Title
Human distal airways contain a multipotent secretory cell that can regenerate alveoli

Journal
Nature, 604(7904)

ISSN
0028-0836

Authors
Basil, Maria C
Cardenas-Diaz, Fabian L
Kathiriya, Jaymin J
et al.

Publication Date
2022-04-07

DOI
10.1038/s41586-022-04552-0

Peer reviewed

Published in final edited form as:

Nature. 2022 April ; 604(7904): 120–126. doi:10.1038/s41586-022-04552-0.

Human distal airways contain a multipotent secretory cell that can regenerate alveoli

Maria C. Basil[1,2,12], **Fabian L. Cardenas-Diaz**[1,2,12], **Jaymin J. Kathiriya**[3], **Michael P. Morley**[1,2,5], **Justine Carl**[1,2], **Alexis N. Brumwell**[3], **Jeremy Katzen**[1,2], **Katherine J. Slovik**[1,2,5], **Apoorva Babu**[1,2,5], **Su Zhou**[1,2], **Madison M. Kremp**[1,2], **Katherine B. McCauley**[6], **Shanru Li**[1,2], **Joseph D. Planer**[1,2], **Shah S. Hussain**[7], **Xiaoming Liu**[8], **Rebecca Windmueller**[2,11], **Yun Ying**[1,2], **Kathleen M. Stewart**[1,2], **Michelle Oyster**[1], **Jason D. Christie**[1,2], **Joshua M. Diamond**[1], **John F. Engelhardt**[8], **Edward Cantu**[2,9], **Steven M. Rowe**[7], **Darrell N. Kotton**[6,10], **Harold A. Chapman**[3,4], **Edward E. Morrisey**[1,2,11,✉]

[1]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[2]Penn-CHOP Lung Biology Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[3]Department of Medicine, University of California, San Francisco, San Francisco, CA, USA.

[4]Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA, USA.

[5]Penn Cardiovascular Institute, University of Pennsylvania, Philadelphia, PA, USA.

[6]Center for Regenerative Medicine, Boston University and Boston Medical Center, Boston, MA, USA.

[7]Department of Medicine and the Gregory Fleming James Cystic Fibrosis Research Center, University of Alabama at Birmingham, Birmingham, AL, USA.

✉**Correspondence and requests for materials** should be addressed to Edward E. Morrisey. emorrise@pennmedicine.upenn.edu.

Author contributions M.C.B., F.L.C.-D., J.J.K., J.C., A.N.B., J.K., K.J.S., S.Z., M.M.K., S.L., J.D.P., S.S.H., Y.Y. and K.M.S. performed human primary cell experiments. K.J.S., K.B.M. and M.C.B. performed the human stem cell experiments. M.C.B., F.L.C.-D., J.J.K., M.P.M., A.B., R.W., D.N.K. and E.E.M. analysed the data. M.C.B., J.K., M.O., J.D.C., J.M.D. and E.C. provided access to human samples. X.L., S.S.H., J.F.E. and S.M.R. provided access to ferret samples. M.C.B. and E.E.M. wrote the first draft of the manuscript. All of the authors contributed to writing the final manuscript.

Code availability

Analysis associated with the current submission used published R packages and a custom R package, which are available at GitHub (https://github.com/Morriseylab/scExtras). The code for the custom graphical scRNA-seq interface is available at GitHub (https://github.com/Morriseylab/scViewer-Lite). For our analysis using published R packages, the following versions were used: R v.4.0.0, CellRanger v.3.1.0, Slingshot v.1.6.1, tradeSeq v.1.2.01, clusterProfiler v.3.16.1, ComplexHeatmap v.2.4.3, MAST v.1.14.0, Seurat v.3.0.1, Sctransform v.0.3.0 and clustree v.0.4.3.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04552-0.

Competing interests The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04552-0.

[8]Department of Anatomy and Cell Biology, Carver College of Medicine, University of Iowa, Iowa City, IA, USA.

[9]Division of Cardiovascular Surgery, Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[10]The Pulmonary Center and Department of Medicine, Boston University and Boston Medical Center, Boston, MA, USA.

[11]Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

[12]These authors contributed equally: Maria C. Basil, Fabian L. Cardenas-Diaz.

Abstract

The human lung differs substantially from its mouse counterpart, resulting in a distinct distal airway architecture affected by disease pathology in chronic obstructive pulmonary disease. In humans, the distal branches of the airway interweave with the alveolar gas-exchange niche, forming an anatomical structure known as the respiratory bronchioles. Owing to the lack of a counterpart in mouse, the cellular and molecular mechanisms that govern respiratory bronchioles in the human lung remain uncharacterized. Here we show that human respiratory bronchioles contain a unique secretory cell population that is distinct from cells in larger proximal airways. Organoid modelling reveals that these respiratory airway secretory (RAS) cells act as unidirectional progenitors for alveolar type 2 cells, which are essential for maintaining and regenerating the alveolar niche. RAS cell lineage differentiation into alveolar type 2 cells is regulated by Notch and Wnt signalling. In chronic obstructive pulmonary disease, RAS cells are altered transcriptionally, corresponding to abnormal alveolar type 2 cell states, which are associated with smoking exposure in both humans and ferrets. These data identify a distinct progenitor in a region of the human lung that is not found in mouse that has a critical role in maintaining the gas-exchange compartment and is altered in chronic lung disease.

Despite the considerable global clinical impact of lung diseases, including chronic obstructive pulmonary disease (COPD), treatments for many chronic and acute lung diseases have seen minimal advances over the past decade[1]. In mice, the conducting airways propagate air distally while removing particulate matter and participating in vital immune functions[2–4]. The conducting airways of mice transition directly into the alveolar space at the bronchoalveolar duct junction. By contrast, in humans and other large mammals, the proximal airways gradually transition to the most distal respiratory airway zone (hereafter, respiratory airways), which comprises the respiratory bronchioles (also known as the transitional bronchioles) and alveolar ducts[5,6]. Little is known about the cellular constituents of the human eespiratory bronchiole[5,7–9]. Emerging evidence suggests that this unique niche is disrupted in several human lung diseases[10] including COPD[11,12], viral infections[13–15] and aging[16].

Here we reveal that human respiratory airways are lined with a unique secretory cell population that is distinct from the canonical secretory cells that are primarily found in proximal upper airways, which we term RAS cells. RAS cells line regions of the respiratory

bronchioles and reside transcriptionally between canonical secretory cells and alveolar type 2 (AT2) cells. We show that RAS cells can serve as a previously undescribed distal lung progenitor population by rapidly and unidirectionally differentiating into AT2 cells, a process controlled by Notch and Wnt signalling. In patients with COPD, the transcriptome of RAS cells is altered, correlating with the presence of altered AT2 cell states in this disease as well as in models of smoking induced injury. Together, these studies identify and characterize a progenitor cell lineage in the respiratory airways of the human lung and its alteration in COPD.

Distinct secretory cells in the human airway

Human respiratory airways contain respiratory bronchioles and the alveolar ducts, which ultimately connect with the alveoli (Fig. 1a, b). This contrasts with mice, in which the conducting airways end bluntly at the bronchoalveolar duct junction (Fig. 1a, b). Single-cell RNA-sequencing (scRNA-seq) analysis was performed on micro-dissected lung tissue from five healthy non-smoker donors to assess distal lung cell diversity (Supplementary Table 1). All expected cellular compartments, including epithelial, mesenchymal and endothelial cells, were identified (Fig. 1c and Extended Data Fig. 1a, b). Epithelial cells were extracted and reclustered for a more in depth analysis[2,3] (Extended Data Fig. 1e).

In this analysis, we observed two secretory cell populations, both expressing the canonical marker *SCGB1A1*, but with a distinct population expressing high levels of *SCGB3A2* (Fig. 1d, e, Extended Data Fig. 1b, h, i). The *SCGB3A2*$^{+}$*SCGB1A1*$^{+}$ population contained a divergent gene expression profile compared with the *SCGB1A1*$^{+}$ population (Fig. 1e, Extended Data Figs. 1f, g and 2a), including several ribosomal genes that are consistent with a synthetically active cell. The analysis revealed that this population comprises a distinct type of secretory cells in the human distal lung that we now call RAS cells.

Proximal airways lack *SCGB3A2*$^{+}$ cells

To determine whether RAS cells were found in the proximal airways of the human lung, we dissected the first segmental airway branch for five subsequent generations and performed scRNA-seq analysis (Fig. 1f, Extended Data Fig. 1c, d). Examination of the proximal airway epithelial populations revealed many of the expected cell types[17,18] (Fig. 1f, Extended Data Figs. 1j, k and 3a, b). The secretory cells in the proximal airways were identified by expression of *SCGB1A1* (Fig. 1g). However, only rare cells in the proximal airways expressed *SCGB3A2* (Fig. 1g, Extended Data Fig. 1j, k).

SCGB3A2^{+} cells in respiratory bronchioles

Immunohistochemistry (IHC) and RNAscope analysis showed that SCGB3A2^{+} cells were found in patches of airway epithelium interspersed with morphologically identified alveoli within human respiratory bronchioles (Fig. 1h, i, Extended Data Figs. 2d and 3c–e). SCGB3A2^{+} cells lined the distal most regions of these airways, whereas SCGB1A1^{+} cells were observed in the terminal bronchioles (Fig. 1h, Extended Data Fig. 3c–e). The

SCGB3A2$^+$ cells were cuboidal in shape and were interspersed with SFTPC$^+$ or LAMP3$^+$ AT2 cells (Fig. 1j, k), and RAGE$^+$ AT1 cells (Extended Data Fig. 2b).

To determine whether SCGB3A2$^+$ RAS cells were found in other species known to contain respiratory bronchioles, we examined expression in ferrets[19–21]. Similar to humans, the respiratory bronchioles of ferrets contained patches of simple low-cuboidal epithelium marked by SCGB3A2 expression interspersed with alveolar outpouchings (Fig. 1l, m). By contrast, mouse conducting airways terminate at the BAJD[22] (Extended Data Fig. 2c).

We next used scRNA-seq analysis of epithelial cells from the same patient to determine that the *SCGB1A1*$^+$ single-positive secretory cells from both the proximal and distal compartments co-clustered, indicating that these populations from proximal and distal human lungs represented similar canonical secretory cells (Extended Data Fig. 4a–c, e). By contrast, *SCGB3A2*$^+$ cells clustered predominately with cells derived from distal airways (Extended Data Fig. 4b, c, f). IHC analysis was performed and SCGB3A2$^+$ cells were not found in any of the proximal large airways (Extended Data Fig. 3f, g), but low cuboidal SCGB3A2$^+$ cells were consistently found in the distal airways, including the respiratory bronchioles (Extended Data Fig. 3f, g). Robust SCGB1A1 expression was observed in the proximal large conducting airway cells (Extended Data Fig. 3f, g).

A RAS cell–AT2 cell progenitor relationship

Comparative gene expression analysis between SCGB1A1$^+$ secretory cells, RAS cells and AT2 cells was performed using model-based analysis of single-cell transcriptomics (MAST)[23]. We chose to focus on the RAS cell–AT2 cell relationship owing to the close linkage in uniform manifold projection (UMAP) space and the described the progenitor function of the AT2 cell lineage[24–26]. This analysis revealed that RAS cells shared expression patterns with both SCGB1A1$^+$ secretory cells and AT2 cells (Extended Data Fig. 4g). Several genes were identified that were either restricted to one population, shared between secretory cells and RAS cells, or shared between RAS cells and AT2 cells (Extended Data Fig. 4d, g).

Trajectory inference analysis using Slingshot[27] showed that the proximal lung epithelium has a series of pseudotemporal relationships (trajectories) that are consistent with previous lineage-tracing experiments in mouse models[17,28] (Extended Data Fig. 5a, b). An analysis of the distal lung epithelium demonstrated a pseudotemporal relationship between RAS cell and AT2 cells (Extended Data Fig. 5d, e), among other putative relationships observed (Extended Data Fig. 5c, f). The gene expression changes along the RAS cell–AT2 cell trajectory highlight several genes that are known to be involved in AT2 cell differentiation and maturation including *WIF1* and *SFTPC*, suggesting that RAS cells may serve as a putative progenitor for the AT2 cell lineage (Extended Data Fig. 5e, g).

iRAS cells can differentiate into iAT2 cells

We next used a previously described SCGB3A2–mCherry human embryonic stem (hES) cell reporter line that can generate SCGB3A2$^+$ human airway epithelial cells to assess the relationship between RAS cells and AT2 cells[29]. We have named these SCGB3A2$^+$

cells induced RAS (iRAS) cells. iRAS cells were grown in organoid cultures, sorted using fluorescence-activated cell sorting (FACS) on the basis of their mCherry expression, and then propagated either under continued airway conditions (airway medium)[29,30] or by transitioning into medium that was previously shown to promote iAT2 cell differentiation (alveolar medium)[31] (Extended Data Fig. 6a). We observed that, when iRAS cells are maintained in airway medium, they maintain *SCGB3A2* expression[29] (Extended Data Fig. 6b, c). By contrast, when purified iRAS cells were transitioned to alveolar medium, *SCGB3A2* expression was rapidly downregulated and expression of *SFTPC* was initiated, and this was associated with appropriate processing of the SFTPC protein and localization into lamellar bodies (Extended Data Fig. 6b–g, Supplementary Fig. 1).

Primary human distal epithelium was compared to iRAS cells and iAT2 cells using reference-based integration to determine similarities between the stem-cell-derived and in vivo cell types (Extended Data Fig. 7a, b). This shows that iRAS cells co-clustered predominately with primary RAS cells, and iAT2 cells co-clustered with primary AT2 cells (Extended Data Fig. 7a–c). Similar assessment was performed using the previously identified SCGB3A2^{+} fetal epithelial progenitor cells[32], indicating again that iRAS cells co-cluster preferentially with primary adult RAS cells (Extended Data Fig. 7f–h). Expression of hallmark genes verifies a strong concordance between primary and stem-cell-derived RAS cells and AT2 cells (Extended Data Fig. 6h, i).

RAS to AT2 transition is unidirectional and rapid

We generated a bilineage fluorescence reporter system that reports SFTPC expression through an enhanced green fluorescent protein (eGFP) reporter in the same cell line containing the SCGB3A2–mCherry reporter (Extended Data Fig. 8a). Using this dual reporter line, we show that when iRAS cells are transitioned from airway to alveolar medium, SFTPC–eGFP expression is noted as early as 72 h after transition, whereas SCGB3A2 expression is rapidly downregulated (Fig. 2a, Extended Data Fig. 8b–g).

To assess whether the rapid transition of iRAS cells into iAT2 cells was a true differentiation event versus temporary cellular plasticity, iAT2 cells derived from this stepwise culture system were returned to airway culture. When iAT2 cells derived from iRAS cells were returned to airway medium, they did not activate SCGB3A2 expression nor form well-defined organoids, indicating a unidirectional differentiation process (Fig. 2b).

The RAS cell–AT2 cell transcriptional relationship

To characterize how iRAS cells differentiate into iAT2 cells, we performed a temporal series of scRNA-seq analyses as iRAS cells differentiated into iAT2 cells at multiple time points (Extended Data Fig. 9a–c). *SCGB3A2* was uniformly expressed in day 0 iRAS cells, which lacked expression of *SFTPC*; however, after only 3 days of differentiation into iAT2 cells, the cells were uniformly *SFTPC*$^{+}$ and lacked *SCGB3A2* expression (Extended Data Fig. 9b, g). Cells from days 3–7 also exhibited higher rates of cell proliferation, as identified by cell cycle scoring[33] (Extended Data Fig. 9b).

Trajectory inference analysis of the concatenated time course was performed and revealed several projected transcriptional relationships, including one which terminated in probably the most mature iAT2 cells based on lower proliferation levels (Extended Data Fig. 9d). Other trajectories varied in expression patterns due to a heterogeneous proliferating culture system[29,31,34,35] (Extended Data Fig. 9d, o). The expression of genes that defined the primary RAS cell–AT2 cell trajectory (Extended Data Fig. 5e, g) was compared with the iRAS cell to iAT2 cell transition (Extended Data Fig. 9e, f), showing that key lineage identifying genes followed similar expression patterns in the iRAS cell to iAT2 cell transition as they do in the primary RAS cell–AT2 cell pseudotemporal trajectory, indicating that the iRAS cell to iAT2 cell transition reflects the in vivo process (Extended Data Fig. 9e, f).

Transcriptional regulators enriched in both primary RAS cells and iRAS cells were compared with primary AT2 and iAT2 cells, respectively (Supplementary Table 2). These gene lists were compared to identify shared transcription factors that are enriched in RAS cells and iRAS cells, and identified targets of Notch signalling (Extended Data Fig. 7d, e). Notch signalling is known to be critical for secretory cell development and homeostasis in mice[36–40]. We next cultured iRAS cells at homeostasis and when undergoing differentiation into iAT2 cells, with the Notch signalling inhibitor DAPT[41] (Fig. 2c, d). Inhibition of Notch in iRAS cell airway cultures led to reduced expression of cardinal RAS-cell-specific genes, including *SCGB3A2* and *SCGB1A1* (Extended Data Fig. 9h, j). By contrast, inhibiting Notch during the iRAS cell to iAT2 cell transition resulted in enhanced *SFTPC*, *ABCA3* and *PBC* expression, as assessed by increased mean fluorescent reporter expression and by quantitative PCR (qPCR) (Extended Data Fig. 9i, k).

Wnt signalling is known to be important for AT2 cell behaviour and regeneration[25,26,42,43], and the Wnt agonist CHIR99021 is a critical component of the iAT2 cell differentiation medium[30,34]. We therefore cultured iRAS cells in airway medium containing CHIR99021 to assess whether the activation of Wnt signalling is sufficient to activate the human iAT2 gene program (Fig. 2e). This resulted in the downregulation of the SCGB3A2–mCherry reporter and upregulated SFTPC–eGFP expression (Fig. 2e and Extended Data Fig. 9l–n). Together, these data show that Notch and Wnt signalling are important in maintaining iRAS cell fate and that the inhibition of Notch and activation of Wnt enhances iRAS cell to iAT2 cell differentiation.

Primary RAS cells can give rise to AT2 cells

A new scRNA-seq dataset using only the distal lung epithelium was used to identify a cell-surface marker to isolate primary RAS cells (Extended Data Fig. 10a, b). This analysis identified CEACAM6 (CD66c) as a possible marker for RAS cells (Extended Data Fig. 10g). IHC analysis revealed a high concordance of SCGB3A2 and CEACAM6 expression in the respiratory bronchioles in vivo (Fig. 3a, b, Extended Data Fig. 10c, d). CEACAM6 has been shown to be expressed on mouse AT2 cells and basal cells[44–46], so the AT2-cell-specific antibody HT2-280 and the basal-cell-specific antibody NGFR were used to isolate a $CEACAM6^{+}HT2\text{-}280^{-}NGFR^{-}$ population using FACS (Extended Data Fig. 10h,

i). Importantly, cytospin analysis showed that <5% of the CEACAM6^{+}HT2-280^{-}NGFR^{-} purified cells expressed the AT2 marker SFTPC (Extended Data Fig. 10e, f).

To determine whether isolated CEACAM6^{+}HT2-280^{-}NGFR^{-} cells could generate human AT2 cells, we generated organoids from these cells using the iRAS cell–iAT2 cell differentiation medium (Fig. 3c). These organoids grew extensively and immunostaining showed that more than 60% of the resultant organoids expressed SFTPC protein after 21 days in culture (Fig. 3d–f).

Altered AT2 cell states in human COPD

To determine whether RAS cells and the RAS cell–AT2 cell cellular axis is disrupted in COPD, we performed scRNA-seq analysis of peripheral lung tissue from a cohort of patients with COPD undergoing lung transplantation (Supplementary Table 1). We identified RAS cells as well as multiple AT2 cell subpopulations (Fig. 4a, b, Extended Data Fig. 11a–e). Two new AT2 cell clusters were found in COPD lungs that did not appear in the normal donor dataset. One of these unique populations expressed an elevated level of the canonical RAS cell marker *SCGB3A2* (Fig. 4b, c). In healthy donors, we observed only rare *SCGB3A2*$^{+}$ AT2 cells that did not form a unique cluster (Fig. 1d, e and Extended Data Fig. 1h). All of the patients with COPD contributed to the *SCGB3A2*$^{+}$ AT2 cell cluster (Extended Data Fig. 11a, b). IHC analysis showed that, although only rare SCGB3A2^{+}LAMP3^{+} double-positive cells were found in the respiratory bronchioles of normal donor lungs, there was a statistically significant increase in these double-positive cells in the lungs of patients with COPD (Fig. 4d, g).

SCGB3A2^{+} AT2 cells and smoke injury

To determine whether the RAS cell to AT2 cell differentiation process was altered in a model of smoking-induced chronic injury, we examined a cohort of patients who were either active smokers or had discontinued tobacco use within the last year and grouped them on the basis of reported exposure to cigarette smoke as measured in pack-years (Supplementary Table 1). We performed IHC analysis of SCGB3A2 and LAMP3 in lung sections from these patients and determined that the percentage of SCGB3A2^{+}LAMP3^{+} cells increased with increasing pack-year history (Fig. 4e, g). We also performed IHC analysis of distal lung parenchyma from ferrets that had been exposed to cigarette smoke twice daily for 6 months[47–49]. These data demonstrate an increased number of SCGB3A2^{+}LAMP3^{+} cells in ferrets exposed to cigarette smoke compared with the untreated control ferrets (Fig. 4f, h). Together, this suggests that cigarette smoke injury results in the emergence of an SCGB3A2^{+} AT2 cell population in ferrets and humans.

Abnormal RAS cell states in human COPD

The normal and COPD peripheral datasets were concatenated, and gene expression differences in normal versus COPD RAS cells revealed changes in multiple categories, including metabolic processes and immune function (Extended Data Fig. 12a–d, l, m). Similar gene expression differences were observed in AT2 cells of patients with COPD

(Extended Data Fig. 12e–h). Endothelial and mesenchymal cells revealed different gene expression patterns between the lungs of healthy individuals and patients with COPD (Extended Data Fig. 11f, g).

Using trajectory analysis, we observed that RAS cells shared a relationship with both the most mature and normal AT2 cell population (trajectory 1; Extended Data Fig. 12i, j), as well as with the *SCGB3A2*$^{+}$ AT2 cell population (trajectory 2; Extended Data Fig. 12i, k). We compared the changes along each of these trajectories (Extended Data Fig. 12j, k), and identified several genes that are expressed preferentially in the generation of *SCGB3A2*$^{+}$ AT2 cells, including genes associated with the development of lung disease, such as *SPP1* (ref.[50]) and *CLEAC7A*[51], and *HSPA6* (ref.[52]). We also observed differences in expression of *BCL2A1* and *ATF3*, which have roles in cell survival and proliferation. Taken together, these analyses highlight the dysfunction of RAS cell to AT2 cell differentiation in COPD, which is consistent with the known anatomical abnormalities that occur in the distal respiratory bronchioles and the alveolar niche in this disease.

Discussion

Here we identified a distinct secretory cell lineage in the distal human lung called RAS cells, which are found in human and ferret respiratory bronchioles. RAS cells lie in a transcriptionally intermediate state between canonical secretory cells and AT2 cells and we show that iRAS cells and RAS cells can rapidly and unidirectionally differentiate into human AT2 cells through Notch and Wnt signalling. Importantly, we demonstrate that the RAS cell–AT2 cell trajectory is altered in COPD, consistent with the presence of unique AT2 cell states, and chronic smoke-induced injury is associated with aberrant RAS cell to AT2 cell differentiation. Together, these data characterize a poorly understood architectural niche in the human lungs that is populated by a distinct cell lineage that has the ability to generate AT2 cells and participate in alveolar regeneration.

Previous studies have described a rare population of epithelial cells in the mouse lung called the bronchioalveolar stem cells, marked by the co-expression of *Scgb1a1* and *Sftpc*, and found at a distinct anatomical niche[53]. In contrast to mouse bronchioalveolar stem cells, human RAS cells are an abundant cell type in respiratory bronchioles in the human lungs. Although our data show that there are rare SCGB3A2^{+}LAMP3^{+} cells in the normal lungs, given their number and distribution in comparison to RAS cells, they may represent cells transitioning from RAS cells to AT2 cells rather than a true sublineage. A recent study also showed that human KDR^{+}HT2-280^{-} airway epithelium can generate AT2 cells, further supporting the idea that a lineage within the human airway epithelium can generate AT2 cells[54]. Although it is important to note that such stem cell models may not reflect all aspects of their in vivo counterparts, our data showing that iRAS cells can generate iAT2 cells support and concur with our primary organoid and scRNA-seq data.

The respiratory bronchioles are a frequent site of injury in the human lungs, from common diseases such as COPD[11,12], to rare disease such as bronchiolitis obliterans[55–57]. The absence of this niche in mice may explain in part why some human respiratory diseases have proven to be difficult to model in mice[58]. Our data demonstrate that the differentiation

trajectory of RAS cells to AT2 cells is aberrant in COPD and that SCGB3A2$^+$ AT2 cells are increased in both humans and ferrets exposed to cigarette smoke. The accumulation of SCGB3A2$^+$ AT2 cells could represent either a more active transitioning or a stall during the transitioning of these cells, leading to increased numbers of these double-positive cells in COPD. Given the critical role that the respiratory bronchioles have in both normal human lung homeostasis and chronic disease, future studies into this region should reveal additional insights into this distinct niche of the human lungs.

Methods

Participants and ethical compliance

The normal samples used for this study were from de-identified non-used lungs donated for organ transplantation through an established protocol (PROPEL, approved by University of Pennsylvania Institutional Review Board) with informed consent in accordance with institutional and NIH procedures, and consent was provided by next of kin or healthcare proxy. COPD tissue was obtained from participants enrolled at the University of Pennsylvania as part of the Prospective Registry of Outcomes in Patients Electing Lung Transplantation (PROPEL) (Penn cohort)[59]. All of the selected patients had severe COPD based on FEV1 measurements, significantly impaired quality of life based on current guidelines for transplantation for COPD, none had evidence of other parenchymal disease and none were on systemic steroids. The institutional review board of the University of Pennsylvania approved this study, and written informed consent was obtained from all participants before inclusion in the study. All patient information was removed before use. This use does not meet the current NIH definition of human participant research, but all relevant guidelines and regulations and all institutional procedures required for human participant research were followed throughout the reported experiments.

Tissue preparation and scRNA-seq

Samples of normal de-identified human lungs from donors who were not matched for lung transplant were obtained as described previously[26] with the following adaptations. For peripheral tissue experiments, a 3 cm × 2 cm piece of distal lung tissue was obtained, pleura and visible airways/blood vessels were dissected away, mechanically minced into ~2 mm pieces and processed into a single-cell suspension as described previously[26]. For proximal airways, the airways were dissected away from the surrounding parenchyma starting at the first generation after the carina and dissected for approximately five generations. Cleaned airways were removed from surrounding lung tissue, cut open along one side to create flattened tubes and incubated in the same digestion buffer as described above for 60 min at 37 °C with frequent manual agitation. Then, 30 min into the incubation, a scalpel was used to gently scrape the surface of the airway epithelium to dislodge cells, and this was repeated at the end of the incubation. The airway scaffold was removed, washed with PBS and discarded. After a single-cell suspension was obtained from the proximal or peripheral tissue, CD45$^+$ immune cells were depleted using MACS LS columns with CD45-microbeads (Miltenyi, 130-045-801) with 2×10^6 cells per column to enhance purity and viability. After CD45 depletion, sorted cells were loaded onto a GemCode instrument (10x Genomics) to

generate single-cell barcoded droplets (GEMs) according to the manufacture's protocol. The resulting libraries were sequenced on an Illumina HiSeq2500 or NovaSeq instrument.

Analysis of scRNA-seq data

Reads were aligned and unique molecular identifier (UMI) counts were obtained using the Cell Ranger pipeline (10x Genomics). For further processing, integration and downstream analysis, Seurat v.3 was used[33]. Cells with less than 200 genes, greater than 2 median absolute deviation above the median and with potential stress signals of greater than 10% mitochondrial reads were removed. The cell cycle phase prediction score was calculated using Seurat function CellCycleScoring, and data were normalized and scaled using the SCTransform function and regressing out the effects of cell cycle score (G2M.Score,S.Score), percentage fraction of mitochondria, number of features per cell and number of UMI per cell. For integration of individual samples, normalized values from SCTransform and the top 3,000 variable genes as anchors were used for canonical correlation analysis (CCA) using Seurat v.3. For the proximal-peripheral integration, a single patient was used and the data were merged and then normalized and scaled using SCTransform. Linear dimension reduction was performed using principal component analysis (PCA), and the number of PCA dimensions was evaluated and selected based on assessment of an ElbowPlot. Data were clustered using the Louvain-graph-based algorithm in R and cluster resolution was chosen on the basis of evaluation by the clustree program[60,61]. The UMAP data reduction algorithm was used to project the cells onto two dimensional coordinates. Subsequently, canonical marker genes were used to identify cellular compartments (epithelium, endothelium, mesenchymal and immune populations). Epithelial clusters were subsetted and clustering and UMAP reduction were repeated. Clusters were then assigned putative cell types based on annotation with canonical marker genes, or from assessment of top cluster-defining genes based on differential expression (using the FindConservedMarkers function in Seurat v.3). For intracluster gene expression differences, the FindConservedMarkers function was used to identify variation between specified clusters and the resultant gene sets were comparted using the MAST method[23]. Trajectory analysis was performed on the UMAP reduction using the R Slingshot package[27] with RAS cells or iRAS cells as an assigned starting point without assigned end points. Differential expression analysis along trajectories was performed using the tradeSeq R Package[62] and single-lineage analysis was performed using the StartVsEnd test[62]. Heat maps were constructed per cluster or with cells arranged in pseudotemporal ordering across the *x* axis, and were generated using the ComplexHeatmap R package. Scatterplots of gene expression versus pseudotime ordering were generated using a custom R function. Primary adult, fetal and hES cell data (as indicated) were integrated using Seurat v.3 reference-based integration with the primary human sample as reference for CCA analysis. For the hESC RUES2-SC temporal analysis integration, cells expressing SCGB3A2 or SFTPC were selected and time points were merged, and SCTransform was used to normalize and scale for number counts, number features and mitochondria (>5% expression). Trajectory analysis was again performed using Slingshot on UMAP reduction, and cluster 0, corresponding to the day 0 time point, was assigned as the starting point with no specified end points. COPD diseased samples were processed the same as the primary normal human samples described. COPD and normal integration was performed with Seurat v.3 with normal human donor

chosen as the reference. Trajectory analysis was performed with intratrajectory comparison using tradeSeq with StartVsEndTest function. GO and WikiPathway enrichment analysis was performed using the clusterProfiler R package[63].

Histology analysis

Tissue was cut from the distal or proximal regions of the human lungs as described above. 2 cm × 2 cm pieces of peripheral parenchyma or 1 cm rings of airway were washed two to five times in cold PBS and then placed in 2% or 4% paraformaldehyde (PFA) overnight. The tissue was washed with PBS and dehydrated to 100% ethanol over the subsequent 12–24 h, paraffin embedded and then sectioned at a thickness of 8 μM. Ferret lungs were fixed in 10% neutral buffered formalin and processed into 10 μm paraffin sections (use approved by University of Iowa Institutional Animal Care and Use Committee, and by the University of Alabama at Birmingham Institutional Animal Care and Use Committee.). Organoids from hES cell experiments grown in Matrigel (Corning, 356231) and used for histology were embedded in Histogel (VWR, 83009-992) or 2–4% low-melting-point agarose (Invitrogen, 16520050), and subsequently fixed in 2% PFA for 30 min followed by washes in PBS, dehydration, paraffin-embedding, sectioning and immunostaining. IHC analysis was performed using the following antibodies on paraffin sections: anti-SCGB1A1 (rat, Novus, MAB4218, 1:200), anti-SCGB3A2 (mouse, Novus, H00117156-M01, 1:100), anti-SCGB3A2 (goat, R and D, RB01, 1:100), anti-SFTPC (rabbit, Millipore, AB3786, 1:100), anti-DC-LAMP (rat, Novus, DDX0191P-100, 1010E1.01, 1:100) and anti-EPCAM (Abcam, ab32392, 1:50). The slides were imaged using the Zeiss LSM 710 Confocal microscope. For cell counting, all of the human experiments included 4–5 patients per group, 6–10 ×20 high-power fields per patient containing a respiratory bronchiole, data were presented as mean per patient and statistical analysis was performed using ANOVA with Tukey's honest significant difference test ($P < 0.05$). For the ferret experiments, there were $n = 4$ animals per group, 5 ×20 high-power fields per animal containing a respiratory bronchiole, data were presented as mean per animal and statistical analysis was performed using two-tailed *t*-tests ($P < 0.05$). Blinding was not possible due to the obvious histological differences observed between samples and controls.

Human ES cell culture

The original iRASC/RUES2-SCGB3A2-mCherry line was previously described[29]. The dual reporter, SFTPC–eGFP and SCGB3A2–mCherry line was generated using the iRASC/RUES2-SCGB3A2-mCherry line and targeting the SFTPC locus using TALEN plasmids and a previously described targeting strategy[31]. The SFTPC-tdTomato targeting vector plasmid was altered to replace tdTomato with eGFP. The resulting targeting vector was used along with the SFTPC TALENs (left) 5′-TAGCACCTGCAGCAAGATGG-3′ and (right) 5′-TCACCGGCGGGCTCTCCATC-3′. The puromycin cassette was excised by transient transfection of a CAG-CRE-GFP plasmid. Karyotype analysis was performed by Cell Line Genetics and was determined to be normal. During maintenance, all ES cell lines were cultured on plates precoated with Matrigel and maintained on mTeSR1 medium (Stem-Cell Technologies) to maintain cell growth and stemness. All cultures were maintained with Primocin (Invitrogen) to prevent mycobacterium infection and were quarterly tested to ensure that they remained mycobacterium free.

ES cell lung epithelial differentiation assays

The protocols described are derived from previous protocols[29–31,35]. In brief, NKX2.1$^+$ lung progenitor cells were generated from ES cells by the induction of definitive endoderm using the STEMdiff Definitive Endoderm Kit (StemCell Technologies) for 72 h. Endoderm was dissociated and passaged in small clumps to Matrigel-coated tissue culture plates in complete serum free differentiation medium (cSFDM) and NKX2.1 primordial lung progenitors were derived as previously described[35]. Lung progenitors were sorted on the basis of expression of CPM (Wako Chemicals, 014-27501, 1:200)[64], and subsequently replated in three-dimensional growth-factor-reduced Matrigel at 50,000 cells per 100 µl of Matrigel. Cultures were maintained at 37 °C for 20 min to allow to stiffen and then medium was added at a sufficient volume to cover the droplets. For the derivation of SCGB3A2–mCherry-positive iRAS cells, lung progenitors were grown for 21 days in airway medium, which consists of cSFDM with FGF2 (rhFGFbasic; R&D), 100 ng ml^{-1} FGF10 (R&D), 50 nM dexamethasone (Sigma-Aldrich), 100 nM 8-bromoadenosine 3′,5′-cyclic monophosphate sodium salt (Sigma-Aldrich), 100 mM 3-isobutyl-1-methylxanthine (IBMX; Sigma-Aldrich) and 2 mM TZV (Selleckchem)[30]. Dispase II (Gibco, 17-105-041) was used to dissociate the Matrigel droplets and then 0.25% trypsin was used to generate a single-cell suspension from the dissociated SCGB3A2–mCherry$^+$ airway organoids. These cells were sorted using FACS and single cells were replated in Matrigel droplets at either 50,000 cells per 100 µl Matrigel droplet in airway medium, or 10,000 cells per 100 µl Matrigel droplet and cultured in alveolar medium, which consists of cSFDM with 3 µM CHIR99021 (Cayman), 10 ng ml^{-1} rhKGF (R&D), dexamethasone, 8-bromoadenosine 3′,5′-cyclic monophosphate sodium salt, 3-isobutyl-1-methylxanthine and 2 mM TZV as above[31]. Subsequent cell collection and sorting were completed as described and FACS was performed on either mCherry$^+$ populations or eGFP$^+$ populations as noted, and the samples were returned to Matrigel droplets as described in airway or alveolar medium. To assess cell differentiation of iAT2 cells to iRAS cells, iAT2 cells were sorted on the basis of eGFP expression and returned to airway medium. For Notch inhibition and Wnt activation experiments 10 µM DAPT (Cayman, 13197) or 3 µM CHIR99021 (Cayman) were added to cultures as indicated. For scRNA-seq experiments, single cells were isolated as described, counted and loaded onto the GemCode Instrument (10x Genomics) as described above.

Organoid imaging and quantification

Organoids were imaged directly within Matrigel droplets to enable visualization of the fluorescent reporters without interruption of culturing using either the Nikon Eclipse Ni wide-field microscope or the Thermo Fisher Scientific EVOS Imaging system. Images were acquired at a low magnification using a ×1.2 lens. For IHC, Matrigel drops were carefully removed from the well plates and embedded in HistoGel (Thermo Fisher Scientific). Samples were placed at 4 °C for 15 min and then the samples were fixed with 2% PFA for 30 min. The samples were washed with PBS then dehydrated and embedded in parafilm before sectioning for antibody staining.

RNA isolation and qPCR analysis

ES cells grown in organoids were collected from Matrigel droplets by dissociation with dispase, and the organoids were pelleted. PureLink RNA Micro Kit (Invitrogen) was used to isolate RNA according to the manufacturer's protocol. Total RNA (200–300 ng) was used for cDNA synthesis using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific). qPCR was performed using 2× Power SYBR green reagents on the QuantStudio 7 Thermocycler (Life Technologies). A list of primers and genes is provided in Supplementary Table 3.

Primary human epithelial organoids

An immune-cell-depleted single-cell suspension of peripheral human lungs was obtained as described previously[65]. Immune-cell-depleted single-cell suspension was stained with the following antibodies: anti-DRAQ7 (Cell Signaling Technologies, 7406S, 1:200) for dead cell exclusion, mouse anti-human CD45 APC-Cy7 (BioLegend, 304014, 1:200), mouse anti-human anti-CD11b-APC-Cy7 (BD, 557754, 1:200), mouse anti-human CD31 APC-Cy7 (BioLegend, 303120, 1:200), mouse anti-human CD326 PE (BioLegend, 324206, 1:200), mouse IgM anti-human HTII-280 (Terrace Biotech, 303118, 1:200), mouse anti-human NGFR-APC (BioLegend, 345108, 1:100) and mouse anti-human CEACAM6-PE (Thermo Fisher Scientific, 12-0667-42, 1:100). Finally, goat anti-mouse IgM Alexa Fluor 488 (Thermo Fisher Scientific, A-21042, 1:1,000) was added to label HTII-280^{+} cells. Cells were sorted using the BD FACSAria II (BD Bioscience) according to the manufacturer's instructions. Next, 10,000 sorted cells were resuspended in 40 µl of growth-factor-reduced Matrigel (Corning, 354230) and placed as droplets directly in cell culture plates. Cell–Matrigel suspension was allowed to polymerize at 37 °C for 20 min followed by addition of the appropriate cell culture medium.

Imaging of primary human organoids

At day 10–14, organoids were collected by incubating for 30 min with 3 U ml^{-1} of dispase II (Thermo Fisher Scientific, 17105041) at 37 °C. Organoids were fixed in 4% PFA at room temperature for 15 min, followed by resuspending in 15 µl of growth-factor-reduced Matrigel. Polymerized organoid–Matrigel suspensions were then embedded in OCT, and 7-µm-thick cryosections were prepared. Cryosections were stained with rabbit anti-SFTPC antibodies (AB3786; Millipore Sigma; 1:2,000) and visualized using AxioImager.M1 (Zeiss).

Flow cytometry, cytospin and immunocytochemistry analysis

As described above, ES cell organoids were extracted from Matrigel droplets and treated with dispase II to allow for the generation of a single-cell suspension followed by dissociation in 0.25% trypsin to generate a single-cell suspension. For primary human cells, a single-cell suspension was generated as described above and AT2 cells were isolated using the MACS multisort kit, MACs LS columns and the following antibodies: anti-HT2-280 (mouse IgM, Terrace, TB-27AHT2-280, 1:50) and anti-mouse IgM microbeads (Miltenyi, 130-047-302, 1:200). For flow cytometry, isolated single cells were assessed directly for fluorescence using the LSR Fortessa (BD) system or sorted using the MoFlo

Astrios (Beckman Coulter), FACSJazz (Beckman Coulter) or FACSAria Fusion (FACSAria fusion) system. Analysis was performed using FlowJo v.10 or v.11. For cytospins and immunocytochemistry, isolated single cells were resuspended at 10,000 cells per 100 μl of PBS, and cytospins were performed onto single glass-coated Shandon cytoslides (VWR), followed by brief fixation in 2% PFA, washing in PBS and antibody staining by permeabilization with 0.1% Triton X-100 and blocking with 0.25 g BSA fraction V in normal goat serum (Jackson, 005-000-121). For SFTPC cytospins, anti-SFTPC antibodies (Millipore Sigma, rabbit, AB3786, <1:100) was used with secondary goat anti-rabbit IgG AF488 (Jackson, 111-545-144, 1:1,000). Slides were imaged using the Nikon Eclipse Ni wide-field microscope.

SDS–PAGE and immunoblotting analysis

ES cells were collected from Matrigel droplets as described above and subsequent protein isolation was performed. Human primary AT2 cells were isolated as described above. Samples were loaded for SDS–PAGE using Novex Bis-Tris gels (NP0301, Thermo Fisher Scientific), then PVDF membranes were processed for immunoblotting using primary antibodies followed by species-specific HRP-conjugated secondary antibodies, and bands were detected using enhanced chemiluminescence (ECL2, 80196, Thermo Fisher Scientific) using the LiCOR Odyssey Imaging Station. The antibodies used were as follows: proSFTPC (NPRO18, gift from M. Beers, 1:3,000) mature SFTPC (Seven Hills, WRAB-76694, 1:2,500) and beta-actin (Sigma Aldrich, A1987, 1:5,000). The secondary antibodies used were goat-anti-mouse HRP (Bio-Rad, 170-6516, 1:10,000) and goat anti-rabbit HRP (Bio-Rad, 170-6515, 1:10,000).

Assessment of ferret airways and cigarette smoke exposure in ferrets

For the assessment of normal ferret airways, the ferret experiment protocol (0031945) in this study conformed to NIH standards and was approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Iowa, and all relevant guidelines and regulations were followed. All of the ferrets were purchased from Marshall Farms and were housed in separate cages under a controlled temperature (20–22 °C) and a light cycle of 16 h–8 h light–dark, with ad libitum access to water and diet obtained from Marshall Farms. Smoking exposure was performed as previously reported[48]. In brief, wild-type adult ferrets (*Mustela putorius furo*), with equal numbers of males and females, from Marshall BioResources were randomly assigned to cigarette smoke exposure or air control groups. Ferrets were restrained in customized nose-only exposure tubes with a 24-port plenum connected to mainstream smoke output and exposed to 60 min of smoke from 1R6F research cigarettes (University of Kentucky) twice daily for 6 months. Ferrets were exposed to 200 $\mu g\ l^{-1}$ of particulate matter (35 ml puffs of 2 s duration at a rate of $3\ l\ s^{-1}$), as previously described[48]. Cigarette smoke was generated by an ADVANCE automated cigarette smoke generator (TSE Systems). Animals were monitored continuously during exposure and gas monitored for O_2, CO_2, CO and particulate matter. Use was approved by the University of Alabama at Birmingham Institutional Animal Care and Use 20232, and all of the relevant guidelines and regulations were followed.

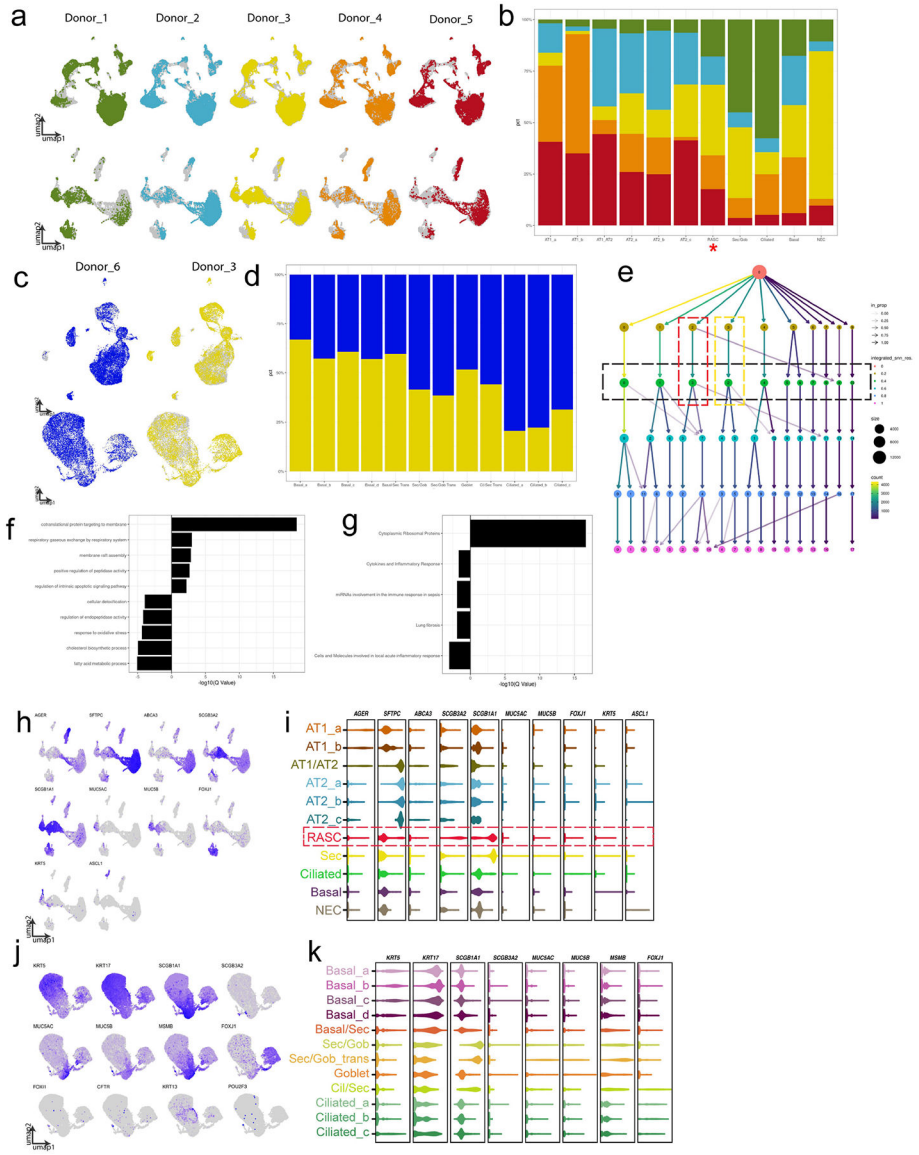
Statistical analysis

No statistical methods were used to predetermine sample size. Statistical analysis was performed in Prism for Mac, and R (statistical analysis for scRNA-seq datasets described above). The results are expressed as the mean ± s.e.m. An unpaired two-tailed Student's *t*-test for groups with equal variance was performed to determine *P* values for two-sided analysis. For experiments with more than two groups, ANOVA was performed and post hoc testing of significant results was performed using Tukey's honest significant difference test. All statistical tests were performed in a two-sided manner and were performed at the 5% significance level (that is, $\alpha = 0.05$) using GraphPad Prism software. The error bars designate s.e.m. unless indicated otherwise.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

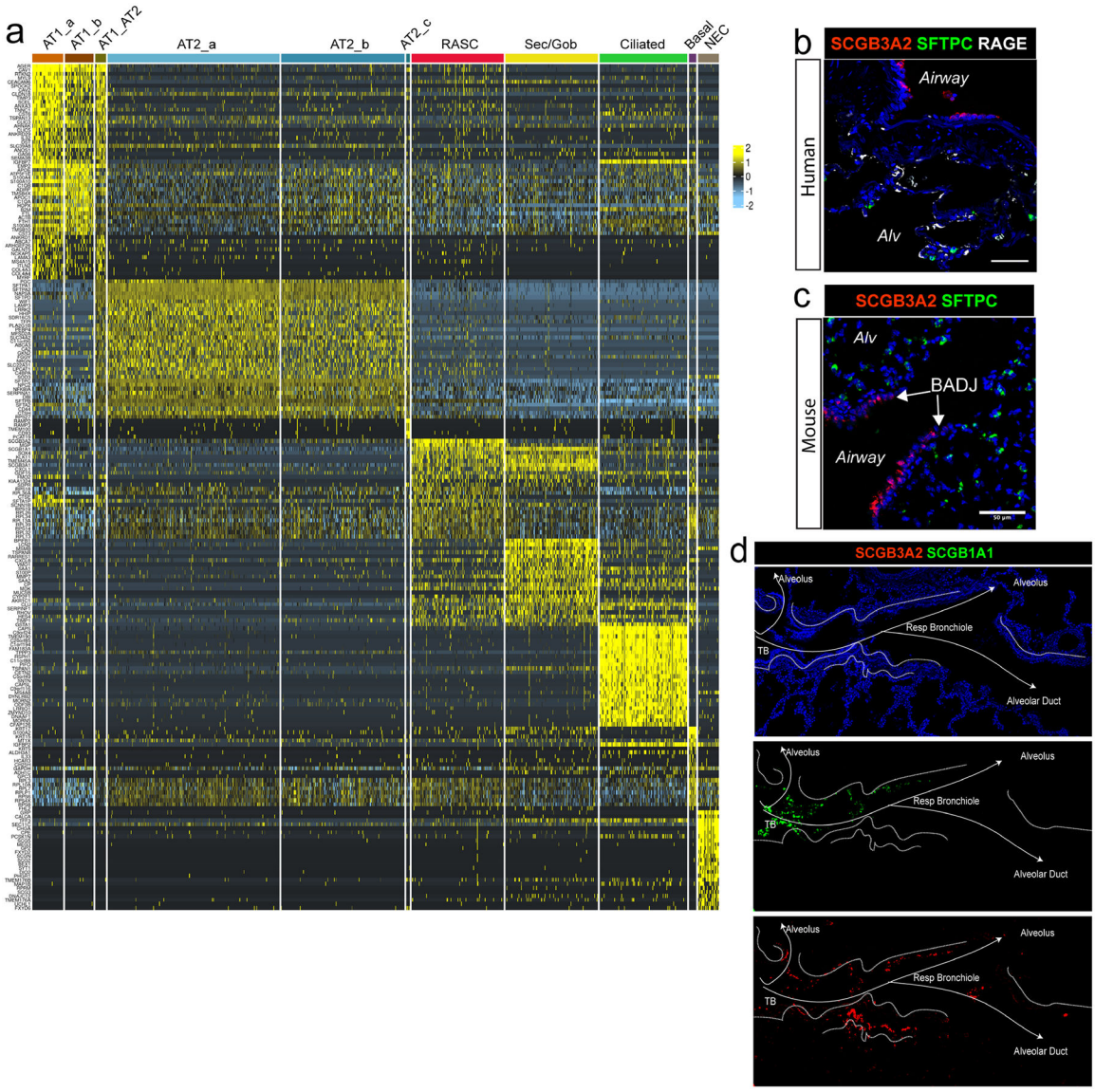
Extended Data



Extended Data Fig. 1 |. All sequenced patients contributed to all observed clusters in normal human epithelium and clusters are identified based on expression of known markers.

A) Distribution of individual patient data is shown overlaying concatenated UMAP of the full distal data set (top row) and the subset of epithelium (bottom row). **B**) The patient contribution to each of the epithelial clusters is shown, demonstrating that each patient contributed to each epithelial population. RAS cell cluster marked with red asterisk. **C**) The UMAPs of each patient that contributed to the proximal data sets is shown for the full data set (top row) and the subset of epithelium (bottom row). **D**) The patient contribution to each epithelial cluster is shown in stacked bar graph. **E**) Output of clustree analysis as used to determine optimal cluster resolution. Green box denotes chosen resolution for epithelial analysis shown, red box denotes RAS cell cluster, yellow box denotes Secretory cell cluster. Gene expression between canonical Secretory cells and RAS cells were compared and **F**)

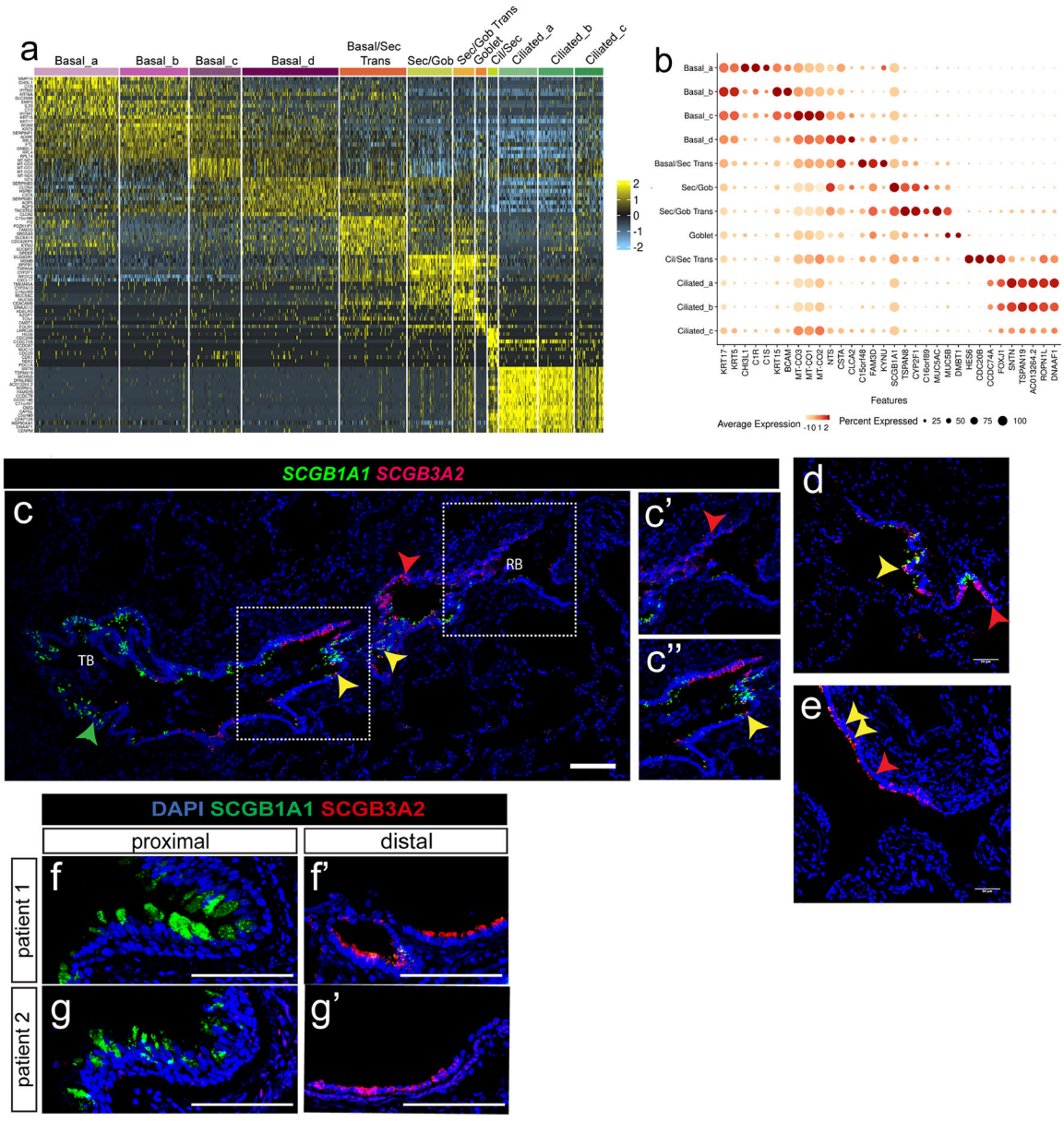
GO analysis and **G**) WikiPathway analysis was performed on differentially expressed genes between RAS cells and Secretory cells. Shown are categories either up or down regulated in RAS cells as compared to Secretory cells. Previously identified cell lineage markers were examined to identify various cell types within the epithelium in distal (**H**, **I**) and proximal (**J**, **K**) epithelial subsets. **H**, **J**) Feature plots show graded gene expression and distribution in UMAP. **I,K**) Violin plots show selected gene expression per cluster.



Extended Data Fig. 2 |. Epithelial clusters within the normal distal human lung each harbor unique gene expression patterns.

A) Heatmap of top cluster defining genes within the distal human lung epithelium demonstrates that *SCGB3A2+* RAS cells have a distinct transcriptional signature. Of note, AT2_c is a very small cluster with a small subset of cluster defining genes that are expressed at low levels and are not identified in most other cells. **B**) IHC of SCGB3A2, LAMP3, and RAGE on human lung distal parenchyma, demonstrating the interdigitation between

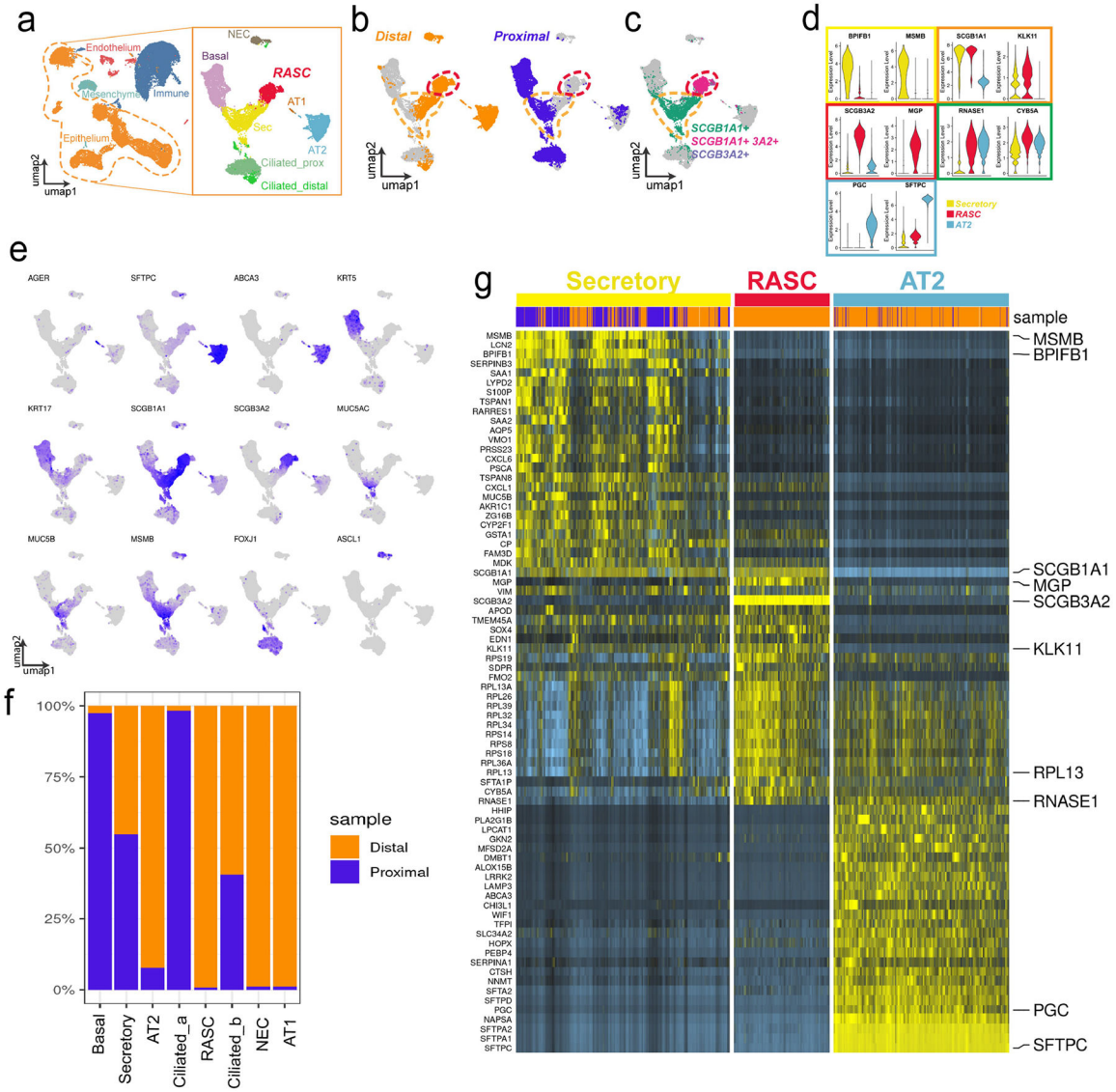
respiratory bronchioles and alveolar units. N = 3. Scale bar = 100 μm. **C**) In contrast to the distal human lung in B, distal mouse lung shows the termination of the mouse terminal bronchiole directly into the alveolar space. N = 2. Scale bar = 50 μm. **D**) Separate channels from IHC presented in Fig. 1h are presented to allow for enhanced evaluation of each protein.



Extended Data Fig. 3 |. Epithelial clusters in the normal human lung proximal airways lack the distinct secretory subtypes found in distal ariways.

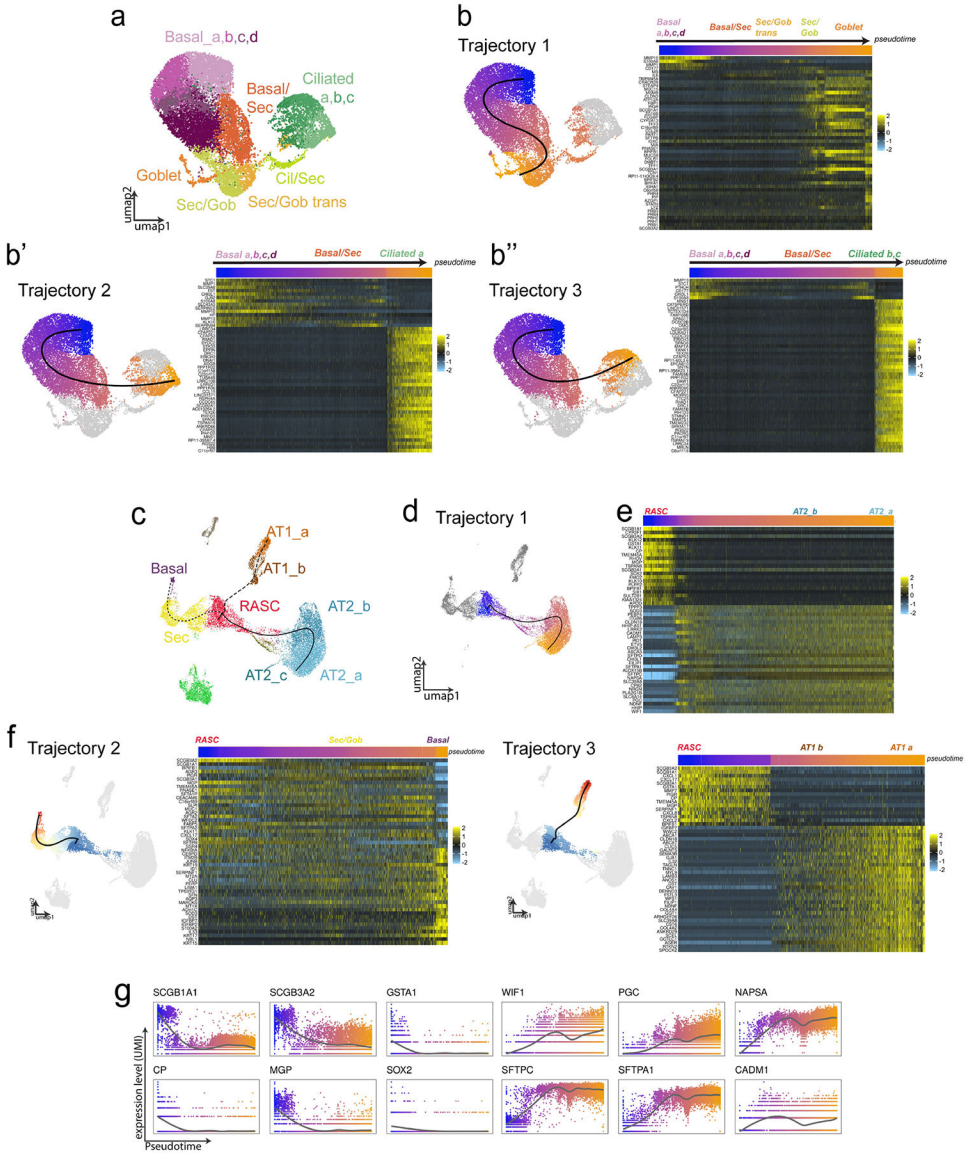
A) Heatmap of cluster defining genes within the proximal human epithelium. **B**) Selected gene expression compared across all proximal epithelial clusters shown as dot plot. **C**) RNAscope analysis for *SCGB1A1* and *SCGB3A2* transcripts reveals a similar distribution to protein expression in distal airways; A distal airway transitioning from terminal

to respiratory bronchiole is shown, highlighting the preserved gradient of *SCGB1A1* expressing cells in larger airways to *SCGB3A2* expressing cells in more distal airways. Scale bar = 500 μm. N = 3. **D**, **E**) Additional airway structure showing intermixing of *SCGB1A1* and *SCGB3A2* positive populations. Yellow arrows point to double positive cells along gradient. While with the scRNA analysis there is *SCGB1A1* expression in *SCGB3A2* positive cells, double positive cells were rarely observed in RNAscope or protein IHC, likely due to a threshold of the technique. Yellow arrows indicate rare double positive cells in a distal human respiratory airway. Red arrows indicate *SCGB3A2* single positive cells. Scale bars = 50 μm. N = 3. **F**, **G**) IHC of SCGB1A1 and SCGB3A2 expression in large and distal airways from representative individual patients (n = 6). Scale bar = 100 μm.



Extended Data Fig. 4 |. Integration of proximal and distal single cell transcriptomes defines the unique gene signature and location of human RASCs.

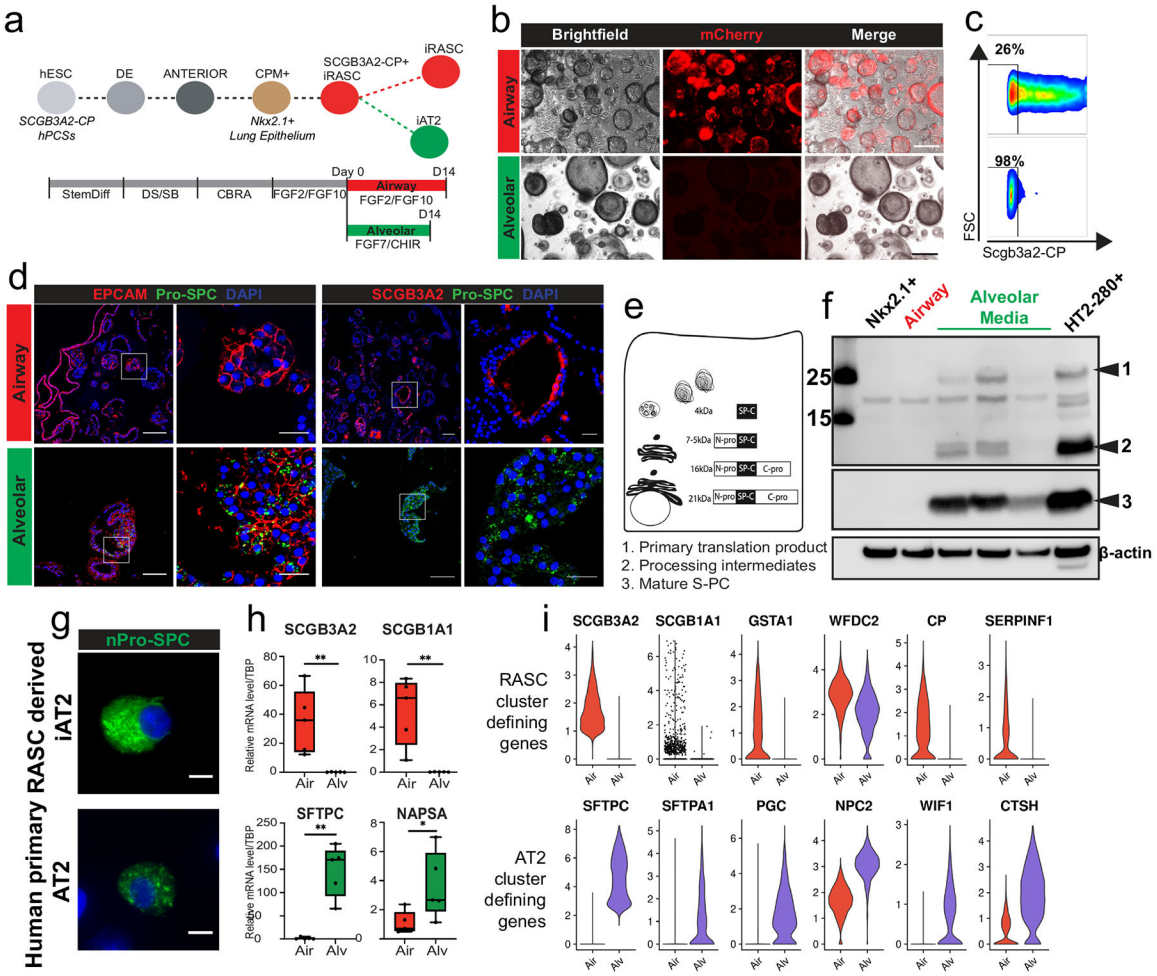
A) Distal and proximal scRNA-sequencing results from the same patient were concatenated, epithelium was isolated and re-clustered, and cell types were identified based on known markers. **B**) Distal and Proximal epithelium contribution to the resultant UMAP. **C**) Expression of *SCGB1A1* and *SCGB3A2*. Secretory cell cluster outlined in yellow, RAS cell clustered outlined in Red. **D**) Violin plots showing expression of selected genes, highlighting unique genes for each cluster, and shared gene expression between secretory and RAS cell clusters (orange box) and RAS cell and AT2 cell clusters (green box). **E**) Feature plots of both the proximal and distal epithelium highlighting key gene expression patterns used to identify known cell subtypes. **F**) Contribution of proximal and distal samples to each cell type, demonstrating that some groups are unique to each region while some are shared. **G**) Heat map showing gene expression of Secretory, RAS cell, and AT2 cell gene signatures from proximal and distal concatenated data set. Of note, these pairwise comparisons were done between the displayed groups only, as such the resultant gene expression is distinct from the cluster-defining genes we observed in Extended Data Fig. 2a, where the comparisons were done between all epithelial populations. Origin of cells shown in blue (proximal) and orange (distal) bar across top of heat map.



Extended Data Fig. 5 |. Trajectory of primary human proximal and distal epithelium demonstrates both anticipated and novel epithelial progenitor relationships.

A) Proximal airways epithelium clustering and resultant cell populations shown for reference for following analysis. **B**) Trajectory analysis in the proximal human airways results in 3 putative lineage relationships. UMAP plots show trajectories, color represents cell localization along pseudotime. Heatmaps showing expression of trajectory defining genes across pseudotime. **C**) Trajectory analysis of scRNAseq gene expression across distal epithelial populations suggests several epithelial relationship, including **D**) a putative relationship between RAS cells and AT2 cells. **E**) Heatmap of defining gene expression changes along pseudotemporal ordering of cells from RAS cell-to-AT2 cell trajectory. Pseudotemporal ordering shown in bar above heatmap. **F**) Additional epithelial trajectories were identified by slingshot analysis within the distal epithelium. For the additional trajectories that are not a focus of this current study, the individual trajectory is shown along with the corresponding heatmap of trajectory defining genes. Top indicated position

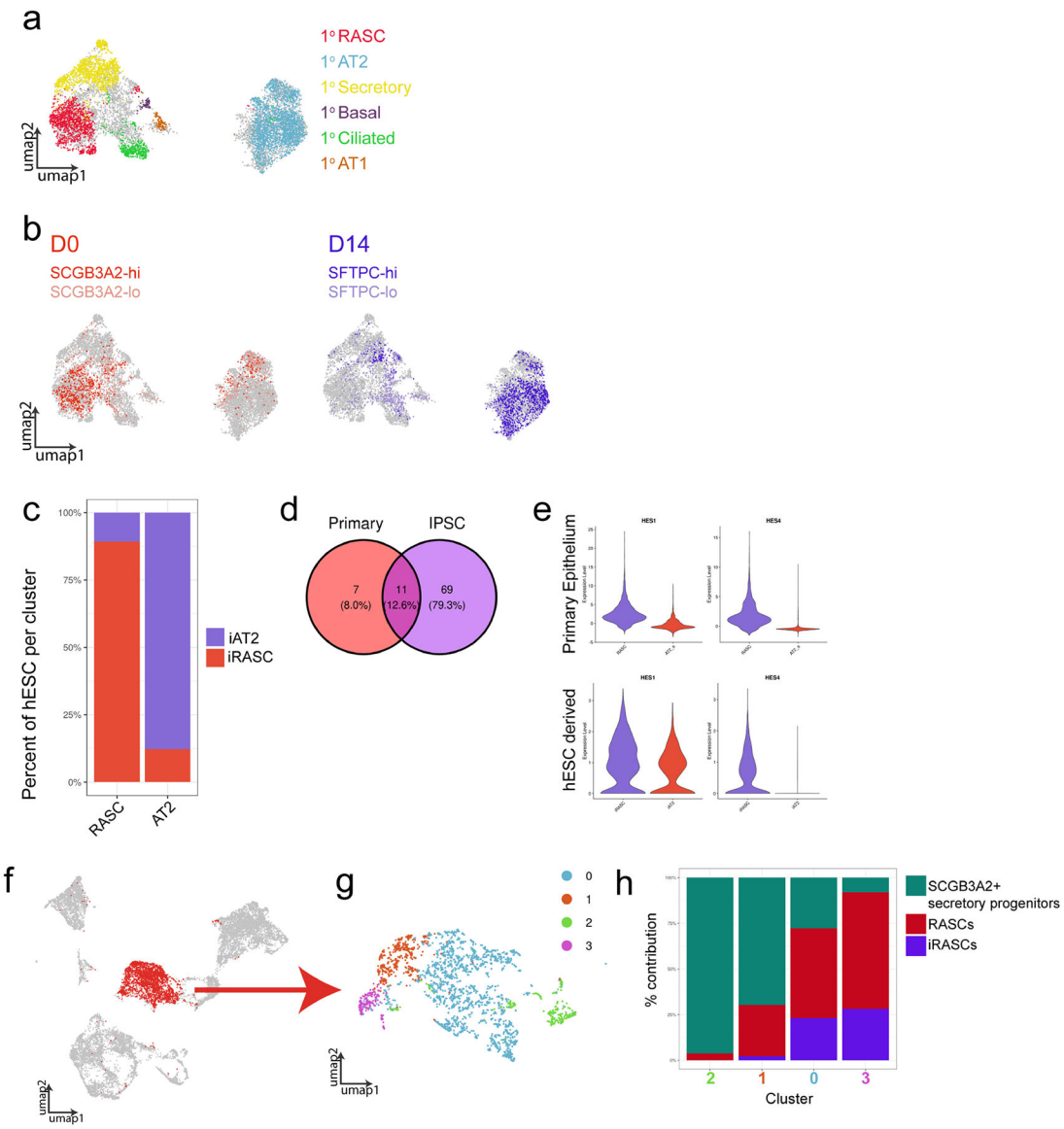
along trajectory as color coded by position of the cell along pseudotemporal ordering. **G**) Gene expression in UMI count of selected genes along pseudotemporal ordering of distal trajectory 1 (panel **d**, **e**) on the x-axis.



Extended Data Fig. 6 |. Human ES cell model of RAS cells demonstrates capacity to differentiate into AT2-like cells *in vitro*.

A) General experimental schematic of development of iRAS cells and propagation in airway or transition to alveolar media. **B**) Brightfield microscopy showing organoid formation and fluorescence after sorted iRAS cells were grown in airway (top) or alveolar (bottom) media. Scale bar = 100 μm. **C**) Corresponding flow cytometry analysis of endogenous mCherry reporter. **D**) Organoids from 3D culture with EPCAM or SCGB3A2 and SFTPC staining demonstrating SFTPC expression in alveolar organoids (bottom) and retention of SCGB3A2 expression in airways organoids (top). Scale bar = 100 μm, 20 μM for enlarged regions. N = 3 for panels **B**–**D**. **E**) Schematic of SFTPC protein processing in AT2 cells and **F**) western blot of primary translation product and processing intermediates (top, bands 1 and 2) and mature S-PC protein (band 3) in NKX2.1 progenitors, iRAS cells propagated in airway media, iRAS cells grown in alveolar media, and primary human AT2 cells (HT2-280+ cells). Molecular weights indicated on left. N = 3. **G**) SFTPC staining in individual iAT2 cells (top) and primary human AT2 cells (bottom) showing punctate nature of SFTPC staining. Scale

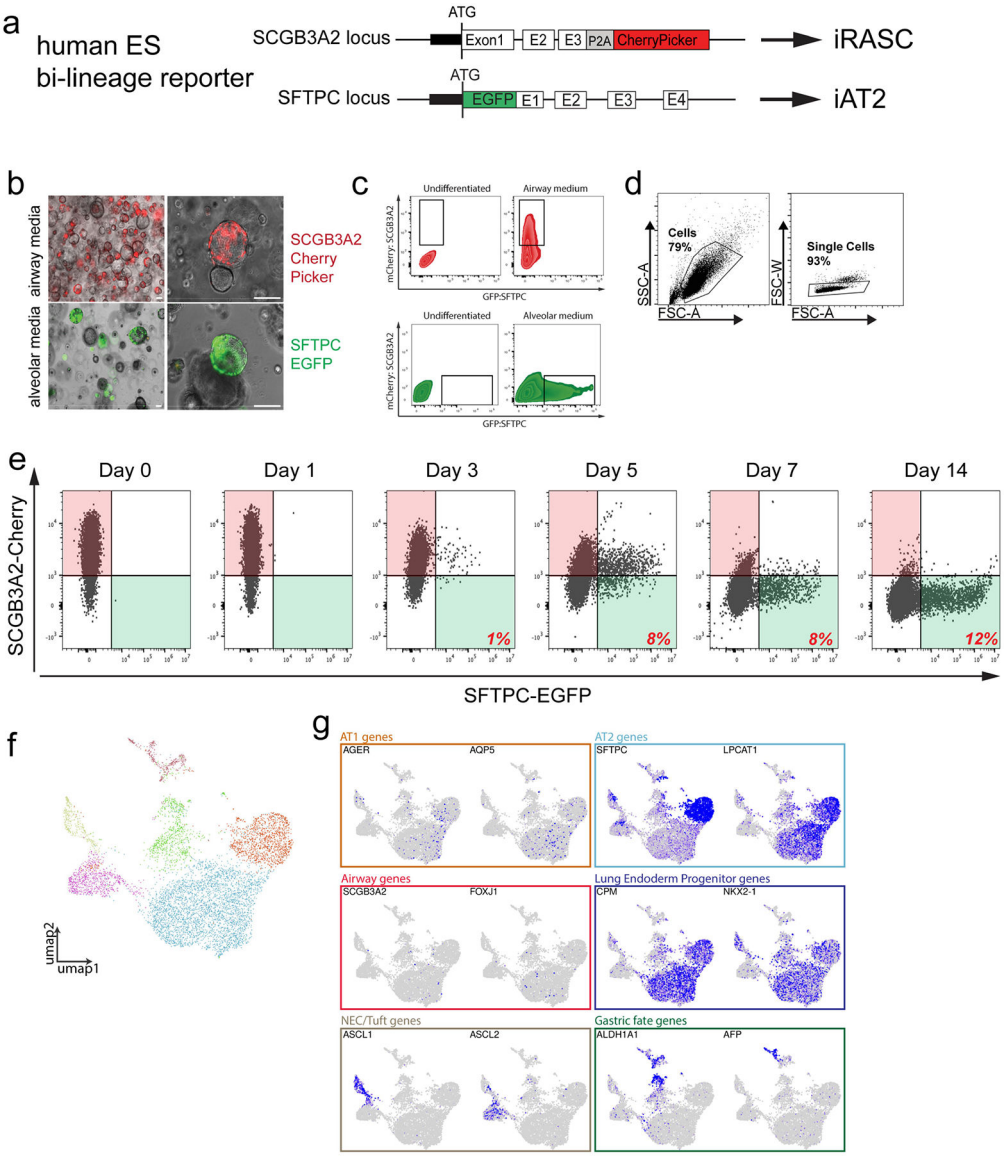
bar 10 μm. N = 4. **H**) q-RT-PCR from iRASC grown in airway (red) or alveolar (green) media demonstrates differential gene expression of known RAS cell and AT2 cell marker genes. Data presented as box with median (bar) and upper and lower quartiles (box bounds) and whiskers for min and max values (n = 5). Unpaired Student's test performed *p < 0.05, **p < 0.01. **I**) Gene expression based on scRNA sequencing of iRAS cells, and iRAS cells after 14 days in alveolar media showing downregulation of primary RAS cell defining genes and upregulation of AT2 defining genes. All N refer to biological replicates.



Extended Data Fig. 7 |. Reference based integration of primary epithelium and hES cell derived epithelial populations.

A) UMAP analysis showing the distribution of primary human epithelial populations within the concatenated data set of primary human epithelium, iRAS cells, and iAT2 cells. Grey cells represent non-primary human cells (ES derived populations). Colors correspond to primary human cell populations as indicated. All primary epithelial populations are shown.

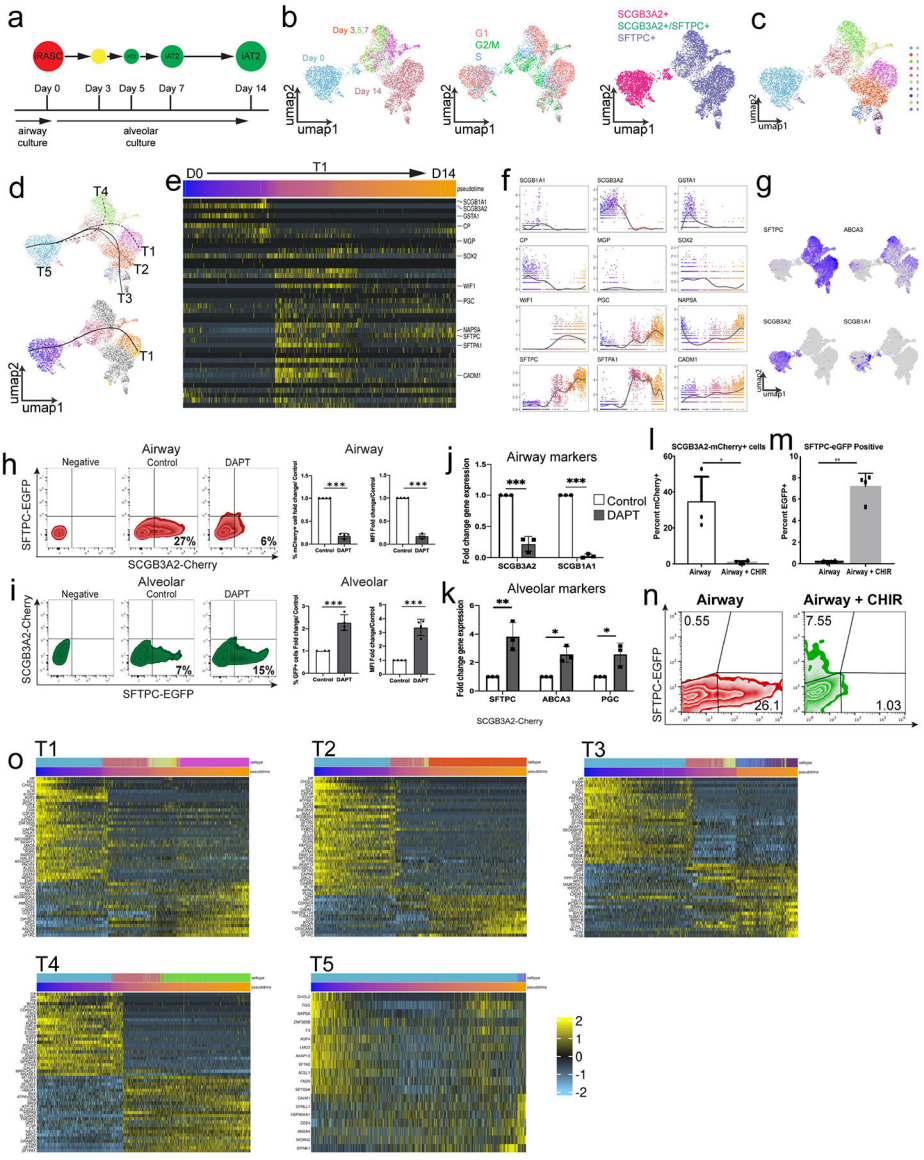
B) UMAP showing distribution of the hES cell populations included in the concatenated data set. Colors correspond to hES cell populations based on gene expression of *SCGB3A2* and *SFTPC* as indicated. **C**) Percentage of hES cells within primary cell defined clusters within concatenated data set. Clusters were identified based on localization of primary human epithelium. **D**) Venn diagram of Transcription factors identified as upregulated in primary RAS and iRAS cells compared to primary AT2 and iAT2 cells, respectively. **E**) Expression of Notch pathway genes *HES1* and *HES4* in primary RAS and iRAS cells compared to AT2 cell counterparts. **F**) Reference based integration of primary adult human epithelium, fetal human lung epithelial from day 11.5, 15, 18, and 21 (the days which included *SCGB3A2*+ secretory cell progenitor populations) from published fetal lung data set[32], and iRAS cells. All three *SCGB3A2*+ populations clustered together, and *SCGB3A2*+ cells are shown in red. **G**) *SCGB3A2*+ cells were selected and re-clustered. **H**) A stacked bar graph of the contribution of each population to each resultant cluster.



Extended Data Fig. 8 |. Dual reporter system demonstrates SFTPC expression in iAT2 cells and highlights dynamics of cell transitions *in vitro*.

A) Schematic of vectors for SFTPC-eGFP targeting in the dual reporter hES cell line. **B**) Brightfield microscopy showing mCherry and eGFP expression in 3D cultures of sorted iRAS cells grown in either airway (top) or alveolar (bottom) media for 14 days. Scale bar = 100 μm. N = 3. **C**) Flow cytometry showing endogenous SCGB3A2-mCherry and SFTPC-eGFP expression in iRAS cells grown in airway (top) or alveolar (bottom) media compared to NKX2.1 progenitor controls. **D**) Gating strategy for all hES cell flow cytometry and FACS experiments shown in Extended Data Figs. 6c, 8b, and 9h, i and n. **E**) Flow cytometry corresponding to experiment in Fig. 2a, of SCGB3A2-mCherry and SFTPC-eGFP expression over time as iRASC were propagated in Alveolar media. N = 3. **F**) UMAP of scRNA-seq analysis of all populations derived from iRAS cells at day 14 of differentiation in alveolar media reveals several clusters. **G**) The resulting culture was heterogenous and included both iAT2 lung endoderm progenitors as well as a small number

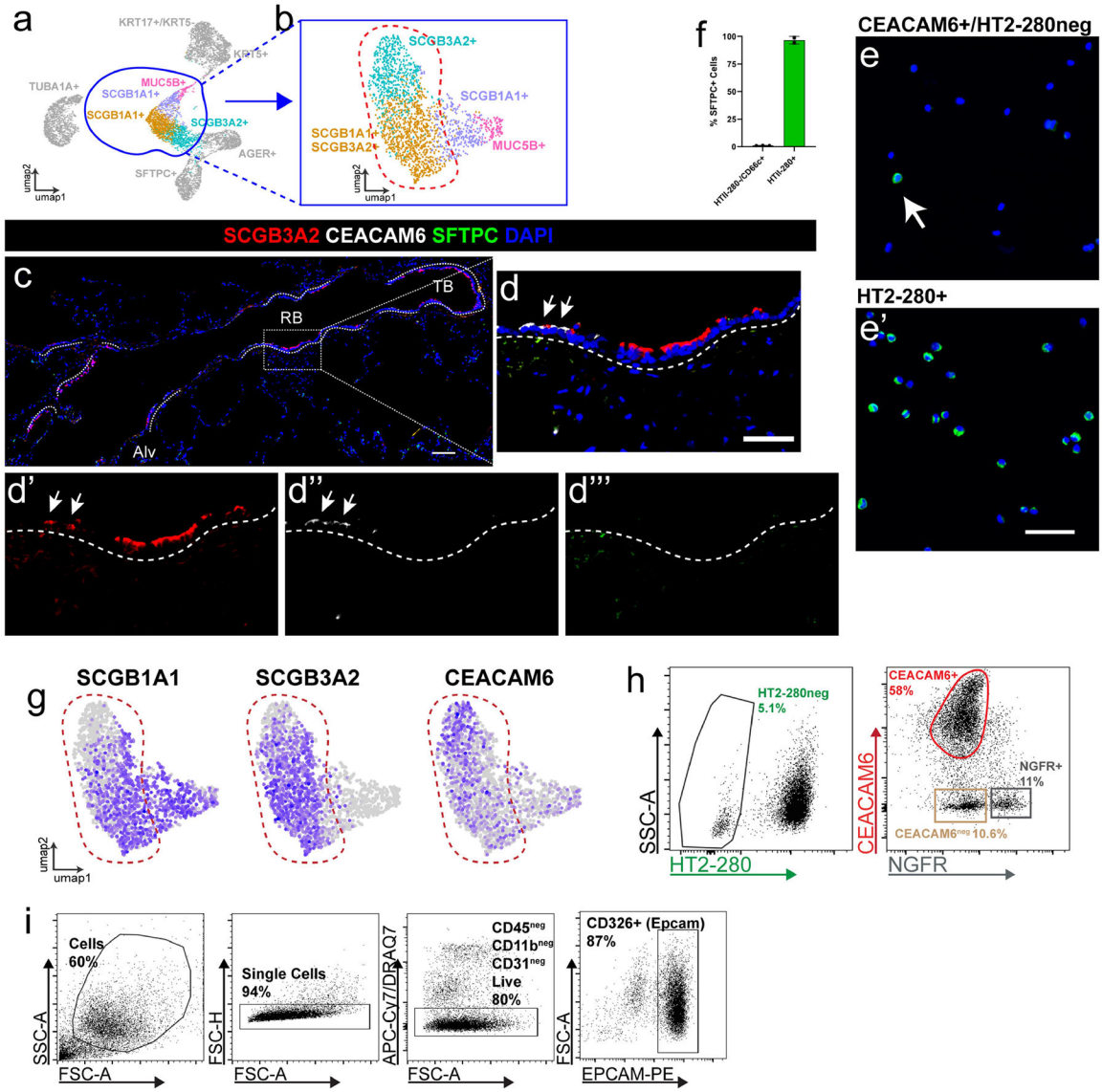
of other foregut endoderm cell types. Feature plots showing expression of canonical AT1 and AT2 cell alveolar epithelial markers, Airway cell markers, Lung endoderm progenitor markers, neuroendocrine and tuft cell markers, and gastric fate markers, allowing putative identification of all observed clusters.



Extended Data Fig. 9 |. Transition of iRAS cells to iAT2 cells is similar to the primary RAS cell-AT2 transition and is partially regulated by Notch and Wnt.

A) Schematic of time-course of scRNA-seq experiment. **B**) Integration of entire time course showing cell origin, cell-cycle phase, and gene expression of *SCGB3A2* and *SFTPC*. **C**) Clustering of complete time course from iRAS to iAT2 cells development shows multiple clusters within the various time points. **D**) Trajectory analysis showed multiple putative pseudotemporal orderings (top), and selected curve for further analysis based on termination in day 14 non-mitotic iAT2 cells (bottom). **E**) Heatmap of iRAS cell to iAT2 cell trajectory displaying genes defining the primary RAS to AT2 cell transition from Extended Data Fig.

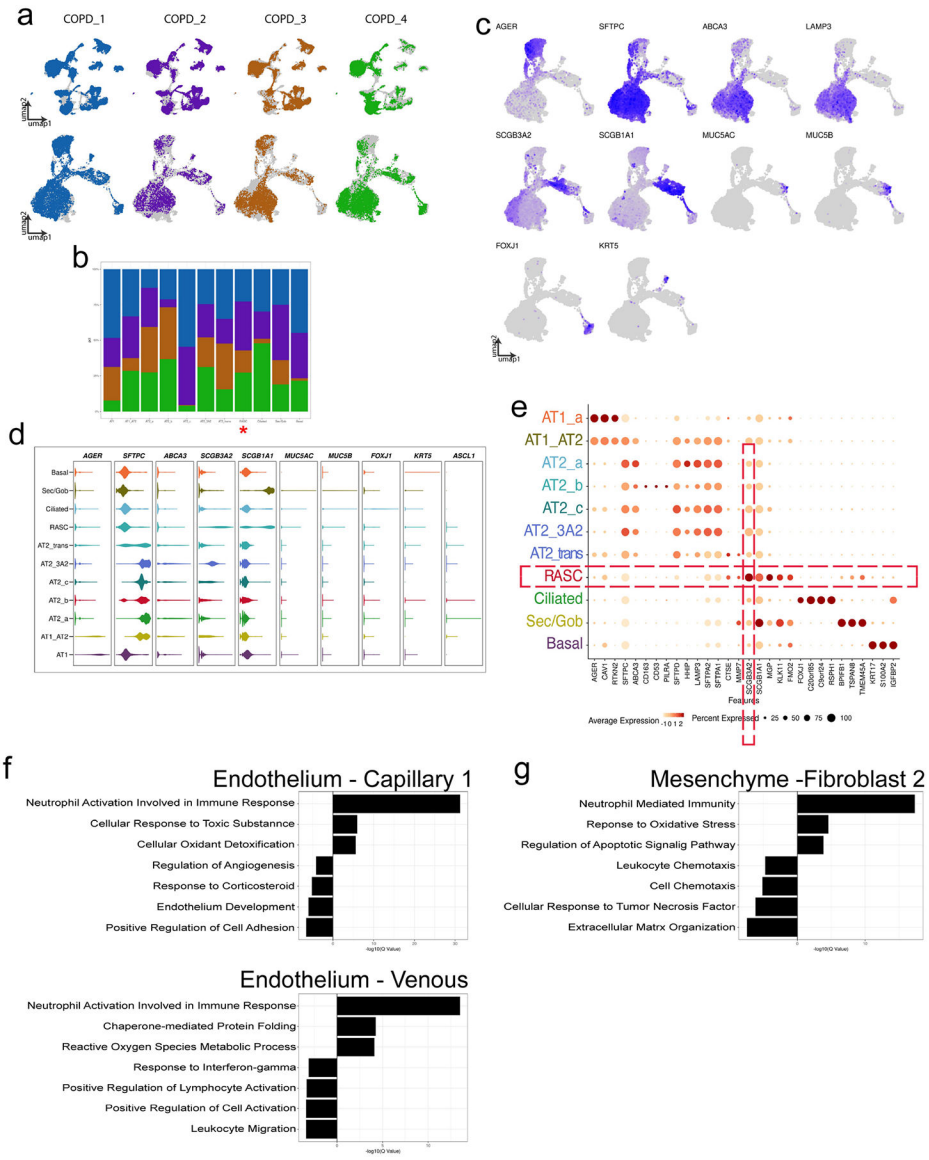
5e. **F**) Expression of a subset of genes identified in primary RAS to AT2 cell transition shown over pseudotime in iRAS to iAT2 cell transition. **G**) Canonical airway and alveolar epithelial marker genes expression within the resultant UMAP. **H**) Flow cytometry of mCherry and eGFP expression in iRAS cells grown in Airway media in the presence or absence of DAPT, and corresponding percent fold change in mCherry expression and MFI. N = 4. **I**) Flow cytometry of mCherry and eGFP expression of iRAS cells grown in Alveolar media in the presence or absence of DAPT showing percent change in eGFP expression and MFI. **J**) q-RT-PCR of bulk populations from iRAS cells grown in airway media with or without DAPT. N = 3. **K**) q-RT-PCR from iRAS cells gown in alveolar media with or without DAPT. N = 3. Quantification of flow cytometry analysis of **L**) mCherry and **M**) eGFP positive single cells in culture after iRAS cells were grown in either Airway media or Airway media supplemented with CHIR99021. Data are represented as mean +/−SD and unpaired two-tailed t-tests perfomed. *$p < 0.05$, **$p < 0.01$. **N**) Representative flow cytometry plots (representative of n = 4). **O**) The heat map of each observed trajectory in panel d is presented with the top trajectory defining genes identified, and cell are ordered by pseudotemporal order on x-axis. All N represent biological replicates.



Extended Data Fig. 10 |. Identification and isolation of CEACAM6+ distal lung epithelial cells and demonstration that SCGB3A2+ cells can be identified by CEACAM6 and isolated from distal lung parenchyma.

A) scRNA sequencing was performed on flow sorted Epcam+ HT2-280^{neg} cells in order to enrich for epithelial cells of the human distal airways. **B**) UMAP of re-clustering of selected secretory airway cell populations demonstrate refined heterogeneity in *SCGB3A2* expressing cells. **C**) View of distal airway showing distribution of *SCGB3A2*, *CEACAM6*, and *SFTPC* expression and **D**) Zoom in of region highlighting *SCGB3A2* and *CEACAM6* staining. White arrows indicate double positive cells. Of note, there are a small minority of CEACAM+/SCGB3A2low cells present in some sections. Scale bars = 100 μm, 50 μm for enlarged regions. N = 5. **E**) Representative cytospins and associated **F**) quantification of CEACAM6+ population for pro-SFTPC reveals that <5% of cells are positive for this canonical AT2 cell marker (n = 3), compared to HT2-280+ cells where over 90% of cells are positive for Pro-SFTPC (n = 2). Data presented as mean +/−SD. White arrow

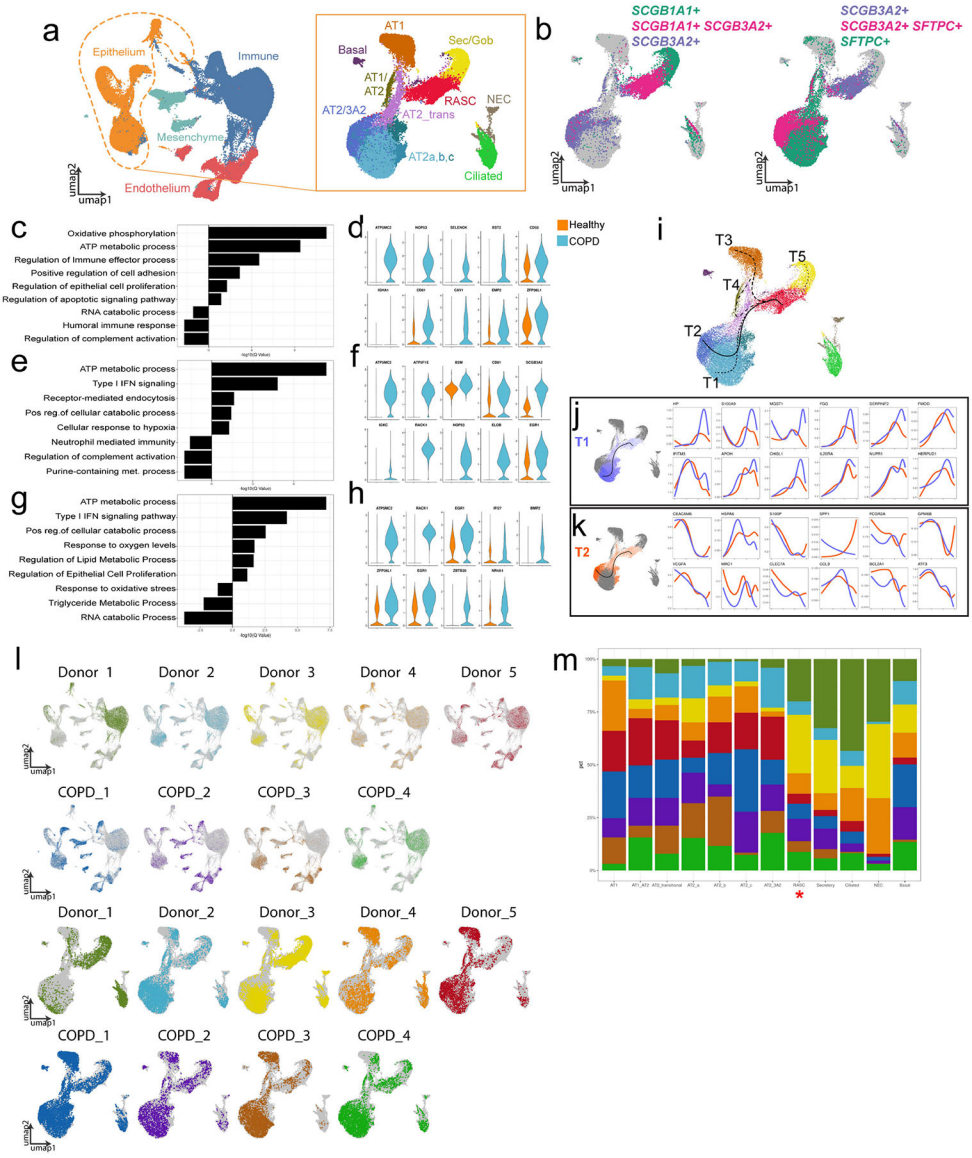
indicates rare SFTPC+ cell in CEACAM+ population. Scale bar = 50 μm. **G**) Feature plots demonstrates that surface marker CEACAM6 expression overlaps SCGB3A2 expression. **H**) FACS approach to isolate CEACAM6+/HT2-280−/NGFR-airway cells. HT2-280 neg population was sorted for a CEACAM6+ and NGFR- population. **I**) Gating strategy for obtaining populations in panel **H**. After gating on single cells, immune cells, endothelial cells and dead cells were excluded prior to selecting EPCAM+ cells for further subsetting as shown in **H**.



Extended Data Fig. 11 |. All COPD patients contributed to all resultant clusters and gene expression within the COPD epithelium highlights the various epithelial cell clusters.

Each patient contribution is shown overlying the entire concatenated UMAP for **A**) the entire data set (top row), and the subset of epithelial cells (bottom row). **B**) Stacked bar graph showing the patient contribution to each epithelial cell cluster, demonstrating that all patients contributed to all resultant clusters. RAS cell cluster marked with red asterisk.

C) Canonical marker genes used to identify epithelial clusters are shown. **D**) Violin plots demonstrate distribution of known canonical marker genes across the various epithelial populations. **E**) Cluster defining genes within the COPD epithelium are shown in dot plot format. There are distinct changes in the transcriptome of non-epithelial populations in COPD. GO analysis of inter-cluster gene expression comparing select endothelial (**F**) and mesenchymal (**G**) populations suggests that the differential gene expression seen in the epithelium is distinct to that population.



Extended Data Fig. 12 |. Epithelial cell gene expression differences in disease and the RAS cell to AT2 cell transition is altered in COPD.

A) Concatenation of normal and COPD peripheral samples and subset of epithelium showing expected epithelial populations. **B**) Identification of RAS cell and SCGB3A2+ AT2 populations based on expression of markers indicated. **C**) GO analysis of intra-cluster gene expression of RAS cells comparing COPD and healthy patient derived cells. **D**)

Violin plots of selected genes contributing to GO processes in (**C**). **E**) GO analysis of inter-cluster gene expression comparing SCGB3A2+ AT2 cells from COPD donors and AT2 cells from healthy donors and **F**) corresponding selected gene expression. **G**) GO analysis of intra-cluster gene expression of AT2 cells from COPD and healthy controls. **H**) Violin plots of selected genes from GO processes in (**G**). For all, up regulated is COPD compared to healthy controls. **I**)Transcriptional inference analysis of the concatenated data set revealing multiple trajectories initiating at RAS cells. **J**, **K**) Comparison of gene expression along trajectory 1 (T1) versus trajectory 2 (T2) demonstrating differential gene expression changes along pseudotemporal ordering between the RAS to AT2 cell trajectory and the RAS to SCGB3A2+ AT2 cell trajectory. **L**) Distribution of individual patient data is shown overlaying concatenated UMAP of the COPD and healthy peripheral data sets. All cells are shown in top two rows, and epithelial subsets in bottom rows. **M**) Stacked bar graphs highlight patient level contribution to each cluster. RAS cells are indicated in red asterisk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the patients who contributed to our study, without their willingness to participate in research, these studies would not be possible; the staff at the Cell and Developmental Microscopy Core at the University of Pennsylvania, the Flow Cytometry Core at The Children's Hospital of Philadelphia and the Animal Model Core of the Cystic Fibrosis Research Center of the University of Alabama at Birmingham for assistance in these studies; and M. Beers, V. Krymskya, S. Millar, A. Vaughan and S. Albelda for their critiques and insights throughout the development of the project. This work was supported by grants from the National Institutes of Health including HL148857, HL087825, HL134745 and HL132999 (to E.E.M.); 5T32HL007586-35 (to M.C.B.); 5R03HL135227-02 (to E.C.); K23 HL121406 (to J.M.D.); K08 HL150226 (to J.K.); DK047967, HL152960 and Federal Contract 75N92019R0014 (to J.F.E.); R35HL135816, P30DK072482, and U01HL152978 (to S.M.R.), R35HL150767 and U01HL134766 (to H.A.C.); and F32HL143931-01A1 and K99HL155785-01 (to J.J.K.). E.E.M. was also supported by the BREATH Consortium/Longfunds of the Netherlands. J.K. was supported by the Parker B. Francis Foundation. F.L.C.-D. was supported by a postdoctoral fellowship from GSK (RA3000034436). The embryonic stem cell experiments in this manuscript were not funded as part of the research agreement between University of Pennsylvania and GSK.

Data availability

All scRNA-seq data generated during this study have been deposited at the Gene Expression Omnibus database (accession numbers GSE168191 and GSE130076). The scRNA-seq datasets can be viewed online (http://bit.ly/2O45FIb). All cell lines and other reagents will be distributed on request. Source data are provided with this paper.

References

1. Burney PG, Patel J, Newson R, Minelli C & Naghavi M Global and regional trends in COPD mortality, 1990–2010. Eur. Respir. J 45, 1239–1247 (2015). [PubMed: 25837037]
2. Basil MC et al. The cellular and physiological basis for lung repair and regeneration: past, present, and future. Cell Stem Cell 26, 482–502 (2020). [PubMed: 32243808]
3. Zepp JA & Morrisey EE Cellular crosstalk in the development and regeneration of the respiratory system. Nat. Rev. Mol. Cell Biol 20, 551–566 (2019). [PubMed: 31217577]
4. Hogan BL et al. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. Cell Stem Cell 15, 123–138 (2014). [PubMed: 25105578]

5. Weibel ER, Sapoval B & Filoche M Design of peripheral airways for efficient gas exchange. Respir. Physiol. Neurobiol 148, 3–21 (2005). [PubMed: 15921964]
6. Weibel ER & Gomez DM Architecture of the human lung. Use of quantitative methods establishes fundamental relations between size and number of lung structures. Science 137, 577–585 (1962). [PubMed: 14005590]
7. Have-Opbroek AAW, Otto-Verberne CJM, Dubbeldam JA & Dÿkman JH The proximal border of the human respiratory unit, as shown by scanning and transmission electron microscopy and light microscopical cytochemistry. Anat. Rec 229, 339–354 (1991). [PubMed: 1708949]
8. Rock JR, Randell SH & Hogan BL Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. Dis. Models Mech 3, 545–556 (2010).
9. Mercer RR, Russell ML, Roggli VL & Crapo JD Cell number and distribution in human and rat airways. Am. J. Respir. Cell Mol. Biol 10, 613–624 (1994). [PubMed: 8003339]
10. Ryu JH, Myers JL & Swensen SJ Bronchiolar disorders. Am. J. Respir. Crit. Care Med 168, 1277–1292 (2003). [PubMed: 14644923]
11. Hogg JC, Macklem PT & Thurlbeck WM Site and nature of airway obstruction in chronic obstructive lung disease. New Engl. J. Med 278, 1355–1360 (1968). [PubMed: 5650164]
12. Koo H-KK et al. Small airways disease in mild and moderate chronic obstructive pulmonary disease: a cross-sectional study. Lancet Respir. Med 6, 591–602 (2018). [PubMed: 30072106]
13. Visscher DW & Myers JL Bronchiolitis: the pathologist's perspective. Proc. Am. Thorac. Soc 3, 41–47 (2006). [PubMed: 16493150]
14. Fretzayas A & Moustaki M Etiology and clinical features of viral bronchiolitis in infancy. World J. Pediatr 13, 293–299 (2017). [PubMed: 28470580]
15. Gómez R, Colás C, Sebastián A & Arribas J Respiratory repercussions in adults with a history of infantile bronchiolitis. Ann. Allergy Asthma Immunol 93, 447–451 (2004). [PubMed: 15562883]
16. Verleden SE et al. Small airway loss in the physiologically ageing lung: a cross-sectional study in unused donor lungs. Lancet Respir. Med 9, 167–174 (2020). [PubMed: 33031747]
17. Montoro DT et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature 560, 319–324 (2018). [PubMed: 30069044]
18. Plasschaert LW et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature 560, 377–381 (2018). [PubMed: 30069046]
19. Hyde DM, Samuelson DA, Blakeney WH & Kosch PC A correlative light microscopy, transmission and scanning electron microscopy study of the ferret lung. Scan. Electron Microsc 3, 891–898 (1979).
20. Sterner-Kock A, Kock M, Braun R & Hyde DM Ozone-induced epithelial injury in the ferret is similar to nonhuman primates. Am. J. Respir. Crit. Care Med 162, 1152–1156 (2000). [PubMed: 10988145]
21. Vinegar A, Sinnett EE, Kosch PC & Miller ML Pulmonary physiology of the ferret and its potential as a model for inhalation toxicology. Lab. Anim. Sci 35, 246—250 (1985). [PubMed: 4021438]
22. Bal HS & Ghoshal NG Morphology of the terminal bronchiolar region of common laboratory mammals. Lab. Anim 22, 76–82 (1988). [PubMed: 3352222]
23. Finak G et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 16, 278 (2015). [PubMed: 26653891]
24. Barkauskas CE et al. Type 2 alveolar cells are stem cells in adult lung. J. Clin. Invest 123, 3025–3036 (2013). [PubMed: 23921127]
25. Nabhan AN, Brownfield DG, Harbury PB, Krasnow MA & Desai TJ Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. Science 359, 1118–1123 (2018). [PubMed: 29420258]
26. Zacharias WJ et al. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. Nature 555, 251–255 (2018). [PubMed: 29489752]
27. Street K et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genom. 19, 477 (2018).

28. Rock JR et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. Proc. Natl Acad. Sci. USA 106, 12771–12775 (2009). [PubMed: 19625615]
29. McCauley KB et al. Single-cell transcriptomic profiling of pluripotent stem cell-derived SCGB3A2^{+} airway epithelium. Stem Cell Rep. 10, 1579–1595 (2018).
30. McCauley KB et al. Efficient derivation of functional human airway epithelium from pluripotent stem cells via temporal regulation of Wnt signaling. Cell Stem Cell 20, 844–857 (2017). [PubMed: 28366587]
31. Jacob A et al. Differentiation of human pluripotent stem cells into functional lung alveolar epithelial cells. Cell Stem Cell 21, 472–488 (2017). [PubMed: 28965766]
32. Miller AJ et al. In vitro and in vivo development of the human airway at single-cell resolution. Dev. Cell 53, 117–128 (2020). [PubMed: 32109386]
33. Butler A, Hoffman P, Smibert P, Papalexi E & Satija R Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol 36, 411–420 (2018). [PubMed: 29608179]
34. Jacob A et al. Derivation of self-renewing lung alveolar epithelial type II cells from human pluripotent stem cells. Nat. Protoc 14, 3303–3332 (2019). [PubMed: 31732721]
35. Hawkins F et al. Prospective isolation of NKX2–1-expressing human lung progenitors derived from pluripotent stem cells. J. Clin. Invest 127, 2277–2294 (2017). [PubMed: 28463226]
36. Guseh JS et al. Notch signaling promotes airway mucous metaplasia and inhibits alveolar development. Development 136, 1751–1759 (2009). [PubMed: 19369400]
37. Morimoto M, Nishinakamura R, Saga Y & Kopan R Different assemblies of Notch receptors coordinate the distribution of the major bronchial Clara, ciliated and neuroendocrine cells. Development 139, 4365–4373 (2012). [PubMed: 23132245]
38. Rock JR et al. Notch-dependent differentiation of adult airway basal stem cells. Cell Stem Cell 8, 639–648 (2011). [PubMed: 21624809]
39. Stupnikov MR, Yang Y, Mori M, Lu J & Cardoso WV Jagged and Delta-like ligands control distinct events during airway progenitor cell differentiation. eLife 8, e50487 (2019). [PubMed: 31631837]
40. Tsao PN et al. Notch signaling controls the balance of ciliated and secretory cell fates in developing airways. Development 136, 2297–2307 (2009). [PubMed: 19502490]
41. Geling A, Steiner H, Willem M, Bally-Cuif L & Haass C A γ-secretase inhibitor blocks Notch signaling in vivo and causes a severe neurogenic phenotype in zebrafish. EMBO Rep. 3, 688–694 (2002). [PubMed: 12101103]
42. Frank DB et al. Emergence of a wave of Wnt signaling that regulates lung alveologenesis by controlling epithelial self-renewal and differentiation. Cell Rep. 17, 2312–2325 (2016). [PubMed: 27880906]
43. Zepp JA et al. Distinct mesenchymal lineages and niches promote epithelial self-renewal and myofibrogenesis in the lung. Cell 170, 1134–1148 (2017). [PubMed: 28886382]
44. Chapin C et al. Distribution and surfactant association of carcinoembryonic cell adhesion molecule 6 in human lung. Am. J. Physiol. Lung Cell. Mol. Physiol 302, L216–L225 (2012). [PubMed: 22037359]
45. Lin SE et al. Expression of human carcinoembryonic antigen-related cell adhesion molecule 6 and alveolar progenitor cells in normal and injured lungs of transgenic mice. Physiol. Rep 3, e12657 (2015). [PubMed: 26702074]
46. Shikotra A et al. A CEACAM6-high airway neutrophil phenotype and ceacam6-high epithelial cells are features of severe asthma. J. Immunol 198, 3307–3317 (2017). [PubMed: 28275137]
47. Lin VY et al. Excess mucus viscosity and airway dehydration impact COPD airway clearance. Eur. Respir. J 55, 1900419 (2020). [PubMed: 31672759]
48. Raju SV et al. A ferret model of COPD-related chronic bronchitis. JCI Insight 1, e87536 (2016). [PubMed: 27699245]
49. Stanford D et al. Airway remodeling in ferrets with cigarette smoke induced COPD using microCT Imaging. Am. J. Physiol. Lung Cell. Mol. Physiol 319, L11–L20 (2020).

50. Ganguly K et al. Secreted phosphoprotein 1 is a determinant of lung function development in mice. Am. J. Respir. Cell Mol 51, 637–651 (2014).
51. Calabrese DR et al. Dectin-1 genetic deficiency predicts chronic lung allograft dysfunction and death. JCI Insight 4, e133083 (2019).
52. Regeling A et al. HSPA6 is an ulcerative colitis susceptibility factor that is induced by cigarette smoke and protects intestinal epithelial cells by stabilizing anti-apoptotic Bcl-XL. Biochim. Biophys. Acta 1862, 788–796 (2016). [PubMed: 26826017]
53. Kim CF et al. Identification of bronchioalveolar stem cells in normal lung and lung cancer. Cell 121, 823–835 (2005). [PubMed: 15960971]
54. Choi J et al. Release of Notch activity coordinated by IL-1β signalling confers differentiation plasticity of airway progenitors via Fosl2 during alveolar regeneration. Nat. Cell Biol 23, 953–966 (2021). [PubMed: 34475534]
55. Todd JL & Palmer SM Bronchiolitis obliterans syndrome: the final frontier for lung transplantation. Chest 140, 502–508 (2011). [PubMed: 21813529]
56. Verleden SE, Sacreas A, Vos R, Vanaudenaerde BM & Verleden GM Advances in understanding bronchiolitis obliterans after lung transplantation. Chest 150, 219–225 (2016). [PubMed: 27212132]
57. Verleden SE et al. The site and nature of airway obstruction after lung transplantation. Am. J. Respir. Crit. Care Med 189, 292–300 (2014). [PubMed: 24354907]
58. Ghorani V, Boskabady MH, Khazdair MR & Kianmeher M Experimental animal models for COPD: a methodological review. Tob. Induc. Dis 15, 25 (2017). [PubMed: 28469539]
59. Diamond JM et al. Clinical risk factors for primary graft dysfunction after lung transplantation. Am. J. Respir. Crit. Care Med 187, 527–534 (2013). [PubMed: 23306540]
60. Kiselev VY et al. SC3: consensus clustering of single-cell RNA-seq data. Nat. Methods 14, 483–486 (2017). [PubMed: 28346451]
61. Zappia L & Oshlack A Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience 7, giy083 (2018).
62. Van den Berge K et al. Trajectory-based differential expression analysis for single-cell sequencing data. Nat. Commun 11, 1201 (2020). [PubMed: 32139671]
63. Wu T et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation 2, 100141 (2021). [PubMed: 34557778]
64. Gotoh S et al. Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells. Stem Cell Rep. 3, 394–403 (2014).
65. Kathiriya JJ et al. Human alveolar type 2 epithelium transdifferentiates into metaplastic KRT5^{+} basal cells. Nat. Cell Biol 24, 10–23 (2022). [PubMed: 34969962]

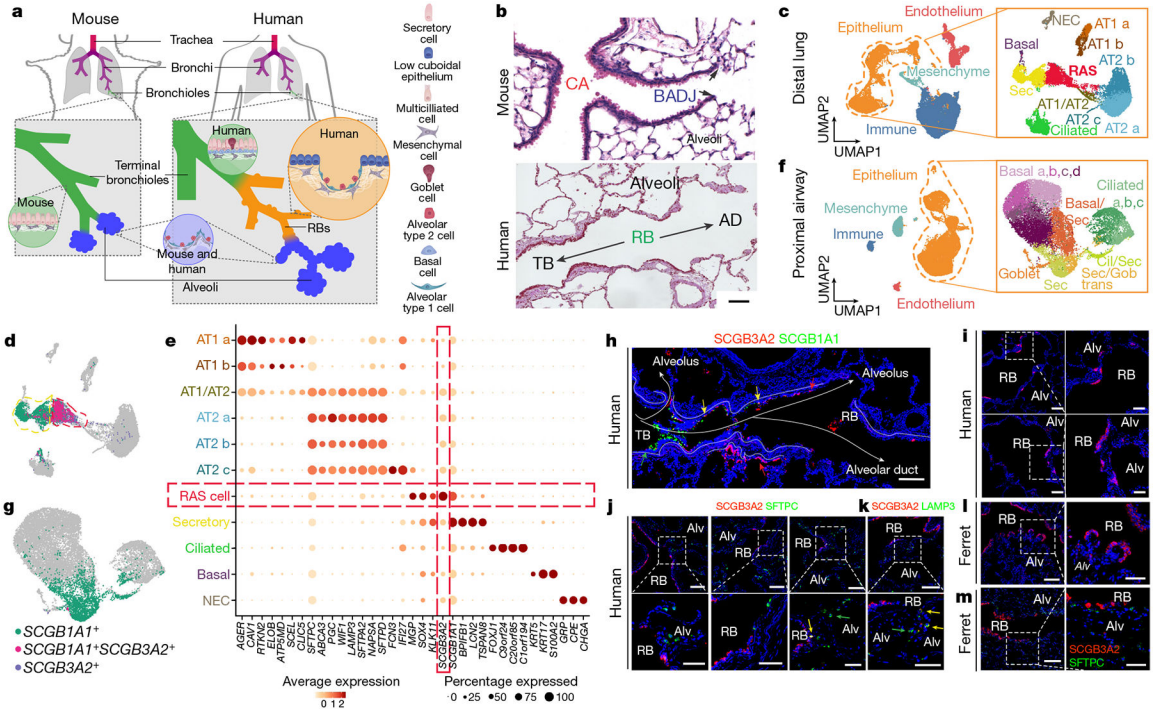


Fig. 1 |. Identification of secretory cell heterogeneity in the distal human lungs.

a, Schematic of the differences between the mouse and human distal airways including respiratory bronchioles (RBs). **b**, Haematoxylin-and-eosin-stained tissue showing the blunt ending of the mouse airways into the alveolar compartment, whereas the transition through the respiratory bronchioles in the human lung is more gradual. AD, alveolar duct; BADJ, bronchioalveolar duct junction; CA, conducting airway; TB, terminal bronchiole. Representative of $n = 10$. Scale bar, 100 μm. **c**, UMAP representation of the scRNA-seq data of the human distal lung with epithelial subset analysis labelled according to the corresponding cell type. NEC, neuroendocrine cell; Sec, secretory. **d**, Bi-gene plots showing expression of *SCGB1A1, SCGB3A2* or cells expressing both of these genes, which identifies RAS cells. **e**, Gene expression dot plot highlighting the unique expression pattern found in RAS cells. **f**, UMAP analysis of scRNA-seq data from proximal human airways with epithelial subset analysis labelled according to the corresponding cell type. Basal/Sec, basal and secretory; Cil/Sec, ciliated and secretory; Sec/Gob, secretory and goblet. **g**, Bi-gene plot showing the expression of *SCGB1A1* and *SCGB3A2* in the proximal epithelium. **h**, SCGB1A1 and SCGB3A2 expression in a distal human airway, showing the progression of SCGB1A1$^+$ and SCGB3A2$^+$ cells along the distal airway axis, interweaving with alveolar tissue. Scale bar, 100 μm. **i**, SCGB3A2 expression in human lungs in the respiratory bronchiole region. **j**, **k**, Co-staining of SCGB3A2 with either SFTPC (**j**) or LAMP3 (**k**) in the distal human lungs. Respiratory bronchioles are interspersed, and areas of alveolar parenchyma were identified by the presence of SFTPC$^+$ or LAMP3$^+$ AT2 cells. The images on the right are magnifications of the regions indicated by the dashed white boxes. The yellow arrows highlight rare double-positive cells, and the green arrows highlight single-positive cells. Scale bars, 100 μm and 50 μm (magnified images). For **h**–**k**, $n = 8$ for each. **l**, **m**, SCGB3A2 staining in the distal parenchyma of adult ferret lungs (**l**), and

SCGB3A2 and SFTPC staining in the distal ferret lung in regions of respiratory bronchioles (**m**). $n = 4$. The images on the right are magnifications of the regions indicated by the dashed white boxes. Scale bars, 100 μm and 50 μm (magnified images). Alv, alveolus.

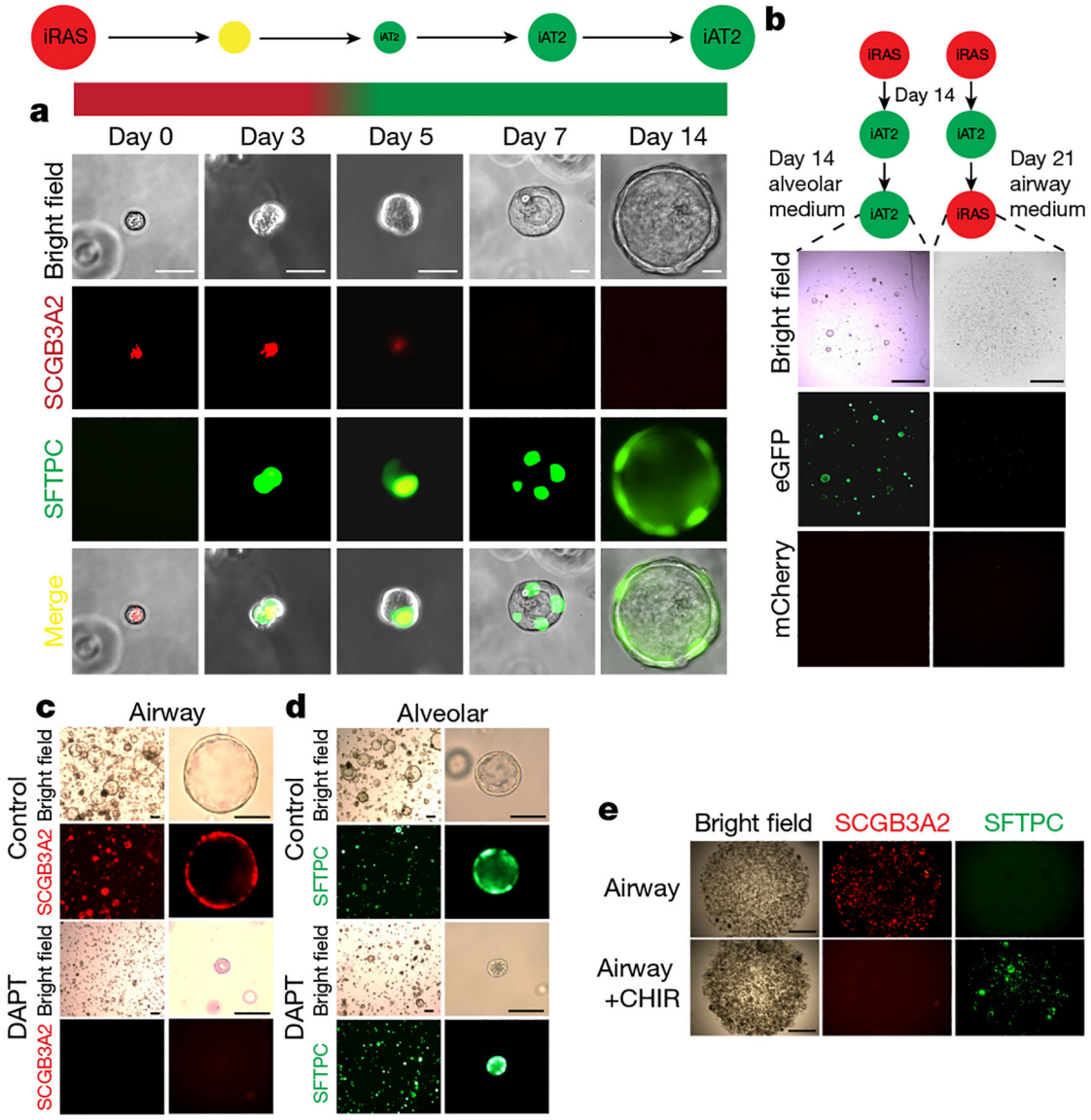


Fig. 2 |. The iRAS cell transition to iAT2 cells is rapid, unidirectional, and regulated in part by Notch and Wnt signalling.

a, Bright-field and fluorescence microscopy analysis of mCherry and eGFP expression over time in iRAS cells grown in alveolar medium. Scale bars, 20 μm. **b**, Bright-field and fluorescence microscopy analysis of mCherry and eGFP expression in iAT2 cells grown in continued alveolar medium (left) or returned to airway medium (right). Scale bars, 200 μm. **c**, **d**, Bright-field and fluorescence microscopy analysis of iRAS cells grown in airway medium (**c**) or alveolar medium (**d**) with vehicle (top) or the γ-secretase inhibitor DAPT (bottom). For **c** and **d**, scale bars, 100 μm. **e**, Bright-field and fluorescence microscopy analysis of iRAS cells grown in airway medium with or without CHIR99021 supplementation. Scale bars, 200 μm. All experiments represent $n = 4$ biological replicates.

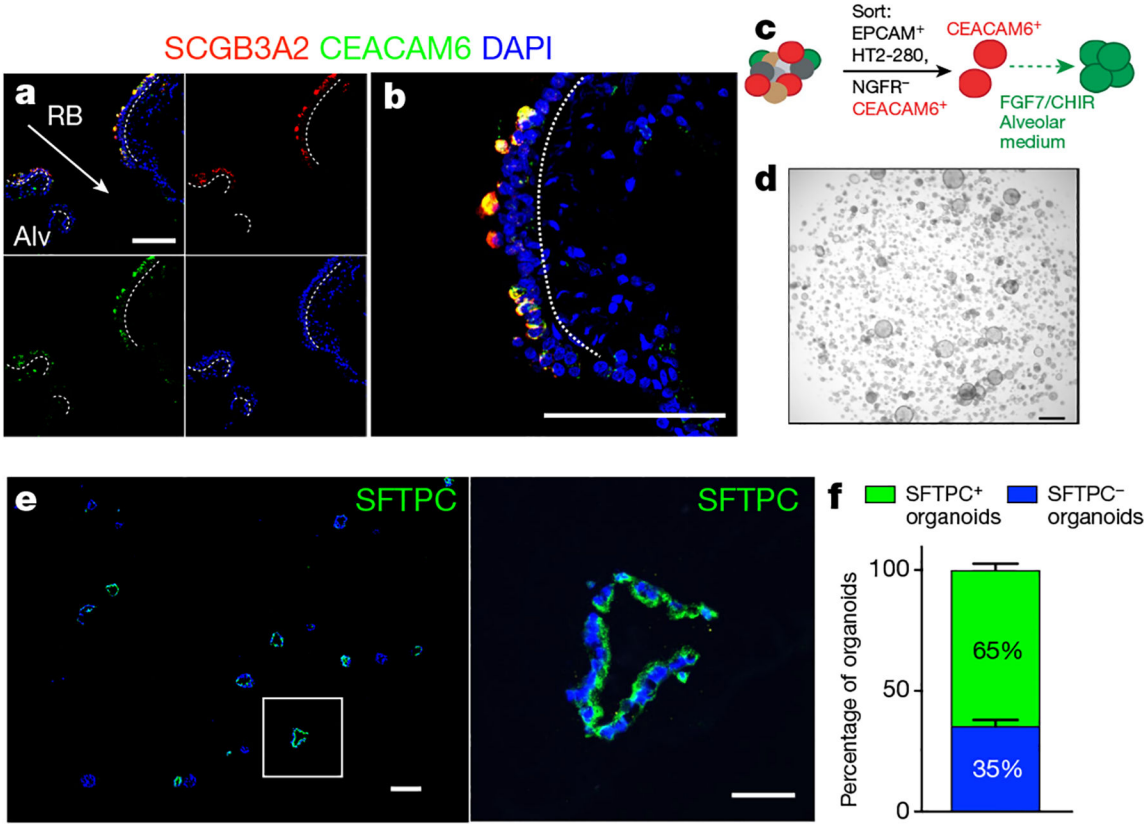


Fig. 3 |. SCGB3A2^{+} cells, identified by CEACAM6, can be isolated from distal lung parenchyma and give rise to SFTPC^{+} organoids ex vivo.

a, Immunofluorescence staining of CEACAM6 and SCGB3A2 reveals co-staining within the distal respiratory bronchioles. **b**, Enlarged image of a respiratory bronchiole from **a**. n = 5. For **a** and **b**, scale bars, 100 μm. **c**, Schematic of the CEACAM6^{+}HT2-280^{-}NGFR^{-} organoid assay. **d**, Bright-field image of CEACAM6^{+} cells grown in Matrigel in alveolar medium with FGF7 and CHIR99021. Scale bar, 1,000 μm. **e**, Whole-mount antibody staining for SFTPC expression in the resulting organoids. Scale bars, 200 μm (wide-field images), 50 μm (individual organoids). n = 3. **f**, Approximately 60% of the organoids contain SFTPC^{+} cells. Data are mean ± s.d. n = 3.

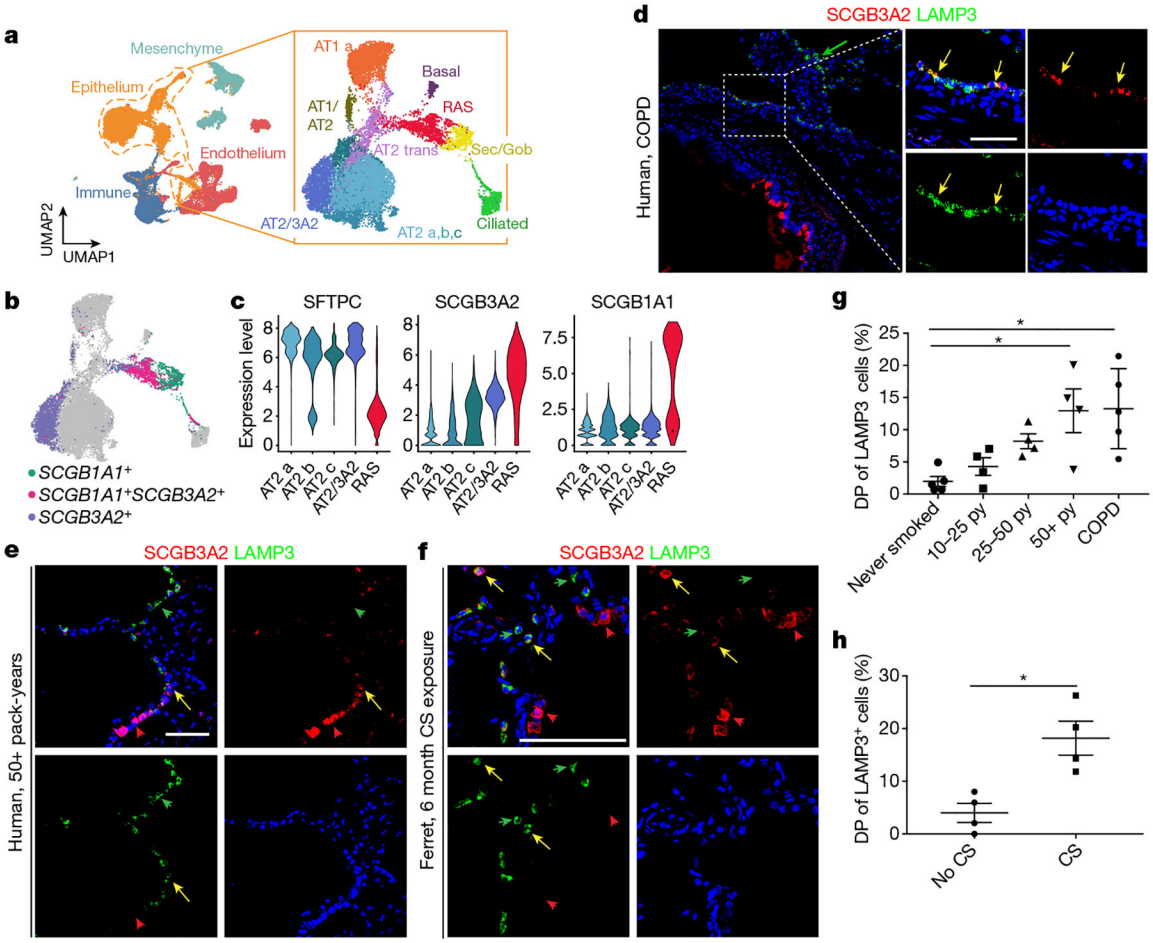


Fig. 4 |. COPD and exposure to cigarette smoke is associated with alternative AT2 cellular states.

a, A scRNA-seq analysis of distal parenchyma from patients with COPD shows all of the expected lineages, and epithelium reclustering shows all of the expected epithelial populations. **b**, RAS cells are identified by the co-expression of *SCGB1A1* and *SCGB3A2*; an AT2 population that expresses higher levels of *SCGB3A2* is noted. **c**, Key gene expression within the AT2 populations, including the *SCGB3A2*$^+$ AT2 cell population (AT2_3A2) and RAS cells. **d**, **e**, LAMP3 and SCGB3A2 expression in the peripheral parenchyma from patients with COPD ($n = 5$) (**d**) and in patients with >50 pack-year exposure to cigarette smoke (CS) but no reported history of respiratory disease ($n = 4$) (**e**). **f**, LAMP3 and SCGB3A2 expression in ferrets exposed to 6 months of cigarette smoke compared with control animals exposed to healthy air. Scale bar, 100 μm. The yellow arrows denote SCGB3A2$^+$LAMP3$^+$ cells; the green and red arrows denote single-positive cells. **g**, Quantification of LAMP3$^+$SCGB3A2$^+$ double-positive (DP) cells as a percentage of LAMP3$^+$ cells stratified by cigarette smoke pack-year (py), and patients with COPD. $n = 4$ patients per group in cohorts exposed to cigarette smoke, and $n = 5$ patients per group for non-smoking individuals and individuals with end-stage COPD. Data are mean ± s.d. Statistical analysis was performed using ordinary one-way analysis of variance (ANOVA) with correction for multiple comparisons; *$P = 0.0132$, **$P = 0.0066$. **h**, Quantification of LAMP3$^+$SCGB3A2$^+$ DP cells as a percentage of LAMP3$^+$ cells in ferrets exposed to 6

months of cigarette smoke daily or unexposed controls. $n = 4$ animals per group. Data are as mean ± s.e.m. Statistical analysis was performed using an unpaired two-tailed *t*-test; $*P = 0.0087$. Scale bars, 100 μm.