Stem Cell Reports
Article

OPEN ACCESS

Single-Cell Transcriptomic Profiling of Pluripotent Stem Cell-Derived SCGB3A2+ Airway Epithelium

Katherine B. McCauley,[1,2] Konstantinos-Dionysios Alysandratos,[1,2] Anjali Jacob,[1,2] Finn Hawkins,[1,2] Ignacio S. Caballero,[1] Marall Vedaie,[1,2] Wenli Yang,[3] Katherine J. Slovik,[3] Michael Morley,[3] Gianni Carraro,[4] Seunghyi Kook,[5] Susan H. Guttentag,[5] Barry R. Stripp,[4] Edward E. Morrisey,[3] and Darrell N. Kotton[1,2,*]

[1]Center for Regenerative Medicine of Boston University and Boston Medical Center, Boston, MA 02118, USA
[2]The Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA
[3]Penn Center for Pulmonary Biology and Institute for Regenerative Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Department of Medicine, Cedars Sinai Medical Center, Los Angeles, CA 90048, USA
[5]Department of Pediatrics, Monroe Carell Jr. Children's Hospital, Vanderbilt University, Nashville, TN 37232, USA
*Correspondence: dkotton@bu.edu
https://doi.org/10.1016/j.stemcr.2018.03.013

SUMMARY

Lung epithelial lineages have been difficult to maintain in pure form *in vitro*, and lineage-specific reporters have proven invaluable for monitoring their emergence from cultured pluripotent stem cells (PSCs). However, reporter constructs for tracking proximal airway lineages generated from PSCs have not been previously available, limiting the characterization of these cells. Here, we engineer mouse and human PSC lines carrying airway secretory lineage reporters that facilitate the tracking, purification, and profiling of this lung subtype. Through bulk and single-cell-based global transcriptomic profiling, we find PSC-derived airway secretory cells are susceptible to phenotypic plasticity exemplified by the tendency to co-express both a proximal airway secretory program as well as an alveolar type 2 cell program, which can be minimized by inhibiting endogenous Wnt signaling. Our results provide global profiles of engineered lung cell fates, a guide for improving their directed differentiation, and a human model of the developing airway.

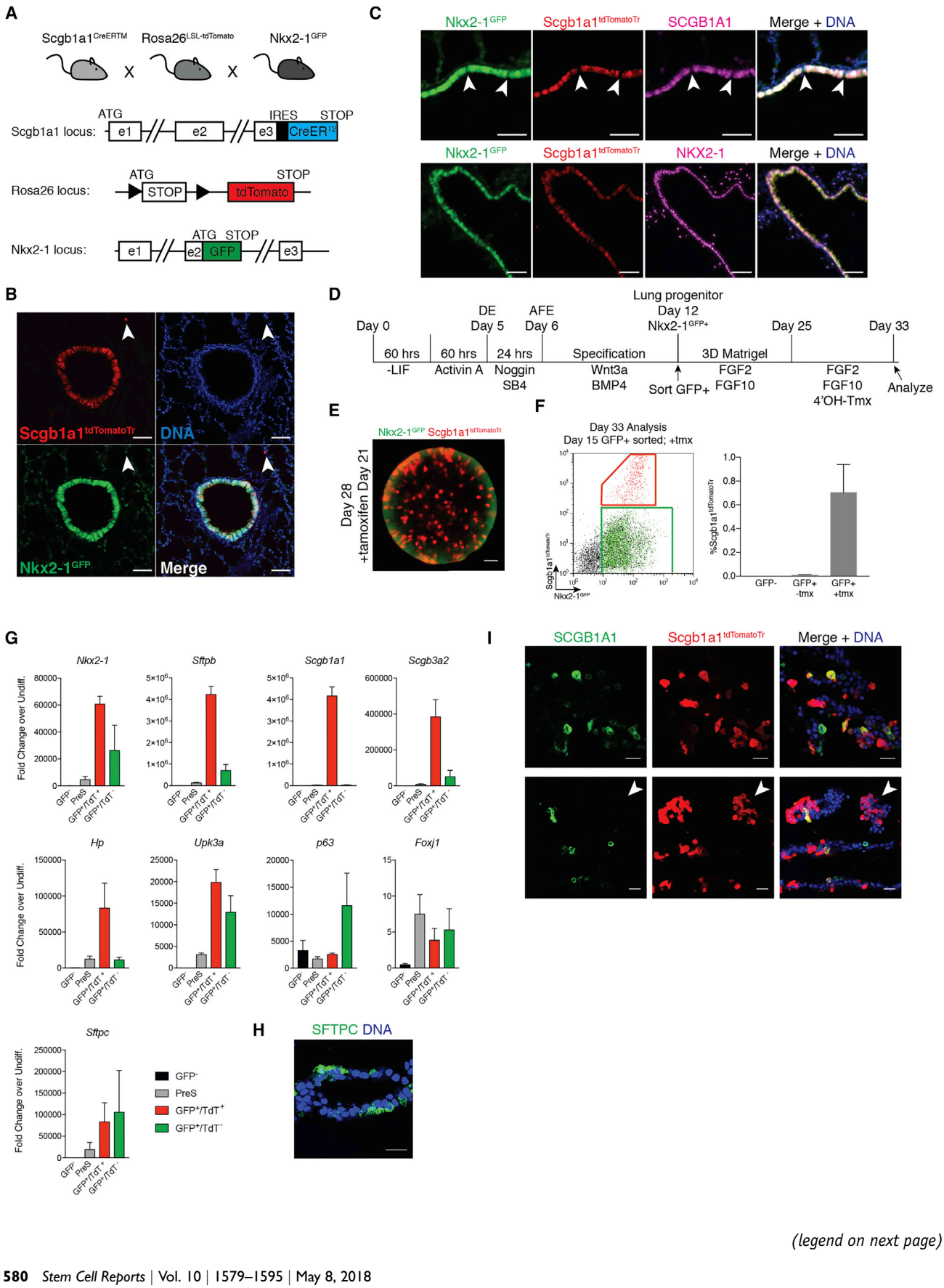
INTRODUCTION

The mature respiratory epithelium comprises two major compartments, the distal alveoli and the proximal airway, both of which include numerous functional cell types. The various cell types of the airways, including goblet, secretory, and ciliated cells, form a mucociliary escalator that protects the lung from pathogens and other inhaled agents. The lung secretory (or club) cell lineage represents an abundant and heterogeneous cell type localized to the conducting airways. These cells provide critical protective functions to the lung by participating in immune modulation, oxidative stress reduction, and xenobiotic metabolism. Club cells are also capable of differentiation to other proximal cell types, are a major known facultative progenitor for ciliated and goblet cells, and can reconstitute the airway epithelium after injury (Boers et al., 1999; Mango et al., 1998; Rawlins et al., 2009).

Differentiated lung club cells are marked by the expression of club cell secretory protein (Cc10/Ccsp/Scgb1a1). In addition to Scgb1a1, secretoglobin family members, including Scgb3a1 and Scgb3a2, are expressed in heterogeneous populations of secretory cells and are important to the function of these cells in immune homeostasis and mediation of oxidant-induced stress (Mango et al., 1998; Reynolds et al., 2002). In particular, expression of Scgb3a2 has been reported in proximal epithelial progenitors in lung development prior to maturation of the Scgb1a1+ club cell lineage, and is the earliest known marker of differentiation to this fate (Guha et al., 2012).

The *in vitro* directed differentiation of pluripotent stem cells (PSCs) via sequential regulation of developmental signaling pathways has been established as a model to study early stages of human development that are otherwise difficult to examine *in vivo*. In addition, lung epithelia generated from patient-specific induced PSCs (iPSCs) using this approach (Huang et al., 2013) have the potential to provide primary-like cells for disease modeling and drug testing. We have previously described a protocol for generating functional proximalized airway spheres expressing markers of multiple airway epithelial cell types, including basal and secretory lineages. These cells can be further differentiated to multiciliated cells in the context of Notch inhibition or air-liquid interface culture (McCauley et al., 2017).

While the expression of lineage-specific markers, including secretory cell markers *SCGB1A1* and *SCGB3A2*, as well as basal cell markers *TP63* and *KRT5*, suggest that PSC-derived airway spheres contain defined lung lineages, questions remain about the identity, heterogeneity, and long-term phenotypic stability of these cells, as well as the establishment and segregation of these separate lineages during airway sphere formation and how these findings relate to *in vivo* murine biology. The PSC model system has suggested that manipulation of key signaling pathways can regulate the sequence of lung endodermal



(legend on next page)

and proximal airway cell fate decisions during development. However, because the precise signals required to maintain these cells are not fully understood, it is likely that the airway derivatives engineered from PSCs *in vitro* may lose or drift in their phenotypes with prolonged periods in culture, as has previously been observed in primary lung epithelial cells. For airway secretory cells it may be particularly difficult to maintain a stable phenotype in culture given the known plasticity displayed by these cells when exposed to distalizing factors *in vivo* in published genetic mouse models (Zhang et al., 2008; Xi et al., 2017; Reynolds et al., 2008) or when primary murine club cells undergo even short periods of *in vitro* culture (Shannon, 1994; Tata et al., 2013; Lee et al., 2017).

Here we address these ongoing questions regarding the derivation of airway epithelial cells from PSCs in general and secretory lineages in particular. We have generated both murine and human PSC-based tools to study secretory lineage specification *in vitro*, demonstrating that both mouse and human PSC-derived airway express lineage-selective markers consistent with the known *in vivo* identity of these cells. Using a new SCGB3A2 PSC reporter system, time-series microarray, and single-cell RNA sequencing (RNA-seq) profiling in comparison with PSC-derived alveolar epithelial cells, we find that PSC-derived airway spheres contain both basal epithelial cells and SCGB3A2+ secretory airway cells. In contrast to PSC-derived distal alveolar epithelial type 2 (AEC2)-like cells and proximal basal-like cells, we find the proximal secretory lineage exhibits plasticity and is susceptible to phenotypic drift, acquiring the co-expression of both proximal secretory and distal alveolar cell programs, including the capacity to generate functional lamellar bodies that process surfactant. These results clarify the identity of the various cell types of the lung epithelium derived from PSCs via our previously described approaches, and further emphasize the utility of global transcriptomic profiling of single cells to reveal the heterogeneity, identity, and potential plasticity of emerging lineages.

RESULTS

We have previously described an approach to generate proximalized airway epithelial spheres from both human and murine pluripotent stem cells (hPSCs and mPSCs, respectively [McCauley et al., 2017; Serra et al., 2017]). We found that a low versus high level of canonical Wnt signaling was a key driver of proximal versus distal patterning, respectively, measured by the emergence of lineages expressing specific proximal and distal markers, including *Scgb1a1* and *Sftpc* (McCauley et al., 2017). Because the proximal airway contains a diversity of cell types, we here sought to derive and purify more defined subsets of airway epithelia from both mPSCs and hPSCs, beginning with airway secretory cells for which there are well established *in vivo* genetic murine reporters or lineage tracers (Rawlins et al., 2009).

Directed Differentiation of Secretory Airway Cells from Murine PSCs

To generate a bifluorescent system able to identify multiple developmental stages in airway secretory cell differentiation, we bred knockin mice carrying lineage reporters or lineage tracers targeted to gene loci known to be sequentially activated during airway differentiation: Nkx2-1^{GFP}, Rosa26$^{LSL\text{-}tdTomato}$, and Scgb1a1^{CreERTM} (hereafter Nkx2.1^{GFP};Scgb1a1^{TomatoTr}). We characterized expression patterns of these fluorochromes both *in vivo* as well in murine iPSCs (miPSCs) generated by reprogramming tail tip fibroblasts (Figures 1A and S1). In adult mice exposed to tamoxifen to induce Scgb1a1 lineage tracing, we observed Scgb1a1 lineage labeling in the vast majority of SCGB1A1 protein-expressing cells (Figures 1B and 1C), as has been

Figure 1. Generation and Directed Differentiation of Nkx2-1^{GFP};Scgb1a1$^{tdTomatoTrace\ (Tr)}$ Mouse and iPSC Lines

(A) Schematic of targeted alleles.

(B) Nkx2-1^{GFP};Scgb1a1^{tdTomatoTr} adult mouse lung post-tamoxifen exposure. The arrowheads indicate an Scgb1a1^{tdTomatoTr}-labeled cell in the alveolar compartment. Scale bars, 50 μm.

(C) Immunofluorescence stains of Nkx2-1^{GFP};Scgb1a1^{tdTomatoTr} adult mouse lungs for SCGB1A1, GFP, tdTomato, or NKX2-1 proteins. The arrowheads indicate SCGB1A1− cells. Scale bars, 50 μm.

(D) Schematic of directed differentiation of miPSCs to lung lineages. DE, definitive endoderm; AFE, anterior foregut endoderm.

(E) Nkx2-1^{GFP};Scgb1a1^{tdTomatoTr} miPSC-derived sphere at day 28 of differentiation to airway epithelium. Scale bar, 50 μm.

(F) Flow cytometry and quantification of Nkx2-1^{GFP};Scgb1a1^{tdTomatoTr} expression on day 33 of differentiation.

(G) qRT-PCR of fold change of gene expression compared with day 0 miPSCs ($2^{-\Delta\Delta Ct}$) in day 33 cells sorted for expression of Nkx2-1^{GFP} and Scgb1a1^{tdTomatoTr}. Bars represent mean ± SD (n = 3 replicates separated at day 5 of differentiation).

(H and I) Representative immunofluorescence of embedded spheres for (H) anti-SFTPC (green) and DNA (Hoechst, blue) and (I) anti-SCGB1A1 (green), tdTomato (red), and DNA (Hoechst, blue). The arrowheads in (I) indicate a tdTomato+ SCGB1A1− cluster. Scale bars, 25 μm.

See also Figure S1.

reported previously (Rawlins et al., 2009). Similarly, we confirmed co-expression of NKX2-1 nuclear protein and the cytoplasmic GFP reporter (Figure 1C). Although both secretory airway and AEC2 cells expressed Nkx2-1^{GFP}, only a minor subset of alveolar cells expressed the Scgb1a1^{tdTomatoTr} reporter (Figure 1B), as has been reported previously (Rawlins et al., 2009). In contrast, all SCGB1A1+ cells within the airway epithelium co-expressed both the GFP and tdTomato^{Tr} reporters, consistent with the expected distribution of these markers in the normal mouse lung (Figures 1B and 1C).

To test the differentiation potential of miPSCs engineered from these mice, we sequentially differentiated them via definitive endoderm and anterior foregut endoderm to Nkx2-1^{GFP+} lung progenitors using our previously reported combination of Wnt and BMP4 activation (Figure 1D) (Longmire et al., 2012; Serra et al., 2017). Quantifying lung lineage specification efficiency by flow cytometry on day 12 of differentiation, we observed Nkx2-1^{GFP} induction in 7.30% ± 0.22% of cells, corresponding to a total cell yield of 6.9 ± 2.6 Nkx2-1^{GFP+} cells per starting input miPSC (Figure 1E). Consistent with our previous publications (Hawkins et al., 2017; Longmire et al., 2012; Serra et al., 2017) the Nkx2-1^{GFP+} progenitors at this day 12 stage are relatively undifferentiated without evidence of *Scgb1a1* mRNA expression (Serra et al., 2017).

We sorted day 12 GFP+ versus GFP− cells for further differentiation (Figure 1D) in 3D culture in proximal airway medium containing fibroblast growth factor 2 (FGF2) and FGF10, previously shown to pattern cells toward airway secretory cells enriched in Scgb1a1 expression (McCauley et al., 2017; Serra et al., 2017). We added tamoxifen starting at day 25 to induce Scgb1a1 lineage tracing.

By 48 hr post-tamoxifen exposure, a subset of Nkx2-1^{GFP+}-derived cells began to co-express both Nkx2-1^{GFP} and Scgb1a1^{tdTomatoTr} (Figure 1E).

On day 33, after 8 days of tamoxifen treatment, 1.05% ± 0.30% of Nkx2-1^{GFP+} cells co-expressed Scgb1a1$^{tdTomatoTr+}$ by flow cytometry, compared with just 0.03% ± 0.02% of vehicle-only-exposed cells (Figure 1F). In contrast, we observed no expression of Scgb1a1^{tdTomatoTr} in any cell deriving from outgrowths of the Nkx2-1^{GFP-} population (Figure 1F).

We next purified Nkx2-1^{GFP+}Scgb1a1$^{tdTomatoTr-}$ and Nkx2-1^{GFP+}Scgb1a1$^{tdTomatoTr+}$ populations to ask whether lineage labeled cells were enriched for markers consistent with club cell identity. By qRT-PCR, we found that Scgb1a1$^{tdTomatoTr+}$ cells expressed *Nkx2-1* and *Sftpb* as well as known club cell genes, *Scgb1a1*, *Scgb3a2*, *Hp*, and *Upk3a* (Figure 1G). Sftpc expression at the protein level within a subset of cells was confirmed by immunostaining (Figure 1H). Immunofluorescence microscopy confirmed co-expression of Scgb1a1^{tdTomatoTr} and SCGB1A1 protein within a subset of Scgb1a1$^{tdTomatoTr+}$ cells and this pattern was present in the majority of spheres examined (Figure 1I). The presence of a subset of Scgb1a1$^{tdTomatoTr+}$ cells that lacked expression of SCGB1A1 protein suggests either the loss of club cell fate during further culture or differentiation into alternative airway lineages as has been reported in prior studies of primary club cell cultures (Tata et al., 2013).

SCGB3A2 hPSC Reporter Line Facilitates the Purification and Characterization of Human Airway Lineages

Having established that miPSC-derived lung epithelial cells in FGF2- and FGF10-supplemented media contain a population of secretory-like cells, we next sought to study human airway secretory cells derived from hPSCs using similar media and a reporter-based approach. Of the secretoglobins, in humans, SCGB3A2 is highly specific to lung tissue (gtexportal.org), and in mice has been reported to be developmentally expressed at an earlier stage than Scgb1a1 (Guha et al., 2012). Further supporting the selection of this marker, we have previously found that *SCGB3A2* expression is upregulated rapidly in the hPSC-derived proximalized epithelial population and is highly specific to the hPSC-derived lung epithelium (McCauley et al., 2017). To validate the use of SCGB3A2 as an airway epithelial marker, we performed immunostaining of human distal conducting airway tissue obtained from 18-month- and 60-year-old donors. In all sections examined, we observed high levels of SCGB3A2 protein expression selectively in the airway epithelium with no detectable signal in distal alveolar tissue (Figure 2A). For these reasons, we chose to design an SCGB3A2 fluorescent reporter to track and purify putative airway epithelial cells derived from hPSCs *in vitro*.

We used CRISPR/Cas9 technology to target a 2A-CherryPicker (membrane-anchored mCherry) fluorochrome cassette to the end of the coding sequence of the endogenous SCGB3A2 locus in normal human embryonic stem cell line RUES2 (henceforth called RUES2-SCGB3A2^{CherryPicker}, or RSC; Figures 2B and S2). We differentiated RSCs using our published differentiation strategy (Hawkins et al., 2017; McCauley et al., 2017) to the lung progenitor stage (day 15, Figure 2C), and sorted NKX2-1-enriched and -depleted populations using our established cell surface marker sorting strategy (CD47^{high}/CD26^{lo}). Lung lineage specification efficiency was measured by flow cytometry for NKX2-1 protein to be 67.4% ± 17.0% (n = 11 differentiations) for this hPSC line, and, as expected, there was no expression of the SCGB3A2^{CherryPicker} reporter at this NKX2-1+ lung progenitor developmental stage (data not shown).

To further validate our previous work identifying a proximal-specific SCGB3A2+ signal (McCauley et al., 2017),

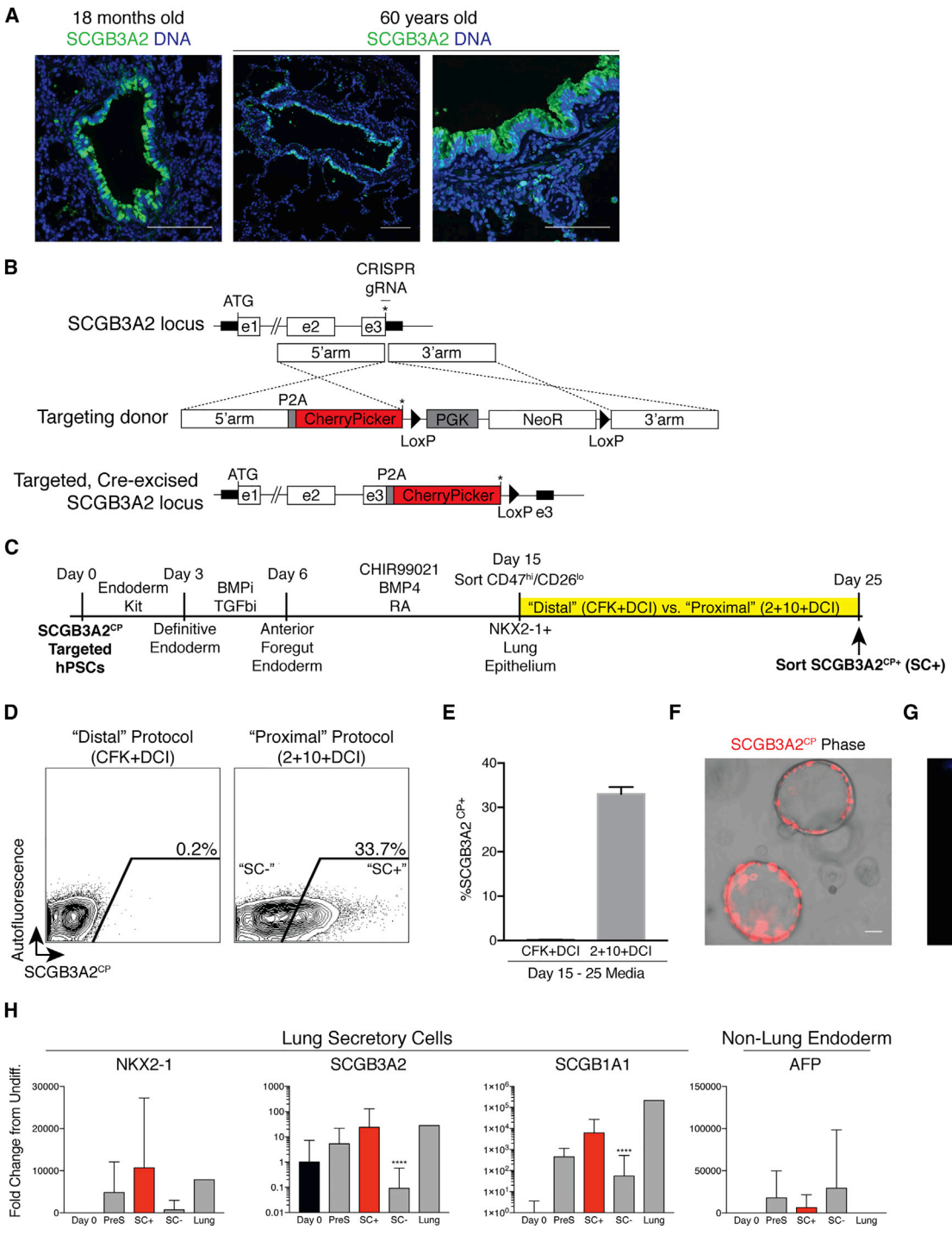


Figure 2. Generation and Characterization of hPSC SCGB3A2^{CherryPicker} Reporter Line

(A) Human distal lung sections from 18-month- and 60-year-old donors, respectively, stained with anti-SCGB3A2 (green). Scale bars, 100 μm.

(B) CRISPR/Cas9 targeting strategy for human SCGB3A2 locus. CP, CherryPicker; *, stop codon.

(C) Schematic of directed differentiation of hPSCs to lung epithelial cell types.

(D) Representative flow cytometry of SCGB3A2^{CP} expression in cells differentiated in CFK+DCI (left) or 2+10+DCI (right).

(legend continued on next page)

sorted day 15 $CD47^{high}/CD26^{lo}$ cells were replated in either our previously published proximal bronchosphere (FGF2, FGF10, and DCI; hereafter 2+10+DCI) versus distal alveolosphere (CHIR, FGF10, KGF, and DCI; hereafter; CFK+DCI) media (Jacob et al., 2017; McCauley et al., 2017). In 2+10+DCI medium, $SCGB3A2^{CherryPicker+}$ (henceforth SC) expression was found to be 33.0% ± 1.57% by day 25, whereas expression in CFK+DCI was just 0.18% ± 0.03% at this time (Figures 2D–2G). SC expression in the $CD47^{low}$ outgrowth in 2+10+DCI was just 8.16% ± 0.87%, supporting the conclusion that cells competent to become SCGB3A2+ epithelia are relatively enriched in the day 15 NKX2-1+ progenitor population and likely derive from this primordial intermediate, in keeping with our mouse PSC *in vitro* model (Figure 1). Sorted human SC+ cells from the proximal medium conditions at day 25 were enriched for expression of airway epithelial genes *SCGB3A2* and *SCGB1A1*. Notably, *NKX2-1* and *SCGB3A2* expression levels in RSC-derived SC+ cells were similar to levels measured in adult human lung biopsy tissue controls (Figure 2H).

Time Course Microarray Analysis of Differentiating Lung Epithelium Reveals Emergence of a Secretory Cell Program

Next we sought to better understand the kinetics of differentiation and the global transcriptomic programs of these SC+-labeled cells as they emerge over prolonged culture periods. Analysis of SC expression over time identified initial reporter activation by microscopy as early as 1 week after sorting and replating NKX2-1+ progenitors (day 22, Figure 3A) and established a peak of reporter expression (39.2% ± 12.5%) around day 36 of differentiation (Figure 3A). The sorted SC+ population at each of four time points between days 28 and 52 expressed both SCGB3A2 and SCGB1A1 (Figure 3B).

We next performed whole transcriptome microarray profiling of cells at various key stages of RSC differentiation to proximal airway: undifferentiated cells (day 0), purified early lung progenitors (day 15 $CD47^{hi}$), early airway progenitors (day 22), and SC+ and SC− late airway populations (day 36, Figures 3C and 3D, Table S2). Comparing day 36 SC+ with SC− cells, we found that the SC+ population was significantly enriched in expression of airway secretory cell genes, including *SCGB3A2* and *SCGB1A1* (the top two most differentially expressed genes, ranked by fold change [FC]) as well as *HP* (Figures 3E and S3A). Other upregulated genes between the SC+ and SC− populations included *SFTA3*, *SFTPA1*, *CYP1B1*, and *NKX2-1* (FDR-adjusted $p \leq 0.05$; hereafter "FDR," FC > 10; Figures 3E and S3A). These lung secretory genes were also upregulated between day 22 and day 36 SC+ (FDR ≤ 0.05; Figures 3F and S3A). We next performed pre-ranked gene set enrichment analysis comparing day 22 with day 36 SC+ and day 36 SC+ with day 36 SC− populations. Multiple pathways were upregulated in SC+ cells (FDR < 0.1) in both comparisons, including pathways known to be important for mature secretory cell function, such as secretion (Hallmark pathways: protein secretion, cholesterol homeostasis), and innate immune modulation (Hallmark pathways: TNF-α/NF-κB signaling, IL-6/STAT3 signaling; interferon-gamma response, and IL-2/STAT5 signaling; Figures S3B–S3D).

We next evaluated whether genes not typically found in developing airway secretory cells were detectable within the hPSC-derived secretory population. In the top 70 genes differentially expressed in the SC+ versus SC− population (FDR ≤ 0.05; FC > 10) we observed surfactants A1 and D (SFTPA1, SFTPD; Figure S3A), consistent with published descriptions of airway surfactants (A, B, and D) detected in both proximal and distal lung epithelia in rodents (Kalina et al., 1992). Expanding the list of differentially expressed genes (FDR ≤ 0.05; FC > 2), we found that the remainder of airway surfactants (SFTPB) were similarly enriched in SC+ cells; however, we also observed enriched expression of AEC2-selective genes, including *SFTPC, LAMP3, ABCA3, LPCAT1, PGC,* and *NAPSA* in this population (Figure 4A). Although expression of these markers in the SC+ population was validated by qRT-PCR, we found the level of expression of *SFTPC* to be significantly lower in SC+ cells than in control hPSC-derived AEC2s generated in "distalizing media" and expressing an $SFTPC^{tdTomato}$ reporter (hereafter iAEC2s [Jacob et al., 2017]). Expression of proximal

(E) Quantification of flow cytometry results from (D). n = 3 biological replicates from cells separated at day 15; similarly validated in 5 repeated differentiations.
(F) Fluorescence image of $SCGB3A2^{CP}$ expression in airway spheres. Scale bar, 50 μm.
(G) Confocal microscopy of CherryPicker ($SCGB3A2^{CP}$) cytoplasmic protein (red) in airway spheres (DNA, Hoechst; blue). Scale bar, 25 μm.
(H) qRT-PCR of fold change of gene expression compared with day 0 hPSCs ($2^{-\Delta\Delta Ct}$) in cells differentiated in each indicated condition and sorted based on expression of $SCGB3A2^{CP}$ (SC+ versus SC−). PreS, presorted/unsorted cells. Bars represent mean ± SD; $p \leq 0.01$, ****$p \leq 0.0001$. Adjusted p values obtained by ordinary one-way ANOVA with Tukey's multiple comparisons test. Asterisks indicate comparison between SC+ and other conditions (SC+, SC−, n = 9 biological replicates of independent wells from 3 independent differentiations for SCGB3A2 and SCGB1A1 and n = 6 biological replicates of independent wells from 2 independent differentiations for AFP and NKX2-1. PreS values are from n = 3 replicates of independent wells from 1 differentiation).
See also Figure S2.

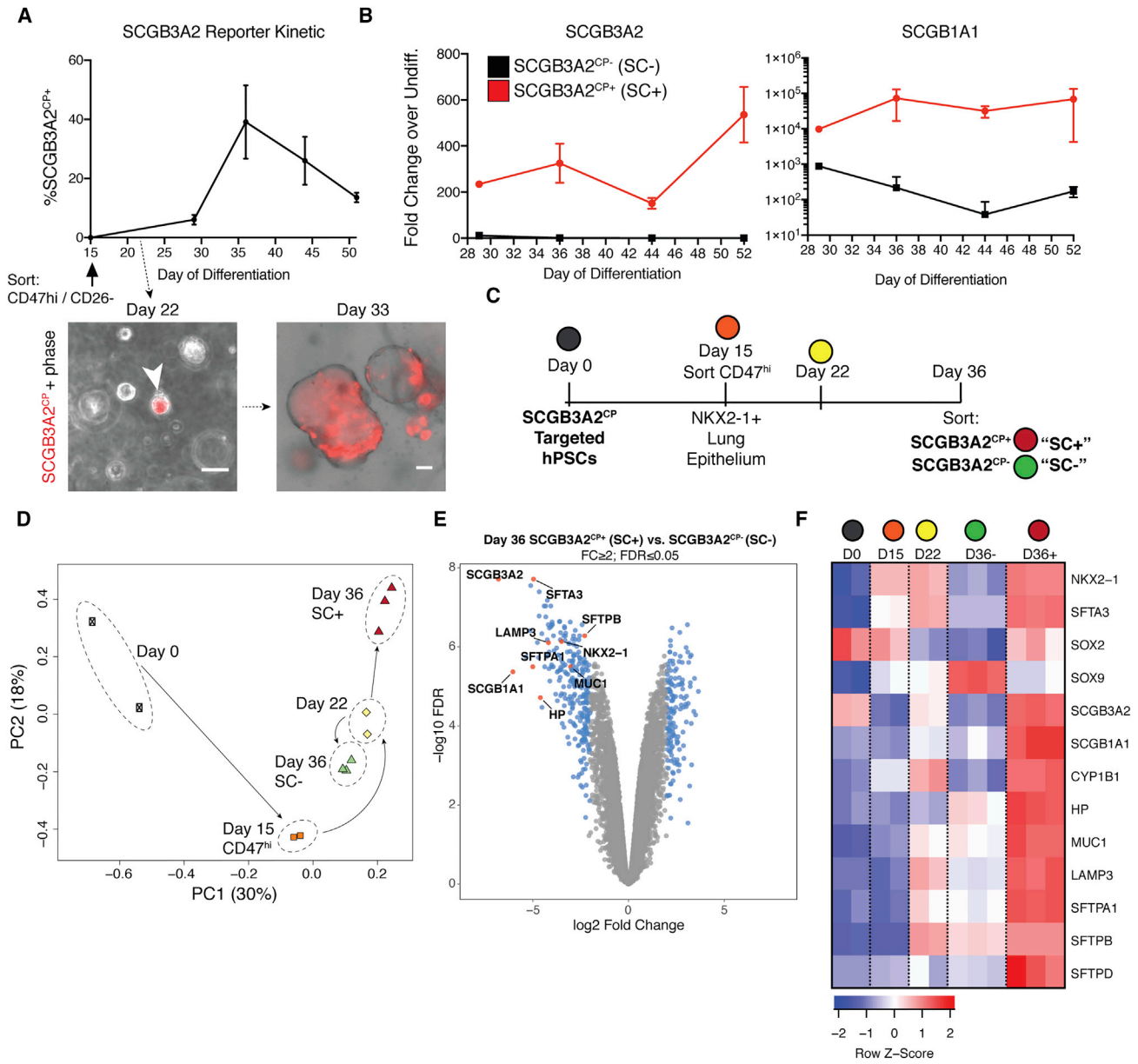


Figure 3. Kinetic Microarray Analysis Reveals Progression of Cells to an SCGB3A2+ Secretory Lineage

(A) Flow cytometry of SCGB3A2^{CP} expression over time with fluorescence microscopy of SCGB3A2^{CP} expression in live airway spheres shown for the indicated time points. Scale bars, 50 μm. The arrow indicates emerging CherryPicker+ sphere.

(B) Kinetic of gene expression (qRT-PCR; $2^{-\Delta\Delta Ct}$). Bars represent mean ± SD (day 29: n = 1; day 36: n = 3; day 44: n = 3; day 52: n = 2 replicates from wells separated after day 15 sort and then sorted for SC+ versus SC− at each indicated time point for qRT-PCR).

(C) Schematic of microarray experiment for (D–F).

(D) Principle component analysis of global transcriptomes from n = 3 independent differentiations (biological replicates) of microarray samples indicated in (C).

(E) Volcano plot of differential gene expression between day 36 SC+ and SC− cells. Indicated genes are a selected subset of those meeting indicated cutoff criteria by pairwise Student's t test.

(F) Heatmap representing relative gene expression over time of selected transcripts related to lung development (FDR ≤ 0.01).

See also Figure S3.

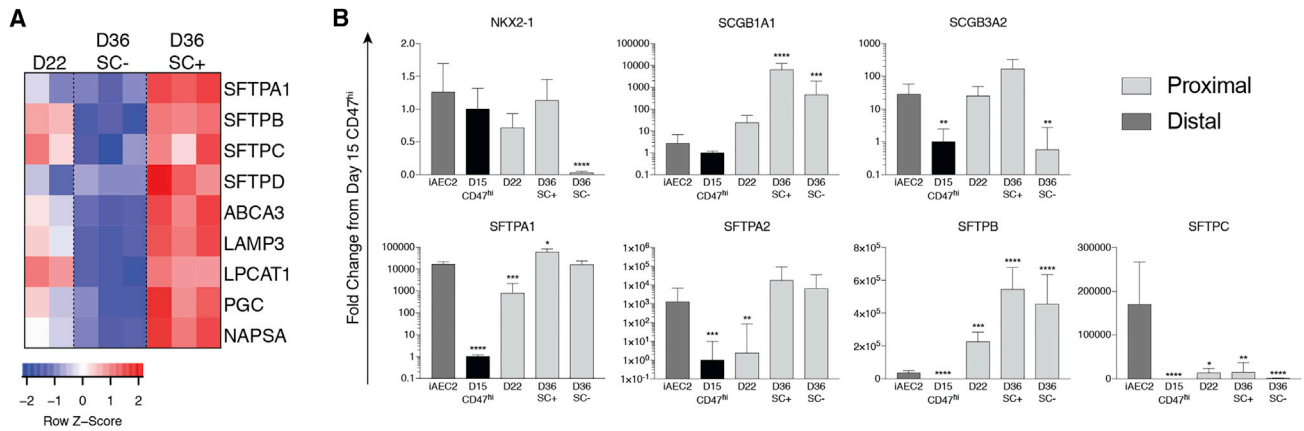


Figure 4. Expression of Proximal and Distal Lung Epithelial Marker Genes in the hPSC-Derived Airway Epithelial Cell Population

(A) Heatmap of AEC2 cell marker genes from microarray analysis across indicated time points (FDR $\leq$ 0.01).

(B) Gene expression (qRT-PCR; $2^{-\Delta\Delta Ct}$) in cells differentiated toward airway epitheilum and sorted for expression of SCGB3A2^{CP} (SC+) or differentiated toward alveolar epithelium and sorted for expression of SFTPC^{tdTomato} (iAEC2s). D, time of differentiation in days. Bars represent mean ± SD, n = 3 biological replicates from independent differentiations. *p $\leq$ 0.05, **p $\leq$ 0.01, ***p $\leq$ 0.001, ****p $\leq$ 0.0001. Adjusted p values obtained by ordinary one-way ANOVA with Tukey's multiple comparisons test. Asterisks indicate comparison between iAEC2s and all other conditions.

marker *SCGB1A1* was significantly higher in the SC+ population compared with iAEC2s (Figure 4B; Table S1).

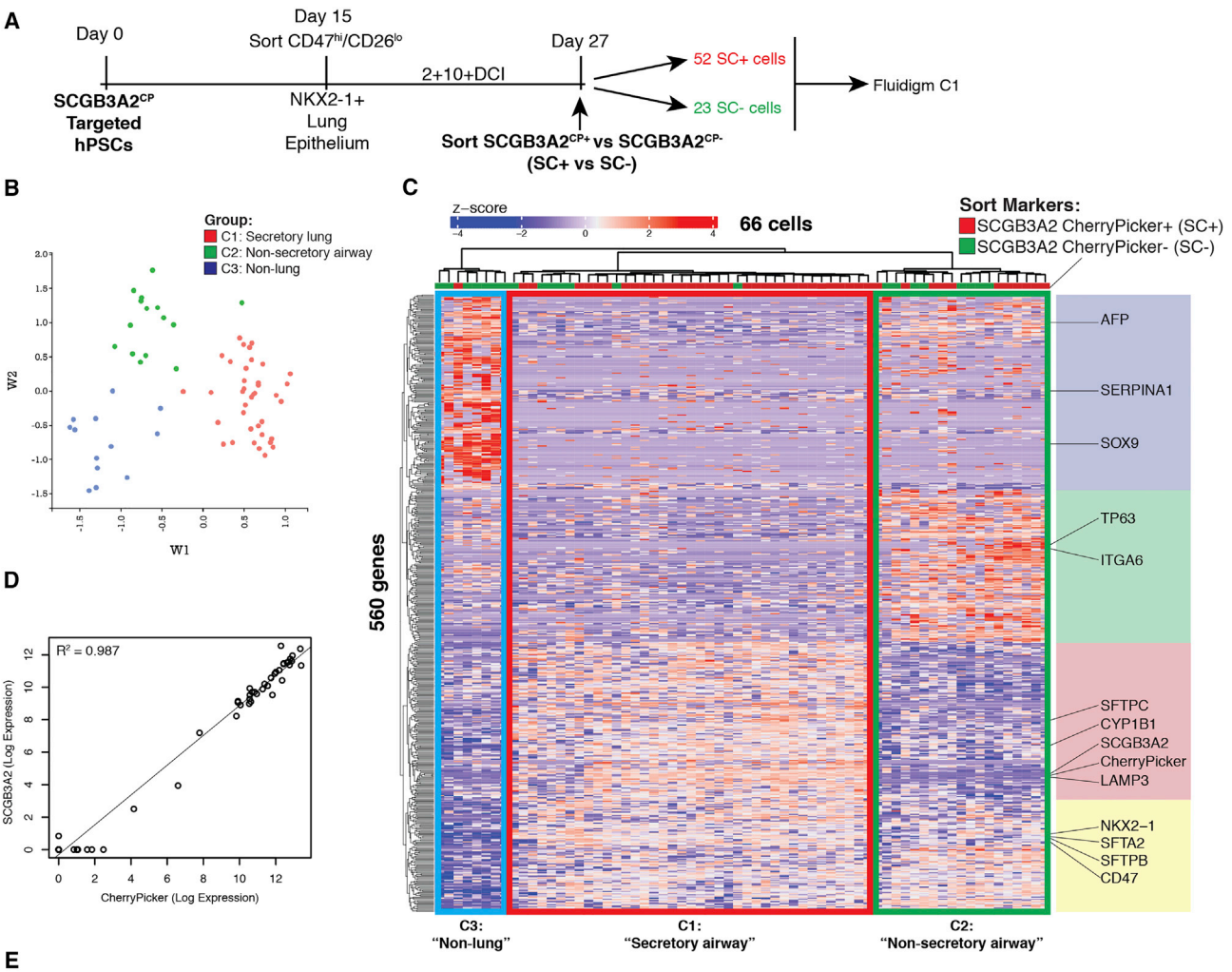
Single-Cell RNA Sequencing Identifies Cell-Type-Specific Clusters in hPSC-Derived Airway Epithelium

We next aimed to determine whether in our "proximal" conditions *SFTPC* and other AEC2 markers were being expressed in rare distal cells within the heterogeneous SC+ population, or the alternative possibility that a distal program might be co-expressed within cells that also express proximal lung markers. To distinguish these potential explanations, we used single-cell mRNA sequencing (RNA-seq) to elucidate the global transcriptomes of hPSC-derived putative airway cells on a per-cell basis. We differentiated hPSCs (RSC line) into airway spheres and purified SC+ and SC− populations at day 27 of differentiation for single-cell RNA-seq (Figure 5A), analyzing 52 SC+ and 24 SC− cells, of which 44 SC+ and 22 SC− passed quality control. Dimensionality reduction (ZINB-WaVE) and hierarchical clustering revealed three distinct cell clusters (Table S3; Figures 5B and 5C). The majority of the SC+-sorted cells clustered together and expressed high levels of both *SCGB3A2* and *CherryPicker* mRNA (correlation R^2 = 0.987, Figure 5D). We therefore categorized the three cell clusters as secretory airway (C1), non-secretory airway (C2), and non-lung (C3), based on expression of known marker genes within these gene clusters (FDR < 0.05), including *SCGB3A2* and *SCGB1A1* (secretory; enriched in cell cluster C1); *TP63* and *ITGA6* (basal; enriched in C2); *NKX2-1*, *SFTA3*, *SFTA2*, *SFTPB*, and *CD47* (enriched in both lung cell clusters: C1 and C2); and *SERPINA1*, *SOX9*, *FGB*, *APOA2*, and *AFP* (liver; enriched in C3) (Figure 5C). C3 contained 8 out of 66 total cells, suggesting that 12.1% of the analyzed cells were non-lung. Notably, only a subset of *SCGB3A2* mRNA+ cells co-expressed the more mature club cell marker *SCGB1A1* (10 out of 35 cells, Figures 5E and S4). A distinct subset of SC+-sorted cells clustered within C2, the non-secretory airway population, expressed high levels of *NKX2-1, TP63,* and *ITGA6*, but low levels of secretory markers, indicating a putative basal cell-like identity (Figure 5E). Despite expressing CherryPicker protein, these TP63+ cells were predominantly SCGB3A2− and did not express CherryPicker mRNA (Figures 5E, S4B, and S4C), suggesting these cells had recently lost expression of the reporter as they acquired a basal cell-like fate.

We next investigated whether any hPSC-derived airway cells co-expressed proximal and distal lung markers, including *SFTPC*. We found that expression of *SFPTC*, along with that of *LPCAT1, LAMP3*, and *NAPSA*, was upregulated in cluster C1 (Figures 5E and S4D) and that *SFTPC* was expressed predominantly in SCGB3A2+ cells (25 out of 31 cells, Figure S4E). Cells in this cluster expressed low levels of distal patterning marker *SOX9* (Figure 5E). In contrast, the basal cell-like population in cluster 2 was not enriched in expression of *SFTPC* or other AEC2-selective genes (Figure 5E).

Single-Cell RNA-Seq of Airway versus Alveolar Spheres Reveals Heterogeneity and Distinctly Patterned Lineages

Having established that *SFTPC* is co-expressed with airway markers in hPSC-derived airway epithelium, we



(legend on next page)

simultaneously derived proximal airway epithelial cells and distal iAEC2s from hPSCs in parallel and profiled these lineages by single-cell RNA-seq of larger numbers of cells after a prolonged period of culture. We differentiated our published bifluorescent iPSC line (BU3) engineered to carry GFP and tdTomato reporters targeted to the NKX2-1 and SFTPC loci, respectively (hereafter BU3 NGST [Jacob et al., 2017]). After 15 days of differentiation into primordial lung progenitors (NKX2-1^{GFP+}/SFTPC$^{tdTomato-}$), we sorted GFP+/tdTomato− cells for further differentiation in parallel in either our published proximal/airway (McCauley et al., 2017) or distal/alveolar (Jacob et al., 2017) 3D culture conditions (Figure 6A). To explore the effect of extended time in culture, we differentiated the sorted GFP+ cells until day 41 to form putative airway or alveolar spheres, passaged each condition one time on day 26 (alveolar) or day 29 (airway), and analyzed 1,392 cells that passed data quality control.

Initial k-means clustering analysis and dimensionality-reduction visualization represented as tSNE plots demonstrated the separation between the airway and iAEC2 medium-derived cells and identified four distal (C1, 2, 3, and 7) and three proximal (C4–6) clusters (Figures 6B, 6C, and S5A). We first confirmed the predicted medium effects by evaluating expression of Wnt target gene *AXIN2*, which was higher in the distal cells cultured in medium containing the Wnt signaling stimulator, CHIR, than in the proximal cells (Figure 6D). We next asked whether any of the identified clusters represented non-lung populations. Mapping of expression of *NKX2-1* revealed that one proximal (C4) and one distal (C3) cluster were NKX2-1^{low}. We found that 234 out of 673 clustered proximal cells (34.8%) and 131 out of 719 clustered distal cells (18.2%) were represented in C4 and C3, respectively, and therefore we estimate these to be the proportion of non-lung cells obtained in each of these conditions in this differentiation. Further statistical analysis of genes differentially expressed in each cluster identified these putative non-lung clusters as gastric epithelium (C3) expressing *GIF*, *TFF1*, *TFF2*, and *CLDN18* in the distal condition and hepatic endoderm (C4) expressing a constellation of known liver genes including *AFP*, *APOA2*, *TTR*, *TF VTN*, *ALB*, and *FGB* in the proximal conditions (Figures 6D and 6E; Table S4), consistent with prior observations (Hawkins et al., 2017; McCauley et al., 2017). A subset of these hepatic-like cells had high levels of proliferation markers (*AURKB*, *MKI67*, and *TOP2A*), suggesting that a subset of this population is mitotic (Figures 6D and S5B; Table S4).

To identify potential lung lineages, we next interrogated all differentially expressed genes (Table S4) and supervised clustered heat maps of the top 50 genes enriched in each cell cluster compared with all cells (Figure 6E). Marker genes of known lung lineages were differentially enriched in identifiable clusters, allowing us to validate our assignments of cell-type identity for the remaining NKX2-1+ clusters. For example, the distal NKX2-1+ clusters C1 and C2 contained AEC2-like cells expressing *SFTPC* (Figure 6D; Table S4) along with other distal lung epithelial markers including *FGF9, BMP4*, *BMP3*, and *FOXP2* (Figure 6E; Table S4), and were depleted of markers of proximal epithelium including *SCGB3A2* and *P63* (Figure 6D). Despite these similarities, we found that C2 had higher expression of *SFTPC* (FC = 1.81, p_{adj} = 0.077) relative to all other clusters, including C1. We therefore asked what distinguished C1 and C7 from C2. We found that C1 expressed high levels of proliferation markers (*AURKB*, *MKI67*, and *TOP2A*), suggesting that this represents a mitotic AEC2-like population (Figures 6D, 6E, and S5B; Table S4). The final NKX2-1+ distal cluster (C7) also had an AEC2-like phenotype and similar gene expression to C1 and C2, but expressed lower levels of key genes including *SFTPC* and *FOXP2* (Figures 6D and 6E; Table S4).

We next examined the NKX2-1+ proximal clusters (C5 and 6). Consistent with our prior single-cell analysis (Figure 5), we found that cluster C5 represented a basal cell population expressing *P63*, *KRT5*, *KRT14*, *KRT17*, and *PAX9* (Figures 6D and 6E; Table S4), and cluster C6 represented an airway secretory cell population enriched in expression of known secretory markers, *SCGB3A2*, *SCGB1A1*, *HP*, *SFTA2*, and *AQP4* (Figures 6D and 6E; Table S4) (Guha et al., 2014; Kalina et al., 1992; Kreda et al., 2001). We similarly identified subpopulations of

Figure 5. Single-Cell RNA-Seq of hPSC-Derived Airway Epithelial Spheres Reveals Subsets of Secretory and Basal-like Cells Expressing Distinct Markers

(A) Experiment schematic.

(B) Dimensionality reduction plot (ZINB-WaVE) of global transcriptomes of the top 1,000 most variable genes across 66 individual cells. Indicated cell groups were identified by k-means clustering analysis.

(C) Row-normalized heatmap of global gene expression using unsupervised hierarchical clustering of all 66 cells based on the 560 genes that passed the thresholds (variance > 3 and ANOVA FDR < 0.05) indicated in (B).

(D) Correlation analysis between the log-expression of *SCGB3A2* and *CherryPicker* mRNA in all 66 cells.

(E) Violin plots showing normalized expression for indicated genes across cell clusters. Colored points indicate sort markers used (red, SC+; green, SC−); violin color indicates assigned cell cluster (red, C1; green, C2; blue, C3).

See also Figure S4.

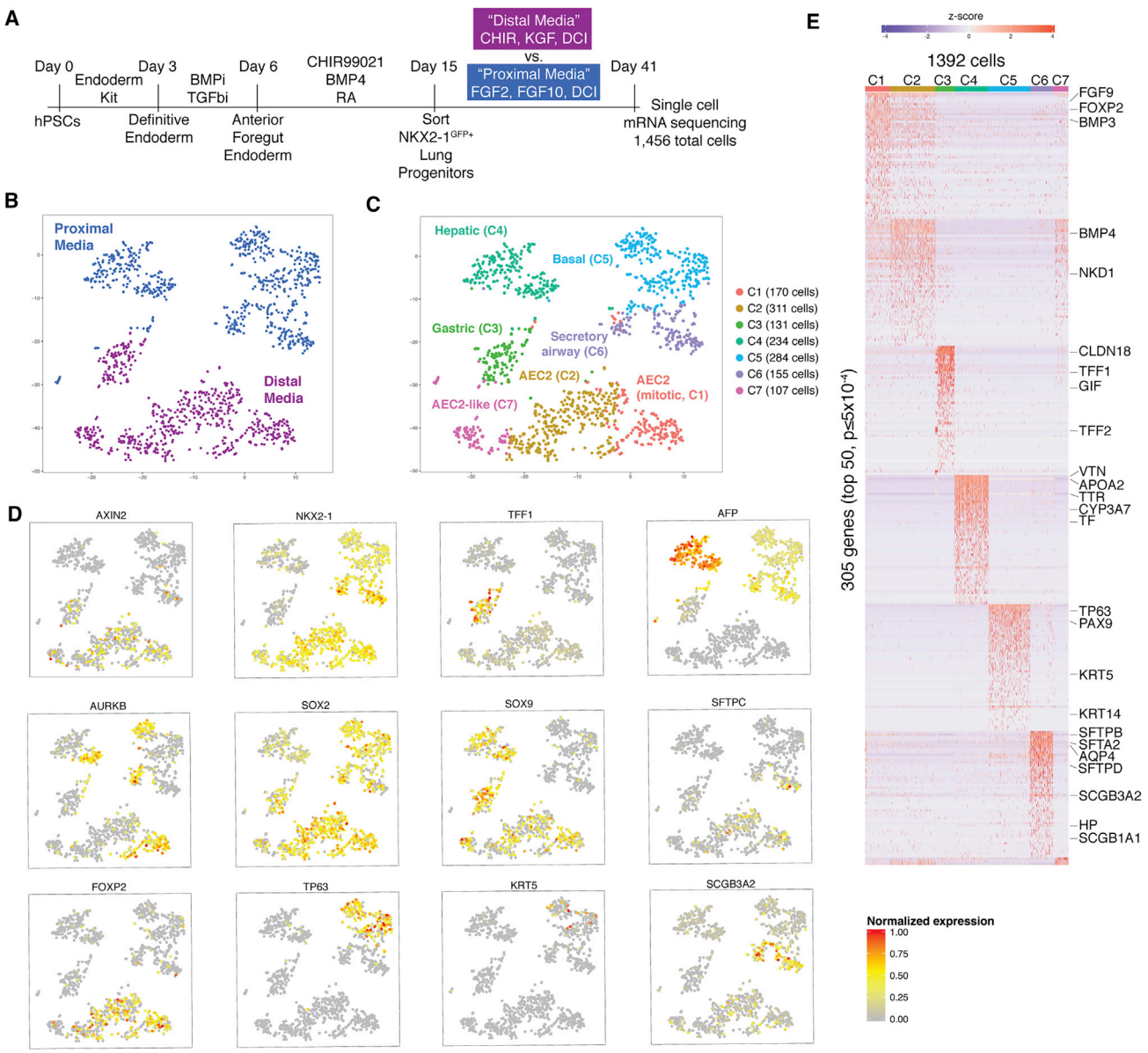


Figure 6. Single-Cell Transcriptomic Comparison of hPSC-Derived Airway and Alveolar Epithelial Cells

(A) Experimental schematic.

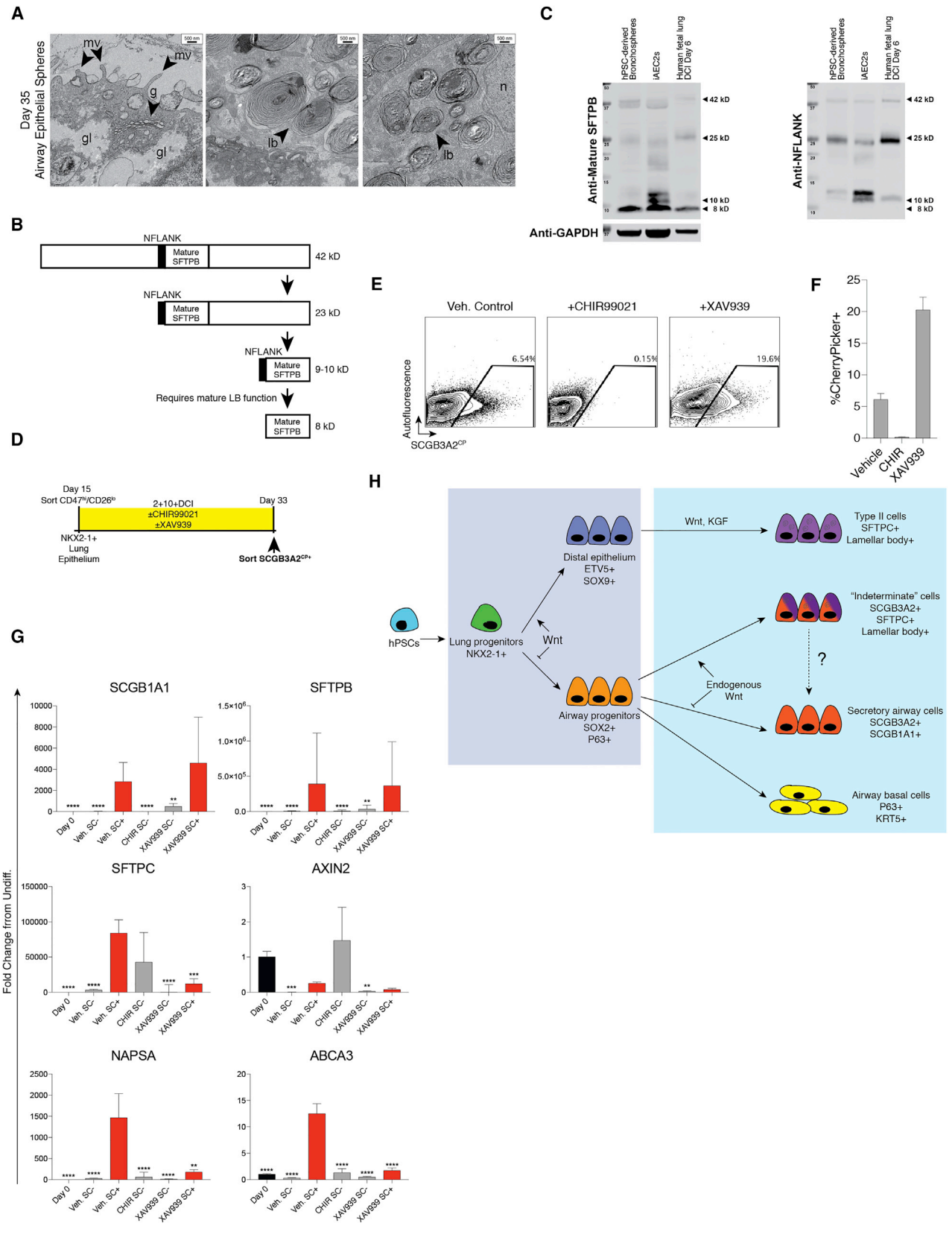
(B–D) tSNE plots of 1,392 cells with (B) culture medium indicated; (C) lineage names assigned to 7 clusters identified by k-means; or (D) overlayed normalized expression of indicated marker genes.

(E) Heatmap of the top 50 differentially expressed genes in each of 7 cell clusters compared with all cells by negative binomial exact test (ranked by FC; FDR < 5×10^{-4}).

See also Figures S5 and S6.

mitotic secretory and basal cells enriched for *AURKB*, *MKI67*, and *TOP2A* (Figures 6D and S5C; Table S4). To exclude the effect of cell-cycle genes on our statistical clustering analyses, we repeated all analyses after removing 280 known cell-cycle genes (Table S5) from the dataset, and found minimal effects on the clustering and no alteration in the cellular identities of the identified clusters (Figures S5C and S5D). We therefore continued all analyses without excluding these genes (Figure 6).

Similar to our prior 66 cell analysis (Figure 5), the airway secretory population shown in Figure 6 again expressed all 5 surfactants (*SFTPA1*, *SFTPA2*, *SFTPB*, *SFTPC*, and *SFTPD*) along with other AEC2-selective transcripts that encode lamellar body genes, including *NAPSA* and *PGC*



(legend on next page)

(Figures 6D and 6E; Table S4). Correlation analysis within the proximal epithelial population further established that 63 out of 146 of the *SFTPC*+ cells within this population also expressed *SCGB3A2*, but only 1 out of 146 expressed *SCGB1A1* (Figure S5E). A direct comparison of iAEC2 with airway secretory gene expression revealed that the majority of the iAEC2 gene expression signature was depleted in the secretory airway population (Figures 6E and S5A; Table S4). In particular, the airway secretory cells expressed lower levels of distal patterning genes, including *SOX9*, *BMP4*, *FOXP2*, *FGF9*, and *ETV5*, than the distal clusters, further demonstrating that the secretory airway cells exhibit an indeterminate phenotype with features of both proximally patterned airway epithelium along with some AEC2 genes, but without markers of early distal patterning.

Finally, we also found that, in spite of the distinct expression of known marker genes and associated transcriptomic signatures indicative of particular known cell types, some genes were also expressed at low levels in unpredicted populations, such as expression of *AFP* and *ALB* in *NKX2-1*+ putative airway epithelial cells (Figures 6D and 6E), when these three markers are not known to be co-expressed *in vivo*. We confirmed the expression of liver markers within the SCGB3A2^{CP+} population by qRT-PCR (data not shown).

A Subset of hPSC-Derived Airway Epithelial Cells Contain Functional Lamellar Bodies

Having identified a population of airway epithelial cells expressing some AEC2-associated genes but lacking a complete AEC2 signature, we next sought to test whether these cells exhibited functional and ultrastructural features consistent with AEC2s. Electron microscopy of plastic sections of hPSC-derived differentiated airway spheres revealed that spheres were heterogeneous and contained a mixed population of cells, some of which exhibited characteristics of airway secretory lineages. These included electron-dense granules and abundant rough endoplasmic reticulum, as well as abundant glycogen and microvilli with associated secretory bodies on the apical surface of some epithelial spheres (Figure 7A). All examined cells lacked any evidence of goblet cell maturation or ciliation. Together, these features are characteristic of immature, differentiating human airway secretory cells (Plopper et al., 1980), and support the conclusion that hPSC-derived airway spheres contain secretory cell types.

As predicted by single-cell RNA-seq, we also found that some cells generated in proximal medium contained structures reminiscent of surfactant-producing lamellar bodies (Figure 7A), with similar features to those we have previously reported for hPSC-derived alveolospheres (Jacob et al., 2017). To test the function of these putative lamellar bodies, and given the high expression of surfactant proteins and the SFTPB processing enzymes *PGC* and *NAPSA* in these cells, we tested whether airway secretory cells were able to process SFTPB pro-protein (pro-SFTPB) to its mature 8 kDa form, a capacity unique to lamellar body-expressing AEC2s *in vivo* following 24 weeks of human gestation (Brasch et al., 2004; Guttentag et al., 1998). By western blot analyses, hPSC-derived airway spheres were capable of processing SFTPB to its mature 8 kDa form (Figures 7B and 7C); however, in comparison with hPSC-derived iAEC2s, when stained with an antibody specific to pro-SFTPB (anti-NFlank) these cells had relatively higher accumulation of the unprocessed 25 kDa form, known to be expressed in airway cells *in vivo* (Kalina et al., 1992) (Figure 7C). This pattern suggested that hPSC-derived airway spheres may contain a heterogeneous mixture of cells, many of which produce pro-SFTPB, but only some of which are able to fully process the pro-protein to its mature form (Figures 7B and 7C), a conclusion supported by the presence of at least two morphological subsets of cells present in airway spheres analyzed by electron microscopy (Figure 7A and data not shown).

Figure 7. Co-expression of Distal Programs in hPSC-Derived Airway Cells Is Reduced by Wnt Inhibition

(A) Representative electron micrograph of ultra-thin sections of epithelial spheres differentiated in proximal medium with characteristic lamellar body structures. gl, glycogen; lb, lamellar bodies; mv, microvilli; n, nucleus; g, Golgi apparatus.

(B) Schematic of surfactant protein B processing in functional lamellar bodies.

(C) Western blots of hPSC-derived airway (bronchospheres) and alveolar spheres (iAEC2s), generated in proximal versus distal medium, respectively; immunostained for total SFTPB (left) and precursor SFTPB protein with retained N-flank (right).

(D) Schematic of experiment testing Wnt inhibition during airway sphere formation.

(E and F) Representative flow cytometry and quantification (E) of day 33 airway spheres treated with vehicle, CHIR, or XAV939 from day 15–33 of differentiation (F). Mean ± SD; n = 3 replicates of independent wells separated at day 15.

(G) Gene expression (qRT-PCR; $2^{-\Delta\Delta Ct}$) in cells differentiated under indicated conditions and sorted based on expression of SCGB3A2^{CP} (SC+ versus SC−). (F and G) Bars represent mean ± SD (n = 3 replicates of independent wells separated at day 15, representative of 2 repeated differentiations on independent PSC lines). **p ≤ 0.01, ***p ≤ 0.001, ****p ≤ 0.0001. Adjusted p values obtained by ordinary one-way ANOVA with Tukey's multiple comparisons test. Asterisks indicate comparison between vehicle-treated SC+ and all other conditions.

(H) Schematic of cell fate decisions in directed differentiation of lung epithelium.

See also Figure S7.

Distal Plasticity of hPSC-Derived Secretory Cells Is Promoted by Exposure to Endogenous and Exogenous Distalizing Soluble Factors

We considered the possibility that expression of distal markers within the proximal secretory population might result from the presence of unappreciated distalizing factors acting over prolonged culture periods. Building on previous work identifying Wnt as a key driver of distal patterning both immediately post-specification (McCauley et al., 2017) as well as in the context of differentiation of AEC2-like cells from primary adult mouse secretory cells in culture (Lee et al., 2017), we sought to test whether endogenous Wnt signaling might play a role in the aberrant expression of *SFTPC* within hPSC-derived airway epithelium.

To test the effect of Wnt on the expression of AEC2 marker genes in proximalized spheres, we tested whether spheres differentiated in proximal medium responded to Wnt activation or inhibition with CHIR or the Wnt inhibitor XAV939, respectively (Figures 7D and 7E). Compared with proximal/airway control medium, Wnt inhibition resulted in maintenance of *SCGB1A1* and *SFTPB* levels as well as marked reduction in AEC2 and lamellar body markers, *SFTPC*, *NAPSA*, and *ABCA3*, within sorted SC+ cells (Figure 7G; Table S1). As expected, CHIR treatment resulted in loss of the SCGB3A2^{CP+} program (Figures 7E–7G) and increased *SOX9*, consistent with canonical Wnt-driven distalization (Figure 7G; Table S1 and data not shown). Lack of cell sorting in this CHIR-containing condition precluded equivalent profiling of the entire iAEC2 program. To determine whether the putative plasticity of proximalized cells in these conditions was limited to a narrow developmental window we repeated this experiment with delayed CHIR addition to the proximalized cells. We found that CHIR treatment, even up to 10 days post-replating in airway medium, resulted in upregulation of SFTPC^{tdTomato} expression (Figure S7), a reporter we have previously shown to be reflective of higher levels of *SFTPC* mRNA expression (Jacob et al., 2017).

This response to CHIR treatment indicates that the putative plasticity of the proximalized cells, measured by changes in SFTPC reporter expression, was sustained for at least 25 total days of differentiation time. In comparison, cells cultured in the absence of CHIR had little to no detectable SFTPC^{tdTomato} expression by microscopy (Figure S7).

DISCUSSION

Our results provide a thorough characterization of hPSC-derived putative airway and alveolar lineages, with a focus on the understudied secretory airway population, using both murine and human lineage reporters. Through single-cell RNA-seq accompanied by ultrastructural and functional surfactant processing studies, our findings emphasize the heterogeneity of lung and foregut endodermal lineages that emerge over time in culture and demonstrate the relative plasticity versus stability of different proximal and distal putative lung lineages derived from PSCs in these conditions.

We find iAEC2s generated in distal medium display overall long-term stability of lineage without phenotypic drift in their transcriptomic programs toward either proximal airway or AEC1 phenotypes. In proximal medium, the putative airway basal cell population also displays a program that is devoid of alternate lung lineages, although low levels of expression of hepatic transcripts can be detected in this population. In marked contrast, airway secretory cells generated from PSCs display lineage plasticity with a tendency to upregulate a Wnt-dependent distal lung program, leading to co-expression of both proximal-secretory and AEC2-like programs. This culminates in the assembly of functional lamellar bodies able to process surfactant, a feature not present in unperturbed primary secretory cells *in vivo* (Figure 7H).

Plasticity in epithelial identity within the proximal lung epithelium has previously been described in the trachea and airways of the mouse during development (Shannon et al., 1998), in settings of forced over-stimulation of Wnt signaling (Zhang et al., 2008; Reynolds et al., 2008), and in the context of epithelial injury and repair (Xi et al., 2017). Particularly relevant to our findings is the recent observation by Lee et al. (2017) that adult murine airway secretory (club) cells placed in co-culture with Wnt-activating mesenchymal cells similarly exhibit upregulation of a distal program within these club cells. Importantly, recent work studying patients with idiopathic pulmonary fibrosis has also identified aberrant proximal and distal transcriptomic programs co-expressed in individual epithelial cells (Xu et al., 2016), suggesting that this phenomenon may have direct relevance to abnormalities in cell fate observed in complex diseases of the lung epithelium. We hypothesize that PSC-derived airway secretory cells similarly represent an *in vitro* proximalized correlate population that is plastic and tends to phenotypically drift toward distal epithelium under specific signaling conditions.

An important question of semantics is whether to assign airway identity to our hPSC-derived SCGB3A2+ cells by referring to them as "plastic secretory airway cells" as opposed to identifying them as "indeterminate cells" given their co-expression of several AEC2 markers. Our single-cell global transcriptomic profiling begins to address this question by establishing that *SCGB3A2+SFTPC+* cells cluster together with putative "mature" secretory airway cells that are *SCGB1A1+/SFTPC−*, express other club cell

markers, and are distinct from the separately clustered iAEC2 population, despite mutual expression of *SFTPC* and lamellar body genes. Although these RNA-seq profiles represent limited temporal snapshots, they support the conclusion that these cells have an underlying secretory airway identity with aberrant upregulation of a functional alveolar epithelial program.

What are the factors responsible for the emergence of a co-expressed distal program within the secretory airway population in our studies or in the above *in vivo* scenarios? Our current study and prior published work suggests that the presence of distalizing factors influences core transcriptional programs and maturation programs in proximal airway cells during the period following lung lineage specification. Even mature secretory cells, such as adult club cells, following naphthalene injury (Reynolds et al., 2008) or when placed in culture (Lee et al., 2017), are susceptible to upregulation of AEC2 programs in the presence of high levels of canonical Wnt signaling. One candidate pathway that may similarly influence the fate of the developing secretory lineage in our PSC model is endogenous Wnt signaling, since blocking Wnt appears to markedly reduce the AEC2 program within the SCGB3A2^{CP+} population.

Two non-lung foregut endodermal subsets have reproducibly emerged from the sorted day 15 population enriched in NKX2-1+ cells: hepatic-like and gastric-like cells that are NKX2-1− (Hawkins et al., 2017). This raises a number of questions that will require further study to definitely address the origins of these cells. It is possible that gastric and hepatic cells simply arise from low levels of NKX2-1− cells present in the NKX2-1+ sorted outgrowth; however, a more interesting possibility is the likelihood that gastric progenitors aberrantly expressing NKX2-1 are generated through the high levels of Wnt signaling stimulated by CHIR during lineage specification, analogous to mouse genetic models, where forced *in vivo* overexpression of β-catenin results in the generation of Nkx2-1+ cells in the anterior stomach (Goss et al., 2009; Harris-Johnson et al., 2009). Notably, the gastric-like cells produced in our experiments emerge in distalizing conditions with sustained CHIR, consistent with previous reports that emergence of the gastric lineage is promoted by Wnt activation (McCracken et al., 2017).

Taken together, our results emphasize the need for sorting strategies to isolate defined lineages emerging within heterogeneous populations over developmental time. In addition, our findings demonstrate the need for global profiling approaches to completely understand the genetic programs of cells derived via directed differentiation, as cells may be susceptible to phenotypic drift in culture. For example, our single-cell RNA-seq results indicate that, in our hands, only a minor subset of the overall NKX2-1+SCGB3A2+ proximal airway population avoids co-expressing distal programs or reaches expression of the more mature secretory marker, SCGB1A1.

In contrast to these plastic or less-differentiated lineages, however, this same global analysis supported our identification of a population of cells expressing a basal cell-like program within airway spheres, and found that these cells exhibit little observed drift toward a distal lung identity. Although the differentiation potential and characteristics of these putative basal-like cells needs to be further elucidated in future studies, this exemplifies the importance and utility of global transcriptomic profiling to precisely and completely identify populations of cells of interest derived from hPSCs.

EXPERIMENTAL PROCEDURES

Further details of experimental protocols, datasets, and reagents can be found online at http://www.kottonlab.com.

Mouse Line Maintenance and Characterization

All studies involving mice were approved by the Institutional Animal Care and Use Committee of Boston University School of Medicine. Mouse strains used are detailed in Supplemental Experimental Procedures.

Mouse iPSC Line Generation and Differentiation

The Nkx2-1^{GFP};Scgb1a1^{CreERTM};Rosa26$^{LSL\text{-}tdTomato}$ iPSC line was generated by transduction of tail tip fibroblasts with a Frt-flanked STEMCCA reprogramming lentiviral vector. Differentiation to lung lineages was performed as described previously (Serra et al., 2017) with further details of reprogramming, line maintenance, and differentiation detailed in the Supplemental Experimental Procedures.

Immunofluorescence Microscopy and qRT-PCR

Details of all immunostaining protocols, antibodies employed, and quantitation of gene expression by qRT-PCR are provided in the Supplemental Experimental Procedures.

hPSC Reporter Line Generation and Maintenance

Human PSC lines were maintained in feeder-free culture in mTeSR1 medium (STEMCELL Technologies). The RUES2-SCGB3A2^{CherryPicker} reporter line was generated by CRISPR/Cas9-based gene editing of RUES2 (gift from Dr. Ali H. Brivanlou, Rockefeller University) as detailed in the Supplemental Experimental Procedures, and the derivation of the NKX2-1^{GFP}/SFTPC^{tdTomato} iPSC line (BU3 NGST) was reported previously (Jacob et al., 2017).

hPSC Directed Differentiation

hPSCs were differentiated to airway or alveolar spheres via endodermal lung progenitors as described previously (Hawkins et al., 2017; Jacob et al., 2017; McCauley et al., 2017) and as detailed in the Supplemental Experimental Procedures. Detailed differentiation

and CD47/CD26 cell sorting protocols can be freely downloaded from http://www.bu.edu/dbin/stemcells/protocols.php.

Single-Cell RNA-Seq and Microarrays

Details of single-cell capture using Fluidigm C1 and 10X Chromium platforms, sequencing of global transcriptomes, and all microarray and RNA-seq bioinformatics analyses are described in the Supplemental Experimental Procedures.

Statistics

Details of all bioinformatics statistics, including lists of differentially expressed genes with fold changes and FDR-adjusted p values are delineated in Tables S2, S3, and S4. Additional relevant statistical methods for qRT-PCR analysis are described in the related figure legends and in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

Global transcriptomic raw datasets can be downloaded from the Gene Expression Omnibus (GEO: GSE103517, GSE103920, and GSE103920) or interrogated through the Bioinformatics Portal at www.kottonlab.com.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and six tables and can be found with this article online at https://doi.org/10.1016/j.stemcr.2018.03.013.

AUTHOR CONTRIBUTIONS

K.B.M. and D.N.K. designed the project, developed the experiments, analyzed the data, and wrote the paper. K.B.M., K.-D.A., A.J., M.V., and F.H. performed differentiations and analyzed the data. K.B.M., W.Y., K.J.S., and E.E.M. generated CRISPR reagents. I.S.C., M.M., and E.E.M. analyzed bioinformatics data. K.B.M., G.C., and B.R.S. performed single-cell RNA-seq experiments. S.K. and S.H.G. performed and analyzed protein immunoblots.

ACKNOWLEDGMENTS

We thank members of the Kotton laboratory, especially Dylan Thomas, for helpful discussions. We thank Anne Hinds and Don Gantz for assistance with electron microscopy. We are grateful to the Boston University (BU) CTSI, Brian Tilton of the BU Flow Cytometry Core (supported by grant 1UL1TR001430), Adam Gower of the BU Microarray Core, Drs. Greg Miller and Marianne James of the CReM (supported by NIH grants R24HL123828 and U01TR001810), and the Cedars-Sinai Genomics Core. We also thank Dr. Charles Plopper for his insightful discussions of airway epithelial ultrastructure. This work was supported by NIH grant awards F31HL129777-01 (to K.B.M.), F31HL134274-01 and TL1TR001410 (to A.J.), and R01HL122442, R01HL095993, R01HL128172, U01HL099997, U01HL134745, and U01HL134766 (to D.N.K.); awards from the Cystic Fibrosis Foundation (CFF HAWKIN15XX0 to F.H. and CFF DAVIS15XX1 to D.N.K.); and an award from the Massachusetts Life Sciences Center to the CReM.

Received: October 24, 2017
Revised: March 15, 2018
Accepted: March 16, 2018
Published: April 12, 2018

REFERENCES

Boers, J.E., Ambergen, A.W., and Thunnissen, F.B. (1999). Number and proliferation of Clara cells in normal human airway epithelium. Am. J. Respir. Crit. Care Med. *159* (5 Pt 1), 1585–1591.

Brasch, F., Johnen, G., Winn-Brasch, A., Guttentag, S.H., Schmiedl, A., Kapp, N., Suzuki, Y., Müller, K.M., Richter, J., Hawgood, S., et al. (2004). Surfactant protein B in type II pneumocytes and intra-alveolar surfactant forms of human lungs. Am. J. Respir. Cell Mol. Biol. *30*, 449–458.

Goss, A.M., Tian, Y., Tsukiyama, T., Cohen, E.D., Zhou, D., Lu, M.M., Yamaguchi, T.P., and Morrisey, E.E. (2009). Wnt2/2b and beta-catenin signaling are necessary and sufficient to specify lung progenitors in the foregut. Dev. Cell *17*, 290–298.

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

Guha, A., Vasconcelos, M., Zhao, R., Gower, A.C., Rajagopal, J., and Cardoso, W.V. (2014). Analysis of notch signaling-dependent gene expression in developing airways reveals diversity of Clara cells. PLoS One *9*, e88848.

Guttentag, S.H., Beers, M.F., Bieler, B.M., and Ballard, P.L. (1998). Surfactant protein B processing in human fetal lung. Am. J. Physiol. *275* (3 Pt 1), L559–L566.

Harris-Johnson, K.S., Domyan, E.T., Vezina, C.M., and Sun, X. (2009). beta-Catenin promotes respiratory progenitor identity in mouse foregut. Proc. Natl. Acad. Sci. USA *106*, 16287–16292.

Hawkins, F., Kramer, P., Jacob, A., Driver, I., Thomas, D.C., McCauley, K.B., Skvir, N., Crane, A.M., Kurmann, A.A., Hollenberg, A.N., et al. (2017). Prospective isolation of NKX2-1-expressing human lung progenitors derived from pluripotent stem cells. J. Clin. Invest. *127*, 2277–2294.

Huang, S.X., Islam, M.N., O'Neill, J., Hu, Z., Yang, Y.G., Chen, Y.W., Mumau, M., Green, M.D., Vunjak-Novakovic, G., Bhattacharya, J., and Snoeck, H.W. (2013). Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. Nat. Biotechnol. *32*, 84–91.

Jacob, A., Morley, M.P., Hawkins, F., McCauley, K.B., Jean, J.-C., Heins, H., Na, C.L., Weaver, T.E., Vedaie, M., Hurley, K., et al. (2017). Differentiation of human pluripotent stem cells into functional lung alveolar epithelial cells. Cell Stem Cell *21*, 472–488.e10.

Kalina, M., Mason, R.J., and Shannon, J.M. (1992). Surfactant protein C is expressed in alveolar type II cells but not in Clara cells of rat lung. Am. J. Respir. Cell Mol. Biol. *6*, 594–600.

Kreda, S.M., Gynn, M.C., Fenstermacher, D.A., Boucher, R.C., and Gabriel, S.E. (2001). Expression and localization of epithelial

aquaporins in the adult human lung. Am. J. Respir. Cell Mol. Biol. *24*, 224–234.

Lee, J.H., Tammela, T., Hofree, M., Choi, J., Marjanovic, N.D., Han, S., Canner, D., Wu, K., Paschini, M., Bhang, D.H., et al. (2017). Anatomically and functionally distinct lung mesenchymal populations marked by Lgr5 and Lgr6. Cell *170*, 1149–1163.e12.

Longmire, T.A., Ikonomou, L., Hawkins, F., Christodoulou, C., Cao, Y., Jean, J.C., Kwok, L.W., Mou, H., Rajagopal, J., Shen, S.S., et al. (2012). Efficient derivation of purified lung and thyroid progenitors from embryonic stem cells. Cell Stem Cell *10*, 398–411.

Mango, G.W., Johnston, C.J., Reynolds, S.D., Finkelstein, J.N., Plopper, C.G., and Stripp, B.R. (1998). Clara cell secretory protein deficiency increases oxidant stress response in conducting airways. Am. J. Physiol. *275* (2 Pt 1), L348–L356.

McCauley, K.B., Hawkins, F., Serra, M., Thomas, D.C., Jacob, A., and Kotton, D.N. (2017). Efficient derivation of functional human airway epithelium from pluripotent stem cells via temporal regulation of Wnt signaling. Cell Stem Cell *20*, 844–857.e6.

McCracken, K.W., Aihara, E., Martin, B., Crawford, C.M., Broda, T., Treguier, J., Zhang, X., Shannon, J.M., Montrose, M.H., and Wells, J.M. (2017). Wnt/β-catenin promotes gastric fundus specification in mice and humans. Nature *541*, 182–187.

Plopper, C.G., Hill, L.H., and Mariassy, A.T. (1980). Ultrastructure of the nonciliated bronchiolar epithelial (Clara) cell of mammalian lung. III. A study of man with comparison of 15 mammalian species. Exp. Lung Res. *1*, 171–180.

Rawlins, E.L., Okubo, T., Xue, Y., Brass, D.M., Auten, R.L., Hasegawa, H., Wang, F., and Hogan, B.L. (2009). The role of Scgb1a1+ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. Cell Stem Cell *4*, 525–534.

Reynolds, S.D., Reynolds, P.R., Pryhuber, G.S., Finder, J.D., and Stripp, B.R. (2002). Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways. Am. J. Respir. Crit. Care Med. *166*, 1498–1509.

Reynolds, S.D., Zemke, A.C., Giangreco, A., Brockway, B.L., Teisanu, R.M., Drake, J.A., Mariani, T., Di, P.Y., Taketo, M.M., and Stripp, B.R. (2008). Conditional stabilization of beta-catenin expands the pool of lung stem cells. Stem Cells *26*, 1337–1346.

Serra, M., Alysandratos, K.-D., Hawkins, F., McCauley, K.B., Jacob, A., Choi, J., Caballero, I.S., Vedaie, M., Kurmann, A.A., Ikonomou, L., et al. (2017). Pluripotent stem cell differentiation reveals distinct developmental pathways regulating lung versus thyroid lineage specification. Development *144*, 3879–3893.

Shannon, J.M. (1994). Induction of alveolar type II cell differentiation in fetal tracheal epithelium by grafted distal lung mesenchyme. Dev. Biol. *166*, 600–614.

Shannon, J.M., Nielsen, L.D., Gebb, S.A., and Randell, S.H. (1998). Mesenchyme specifies epithelial differentiation in reciprocal recombinants of embryonic lung and trachea. Dev. Dyn. *212*, 482–494.

Tata, P.R., Mou, H., Pardo-Saganta, A., Zhao, R., Prabhu, M., Law, B.M., Vinarsky, V., Cho, J.L., Breton, S., Sahay, A., et al. (2013). Dedifferentiation of committed epithelial cells into stem cells in vivo. Nature *503*, 218–223.

Xi, Y., Kim, T., Brumwell, A.N., Driver, I.H., Wei, Y., Tan, V., Jackson, J.R., Xu, J., Lee, D.K., Gotts, J.E., et al. (2017). Local lung hypoxia determines epithelial fate decisions during alveolar regeneration. Nat. Cell Biol. *19*, 904–914.

Xu, Y., Mizuno, T., Sridharan, A., Du, Y., Guo, M., Tang, J., Wikenheiser-Brokamp, K.A., Perl, A.T., Funari, V.A., Gokey, J.J., et al. (2016). Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. JCI Insight *1*, e90558.

Zhang, Y., Goss, A.M., Cohen, E.D., Kadzik, R., Lepore, J.J., Muthukumaraswamy, K., Yang, J., DeMayo, F.J., Whitsett, J.A., Parmacek, M.S., and Morrisey, E.E. (2008). A Gata6-Wnt pathway required for epithelial stem cell development and airway regeneration. Nat. Genet. *40*, 862–870.