 5I).

### scATAC-seq analysis identifies putative cell fate regulators

Single cell ATAC-seq provides an independent method of assessing cellular-level gene regulation based on open chromatin regions and allows cell-type-specific TFs to be predicted. After tissue dissociation, the single-cell suspensions were split, and half of the cells were processed for nuclear isolation and scATAC-seq (Figure 1A). Following quality control and doublet removal, 67 scATAC-seq clusters comprising ~100K cells were obtained, and label transfer was used to annotate scATAC-seq clusters based

on our scRNA-seq data (Figure 6A). Not every cell state detected by scRNA-seq was distinguishable by scATAC-seq, consistent with previous work.<sup>13,44</sup> For example, separate early-tip, stalk, and airway progenitor clusters were discerned by scRNA-seq (Figure 2A), but a combined cluster with strong similarity to all three cell types was detected by scATAC-seq (Figure 6A). Nevertheless, there was broad agreement between the scRNA-seq and ATAC-seq data in terms of capturing cell types, including many of the novel/lesser-known cell types we identified by scRNA-seq (mid and late tip, mid and late airway progenitors, GHRL<sup>+</sup> NE, MUC16<sup>+</sup> ciliated, dueterosomal, airway fibroblasts, aerocytes, and SCG3<sup>+</sup> lymphatic endothelial cells).

We analyzed TF binding motifs in the unique/enriched open chromatin regions in each cluster and plotted the top TF motifs per cell type (Table S4). As expected, TFs belonging to the same family are frequently enriched in the same cell type due to similarities in their binding motifs. This analysis revealed some expected TF signatures, for example TCF21 in the fibroblasts,<sup>45</sup> GRHL, and FOXA1/2 in epithelium,<sup>46,47</sup> and SOX17 in arterial endothelium.<sup>48</sup> Examining epithelial cells and focusing on TFs expressed in the corresponding cell type in the scRNA-seq data (Figures 6B and 6C, marked by asterisk in 6B), TEAD motifs were enriched in mid-stalk cells, consistent with a key role for Yap,<sup>49</sup> NKX2.1 in AT1/AT2 cells,<sup>50</sup> KLF factors in secretory cells and AT1/AT2,<sup>51</sup> and TP63 in basal cells.<sup>52</sup> Unexpected TF signatures included HNF1B in late-tip cells and ZBTB7A in early-tip/stalk/airway progenitors. We focused on the pulmonary and GHRL<sup>+</sup> NE cells, which cluster closely (Figures 2A and 6A). ASCL1 is required for mouse NE cell differentiation,<sup>53,54</sup> and this motif is strongly associated with both pulmonary and GHRL<sup>+</sup> NE cells (Figure 6B). However, both cell types also respectively have specific TF motifs including NEUROD1 and RFX6 in the GHRL<sup>+</sup> NEs, and TCF4 and ID in the pulmonary NEs (Figure 6B). Consistent with this, there are distinct, unique regions of open chromatin, especially in the neighborhood of cell-type-specific genes such as GRP and GHRL (Figure 6D and 6E).

We have produced a high-resolution scATAC-seq dataset for the developing human lungs, which is highly consistent with our scRNA-seq data. Mining these data provides hypotheses for lineage-determining TFs in lung development.

### Transcriptional control of neuroendocrine cell subtype formation

Pulmonary NE and GHRL<sup>+</sup> NE cells share the expression of many TFs and open chromatin regions but are transcriptionally distinct. In our scRNA-seq data, they were both observed along a maturation trajectory and shared classical NE markers (*CHGA*, *SYP*), but differed in TF and hormone expression (Figures 7A and 7B). A third NE population (intermediate NE) clustered between pulmonary and GHRL<sup>+</sup> NE cells with intermediate gene expression (Figures 7A and 7B), although it did contain a small number

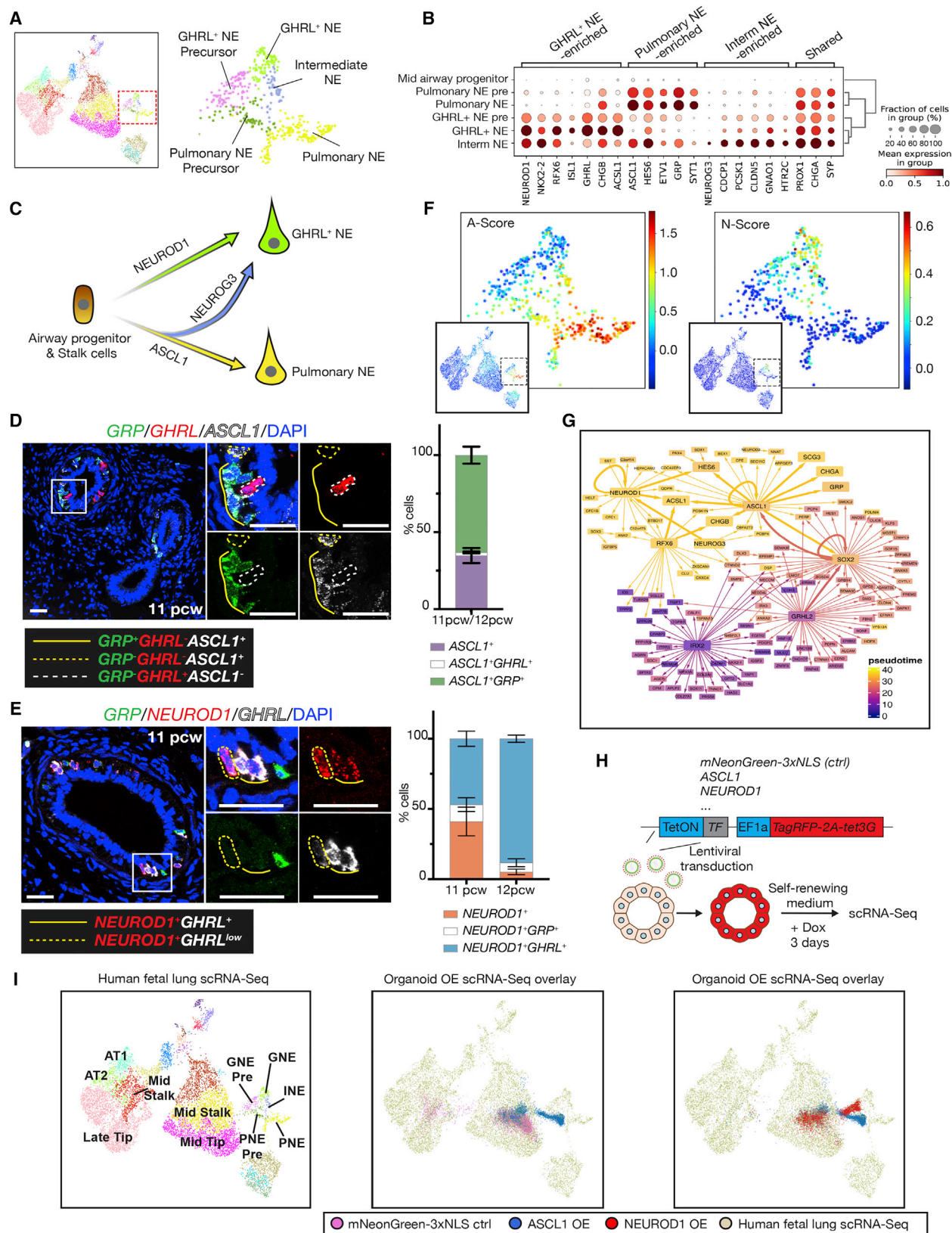
#### Figure 6. DNA accessibility and motif enrichment revealed by scATAC-seq

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

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

(C) Expression dot plot of the concordant transcription factors from (B) in epithelial cell types.

(D and E) Read coverage tracks of *in silico* aggregated “pseudo-bulk” epithelial clusters over the GRP locus (D) and GHRL locus (E). See also Table S4.



(legend on next page)

of cells expressing the unique marker *NEUROG3*. Pseudotime trajectory analysis suggested that pulmonary NE and *GHRL*<sup>+</sup> NE cells were derived from airway progenitors/stalk cells and that intermediate NEs are an additional transition population (Figures S8A and S8B). Transition states between pulmonary NE and *GHRL*<sup>+</sup> NE were observed in sections (Figure S8C). We therefore postulated that pulmonary NE precursors could acquire *NEUROG3* and convert to *GHRL*<sup>+</sup> NE fate (Figure 7C), or vice-versa—*GHRL*<sup>+</sup> precursors converting to pulmonary NE fate. In sections, *ASCL1* was co-expressed with *GRP*, but rarely with *GHRL*. We also observed *ASCL1* single-positive cells, likely representing pulmonary NE precursors (Figure 7D). *NEUROD1* was co-expressed with *GHRL* but also observed with *GRP* (Figure 7E), whereas *NEUROG3* was co-expressed with *ASCL1* and/or *NEUROD1*, supporting a role in a transition population (Figure S8D).

Differential expression of *ASCL1* and *NEUROD1* defines A- and N-type human SCLC, which likely derives from NE cells.<sup>55</sup> Interestingly, these two TFs coincide with the scRNA-seq marker genes and scATAC-seq TF motif enrichment of our fetal NE cells (Figures 6B and 7B). We generated SCLC feature gene lists<sup>18</sup> and performed gene signature scoring, showing that the A-type signature resembles pulmonary NEs, whereas the N type resembles *GHRL*<sup>+</sup> NEs (Figure 7F). These data suggest that either there are two different NE cells of origin for human SCLCs or that SCLCs reuse developmental mechanisms, as suggested by some mouse models.<sup>56</sup> We have been unable to detect *GHRL*<sup>+</sup> NEs in the adult airways using HCR (5 biological replicates). However, a small number of *GHRL*<sup>+</sup> cells are present within a tuft cell cluster in an integrated adult lung cell atlas containing 2.2 million cells,<sup>57</sup> suggesting that *GHRL*<sup>+</sup> NEs could be a rare cell state in the adult airways. Given their relevance to human disease states, we used our single-cell atlas to predict NE lineage-defining TFs and test these using our organoid system. We reasoned that overexpression of lineage-defining TFs in lung tip organoids<sup>8,58</sup> would promote cell-type-specific differentiation.

Multiple TFs were differentially expressed between pulmonary NE and *GHRL*<sup>+</sup> NE cells (Figure 7B). We used SCENIC analysis of gene regulatory networks (GRNs)<sup>59</sup> along a predicted airway progenitor to *GHRL*<sup>+</sup> NE trajectory (Figures S8A and S8B) to identify putative lineage-defining TFs (Figure 7G). *ASCL1*, *NEUROD1*, and *NEUROG3* all emerged as potential key nodes

and are required for endocrine cell differentiation in various organs.<sup>53,54,60,61</sup> We also selected the *GHRL*<sup>+</sup> NE-specific *RFX6* (Figure S8E) and *NKX2.2* (Figure 7B), the pan-NE *PROX1* (Figure 7B), and, as controls, the basal cell-specific TFs *DeltaNTP63*, *TFAP2A*, *PAX9*, and *mNeonGreen-3xNLS*. Overexpression of *PROX1* or *NKX2.2* did not result in NE gene upregulation based on qRT-PCR (data not shown), and these TFs were not followed up. The other factors resulted in increased expression of basal or NE markers compared to *mNeonGreen-3xNLS* controls, and the experiments were repeated using scRNA-seq. Individual TFs were overexpressed from a doxycycline-inducible construct for 3 days, and organoids were maintained in the self-renewing (tip cell-promoting) medium throughout to rigorously assay the lineage-determining competence of the TF (Figures 7H and S8F), followed by scRNA-seq.

When mapped to epithelial cells of our fetal lung atlas, the majority of the *mNeonGreen-3xNLS* expressing organoid cells projected to mid-tip or stalk cells as expected (Figure 7I), whereas overexpression of *DeltaNTP63* resulted in basal cell-like lineages (Figure S8G) consistent with a previous report.<sup>62</sup> Overexpression of *RFX6*, *TFAP2A*, or *PAX9* did not result in the predicted lineage progression at a transcriptome level (Figure S8G). However, *ASCL1*-overexpressing organoids progressed into pulmonary NE precursors (Figure 7I), and *NEUROD1* overexpression promoted differentiation into *GHRL*<sup>+</sup> NE precursors (Figure 7I). *NEUROG3* overexpression also led to *GHRL*<sup>+</sup> NE precursor formation (Figure S8G), suggesting that the *GHRL*<sup>+</sup> NE lineage is the destination of the intermediate NE population (Figure 7C).

The 5' differences between the transgenes and endogenous TFs allowed us to distinguish these transcripts and infer gene regulation hierarchy. We observed autoregulation of *ASCL1*, *NEUROD1*, *NEUROG3*, and *RFX6* (Figure S8H). By contrast, *NKX2.2* and *PROX1* were upregulated by other TFs, indicating they are relatively low in the hierarchy (Figure S8H). *NKX2.2* and *PROX1* expression in the organoid assay matched their expression in NE cells *in vivo* (Figures 7B and S8H), showing that this assay recapitulated key features of the TF network. These experiments tested GRN predictions from the single-cell atlas, confirmed the predicted lineage trajectory, and provided a foundation for studying human SCLC. This is significant given that there is no evidence that *GHRL*<sup>+</sup> NE cells are present in mice,<sup>18</sup> making the use of mouse models difficult.

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

- (A) Zoom-in UMAP plot of NE lineages.  
 (B) Dot plot showing selected gene expression in NE lineages.  
 (C) Schematic model of NE lineage formation.  
 (D) Left: HCR, *GRP* (green), *GHRL* (red), *ASCL1* (white). Right: mean  $\pm$  SEM of *ASCL1*<sup>+</sup> cell types, N = 3 human fetal lungs, n = 243 *ASCL1*<sup>+</sup> cells.  
 (E) Left: HCR, *GRP* (green), *NEUROD1* (red), *GHRL* (white). Right: mean  $\pm$  SEM of *NEUROD1*<sup>+</sup> cell types: N = 2, 11 pcw human fetal lungs, n = 129; N = 3, 12 pcw human fetal lungs, n = 132. Scale bars, 25  $\mu$ m.  
 (F) Gene signature scoring of A-type and N-type SCLC features in the epithelial UMAP.  
 (G) Scenic analysis of predicted TF network governing mid tip progenitor cells to pulmonary NE and *GHRL*<sup>+</sup> NE. Trajectory and color coding match Figures S8A and S8B.  
 (H) Organoids from 8 pcw human fetal lungs were transduced with Doxycycline (Dox)-inducible TF, or *mNeonGreen-NLS*, lentivirus. Transduced organoids were isolated by flow cytometry based on TagRFP expression, seeded in Matrigel for 10–13 days prior to Dox treatment. Organoid cells were harvested 3 days post-Dox for scRNA-Seq. N = 3 organoid lines.  
 (I) Left: reference UMAP of primary human fetal lung epithelium. Mid and right: scRNA-Seq of organoids overexpressing *mNeonGreen-NLS*, *ASCL1*, or *NEUROD1* projected onto the primary data. See also Figure S8.



## DISCUSSION

Using a combination of single-cell and spatial approaches, we have identified 144 cell types, or states, in the developing human lungs across the 5–22 pcw period. We take advantage of a known proximal-distal gradient in epithelial differentiation to identify progenitor and differentiating states in the developing airway, including a neuroendocrine cell subtype related to SCLC. Moreover, analysis of the mesenchymal compartment identified three niche regions with distinct signaling interactions, allowing us to identify signaling conditions that are sufficient for airway differentiation of human embryonic lung organoids. We tested GRN predictions for NE cell differentiation in an organoid system, allowing us to identify lineage-defining TFs and provide directionality to the inferred differentiation trajectory. This study provides a paradigm for combining single-cell datasets with spatial analysis of the tissue and functional analyses in a human organoid system to provide mechanistic insights into human development.

Our data suggest that at all stages of lung development, cells exit the tip and enter a stalk state prior to differentiation. We propose that human alveolar epithelial differentiation also follows this model, using a tip-stalk-AT2 or AT1 fate decision pattern (Figure 3). This is different to the prevailing cellular models of mouse alveolar development: early cell fate restriction<sup>17,63</sup> and bipotent progenitors with AT1/2 characteristics.<sup>64</sup>

Airway, adventitial, and alveolar fibroblasts are localized in distinct niche regions and participate in different signaling interactions. Airway and adventitial fibroblasts both express unique combinations of signaling molecules and also form physical barriers between the neighboring airway epithelium or vascular endothelium and the widespread alveolar fibroblasts (Figures 4 and 5). Similarly, we characterize a population of myofibroblasts that contacts the developing epithelial stalk region and expresses high levels of the secreted Wnt-inhibitor, *NOTUM* (Figure S7O), whereas alveolar fibroblasts express high levels of the canonical *WNT2* ligand (Figure 4). In a separate study, using surface markers identified in this single-cell atlas, we specifically isolated alveolar fibroblasts and myofibroblast 2 cells for co-culture experiments with late-tip organoids.<sup>29</sup> Those experiments confirmed that a three-way signaling interaction between alveolar fibroblasts, myofibroblast 2 cells, and late-tip cells can control human AT2 spatial patterning.

We find that GHRL<sup>+</sup> NE cells are transcriptionally similar to the NEUROD1<sup>+</sup> N subtype of SCLC (Figure 7). Our functional analyses of NE cell differentiation in organoids will provide tools to test these hypotheses. Mouse studies show that fetal transcriptional and chromatin cell states are accessed during the normal process of tissue regeneration and may contribute to neoplasm in chronic inflammation.<sup>65,66</sup> Detailed ATAC-seq datasets are not yet available for human lung disease. Our high-quality ATAC-seq atlas will provide a baseline for further analyses when adult chromatin accessibility lung atlases are published. In summary, our multi-component atlas is a community resource for future analyses of human development, regeneration, and disease.

## Limitations of the study

We provide a carefully annotated, descriptive cell atlas resource. Many conclusions are derived from trajectory inference or TF

binding site analyses and require future validation. Trajectory inference analyses are largely based on transcriptomic similarities without ground-truth directionality, or are unable to handle complex expression kinetics in groups of genes.<sup>67</sup> For these reasons, we fed Monocle3<sup>68</sup> with starting and end points guided by known biological features of the data (age and spatial arrangement of cells). Furthermore, validation assays for lineage analysis in human systems rely on *in vitro* experiments. These usually define differentiation competence and do not necessarily mean that a specific differentiation route occurs *in vivo*. The clustering of our scRNA-seq and scATAC-seq data are in broad agreement. However, many motifs enriched in cell-type-specific peaks belong to TFs not detected by scRNA-seq. This discordance might be due to differing sensitivity of the two assays, transcription factor latency, and the incompleteness of the motif databases.

We have compared the identity of fetal and adult human lung cells and have seen many fetal-adult similarities. Nevertheless, there are approximately three decades between the oldest fetal and youngest adult human lung samples sequenced, including a rapid period of postnatal growth and morphogenesis, puberty, and unknown infections/environmental insults. It will be important to sequence additional lungs and, when possible, to fill the age gap. Moreover, our mouse-human fetal lung cell comparisons are affected by both technical (experimental protocols and annotation granularity) and biological differences (size and gestation rate). It will be informative in the future to make comparisons with a range of fetal lungs, including larger, long-developing species such as pig and sheep, to distinguish between differences due to species, size, and gestation period.

## STAR★METHODS

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Human lung tissue
- **METHOD DETAILS**
  - Cell isolation for 10X single cell RNA and ATAC seq
  - Human fetal lung organoid maintenance
  - Human fetal lung organoid bronchiolar differentiation
  - Isolation and airway differentiation of SCGB3A2+ distal and proximal airway cells
  - RNA extraction, cDNA synthesis, and qRT-PCR analysis
  - Human fetal lung organoid immunofluorescence
  - Plasmid cloning
  - Lentivirus packaging
  - Lentivirus transduction
  - Overexpression of transcription factors and scRNA-Seq

- *In situ* hybridization chain reaction and immunofluorescence
- Library generation and sequencing
- Visium spatial transcriptomics
- Reads mapping and quantification
- VDJ analysis
- Single-cell RNA-seq processing and cell type annotation
- Artefact evaluation and removal for scRNA-seq data
- Visium spatial transcriptomics data analysis
- Differential gene expression along trajectories
- CellPhoneDB analysis
- Velocity analysis
- Gene regulatory network analysis
- Comparing fetal neuroendocrine transcriptome with SCLC
- Comparing scRNA-seq datasets of the fetal lung and other studies
- Single-cell ATAC-seq processing and annotation
- Comparing organoid scRNA-seq with fetal lung scRNA-seq
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - HCR image analysis
  - Statistical analysis for cell-type composition biases
  - Marker gene calculation

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2022.11.005>.

## ACKNOWLEDGMENTS

We would like to acknowledge the Gurdon Institute Imaging Facility, and the Cellular Genetics IT and Phenotyping group, New Pipeline Group and DNA pipelines of Sanger Institute; Menna Clatworthy and Muzz Haniffa and R.E., C.S., E.D., I.G., M.H., C.D., and W.S. for discussions on cell-type annotations; and M.P., A.P., S.L., W.K.T., P.M., and C.T. for informatics support. K.L. is supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1A6A3A03012122). D.S. is supported by a Wellcome Trust PhD studentship (109146/Z/15/Z) and the Department of Pathology, University of Cambridge. P.H. holds a non-stipendiary research fellowship at St Edmund's College, University of Cambridge. E.M. is supported by ESPOD fellowship of EMBL-EBI and Sanger Institute. J.P.P. is supported by the MSCA Postdoctoral Fellowship. E.L.R. is supported by the MRC (MR/P009581/1; MR/S035907/1) and acknowledges the Gurdon Institute Core support from the Wellcome Trust (203144/Z/16/Z) and Cancer Research UK (C6946/A24843). Z.D. is supported by a Wellcome Trust PhD studentship (222275/Z/20/Z). R.A.B. is supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) and was an NIHR senior investigator. K.B.M., J.C.M., and S.A.T. acknowledge funding from the MRC (MR/S035907/1) and from Wellcome (WT211276/Z/18/Z and Sanger core grant WT206194). M.Z.N. 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, which received a proportion of funding from the Department of Health's NIHR Biomedical Research Centre's funding scheme.

## AUTHOR CONTRIBUTIONS

Conceptualization: P.H., K.L., D.S., K.B.M., and E.L.R. Methodology: P.H., K.L., and D.S. Software: P.H. Formal Analysis: P.H., J.P.P., K.P., and Z.K.T.

Investigation: K.L., D.S., Q.J., Z.D., L.B., L.R., L.M., M.D., A.W., and M.Y. Resources: E.M., X.H., R.A.B., and S.M.J. Data Curation: P.H., K.L., D.S., E.M., Z.K.T., E.D., C.S., and I.G. Writing – original draft: P.H., K.L., D.S., J.P.P., K.B.M., and E.L.R. Writing – review and editing: P.H., K.L., D.S., K.B.M., E.L.R., S.A.T., and J.B.M. Supervision: M.Z.N., R.A.B., S.A.T., J.B.M., K.B.M., and E.L.R. Funding Acquisition: K.L., E.M., J.P.P., R.A.B., M.Z.N., S.A.T., J.B.M., K.B.M., and E.L.R.

## DECLARATION OF INTERESTS

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

## INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community.

Received: December 13, 2021

Revised: August 11, 2022

Accepted: November 3, 2022

Published: December 8, 2022

## REFERENCES

1. Carraro, G., and Stripp, B.R. (2022). Insights gained in the pathology of lung disease through single-cell transcriptomics. *J. Pathol.* 257, 494–500.
2. Blenkinsopp, W.K. (1967). Proliferation of respiratory tract epithelium in the rat. *Exp. Cell Res.* 46, 144–154.
3. Rawlins, E.L., and Hogan, B.L.M. (2008). Ciliated epithelial cell lifespan in the mouse trachea and lung. *Am. J. Physiol. Lung Cell Mol. Physiol.* 295, L231–L234.
4. Burri, P.H. (1984). Fetal and postnatal development of the lung. *Annu. Rev. Physiol.* 46, 617–628.
5. Nikolić, M.Z., Sun, D., and Rawlins, E.L. (2018). Human lung development: recent progress and new challenges. *Development* 145, dev163485. <https://doi.org/10.1242/dev.163485>.
6. Rawlins, E.L., Clark, C.P., Xue, Y., and Hogan, B.L.M. (2009). The Id2+ distal tip lung epithelium contains individual multipotent embryonic progenitor cells. *Development* 136, 3741–3745.
7. Alanis, D.M., Chang, D.R., Akiyama, H., Krasnow, M.A., and Chen, J. (2014). Two nested developmental waves demarcate a compartment boundary in the mouse lung. *Nat. Commun.* 5, 3923–4015.
8. Nikolić, M.Z., Cariti, O., Jeng, Q., Johnson, J.A., Sun, D., Howell, K.J., Brady, J.L., Laresgoiti, U., Allen, G., Butler, R., and Zillbauer, M. (2017). Human embryonic lung epithelial tips are multipotent progenitors that can be expanded in vitro as long-term self-renewing organoids. *Elife* 6, e26575. <https://doi.org/10.7554/eLife.26575>.
9. Miller, A.J., Hill, D.R., Nagy, M.S., Aoki, Y., Dye, B.R., Chin, A.M., Huang, S., Zhu, F., White, E.S., Lama, V., and Spence, J.R. (2018). In vitro induction and in vivo engraftment of lung bud tip progenitor cells derived from human pluripotent stem cells. *Stem Cell Rep.* 10, 101–119.
10. Rawlins, E.L., Ostrowski, L.E., Randell, S.H., and Hogan, B.L.M. (2007). Lung development and repair: contribution of the ciliated lineage. *Proc. Natl. Acad. Sci. USA* 104, 410–417.
11. Trivedi, V., Choi, H.M.T., Fraser, S.E., and Pierce, N.A. (2018). Multidimensional quantitative analysis of mRNA expression within intact

- vertebrate embryos. *Development* 145, dev156869. <https://doi.org/10.1242/dev.156869>.
12. He, P., Williams, B.A., Trout, D., Marinov, G.K., Amrhein, H., Berghella, L., Goh, S.T., Plajzer-Frick, I., Afzal, V., and Pennacchio, L.A. (2020 Jul). The changing mouse embryo transcriptome at whole tissue and single-cell resolution. *Nature* 583, 760–767.
13. Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A., Aldinger, K.A., Blecher-Gonen, R., Zhang, F., and Spielmann, M. (2020). A human cell atlas of fetal gene expression. *Science* 370, eaba7721. <https://science.sciencemag.org/content/370/6518/eaba7721>.
14. Madisson, E., Oliver, A.J., Kleshchevnikov, V., Wilbrey-Clark, A., Polanski, K., Orsi, A.R., Mamanova, L., Bolt, L., Richoz, N., Elmentaite, R., and Pett, J.P. (2021). A spatial multi-omics atlas of the human lung reveals a novel immune cell survival niche. Preprint at bioRxiv. <https://www.biorxiv.org/content/10.1101/2021.11.26.470108v1>.
15. Cutz, E., Gillan, J.E., and Bryan, A.C. (1985). Neuroendocrine cells in the developing human lung: morphologic and functional considerations. *Pediatr. Pulmonol.* 1, S21–S29.
16. Negretti, N.M., Plosa, E.J., Benjamin, J.T., Schuler, B.A., Christian Habermann, A., Jetter, C., Gulleman, P., Taylor, C.J., Nichols, D., Matlock, B.K., and Guttentag, S.H. (2021). A Single cell atlas of lung development. Preprint at bioRxiv. <https://www.biorxiv.org/content/10.1101/2021.01.21.427641v3>.
17. Zepp, J.A., Morley, M.P., Loebel, C., Kremp, M.M., Chaudhry, F.N., Basil, M.C., Leach, J.P., Liberti, D.C., Niethamer, T.K., and Ying, Y. (2021). Genomic, epigenomic, and biophysical cues controlling the emergence of the lung alveolus. *Science* 371, eabc3172. <https://doi.org/10.1126/science.abc3172>.
18. Borromeo, M.D., Savage, T.K., Kolipara, R.K., He, M., Augustyn, A., Osborne, J.K., Girard, L., Minna, J.D., Gazdar, A.F., Cobb, M.H., and Johnson, J.E. (2016). ASCL1 and NEUROD1 reveal heterogeneity in pulmonary neuroendocrine tumors and regulate distinct genetic programs. *Cell Rep.* 16, 1259–1272.
19. Deprez, M., Zaragosi, L.E., Truchi, M., Becavin, C., Ruiz García, S., Arguel, M.J., Plaisant, M., Magnone, V., Lebrigand, K., and Abelanet, S. (2020). A single-cell atlas of the human healthy airways. *Am. J. Respir. Crit. Care Med.* 202, 1636–1645.
20. Carraro, G., Langerman, J., Sabri, S., Lorenzana, Z., Purkayastha, A., Zhang, G., Konda, B., Aros, C.J., Calvert, B.A., and Szymaniak, A. (2021). Transcriptional analysis of cystic fibrosis airways at single-cell resolution reveals altered epithelial cell states and composition. *Nat. Med.* 27, 806–814.
21. Vieira Braga, F.A., Kar, G., Berg, M., Carpaij, O.A., Polanski, K., Simon, L.M., Brouwer, S., Gomes, T., Hesse, L., and Jiang, J. (2019). A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* 25, 1153–1163.
22. Yao, E., Lin, C., Wu, Q., Zhang, K., Song, H., and Chuang, P.T. (2018). Notch signaling controls transdifferentiation of pulmonary neuroendocrine cells in response to lung injury. *Stem Cell.* 36, 377–391.
23. Ouadah, Y., Rojas, E.R., Riordan, D.P., Capostagno, S., Kuo, C.S., and Krasnow, M.A. (2019). Rare pulmonary neuroendocrine cells are stem cells regulated by Rb, p53, and Notch. *Cell* 179, 403–416.e23.
24. Kaarteenaho, R., Merikallio, H., Lehtonen, S., Harju, T., and Soini, Y. (2010). Divergent expression of claudin -1, -3, -4, -5 and -7 in developing human lung. *Respir. Res.* 11, 59.
25. Guha, A., Vasconcelos, M., Cai, Y., Yoneda, M., Hinds, A., Qian, J., Li, G., Dickel, L., Johnson, J.E., and Kimura, S. (2012). Neuroepithelial body microenvironment is a niche for a distinct subset of Clara-like precursors in the developing airways. *Proc. Natl. Acad. Sci. USA* 109, 12592–12597.
26. Yang, Y., Riccio, P., Schotsaert, M., Mori, M., Lu, J., Lee, D.K., García-Sastre, A., Xu, J., and Cardoso, W.V. (2018). Spatial-temporal lineage restrictions of embryonic p63+ progenitors establish distinct stem cell pools in adult airways. *Dev. Cell* 44, 752–761.e4.
27. Hawkins, F.J., Suzuki, S., Beermann, M.L., Barilla, C., Wang, R., Villacorta-Martin, C., Berical, A., Jean, J.C., Le Suer, J., and Matte, T. (2021). Derivation of airway basal stem cells from human pluripotent stem cells. *Cell Stem Cell* 28, 79–95.e8.
28. Negretti, N.M., Plosa, E.J., Benjamin, J.T., Schuler, B.A., Habermann, A.C., Jetter, C.S., Gulleman, P., Bunn, C., Hackett, A.N., and Ransom, M. (2021). A single-cell atlas of mouse lung development. *Development* 148, dev199512. <https://doi.org/10.1242/dev.199512>.
29. Lim, K., Donovan, A.P.A., Tang, W., Sun, D., He, P., Teichmann, S.A., Marioni, J.C., Meyer, K.B., Brand, A.H., and Rawlins, E.L. (2022). Organoid modelling of human fetal lung alveolar development reveals mechanisms of cell fate patterning and neonatal respiratory disease. *Cell Stem Cell*. Published online December 8, 2022. <https://doi.org/10.1016/j.stem.2022.11.013>.
30. Gillich, A., Zhang, F., Farmer, C.G., Travaglini, K.J., Tan, S.Y., Gu, M., Zhou, B., Feinstein, J.A., Krasnow, M.A., and Metzger, R.J. (2020). Capillary cell-type specialization in the alveolus. *Nature* 586, 785–789.
31. Vila Ellis, L., Cain, M.P., Hutchison, V., Flodby, P., Crandall, E.D., Borok, Z., Zhou, B., Ostrin, E.J., Wythe, J.D., and Chen, J. (2020). Epithelial vegfa specifies a distinct endothelial population in the mouse lung. *Dev. Cell* 52, 617–630.e6.
32. Schupp, J.C., Adams, T.S., Cosme, C., Jr., Raredon, M.S.B., Yuan, Y., Omote, N., Poli, S., Chioccioli, M., Rose, K.A., and Manning, E.P. (2021). Integrated single-cell atlas of endothelial cells of the human lung. *Circulation* 144, 286–302.
33. Takeda, A., Holmén, M., Dermadi, D., Pan, J., Brulois, K.F., Kaukonen, R., Lönnberg, T., Boström, P., Koskivuo, I., and Irla, H. (2019). Single-cell survey of human lymphatics unveils marked endothelial cell heterogeneity and mechanisms of homing for neutrophils. *Immunity* 51, 561–572.e5.
34. Frid, M.G., Dempsey, E.C., Durmowicz, A.G., and Stenmark, K.R. (1997 Jul). Smooth muscle cell heterogeneity in pulmonary and systemic vessels. Importance in vascular disease. *Arterioscler. Thromb. Vasc. Biol.* 17, 1203–1209.
35. Hein, R.F.C., Wu, J.H., Holloway, E.M., Frum, T., Conchola, A.S., Tsai, Y.H., Wu, A., Fine, A.S., Miller, A.J., Szenker-Ravi, E., and Yan, K.S. (2022). R-SPONDIN2+ mesenchymal cells form the bud tip progenitor niche during human lung development. *Dev. Cell*. <https://doi.org/10.1016/j.devcel.2022.05.010>.
36. Kumar, M.E., Bogard, P.E., Espinoza, F.H., Menke, D.B., Kingsley, D.M., and Krasnow, M.A. (2014). Mesenchymal cells. Defining a mesenchymal progenitor niche at single-cell resolution. *Science* 346, 1258810.
37. Dahlgren, M.W., and Molofsky, A.B. (2019 Oct). Adventitial cuffs: regional hubs for tissue immunity. *Trends Immunol.* 40, 877–887.
38. Travaglini, K.J., Nabhan, A.N., Penland, L., Sinha, R., Gillich, A., Sit, R.V., Chang, S., Conley, S.D., Mori, Y., and Seita, J. (2020). A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 587, 619–625.
39. Li, R., Li, X., Hagood, J., Zhu, M.S., and Sun, X. (2020). Myofibroblast contraction is essential for generating and regenerating the gas-exchange surface. *J. Clin. Invest.* 130, 2859–2871.
40. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* 15, 1484–1506.
41. Herbert, S.P., and Stainier, D.Y.R. (2011). Molecular control of endothelial cell behaviour during blood vessel morphogenesis. *Nat. Rev. Mol. Cell Biol.* 12, 551–564.
42. Miller, A.J., Yu, Q., Czerwinski, M., Tsai, Y.H., Conway, R.F., Wu, A., Holloway, E.M., Walker, T., Glass, I.A., and Treutlein, B. (2020). In vitro and



- in vivo development of the human airway at single-cell resolution. *Dev. Cell* 53, 117–128.e6.
43. Mou, H., Vinarsky, V., Tata, P.R., Brazauskas, K., Choi, S.H., Crooke, A.K., Zhang, B., Solomon, G.M., Turner, B., and Bihler, H. (2016). Dual SMAD Signaling Inhibition Enables Long-Term Expansion of Diverse Epithelial Basal Cells. *Cell Stem Cell* 19, 217–231.
  44. Domcke, S., Hill, A.J., Daza, R.M., Cao, J., O'Day, D.R., and Pliner, H.A. (2020). A human cell atlas of fetal chromatin accessibility. *Science* 370, eaba7612. <https://science.sciencemag.org/content/370/6518/eaba7612.abstract>.
  45. Quaggin, S.E., Schwartz, L., Cui, S., Igarashi, P., Deimling, J., Post, M., and Rossant, J. (1999). The basic-helix-loop-helix protein pod1 is critically important for kidney and lung organogenesis. *Development* 126, 5771–5783.
  46. Gao, X., Vockley, C.M., Pauli, F., Newberry, K.M., Xue, Y., Randell, S.H., Reddy, T.E., and Hogan, B.L.M. (2013). Evidence for multiple roles for grainyhead-like 2 in the establishment and maintenance of human mucociliary airway epithelium. *Proc. Natl. Acad. Sci. USA* 110, 9356–9361.
  47. Wan, H., Dingle, S., Xu, Y., Besnard, V., Kaestner, K.H., Ang, S.L., Wert, S., Stahlman, M.T., and Whitsett, J.A. (2005). Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J. Biol. Chem.* 280, 13809–13816.
  48. Corada, M., Orsenigo, F., Morini, M.F., Pitulescu, M.E., Bhat, G., Nyqvist, D., Breviaro, F., Conti, V., Briot, A., and Iruela-Arispe, M.L. (2013). Sox17 is indispensable for acquisition and maintenance of arterial identity. *Nat. Commun.* 4, 2609.
  49. van Soldt, B.J., Qian, J., Li, J., Tang, N., Lu, J., and Cardoso, W.V. (2019). Yap and its subcellular localization have distinct compartment-specific roles in the developing lung. *Development* 146, dev175810. <https://doi.org/10.1242/dev.175810>.
  50. Kimura, S., Ostrin, E.J., and Chen, J. (2019). Transcriptional control of lung alveolar type 1 cell development and maintenance by NK homeobox 2-1. *Proc. Natl. Acad. Sci. USA* 116, 20545–20555. <https://www.pnas.org/content/116/41/20545.short>.
  51. Liberti, D.C., Liberti, W.A., III, Kremp, M.M., Penkala, I.J., Cardenas-Diaz, F.L., Morley, M.P., Babu, A., Zhou, S., Fernandez, R.J., III, and Morrissey, E.E. (2022). Klf5 defines alveolar epithelial type 1 cell lineage commitment during lung development and regeneration. *Dev. Cell* 57, 1742–1757. <https://doi.org/10.1016/j.devcel.2022.06.007>.
  52. Rock, J.R., Onaitis, M.W., Rawlins, E.L., Lu, Y., Clark, C.P., Xue, Y., Randell, S.H., and Hogan, B.L.M. (2009). Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl. Acad. Sci. USA* 106, 12771–12775.
  53. Ito, T., Udaka, N., Yazawa, T., Okudela, K., Hayashi, H., Sudo, T., Guillemot, F., Kageyama, R., and Kitamura, H. (2000). Basic helix-loop-helix transcription factors regulate the neuroendocrine differentiation of fetal mouse pulmonary epithelium. *Development* 127, 3913–3921.
  54. Borges, M., Linnoila, R.I., van de Velde, H.J., Chen, H., Nelkin, B.D., Mabry, M., Baylin, S.B., and Ball, D.W. (1997). An achaete-scute homologue essential for neuroendocrine differentiation in the lung. *Nature* 386, 852–855.
  55. Gay, C.M., Stewart, C.A., Park, E.M., Diao, L., Groves, S.M., Heeke, S., Nabet, B.Y., Fujimoto, J., Solis, L.M., and Lu, W. (2021). Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. *Cancer Cell* 39, 346–360.e7.
  56. Ireland, A.S., Micinski, A.M., Kastner, D.W., Guo, B., Wait, S.J., Spainhower, K.B., Conley, C.C., Chen, O.S., Guthrie, M.R., and Soltero, D. (2020). MYC drives temporal evolution of small cell lung cancer subtypes by reprogramming neuroendocrine fate. *Cancer Cell* 38, 60–78.e12.
  57. Sikkema, L., Strobl, D., Zappia, L., Madisson, E., Markov, N.S., Zargosi, L., Ansari, M., Arguel, M.J., Apperloo, L., Becavin, C., and Berg, M. (2022). An integrated cell atlas of the human lung in health and disease. Preprint at bioRxiv. <https://www.biorxiv.org/content/10.1101/2022.03.10.483747v1?ct=>.
  58. Sun, D., Evans, L., Perrone, F., Sokleva, V., Lim, K., Rezakhani, S., Lutolf, M., Zilbauer, M., and Rawlins, E.L. (2021). A functional genetic toolbox for human tissue-derived organoids. *Elife* 10, e67886. <https://doi.org/10.7554/eLife.67886>.
  59. Aibar, S., González-Bias, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., and Aerts, J. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* 14, 1083–1086.
  60. Mellitzer, G., Beucher, A., Lobstein, V., Michel, P., Robine, S., Kedinger, M., and Gradwohl, G. (2010). Loss of enteroendocrine cells in mice alters lipid absorption and glucose homeostasis and impairs postnatal survival. *J. Clin. Invest.* 120, 1708–1721.
  61. Naya, F.J., Huang, H.P., Qiu, Y., Mutoh, H., DeMayo, F.J., Leiter, A.B., and Tsai, M.J. (1997). Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/neuroD-deficient mice. *Genes Dev.* 11, 2323–2334.
  62. Warner, S.M.B., Hackett, T.L., Shaheen, F., Hallstrand, T.S., Kicic, A., Stick, S.M., and Knight, D.A. (2013). Transcription factor p63 regulates key genes and wound repair in human airway epithelial basal cells. *Am. J. Respir. Cell Mol. Biol.* 49, 978–988.
  63. Frank, D.B., Penkala, I.J., Zepp, J.A., Sivakumar, A., Linares-Saldana, R., Zacharias, W.J., Stolz, K.G., Pankin, J., Lu, M., and Wang, Q. (2019). Early lineage specification defines alveolar epithelial ontogeny in the murine lung. *Proc. Natl. Acad. Sci. USA* 116, 4362–4371.
  64. Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509, 371–375.
  65. Larsen, H.L., and Jensen, K.B. (2021). Reprogramming cellular identity during intestinal regeneration. *Curr. Opin. Genet. Dev.* 70, 40–47.
  66. Jadhav, U., Saxena, M., O'Neill, N.K., Saadatpour, A., Yuan, G.C., Herbert, Z., Murata, K., and Shivdasani, R.A. (2017). Dynamic Reorganization of Chromatin Accessibility Signatures during Dedifferentiation of Secretory Precursors into Lgr5+ Intestinal Stem Cells. *Cell Stem Cell* 21, 65–77.e5.
  67. Barile M., Imaz-Rosshandler I., Inzani I., Ghazanfar S., Nichols J., Marioni J.C., Guibentif, C., and Göttgens, B. Coordinated Changes in Gene Expression Kinetics Underlie both Mouse and Human Erythroid Maturation. *Genome Biol.*, 22(1), pp.1-22 <https://doi.org/10.1101/2020.12.21.423773>
  68. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386.
  69. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* 145, dev165753. <https://doi.org/10.1242/dev.165753>.
  70. Mimitou, E.P., Cheng, A., Montalbano, A., Hao, S., Stoeckius, M., Legut, M., Roush, T., Herrera, A., Papalexi, E., and Ouyang, Z. (2019). Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* 16, 409–412.
  71. Conde C.D., Domínguez Conde C., Xu C., Jarvis L.B., Gomes T., Howlett S.K., Rainbow, D.B., Suchanek, O., King, H.W., and Mamanova, L. Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture in humans Preprint at bioRxiv. Available from: <https://doi.org/10.1101/2021.04.28.441762>
  72. Kaminow B., Yunusov D., and Dobin A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq

- data .Preprint at bioRxiv <https://www.biorxiv.org/content/10.1101/2021.05.05.442755.abstract>
73. Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., and participants in the 1st Human Cell Atlas Jamboree, and Marioni, J.C. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* 20, 63.
  74. Gadala-Maria, D., Yaari, G., Uduman, M., and Kleinstein, S.H. (2015). Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *Proc. Natl. Acad. Sci. USA* 112, E862–E870.
  75. Stephenson, E., Reynolds, G., Botting, R.A., Calero-Nieto, F.J., Morgan, M.D., Tuong, Z.K., Bach, K., Sungnak, W., Worlock, K.B., and Yoshida, M. (2021). Single-cell multi-omics analysis of the immune response in COVID-19. *Nat. Med.* 27, 904–916.
  76. Wolf, F.A., Alexander Wolf, F., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 1–5. <https://doi.org/10.1186/s13059-017-1382-0>.
  77. Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C.L., Lindberg, E.L., Kanda, M., Polanski, K., Heinig, M., and Lee, M. (2020). Cells of the adult human heart. *Nature* 588, 466–472.
  78. Madissoon, E., Wilbrey-Clark, A., Miragaia, R.J., Saeb-Parsy, K., Mahbubani, K.T., Georgakopoulos, N., Harding, P., Polanski, K., Huang, N., and Nowicki-Osuch, K. (2019). scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biol.* 21, 1.
  79. Popescu, D.M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E.F., Polanski, K., Goh, I., and Efremova, M. (2019). Decoding human fetal liver haematopoiesis. *Nature* 574, 365–371.
  80. Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B., Park, J.E., Stephenson, E., Polarski, K., and Goncalves, A. (2018). Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* 563, 347–353.
  81. Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H.W., Li, T., Elmentaite, R., Lomakin, A., Kedlian, V., and Gayoso, A. (2022). Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* 40, 661–671.
  82. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., III, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e21.
  83. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 20, 296.
  84. Polarski, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., and Park, J.E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* 36, 964–965.
  85. Haghverdi, L., Büttner, F., and Theis, F.J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* 31, 2989–2998.
  86. Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* 20, 59.
  87. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* 9, 1–2. <https://doi.org/10.1038/s41598-019-41695-z>.
  88. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., and Steemers, F.J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566, 496–502. <https://doi.org/10.1038/s41586-019-0969-x>.
  89. Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika* 37, 17–23.
  90. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>.
  91. Hahsler, M., Hornik, K., and Buchta, C. (2008). Getting things in order: an introduction to the R Package seriation. *J. Stat. Softw.* 25, 1–34. <https://doi.org/10.18637/jss.v025.i03>.
  92. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastrioti, M.E., Lönnerberg, P., and Furlan, A. (2018). RNA velocity of single cells. *Nature* 560, 494–498.
  93. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* 38, 1408–1414.
  94. Van de Sande, B., Flierin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., and Rouchon, Q. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* 15, 2247–2276.
  95. Moerman, T., Aibar Santos, S., Bravo González-Blas, C., Simm, J., Moreau, Y., Aerts, J., and Aerts, S. (2019). GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* 35, 2159–2161.
  96. Imrichová, H., Hulselmans, G., Atak, Z.K., Potier, D., and Aerts, S. (2015). i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res.* 43, W57–W64.
  97. Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. *Nat. Methods* 15, 1053–1058.
  98. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., and Vanderplas, J. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
  99. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* 53, 403–411.
  100. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS A J. Integr. Biol.* 16, 284–287. <https://doi.org/10.1089/omi.2011.0118>.
  101. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675.
  102. Heaton, H., Talman, A.M., Knights, A., Imaz, M., Gaffney, D.J., Durbin, R., Hemberg, M., and Lawnczak, M.K.N. (2020). Souporell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* 17, 615–620.
  103. Megill, C., Martin, B., Weaver, C., Bell, S., Prins, L., Badajoz, S., McCandless, B., Pisco, A.O., Kinsella, M., Griffin, F., and Kiggins, J. (2021). cellx-gene: a performant, scalable exploration platform for high dimensional sparse matrices. Preprint at bioRxiv. <https://www.biorxiv.org/content/10.1101/2021.04.05.438318v1.abstract>.
  104. Young, M.D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* 9, gaa151. <https://academic.oup.com/gigascience/article/9/12/gaa151/6049831>.
  105. Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* 8, 281–291.e9.
  106. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Mouse monoclonal anti-ACTA2	Thermo Fisher Scientific	Cat#MA1-06110; RRID: AB_557419
PE-conjugated Mouse monoclonal anti-THBD (CD141)	BioLegend	Cat#344104; RRID: AB_2255842
Rabbit monoclonal anti-PDGFRα	Cell Signaling Technology	Cat#3174; RRID: AB_2162345
Rabbit polyclonal anti-S100A4	Proteintech	Cat#16105-1-AP; RRID: AB_11042591
APC-conjugated Rat monoclonal anti-CD44	Thermo Fisher Scientific	Cat#17-0441-82; RRID:AB_469390
Rabbit polyclonal anti-SOX9	Merck	Cat#AB5535; RRID:AB_2239761
Sheep polyclonal anti-PDPN	R&D systems	Cat#AF3670; RRID:AB_2162070
Rat monoclonal anti-E-cadherin	Thermo Fisher Scientific	Cat#13-1900; RRID:AB_2533005
Chicken polyclonal anti-KRT5	BioLegend	Cat#905901; RRID:AB_2565054
Rabbit polyclonal anti-SCGB1A1	Proteintech	Cat#10490-1-AP; RRID:AB_2183285
Mouse monoclonal anti-FOXJ1	Thermo Fisher Scientific	Cat#14-9965-80; RRID:AB_1548836
Rabbit monoclonal anti-SCGB3A2	Abcam	Cat#14-9965-80; RRID: N/A
Mouse polyclonal anti-SCGB3A1	Novus Biological	Cat#MAB27901; RRID:N/A
<b>Biological samples</b>		
Organoid line: HDBR-L 13393, 15909	HDBR London	N/A
Organoid line: BRC 1943, 1915, 2174, 2315, 2316	Brain Repair Center, University of Cambridge	N/A
<b>Chemicals, peptides, and recombinant proteins</b>		
Proteinase K solution	Thermo Fisher Scientific	Cat#AM2546
N2 supplement	Thermo Fisher Scientific	Cat#17502001
B27 supplement	Thermo Fisher Scientific	Cat#12587001
N-acetylcysteine	Merck	Cat#A9165
EGF	PeproTech	Cat#AF-100-15
FGF10	PeproTech	Cat#100-26
FGF7	PeproTech	Cat#100-19
Noggin	PeproTech	Cat#120-10C
R-spondin	Stem Cell Institute, University of Cambridge	
CHIR99021	Stem Cell Institute, University of Cambridge	
SB431542	bio-technie	1614
cAMP	Merck	B5386
IBMX	Merck	I5879
Y-27632	Merck	688000
Dexamethasone	Merck	D4902
Doxycycline	Merck	D9891
<b>Critical commercial assays</b>		
Chromium Single Cell V(D)J Kits (v1)	10X genomics	
Visium Spatial Gene Expression Slide & Reagents Kit	10X genomics	
Chromium Next GEM Single Cell ATAC Kits (v1)	10X genomics	
In-Fusion® HD Cloning Plus	Takara	638910

(Continued on next page)



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Deposited data</b>		
scRNA-seq and scV(D)J of lung tissue	ArrayExpress	E-MTAB-11278
scRNA-seq of lung organoids	ArrayExpress	E-MTAB-11267
Visium spatial transcriptomics	ArrayExpress	E-MTAB-11265
scATAC-seq of lung tissue	ArrayExpress	E-MTAB-11266
<b>Oligonucleotides</b>		
Primer: TCR $\gamma/\delta$ library PCR1-R1_hTRDC: AGCTTGACAGCATTGTACTTCC	Mimitou et al. <sup>70</sup>	N/A
Primer: TCR $\gamma/\delta$ library PCR1-R1_hTRGC TGTGTCGTTAGTCTTCATGGTGTCC	Mimitou et al. <sup>70</sup>	N/A
Primer: TCR $\gamma/\delta$ library PCR2-R2_hTRDC TCCTTACCAGACAAGCGAC	Mimitou et al. <sup>70</sup>	N/A
Primer: TCR $\gamma/\delta$ library PCR2-R2_hTRGC GATCCCAGAATCGTGTGCTC	Mimitou et al. <sup>70</sup>	N/A
SI-PCR primer: AATGATACGGCGACCACCG AGATCTACACTCTTCCCTACACGACGC*T°C	Mimitou et al. <sup>70</sup>	N/A
<b>Recombinant DNA</b>		
Plasmid: pLenti-tetON-KRAB-dCas9-DHFR-EF1a-TagRFP-2A-tet3G	Sun et al. <sup>58</sup>	Addgene: #167935
Plasmid: pLenti-tetON-mNeonGreen-3XNLS-EF1a-TagRFP-2A-tet3G	this manuscript	N/A
Plasmid: pLenti-tetON-ASCL1-EF1a-TagRFP-2A-tet3G	this manuscript	N/A
Plasmid: pLenti-tetON-NEUROD1-EF1a-TagRFP-2A-tet3G	this manuscript	N/A
Plasmid: pLenti-tetON-NEUROG3-EF1a-TagRFP-2A-tet3G	this manuscript	N/A
Plasmid: pLenti-tetON-RFX6-EF1a-TagRFP-2A-tet3G	this manuscript	N/A
Plasmid: pLenti-tetON-TFAP2A-EF1a-TagRFP-2A-tet3G	this manuscript	N/A
Plasmid: pLenti-tetON-DeltaNP63-EF1a-TagRFP-2A-tet3G	this manuscript	N/A
Plasmid: pLenti-tetON-PAX9-EF1a-TagRFP-2A-tet3G	this manuscript	N/A
Plasmid: pLenti-tetON-NKX2-2-EF1a-TagRFP-2A-tet3G	this manuscript	N/A
Plasmid: pLenti-tetON-PROX1-EF1a-TagRFP-2A-tet3G	this manuscript	N/A
<b>Software and algorithms</b>		
python-genomics	this manuscript	<a href="https://github.com/brianpenghe/python-genomics">https://github.com/brianpenghe/python-genomics</a>
Seurat3-plus	this manuscript	<a href="https://github.com/brianpenghe/Seurat3-plus">https://github.com/brianpenghe/Seurat3-plus</a>
ImageJ (version: 2.1.0)	<sup>101</sup>	<a href="https://imagej.nih.gov/ij/">https://imagej.nih.gov/ij/</a> ; RRID:SCR_003070
GraphPad Prism software (version: 9.1.0)	GraphPad Software Inc.	GraphPad Prism ( <a href="https://graphpad.com/">https://graphpad.com/</a> ); RRID:SCR_015807
FlowJo software (version: 10.0.0)	FlowJo, LLC	FlowJo ( <a href="https://www.flowjo.com/">https://www.flowjo.com/</a> ); RRID:SCR_008520
Scanpy (version: 1.5.0, 1.8.1)	Wolf et al. <sup>76</sup>	<a href="https://github.com/theislab/scanpy">https://github.com/theislab/scanpy</a>
bbknn (version: 1.5.1)	Polański et al. <sup>84</sup>	<a href="https://github.com/Teichlab/bbknn">https://github.com/Teichlab/bbknn</a>
Scvelo (version 0.2.3)	Bergen et al. <sup>93</sup>	<a href="https://github.com/theislab/scvelo">https://github.com/theislab/scvelo</a>
Monocle 3 (version: 1.0.0)	<sup>68,88</sup>	<a href="https://github.com/cole-trapnell-lab/monocle3">https://github.com/cole-trapnell-lab/monocle3</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pySCENIC (version: 0.11.2)	<sup>59,94</sup>	<a href="https://github.com/aertslab/pySCENIC">https://github.com/aertslab/pySCENIC</a>
ComplexHeatmap (version 2.6.2)	Gu et al. <sup>90</sup>	<a href="https://github.com/jokergoo/ComplexHeatmap">https://github.com/jokergoo/ComplexHeatmap</a>
seriation (version: 1.3.0)	Hahsler et al. <sup>91</sup>	<a href="https://github.com/mhahsler/seriation">https://github.com/mhahsler/seriation</a>
souporcell (version: 2.0)	Heaton et al. <sup>102</sup>	<a href="https://github.com/wheaton5/souporcell">https://github.com/wheaton5/souporcell</a>
ArchR (version: 1.0.1)	Granja et al. <sup>99</sup>	<a href="https://github.com/GreenleafLab/ArchR">https://github.com/GreenleafLab/ArchR</a>
cellxgene (version: 0.16.7)	Megill et al. <sup>103</sup>	<a href="https://github.com/chanzuckerberg/cellxgene">https://github.com/chanzuckerberg/cellxgene</a>
clusterProfiler (version: 3.18.1)	Yu et al. <sup>100</sup>	<a href="https://github.com/YuLab-SMU/clusterProfiler">https://github.com/YuLab-SMU/clusterProfiler</a>
STARsolo (version: 2.7.3a)	Kaminow et al. <sup>72</sup>	<a href="https://github.com/alexdobin/STAR/blob/master/docs/STARsolo.md">https://github.com/alexdobin/STAR/blob/master/docs/STARsolo.md</a>
EmptyDrop	Lun et al. <sup>73</sup>	<a href="https://github.com/MarioniLab/DropletUtils">https://github.com/MarioniLab/DropletUtils</a>
cellranger (versions: 3.0.2, 4.0.0)	10X genomics	<a href="https://github.com/10XGenomics/cellranger">https://github.com/10XGenomics/cellranger</a>
cellranger-atac (version: 1.2.0)	10X genomics	<a href="https://github.com/10XGenomics/cellranger-atac">https://github.com/10XGenomics/cellranger-atac</a>
SoupX (version: 1.4.5)	Young et al. <sup>104</sup>	<a href="https://github.com/constantAmateur/SoupX">https://github.com/constantAmateur/SoupX</a>
dandelion (version: 0.1.10)	Stephenson et al. <sup>75</sup>	<a href="https://github.com/zktuong/dandelion">https://github.com/zktuong/dandelion</a>
Scrublet (version 0.2.1)	Wolock et al. <sup>105</sup>	<a href="https://github.com/swolock/scrublet">https://github.com/swolock/scrublet</a>
macs2 (version: 2.2.7.1)	Zhang et al. <sup>106</sup>	<a href="https://github.com/mac3-project/MACS">https://github.com/mac3-project/MACS</a>
Space Ranger (version: 1.1.0)	10X genomics	<a href="https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/what-is-space-ranger">https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/what-is-space-ranger</a>
Seurat (version 3.2.2)	Stuart et al. <sup>82</sup>	<a href="https://github.com/satijalab/seurat">https://github.com/satijalab/seurat</a>
sklearn (version: 0.24.2)	Pedregosa et al. <sup>98</sup>	<a href="https://github.com/scikit-learn/scikit-learn">https://github.com/scikit-learn/scikit-learn</a>
CellPhoneDB(version: 2.1.7)	Vento-Tormo et al. <sup>80</sup>	<a href="https://github.com/Teichlab/cellphonedb/">https://github.com/Teichlab/cellphonedb/</a>

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Emma L. Rawlins ([elr21@cam.ac.uk](mailto:elr21@cam.ac.uk)).

### Materials availability

Human lung organoid lines used in this study are available from the [lead contact](#), Emma L. Rawlins ([elr21@cam.ac.uk](mailto:elr21@cam.ac.uk)), with a completed Materials Transfer Agreement.

### Data and code availability

- Sequencing data have been deposited at ArrayExpress and ENA and are publicly available. Accession numbers are listed in the [key resources table](#). Processed sequencing data and microscopy data reported in this paper are available at <https://fetal-lung.cellgeni.sanger.ac.uk/>. ATAC-seq pseudobulk coverage profiles can be browsed at [https://genome.ucsc.edu/s/brianpenghe/scATAC\\_fetal\\_lung20211206](https://genome.ucsc.edu/s/brianpenghe/scATAC_fetal_lung20211206)
- All original code has been deposited at GitHub and is publicly available as of the date of publication. Links are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this work is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Human lung tissue

Human embryonic and fetal lung tissues were provided from terminations of pregnancy from Cambridge University Hospitals NHS Foundation Trust under permission from NHS Research Ethical Committee (96/085) and the MRC/Wellcome Trust Human Developmental Biology Resource (London and Newcastle, University College London (UCL) site REC reference: 18/LO/0822; Newcastle site REC reference: 18/NE/0290; Project 200419; [www.hdbbr.org](http://www.hdbbr.org)). Sample age ranged from 4 to 23 weeks of gestation (post-conception weeks; pcw). Stages of the samples were determined according to their external physical appearance and measurements. Sample names and gestational ages are listed in [Table S3](#). None of the samples used for the current study had any known genetic

abnormalities. Sample gender was unknown at the time of collection, but molecularly-inferred sample gender is available on the web interface (<https://lungcellatlas.org>).

Ethical approval for the adult human lung samples was given by the South Central Hampshire B Research Ethics Committee (REC reference 18/SC/0514, IRAS project: 245471) administered through the University College London Hospitals NHS Foundation Trust. Human adult lung samples were also obtained from Royal Papworth Hospital Tissue Research Bank (REC reference: 18/EE/0269).

## METHOD DETAILS

### Cell isolation for 10X single cell RNA and ATAC seq

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

### Human fetal lung organoid maintenance

Human fetal lung organoids were derived and maintained as previously described.<sup>8</sup> In brief, human foetal lung tissues were treated with Dispase (8 U/ml Thermo Fisher Scientific, 17105041) at room temperature (RT) for 2 min to digest mesenchymal connections. Mesothelium and mesenchymal cells were carefully removed by needles. Branching epithelial tips were micro-dissected by needles, transferred into Matrigel (356231, Corning) and seeded in a 24 well low-attachment plate (M9312-100EA, Greiner) with 4–5 tips per 50  $\mu$ L Matrigel dome per well. The plate was incubated at 37°C for 5–10 min until the Matrigel domes solidified. 600  $\mu$ L of self-renewing medium containing: N2 (1: 100, Thermo Fisher Scientific, 17502001), B27 (1: 50, Thermo Fisher Scientific, 12587001), N-acetylcysteine (1.25 mM), EGF (50 ng/mL, PeproTech, AF-100-15), FGF10 (100 ng/mL, PeproTech, 100-26), FGF7 (100 ng/mL, PeproTech, 100-19), Noggin (100 ng/mL, PeproTech, 120-10C), R-spondin (5% v/v, Stem Cell Institute, University of Cambridge), CHIR99021 (3  $\mu$ M, Stem Cell Institute, University of Cambridge) and SB 431542 (10  $\mu$ M, bio-technie, 1614), was added. Organoids were cultured under standard tissue culture conditions (37°C, 5% CO<sub>2</sub>), maintained in self-renewing medium and passaged by mechanically breaking using P200 pipettes every 4–7 days.

### Human fetal lung organoid bronchiolar differentiation

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

### Isolation and airway differentiation of SCGB3A2+ distal and proximal airway cells

Human fetal lungs at 8–11 pcw were carefully separated, and tip/stalk, distal airway, and proximal airway regions were further dissected using fine forceps under a dissecting microscope (Figure S5G). The tissue fragments were enzymatically digested into single cells by treating them in dissociation solution containing 0.125 mg/mL Collagenase, 1 U/ml Dispase and 0.1 U/ $\mu$ L DNAase, in a rotating incubator for 20 min at 37°C. The cells were treated with 1X RBC lysis buffer (Thermo Fisher, 00-4300-54), and were enriched by CD326 MACS beads according to the manufacturer's instructions. The enriched epithelial cells from distal and proximal regions were infected with a lentivirus harbouring SCGB3A2 promoter-driven EGFP with EF1a promoter driven-TagRFP. Next, the infected TagRFP cells were sorted by EGFP expression by FACS and analysed by qRT-PCR after 48 h (Figures S5G and S5H). The sorted distal and proximal SCGB3A2-GFP positive cells were cultured for 28 and 45 days in the airway differentiation medium.

### RNA extraction, cDNA synthesis, and qRT-PCR analysis

The cultured lung organoids were collected and lysed. Total RNA was extracted according to the RNeasy Mini Kit (Qiagen, 74004) procedure. cDNA synthesis was performed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, 4368814) and



the synthesised cDNA was diluted 1:20 for the qRT-PCR reaction (SYBR Green PCR Master Mix; Applied Biosystems, 4309155). Primer sequence information is listed in [Table S5](#). Data were presented as fold-change, calculated by ddCt method, using ACTB as a reference gene control.

### Human fetal lung organoid immunofluorescence

The differentiated organoids were released from the BME and fixed in 4% PFA at 4°C for 30 min. Then the organoids were washed in PBS, incubated in 0.3% PBTX (0.3% Triton X-100 in PBS) at 4°C for 1 h, and blocked (1% bovine serum albumin, 5% normal donkey serum, 0.3% Triton X-100 in PBS) at 4°C overnight. The organoids were incubated with primary antibodies: SCGB3A2 (1:800, Abcam, ab181853), KRT5 (1: 500, BioLegend, 905901), E-cadherin (1: 500; Thermo Fisher Scientific, 13-1900), SOX9 (1: 400; Merck, AB5535), SCGB1A1 (1: 800, Proteintech, 10490-1-AP), SCGB3A1 (1:200, Novus Biological, MAB27901), FOXJ1 (1: 300, eBioscience, 14-9965-80) at 4°C overnight. The organoids were washed by PBS and further incubated with secondary antibodies (donkey anti-chicken 488, 1: 1000, Jackson ImmunoResearch, 703-545-155; donkey anti-mouse 594, 1: 1000, Invitrogen, A-21203; donkey anti-rabbit 594, 1: 1000, Invitrogen, A-21207; donkey anti-rat 647, 1: 1000, Jackson ImmunoResearch, 712-605-153; donkey anti-rabbit 647, 1: 1000, Invitrogen, A-31573). After DAPI staining (1 µg/mL) at 4°C for 1 hour, the organoids were processed through a thiodiethanol series (25%, 50%, 75% and 97% v/v concentration in PBS) at 4°C for imaging.

### Plasmid cloning

cDNAs for genes ASCL1, NEUROD1, NEUROG3, RFX6 and PAX9 were purchased from Genscript. cDNAs for gene TFAP2A and mNeonGreen-3XNLS were gifts from Azim Surani's Group. cDNA for DeltaNTP63 was purchased from IDT as a gBlock fragment. cDNA sequences were cloned into a Doxycycline inducible vector pLenti-tetON-KRAB-dCas9-DHFR-EF1a-TagRFP-2A-tet3G (Addgene: #167935)<sup>58</sup> using XhoI and BamHI sites by Infusion cloning (Takara, 638910).

A promoter region (chr5:147,878,065 + 147,878,803; 739 bp) of SCGB3A2 was amplified using primers: 5'-AATTGAATCCCA GGTTTTTCAAAGACACT-3' and 5'-GACAGTTATCTGGGATATTTTCAGGAGTTT-3'. The amplicons were cloned into a lentiviral vector, pLenti-(promoter)-EGFP/EF1a-TagRFP by Infusion (Takara, 638909). Plasmids used in this study will be deposited to Addgene.

### Lentivirus packaging

We packaged the lentivirus as described previously.<sup>58</sup> In brief, HEK293T cells were grown in 10-cm dishes to 70–80% confluence. Lentiviral vector (10 µg) was co-transfected with packaging MD2.G (3 µg, Addgene plasmid # 12259), psPAX2 (6 µg, Addgene plasmid # 12260) and pAdVantage (3 µg, E1711, Promega) using Lipofectamine 2000 Transfection Reagent (11668019, Thermo Fisher Scientific) according to manufacturer's protocol. Medium was refreshed the next morning. Lentivirus containing cell medium was harvested at 24 and 48 h after medium refreshing and pooled together. Cell fragments were removed by 300X g centrifugation. Supernatant was then passed through a 0.45 µm filter. Lentivirus was concentrated using Lenti-X™ Concentrator (631232, Takara) according to the manufacture's instructions. Lentivirus pellets were dissolved in 400 µL PBS, aliquoted and frozen in –80°C.

### Lentivirus transduction

Lentivirus transduction was performed as previously described.<sup>58</sup> In brief, human fetal lung organoids derived from 3 independent donors were incubated with prewarmed TrypLE for 10 min with trituration after 5 min. Organoid single cells and small fragments were collected, counted, pelleted and resuspended to around 100K cells/500 µL self-renewing medium with ROCKi (10 µM Y-27632). 0.5–2 µL of lentivirus was added and incubated overnight. Organoid cells were harvested the next morning, pelleted and re-seeded into Matrigel.

### Overexpression of transcription factors and scRNA-Seq

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

### In situ hybridization chain reaction and immunofluorescence

*In situ* HCR v3.0 was performed according to the manufacturer's protocol (Molecular Instruments).<sup>69</sup> Probes were designed according to the manual, and amplifiers with buffers were supplied by Molecular Instruments. All the sequence information of the probes is listed in [Table S5](#). In brief, the frozen human tissue sections fixed in 4% PFA/DEPC-treated PBS were cut into 20 µm slices and rinsed in nuclease-free ultrapure water, followed by 10 µg/mL proteinase K solution (Thermo Fisher Scientific, AM2546) for 2 min at 37°C. For *in situ* HCR with immunostaining, the tissue slices were permeabilized in 0.3% Triton-X/DEPC-treated PBS for 5 min at room temperature, avoiding the treatment of the proteinase K solution. Next, the tissue slices were incubated with 2 pmol of probes at 37°C overnight. After washing, the slices were treated with 6 pmol of the amplifiers at room temperature overnight. The amplifiers,

consisting of a pair of hairpins conjugated to fluorophores, Alexa 488, 546, or 647, were used at final concentration of 0.03  $\mu$ M. Then, excess hairpins were rinsed in 5X SSC (sodium chloride sodium citrate) solution containing 0.1% Triton X-100. Nuclei were counterstained with DAPI. For the immunostaining following the *in situ* HCR, the tissue slices were incubated with a blocking solution containing 5% NDS, 1% BSA, 0.1% Triton-X in DEPC-treated PBS at room temperature for 1 h after the hairpin amplification. After rinsing with DEPC-treated PBS, treated with primary antibodies against ACTA2 (1:500; Thermo Fisher Scientific, MA1-06110), THBD (1:100; PE-conjugated; BioLegend, 344104), PDGFRA (1:200; Cell Signaling Technology, 3174), S100A4 (1:200; Proteintech, 16105-1-AP), CD44 (1:200; Thermo Fisher Scientific, 17-0441-82), SOX9 (1: 200, Merck, AB5535), PDPN (1:200; R&D Systems, AF3670), or E-cadherin (1: 500; Thermo Fisher Scientific, 13-1900) overnight. Secondary antibodies were treated for 3 h at room temperature. The tissue was washed three times in DEPC-treated PBS at room temperature and counterstained with DAPI. Images were collected under Leica SP8 confocal microscope.

### Library generation and sequencing

Chromium Single Cell 5' V(D)J Reagent Kits (V1.0 chemistry) were used for scRNAseq library construction. Gene expression libraries (GEX) and V(D)J libraries were prepared according to the manufacturer's protocol (10X Genomics) using individual Chromium i7 Sample Indices. Libraries for gamma/delta TCR variable regions were amplified as previously described.<sup>70,71</sup> GEX and V(D)J were pooled in 1:0.1 ratio respectively and sequenced on a NovaSeq 6000 S4 or Illumina HiSeq 4000 Flowcell (paired-end (PE), 150-bp reads) aiming for a minimum of 50,000 PE reads per cell for GEX libraries and 5,000 PE reads per cell for V(D)J libraries.

### Visium spatial transcriptomics

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

### Reads mapping and quantification

scRNA-seq data were mapped with STARsolo 2.7.3a<sup>72</sup> to the 10X distributed GRCh38 reference, version 3.0.0, derived from Ensembl 93. Cell calling was post-processed with an implementation of EmptyDrops<sup>73</sup> extracted from Cell Ranger 3.0.2 (distributed as empty drops on PyPi). For transduced organoid cells, exogenous genes were added to the reference as appropriate for organoids, with the transgene sequence truncated (length(R2)-1) bp after the end of the synthetic promoter to avoid reads from endogenous transcripts being mapped onto transgenes. For single-cell V(D)J data, reads were mapped with Cell Ranger 4.0.0 to the 10X distributed VDJ reference, version 4.0.0. Visium reads were mapped with Space Ranger 1.1.0 to the 10X distributed GRCh38 reference, version 3.0.0, derived from Ensembl 93 for consistency with the single cell data. scATAC reads were mapped with Cellranger-atac 1.2.0 to reference GRCh38-1.2.0.

### VDJ analysis

Both TCR and BCR contigs contained in respective all\_contigs.fasta and all\_contig\_annotations.csv files were re-annotated with igblastn (v1.17.1) using reference sequences curated from IMGT database (downloaded 01-Aug-2021) as per described with changeo (v1.0.0). For BCR contigs, heavy chain constant region calls were re-annotated using blastn (v2.12.0+) against curated sequences of CH1 regions corresponding to respective isotype classes from IMGT. BCR heavy chain V-gene alleles were corrected for individual genotypes using tigger (v1.0.0).<sup>74</sup> Contigs were then filtered for basic quality control as described previously.<sup>75</sup> Briefly, the following occurrences would lead to removal of contigs from further analysis: i) contigs were annotated with V, D, J or constant gene calls that are not from the same locus; ii) multiple long/heavy chain contigs present in the same cell; iii) there were only short/light chain contigs in a cell; and/or iv) there are multiple short/light chain contigs in a cell. Cells with multiple contigs were nevertheless retained if a) contigs were assessed to have identical V(D)J sequences but were assigned to a different contig by cellranger-udj (presumably due to differences in non-V(D)J elements); b) UMI count differences were large in which case the contig with the highest UMI count is retained; and c) only IgM and IgD were both assigned to a cell. These checks were all performed using dandelion<sup>75</sup> singularity container (v0.1.10).

### Single-cell RNA-seq processing and cell type annotation

Count matrices were loaded into Scanpy and concatenated. Cells expressing no more than 200 genes, and genes detected in no more than 5 cells, were removed. Cells having more than 20% of their reads mapped to mitochondria were also discarded. Counts were then divided by total counts and multiplied by a factor of 10000, followed by log transformation, all implemented in Scanpy's default setting.<sup>76</sup>  $Y_{ij} = \ln \left( \frac{X_{ij}}{\sum_{i=1}^n X_{ij}} \cdot 10000 + 1 \right)$ , where  $X_{ij}$  is the raw count of  $i^{\text{th}}$  gene in  $j^{\text{th}}$  cell.

Feature genes were selected in three steps: For each sample, highly variable genes were calculated using Scanpy's default settings that extract genes with highest dispersion (variance divided by mean) values of log-transformed counts. Next, highly correlated genes for each sample were extracted using the DeepTree algorithm described in,<sup>12</sup> reimplemented in our python-genomics toolkit. Genes extracted in at least two samples were merged as the final feature gene list. The log-transformed counts of these genes were then scaled after cell-cycle scores were regressed out using Scanpy's default scoring and regression functions. Using the top 50 PCs and 10 neighbors with resolution at 0.01, initial clustering was generated, yielding 10 major clusters (Figure S1C) corresponding to different compartments. These clusters were subsequently and recursively subclustered, curated and annotated manually (Figure S1D). Annotation was based on markers summarised in Table S1.

### Artefact evaluation and removal for scRNA-seq data

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

Maternally derived cells were evaluated based on SNP variations between the transcribed paternal genome in the fetuses and the maternal counterparts in the maternal cells. To do this, we indexed and pooled samples from the same donor into "Supersamples". Then we applied Souporcell to compare known common variants captured in scRNA-seq reads, setting the sample number to 2. Supersamples without maternal cells would split into two equal-sized groups while other supersamples would putatively capture maternal cells as a minor genotype group (Figure S1F). Based on this analysis, maternal-like cells do not contribute to scRNA-seq clusters (Figure S1J) and were thus kept for downstream analysis. For libraries with two multiplexed donors, we only used the Souporcell workflow to demultiplex the donors without maternal genetic detection.

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

An additional four clusters of contaminants coming from other organs were further removed (Figure S1H). These were cardiomyocytes (ACTN2<sup>+</sup> MYH6<sup>+</sup>),<sup>77</sup> esophagus epithelial cells (SOX2<sup>+</sup> TP63<sup>+</sup> TRH<sup>+</sup>),<sup>78</sup> APOA1<sup>+</sup> APOA2<sup>+</sup>)<sup>79</sup> and cytotrophoblasts from the placenta (PAGE4<sup>+</sup> GSTA3<sup>+</sup>).<sup>80</sup>

### Visium spatial transcriptomics data analysis

Two methods were used side by side to predict cell compositions of the Visium datasets. Mapped Visium and filtered scRNA-seq data (removing cell types that have fewer than 20 cells) were both fed into the default pipeline of the cell2location algorithm,<sup>81</sup> with the default detection alpha set to 20. The q05\_cell\_abundance was used as a conservative estimate of cell abundance in each voxel. This method was used to generate figure panels in this manuscript. In the alternative method, mapped Visium count matrices and scRNA-seq count matrices (after artefact removal) were both imported into Seurat 3<sup>82</sup> and transformed using SCTransform,<sup>83</sup> with mitochondria percentage of scRNA-seq data regressed out. Next, the scRNA-seq data were subsetted into a "pcw11,15,18" subgroup and a "pcw18,20,22" subgroup for cell-type prediction. The prediction was done for each Visium library using its corresponding scRNA-seq subgroup following the default label transfer pipeline of Seurat using the top 50 PCs. We provide the results of both methods on our data portal.

### Differential gene expression along trajectories

The single cell transcriptomics data was preprocessed using Scanpy<sup>76</sup> version 1.8.1. The cell cycle effect was regressed out using scanpy.pp.regress\_out and batch correction was performed using bbknn,<sup>84</sup> before denoising the knn-graph using diffusion maps<sup>85</sup> with scanpy.tl.diffmap and applying PAGA<sup>86</sup> with scanpy.tl.paga to examine the connectivities between cell types. The final UMAPs were computed using the results of PAGA on Leiden<sup>87</sup> clusters as previously described.<sup>86</sup> Data and UMAPs were exported into R, and monocle3<sup>68,88</sup> was used to find a principal graph and define pseudotime. Differentially expressed genes were then computed along pseudotime using a graph-based test (morans' I)<sup>88,89</sup> and the principal graph in monocle3, which allows identification of genes upregulated at any point in pseudotime. The results were visualised with heatmaps using the complexHeatmap<sup>90</sup> and seriation<sup>91</sup> packages, after smoothing gene expression with smoothing splines in R (smooth.spline, df = 12).

### CellPhoneDB analysis

Filtered single-cell RNA-seq data were partitioned into early- (5–6pcw), middle- (9–11) and late-stage (15–22) subsets and grouped into broad cell types. These datasets were used as input for CellPhoneDB<sup>90</sup> (command: cellphonedb method statistical\_analysis -database v2.0.0 -threads 20 -counts-data gene\_name -project-name FetalLungBroad -subsampling -subsampling-log False -subsampling-num-cells = \$TotalCellNumber -iterations = 10000 -result-precision = 4). Interaction pairs were manually curated from the outputs.



### Velocity analysis

Velocity analysis<sup>92</sup> was performed using scvelo<sup>93</sup> version 0.2.3. The preprocessed dataset was merged with spliced and unspliced read counts computed with velocyto, before using scvelo.pp.moments, scvelo.tl.velocity and scvelo.tl.velocity\_graph to compute velocities using the stochastic mode in scvelo.

### Gene regulatory network analysis

The Scenic pipeline<sup>59,94</sup> was used (pySCENIC version 0.11.2) to predict transcription factors and putative target genes regulated throughout neuroendocrine cell differentiation. First, gene regulatory interactions were calculated based on co-expression across the single cell dataset with GRNBoost2,<sup>95</sup> followed by pruning interactions using known TF binding motifs and the construction of dataset specific regulatory modules (regulons).<sup>96</sup> Regulons were then scored in each individual cell using AUCell. Cells of the neuroendocrine differentiation trajectory computed with monocle3 (as described above) were selected. The regulon target genes were filtered for differentially expressed genes along pseudotime for this trajectory. A network of TFs and target genes was then constructed by linking individual regulons.

### Comparing fetal neuroendocrine transcriptome with SCLC

A-type and N-type signatures were selected from previous data 'ASCL1High and NEUROD1High Gene Signatures and the Stratified Primary Tumor Samples'.<sup>18</sup> Top 10 genes with the highest fold enrichment were selected to score epithelial cells, using Scanpy's tl.score\_genes function.

### Comparing scRNA-seq datasets of the fetal lung and other studies

Annotated scRNA-seq adult lung datasets,<sup>14</sup> the multi-organ scRNA-seq dataset,<sup>13</sup> and the mouse scRNA-seq dataset<sup>17</sup> were downloaded. Orthologs were translated from mouse to human counterparts using ENSEMBL biomart. scVI<sup>97</sup> was used to integrate our fetal lung scRNA-seq and the Madisson et al. and Zepp et al. data (human-mouse orthologs only), with sample IDs and project IDs both included as categorical covariate keys (other parameters: n\_latent = 30, encode\_covariates = True, dropout\_rate = 0.2, n\_layers = 2, early\_stopping = True, train\_size = 0.9, early\_stopping\_patience = 45, max\_epochs = 400, batch\_size = 1024, limit\_train\_batches = 20, use\_gpu = True). The latent variables calculated by xcVI were fed into Scanpy's pl.correlation\_matrix function to calculate and visualise correlation scores. A logistic regression model was trained based on the fine-grained cell-types for each of the multi-organ data, using sklearn.linear\_model.LogisticRegression.<sup>98</sup> The trained model was then used to predict the cell types of single-cell transcriptomic profiles of the fetal lung (Figure S1O).

### Single-cell ATAC-seq processing and annotation

Cellranger-atac outputs were loaded into and processed by ArchR.<sup>99</sup> The top 50 dimensions were used for LSI and no batch effect was carried out to preserve weak biological features. Doublets were removed using ArchR's default settings. Cells with TSSEnrichment score <8 or ReadsInTSS <1000 were discarded. Initial clustering was performed at resolution = 0.01 to be consistent with scRNA-seq, resulting in 7 large clusters corresponding to compartments. These clusters were further subclustered, similar to the workflow for scRNA-seq.

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

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

### Comparing organoid scRNA-seq with fetal lung scRNA-seq

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

## QUANTIFICATION AND STATISTICAL ANALYSIS

### HCR image analysis

*In situ* HCR images were analyzed using ImageJ (<https://imagej.nih.gov/ij/>) for quantification and statistical analysis. Cells expressing airway lineage markers along distal to proximal airway axis at different ages, mid (10–12 pcw) and late (15–21 pcw) stages, were counted (Figure S4B). For measuring the proportion of proximal secretory lineage cells within proximal cartilaginous airway regions, the fetal tissue sections at 10–12, 15–16, and 19–21 pcw were analysed based on expression patterns of SCGB3A2, SCGB3A1, and/or SCGB1A1 (Figure S4C). Mean, SD, 1-way ANOVA, and 2-way ANOVA were calculated using the Prism software (GraphPad Prism). Significance was evaluated by 1 or 2-way ANOVA with Tukey multiple comparison post-test; ns = not significant, \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001.

### Statistical analysis for cell-type composition biases

Chi-squared test of independence was performed for sample gestation age, cell-cycle stage and proximal/distal dissection regions against cell type categories. For proximal/distal biases, Fisher exact test was used for each cell type and Benjamini-Hochberg correction was performed for multiple testing.

We also visualised the effect size for cell composition biases over developmental age and proximal/distal dissection in Figure S1O. After removing the clusters that are specific to the very early stages in low abundance (such as neuronal clusters), cell number counts were normalised against total number counts per stage. The mean developmental stage for each cluster was calculated based on the empirical distribution based on the aforementioned normalised counts, denoted by  $x$ . The weighted probability  $y$  of proximal representation was calculated as the frequency of cells from proximal samples normalised against total numbers of cells from proximal samples, ignoring whole-lung samples. The  $x$  and  $y$  values were calculated for Figure S1O.

$$x_t = \sum_{s=5}^{22} sp_{s|t}, \text{ where } p_{s|t} = \frac{C_{t,s} / \sum_{t=1}^{tn} C_{t,s}}{\sum_{s=5}^{22} (C_{t,s} / \sum_{t=1}^{tn} C_{t,s})}$$

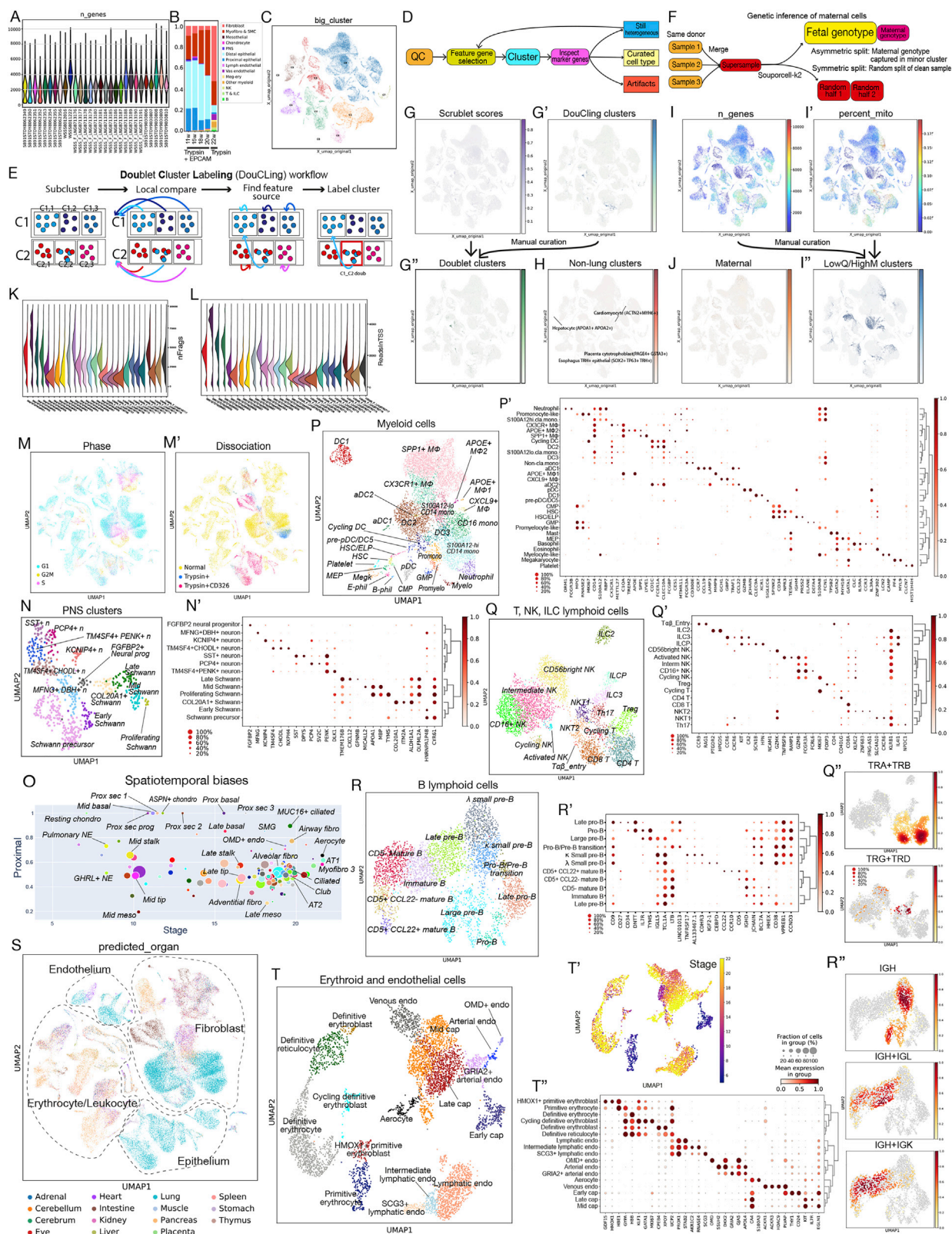
$$y_t = \frac{C_{t,prox} / C_{prox}}{C_{t,prox} / C_{prox} + C_{t,dist} / C_{dist}}$$

where  $(x_t, y_t)$  are the  $x$  and  $y$  coordinates of a cell type  $t$ ,  $s$  is the post conception week,  $C_{t,s}$  is the number of cells labelled as cell type  $t$  at stage  $s$ ,  $C_{t,prox}$  and  $C_{t,dist}$  are the numbers of cells labelled as cell type  $t$  coming from proximal and distal samples, respectively.

### Marker gene calculation

Ambient RNA was removed with SoupX 1.4.5 with default parameters. Using the corrected count matrices, Scanpy.tl.rank\_genes\_groups was applied with default settings but keeping all the genes. These ranked genes were then filtered using Scanpy.tl.filter\_rank\_genes\_groups with `max_out_group_fraction = 0.25` and `min_fold_change = 2`. To compare specific cell types in the same compartment, Scanpy.tl.rank\_genes\_groups was applied for each cell type with only the other cell types of this compartment as a reference. Over-representation analysis (hypergeometric test) with gene sets from GO BP, KEGG and MSigDB was performed using the clusterProfiler R package.<sup>100</sup>

# Supplemental figures

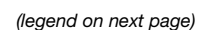


(legend on next page)



**Figure S1. Quality control for scRNA-seq and scATAC-seq data and clustering overview of 144 cell types or cell states, related to Figures 1 and S7**

- (A) Distributions of the number of genes detected per cell, grouped by 10X libraries.
- (B) Proportions of broad cell types in samples treated with Trypsin and Trypsin plus EPCAM enrichment following color codes in Figure 1C.
- (C) Initial clusters of data-separating compartments, before subclustering.
- (D–F) Workflows of the recursive subclustering method (D), the Doublet Cluster Labeling (DouCLing) method to identify doublet-driven clusters (E), and inference of maternal cells using Soupcorell (F).
- (G) Doublet scores by Scrublet (G), inferred doublet clusters by DouCLing (G'), and cells in curated doublet clusters (G'').
- (H) Cells in clusters of cells coming from other organs. Marker genes in parentheses.
- (I) Number of genes (I), percentage of mitochondrial reads (I'), and cells in curated low-quality cell clusters (I'') were projected on UMAP.
- (J) Inferred maternal cells.
- (K and L) scATAC-seq quality metrics of fragment detection per cell (K) and reads mapped in transcription-start sites (L).
- (M) All of the curated 144 clusters of single cells projected on UMAP space of transcriptomes, colored by inferred cell-cycle phase (M), and dissociation/enrichment strategy (M').
- (N) Cells from the initial PNS cluster (C7) projected on UMAP space of transcriptomes, colored by cell type/state (N) and selected feature genes of cell types/states (N') in the initial PNS cluster.
- (O) Spatiotemporal biases of cell types. Cell types are shown as dots, with *x* representing the weighted average of developmental stages, *y* representing the score of proximal enrichment, and the size corresponding to the cluster size.
- (P) UMAP embedding (P) and dot plots (P') of myeloid cell types/states.
- (Q) UMAP embedding (Q), dot plots (Q'), and enrichment of each class of immune receptors based on abTCR, gdTCR, and BCR-enriched scRNA-seq (Q''), in T, NK, and ILC lymphoid cell compartments.
- (R) UMAP embedding (R), dot plots (R'), and enrichment of immune receptors based on BCR-enriched scRNA-seq (R''), in B lymphoid cell compartment.
- (S) Predicted organ-of-source with highest scores for cells shown in Figure 1, based on the reference atlas in Cao et al.<sup>13</sup>
- (T) UMAP visualization of erythroid and endothelial cells colored by cell types/states (T), stages (T'), and dot plot describing differential marker gene expression level by cell type (T'').

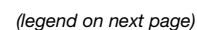


---

**Figure S2. Comparing fetal lung scRNA-seq with adult human and mouse lung scRNA-seq, related to [Figure 1](#)**

(A–F) Correlations of scVI latent variables between human fetal lung cell clusters and those of previously annotated adult cell clusters<sup>1</sup> (A–C) and mouse lung cell clusters<sup>2</sup> (D–F), focusing on epithelial (A, D), fibroblast (B, E), and endothelial (C, F) compartments.  
(G) Expression of genes shared or unique to fetal/adult lung AT1/AT2 cell clusters.





**Figure S3. Spatial analysis of airway epithelial cells in the developing human lungs by *in situ* HCR, related to Figure 2**

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

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

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

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

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

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

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

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

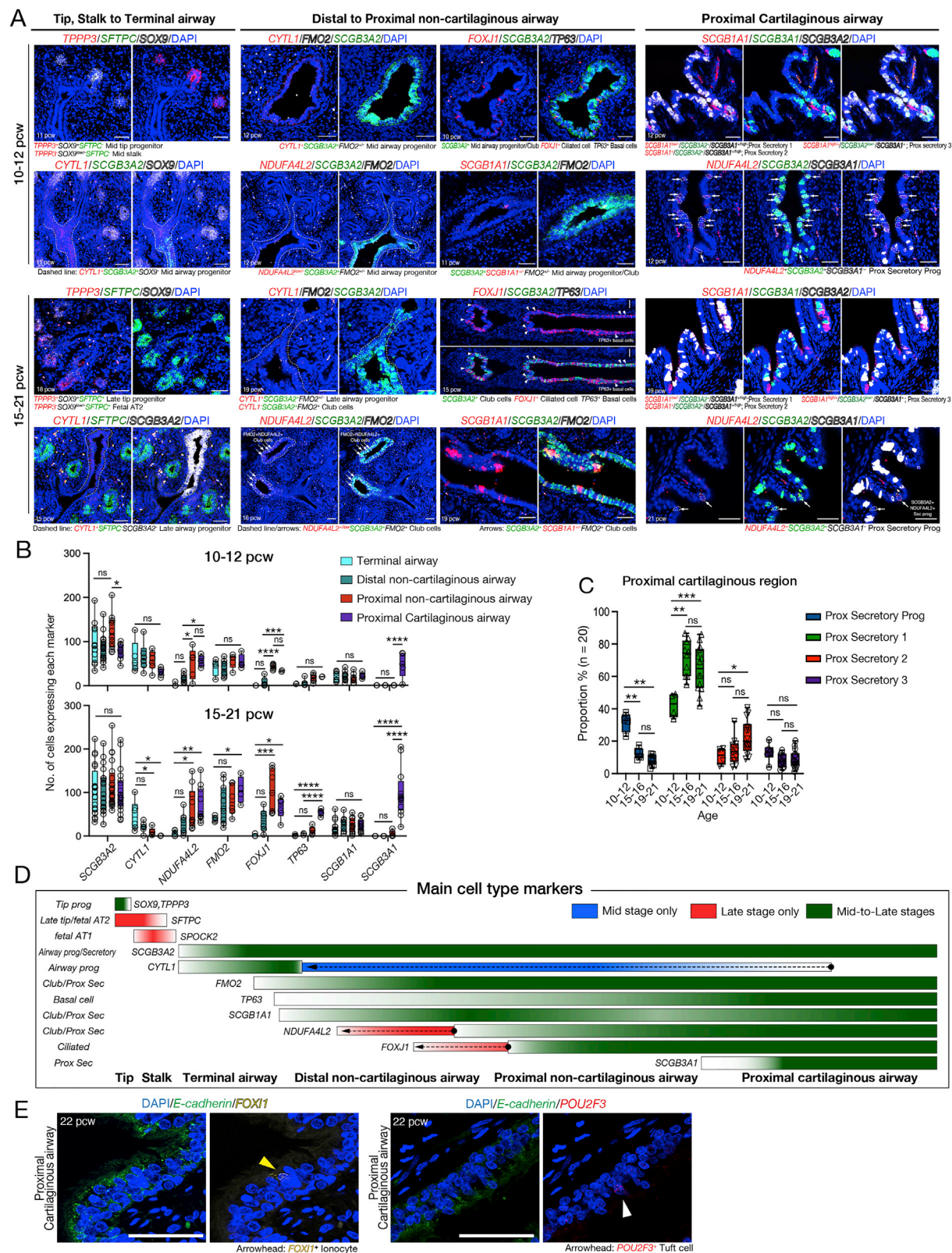
(I)  $MUC16^+$  ciliated cells (dashed line), ciliated cells (dashed circle), and secretory cells (arrow) located in the proximal cartilaginous airway regions of fetal lungs at 19 pcw. The  $MUC16^+$  ciliated cells express  $MUC16$  (white) with a weak level of  $FOXJ1$  (red), whereas the ciliated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

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

(K)  $ASCL1^+$  pulmonary neuroendocrine (arrow) and  $MUC5AC^+/ASCL1^+$  progenitors (arrowhead) in the non-cartilaginous airway regions of fetal lung at 12 pcw.  $MUC5AC$  (green),  $ASCL1$  (red),  $SCGB3A2$  (white). DAPI, nuclei. Scale bars, 50  $\mu$ m.

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





(legend on next page)



**Figure S4. Spatiotemporal location, distribution, and quantification of major epithelial cell types along the distal-to-proximal axis of the developing lungs, related to Figure 2**

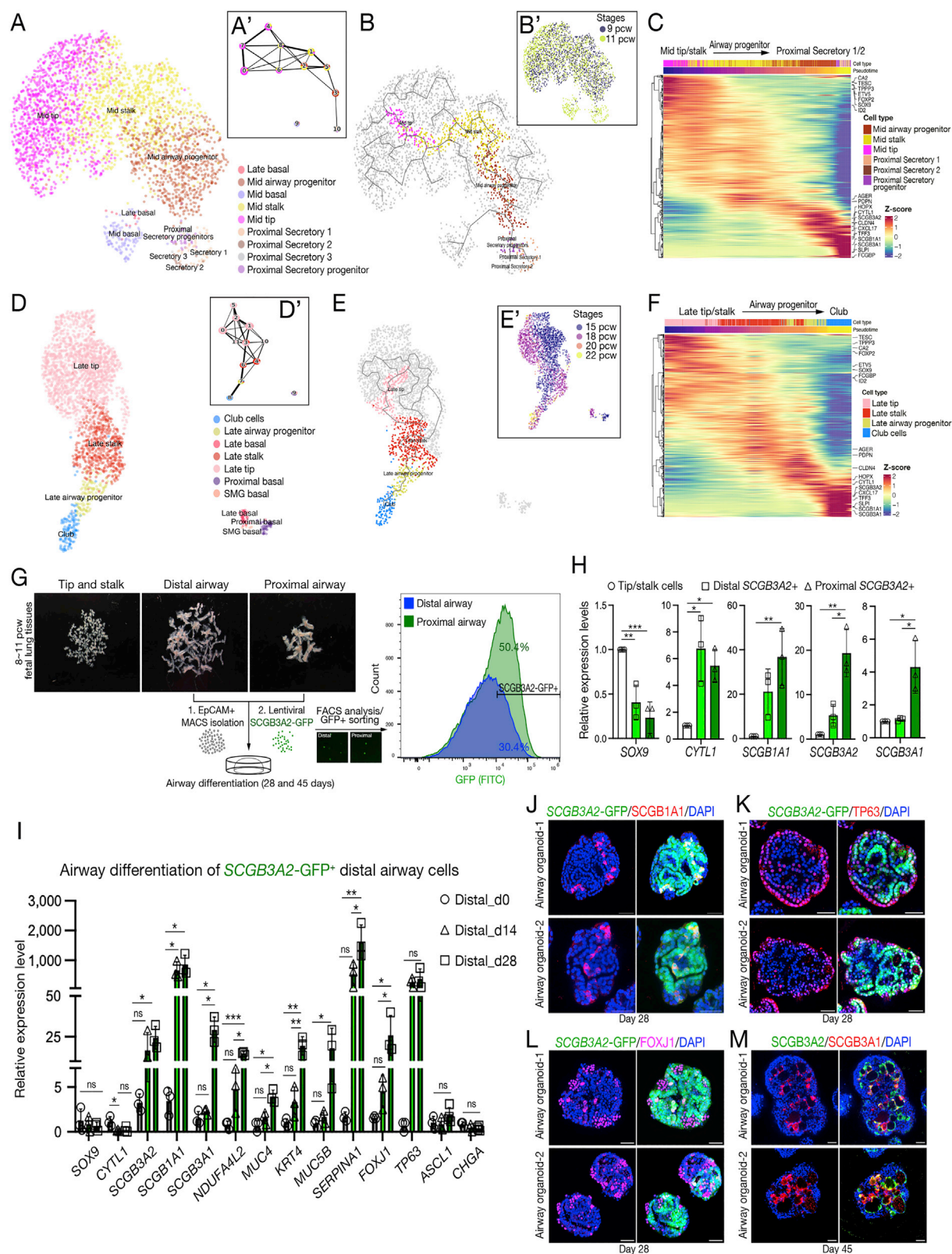
(A) *In situ* HCR analysis of fetal human lung tissues at mid (10–12 pcw) and late (15–21 pcw) stages, showing spatiotemporal location and distribution of major epithelial cell types along the distal to proximal axis of the developing lungs. The lung regions were divided for imaging into tip, stalk to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

(B) Quantification of cells expressing marker genes of airway lineages along the airway regions at mid (10–12 pcw, upper) and late (15–21 pcw, lower) stages. SCGB3A2, airway progenitors/all secretory lineage cells; *CYTL1*, airway progenitor cells; *NDUFA4L2*, club/proximal secretory cells; *FMO2*, club/proximal secretory cells; *FOXJ1*, ciliated cells; *TP63*, basal cells; *SCGB1A1*, club/proximal secretory cells; *SCGB3A1*, proximal secretory cell subtypes 1–3. Significance was evaluated by 1-way ANOVA with Tukey multiple comparison post-test; n = 3 biological replicates; ns: not significant, \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001.

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

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

(E) *In situ* HCR analysis of rare cell type markers of putative ionocytes (*FOXI1* yellow) and tuft cells (*POU2F3*, red). *E-cadherin*, green. DAPI, nuclei. Scale bar, 50 μm.



(legend on next page)

**Figure S5. Trajectory analysis of airway lineage differentiation via airway progenitor cells in the developing human lung, related to Figure 2**

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

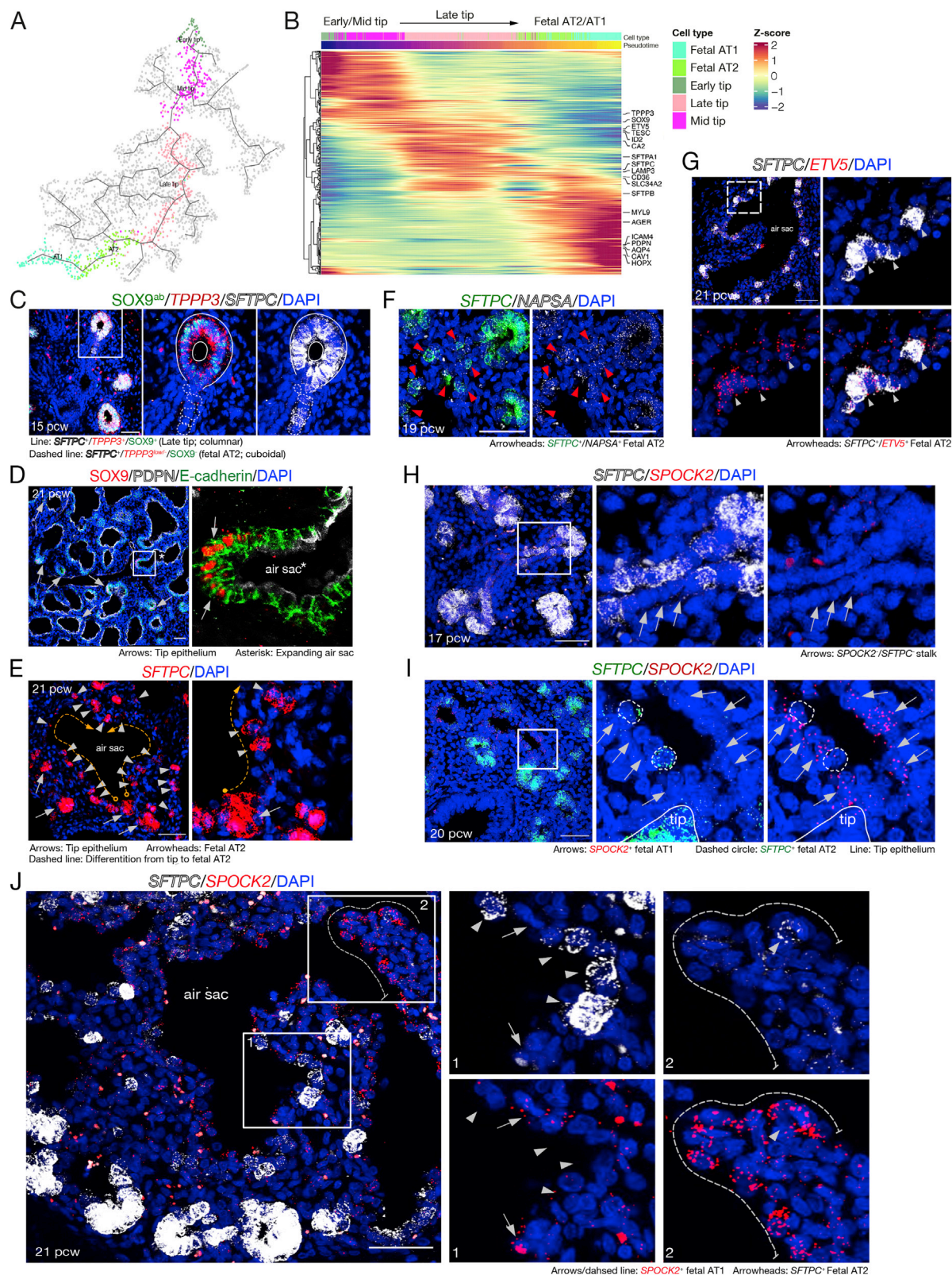
(B and C) Trajectory UMAPs, by cell type (B) and stages (B'), and the relevant gene expression heatmap (C) displaying the selected lineage trajectory from mid-tip to proximal secretory cell subtypes 1 and 2, analyzed by Monocle 3. (Note that the gray lines in UMAP indicate all of the predicted differentiation paths from a user-defined starting point.). (D, D') UMAP visualization (D) and PAGA analysis (D') of a lineage trajectory from late tip, late stalk, and late airway progenitor, to club cells. Basal cells, including late basal, proximal basal, and SMG basal cells were shown to be left out of the trajectory as they do not connect clearly to the other cell types in this analysis (D').

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

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

(H) Gene expression profile of the freshly purified *SCGB3A2*-GFP positive cells derived from distal and proximal airway tissues were investigated by qRT-PCR and compared with dissected tip cells. *SOX9*, distal tip progenitor marker. *CYTL1*, airway progenitor marker. *SCGB1A1* and *SCGB3A2*, airway/secretory cell lineage markers. *SCGB3A1*, proximal secretory cell marker. Data was normalized to *SCGB3A2*-GFP negative cells derived from distal tip/stalk tissues; mean  $\pm$  SD of 3 biological replicates. Significance was evaluated by 1-way ANOVA with Tukey multiple comparison post-test; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . (I) Gene expression analysis of the *in vitro* cultured *SCGB3A2*-GFP positive cells (airway progenitors) derived from distal airway tissues by qRT-PCR. Airway organoids were formed from the *SCGB3A2*-GFP positive cells and collected at Day 0, 14, and 28 days after culture for the analysis. Data were normalized to *SCGB3A2*-GFP negative cells derived from distal tip/stalk tissues; mean  $\pm$  SD of 4 biological replicates. Significance was evaluated by 1-way ANOVA with Tukey multiple comparison post-test; ns: not significant, \* $p < 0.05$ , \*\* $p < 0.01$ . (J-M) Immunofluorescence analysis of two biologically independent, *SCGB3A2*-GFP<sup>+</sup> cell-derived airway organoids cultured in the airway differentiation medium for 28 (J-L) and 45 (M) days. *SCGB1A1* (J, red), airway progenitor/secretory cell marker. TP63 (K, red), basal cell marker. FOXJ1 (L, magenta), ciliated cell marker. *SCGB3A1* (M, red), proximal secretory cell marker. DAPI, nuclei. Scale bar, 50  $\mu$ m.





(legend on next page)

**Figure S6. Late epithelial tip cells differentiate to AT2 and AT1 cells, related to Figure 3**

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

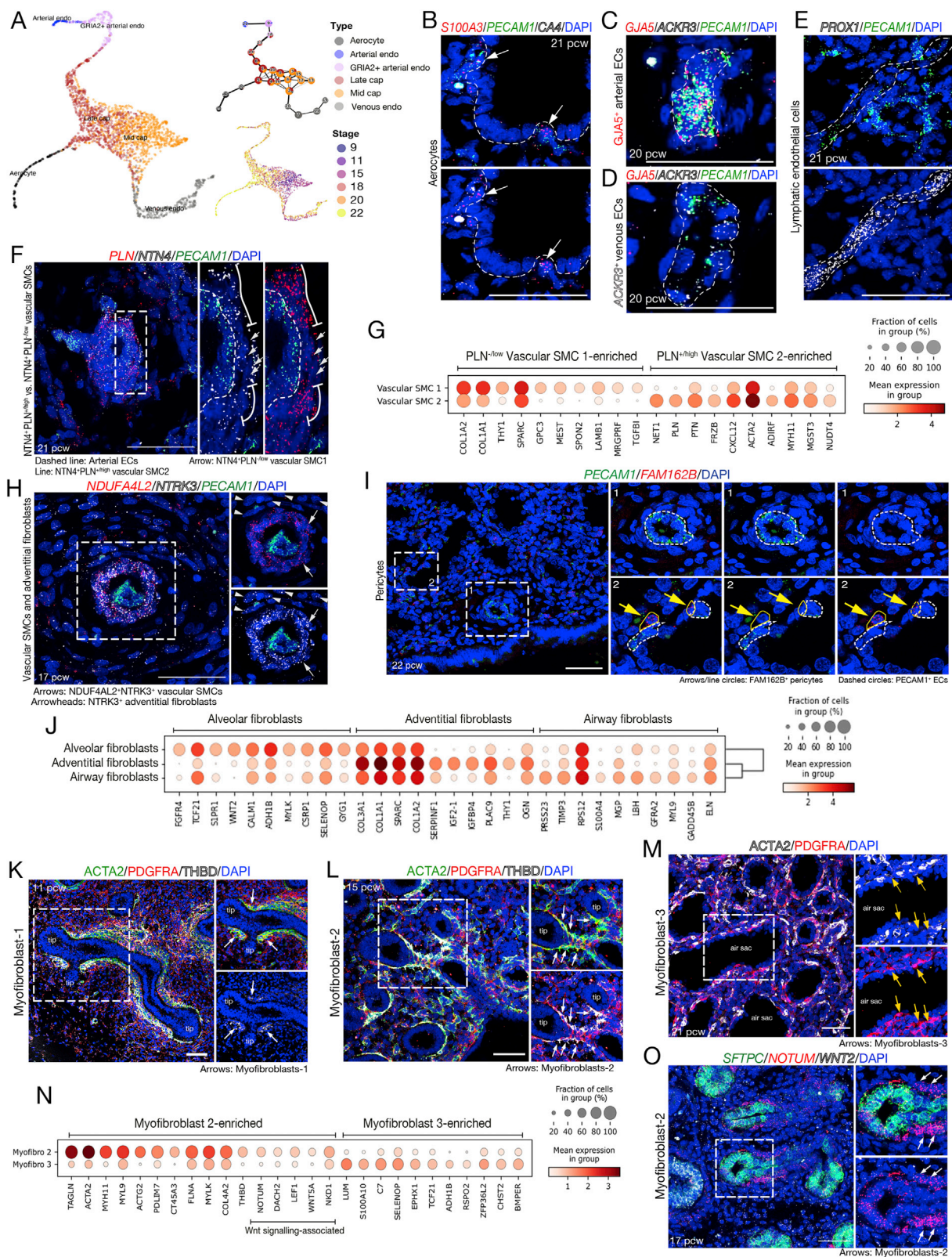
(C) *In situ* HCR (TPPP3 and *SFTPC*) and immunostaining (SOX9) analysis of 15 pcw fetal lung, describing  $SOX9^{+}TPPP3^{+}SFTPC^{+}$  tip epithelial progenitors (lines) and  $SOX9^{-}TPPP3^{-}SFTPC^{+}$  fetal AT2 cell population (dashed circles) lining the stalk.

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

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

(H–J) *In situ* HCR analysis of distal lung regions at 17 (H), 20 (I), and 21 (J) pcw, visualizing *SFTPC*<sup>−</sup>/*SPOCK2*<sup>−</sup> stalk cells and *SFTPC*<sup>−</sup> fetal AT1 cells (arrows). *SFTPC*<sup>−</sup>/*SPOCK2*<sup>−</sup> stalk cells at 17 pcw (H) began to express *SPOCK2* (red) at 20 pcw (I) and further developed to future AT1 cells (*SFTPC*<sup>−</sup>/*SPOCK2*<sup>+</sup>) at 21 pcw (J). Dashed circles (I) and arrowheads (J) indicate fetal AT2 cells. Dashed line (J) shows fetal AT1 cells lining the developing air sacs. DAPI, nuclei. Scale bars, 50 μm.



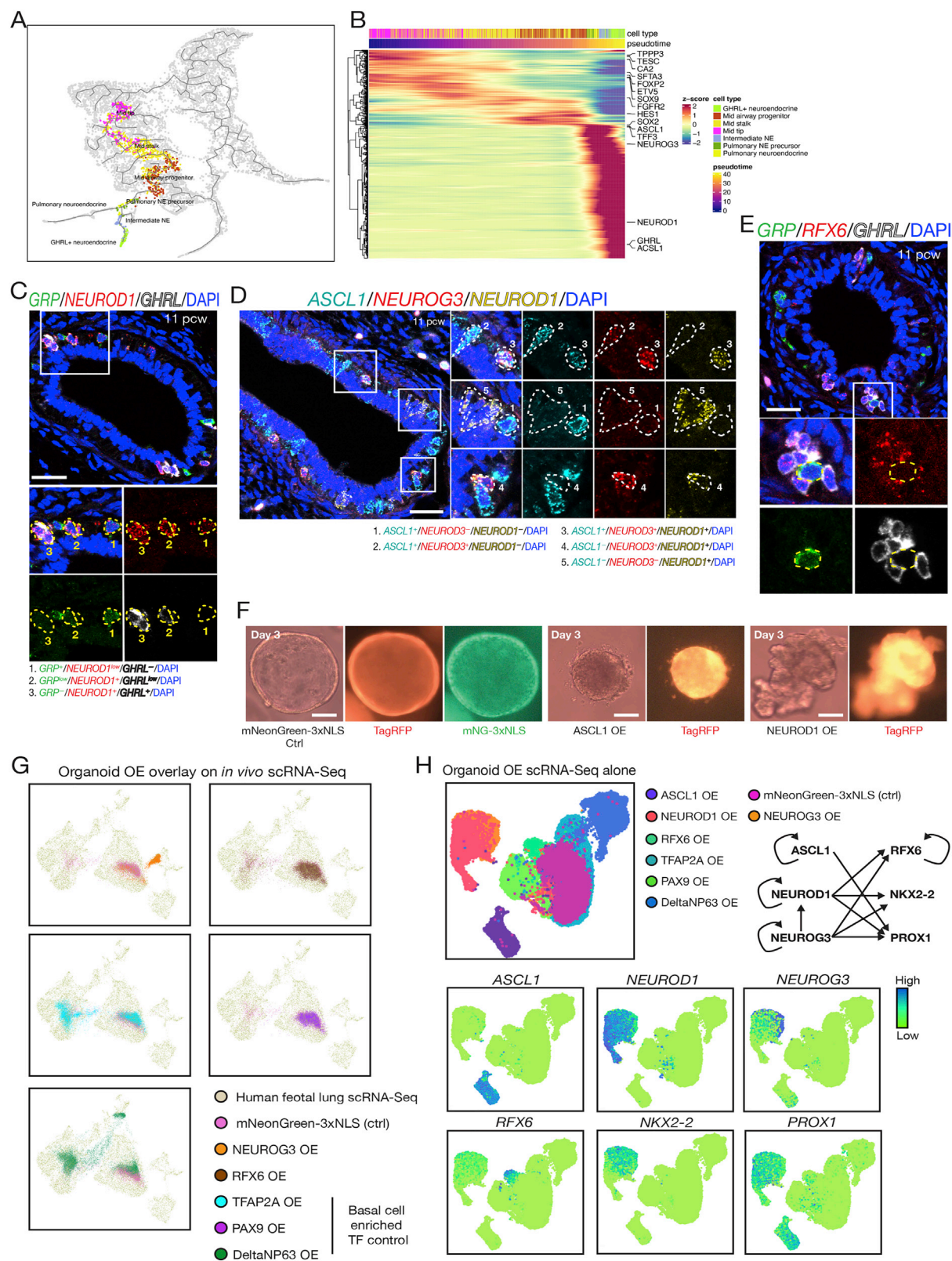


(legend on next page)



**Figure S7. Spatial analysis of endothelial and mesenchymal cell types in the developing human lungs by *in situ* HCR assay and immunostaining, related to Figure 4**

- (A) Trajectory UMAP and PAGA plot visualizing potential endothelial cell lineage hierarchy from Mid/Late capillary endothelial cells to arterial endothelial cells, aerocytes, or venous endothelial cells colored by cell types and stages.
- (B–E) *In situ* HCR analysis of distal lung regions at 20 (C, D), and 21 (B, E) pcw.
- (B) Aerocytes (*S100A3*<sup>+</sup> red/*CA4*<sup>+</sup> white), capillary endothelium (*CA4*<sup>+</sup> white), and all endothelial cells (*PECAM*<sup>+</sup>, green).
- (C) Arterial endothelial cells (*GJA5*<sup>+</sup> red), (D) venous endothelial cells (*ACKR3*<sup>+</sup> white), and all endothelial cells (*PECAM*<sup>+</sup>, green).
- (E) Lymphatic endothelial cells (*PROX1*<sup>+</sup> white) and all endothelial cells (*PECAM*<sup>+</sup>, green). DAPI, nuclei. Scale bars, 50  $\mu$ m.
- (F) Vascular SMC 1 and 2 are surrounding arterial endothelial cells (*PECAM1*<sup>+</sup>, dashed line), each marked by *NTN4*<sup>+</sup>/*PLN*<sup>low</sup> (vSMC 1, arrows) and *NTN4*<sup>+</sup>/*PLN*<sup>high</sup> (vSMC 2, lines).
- (G) Dot plot describing differential gene expression between vascular SMC 1 and 2.
- (H) Vascular SMCs and adventitial fibroblasts in 17 pcw fetal lung. *NDUF4AL2*<sup>+</sup> red/*NTRK3*<sup>+</sup> vSMCs (arrows) are surrounded by *NDUF4AL2*<sup>+</sup>/*NTRK3*<sup>+</sup> adventitial fibroblasts (arrowheads). *PECAM1* (green) indicates an endothelial cell tube.
- (I) *FAM162B*<sup>+</sup> pericytes (red) are surrounding *PECAM1*<sup>+</sup> endothelial cells (green) in the microvascular regions.
- (J) Dot plot describing differential marker gene expression level between alveolar, adventitial and airway fibroblasts.
- (K–M) Immunostaining of fetal lung tissues at 11 (K), 15 (L), and 21 (M) pcw, to visualize myofibroblast populations: Myofibroblast-1 (K) and –2 (L) surrounding the developing stalk epithelial tubes, and Myofibroblast-3 (M) surrounding the developing air sacs. *ACTA2*<sup>+</sup>/*PDGFRA*<sup>+</sup> Myofibroblast-1 (THBD<sup>weak</sup>; K) and –2 (THBD<sup>high</sup>, arrows; L). *PDGFRA*<sup>+</sup> Myofibroblast-3 at 21 pcw does not express *ACTA2* (arrows; M).
- (N) Dot plot describing differential gene expression level between myofibroblast-2 and -3. The myofibroblast-2 population showed enriched expression of Wnt signaling associated genes, e.g., *NOTUM*, *LEF1*, and *DACH2*.
- (O) *In situ* HCR assay of 17 pcw fetal lung tissues. Myofibroblast-2 expresses *NOTUM* (red), a Wnt antagonist, to block local Wnt signals from alveolar fibroblasts (white, *WNT2*) to the stalk epithelium. DAPI, nuclei. Scale bars, 50  $\mu$ m.



(legend on next page)

**Figure S8. Transcription factor regulatory network controlling NE subtypes, related to Figure 7**

- (A) Selected trajectory from mid-tip cells to *GHRL*<sup>+</sup> NE cells via Intermediate NEs, a transition cell population.
- (B) Heatmap of genes differentially expressed along the trajectory.
- (C) Representative HCR images showing the transition between two types of NE cells. *GRP* (green), *NEUROD1* (red), *GHRL* (white). #1 labeled *GRP*<sup>+</sup>*NEUROD1*<sup>low</sup>*GHRL*<sup>-</sup> cells, which have just started the transition from *GRP*<sup>+</sup> pulmonary NE/precursor cells. #2 labeled *GRP*<sup>low</sup>*NEUROD1*<sup>+</sup>*GHRL*<sup>low</sup> cells, in transition to *GHRL*<sup>+</sup> NE cells. #3 labeled *GRP*<sup>+</sup>*NEUROD1*<sup>+</sup>*GHRL*<sup>+</sup>, *GHRL*<sup>+</sup> NE cells. Right: Mean  $\pm$  SEM of *NEUROD1*<sup>+</sup> cell types. 11 pcw: N = 2 fetal lungs, n = 129 *NEUROD1*<sup>+</sup> cells; 12 pcw N = 3 fetal lungs, n = 132 *NEUROD1*<sup>+</sup> cells. Scale bars, 25  $\mu$ m in all panels.
- (D) Representative HCR images showing *NEUROG3* co-expression with *ASCL1* and *NEUROD1*. Dashed white lines label representative cells showing different combinations of the three transcription factors, further indicated by #1-#5 labeling. *ASCL1* (cyan), *NEUROG3* (red), *NEUROD1* (yellow).
- (E) Representative HCR images showing *RFX6* expression in *GHRL*<sup>+</sup> NE cells. Dash yellow line labeled *GRP*<sup>+</sup>*RFX6*<sup>-</sup> pulmonary NE cells. Scale bars, 25  $\mu$ m in all panels.
- (F) Representative epifluorescent microscopic images showing organoid morphology after 3 days of mNeonGreen-3xNLS (control), *ASCL1*, or *NEUROD1* overexpression.
- (G) scRNA-seq results of organoid transcription factor overexpression overlay on human fetal lung scRNA-seq as a reference.
- (H) scRNA-seq results of transcription factor overexpression; organoid data only in the UMAP. Selected transcription factor expression was shown in the middle panel. A regulatory network of the selected transcription factors were drawn based on the organoid OE data at the bottom of the panel. (Note that the arrows do not necessarily denote direct interactions).