ORIGINAL RESEARCH

# Heterogeneity in Human Induced Pluripotent Stem Cell–derived Alveolar Epithelial Type II Cells Revealed with ABCA3/SFTPC Reporters

Yuliang L. Sun[1,2], Killian Hurley[1,2,3,4], Carlos Villacorta-Martin[1], Jessie Huang[1,2], Anne Hinds[2], Krithi Gopalan[1], Ignacio S. Caballero[1], Scott J. Russo[5], Joseph A. Kitzmiller[6], Jeffrey A. Whitsett[6], Michael F. Beers[5], and Darrell N. Kotton[1,2]

[1]Center for Regenerative Medicine of Boston University and Boston Medical Center, Boston, Massachusetts; [2]The Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston, Massachusetts; [3]Department of Medicine, Royal College of Surgeons in Ireland, Education and Research Centre, Beaumont Hospital, Dublin, Ireland; [4]Tissue Engineering Research Group, Royal College of Surgeons in Ireland, Dublin, Ireland; [5]Penn Center for Pulmonary Biology, University of Pennsylvania, Philadelphia, Pennsylvania; and [6]Division of Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio

## Abstract

Alveolar epithelial type 2 cells (AEC2s), the facultative progenitors of lung alveoli, are typically identified through the use of the canonical markers, SFTPC and ABCA3. Self-renewing AEC2-like cells have been generated from human induced pluripotent stem cells (iPSCs) through the use of knock-in SFTPC fluorochrome reporters. However, developmentally, SFTPC expression onset begins in the fetal distal lung bud tip and thus is not specific to mature AEC2s. Furthermore, SFTPC reporters appear to identify only those iPSC-derived AEC2s (iAEC2s) expressing the highest SFTPC levels. Here, we generate an ABCA3 knock-in GFP fusion reporter (ABCA3:GFP) that enables the purification of iAEC2s while allowing visualization of lamellar bodies, organelles associated with AEC2 maturation. Using an SFTPC$^{tdTomato}$ and ABCA3:GFP bifluorescent line for *in vitro* distal lung–directed differentiation, we observe later onset of ABCA3:GFP expression and broader identification of the subsequently emerging iAEC2 population based on ABCA3:GFP expression compared with SFTPC$^{tdTomato}$ expression. Comparing ABCA3:GFP/SFTPC$^{tdTomato}$ double-positive with ABCA3:GFP single-positive (SP) cells by RNA sequencing and functional studies reveals iAEC2 cellular heterogeneity with both populations functionally processing surfactant proteins but the SP cells exhibiting faster growth kinetics, increased clonogenicity, increased expression of progenitor markers, lower levels of SFTPC expression, and lower levels of AEC2 maturation markers. Over time, we observe that each population (double-positive and SP) gives rise to the other and each can serve as the parents of indefinitely self-renewing iAEC2 progeny. Our results indicate that iAEC2s are a heterogeneous population of cells with differing proliferation versus maturation properties, the majority of which can be tracked and purified using the ABCA3:GFP reporter or surrogate cell surface proteins, such as SLC34A2 and CPM.

**Keywords:** human pluripotent stem cells; ATP binding cassette A3; surfactant; alveolar epithelial cells; gene editing

Alveolar epithelial type II cells (AEC2s) are the facultative progenitors of lung alveoli as they maintain the alveolar epithelium through self-renewal or differentiation into alveolar epithelial type I cells (AEC1s) (1–5). AEC2s also serve important physiological roles in the distal lung such as secreting pulmonary surfactant to modulate alveolar surface tension. Mutations in genes highly expressed within AEC2s, such as *SFTPC* (surfactant protein C), *ABCA3* (ATP Binding cassette A3), or *SFTPB* (surfactant protein

(Received in original form June 18, 2020; accepted in final form June 7, 2021)

Supported by the National Institutes of Health grants F30 HL142169, TL1 TR001410, R01 HL095993, U01 HL148692, U01HL134745, U01TR001810, N01 75N92020C00005, and R01 HL145408 and Department of Veterans Affairs Merit Review 2I01BX001176.

Author Contributions: Y.L.S. and D.N.K. designed the project, performed experiments, and wrote the manuscript. K.H. performed the Fluidigm single-cell RNA sequencing experiment. C.V.-M. and I.S.C. performed RNA sequencing analyses. J.H. performed the EdU staining experiment. A.H. performed Transmission electron microscopy. K.G. performed differentiation experiments. S.J.R. and M.F.B. performed Western blot. J.A.K. and J.A.W. provided primary human lung and performed staining.

Correspondence and requests for reprints should be addressed to Darrell N. Kotton, M.D., Boston University School of Medicine, 72 East Concord St., R-304, Boston, MA 02118. E-mail: dkotton@bu.edu.

This article has a related editorial.

This article has a data supplement, which is accessible from this issue's table of contents online at www.atsjournals.org.

Am J Respir Cell Mol Biol Vol 65, Iss 4, pp 442–460, October 2021

Originally Published in Press as DOI: 10.1165/rcmb.2020-0259OC on June 8, 2021
Internet address: www.atsjournals.org

B), can result in pediatric interstitial lung disease and neonatal respiratory distress (6–10), emphasizing their key role in distal lung homeostasis. However, uncovering the mechanisms regulating the inception of disease in AEC2s has been hindered by limited accessibility to primary and patient-specific cells and difficulties in maintaining the AEC2 phenotype *in vitro* (11, 12).

Adding complexity to the study of AEC2s is the recent realization that both the fetal and adult AEC2 populations in mice may be heterogeneous, with the discovery of distinct ACE2 subsets that appear to be more proliferative and more responsive to Wnt signals than the bulk AEC2 population (13–15). In mouse development there also appears to be a gradient of SFTPC expression that identifies less mature developing AEC2 progenitors compared with more mature AEC2s that express higher levels of all surfactants, including SFTPC, based on single-cell RNA sequencing profiles (16). These recent findings raise questions on whether similar heterogeneity also exists in human AEC2s (12) and if identifiable subsets of AEC2s play possible roles during lung development or repair after injury.

To begin to address human AEC2 heterogeneity as well as the putative role of AEC2s in disease pathogenesis, our laboratory and others have derived primary AEC2-like cells (iAEC2s) *in vitro* from human pluripotent stem cells (PSCs) (17–19), using directed differentiation toward a distal lung fate. Directed differentiation *in vitro* emulates key *in vivo* lung developmental milestones and can be monitored by the use of knock-in fluorochrome reporters such as NKX2-1$^{GFP}$ (20) and SFTPC$^{tdTomato}$ (17, 18, 21) that identify the progression of developing cells from early primordial lung progenitors (22) to distal alveolar epithelium (17, 18). Differentiation of PSC lines containing these reporters gives rise to a population of NKX2-1$^{GFP}$/SFTPC$^{tdTomato}$ double-positive (DP) cells that are transcriptionally similar to primary adult AEC2s with some, but not all, cells having lamellar bodies that process and secrete surfactant proteins and phospholipids (17, 18).

While the SFTPC$^{tdTomato}$ reporter has facilitated the isolation of cells, some of which are similar to adult primary AEC2s, *SFTPC* mRNA is also expressed in the developing lung in the distal lung bud tip as well as in AEC2s (23, 24), complicating the identification of maturing AEC2s from a heterogeneous population of developmental precursors. Other potential AEC2 markers include NKX2-1 and SFTPB, both of which are also expressed in other lung epithelial lineages (23, 25), whereas genes such as *PGC* (progastricsin) are highly expressed in mature AEC2s but are also expressed in nonlung endodermal tissues such as the stomach (26), undermining their utility as specific AEC2 markers.

*ABCA3* is a promising candidate as a more stage- and lineage-specific marker because of its AEC2-specific expression within the adult lung and expression later in fetal lung development, with ABCA3 protein not detected before Week 28 of gestation (27), compared with *SFTPC* mRNA and proprotein expression by Week 15 (23). Functionally, ABCA3 protein is localized to the limiting (outer) membranes of lamellar bodies (LBs), where it serves as the major transporter of surfactant phospholipids. Combined with its later onset of expression in the developing lung, ABCA3's critical role in surfactant metabolism as well as its protein localization pattern presents an opportunity to engineer a fusion reporter enabling the visualization of LB biogenesis, a widely acknowledged maturation milestone in AEC2 fetal development.

Here, we address the question of AEC2 cellular heterogeneity and maturation in our human PSC-derived distal lung epithelial *in vitro* model system by engineering combinatorial reporter systems to identify and isolate distinct populations of PSC-derived AEC2s for transcriptomic and functional characterizations. We first conduct single-cell RNA sequencing (RNA-seq) on PSC-derived distal lung epithelial cells containing NKX2-1 and SFTPC bifluorescent reporters, revealing two transcriptionally similar AEC2 populations expressing similar levels of most canonical AEC2 genes, including *ABCA3*, but different levels of *SFTPC* and proliferative genes. We find a gradient of *SFTPC* mRNA expression in iAEC2s where only those iAEC2s expressing the highest levels of the SFTPC locus express detectable fluorescence from the SFTPC$^{tdTomato}$ reporter. To encompass the cells that seemingly shared the AEC2 program missed by this tdTomato reporter, we developed a new ABCA3:GFP fusion reporter that localizes GFP to the limiting membrane of intracellular vesicles reminiscent of LBs in PSC-derived AEC2s after distal lung differentiation. Using the new ABCA3 fusion reporter in conjunction with the SFTPC reporter to distinguish AEC2-like cells from other lung epithelial lineages, we first observed ABCA3 expression occurring after SFTPC expression, as expected, followed by emergence of distinct populations of cells expressing both reporters as well as a larger population singularly expressing ABCA3:GFP fluorescence characterized by lower levels of some AEC2 maturation markers. Although both reporter populations functionally processed surfactant proteins, the GFP single-positive (SP) population expressed higher levels of transcripts associated with proliferation, faster growth kinetics, and increased clonogenicity. Taken together, our results indicate PSC-derived human AEC2-like cells are a heterogeneous population of cells with differing proliferative and maturation properties.

## Methods

For detailed methods regarding qRT-PCR, immunostainings, Western blot analysis, RNA-seq profiling of PSC-derived populations, and EM, please *see* the data supplement.

### Generation of Knock-in ABCA3:GFP PSC Reporter System for Isolation of iAEC2s

Targeting of embryonic stem cell (ESC) (RUES2) and iPSC (BU3) lines with the knock-in ABCA3:GFP fusion reporter was done using CRISPR/Cas9 gene editing as detailed in the data supplement. In brief, two plasmid constructs were introduced to PSCs by nucleofection (Lonza): (1) a plasmid construct constitutively expressing Cas9 driven by the CBh promoter and a 20 bp gRNA sequence targeting the stop codon of the endogenous *ABCA3* locus driven by the U6 promoter and (2) a donor plasmid construct containing arms of homology, sequences encoding GFP, and a floxed PGK promoter-driven puromycin resistance cassette. Following antibiotic resistance selection of transfected clones and confirmation of *ABCA3* locus targeting by Sanger sequencing, the floxed puromycin resistance cassette was excised by transfection of a plasmid encoding Cre recombinase. All targeted PSC lines (RUES2-STAG, BU3$^{ABCA3:GFP}$) were checked for

normal karyotype by G-banding analysis (Cell Line Genetics).

### Directed Differentiation of PSCs to Distal Lung Epithelium

Directed differentiation was performed as detailed in the data supplement and previously published (17, 18, 20). Day 0 PSCs were differentiated into definitive endoderm (Day 0–3) using StemDiff Endoderm Kit (Stem Cell Technologies, 05110), toward anterior foregut endoderm (Day 3–6) using DS/SB media (2 μM dorsomorphin, Stemgent, 040024; 10 μM SB431543, Biotechne, 1614), and then further specified into NKX2-1+ lung epithelial progenitors using CBRa media (3 μM CHIR99021, Biotechne, 4423; 10 ng/ml rhBMP4, BioTechne, 314BP; 100 nM retinoic acid, Sigma-Aldrich, R2625; Day 6–15). On Day 15, NKX2-1 expressing lung epithelial progenitors were sorted either by NKX2-1$^{GFP}$(BU3-NGST line) or using CD47$^{hi}$/CD26$^{lo}$ sorting to enrich for NKX2-1 (20), and fed with distalizing CK+DCI media (3 μM CHIR99021, 10 ng/ml KGF, 50 nM dexamethasone, 0.1 mM cyclic AMP, and 0.1 mM IBMX).

## Results

### Directed Differentiation of Distal Lung Epithelium from Human iPSCs Results in Heterogeneous Cell Populations Expressing a Gradient of *SFTPC* mRNA

To profile the heterogeneity of distal lung epithelial cells arising from PSCs during *in vitro* distal lung differentiation, we differentiated a previously published human iPSC line containing NKX2-1$^{GFP}$ and SFTPC$^{tdTomato}$ reporters (hereafter referred to as BU3-NGST) using our previously published protocol (17, 18). On Day 15 of differentiation, we purified NKX2-1$^{GFP+}$ putative lung primordial progenitors (20) by flow cytometry and replated this population in 3D Matrigel in the presence of a culture medium (17) designed to promote distal alveolar epithelial differentiation (Figure 1A). Consistent with our prior publications (17, 18, 21), monolayered epithelial spheres emerged from the sorted cells, and NKX2-1$^{GFP}$/SFTPC$^{tdTomato}$ DP cells were first detectable 48 hours later (17 days total differentiation time). By Day 33 of differentiation, these cells made up ~4.5% of the total population. GFP+/tdTomato+ cells represented a subset of the 68% NKX2-1$^{GFP}$+ cell population at this time point (Figure 1B), suggesting heterogeneity in the population of putative lung epithelial (NKX2-1+) cells generated by this approach.

To determine the transcriptional differences between the NKX2-1$^{GFP}$/SFTPC$^{tdTomato}$ DP and NKX2-1$^{GFP}$ SP cells, we sorted pure cells from each population for single-cell RNA sequencing (scRNA-seq). Analyses of sequencing data revealed 147 differentially expressed genes (DEGs) between the two populations but no distinct clustering by Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction analysis (Figure 1C, Table E1 in the data supplement, absolute log2fold change [log2FC] $>$ 1, false discovery rate [FDR] $<$ 0.05) with SFTPC being the top upregulated gene in the NKX2-1$^{GFP}$/SFTPC$^{tdTomato}$ DP population and *TOP2A* being the top downregulated gene (Figure 1D). Although differentially expressed, this canonical transcript of AEC2 identity was detectably expressed in both populations, suggesting that both populations likely contained distal lung epithelium. Consistent with this likelihood, additional AEC2 genes were expressed at similar levels (FDR $>$ 0.05) between the two populations, including *NKX2-1*, *LPCAT1*, *NAPSA*, and *ABCA3* (Figure 1D). A subset of AEC2 genes, however, were upregulated in the DP population, including *SFTPB*, *CTSH*, and the AEC2 maturation marker *PGC* (21) (Figure 1E, FDR $<$ 0.05). In the NKX2-1$^{GFP}$ SP population, upregulated genes were those associated with cell division and DNA synthesis, including *TOP2A*, *MKI67*, *CENPF*, and *MCM3* (Figure 1F, FDR $<$ 0.05). These results suggested a shared AEC2 program in both populations, with the NKX2-1$^{GFP}$ SP population comprising a potentially less mature and more proliferative subset of putative iAEC2s.

### ABCA3 Knock-in GFP Fusion Reporter Enables Isolation and Tracking of iAEC2s and Visualization of LBs

Because the AEC2-specific transcript, ABCA3, exhibited shared expression in both NKX2-1$^{GFP}$+ and NKX2-1$^{GFP}$+/SFTPC$^{tdTomato}$+ populations (Figure 1D) and ABCA3 protein function is known to be necessary for lamellar body biogenesis and AEC2 surfactant metabolism *in vivo* (28), we speculated that visualization of ABCA3 protein might serve to identify iPSC-derived AEC2s in our model system with higher sensitivity than our SFTPC$^{tdTomato}$ reporter. We also hypothesized that tracking localization of ABCA3 protein might facilitate the monitoring of AEC2 functional maturation because *ABCA3* is both a canonical identity gene in AEC2s and a required surfactant phospholipid transporter known to localize to the limiting membrane of LBs in mature AEC2s (28–31).

To generate a new iPSC line carrying an ABCA3 reporter that simultaneously enables the visualization of its intracellular trafficking pattern, we used CRISPR/Cas9 gene editing to insert a knock-in GFP reporter at the stop codon of the endogenous *ABCA3* locus in our published normal human BU3 iPSC line (20) and in the human embryonic stem cell (ESC) line, RUES2, resulting in an ABCA3 translational product that is fused to the GFP protein (hereafter ABCA3:GFP; Figure 2A). Next, we differentiated BU3 iPSCs containing the ABCA3:GFP fusion reporter (BU3AG23CR6 human iPSC line, hereafter BU3$^{ABCA3:GFP}$) using our protocol promoting distal alveolar epithelial development in 3D culture (Figure 2B) and observed the emergence of cells expressing ABCA3:GFP within monolayered epithelial spheres (Figure 2C). Coimmunofluorescent staining of these epithelial spheres using antihuman ABCA3 antibodies and anti-GFP antibodies revealed colocalization of the GFP fusion reporter with the ABCA3 protein (Figure 2D). Next, we quantified GFP expression kinetics, with first onset visualized around Day 22 and detected by flow cytometry in 8.95% ± 0.87% of all cells by Day 31 of differentiation (Figures 2C–2J). By high resolution confocal microscopy, we found that the ABCA3:GFP fusion protein also localized to cytoplasmic vesicles, which costained with the LB-associated marker, LAMP3 (Figures 2E and 2F), reminiscent of ABCA3′s endogenous localization to LBs in primary AEC2s of an 8-year-old human lung (Figure 2G). Notably, the ABCA3:GFP+ vesicles appeared to localize to the apical end of each cell toward the lumens of these spheres, a possible indication of the presence of mature LBs poised for exocytosis into the alveolar lumen. Further imaging using toluidine blue staining for lamellar bodies revealed LBs in the sphere lumen (Figure 2H). Electron micrographs of BU3$^{ABCA3:GFP}$-derived spheres validated the presence of mature LBs and their precursor organelles, multivesicular bodies (Figure 2I). To perform an unbiased assessment of ABCA3:GFP-expressing cells, we performed bulk RNA sequencing of sorted

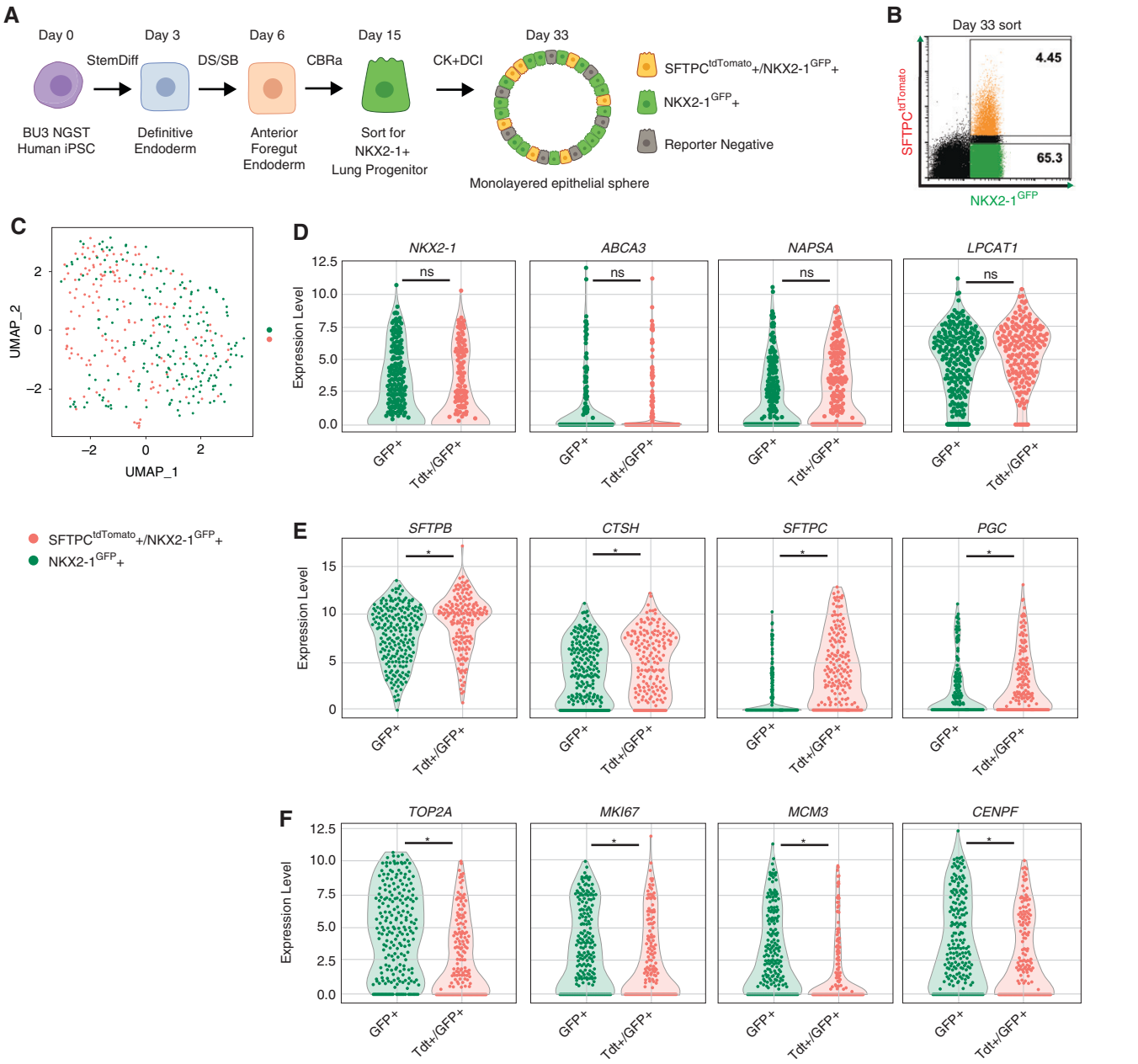


**Figure 1.** SFTPC$^{tdTomato}$ and NKX2-1$^{GFP}$ reporters capture distinct populations of induced pluripotent stem cell (iPSC)–derived distal lung epithelial cells. (*A*) Schematic detailing key steps of *in vitro* distal lung–directed differentiation of BU3-NGST human iPSCs. (*B*) Flow cytometry of Day 33 BU3-NGST epithelial spheres indicating sort gates for SFTPC$^{tdTomato}$+/NKX2-1$^{GFP}$+ double-positive (DP) and NKX2-1$^{GFP}$ single-positive (SP) cells used for single-cell RNA sequencing profiling. (*C*) Uniform Manifold Approximation and Projection (UMAP) plot of transcriptomic profiles of single cells captured in *B* using the Fluidigm platform. (*D*) Violin plots showing normalized expression for select lung and AEC2 genes not differentially expressed between the SFTPC$^{tdTomato}$+/NKX2-1$^{GFP}$+ DP and NKX2-1$^{GFP}$ SP cells. (*E*) Violin plots showing normalized expression for indicated genes significantly upregulated in the SFTPC$^{tdTomato}$+/NKX2-1$^{GFP}$+ DP cells. *False discovery rate (FDR) < 0.05. (*F*) Violin plots showing normalized expression for indicated genes associated with cellular proliferation and division, upregulated in NKX2-1$^{GFP}$ SP cells. *FDR < 0.05. AEC2 = alveolar epithelial type 2 cell; B = BMP4; C = CHIR; CI = cyclic AMP and IBMX; D = dexamethasone; DS = dorsomorphin; K = KGF; ns = not significant, FDR > 0.05; Ra = retinoic acid; SB = SB431543; StemDiff = StemDiff Endoderm Kit.

ABCA3:GFP+ versus GFP− cells on Day 31 of differentiation ($n = 3$; Figures 2B, 2J, and 2K). Eight hundred thirty genes were differentially upregulated in ABCA3:GFP+ cells (FDR < 0.01; Table E2). Further analysis of the top 50 upregulated transcripts in the GFP+ population revealed enrichment of genes associated with LB and LB precursor organelles (Enrichr GO: cellular compartment analysis) (Figure 2M). Among the top 50 upregulated genes were AEC2 genes including *ABCA3, SFTPB, SFTPA2, ADGRF5, NAPSA, SLC34A2, LAMP3, CTSH, CEACAM6,* and *LPCAT1* (Figure 2K). ABCA3 was the number one most

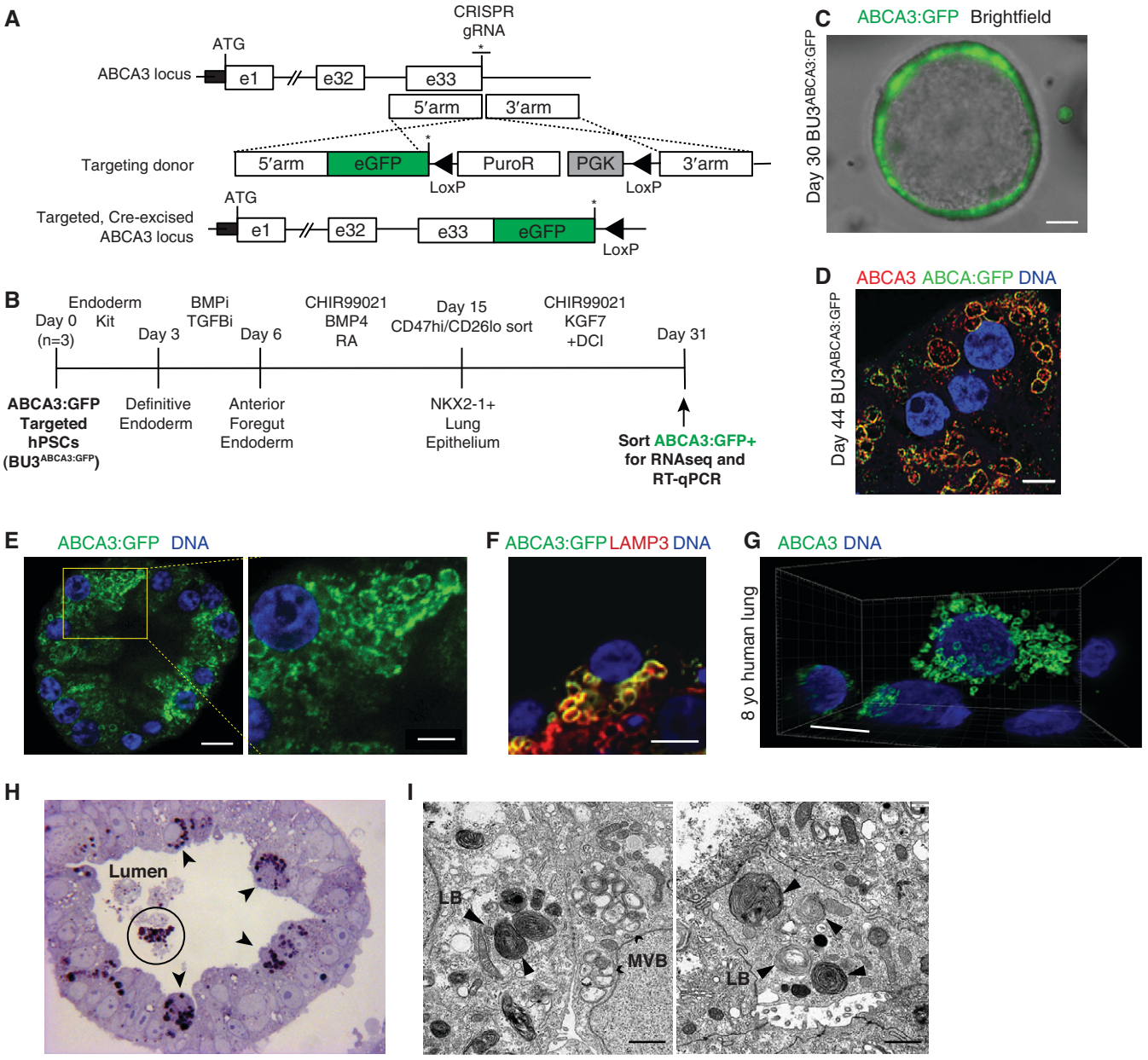


**Figure 2.** ABCA3:GFP knock-in fusion reporter enables isolation of iAEC2s and intracellular localization to lamellar bodies. (*A*) CRISPR/Cas9-based gene editing strategy to introduce a GFP reporter fused to the end of the endogenous ABCA3 coding sequence. *Stop codon. (*B*) Schematic of distal lung–directed differentiation using the BU3$^{ABCA3:GFP}$ iPSC line. (*C*) Representative fluorescence microscopy image of ABCA3:GFP expression in Day 30 BU3$^{ABCA3:GFP}$ spheres. Scale bar, 50 μm. (*D*) Confocal microscopy of Day 44 BU3ABCA3:GFP iAEC2 showing colocalization of ABCA3 protein (red represents anti-ABCA3 immunostaining) and GFP (green represents anti-GFP immunostaining). Scale bar, 5 μm. (*E*) Confocal microscopy showing intracellular localization of GFP (ABCA3:GFP, green) in a Day 88 epithelial sphere. Scale bars,10 μm. (*F*) Confocal microscopy showing colocalization of GFP (ABCA3:GFP, green) with LAMP3 (red) in an epithelial sphere. Scale bar, 10 μm. (*G*) Confocal microscopy of 8-year-old human distal lung immunostained for ABCA3 protein (green). Scale bar, 10 μm. (*H*) Plastic section of Day 88 ABCA3:GFP-expressing epithelial sphere stained with toluidine blue. Arrowheads indicate putative lamellar bodies (dark purple inclusions). Circle indicates intralumenal lamellar bodies consistent with putative secretion. (*I*) Transmission EM of BU3$^{ABCA3:GFP}$ epithelial sphere. Arrowheads indicate lamellar bodies; chevrons indicate multivesicular bodies. Scale bars, 500 nm. (*J*) Flow cytometry of iPSC-derived cells (Day 31 of differentiation of the BU3$^{ABCA3:GFP}$ line) with GFP+ versus GFP− sort gates for bulk RNA sequencing. (*K*) Heatmap comparing relative gene expression of top 50 differentially expressed genes in sorted GFP+ and GFP− cells (FDR < 0.01). (*L*) Gene expression (qRT-PCR; $2^{-\Delta\Delta Ct}$) of Day 31 BU3$^{ABCA3:GFP}$ cells sorted as in *I* and compared with Week 21 distal human fetal lung epithelium (HFL). Bars represent mean fold change ($2^{-\Delta\Delta Ct}$) ± SE, $n = 3$ biological replicates separated on Day 15. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and ****$P < 0.0001$. (*M*) Bar graph of combined enrichment score of top 50 differentially expressed genes analyzed by Enrichr GO: cellular compartment analysis (FDR < 0.05). iAEC2 = iPSC-derived AEC2; LB = lamellar body; MVB = multivesicular body; ns = not significant by unpaired, two-tailed Student's *t* test.

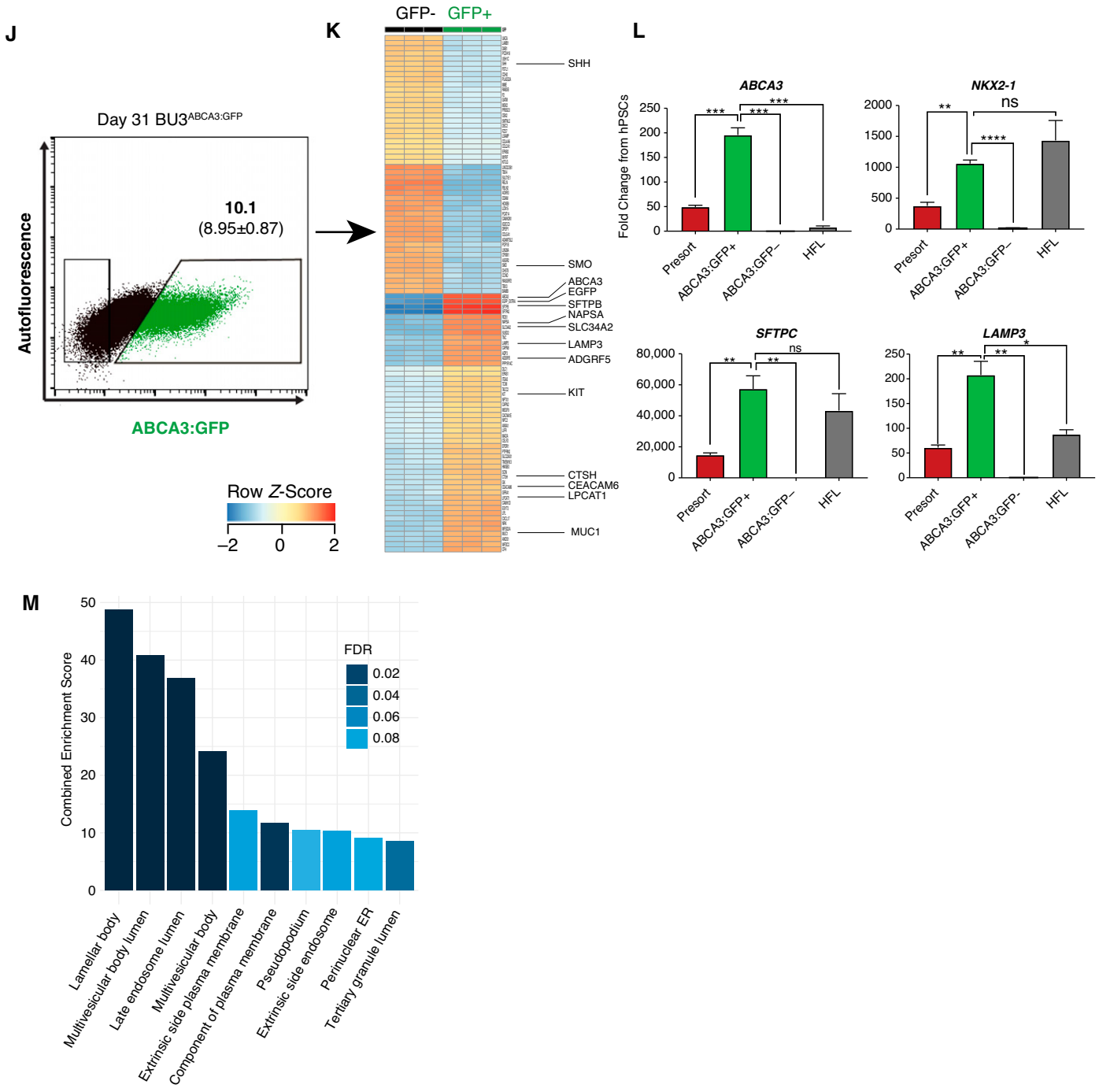


**Figure 2.** (*Continued*).

upregulated gene, further supporting the faithfulness of the ABCA3:GFP reporter. We validated the enrichment of selected alveolar epithelial genes (*NKX2-1, ABCA3, SFTPC, LAMP3*, and *LPCAT1*) in ABCA3:GFP+ cells by qRT-PCR and found their expression levels were at or above the levels of primary fetal alveolar epithelial control cells (Figure 2L).

Transcriptomic analysis also revealed *SLC34A2*, a sodium phosphate transporter, as a candidate surrogate cell surface marker expressed in ABCA3+ cells, consistent with recent findings from Korogi and colleagues (32). We prepared Day 151 BU3$^{ABCA3:GFP}$ cells and found that sorting cells stained with anti-SLC34A2 antibody resulted in enrichment of iAEC2s, with >90% of SLC34A2-high cells expressing the ABCA3:GFP reporter (Figure E1A). Moreover, SLC34A2 sorting of Day 28 cells from triplicate differentiations of the BU3$^{ABCA3:GFP}$ line resulted in the enrichment of cells expressing *SLC34A2, ABCA3, SFTPC, NAPSA*, and *PGC* transcripts, compared with presort, mixed populations (Figures E1B and E1C). Taken

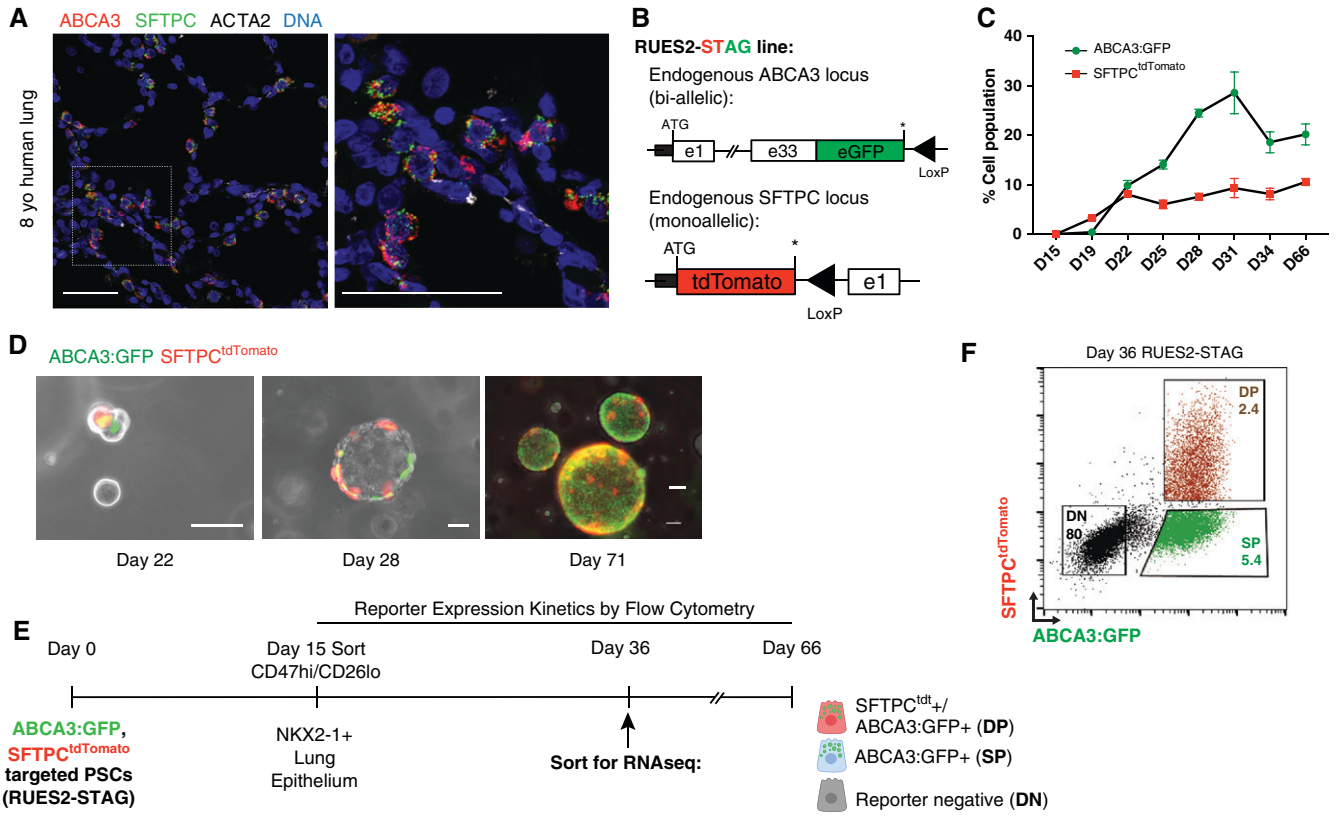


**Figure 3.** iAEC2 heterogeneity identified using ABCA3:GFP/SFTPC$^{tdTomato}$ bifluorescent reporter line. (*A*) Immunofluorescence microscopy of 8-year-old human lung showing coexpression of ABCA3 (red) and SFTPC (green) proteins in AEC2s. Scale bar, 40 μm. (*B*) Schematic showing biallelic targeting of GFP fusion reporter to the endogenous ABCA3 locus and monoallelic targeting of tdTomato reporter to the endogenous SFTPC locus in the RUES2 human ESC line, resulting in the generation of the bifluorescent RUES2-STAG line. *Stop codon. (*C*) ABCA3:GFP and SFTPC$^{tdTomato}$ reporter expression kinetics by flow cytometry analysis from Day 15 to Day 66 of *in vitro* distal lung–directed differentiation. Dots represent mean cell population ± SE, biological replicates ($n = 3$), separated at Day 15. (*D*) Representative fluorescence microscope images of RUES2-STAG differentiation at indicated time points. ABCA3:GFP (green), SFTPC$^{tdTomato}$ (red). Scale bars, 300 μM. (*E*) Time course of RUES2-STAG differentiation. (*F*) Representative flow cytometry analysis showing sorting gates for reporter double-negative (DN), reporter DP, and reporter SP Day 36 RUES2-STAG cells harvested for bulk RNA sequencing analysis. (*G*) Heatmap comparing relative gene expression of differentially expressed genes in sorted populations in *E* (FDR < 0.05). (*H*) Heatmap comparing relative gene expression of differentially expressed genes in sorted reporter-positive (DP, SP) populations (FDR < 0.05). (*I*) Heatmap comparing relative gene expression of differentially expressed genes associated with cell cycle in sorted reporter-positive (DP, SP) populations (FDR < 0.05). (*J*) Bar graph showing average expression (FPKM) of select genes between DP and SP populations. Error bars represent SE. (*K*) Gene expression profiles (by qRT-PCR; $2^{-\Delta\Delta Ct}$) of the same Day 36 RUES2-STAG cells sorted in *E* for select canonical AEC2 genes compared with Week 21 distal HFL control. Bars represent mean fold change ($2^{-\Delta\Delta Ct}$) ± SE, $n = 3$ biological replicates separated on Day 15. *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$. (*L*) Bar graph showing normalized enrichment score of indicated hallmark gene sets enriched by DP and SP cells (FDR < 0.05). ns = not significant by unpaired, two-tailed Student's *t* test.

together, these results further support SLC34A2 as an effective surrogate cell surface marker that can be used to isolate iAEC2s expressing ABCA3, as recently described by Korogi and colleagues (32).

### iAEC2s Derived from a Bifluorescent Reporter PSC Line (Carrying ABCA3:GFP/SFTPC$^{tdTomato}$) Allows the Study of Alveolar Epithelial Cell Developmental Kinetics

*SFTPC* message is expressed in the developing human distal lung tip by Week 15 of gestation (23), and ABCA3 protein expression has been described as first detectable at some point between Week 23 (undetectable) and Week 28 (detectable) of gestation (27). Both proteins are coexpressed in AEC2s as demonstrated in our immunofluorescent staining of 8-year-old human lung (Figure 3A). To see if our reporters would recapitulate the temporal sequence of SFTPC and ABCA3 expression observed *in vivo*, we generated a new human ESC line bifluorescently targeted with the SFTPC$^{tdTomato}$ and ABCA3:GFP reporter, each targeted to the endogenous *SFTPC* and *ABCA3* loci, respectively (RUES2ST15AG4CR8 human ESC line, hereafter RUES2-STAG; Figure 3B). To document the emergence and frequency of expression of each reporter, we harvested distal differentiations of RUES2-STAG from Day 15 to Day 66 for flow cytometry (Figures 3C and 3E). On Day 19, we first observed the emergence of SFTPC$^{tdTomato}$+ cells, followed by emergence of ABCA3:GFP on Day 22 (Figures 3C and 3D), consistent with the known sequence of expression of SFTPC and ABCA3 *in vivo* during human lung development. The frequency of SFTPC$^{tdTomato}$ reporter expression remained

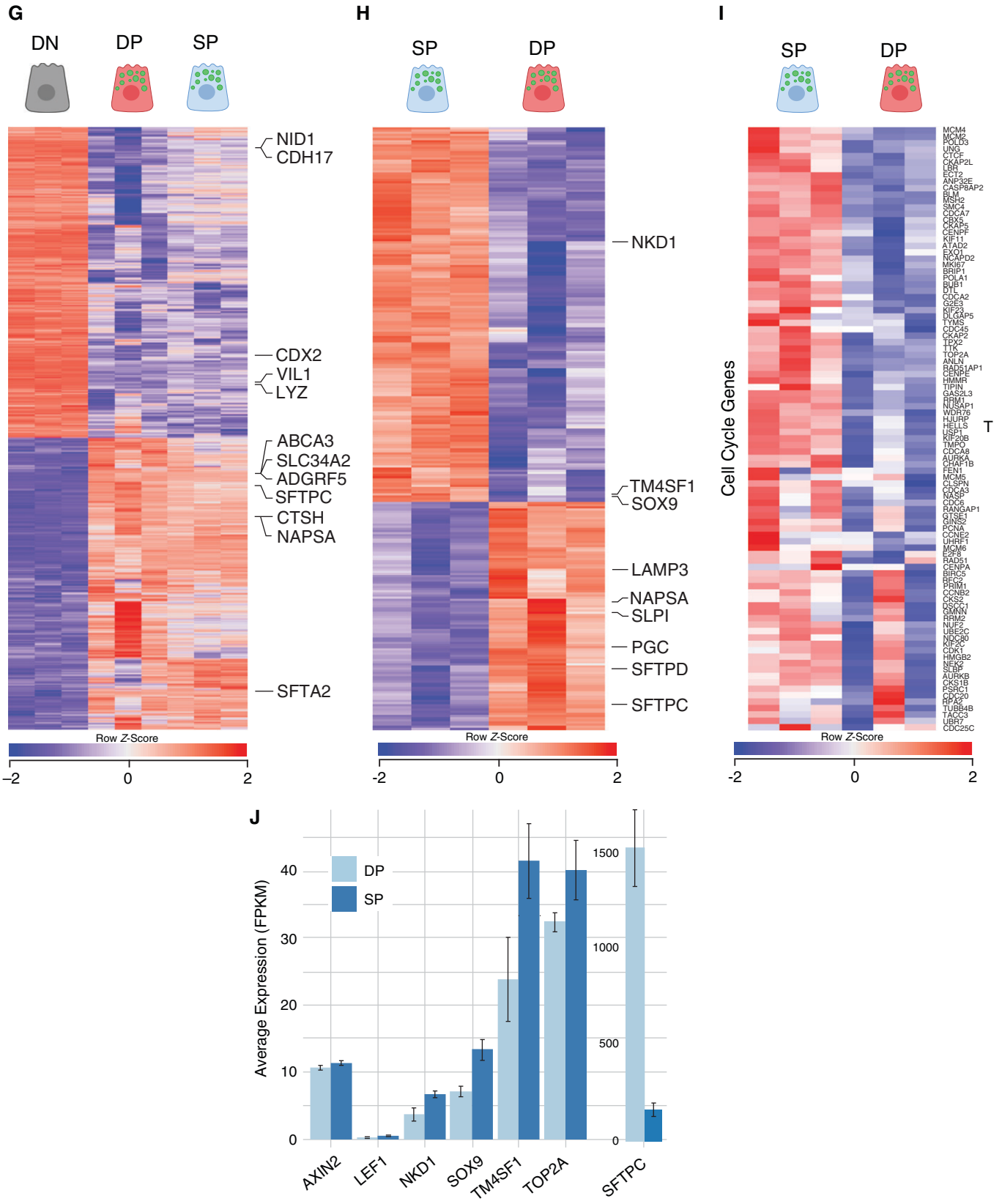


**Figure 3.** (*Continued*).

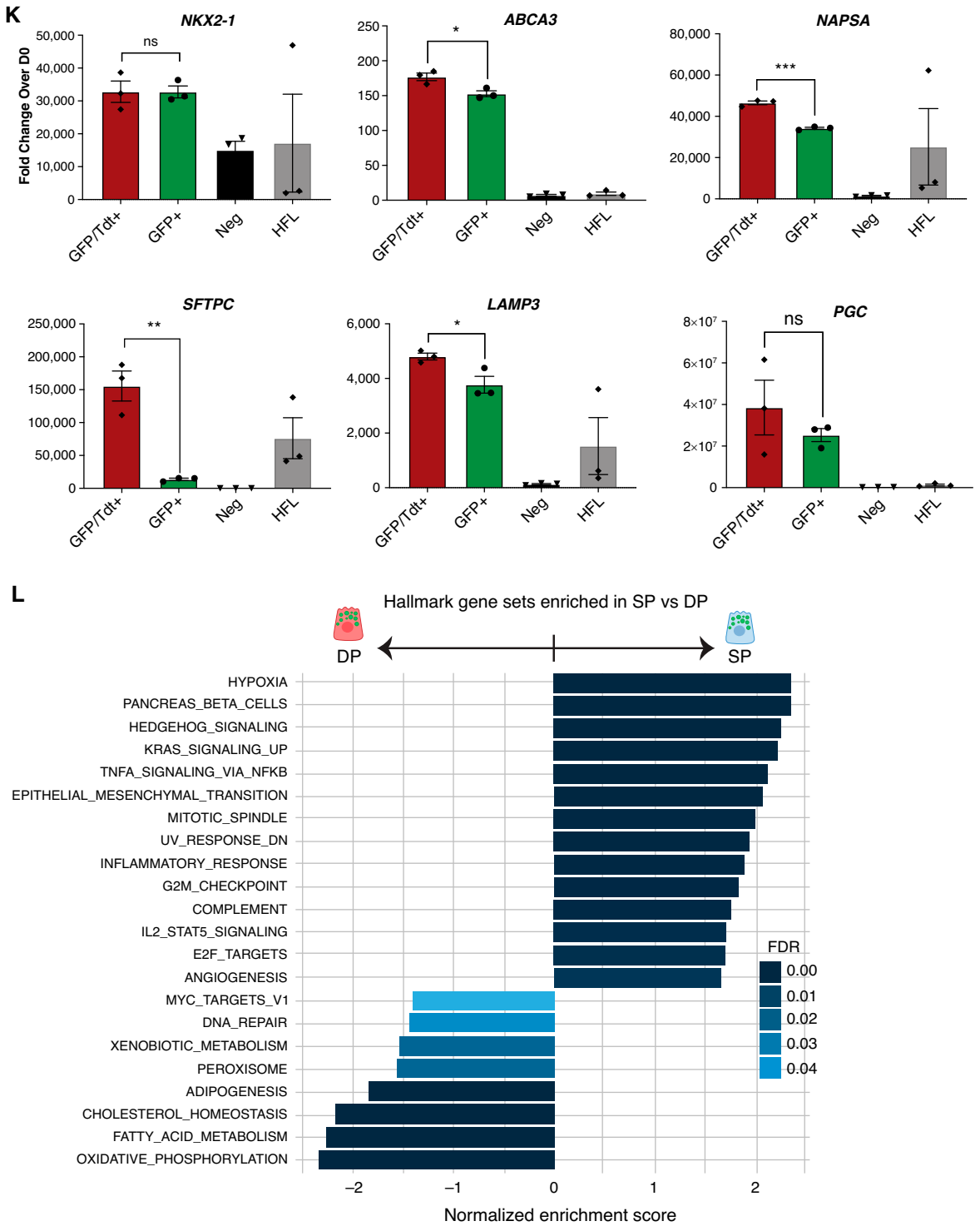


**Figure 3.** (*Continued*).

relatively stable throughout the differentiation (8.02% ± 0.78% on Day 22 to 10.55% ± 1.22% on Day 66 of the total cell population). In contrast, the frequency of ABCA3:GFP-expressing cells increased with time, overtaking the tdTomato+ frequency and peaking on Day 31 at 28.57% ± 7.3% (compared with the frequency of $SFTPC^{tdTomato}$+ cells at this time point of 9.31% ± 3.34%; Figure 3C). This increased frequency of ABCA3:GFP-expressing iAEC2s is consistent with our initial scRNA-seq data indicating ABCA3 message in the $NKX2\text{-}1^{GFP}$ SP population (Figure 1D), highlighting the capability of the ABCA3:GFP reporter to

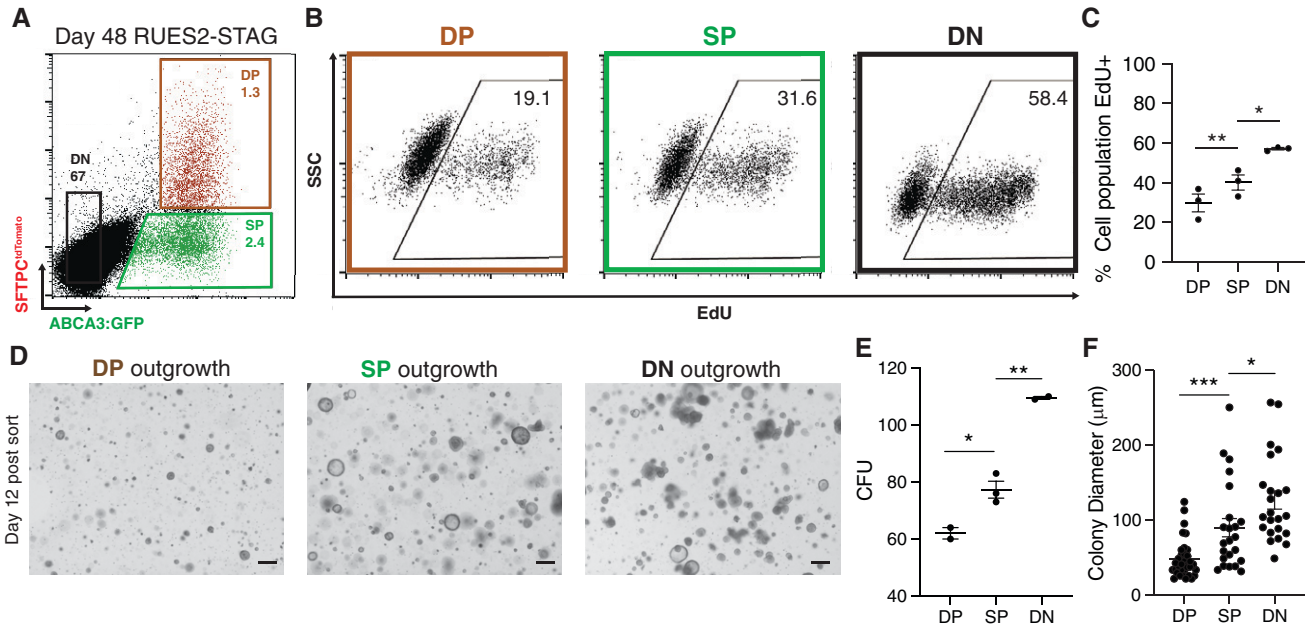


**Figure 4.** ABCA3:GFP SP population is a more clonogenic and proliferative subset of iAEC2s. (*A*) Representative flow cytometry analysis of Day 48 RUES2-STAG spheres sorted for EdU incorporation and colony formation assays. (*B*) Representative flow cytometry analysis of EdU incorporation by DP, SP, and DN sorted cells. (*C*) Dot plots showing percentage of EdU incorporation by indicated cell population. Mean ± SE, $n = 3$ biological triplicate separated at Day 0. *$P < 0.05$ and **$P < 0.01$, by paired, two-tailed Student's *t* test. (*D*) Representative brightfield images of DP, SP, and DN colony outgrowths 12 days after sort. Scale bars, 300 μm. (*E*) Dot plots showing colony formation unit measurements 12 days after sort in indicated populations. Mean ± SE, $n = 3$ biological triplicate separated at Day 0. *$P < 0.05$ and **$P < 0.01$, by unpaired, two-tailed Student's *t* test. (*F*) Dot plots showing colony diameter measurements 12 days after sort in indicated populations. Mean ± SE, $n = 3$ biological triplicate separated at Day 0. *$P < 0.05$ and ***$P < 0.001$, by unpaired, two-tailed Student's *t* test. CFU = colony formation unit; SSC = side scatter.

identify a potentially larger population of putative iAEC2s compared with the SFTPC$^{tdTomato}$ reporter.

### iAEC2 Heterogeneity Is Supported by Transcriptomic Comparisons between ABCA3:GFP/SFTPC$^{tdTomato}$ Reporter DP and ABCA3:GFP SP iAEC2s

Within the lung, SFTPC and ABCA3 are both expressed in mature AEC2s where their protein products can be found in LBs. In our differentiations using the bifluorescent reporter line, we found that after Day 22 of differentiation, almost all SFTPC$^{tdTomato}$+ cells coexpressed ABCA3:GFP+ (Figures 3D–3F). Expression of the two reporters identified three distinct populations: ABCA3:GFP/SFTPC$^{tdTomato}$ DP, ABCA3:GFP SP, and reporter double-negative (DN) populations (Figures 3E and 3F). Bulk RNA sequencing of each of these three populations, purified by cell sorting (Day 36; $n = 3$ replicates) revealed 4,023 DEGs in a pairwise comparison between reporter-positive (DP and SP combined) and reporter DN populations (DEG FDR < 0.05, abs log2FC > 1; Figure 3G; Table E3) with enrichment of the AEC2 program in both SP and DP populations compared with the DN cells (*see* Figure 3G *SFTPC, ADGRF5, SLC34A2, CTSH, NAPSA, SFTA2*, and *ABCA3*). In contrast, the DN population was enriched in nonlung endodermal genes such as *CDX2, CDH17, VIL1, TFF1*, and *AFP*, consistent with our prior reports of some degree of instability or plasticity in iPSC-derived "lung" cells over time after sorting cells at early stages of lung-directed differentiation (18, 21, 22).

We next performed a pairwise comparison between the reporter DP and SP populations (Figures 3H–3L). Consistent with our scRNA-seq data (Figure 1), we found high transcriptomic similarities with only 280 DEGs (FDR < 0.05; Table E3). As expected, SFTPC was the top upregulated gene in the reporter DP population but was expressed in both DP and SP populations (Figures 3H and 3K). Among the differentially upregulated genes in DP cells were surfactant-related genes such as *SFTPC, NAPSA*, and *SFTPB*, whereas markers of distal lung epithelial progenitors such as *SOX9* and *TM4SF1* and markers of cell proliferation such as *TOP2A* and *MKI67* were differentially upregulated in the ABCA3:GFP SP population (Figures 3H–3J; Table E3). qRT-PCR expression assays measuring key AEC2 genes revealed the same *NKX2-1* levels between the two reporter-positive populations but slightly higher expression levels of several key ACE2 surfactant-related genes such as *ABCA3, SFTPC, NAPSA*, and *LAMP3* in the ABCA3:GFP/SFTPC$^{tdTomato}$ reporter DP population (Figure 3K). Gene set enrichment analysis revealed upregulation in DP cells of pathways associated with surfactant lipid processing in AEC2s (17, 33, 34). For example, lipid metabolism and recycling pathways represented four out of the top five pathways including fatty acid metabolism, cholesterol homeostasis, adipogenesis, and peroxisome pathways (FDR < 0.05; Figure 3L). In contrast, in the SP ABCA3:GFP population, proliferation-related pathways were enriched, such as mitotic spindle, G2M checkpoints, E2F targets, hedgehog signaling, and KRAS signaling, (FDR ≤ 0.001), suggestive of a more proliferative subset of iAEC2s. Taken together, these results, similar

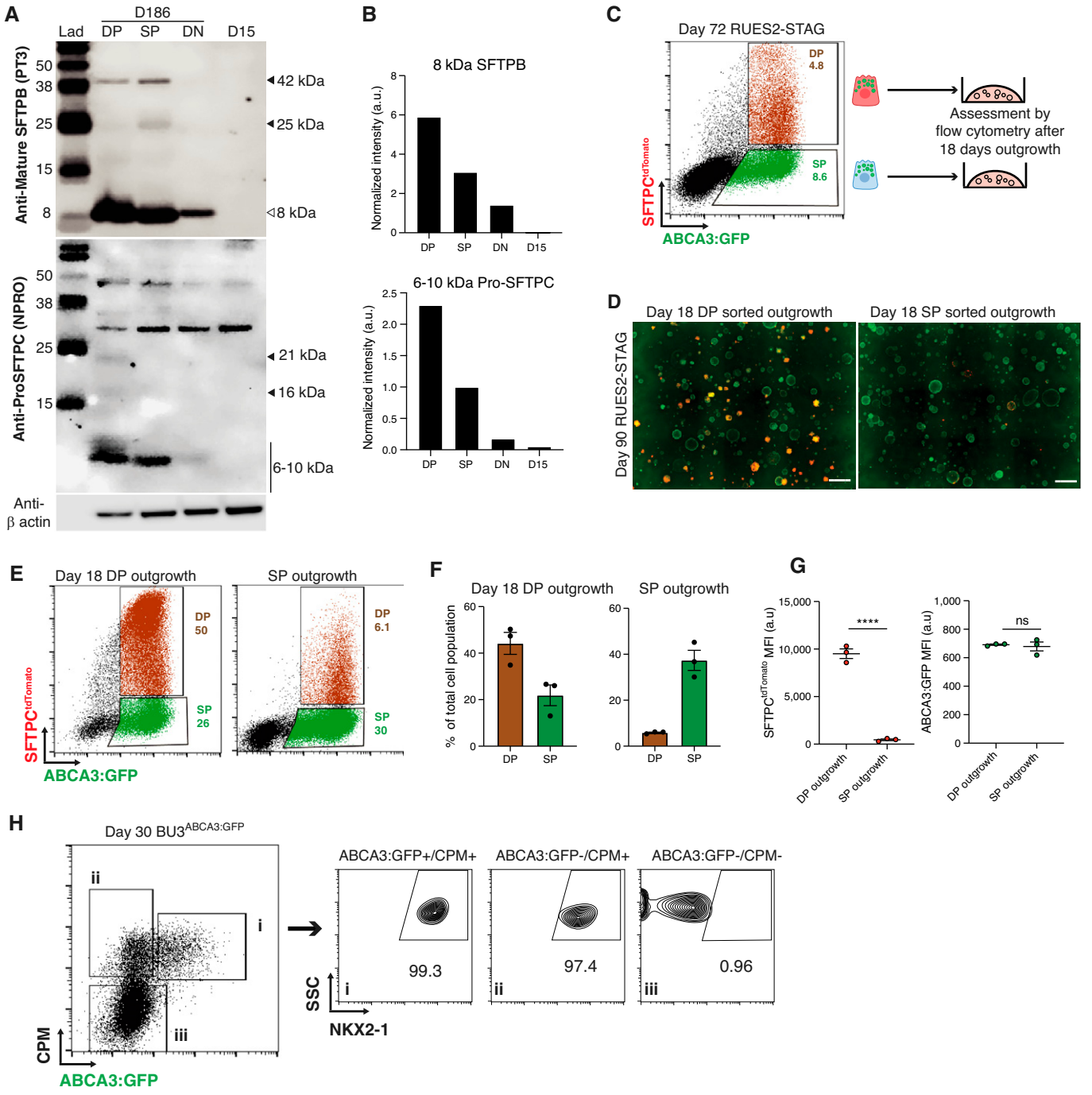


**Figure 5.** SFTPC$^{tdTomato}$/ABCA3:GFP DP and ABCA3:GFP SP populations are interchangeable states of iAEC2s that process surfactant proteins. (*A*) Top: Western blot for mature SFTPB protein (immunoblotted with PT3 antibody) on sorted Day 186 DP, SP, DN, RUES2-STAG cells and Day 15 NKX2-1 + progenitor cells (D15, negative control). Bands showing 42-kD unprocessed SFTPB precursor protein and the 25-kD intermediate protein (black arrowheads) are present in DP and SP populations and absent in the DN and D15 samples. Bands showing mature 8-kD SFTPB (white arrowhead) are present in DP, SP, and DN samples but not in D15 controls. Bottom: Western blot for proSFTPC processing (immunoblotted with NPRO antibody) on indicated samples. Bands showing 21-kD and 16-kD proSFTPC processing intermediates (black arrowheads) are faintly present in DP cells. Bands showing processed 6- to 10-kD proSFTPC protein (vertical line) are only present in DP and SP cells. (*B*) Bar graphs showing β-actin–normalized intensities of processed 8-kD SFTPB (top) and 6- to 10-kD proSFTPC (bottom) proteins in DP, SP, DN, and D15 cells. (*C*) Representative flow cytometry analysis of Day 72 RUES2-STAG DP and SP iAEC2s sorted for outgrowth studies in 3D Matrigel after 18 days; $n = 3$, biological replicates separated at Day 0. (*D*) Representative stitched fluorescent images of Day 90 RUES2-STAG epithelial spheres 18 days after DP sort (left) showing smaller and brighter reporter DP spheres and SP sort (right) showing dimmer and fewer reporter DP spheres. Scale bars, 1.2 mm. (*E*) Representative flow cytometry analysis of Day 90 RUES2-STAG epithelial spheres 18 days after DP

to our scRNA-seq data (Figure 1), further support heterogeneity in the iAEC2 population evident as two populations of cells that are definable by an inverse relationship between proliferation genes and surfactant- or maturation-related genes, including *SFTPC*.

### ABCA3:GFP SP Population Is a More Clonogenic and Proliferative Subset of iAEC2s

An emerging literature has suggested heterogeneity in AEC2s *in vivo* in mouse lungs, during fetal lung development, adult homeostasis, or adult lung injury. For example, a distinct subset of AEC2s *in vivo* was recently reported and defined based on proliferative/progenitor potential, expression of the cell surface marker Tm4sf1, or high Wnt signaling levels (13–15). Hence, we profiled these same characteristics in our DP and SP iAEC2 populations, quantifying expression levels of Wnt signaling targets, *AXIN2, LEF1*, and *NKD1*, as well as *TM4SF1*, finding enrichment in SP cells (Figure 3J), with *NKD1* and *TM4SF1* represented in the top 50 significantly upregulated genes in SP cells (Figure 3H). Gene set enrichment analysis also revealed a trend toward Wnt/β-catenin signaling enrichment in SP cells although above our FDR $< 0.05$ cutoff (FDR = 0.109, data not shown). This trend is consistent with publications from our group (17) and others (13, 15) that Wnt signaling is higher in AEC2 progenitors and downregulation of Wnt signaling promotes the maturation of primary and iPSC-derived AEC2s.

Because our results suggested SP cells may be a more progenitor-like, proliferative population than DP cells, we performed a series of functional studies to test this possibility. First, we performed EdU incorporation assays to formally quantify proliferation in each iPSC-derived population (Figures 4A–4C). We found that 40.13% ± 6.64% of ABCA3:GFP SP cells stained EdU positive whereas the SFTPC$^{tdTomato}$/ABCA3:GFP DP population had significantly lower EdU incorporation (29.8% ± 7.83%; Figures 4B and 4C).

Next, to see whether measurable differences in proliferation translated to functional differences in epithelial sphere outgrowth or clonogenicity, we measured colony-forming efficiencies from single-cell suspensions prepared from enzymatically digested spheres that first underwent sorting on Day 29 to purify double-reporter-positive, single-ABCA3:GFP-positive, and reporter DN populations. We have previously published that epithelial sphere outgrowth in these conditions is clonal, arising from the proliferation of single cells (17, 18). After 12 days of clonal outgrowth, we found that a smaller number of colonies (62 ± 2.83) arose from the DP sorted outgrowth compared with the ABCA3:GFP SP sorted outgrowth (77.33 ± 5.13) (Figures 4D and 4E). Moreover, measurements of colony size revealed larger colony diameter in the ABCA3:GFP SP sorted outgrowth compared with the outgrowth from the reporter DP population (Figure 4F), further suggesting increased proliferation of SP compared with DP iAEC2s. These results were consistent with fluorescence imaging of a mixed population of iAEC2s expressing both populations in which the dual-reporter-positive spheres were smaller than the single-ABCA3:GFP-positive spheres (data not shown and Figure 5).

### SFTPC$^{tdTomato}$/ABCA3:GFP DP and ABCA3:GFP SP iAEC2s Can Repopulate Each Other and Are Capable of Surfactant Processing

Given the higher levels of transcripts and enrichment of pathways associated with surfactant metabolism in the DP populations (Figure 3), we first considered the possibility that GFP+ SP cells represent a more progenitor-like and less mature subset of iAEC2s that are potentially limited in their capacity to process surfactant proteins, a hallmark function of mature AEC2s, compared with the DP population. To assess this possibility, we sorted pure populations of DP, SP, and DN populations and tested their ability to process surfactant precursor proteins (proSFTPB and proSFTPC; Day 186 of differentiation) compared with sorted Day 15 NKX2-1+ lung epithelial progenitors. *In vivo*, proSFTPB is processed from its 42-kD precursor protein to its 8-kD mature form only in the presence of functional multivesicular bodies and LBs in post–Week 24 gestational human AEC2s (35). Western blot analysis of the sorted samples using an antibody specific to processed proSFTPB intermediates, including the final 8-kD mature peptide (36), revealed that whereas the DN sorted cells and the Day 15 lung epithelial progenitor cells expectedly expressed little to no mature 8-kD SFTPB protein, both the DP and SP sorted cells were able to process mature SFTPB, albeit at slightly higher levels in the DP sorted cells (Figures 5A and 5B). Next, similar analysis using antibodies specific to proSFTPC to detect processing of the 21-kD precursor peptide through 16-kD and 6- to 10-kD proSFTPC intermediates (37) confirmed 6- to 10-kD proSFTPC processing in both DP and SP populations but none in DN or Day 15 controls, with a slightly higher level of processed proSFTPC in the DP than in the SP population (Figures 5A and 5B). Combined, these data indicate that both the DP and SP cells are iAEC2s that can functionally process surfactant proteins with slightly higher levels of processed SFTPB and SFTPC in the DP population consistent with a more mature cell state.

Because both the DP and SP populations appeared to have surfactant processing function, even if quantitatively different in capacity, we next considered the possibility that SP and DP cells represented interchangeable states of iAEC2s rather than distinct populations of immature developmental precursors and more mature progeny. To determine whether each sorted population would retain their original reporter expression or repopulate other populations, Day 44 spheres were sorted into pure SFTPC$^{tdTomato}$/ABCA3:GFP DP versus ABCA3:GFP SP populations for replating and outgrowth (Figure 5C). After 18 days of

**Figure 5.** (*Continued*). sort (left) and SP sort (right). (*F*) Dot plots showing percentage of total cell population expressing both reporters (DP) and single ABCA3:GFP reporter in DP sorted outgrowth (left) and SP sorted outgrowth (right). Mean ± SE, $n = 3$ biological replicated separated at Day 0. (*G*) Dot plots showing mean fluorescent intensity measurements of SFTPC$^{tdTomato}$ (left) and ABCA3:GFP (right) of DP and SP sorted outgrowth. ****$P < 0.0001$. (*H*) Left: representative flow cytometry analysis of Day 30 BU3$^{ABCA3:GFP}$ cells stained with CPM antibody and sorted into ABCA3:GFP/CPM DP (*i*), CPM SP (*ii*), and CPM/ABCA3:GFP DN (*iii*). Right: Flow cytometry analysis for NKX2-1 intracellular staining of *i–iii* sorted cells showing enrichment of NKX2-1 in ABCA3:GFP/CPM DP (*i*) and CPM SP cells (*ii*) but not in ABCA3:GFP/CPM DN cells (*iii*). a.u. = arbitrary unit; MFI = mean fluorescent intensity; ns = not significant, by unpaired, two-tailed Student's *t* test.

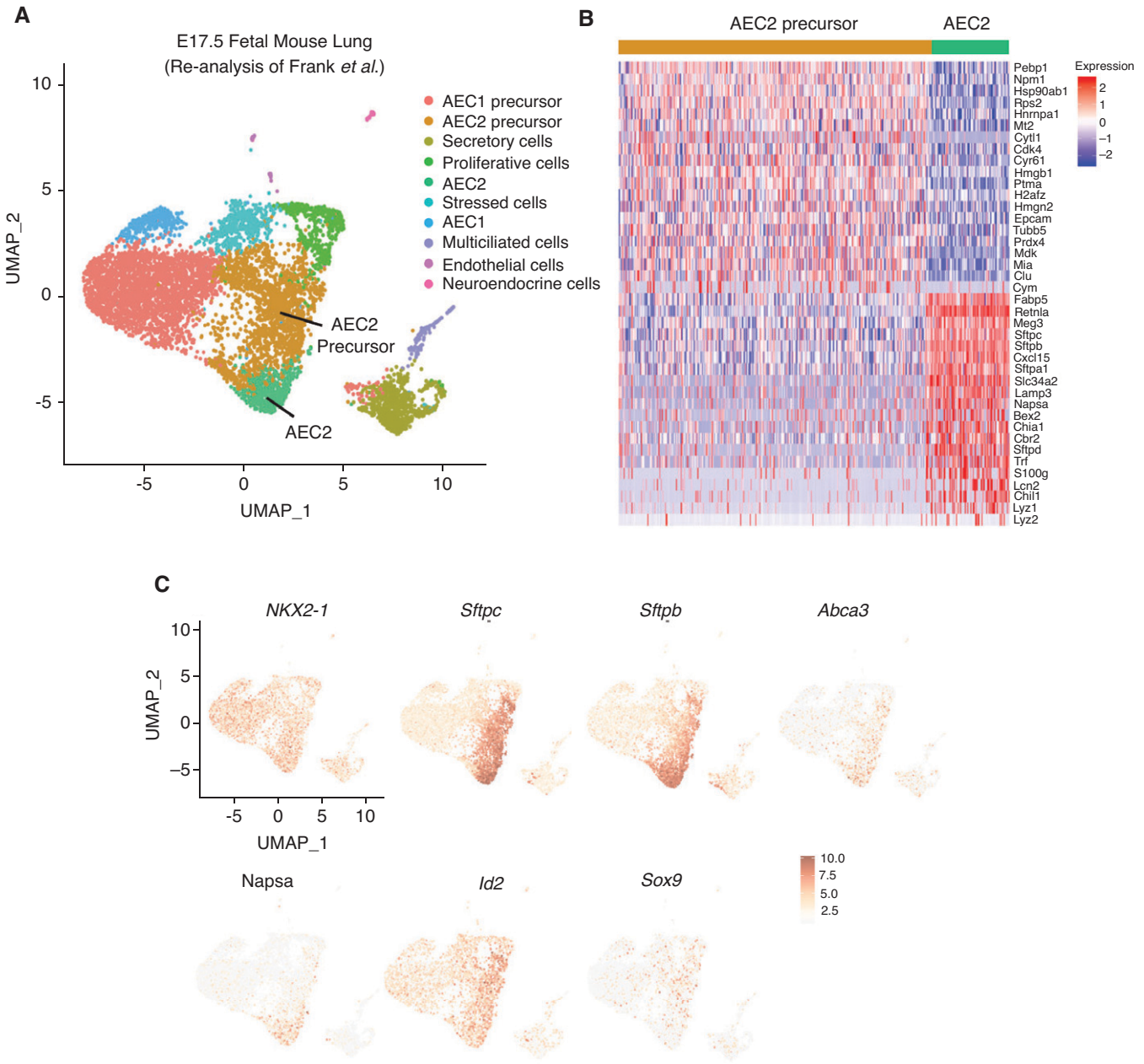


**Figure 6.** AEC2 heterogeneity in SFTPC expression exists in developing mouse and human lungs and in adult normal and diseased alveolar epithelium. (*A*) UMAP clustering of 7,106 Nkx2.1+ epithelial cells from E17.5 mouse lungs (Frank and colleagues, 2019). (*B*) Heatmap comparing relative gene expression of top differentially expressed genes in AEC2 precursor and AEC2 populations identified in Frank and colleagues (2019). (Top 20 upregulated and top 20 downregulated genes [ranked by fold change (FC); filtered by FDR < 0.05] between the two cell clusters indicated in *A* are shown). (*C*) UMAP overlay of AEC2 and distal bud tip markers. (*D*) Violin plots showing normalized expression for indicated genes associated with AEC2 and distal bud tip markers across precursor and AEC2 cell types identified in Frank and colleagues (2019). (*E*) Bar graphs showing Enrichr analysis of indicated collections of gene sets using the top 70 upregulated transcripts in AEC2 or AEC2 precursor clusters. Relevant pathways are highlighted in bold. (*F*) Heatmap of top 50 nonredundant differentially expressed genes in distal lung epithelial cells at three time points of human gestation (Weeks 11.5, 15, and 18; ranked by FC and filtered by FDR < 0.05), reanalyzed from single-cell RNA sequencing data sets previously published in Miller and colleagues (2020). Corresponding violin plots showing heterogeneity in normalized mRNA levels of SFTPC, SFTPB, and SOX9 mRNA expression levels (but not NKX2-1) for each time point. (*G*) Violin plots showing normalized expression levels of SFTPC mRNA in adult human quiescent AEC2s (AT2), proliferating distal epithelial cells, and transitional AEC2s (transitional AT2) from healthy ($n = 5$) or diseased (interstitial lung disease; $n = 6$) human donor lungs, based on expression levels in the three cell populations as originally defined by Habermann and colleagues (2020), or based on reanalysis (*H*) of the adult human AEC2s defined by Reyfman and colleagues (2019) from healthy ($n = 8$) or diseased ($n = 8$) adult human lungs.

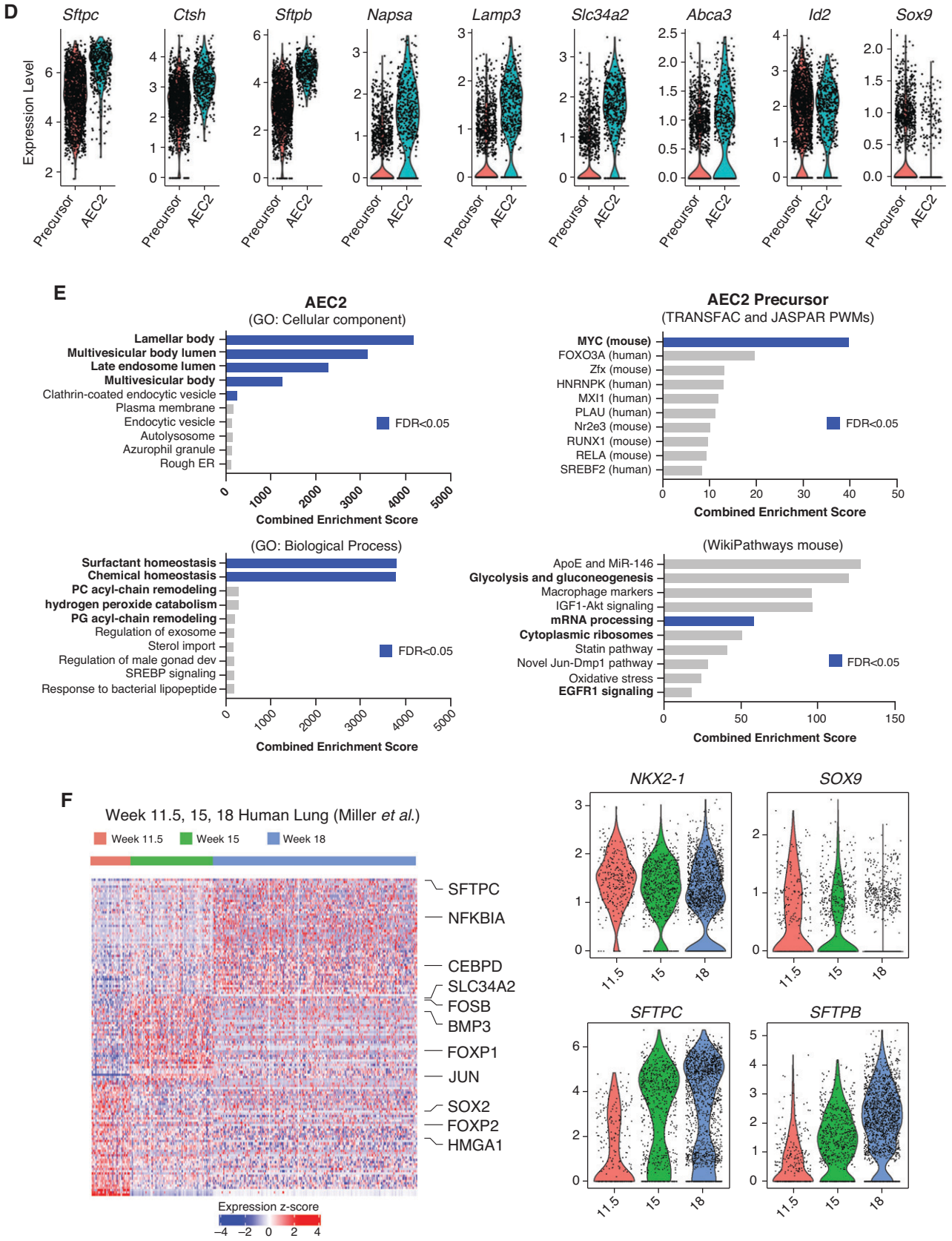


**Figure 6.** (*Continued*).

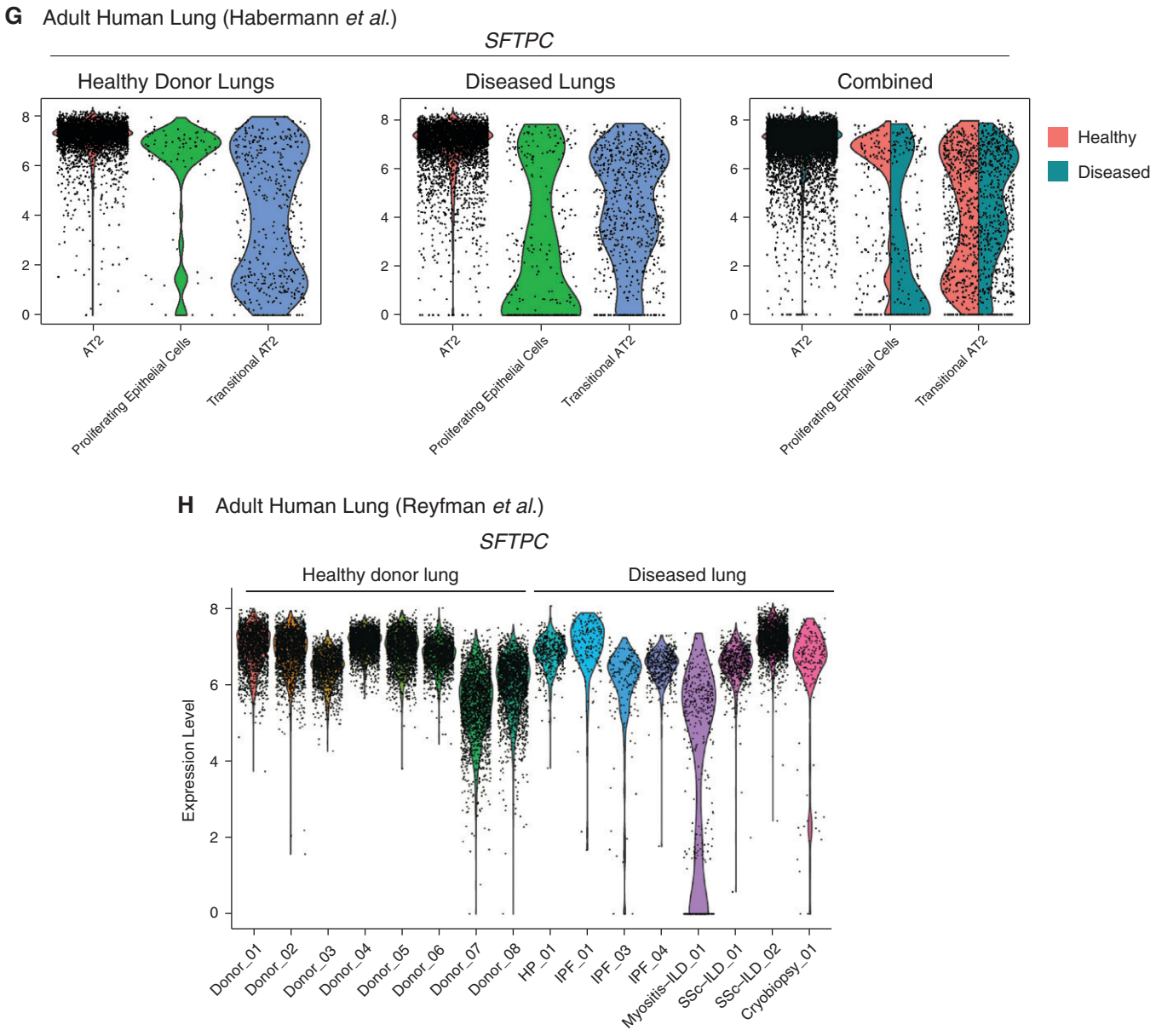


**Figure 6.** (*Continued*).

outgrowth (Figures 5D and 5E), the DP sorted cells had retained dual-reporter expression in 50.6% ± 11.23% of the cells. From the same outgrowth, single-ABCA3:GFP-positive population also emerged, making up a fraction (18.27% ± 3.73%) of the overall population. In the outgrowth from ABCA3:GFP SP sorted cells, we found retention of the original SP population phenotype in a subset of cells (35.67% ± 3.78% of the overall population) and an emergence of the $SFTPC^{tdTomato}$/ABCA3:GFP DP population in 6.29% ± 0.65% of the population. Notably, although the mean fluorescence intensity of the ABCA3:GFP+ populations did not change between the two sorted outgrowths, the mean tdTomato signal was significantly lower in the ABCA3:GFP SP sorted population, suggestive of either lower *SFTPC* locus activity or recent activation of the locus (Figures 5D, 5E, and 5G). Taken together, these data demonstrate the propensity for each sorted population to self-renew, maintaining the DP or SP phenotype of the parent population, but also some ability to repopulate other reporter-positive populations, suggesting some interchangeability.

Notably, in the reporter sorted outgrowths, we also observed a loss of all reporter expression in 27.73% ± 15.18% and 54.6% ± 3.65% of the sorted outgrowths of DP and ABCA3:GFP SP cells, respectively, consistent with either instability in some iPSC-derived "lung" cells over time (19, 28) or reversion of cells to a less mature lung progenitor phenotype lacking in AEC2 gene expression. Regardless, these trends suggest that more mature cells (identified by DP reporter expression) are more stable in culture in terms of their iAEC2 phenotype than SP cells, which appear more prone to lose their AEC2-specific reporter expression.

Furthermore, these profiles further emphasize heterogeneity of the iAEC2 population at late stages of directed differentiation.

Finally, to reevaluate cellular heterogeneity at early stages of the directed differentiation protocol, we repeated our directed differentiation protocol using the BU3$^{ABCA3:GFP}$ line, purifying NKX2-1+ Day 15 progenitors through CD47$^{hi}$/CD26$^{lo}$ sorting (20) for replating in distalizing medium (18). We then examined the resulting outgrowth for expression of ABCA3:GFP, NKX2-1, and CPM, a surrogate marker for NKX2-1 previously described by Gotoh and colleagues (38). At Day 30, the resulting outgrowth of epithelial spheres contained three populations based on levels of GFP and cell surface CPM protein expression (Figure 5H). Although ABCA3:GFP+ iAEC2s made up 19.97% ± 6.07% of the total CPM+ population, a subset of CPM-positive cells (37.87% ± 0.75%) did not express ABCA3:GFP (Figure 5H), suggesting the potential presence of NKX2-1+ lung epithelium that did not exhibit the iAEC2 phenotype (suggested by lack of ABCA3:GFP expression). NKX2-1 intracellular protein staining of each sorted population (Figure 5H, *i–iii*) confirmed positive NKX2-1 staining in both the ABCA3:GFP/CPM DP and CPM SP populations and negative NKX2-1 staining for the ABAC3:GFP/CPM DN population. Taken together with our findings from the bifluorescent BU3-NGST line profiled in Figure 1, these data provide additional evidence for the existence of heterogeneous NKX2-1+ lung epithelial subsets present at early stages (Day 30) of the distal directed differentiation protocol, some of which lack high levels of expression of AEC2 markers, such as ABCA3 and SFTPC.

### AEC2 Heterogeneity in Fetal and Adult Lung *In Vivo*

Based on our data demonstrating heterogeneity within early and late stages of iPSC-derived distalized lung epithelial cells, including an inverse relationship between maturation and proliferation and highlighted by a gradient of SFTPC-high to SFTPC-low–expressing cells, we considered whether heterogeneous subsets of AEC2s exist *in vivo* in mice and human lungs, either during embryological lung development or in adults. Thus, we reanalyzed available single-cell RNA-seq data sets of E17.5 mouse developing lung from Frank and colleagues (16), human developing lungs from Miller and colleagues (Weeks 11.5, 15, and 18 of gestation) (39), and adult normal or diseased human lungs from Habermann and colleagues and Reyfman and colleagues (40, 41). From the E17.5 mouse data set containing all distal NKX2-1+ epithelial cell types, we first identified two cell clusters described in Frank and colleagues (16) as AEC2s (481 cells) and AEC2 precursors (1,765 cells), both of which expressed canonical AEC2 transcripts (Figures 6A–6D). Direct comparison of the two AEC2 clusters revealed 512 DEGs (FDR < 0.05) and higher expression levels of canonical AEC2 transcripts in the AEC2 cluster compared with its precursor cells, including *Sftpc*, the number 12 most upregulated gene, as well as *Ctsh*, *Sftpb*, *Napsa*, *Sftpd*, *Lamp3*, *Slc34a2*, and *Abca3* (Figures 6B and 6D). In contrast, the precursor cells expressed higher levels of *Id2* and *Sox9*, transcripts associated with AEC2 progenitors (Figures 6C and 6D). These findings were consistent with the trend observed in reporter DP and SP PSC-derived AEC2 populations described above (Figures 3–5), with the DP population expressing higher levels of canonical AEC2 transcripts including *SFTPC* and *ABCA3*. Further analysis of the top 70 upregulated transcripts in the AEC2 cluster (FDR < 0.05) revealed gene enrichment associated with LB and LB precursor organelles (Enrichr GO: cellular compartment analysis; Figure 6E) and pathways associated with surfactant homeostasis and phosphatidylcholine acyl-chain remodeling (Enrichr GO: biological process), suggestive of a cellular trajectory toward maturing AEC2s and surfactant synthesis. In contrast, the top 70 upregulated transcripts in the AEC2 precursor cluster (FDR < 0.05) were enriched in genes regulated by MYC (Enrichr: TRANSFAC and JASPAR PWMs; Figure 6E) and pathways associated with increased cellular metabolism and anabolism such as mRNA processing, glycolysis, gluconeogenesis, cytoplasmic ribosomal proteins, and growth factor signaling pathways (Enrichr: WikiPathways mouse). Combined with lower expression levels of canonical AEC2 genes, including *Sftpc*, these enriched pathways are suggestive of a developmentally more immature AEC2 population.

To assess AEC2 heterogeneity in human developing or adult lungs, we next analyzed three recently published scRNA-seq data sets derived from multiple fetal, adult healthy donor, or diseased human lungs (35, 36, 39) by comparing *SFTPC* transcript levels in either the developing distal lung epithelium or the AEC2s that had been predefined in each data set. First, we examined all developing human fetal distal lung epithelia, defined based on cell clusters that express the canonical marker, *NKX2-1* (as detailed in the data supplement) and identified the top 50 nonredundant DEGs at three time points of human gestation (Weeks 11.5, 15, and 18; ranked by FC and filtered by FDR < 0.05; Figure 6F). *SFTPC* was among these top transcripts increasing over time in the distal lung epithelium, together with the transcription factor *CEBPD* and the mature AEC2 marker *SLC34A2* (Figure 6F). Next, we quantified expression levels of developing alveolar epithelial cell markers *SFTPC*, *SFTPB*, and *SOX9* at each gestational time point (represented as violin plots; Figure 6F). Consistent with *in vivo* mouse data (Figures 6A–6D), the developing human lung (Figure 6F) displayed increasing expression of AEC2 markers *SFTPC* and *SFTPB* and decreasing expression of alveolar progenitor marker SOX9 over developmental time. Notably, we observed a bimodal distribution of *SFTPC* mRNA expression (Figure 6F) across all time points, suggesting that SFTPC heterogeneity is detectable *in vivo* during human lung development.

To evaluate *SFTPC* expression heterogeneity in diseased versus healthy adult lung tissues, we examined the three distal lung epithelial populations predefined in Haberman and colleagues (40) as *1*) AEC2s (or AT2), *2*) proliferating epithelial cells, or *3*) "transitional AEC2s" (or "transitional AT2 s" based on expression of mixed AEC2 and AEC1 transcriptomic signatures). Consistent with current understanding of the quiescent nature of the healthy adult lung, the vast majority of AEC2s expressed high levels of *SFTPC* mRNA (Figure 6G) and no markers of proliferation (not shown), including in the diseased lung. However, we also identified *SFTPC* expression heterogeneity, evident in the transitional ACE2 subpopulations in both the diseased and healthy donor lungs as well as in the proliferative subset of the diseased lung (Figure 6G). We further found evidence of *SFTPC* expression variability within and between individuals by analyzing the scRNA-seq profiles of AEC2s previously captured by Reyfman and colleagues (41) from multiple healthy ($n = 8$) or diseased ($n = 8$) lungs (Figure 6H). Taken together, the above findings are suggestive of heterogeneity in terms of *SFTPC* expression

in ACE2 subpopulations in developing lungs as well as in normal and diseased adult human lungs, and these results are consistent with the more recently published scRNA-seq–based molecular lung atlas of Travaglini and colleagues (42), who reported heterogeneous *SFTPC* expression in human adult AEC2s, in which higher levels of SFTPC are accompanied by higher expression of maturation markers (e.g., *PGC* and *SFTPB*), whereas SFTPC-low AEC2s are enriched in Wnt signaling genes.

## Discussion

Our results provide a detailed characterization of human PSC-derived distal lung epithelium using fluorescent knock-in reporters for *NKX2-1, SFTPC*, and *ABCA3*. Using combinations of these fluorescent reporters across two PSC lines, we prepared single-cell and bulk RNA sequencing profiles of iAEC2s accompanied by functional and surfactant processing studies. These data demonstrate that iAEC2s exist *in vitro* in interchanging and heterogeneous populations largely distinguished based on their proliferative potential and maturation states, which appear to be inversely correlated.

Consistent with our previous reports (17, 18), we find the SFTPC$^{tdTomato}$ reporter to be a useful tool for identifying and isolating SFTPC+ iAEC2s from our distal lung differentiation protocol (17, 18, 21). However, the tdTomato reporter appears to selectively identify those cells expressing the highest levels of *SFTPC* mRNA, resulting in decreased sensitivity for identifying all iAEC2s. We demonstrate that targeting another canonical AEC2 locus, *ABCA3*, with a GFP fusion reporter enabled the identification of the majority of AEC2-like cells, including those missed by SFTPC$^{tdTomato}$, providing a solution for those wishing to track and purify all putative iAEC2s, regardless of SFTPC expression gradients. Based on transcriptomic signatures characterized by expression of canonical AEC2 transcripts in GFP+ cells, together with functional capacity to process proSFTPB and proSFTPC surfactant proteins, we find ABCA3:GFP+ cells to be worthy of the name "iAEC2s," regardless of whether they exhibit high versus low levels of select AEC2 transcripts such as SFTPC.

An important question to consider for the two AEC2-specific reporter-positive populations (DP, SP) described in this study is whether they are different functional subpopulations of iAEC2s as opposed to different proliferative states of the same interchangeable cell type. Although our data indicate that, when sorted on either DP or SP expression of AEC2 reporters, both populations maintained significant proliferative differences for at least 18 days, we also observed some degree of interchangeability, suggesting that each cell type is not entirely distinct.

In our study, we observe significant heterogeneity among ABCA3+ iAEC2s in terms of expression of not only the SFTPC$^{tdTomato}$ reporter but *SFTPC* mRNA levels as well. We have analyzed available scRNA-seq data sets procured from mouse and human distal lungs *in vivo* that might suggest similarly heterogeneous *SFTPC* expression levels in primary adult AEC2s. We found heterogenous *SFTPC* expression levels in developing AEC2s in fetal mice and in AEC2s of developing and adult humans with diseases affecting the distal lung, such as idiopathic pulmonary fibrosis. Although some features of our *in vitro* AEC2 model, such as the inverse relationship between proliferation and maturation, appear to recapitulate some patterns observed during *in vivo* development and disease, further work will be needed to define how the iPSC *in vitro* model might be used to study the mechanisms responsible for these processes and whether iAEC2 heterogeneity merely reflects the continual presence of incompletely matured fetal-like cells in current culture media or the interconversion of progenitor–progeny *ACE2* subtypes that might be recapitulated *in vivo* in the adult lung after injury. Our observations likely indicate that a greater level of AEC2 heterogeneity may occur in the context of lung injury and repair or in settings requiring proliferation of AEC2s. A potential pitfall associated with the AEC2 *in vivo* profiles published to date, including those reanalyzed by us here, may be that they rely on SFTPC as a marker for AEC2s either to trace cells that once highly expressed the SFTPC locus or to define AEC2s, possibly missing vital biology of otherwise functional AEC2s or their progenitors that minimally or only transiently expressed SFTPC, diminishing our understanding of true AEC2 heterogeneity *in vivo*. The work of Chapman and colleagues is one example of an alternative approach that has been used in mice to search for "lineage negative epithelial progenitors" or other cellular sources of AEC2s *in vivo* that need not rely on SFTPC lineage tracing (43, 44).

An inverse relationship between proliferative or progenitor states and maturation states in cell biology has been recognized in several contexts, such as in the intestinal epithelium and during neural development (45, 46). The presence of a distinct, more proliferative and Wnt-responsive alveolar progenitor-like cell has previously been described in the context of developing, homeostatic, and injured adult mouse lungs (13–15), and the role of downregulating Wnt in promoting AEC2 maturation, at the expense of proliferation, has been demonstrated previously by us and others (13, 17). To assess whether ABCA3:GFP SP iAEC2s might similarly represent more Wnt-responsive progenitor-like cells as described in mice, we compared DP and SP populations revealing enrichment of *TM4SF1* in the SP cells, a marker of mouse alveolar epithelial progenitors described by Zacharias and colleagues (14). Although we did not observe differences in *AXIN2* expression levels, we found significant enrichment of *NKD1*, a canonical Wnt target gene in the SP population, possibly indicative of increased Wnt signaling. These findings from our *in vitro* iAEC2 models are consistent with the recently published molecular atlas of the adult human lung in which Travaglini and colleagues found that adult human AEC2s expressing lower levels of SFTPC *in vivo* are enriched in Wnt target genes whereas those expressing higher levels of SFTPC are enriched in expression of maturation markers, *PGC* and *SFTPB* (42). Although a healthy quiescent lung may have a more homogeneous population of AEC2s overall, it remains possible that during lung injury or lung development, alveolar progenitor cells, driven in part by Wnt signals, gain functional, proliferative advantages critical for lung regeneration or development. Additional signaling pathways enriched in the ABCA3:GFP+ SP

population, such as KRAS signaling in our studies, may similarly drive a proliferative, progenitor ACE2 state at the expense of maturation, and thus, it may be informative to determine whether diseased states, such as lung adenocarcinoma, which are known to exhibit hyperactive KRAS signaling, display this phenotype in response to oncogenic KRAS mutations arising in AEC2s or their progeny, as was recently observed in models of early lung adenocarcinoma using mouse primary AEC2s and human iAEC2s (47). It will also be important to test whether SFTPC-high versus -low AEC2s differ in their potential to form AEC1s, a capacity that cannot be easily quantified currently in most *in vitro* systems using either iAEC2s or primary human AEC2s cultured as alveolospheres (1), because they appear to inefficiently give rise to AEC1s in our hands (18).

Taken together, our studies identifying multiple subsets of iAEC2s in our human *in vitro* model system highlight the importance of relying not solely on SFTPC but rather a broad spectrum of AEC2 markers to study the potentially heterogeneous population of AEC2s associated with development or diseased states. Moreover, in combination with our iAEC2 model, the visualization of intracellular LBs highlighted by our ABCA3:GFP fusion reporter system is likely to aid in the study of various pulmonary diseases, including those that affect lamellar body morphology, such as the childhood interstitial lung disease caused by *ABCA3* mutations or Hermansky-Pudlak syndrome. ■

**Author disclosures** are available with the text of this article at www.atsjournals.org.

**Acknowledgment:** The authors thank all members of the Kotton Laboratory for insightful discussions. They are grateful for the technical assistance of Ashley LeClerc and Yuriy Alekseyev of the Boston University School of Medicine (BUSM) Microarray and Sequencing Core Facility and Brian R. Tilton of the BUSM Flow Cytometry Core. For facility management, they thank Greg Miller, CReM (Center for Regenerative Medicine) Laboratory Manager, and Marianne James, CReM iPSC Core Manager.

## References

1. Barkauskas CE, Cronce MJ, Rackley CR, Bowie EJ, Keene DR, Stripp BR, *et al.* Type 2 alveolar cells are stem cells in adult lung. *J Clin Invest* 2013;123:3025–3036.
2. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* 2014;507:190–194.
3. Williams MC, Mason RJ. Development of the type II cell in the fetal rat lung. *Am Rev Respir Dis* 1977;115:37–47.
4. Evans MJ, Cabral LJ, Stephens RJ, Freeman G. Transformation of alveolar type 2 cells to type 1 cells following exposure to NO2. *Exp Mol Pathol* 1975;22:142–150.
5. Adamson IY, Bowden DH. The type 2 cell as progenitor of alveolar epithelial regeneration. A cytodynamic study in mice after exposure to oxygen. *Lab Invest* 1974;30:35–42.
6. Nogee LM, Garnier G, Dietz HC, Singer L, Murphy AM, deMello DE, *et al.* A mutation in the surfactant protein B gene responsible for fatal neonatal respiratory disease in multiple kindreds. *J Clin Invest* 1994;93:1860–1863.
7. Nogee LM, Dunbar AE III, Wert SE, Askin F, Hamvas A, Whitsett JA. A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *N Engl J Med* 2001;344:573–579.
8. Whitsett JA, Wert SE, Weaver TE. Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease. *Annu Rev Med* 2010;61:105–119.
9. Shulenin S, Nogee LM, Annilo T, Wert SE, Whitsett JA, Dean M. ABCA3 gene mutations in newborns with fatal surfactant deficiency. *N Engl J Med* 2004;350:1296–1303.
10. Nogee LM. Genetic basis of children's interstitial lung disease. *Pediatr Allergy Immunol Pulmonol* 2010;23:15–24.
11. Ehrhardt C, Kim KJ, Lehr CM. Isolation and culture of human alveolar epithelial cells. *Methods Mol Med* 2005;107:207–216.
12. Dobbs LG. Isolation and culture of alveolar type II cells. *Am J Physiol* 1990;258:L134–L147.
13. Frank DB, Peng T, Zepp JA, Snitow M, Vincent TL, Penkala IJ, *et al.* Emergence of a wave of Wnt signaling that regulates lung alveologenesis by controlling epithelial self-renewal and differentiation. *Cell Rep* 2016;17:2312–2325.
14. Zacharias WJ, Frank DB, Zepp JA, Morley MP, Alkhaleel FA, Kong J, *et al.* Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* 2018;555:251–255.
15. Nabhan AN, Brownfield DG, Harbury PB, Krasnow MA, Desai TJ. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* 2018;359:1118–1123.
16. Frank DB, Penkala IJ, Zepp JA, Sivakumar A, Linares-Saldana R, Zacharias WJ, *et al.* Early lineage specification defines alveolar epithelial ontogeny in the murine lung. *Proc Natl Acad Sci USA* 2019;116:4362–4371.
17. Jacob A, Morley M, Hawkins F, McCauley KB, Jean JC, Heins H, *et al.* Differentiation of human pluripotent stem cells into functional lung alveolar epithelial cells. *Cell Stem Cell* 2017;21:472–488.e10.
18. Jacob A, Vedaie M, Roberts DA, Thomas DC, Villacorta-Martin C, Alysandratos KD, *et al.* Derivation of self-renewing lung alveolar epithelial type II cells from human pluripotent stem cells. *Nat Protoc* 2019;14:3303–3332.
19. Yamamoto Y, Gotoh S, Korogi Y, Seki M, Konishi S, Ikeo S, *et al.* Long-term expansion of alveolar stem cells derived from human iPS cells in organoids. *Nat Methods* 2017;14:1097–1106.
20. Hawkins F, Kramer P, Jacob A, Driver I, Thomas DC, McCauley KB, *et al.* Prospective isolation of NKX2-1-expressing human lung progenitors derived from pluripotent stem cells. *J Clin Invest* 2017;127:2277–2294.
21. Hurley K, Ding J, Villacorta-Martin C, Herriges MJ, Jacob A, Vedaie M, *et al.* Reconstructed single-cell fate trajectories define lineage plasticity windows during differentiation of human PSC-derived distal lung progenitors. *Cell Stem Cell* 2020;26:593–608.e8.
22. Ikonomou L, Herriges MJ, Lewandowski SL, Marsland R III, Villacorta-Martin C, Caballero IS, *et al.* The *in vivo* genetic program of murine primordial lung epithelial progenitors. *Nat Commun* 2020;11:635.
23. Khoor A, Stahlman MT, Gray ME, Whitsett JA. Temporal-spatial distribution of SP-B and SP-C proteins and mRNAs in developing respiratory epithelium of human lung. *J Histochem Cytochem* 1994;42:1187–1199.
24. Wert SE, Glasser SW, Korfhagen TR, Whitsett JA. Transcriptional elements from the human SP-C gene direct expression in the primordial respiratory epithelium of transgenic mice. *Dev Biol* 1993;156:426–443.
25. McCauley KB, Hawkins F, Serra M, Thomas DC, Jacob A, Kotton DN. Efficient derivation of functional human airway epithelium from pluripotent stem cells via temporal regulation of Wnt signaling. *Cell Stem Cell* 2017;20:844–857.e6.
26. Foster C, Aktar A, Kopf D, Zhang P, Guttentag S. Pepsinogen C: a type 2 cell-specific protease. *Am J Physiol Lung Cell Mol Physiol* 2004;286:L382–L387.
27. Stahlman MT, Besnard V, Wert SE, Weaver TE, Dingle S, Xu Y, *et al.* Expression of ABCA3 in developing lung and other tissues. *J Histochem Cytochem* 2007;55:71–83.
28. Cheong N, Zhang H, Madesh M, Zhao M, Yu K, Dodia C, *et al.* ABCA3 is critical for lamellar body biogenesis in vivo. *J Biol Chem* 2007;282:23811–23817.
29. Mulugeta S, Gray JM, Notarfrancesco KL, Gonzales LW, Koval M, Feinstein SI, *et al.* Identification of LBM180, a lamellar body limiting

membrane protein of alveolar type II cells, as the ABC transporter protein ABCA3. *J Biol Chem* 2002;277:22147–22155.
30. Ban N, Matsumura Y, Sakai H, Takanezawa Y, Sasaki M, Arai H, *et al.* ABCA3 as a lipid transporter in pulmonary surfactant biogenesis. *J Biol Chem* 2007;282:9628–9634.
31. Yamano G, Funahashi H, Kawanami O, Zhao LX, Ban N, Uchida Y, *et al.* ABCA3 is a lamellar body membrane protein in human lung alveolar type II cells. *FEBS Lett* 2001;508:221–225.
32. Korogi Y, Gotoh S, Ikeo S, Yamamoto Y, Sone N, Tamai K, *et al. In vitro* disease modeling of Hermansky-Pudlak syndrome type 2 using human induced pluripotent stem cell-derived alveolar organoids. *Stem Cell Reports* 2019;12:431–440.
33. Haagsman HP, van Golde LMG. Synthesis and assembly of lung surfactant. *Annu Rev Physiol* 1991;53:441–464.
34. Agassandian M, Mallampalli RK. Surfactant phospholipid metabolism. *Biochim Biophys Acta* 2013;1831:612–625.
35. Brasch F, Johnen G, Winn-Brasch A, Guttentag SH, Schmiedl A, Kapp N, *et al.* Surfactant protein B in type II pneumocytes and intra-alveolar surfactant forms of human lungs. *Am J Respir Cell Mol Biol* 2004;30: 449–458.
36. Beers MF, Bates SR, Fisher AB. Differential extraction for the rapid purification of bovine surfactant protein B. *Am J Physiol* 1992;262: L773–L778.
37. Beers MF, Kim CY, Dodia C, Fisher AB. Localization, synthesis, and processing of surfactant protein SP-C in rat lung analyzed by epitope-specific antipeptide antibodies. *J Biol Chem* 1994;269:20318–20328.
38. Gotoh S, Ito I, Nagasaki T, Yamamoto Y, Konishi S, Korogi Y, *et al.* Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells. *Stem Cell Reports* 2014;3: 394–403.
39. Miller AJ, Yu Q, Czerwinski M, Tsai YH, Conway RF, Wu A, *et al. In vitro* and *in vivo* development of the human airway at single-cell resolution. *Dev Cell* 2020;54:818.
40. Habermann AC, Gutierrez AJ, Bui LT, Yahn SL, Winters NI, Calvi CL, *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* 2020;6:eaba1972.
41. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, *et al.* Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am J Respir Crit Care Med* 2019;199:1517–1536.
42. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 2020;587:619–625.
43. Kathiriya JJ, Brumwell AN, Jackson JR, Tang X, Chapman HA. Distinct airway epithelial stem cells hide among club cells but mobilize to promote alveolar regeneration. *Cell Stem Cell* 2020;26: 346–358.e4.
44. Vaughan AE, Brumwell AN, Xi Y, Gotts JE, Brownfield DG, Treutlein B, *et al.* Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature* 2015;517:621–625.
45. Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 2007;449:1003–1007.
46. Homem CC, Repic M, Knoblich JA. Proliferation control in neural stem and progenitor cells. *Nat Rev Neurosci* 2015;16: 647–659.
47. Dost AFM, Moye AL, Vedaie M, Tran LM, Fung E, Heinze D, *et al.* Organoids model transcriptional hallmarks of oncogenic KRAS activation in lung epithelial progenitor cells. *Cell Stem Cell* 2020;27: 663–678.e8.