The Journal of Clinical Investigation RESEARCH ARTICLE

Prospective isolation of NKX2-1–expressing human lung progenitors derived from pluripotent stem cells

Finn Hawkins,[1,2] Philipp Kramer,[3] Anjali Jacob,[1,2] Ian Driver,[4] Dylan C. Thomas,[1] Katherine B. McCauley,[1,2] Nicholas Skvir,[1] Ana M. Crane,[3] Anita A. Kurmann,[1,5] Anthony N. Hollenberg,[5] Sinead Nguyen,[1] Brandon G. Wong,[6] Ahmad S. Khalil,[6,7] Sarah X.L. Huang,[3,8] Susan Guttentag,[9] Jason R. Rock,[4] John M. Shannon,[10] Brian R. Davis,[3] and Darrell N. Kotton[1,2]

[1]Center for Regenerative Medicine, and [2]The Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA. [3]Center for Stem Cell and Regenerative Medicine, Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas, USA. [4]Department of Anatomy, UCSF, San Francisco, California, USA. [5]Division of Endocrinology, Diabetes and Metabolism, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts, USA. [6]Department of Biomedical Engineering and Biological Design Center, Boston University, Boston, Massachusetts, USA. [7]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts, USA. [8]Columbia Center for Translational Immunology & Columbia Center for Human Development, Columbia University Medical Center, New York, New York, USA. [9]Department of Pediatrics, Monroe Carell Jr. Children's Hospital, Vanderbilt University, Nashville, Tennessee, USA. [10]Division of Pulmonary Biology, Cincinnati Children's Hospital, Cincinnati, Ohio, USA.

It has been postulated that during human fetal development, all cells of the lung epithelium derive from embryonic, endodermal, NK2 homeobox 1–expressing (NKX2-1$^+$) precursor cells. However, this hypothesis has not been formally tested owing to an inability to purify or track these progenitors for detailed characterization. Here we have engineered and developmentally differentiated NKX2-1^{GFP} reporter pluripotent stem cells (PSCs) in vitro to generate and isolate human primordial lung progenitors that express NKX2-1 but are initially devoid of differentiated lung lineage markers. After sorting to purity, these primordial lung progenitors exhibited lung epithelial maturation. In the absence of mesenchymal coculture support, this NKX2-1$^+$ population was able to generate epithelial-only spheroids in defined 3D cultures. Alternatively, when recombined with fetal mouse lung mesenchyme, the cells recapitulated epithelial-mesenchymal developing lung interactions. We imaged these progenitors in real time and performed time-series global transcriptomic profiling and single-cell RNA sequencing as they moved through the earliest moments of lung lineage specification. The profiles indicated that evolutionarily conserved, stage-dependent gene signatures of early lung development are expressed in primordial human lung progenitors and revealed a CD47^{hi}CD26^{lo} cell surface phenotype that allows their prospective isolation from untargeted, patient-specific PSCs for further in vitro differentiation and future applications in regenerative medicine.

Introduction

Little is known about the early stages of human lung development, preventing an understanding of whether successful healing from adult lung injury involves recapitulation of embryonic mechanisms and limiting approaches for generating lung progenitors from pluripotent cells in vitro. Inbred mouse models have begun to define the mechanisms regulating lung specification and patterning, but how this breadth of work applies to human lung development is unknown. Current claims suggest that all cells of the postnatal mammalian lung epithelium derive from embryonic *NKX2-1*$^+$ progenitors; however, scant literature exists formally testing this hypothesis either in mice or humans. Support for this paradigm derives mainly from the observation that *Nkx2-1* is the first gene locus known to be activated in cells of the endodermal lung primordium (1, 2). NKX2-1–null mutant mice have hypoplastic lungs that fail to mature, and human children with *NKX2-1* mutations develop respiratory insufficiency, hypothyroidism, and neurological impairment (3), but these observations do not necessarily indicate that all lung epithelial cells derive via NKX2-1$^+$ progenitor intermediates. Since lung lineage specification is thought to occur in relatively few endodermal cells during a narrow developmental time period in vivo, it has been difficult to gain access to these cells in human embryos or to follow their cell fate decisions in real time. Hence, we sought to interrogate the earliest moments of human lung lineage specification by engineering an in vitro system that would allow the isolation and differentiation of pure populations of NKX2-1$^+$ putative human lung progenitors. Given the known capacity of mouse pluripotent stem cells (PSCs) to form all cell types, including lung lineages, after transfer into mouse blastocyst embryos and the known broad differentiation repertoire of human PSCs in vitro, we based this system on the in vitro differentiation of PSCs.

Initial published attempts at deriving lung epithelium from PSCs relied on the presence of drug-resistance genes or used incompletely defined media (4–6), resulting in inefficient induction of selected lung markers. Subsequently, a number of groups, including our own, had more success by broadly attempting to recapitulate the key milestones of embryonic lung development in vitro through the exogenous addition of sequential combina-

---

Authorship note: F. Hawkins and P. Kramer are co–first authors. B.R. Davis and D.N. Kotton are co–senior authors. A.A. Kurmann is deceased.

Conflict of interest: The authors have declared that no conflict of interest exists.

Submitted: August 8, 2016; **Accepted:** March 2, 2017.

Reference information: *J Clin Invest.* 2017;127(6):2277–2294. https://doi.org/10.1172/JCI89950.

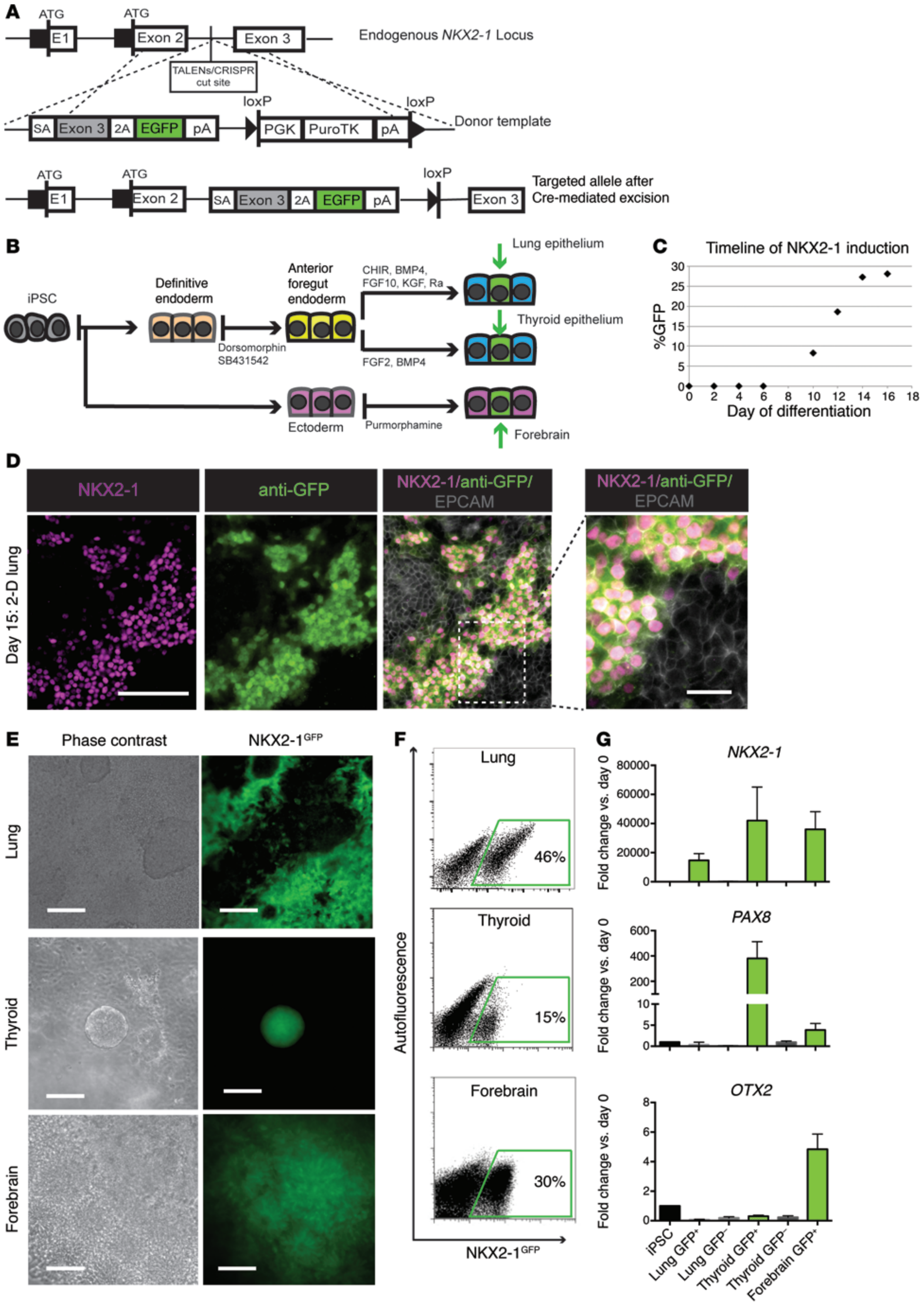

A
ATG
E1
ATG
Exon 2
Exon 3
Endogenous NKX2-1 Locus
TALENs/CRISPR cut site
loxP
loxP
SA
Exon 3
2A
EGFP
pA
PGK
PuroTK
pA
Donor template
ATG
E1
ATG
Exon 2
SA
Exon 3
2A
EGFP
pA
loxP
Exon 3
Targeted allele after Cre-mediated excision
B
iPSC
Definitive endoderm
Dorsomorphin SB431542
Anterior foregut endoderm
CHIR, BMP4, FGF10, KGF, Ra
FGF2, BMP4
Lung epithelium
Thyroid epithelium
Ectoderm
Purmorphamine
Forebrain
C
Timeline of NKX2-1 induction
%GFP
30
25
20
15
10
5
0
0
2
4
6
8
10
12
14
16
18
Day of differentiation
D
NKX2-1
anti-GFP
NKX2-1/anti-GFP/ EPCAM
NKX2-1/anti-GFP/ EPCAM
Day 15: 2-D lung
E
Phase contrast
NKX2-1GFP
Lung
Thyroid
Forebrain
F
Lung
46%
Thyroid
15%
Forebrain
30%
Autofluorescence
NKX2-1GFP
G
NKX2-1
Fold change vs. day 0
80000
60000
40000
20000
0
PAX8
Fold change vs. day 0
600
400
200
10
5
0
OTX2
Fold change vs. day 0
8
6
4
2
0
iPSC
Lung GFP+
Lung GFP–
Thyroid GFP+
Thyroid GFP–
Forebrain GFP+

Figure 1. Purification of human NKX2-1$^+$ lineages derived from ESCs/iPSCs using NKX2-1^{GFP} reporters. (**A**) Gene editing strategy based on TALENs or CRISPR technology to target a GFP reporter to the human *NKX2-1* locus to engineer NKX2-1^{GFP} iPSC/ESC lines. See also Supplemental Figure 1. (**B**) Schematic overview of in vitro directed differentiation of ESCs/iPSCs into NKX2-1$^+$ lineages: endodermal lung or thyroid epithelia versus ectodermal forebrain. (**C**) Representative timeline of GFP expression measured by flow cytometry during lung-directed differentiation (C17). (**D**) Immunostaining of day 15 lung-directed differentiation for NKX2-1, GFP (anti-GFP), and EPCAM. Scale bar (left panels): 100 μm. Right panel is a zoom in (white dashed-line box). Scale bar: 25 μm (C17). (**E**) Phase contrast and fluorescence microscopy of C17 iPSC-derived cells generated in the lung (day 15), thyroid (day 27), and forebrain (day 10) protocols. See also Supplemental Figure 2. (**F**) Flow cytometric analysis of lung (day 15), thyroid (day 17), and forebrain (day 14) protocols (C17). (**G**) Fold change, compared with day 0, of mRNA expression of sorted NKX2-1^{GFP+} and NKX2-1^{GFP-} cells from those time points by RT-qPCR; quantified as 2$^{(-\Delta\Delta CT)}$, *n* = 3 (C17).

tions of growth factors (7–11). By differentiating iPSCs into definitive endoderm, patterning this endoderm via inhibition of TGF-β and BMP signaling (7), and then adding various combinations of Wnts, FGFs, BMPs, and retinoic acid (Ra) these groups demonstrated the in vitro derivation of cultures expressing a broad array of lung epithelial markers. However, the characterization of the cells derived at different stages of these protocols suggested that heterogeneous cell types were present (12, 13). The most recently published directed-differentiation protocols describe variable efficiencies of induction of NKX2-1$^+$ cells from embryonic stem cells (ESCs) or induced PSCs (iPSCs) ranging from approximately 36% to 86% (9, 10, 14). Such heterogeneity limits the utility of these cultures for downstream applications and has caused uncertainty as to whether subsequent lung lineages derive directly from these early endodermal NKX2-1$^+$ precursors. To derive more mature lung cell types from iPSCs, some groups have employed prolonged in vitro cultures, murine kidney capsule or subcutaneous transplantations, or coculture with lung mesenchyme (LgM) (9–11, 14–16). These results suggest that at some point during differentiation of iPSCs, progenitor intermediates with competence for forming mature lung cells likely emerge. However, it has not previously been possible to isolate these cells for characterization or to properly test their differentiation repertoire.

To perform a detailed profiling of candidate human lung progenitors we targeted the *NKX2-1* human locus with a fluorescent reporter, enabling isolation of the earliest identifiable putative lung-lineage-committed cells derived from PSCs. In addition, these same reporter iPSC lines facilitate the derivation and purification of alternate developing human progenitors that express NKX2-1, such as endodermal thyroid–like and ectodermal forebrain–like lineages. After directed differentiation of these PSCs in defined media designed to promote lung rather than thyroid or forebrain development, we demonstrate that the NKX2-1$^+$ endodermal population is highly enriched in undifferentiated (primordial) progenitors that are competent at expressing a broad repertoire of lung epithelial marker genes, supporting the paradigm that the human lung epithelium derives from embryonic NKX2-1$^+$ progenitors. We provide population-based as well as single-cell global transcriptomic profiles that define developmental stage–specific gene signatures of iPSC-derived lung progenitors. These signatures suggest that an evolutionarily conserved lung developmental program exists in both mice and humans. Furthermore, these signatures reveal that NKX2-1$^+$ human lung progenitors can be prospectively isolated from patient-specific iPSCs based on the cell surface phenotype CD47^{hi}CD26^{lo}. Ultimately, the detailed characterization and purification of human lung progenitors presented here should provide access to an inexhaustible supply of these cells for disease modeling, future cell-based therapies, and basic developmental studies.

Results

An NKX2-1^{GFP} reporter enables purification of human lung, thyroid, and forebrain lineages. In order to generate a tool for the identification and purification of candidate human lung progenitors, we used gene editing technologies to target an enhanced green fluorescence reporter (*GFP*) gene to the endogenous human *NKX2-1* locus in multiple human PSC lines. Prior reports of targeting *GFP* to the *NKX2-1* locus in human PSCs for the derivation of forebrain lineages resulted in *NKX2-1* haploinsufficiency (17). Hence, pursuing a strategy designed to retain intact expression of targeted loci without haploinsufficiency (Figure 1A and Supplemental Figure 1A; supplemental material available online with this article; https://doi.org/10.1172/JCI89950DS1), we targeted an exon3-2A-GFP cassette to the second intron of *NKX2-1* using either transcription activator–like effector nucleases (TALENs) or CRISPR-Cas9 tools deployed in ESCs (H9) or in our previously published iPSC lines: cystic fibrosis patient–specific C17 iPSCs (18) and normal BU3 iPSCs (Figure 1 and Supplemental Figure 1) (19). The resulting NKX2-1^{GFP} reporter PSC clones (hereafter H9NKX2-1^{GFP}, C17NKX2-1^{GFP}, and BU3NKX2-1^{GFP}) demonstrated successful mono- and bi-allelic integration of the donor template by PCR (Supplemental Figure 1B). For further profiling we selected 1 homozygous targeted clone for each of the TALENs-targeted lines (H9 and C17) as well as 1 homozygous CRISPR-Cas9–targeted BU3 clone (Supplemental Figure 1, B and C).

To differentiate each targeted PSC line, we tested protocols previously developed by us and others for the in vitro directed differentiation of human PSCs into the 3 lineages known to express NKX2-1: neural/forebrain (20, 21), lung (7–9), or thyroid (19) (Figure 1B). Protocols for lung and thyroid both required generating anterior foregut–like endoderm followed by the addition of Chir99021 (CHIR), BMP4, KGF, FGF10, and Ra for lung versus BMP4 and FGF2 for thyroid. Consistent with prior publications (9), the percentage of NKX2-1$^+$ cells on day 15 of the lung-directed differentiation was typically 41% ± 21% (mean ± SD) when using the ESC line RUES2 (data not shown). However, when other cells lines including H9 and C17 were differentiated the percentage of NKX2-1–expressing cells was initially less than 1% (data not shown). Thus, we developed a methodology for the optimization of lung differentiation for each cell line by altering the duration of endoderm induction, the cell density of replated endoderm, and the duration of TGF-β/BMP inhibition in order to augment the percentage of NKX2-1$^+$ cells emerging by day 15 (Supplemental Figure 1H and Supplemental Figure 2, A and B). Employing the new lung differentiation protocol optimized for each clone, GFP expression was first detected between days 8 and 10 of differentiation and the percentage of GFP$^+$ cells peaked between days 12 and 16, with an efficiency of 25.6% ± 9.5% for C17NKX2-1^{GFP} (Figure 1, C and F), 29.5% ± 4.4% for BU3NKX2-1^{GFP}, and 19.8%

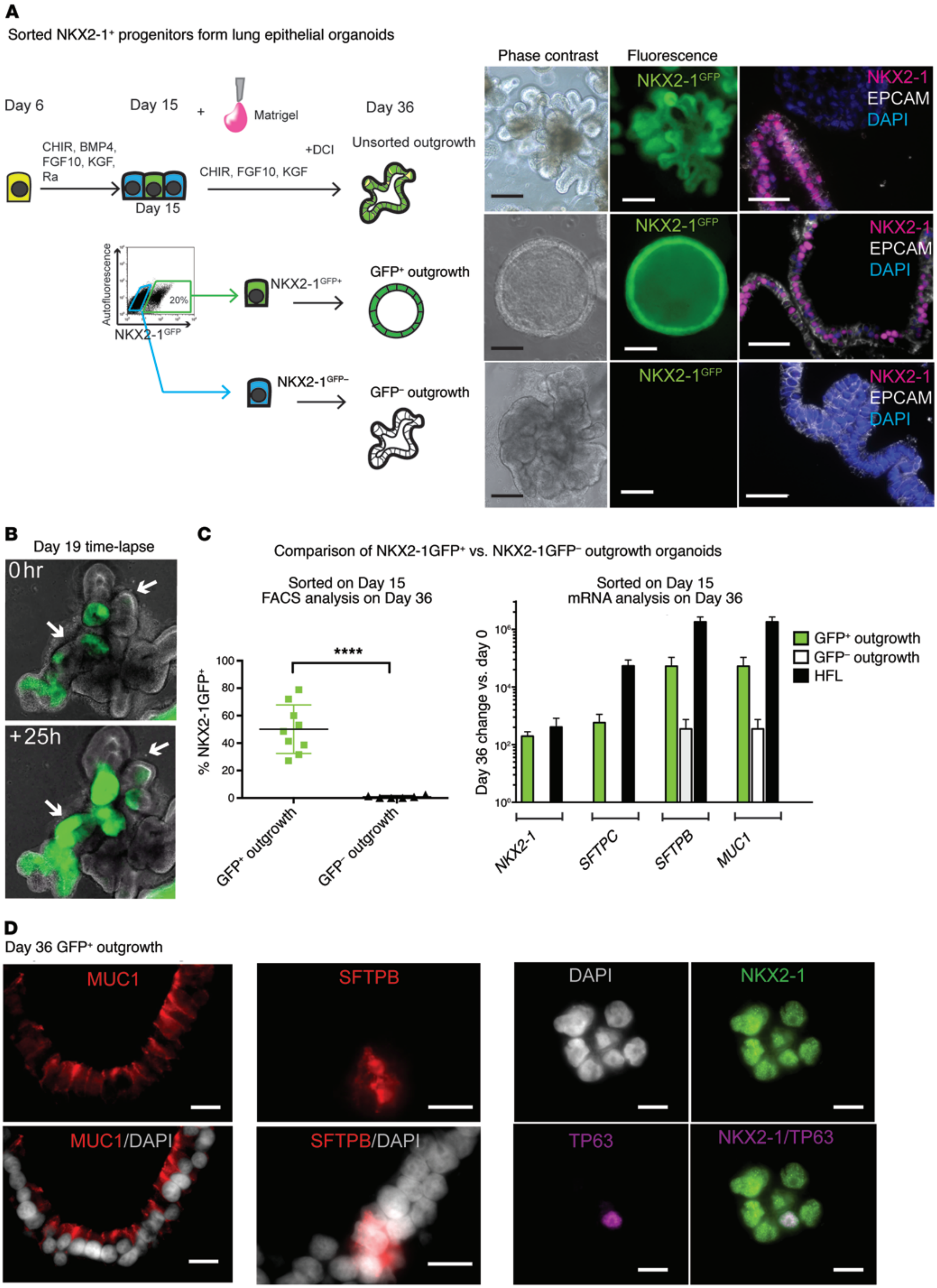

A
Sorted NKX2-1⁺ progenitors form lung epithelial organoids
Day 6
Day 15
+
Matrigel
Day 36
CHIR, BMP4, FGF10, KGF, Ra
Day 15
+DCI
CHIR, FGF10, KGF
Unsorted outgrowth
Autofluorescence
20%
NKX2-1GFP
NKX2-1GFP+
GFP⁺ outgrowth
NKX2-1GFP–
GFP⁻ outgrowth
Phase contrast
Fluorescence
NKX2-1GFP
NKX2-1
EPCAM
DAPI
B
Day 19 time-lapse
0 hr
+25h
C
Comparison of NKX2-1GFP⁺ vs. NKX2-1GFP⁻ outgrowth organoids
Sorted on Day 15
FACS analysis on Day 36
% NKX2-1GFP⁺
0
20
40
60
80
100
****
GFP⁺ outgrowth
GFP⁻ outgrowth
Sorted on Day 15
mRNA analysis on Day 36
Day 36 change vs. day 0
10⁰
10²
10⁴
10⁶
NKX2-1
SFTPC
SFTPB
MUC1
GFP⁺ outgrowth
GFP⁻ outgrowth
HFL
D
Day 36 GFP⁺ outgrowth
MUC1
MUC1/DAPI
SFTPB
SFTPB/DAPI
DAPI
NKX2-1
TP63
NKX2-1/TP63

Figure 2. Sorted iPSC-derived NKX2-1^{GFP+} cells exhibit lung progenitor potential and ability to form epithelial spheroids in 3D culture. (**A**) Schematic overview of organoid generation with representative phase contrast and GFP fluorescence microscopy images (day 25–28) as well as immunostaining (day 36) for NKX2-1 and EPCAM proteins. Cell nuclei are counterstained with DAPI. Each panel shows outgrowth in 3D culture of structures arising from iPSC-derived cells that were either unsorted or sorted on day 15 as GFP$^+$ versus GFP$^-$ populations (C17). Scale bars: 100 μm (left and center columns) and 20 μm (right column). (**B**) Time-lapse microscopy (merged GFP fluorescence and phase contrast) of unsorted organoids over 25 hours. Arrows indicate epithelial organoids undergoing induction of the GFP reporter in real time (C17). See also Supplemental Video file. (**C**) FACS quantification on day 36 of the percentage of cells expressing GFP in the outgrowth wells shown in **A** (C17). Data indicate individual biological replicates (squares and triangles) with mean ± SD, $n = 6$ biological replicates. ****$P \leq 0.0001$ by Student's t test. Fold change [RT-qPCR; $2^{(-\Delta\Delta CT)}$] in mRNA expression on day 36 compared with day 0 for each GFP$^+$ versus GFP$^-$ outgrowth compared with fetal lung control tissue ($n = 3$ biological replicates) (C17). (**D**) Immunostaining of GFP$^+$ outgrowth organoids on day 36 for MUC1, SFTPB, and TP63 (C17). Nuclei are counterstained with DAPI. Scale bars: 10 μm.

± 13.2% for H9NKX2-1^{GFP} (mean ± SD, data not shown). Day 15 of differentiation was selected for further characterization of GFP$^+$ cells in the lung protocol. Immunostaining for cytoplasmic GFP and nuclear NKX2-1 protein indicated faithful and specific expression of the GFP reporter in NKX2-1$^+$ cells (Figure 1D). We confirmed that NKX2-1 protein levels were not significantly perturbed by our targeting strategy by comparing homozygously targeted BU3 NKX2-1^{GFP} iPSCs with nontargeted parental control iPSCs (Supplemental Figure 1G). Sorting GFP$^+$ cells from each of the 3 differentiation protocols enriched for cells expressing *NKX2-1* mRNA (Figure 1, F and G). GFP$^+$ cells from thyroid and forebrain protocols were selectively enriched in expression of early lineage markers: *PAX8*, *HHEX*, and *FOXE1* for thyroid and *OTX2*, *OTX1*, *SOX1*, *PAX6*, and *SIX3* for forebrain (Figure 1G, Supplemental Figure 2, C–E and data not shown). The NKX2-1^{GFP+} cells isolated on day 15 in the lung protocol were *NKX2-1$^+$ PAX8$^-$* (Supplemental Figure 2E) and importantly lacked detectable expression of markers of lung epithelial differentiation (e.g., *SFTPC* or *SCGB1A1*; data not shown), raising the possibility that they may have undergone lung lineage specification but were still undifferentiated or primordial. Lung-specific progenitor markers or transcriptomic signatures are not known at this primordial developmental stage. Hence, to test the hypothesis that day 15 NKX2-1^{GFP+} cells represented lung epithelial progenitors, we next interrogated their competence for subsequently expressing markers of more differentiated lung epithelium.

Purified iPSC-derived NKX2-1^{GFP+} cells exhibit lung progenitor potential. Employing our human lung–directed differentiation protocol and the C17NKX2-1^{GFP} iPSC line, we tested whether differentiated lung epithelial cells derive directly from NKX2-1$^+$ progenitors. We sorted NKX2-1^{GFP+} versus NKX2-1^{GFP-} cells on day 15 and plated each population (vs. unsorted controls) in 3D Matrigel in serum-free media supplemented with factors we and others have previously shown to support lung epithelial differentiation (8, 9): CHIR, KGF, and FGF10 for 7 days followed by the addition of dexamethasone, cAMP, and IBMX until day 36 (Figure 2A). We observed the outgrowth of proliferating cell aggregates over the next 2 to 3 weeks (hereafter referred to as organoids) (Figure 2A). Unsorted day 15 cells plated as clumps gave rise to lobular organoids with GFP$^+$ and GFP$^-$ areas on day 36 (Figure 2A), and GFP$^+$ cells could be followed in real time in these unsorted cultures by time-lapse photography (Figure 2B and Supplemental Video 1). In contrast, when unsorted cells were plated as single-cell suspensions, simpler spherical organoids formed (data not shown). Immunostaining of the unsorted organoids demonstrated areas of monolayered NKX2-1$^+$EPCAM$^+$ epithelium surrounding an inner lumen but also areas and organoids that were NKX2-1$^-$ (Figure 2A). Sorted day 15 GFP$^+$ cells gave rise to GFP$^+$ aggregates in 3D culture, with 50.1% ± 17.6% (mean ± SD; $n = 6$ runs) of the progeny remaining GFP$^+$ by flow cytometry on day 36 (GFP$^+$ outgrowth) (Figure 2C). Sorted day 15 GFP$^-$ cells remained GFP$^-$ on day 36 (99.8% ± 0.2%) (GFP$^-$ outgrowth) (Figure 2C). The GFP$^+$ outgrowth formed predominantly NKX2-1$^+$EPCAM$^+$ spheroids, whereas the GFP$^-$ outgrowth formed EPCAM$^+$ and EPCAM$^-$ organoids that were uniformly NKX2-1$^-$ (Figure 2A). The sorted NKX2-1^{GFP+} progenitors on day 15 comprised the entirety of cells competent at subsequently expressing the lung-specific marker SFTPC by day 36 (Figure 2C and data not shown), suggesting this population contained lung progenitors. The GFP$^+$ outgrowth was also highly enriched for cells competent at expressing lung markers *SFTPB* and *MUC1*, although these markers are known to have less lung specificity than SFTPC (Figure 2C). Immunostaining confirmed discrete populations of cells expressing SFTPB and MUC1 proteins in the GFP$^+$ outgrowth (Figure 2D). In contrast, TP63$^+$ cells were present in both GFP$^+$ and GFP$^-$ outgrowths, but were more prevalent in GFP$^-$ outgrowths, suggesting TP63 is not a lung-specific marker in this system. Consistent with this interpretation, TP63$^+$ cells in the GFP$^+$ outgrowth coexpressed NKX2-1, whereas TP63$^+$ cells in the GFP$^-$ outgrowth did not express NKX2-1 (Figure 2D and Supplemental Figure 3, A and B). Reverse transcription quantitative PCR (RT-qPCR) confirmed significantly higher levels of *TP63* and the esophageal marker, *PITX1*, in the GFP$^-$ outgrowth, raising the possibility that the NKX2-1$^-$TP63$^+$ cells might represent alternative foregut derivatives, such as developing esophageal epithelium (Supplemental Figure 3A) (22).

Fetal LgM augments distal lung differentiation in iPSC-derived lung organoids. Having established that NKX2-1$^+$ primordial progenitors could be induced to upregulate markers of lung epithelial lineages without mesenchymal coculture support, we next sought to determine whether these progenitors might also respond to developmental cues provided by primary embryonic LgM. Lung epithelial-mesenchymal interactions are essential for lung epithelial growth, branching, and differentiation (23–25). For example, separating and recombining rat embryonic lung epithelium and mesenchyme (recombinants) has previously revealed both the importance of LgM and the stage-specific plasticity of the developing lung epithelium in response to mesenchymal signals (23). Hence, we asked whether iPSC-derived NKX2-1$^+$ cells are competent at responding to developing mouse LgM and if distal lung epithelial gene expression might be induced by distal LgM compared with bronchial mesenchyme or standard directed-differentiation conditions without mesenchyme (Figure 3A). Lung organoids generated from unsorted day 15 human iPSCs were cultured in 3D conditions until day 21. Either NKX2-1^{GFP+} or NKX2-1^{GFP-} areas were microdissected and recombined with E12 mouse lung

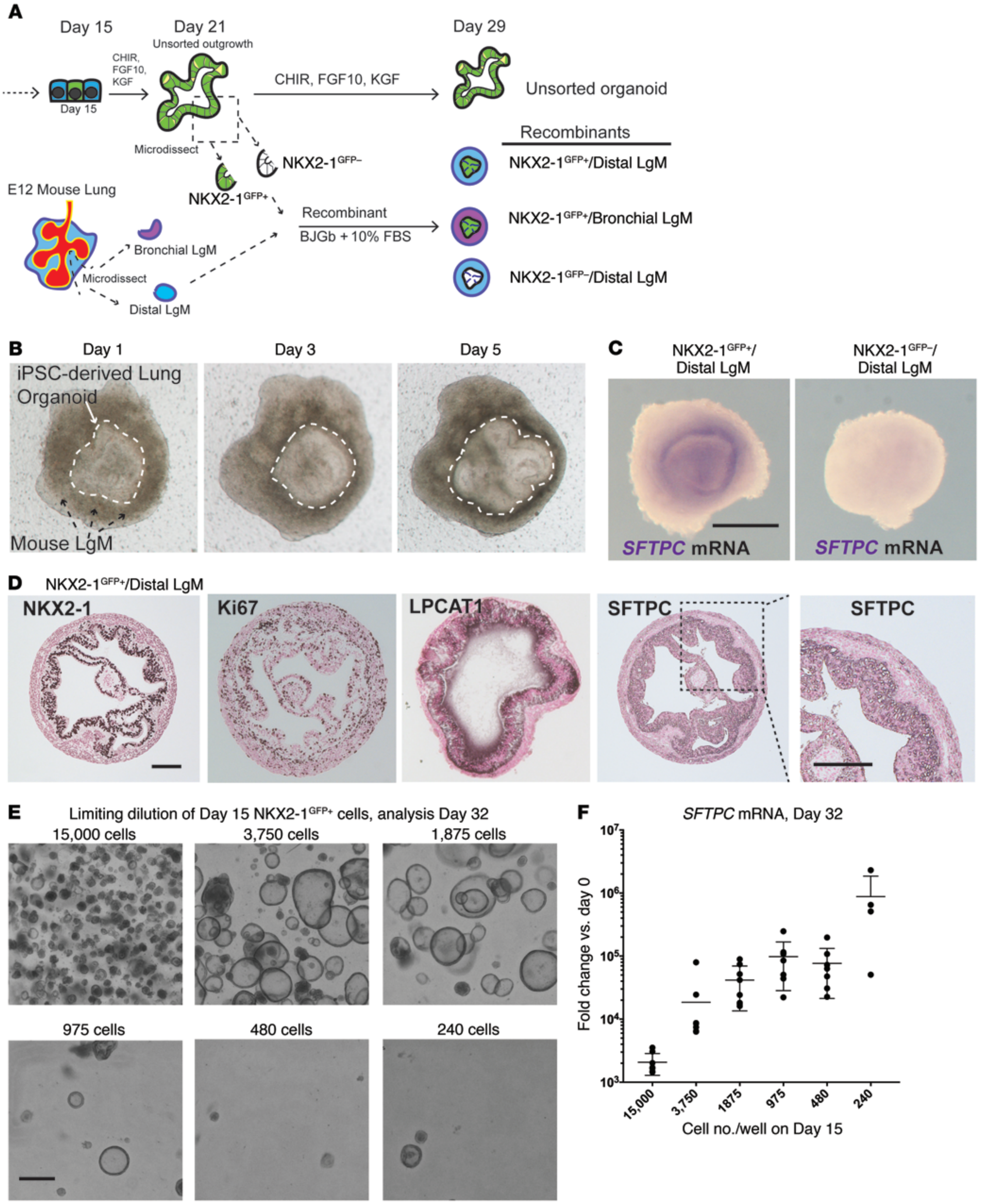

A
Day 15
Day 21
Unsorted outgrowth
Day 29
CHIR, FGF10, KGF
Day 15
CHIR, FGF10, KGF
Unsorted organoid
Microdissect
NKX2-1GFP–
NKX2-1GFP+
Recombinants
NKX2-1GFP+/Distal LgM
NKX2-1GFP+/Bronchial LgM
NKX2-1GFP–/Distal LgM
E12 Mouse Lung
Bronchial LgM
Microdissect
Distal LgM
Recombinant
BJGb + 10% FBS
B
Day 1
Day 3
Day 5
iPSC-derived Lung Organoid
Mouse LgM
C
NKX2-1GFP+/ Distal LgM
NKX2-1GFP–/ Distal LgM
SFTPC mRNA
SFTPC mRNA
D
NKX2-1GFP+/Distal LgM
NKX2-1
Ki67
LPCAT1
SFTPC
SFTPC
E
Limiting dilution of Day 15 NKX2-1GFP+ cells, analysis Day 32
15,000 cells
3,750 cells
1,875 cells
975 cells
480 cells
240 cells
F
SFTPC mRNA, Day 32
Fold change vs. day 0
10^7
10^6
10^5
10^4
10^3
15,000
3,750
1875
975
480
240
Cell no./well on Day 15

Figure 3. Mouse-human recombinant cultures demonstrate that fetal mouse lung mesenchyme augments distal lung differentiation in iPSC-derived human lung organoids. (**A**) Schematic of microdissecting and combining iPSC-derived lung organoids with E12 mouse lung mesenchyme (LgM). (**B**) Light microscopy of the same recombinant on days 1, 3, and 5 of in vitro culture. White dashed lines indicate boundaries of mouse LgM and human iPSC-derived epithelium. (**C**) *SFTPC* mRNA expression (purple) assessed by in situ hybridization using an anti–human *SFTPC* probe to stain recombinants generated with GFP$^+$ versus GFP$^-$ human iPSC–derived organoids recombined with distal mouse LgM. Scale bar: 500 μm. (**D**) Immunostaining of NKX2-1^{GFP+}/distal LgM recombinants for NKX2-1, Ki67, LPCAT1, and pro-SFTPC proteins (with zoom). Brown = immunoperoxidase product after DAB exposure. Scale bars: 200 μm. (**E**) Representative phase microscopy of day 32 organoids grown from day 15 GFP$^+$–sorted progenitors plated at limiting dilution. Shown beneath each image are the cell numbers plated per well of a 96-well plate on day 15. Scale bar: 500 μm. (**F**) Fold change of human *SFTPC* mRNA expression in day 32 organoids, generated from sorted day 15 NKX2-1^{GFP+} cells at concentrations ranging from 15,000 to 240 cells per well, compared with day 0 by RT-qPCR using the $2^{(-\Delta\Delta CT)}$ method. Lines with error bars indicate mean ± SD, $n = 8$ biological replicates except for sample 240, where only 4 samples had appreciable RNA. **B**–**D** were performed with C17, **E** and **F** with BU3.

distal or bronchial mesenchyme (Figure 3A). After 5 to 7 further days in culture as recombinants, we observed continued growth of the human GFP$^+$ aggregates in response to mouse distal LgM, continued robust expression of nuclear human NKX2-1 protein, expression of proliferation marker Ki67, and no detectable expression of thyroid markers (Figure 3, B–D and Supplemental Figure 3C). Importantly, we noted induction of cytoplasmic human pro-SFTPC and LPCAT1 protein expression in the majority of the human iPSC–derived cells by immunostaining, with validation of human *SFTPC* mRNA expression by both in situ hybridization as well as RT-qPCR (Figure 3, B–D and Supplemental Figure 3, D and E). Levels of *SFTPC* induction were higher in organoids recombined with distal LgM than in those continued solely through directed differentiation without recombination (Supplemental Figure 3D). Since E12 mouse LgM presumably lacks the signals needed for full maturation of the alveolar epithelium, which normally begins in the mouse at E18.5, human cells in these recombinants did not exhibit robust induction of transcripts associated with mature lamellar body biogenesis, such as *LAMP3* (data not shown), and did not appreciably display lamellar body–shaped inclusions by microscopy, as expected. Bronchial mesenchyme did not induce *SFTPC* expression or proximal lung epithelial markers (*SOX2*, *TP63*, or *SCGB3A1*) in NKX2-1^{GFP+} cells (Supplemental Figure 3D and data not shown). In addition, control recombinants generated using GFP$^-$ organoids were not competent at inducing either human NKX2-1 or SFTPC expression (Figure 3C and Supplemental Figure 3D). These results suggest that human iPSC–derived NKX2-1$^+$ lung progenitors respond to developing distal lung mesenchymal cues, findings in keeping with our prior observations that recombining rat distal LgM with isolated early distal primary lung epithelium induces distal alveolar marker gene expression (25).

The expression of SFTPC in response to LgM in the majority of epithelial cells analyzed raised the question of whether SFTPC$^+$ cells were being derived by the selective outgrowth of rare distal lung–competent day 15 precursors versus the possibility that distal lung–competent progenitors might be common within the NKX2-1$^+$ day 15 population. To distinguish these possibilities, we purified day 15 NKX2-1^{GFP+} cells and replated increasingly dilute numbers of cells (15,000 cells down to 240 cells per well of a 96-well plate) for further directed differentiation from day 15 to day 32 in 3D cultures without mesenchymal coculture support. This limiting-dilution assay should result in declining SFTPC competence with dilution if only rare distal progenitors are present within the day 15 NKX2-1$^+$ population (Figure 3, E and F). On day 32 we observed that lower cell numbers, plated at limiting dilution, resulted in stable to increased *SFTPC* mRNA expression, consistent with the existence of common rather than rare distal lung–competent progenitors within the NKX2-1$^+$ day 15 population and suggesting inhibition of distal differentiation in increasingly dense replating conditions in epithelial-only sphere outgrowths.

Global gene expression kinetics of early human lung development modeled by directed differentiation of human PSCs. We next sought to define the fundamental programs of early human NKX2-1$^+$ lung progenitors and their global gene expression kinetics during the course of in vitro directed differentiation. We prepared time series microarray expression profiles representing the following 5 key stages of iPSC lung-directed differentiation (Figure 4A): undifferentiated iPSCs (day 0), definitive endoderm (day 3), anterior foregut–like endoderm (day 6), sorted NKX2-1^{GFP+} and NKX2-1^{GFP-} primordial progenitors (day 15), and sorted NKX2-1^{GFP+} and NKX2-1^{GFP-} differentiated cells (day 28). For positive and negative controls we included primary distal fetal lung epithelial cells (21 weeks of human gestation) and forebrain-like iPSC-derived neural NKX2-1^{GFP+} cells (as shown in Figure 1B), respectively (Supplemental Table 1).

Principal component analysis (PCA) indicated that the global transcriptome of NKX2-1^{GFP+} lung cells was easily distinguished from NKX2-1^{GFP+} neural cells (Figure 4B). Unsupervised hierarchical clustering of all 27 samples based on the top ~1,000 transcripts differentially expressed as determined by ANOVA (1,032 genes at $P < 5 \times 10^{-13}$) revealed that day 15 and 28 cells prepared in the lung-directed differentiation protocol clustered closer to distal fetal lung epithelial controls than to endoderm or neural NKX2-1$^+$ cells (Figure 4C). The transcriptional profile of neural NKX2-1^{GFP+} cells compared with lung NKX2-1^{GFP+} cells (day 15) included 4,329 differentially expressed transcripts (FDR-adjusted $P < 0.01$) and a distinct set of transcription factors including *SIX3*, *DBX1*, *OTX1*, *OTX2*, *FOXD1*, *PAX6*, and *LHX5* (ranked by fold change [FC], filtered by FC > 5, FDR < 0.01, and gene ontology [GO] classification GO:0003700: “transcription factor activity, sequence-specific DNA binding”; Figure 4D), further emphasizing the marked differences between these early NKX2-1$^+$ forebrain and lung progenitors. Directed-differentiation protocols have previously used *FOXA2* expression as a marker to define an NKX2-1$^+$ population as endodermal and *TUJ1* to indicate neuroectodermal fate, but we found that *FOXA2* is expressed in both neuronal and lung populations (Figure 4F), and that *TUJ1* is not highly expressed in the neuronal NKX2-1$^+$ population, suggesting neither marker is useful in distinguishing NKX2-1$^+$ neural from lung lineages. In contrast, the top 10 transcription factors differentially expressed in day 15 lung NKX2-1^{GFP+} versus neuronal NKX2-1^{GFP+} cells include 6 genes expressed in the developing lung in vivo (*GRHL2*, *ELF3*, *GATA6*,

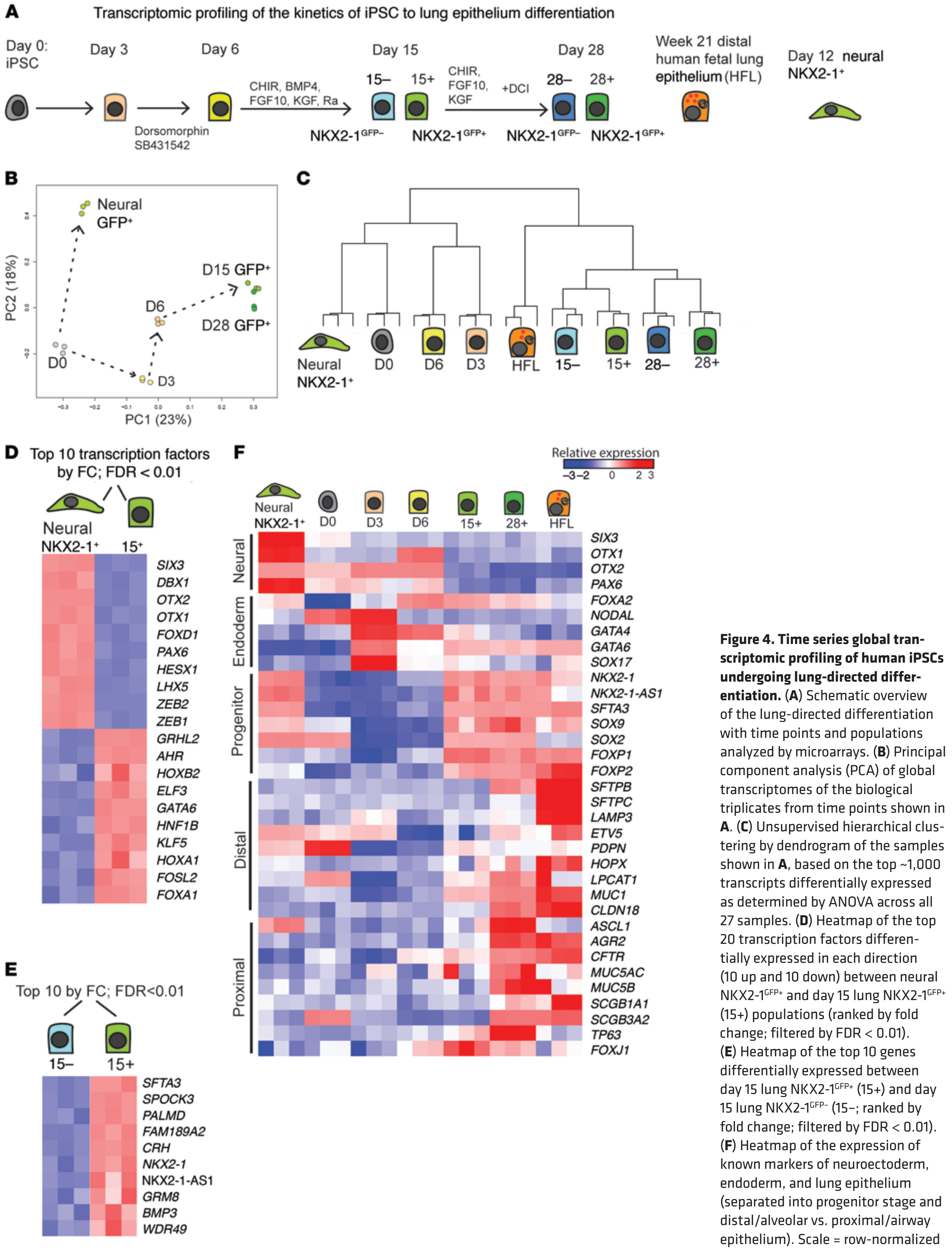
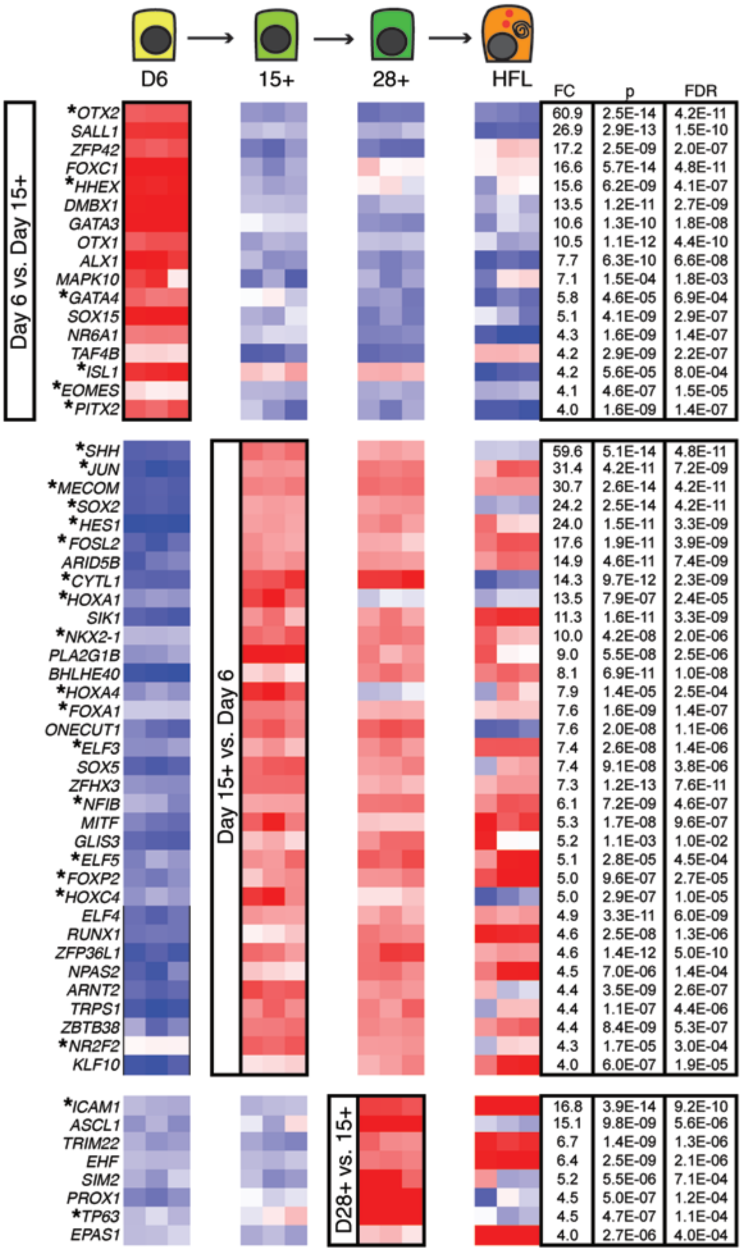


Figure 4. Time series global transcriptomic profiling of human iPSCs undergoing lung-directed differentiation. (**A**) Schematic overview of the lung-directed differentiation with time points and populations analyzed by microarrays. (**B**) Principal component analysis (PCA) of global transcriptomes of the biological triplicates from time points shown in **A**. (**C**) Unsupervised hierarchical clustering by dendrogram of the samples shown in **A**, based on the top ~1,000 transcripts differentially expressed as determined by ANOVA across all 27 samples. (**D**) Heatmap of the top 20 transcription factors differentially expressed in each direction (10 up and 10 down) between neural NKX2-1^{GFP+} and day 15 lung NKX2-1^{GFP+} (15+) populations (ranked by fold change; filtered by FDR < 0.01). (**E**) Heatmap of the top 10 genes differentially expressed between day 15 lung NKX2-1^{GFP+} (15+) and day 15 lung NKX2-1^{GFP-} (15–; ranked by fold change; filtered by FDR < 0.01). (**F**) Heatmap of the expression of known markers of neuroectoderm, endoderm, and lung epithelium (separated into progenitor stage and distal/alveolar vs. proximal/airway epithelium). Scale = row-normalized log2 expression. All differentiation samples were from C17 iPSCs.



	FC	p	FDR
*OTX2	60.9	2.5E-14	4.2E-11
SALL1	26.9	2.9E-13	1.5E-10
ZFP42	17.2	2.5E-09	2.0E-07
FOXC1	16.6	5.7E-14	4.8E-11
*HHEX	15.6	6.2E-09	4.1E-07
DMBX1	13.5	1.2E-11	2.7E-09
GATA3	10.6	1.3E-10	1.8E-08
OTX1	10.5	1.1E-12	4.4E-10
ALX1	7.7	6.3E-10	6.6E-08
MAPK10	7.1	1.5E-04	1.8E-03
*GATA4	5.8	4.6E-05	6.9E-04
SOX15	5.1	4.1E-09	2.9E-07
NR6A1	4.3	1.6E-09	1.4E-07
TAF4B	4.2	2.9E-09	2.2E-07
*ISL1	4.2	5.6E-05	8.0E-04
*EOMES	4.1	4.6E-07	1.5E-05
*PITX2	4.0	1.6E-09	1.4E-07
*SHH	59.6	5.1E-14	4.8E-11
*JUN	31.4	4.2E-11	7.2E-09
*MECOM	30.7	2.6E-14	4.2E-11
*SOX2	24.2	2.5E-14	4.2E-11
*HES1	24.0	1.5E-11	3.3E-09
*FOSL2	17.6	1.9E-11	3.9E-09
ARID5B	14.9	4.6E-11	7.4E-09
*CYTL1	14.3	9.7E-12	2.3E-09
*HOXA1	13.5	7.9E-07	2.4E-05
SIK1	11.3	1.6E-11	3.3E-09
*NKX2-1	10.0	4.2E-08	2.0E-06
PLA2G1B	9.0	5.5E-08	2.5E-06
BHLHE40	8.1	6.9E-11	1.0E-08
*HOXA4	7.9	1.4E-05	2.5E-04
*FOXA1	7.6	1.6E-09	1.4E-07
ONECUT1	7.6	2.0E-08	1.1E-06
*ELF3	7.4	2.6E-08	1.4E-06
SOX5	7.4	9.1E-08	3.8E-06
ZFHX3	7.3	1.2E-13	7.6E-11
*NFIB	6.1	7.2E-09	4.6E-07
MITF	5.3	1.7E-08	9.6E-07
GLIS3	5.2	1.1E-03	1.0E-02
*ELF5	5.1	2.8E-05	4.5E-04
*FOXP2	5.0	9.6E-07	2.7E-05
*HOXC4	5.0	2.9E-07	1.0E-05
ELF4	4.9	3.3E-11	6.0E-09
RUNX1	4.6	2.5E-08	1.3E-06
ZFP36L1	4.6	1.4E-12	5.0E-10
NPAS2	4.5	7.0E-06	1.4E-04
ARNT2	4.4	3.5E-09	2.6E-07
TRPS1	4.4	1.1E-07	4.4E-06
ZBTB38	4.4	8.4E-09	5.3E-07
*NR2F2	4.3	1.7E-05	3.0E-04
KLF10	4.0	6.0E-07	1.9E-05
*ICAM1	16.8	3.9E-14	9.2E-10
ASCL1	15.1	9.8E-09	5.6E-06
TRIM22	6.7	1.4E-09	1.3E-06
EHF	6.4	2.5E-09	2.1E-06
SIM2	5.2	5.5E-06	7.1E-04
PROX1	4.5	5.0E-07	1.2E-04
*TP63	4.5	4.7E-07	1.1E-04
EPAS1	4.0	2.7E-06	4.0E-04

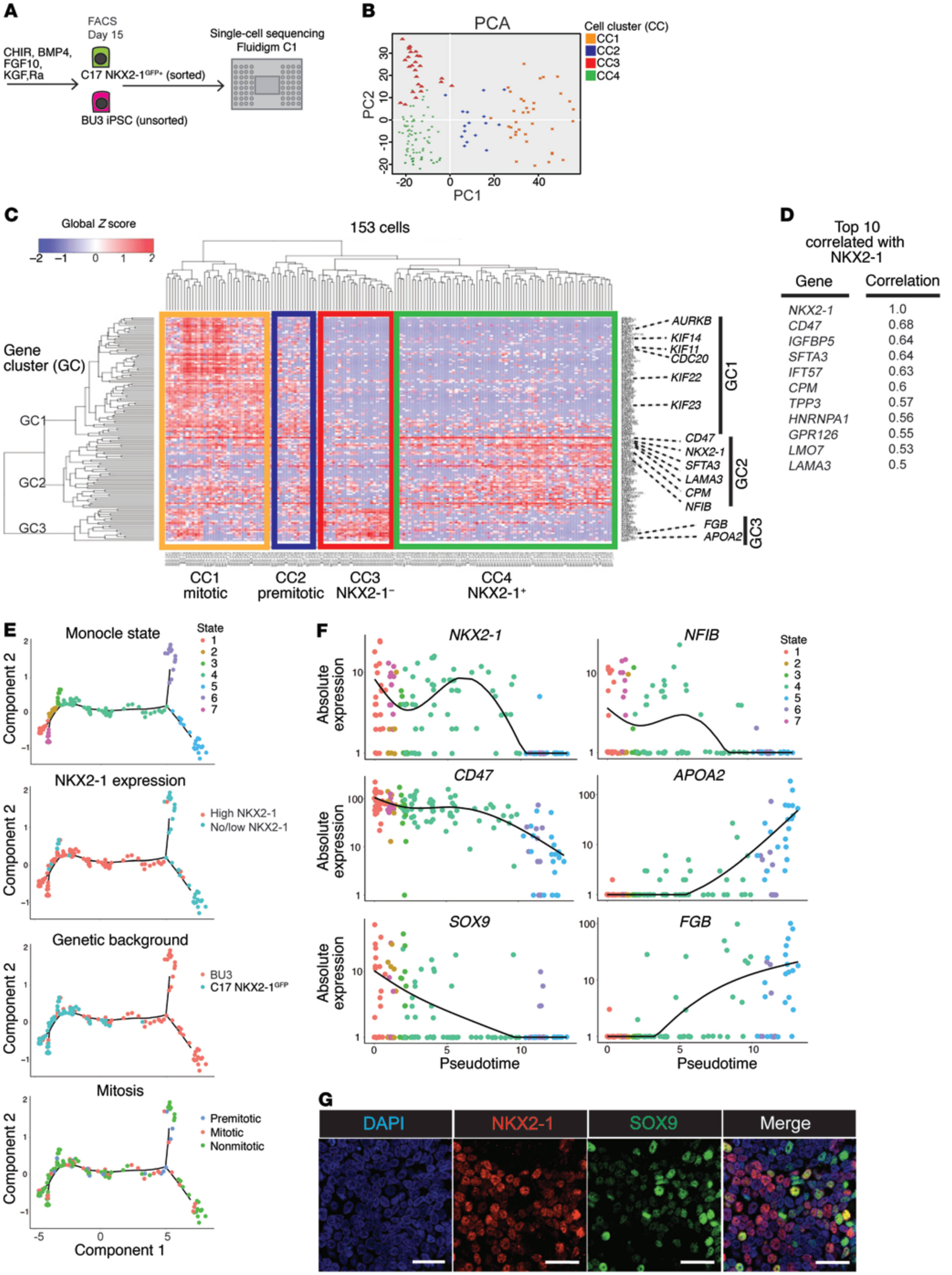
Figure 5. Transcriptomic signatures of iPSC-derived anterior foregut endoderm and lung progenitors, focused on genes associated with transcription factor activity. Heatmap of the top transcription factors or genes with transcription factor activity GO terms that are differentially expressed (filtered by fold change > 4) between successive stages of lung differentiation starting from day 6 and compared with human fetal lung. Vertical text box and heatmap with black outline identify the signature of each stage being analyzed. *Indicates the gene was previously described in foregut and/or lung development (see Supplemental Table 3 for further publication details). See also Supplemental Figure 4. Day 6, day 15, and day 28 time points were derived from C17 iPSCs.

HNF1B, *HOXA1*, and *FOXA1* (Figure 4D) (26–28), suggesting a constellation of transcripts better able to distinguish these 2 iPSC-derived populations.

To interrogate the differences between day 15 NKX2-1^{GFP+} and NKX2-1$^{GFP–}$ cells at the primordial progenitor stage, we ranked the top 10 differentially expressed genes by FC (Figure 4E). We found that *NKX2-1* as well as neighboring long noncoding RNAs (lncRNAs) *SFTA3* and *NKX2-1AS* were highly upregulated in the GFP^{+} population. Also in this top 10 list were genes (*BMP3*, *CRH*, and *SPOCK3*) previously described in lung development (Figure 4E, with validation by RT-qPCR in S4B) (28–34). The finding that *SFTA3* (aka NKX2-1–associated noncoding intergenic RNA [*NANCI*]) is the top differentially expressed transcript in the genome distinguishing day 15 NKX2-1^{GFP+} cells is in keeping with recent reports that in developing mouse lungs this transcript is coexpressed with *Nkx2-1* and shares the same regional and temporal expression pattern (35). To the best of our knowledge, 4 genes in this list (*PALMD*, *FAM189A2*, *GRM8*, and *WDR49*) have not been previously identified in the lung epithelium.

From our microarray datasets we selected known markers of definitive endoderm, forebrain, and 24 genes of known importance in the developing lung epithelium and profiled their expression patterns over the course of directed differentiation in comparison with human primary fetal epithelial lung control cells (HFL) (Figure 4F). Endodermal markers, such as *GATA4*, *GATA6*, *SOX17*, *NODAL*, and *FOXA2* were upregulated early during endodermal differentiation, with retained expression of *GATA6* and *FOXA2* in day 15 and day 28 NKX2-1^{GFP+} cells. In contrast, the transcripts *NKX2-1*, *SFTA3*, *SOX9*, and *FOXP* family members were low or absent prior to day 6, and their clear emergence in the day 15 GFP^{+} population is consistent with their published expression during the early lung progenitor period in mouse lung development (35, 36). These findings, together with the lack of mature lung marker gene expression in day 15 GFP^{+} cells (low *SCGB1A1*, *SCGB3A2*, *TP63*, *SFTPB*, and *SFTPC*), further suggest the day 15 GFP^{+} population represents a relatively undifferentiated or primordial lung progenitor population, as has been observed in early NKX2-1^{+} progenitors in developing mouse cells in vivo (28). In contrast to day 15, by day 28 the GFP^{+} population had begun to express markers known to be enriched in maturing alveolar epithelial cells (*ETV5*, *CLDN18*, *LPCAT1*, *MUC1*, *SFTPB*, and low *SFTPC*) or in airway epithelia, such as basal (*TP63*), secretory (*SCGB3A2*, *MUC5B*, *MUC5AC*, and *AGR2*), and neuroendocrine (*ASCL1*) cells. PDPN, which has occasionally been referred to as a PSC-derived type 1 pneumocyte marker in prior publications (8, 9), was actually expressed in day 0 as well as day 15 GFP^{+} cells, consistent with its expression patterns in developing mice where it is robustly expressed in both the foregut endoderm and the developing pseudoglandular lung epithelium prior to the emergence of type 1 cells (37).

Next we sought to identify unbiased gene signatures of primordial (day 15) and maturing (day 28) NKX2-1^{GFP+} cells. We generated lists of the top 100 differentially expressed genes (ranked by FC, filtered by FDR < 0.01) of each sample across multiple comparisons and identified a common gene set for each sample (Supplemental Figure 4 and Supplemental Table 2). The day 28 GFP$^{–}$ population was enriched for diverse but predominantly liver (*APOA2*, *FGB*, *AFP*, *CDH17*, and *TF*) and intestinal (*CDX2*, *CDH17*, and *GIF*) markers (Supplemental Table 2). In addition to



Gene	Correlation
NKX2-1	1.0
CD47	0.68
IGFBP5	0.64
SFTA3	0.64
IFT57	0.63
CPM	0.6
TPP3	0.57
HNRNPA1	0.56
GPR126	0.55
LMO7	0.53
LAMA3	0.5

Figure 6. Single-cell RNA sequencing of sorted and unsorted iPSC-derived cells reveals NKX2-1⁺ lung and NKX2-1⁻ non-lung lineages and suggests markers for their identification. (**A**) Schematic of the single-cell capture and global RNA sequencing of sorted C17 NKX2-1^{GFP+} and unsorted BU3 iPSCs on day 15 of lung-directed differentiation. (**B**) PCA of the top 150 most variant genes of all sequenced cells reveals 4 cell clusters, color coded to match the clusters shown in panel **C**. (**C**) Heatmap of gene expression (global *Z* score) with unsupervised hierarchical clustering of all 153 cells (*x* axis) and the top 150 most variant genes (*y* axis). Dendrograms as well as colored boxes indicate 4 cell clusters (CC1–4; matching colors shown in panel **B**). *y*-axis dendrograms and thick black lines indicate 3 gene clusters (GC1–3). Key genes indicated on right. CC1 = orange, CC2 = blue, CC3 = red, CC4 = green. (**D**) Top 10 genes correlated with NKX2-1 expression. (**E**) Unsupervised cell clustering using Monocle's pseudotime spanning tree analysis reveals 7 cell states. Individual cells are labeled in subsequent panels by NKX2-1 level, genetic background (iPSC clone), or cell cycle (mitosis), respectively. (**F**) Pseudotime plots of expression levels of *NKX2-1*, *CD47*, *SOX9*, *NFIB*, *FGB*, and *APOA2* with cells colored based on the 7 states determined in **E**. (**G**) Immunostaining of RUES2-derived day 15 cells for NKX2-1 (red) and SOX9 (green) nuclear proteins. Nuclei counterstained with DAPI. Scale bars: 25 μm.

NKX2-1, *SFTA3*, *CPM*, *NFIB*, and *CRH*, which are all expressed in primordial lung progenitors, the maturing lung cells (day 28 GFP⁺) expressed higher levels of *SCGB3A2*, *SFTPB*, *TP63*, *ICAM1*, *IL8*, and *ITGB6* (Supplemental Figure 4 and Supplemental Table 2). *SCGB3A2* was the most differentially expressed transcript of 23,786 probesets ranked by FC (day 28 GFP⁺ vs. GFP⁻ groups; FC = 76.6; FDR-adjusted $P = 1.6 \times 10^{-9}$). *SFTPC* was upregulated by day 28 (GFP⁺), but not yet at levels equivalent to HFL, findings in keeping with our recombinant experiments, which suggested that current differentiation protocols without the use of primary mesenchyme have not yet been optimized for efficient and full distal alveolar maturation (Figure 3, C and D, Supplemental Figure 3D, Figure 4F, and Supplemental Table 2).

Transcription factors play critical roles in organogenesis including in lung development (27). To identify candidate genes that control human lung specification and development, we screened for enrichment of transcription factors or regulators of transcription prior to lung specification (day 6) and at different stages of lung maturity (day 15 NKX2-1^{GFP+}, day 28 NKX2-1^{GFP+}, and HFL; Figure 5). To identify genes of interest, we ranked the significantly differentially expressed genes by FC (FC > 4; FDR < 0.01) and filtered genes based on GO classification for transcription factor activity (GO:0003700, "transcription factor activity"). The majority of the most highly differentially expressed genes in day 6 anterior foregut–like endoderm were known transcription factors of the foregut endoderm previously described in *Xenopus* and mouse model systems: *HHEX*, *GATA3*, *GATA4*, *FOXC1*, *EOMES*, *OTX1*, *OTX2*, *ISL1*, and *PITX2* (28, 38, 39) (Figure 5, Supplemental Figure 4C, and Supplemental Table 3). In comparison with day 6, the day 15 NKX2-1^{GFP+} population was enriched for many transcription factors known to be present in the developing mouse lung (*JUN*, *MECOM*, *SOX2*, *HES1*, *HOXA1*, *NKX2-1*, *FOXA1*, *ELF3*, *ELF5*, *NFIB*, and *FOXP2*) (26–28, 38–44) (Figure 5, Supplemental Figure 4D, and Supplemental Table 3). In addition to *HOXA1*, a number of other HOX genes were upregulated in day 15 NKX2-1^{GFP+} samples (*HOXA4* and *HOXC4*). *SHH*, essential for normal lung development in mice (27), was the most highly expressed gene in the day 15 samples in our analysis (FC = 60, FDR = 4.8×10^{-11}). Day 28 NKX2-1^{GFP+} cells expressed higher levels of transcription factors associated with basal cells (*TP63*) and neuroendocrine cells (*ASCL1*). Taken together, these time series data provide unbiased stage-dependent signatures of the putative transcriptomic programs of human lung progenitors and their differentiated progeny as they emerge during developmental directed differentiation. Moreover, these signatures reveal that many evolutionarily conserved transcription factors, previously observed in developing *Xenopus* and mouse lung endoderm in vivo, also are differentially expressed in the iPSC human lung development model system.

Single-cell RNA sequencing and surface marker profiling of day 15 iPSC-derived lung progenitors. While transcriptomic profiles of purified groups of cells allows a deep understanding of the genetic program of the NKX2-1⁺ progenitor population, it does not allow interrogation of the heterogeneity of these programs at the individual cell level. Hence, we next sought to profile the transcriptomic programs of individual iPSC-derived cells at the day 15 stage of lung differentiation employing the C17 NKX2-1^{GFP} targeted line as well as the untargeted iPSC line, BU3 (19). We performed RNA sequencing (RNA-Seq) of 84 BU3 iPSC-derived cells without any cell sorting and 69 C17 iPSC-derived cells, sorted based on NKX2-1^{GFP+} expression (Figure 6A).

PCA (Figure 6B) as well as unsupervised hierarchical clustering analysis (Figure 6C) both suggested that 4 broad clusters of cells (hereafter CC1-4) were present on day 15, and these cell clusters could be distinguished based on 3 broad gene clusters (GC1-3). Notably, 3 of the 4 clusters (CC1, CC2, and CC4) expressed high levels of *NKX2-1* (Figure 6C), whereas cluster CC3 exhibited an absence of transcripts encoded by the *NKX2-1* locus or its neighboring locus, *SFTA3*. *NKX2-1*–expressing clusters were most robustly distinguished by whether they exhibited mitotic (CC1 and CC2) or noncycling (CC4) gene signatures. For example, CC1 and CC2 were highly enriched for the expression of genes associated with mitosis or cytokinesis (e.g., *KIF11*, *KIF14*, *KIF22*, *KIF23*, *CDC20*, and *AURKB*). Thus, CC1 was labeled mitotic and CC2 was labeled premitotic based on slightly lower expression levels of these markers in the latter cluster, whereas CC3 and CC4 did not appear to be in active cycle at the time of cell capture for analysis. Importantly, NKX2-1⁺ cells clustered together (CC4) regardless of whether they were sorted GFP⁺ C17 iPSCs or unsorted BU3 iPSCs. Furthermore, only 1 GFP⁺ sorted cell could be found misclustering among the 26 cells that comprised the NKX2-1–negative cluster (CC3), and as expected 25 out of 26 cells found in this NKX2-1–negative cluster derived from the unsorted BU3 iPSCs (Figure 6C).

We undertook 3 approaches to interrogate the gene expression differences that distinguished each cell cluster. First, we used unsupervised hierarchical clustering of the top 150 most variant genes (Figure 6C; *y*-axis dendrograms; hereafter gene clusters GC1–3). We found that GC1 was highly enriched for cell-cycle regulation genes (including *AURKB*, *BIRC5*, *BUB1*, *CCNB1*, *CCNB2*, *CENPE*, *CENPF*, *KIF11*, *KIF14*, *KIF22*, *KIF23*, *KIF22c*, *MELK*, and *TOP2A*; Figure 6C), further supporting the interpretation that changes in genes of cytokinesis and cell cycle dominate the first level of clustering of day 15 cells. However, 2 additional distinct gene clusters were also apparent, most notably GC2, including *NKX2-1*, *SFTA3*, *NFIB*,

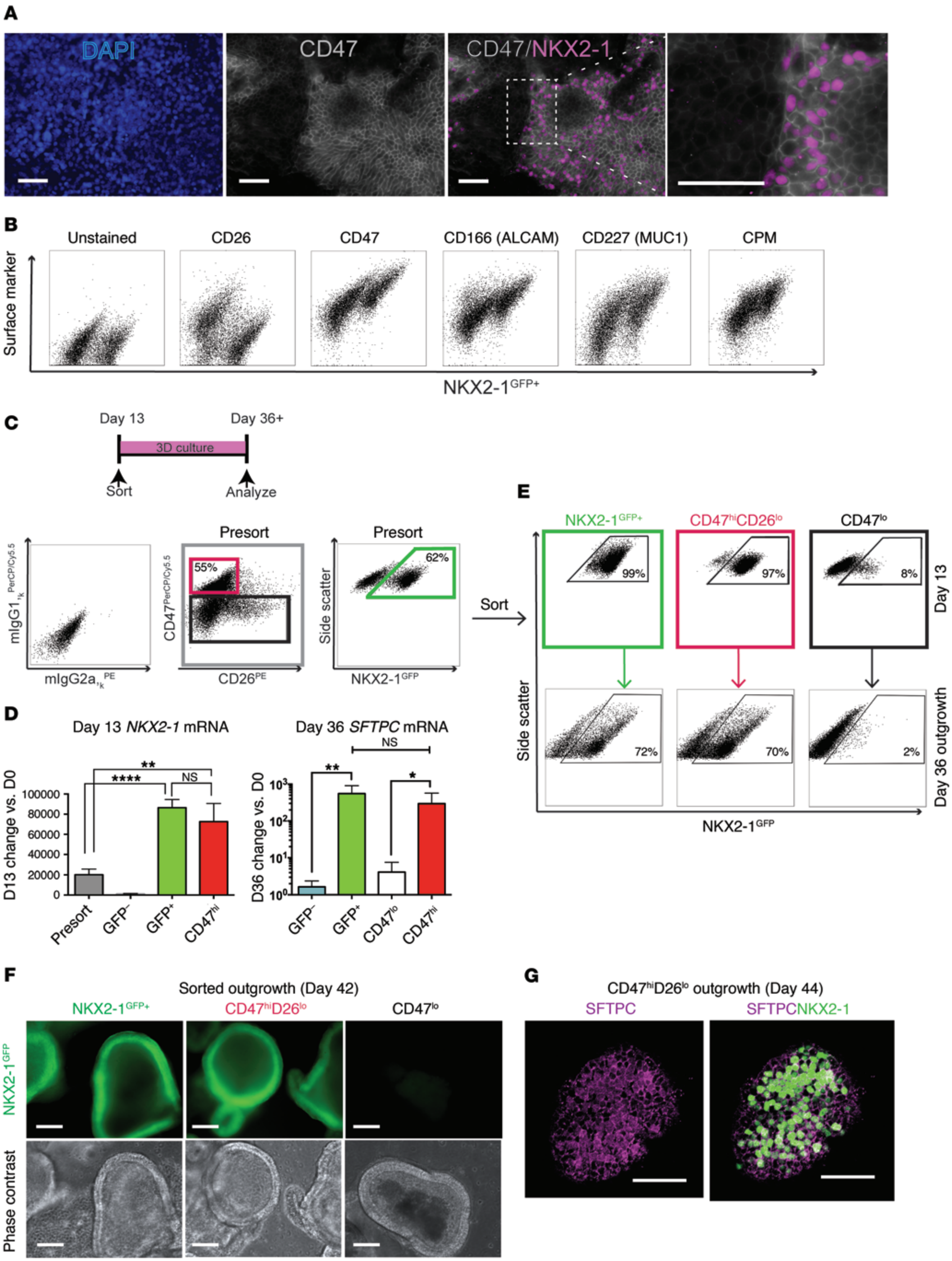

A
DAPI
CD47
CD47/NKX2-1
B
Unstained
CD26
CD47
CD166 (ALCAM)
CD227 (MUC1)
CPM
Surface marker
NKX2-1GFP+
C
Day 13
Day 36+
3D culture
Sort
Analyze
mIgG1,k PerCP/Cy5.5
mIgG2a,k PE
Presort
55%
CD47PerCP/Cy5.5
CD26PE
Presort
62%
Side scatter
NKX2-1GFP
Sort
E
NKX2-1GFP+
CD47hiCD26lo
CD47lo
99%
97%
8%
Day 13
72%
70%
2%
Day 36 outgrowth
Side scatter
NKX2-1GFP
D
Day 13 NKX2-1 mRNA
D13 change vs. D0
100000
80000
60000
40000
20000
0
****
**
NS
Presort
GFP–
GFP+
CD47hi
Day 36 SFTPC mRNA
D36 change vs. D0
10³
10²
10¹
10⁰
**
NS
*
GFP–
GFP+
CD47lo
CD47hi
F
Sorted outgrowth (Day 42)
NKX2-1GFP+
CD47hiD26lo
CD47lo
NKX2-1GFP
Phase contrast
G
CD47hiD26lo outgrowth (Day 44)
SFTPC
SFTPCNKX2-1

Figure 7. Cell surface profiling and prospective isolation of iPSC-derived NKX2-1$^+$ primordial lung progenitors by CD47^{hi}C26^{lo} cell sorting. (**A**) Day 15 iPSCs after lung-directed differentiation, immunostained for CD47 and NKX2-1 proteins (C17). Nuclei are counterstained with DAPI. Scale bars: 50 μm. (**B**) Flow cytometry dot plots of 4 cell surface markers identified in a screen of 243 surface markers on day 15 of lung-directed differentiation (C17); CD26 is depleted while CD47, ALCAM, MUC1, and CPM exhibit higher expression in the GFP$^+$ population. (**C**) Schematic of experimental data for **C**, **D**, and **E**. Flow cytometry dot plots of live day 13 cells (BU3) indicates staining with isotype control antibodies (left panel) or antibodies against CD47 and CD26. Sort gates identify presorted cells (gray box) versus a CD47^{hi}CD26^{lo} population (red box) or a CD47^{lo} population (black box) profiled in **D** and **E**. (**D**) Fold change of *NKX2-1* mRNA (left graph, *n* = 4) in each indicated day 13 population, and *SFTPC* mRNA (right graph, *n* = 3) expression in the outgrowth of each indicated population on day 36 compared with day 0 iPSCs by RT-qPCR; $2^{(-\Delta\Delta CT)}$. Data indicate the mean ± SD. $^*P \leq 0.05$, $^{**}P \leq 0.01$, $^{****}P \leq 0.0001$ by Student's *t* test. (**E**) GFP expression quantified in each of the gates shown in **C**: day 13 presort = 62% NKX2-1^{GFP+}; CD47^{hi}CD26^{lo} = 97% GFP$^+$; CD47^{lo} = 8% GFP$^+$. Lower panel is FACS of day 36 outgrowth of each indicated day 13 sorted population: GFP$^+$, CD47^{hi}CD26^{lo} versus CD47^{lo}. (**F**) Phase contrast and fluorescence microscopy (GFP) of day 42 organoids derived from day 13 sorted GFP$^+$, CD47^{hi}CD26^{lo}, and CD47^{lo} populations from **E**. Scale bars: 100 μm. (**G**) Confocal microscopy of outgrowth organoids (C17), sorted on day 13 based on CD47^{hi}CD26^{lo} and analyzed on day 44 by coimmunostaining for NKX2-1 (green) and pro-SFTPC (purple). The individual panels for this merge image are contained in Supplemental Figure 7E. Scale bars: 25 μm.

CD47, *WNT5A*, *CPM*, and *LAMA3*; and GC3 which was associated with the *NKX2-1*–negative/*SFTA3*-negative cells of CC3 (Figure 6C). This analysis suggested GC2 genes as potential markers associated with *NKX2-1*$^+$ lung cells, and GC3 genes as potential markers associated with non-lung (*NKX2-1*–negative) cells. Consistent with this observation, the top 10 genes most highly correlated with *NKX2-1* expression across individual cells were GC2 genes, including *CD47*, *SFTA3*, *CPM*, and *LAMA3* (Figure 6D). In contrast, GC3 was enriched in liver-lineage genes (*APOA2* and *FGB*) as well as nonspecific mesenchymal genes (*COL19A1* and *S100A10*). We have previously reported that in iPSC-derived hepatic cells, *FGB* represents the most upregulated transcript in the genome during hepatic-directed differentiation (45, 46). Furthermore, in postnatal human tissues both *APOA2* and *FGB* are transcripts specifically enriched in liver cells (gtexportal.org). Significantly, all *NKX2-1*–negative cells of putative hepatic lineage (20 out of 20 *APOA2*$^+$ cells; Figure 6C) were solely composed of unsorted BU3 iPSCs, indicating that sorting on the NKX2-1^{GFP+} marker successfully depleted any contaminating hepatic cells in this protocol. The majority of cells (23 of 37) in the mitotic and premitotic groups clustered with NKX2-1$^+$CD47$^+$ cells when hierarchical clustering was run again after cell cycle genes were removed (Supplemental Figure 5A).

Second, we applied the Monocle computational algorithm (47) to our single-cell dataset in an effort to identify, in an unbiased manner and irrespective of cell cycle, cell subtypes, or intermediate states. Because a particular challenge in single-cell RNA-Seq experiments is the high cell-to-cell variation observed in most genes, including key developmental regulators, during differentiation (48–51), Monocle was developed to improve the resolution of individual transcriptomes and allow the ordering of cells by potential progress through a biological process without relying on known lineage markers (47). Hence, we used Monocle to order all day 15 cells in pseudotime: an abstract, semiquantitative measure of progress through a biological process (Supplemental Figure 5B). In this analysis, pseudotime represents a computational, high-dimensional ordering of the transcriptional spectrum of differentiating cells, accounting for the likelihood that day 15 cultures contain diverse cell types at various stages of differentiation. In light of the dominant effect of proliferation on the first-level analysis (Figure 6C), we excluded cell-cycle genes from the Monocle analysis. Unsupervised cell clustering revealed 7 Monocle states (labeled states 1–7; Figure 6E and Supplemental Figure 5B), composed of 24, 8, 10, 57, 21, 13, and 12 cells, respectively. Next, we labeled cells in each state based on expression levels of *NKX2-1* (high vs. no/low expression; Figure 6E) and determined that cells in states 5 and 6 clustered separately because they expressed no or low levels of NKX2-1, whereas cells in states 1, 2, 4, and 7 were almost entirely *NKX2-1* high. In contrast, most states were neither defined by cell cycle effects (mitotic state) nor genetic background (cell origin) of each iPSC line, with the exception of states 5 and 6, which were entirely composed of unsorted BU3 cells, as expected (Figure 6E). Consistent with our clustered heatmap results (Figure 6C), we found that states 1, 2, 4, and 7 cells expressed higher levels of *NKX2-1*, *CD47*, and *SFTA3*, consistent with a lung signature, whereas state 5 cells were enriched for *APOA2* and *FGB*, consistent with a fetal liver signature (Figure 6, Supplemental Figure 5D, and data not shown). The lineage identity of state 6 cells, enriched for transcription factors including MSX1, EGLN3, and OTX2, was uncertain. The presence of discrete NKX2-1$^+$ states suggested some degree of either temporal or lineage heterogeneity within the overall NKX2-1 population. For example, *SOX9* expression varied across cells in *CD47*$^+$ *NKX2-1*$^+$ states. State 1 and 2 cells were significantly enriched for SOX9 expression, whereas state 4 cells were not. State 3 cells were highly enriched for SOX9 but expressed lower levels of NKX2-1. Using Monocle to examine genes that follow similar expression trends when cells are ordered in pseudotime, we identified genes with increasing expression towards a lung phenotype (including *SOX9*, *NKX2-1*, *CD47*, *NFIB*, *LAMA3*, and *SFTA3*) and conversely genes with increasing expression towards a liver phenotype (*DLK1*, *AFP*, *MSX1*, *FTL1*, *FN1*, *FGB*, and *APOA2*) (Figure 6F and Supplemental Figure 5, C and D). The expression of SOX9 at variable but easily detected levels in the majority of NKX2-1$^+$ putative lung cells was confirmed at the protein level by immunostaining (Figure 6G), and only a minor subset of NKX2-1$^+$ cells expressed high levels of the proximal airway patterning marker SOX2 without SOX9 (either by immunostaining or by supervised hierarchical clustering of single-cell transcriptomes; Supplemental Figure 6A). This predominance of the distal progenitor marker, SOX9, in day 15 NKX2-1$^+$ cells is consistent with the efficient distal alveolar differentiation competence of the NKX2-1$^+$ progenitor population observed in our recombinant cultures in Figure 3. Importantly, mature distal or proximal markers (e.g., SFTPC and SCGB1A1) were not detected in any cell on day 15 (Supplemental Figure 6A).

Finally, we performed repeat ANOVA with hierarchical cell and gene clustering of the 97 cells that were not in active mitosis (focusing solely on CC3 and CC4; Figure 6C and Supplemental Figure 6B). This analysis identified 4 cell subgroups (SG1–4;

Supplemental Figure 6B). The largest subgroup of cells, SG2, expressed key genes of the early developing lung (*SFTA3*, *NFIB*, and *WNT5A*) and, in keeping with our population-based transcriptomic profiles, they lacked detectable expression of markers of lung maturation (*SFTPC*, *SFTPB*, *SCGB3A2*, *ASCL1*, *FOXJ1*, or *SCGB1A1*) (Supplemental Figure 6B). In the 3 remaining minor subgroups (SG1, 3, and 4; comprising predominantly NKX2-1–negative BU3 cells), we found differentially expressed genes suggestive of non-lung endoderm or undetermined identity. Gene sets most significantly correlated with SG1 and SG3 were consistent with hepatic lineages (Supplemental Figure 6B). Significance testing of SG2 versus SG3 demonstrated that *CD47* was the most highly differentially expressed gene in SG2 (ranked by either P value or correlation with *NKX2-1* expression; Supplemental Figure 6C), followed by *IGFBP5*, *SFTA3*, *EIF1AY*, *LAMA3*, *CPM*, *SOX9*, and *LMO7* (Supplemental Figure 6C). *LMO7* is a known target of FGF10 that is upregulated in early developing mouse lung epithelium (52), and *NKX2-1*, *SFTA3*, *CPM*, and *SOX9* are all known to be enriched in developing mouse and human lung epithelia.

To determine whether key markers identified in our single-cell RNA-Seq or microarray analyses are expressed in the developing human lung epithelium in vivo, we analyzed available microarray data of human fetal lungs ranging from 53 to 154 days of gestation (53). Consistent with our PSC in vitro model system, we observed increasing in vivo expression with time of known lung differentiation markers (*SFTPC*, *SFTPB*, and *LAMP3*), absence at any time point of non-lung markers (*APOA2* and *CDX2*), and early, unchanging expression of *NKX2-1*, *CD47*, *NFIB*, *HOXA3*, and *JUN* (Supplemental Figure 6D). In addition, we confirmed early developmental CD47 protein expression in NKX2-1^{+} epithelial cells in vivo by immunostaining week 10 human fetal lung (Supplemental Figure 6E).

Taken together, our results provided an improved understanding of the heterogeneity of iPSC-derived cells emerging with lung-directed differentiation, supported the utility of NKX2-1^{GFP+} sorting to deplete non-lung endodermal lineages that contribute to this heterogeneity, and suggested transcripts associated with NKX2-1^{+} cells in this in vitro model system.

Prospective isolation of iPSC-derived NKX2-1^{+} primordial lung progenitors by CD47^{hi} cell sorting. Because our single-cell RNA-Seq profiles revealed that *CD47* was the transcript in the genome most highly correlated with *NKX2-1* (Figure 6D), we sought to determine whether NKX2-1^{+} primordial progenitor cells might be prospectively isolated based on cell surface protein expression of CD47 without the need for a GFP knockin reporter. Using both immunofluorescence microscopy as well as FACS of day 13–15 unsorted PSCs (C17, BU3, and RUES2 lines), we observed that the brightest CD47^{+} cells selectively coexpressed NKX2-1 nuclear protein as well as the NKX2-1^{GFP} reporter (Figure 7, A and B and Supplemental Figure 7).

In independent experiments we screened day 15 iPSC-derived NKX2-1^{GFP+} progenitors by FACS using a panel of 243 antibodies, validating that CD47 was in the top 4 cell surface markers most associated with GFP^{+} expression (Figure 7B). Notably, this screen also confirmed that: (a) NKX2-1^{GFP+} cells are EPCAM positive (Supplemental Figure 7A); (b) CPM, recently published as a marker of iPSC-derived NKX2-1^{+} cells (16) and highly associated with NKX2-1 in our single-cell RNA-Seq (Figure 6D), indeed costains most NKX2-1^{GFP+} cells (Figure 7B), although it is also associated with NKX2-1^{-} hepatic cells (54) that emerge at low levels in this protocol (Figure 6F and Supplemental Figures 5B and 6D); and (c) ALCAM (CD166) and MUC1 (CD227) are 2 additional candidate markers that selectively identify NKX2-1^{GFP+} cells at this stage (Figure 7B). Importantly, our antibody screen also suggested CD26 as a negative-selection marker since the brightest CD26^{+} cells were lower in expression of the NKX2-1^{GFP} reporter (Figure 7B). When sorting day 13–15 cells based solely on CD47^{hi}CD26^{lo} gating, we found significant enrichment for NKX2-1^{+} cells: 89% ± 4.1% of cells (mean ± SD; $n = 11$ runs) expressed NKX2-1 nuclear protein as well as the NKX2-1^{GFP} reporter compared with CD47^{lo} cells, which were depleted of NKX2-1 expression (Figure 7, C–F). These findings were validated for both human ESCs (RUES2) as well as multiple iPSC lines (C17, BU3, 100-3, RC202, and RC204; refs. 18, 19, 55 and data not shown), despite varying efficiencies of NKX2-1^{+} induction in any given differentiation run (Supplemental Figure 7, C and D). For example, regardless of whether a line differentiated to lung with low or high efficiency (e.g., 13% vs. 56% NKX2-1^{+} in 3 separate runs for C17; or 50% for RUES2), in each case CD47^{hi}CD26^{lo} gating provided significant enrichment in NKX2-1^{+} cells (7-fold vs. 2-fold enrichment for C17 and 2-fold for RUES2), resulting in populations approximately 90% pure for NKX2-1 expression in each run. Furthermore, human ESCs or human iPSCs sorted solely on CD47^{hi}CD26^{lo} gating produced predominantly NKX2-1^{+} spheroids in 3D culture that expressed lung differentiation markers including pro-SFTPC protein and *SFTPC* mRNA at levels similar to those of sorted NKX2-1^{GFP+} cells (Figure 7, D and G).

Discussion

Our results indicate that iPSC-derived lung epithelial cells originate from identifiable NKX2-1^{+} progenitors. Through the use of an NKX2-1–targeted GFP reporter these progenitors can be sorted and then further differentiated without mesenchymal coculture support in 3D Matrigel culture. Importantly, human NKX2-1^{+} progenitors derived with our methods undergo efficient distal SFTPC^{+} differentiation and proliferation after recombinant culture with primary distal embryonic mouse LgM. We have labeled these NKX2-1^{+} cells, which emerge between days 8 and 15 of iPSC differentiation, primordial progenitors because they express a transcriptome that includes the earliest transcripts known to emerge during the endodermal and primary lung bud stages of mammalian development (*NKX2-1*, *SFTA3*, *SOX9*, and *SOX2*), but they are otherwise lacking in transcripts associated with differentiated/maturing lung epithelia, most of which emerge during the later pseudoglandular stage of lung development.

Our findings support a paradigm in which the human lung epithelium derives directly from NKX2-1^{+} endodermal progenitors rather than from alternate cells, because sorting human NKX2-1^{+} cells at the primordial stage highly enriches for cells competent at further differentiating into lung epithelia, while depleting this population significantly depletes cells competent at forming lung. Given the difficulty in accessing and tracking live human fetal cells in vivo during the earliest stages of lung development (approximately 1 month of human gestation), our in vitro model enables the purification, tracking, and visualization of cells undergoing the earliest moments of human lung cell fate decisions, a time period in human lung development that remains elusive to scientific study.

We provide evidence that the genetic control of early human lung development is similar to that in mouse. Indeed, our finding that human iPSC-derived lung epithelial progenitors respond to inductive differentiation cues provided by developing mouse LgM suggests that an evolutionarily conserved biology is common to early mouse and human lung development. Furthermore, our stage-dependent transcription factor signatures for developing lung, thyroid, and forebrain revealed by our in vitro iPSC model provides important validation in a human system of many of the gene ensembles previously identified in mice. For example, we found that many genes and transcription factors of murine lung development are expressed in iPSC-derived human lung progenitors, including *NKX2-1*, *SHH*, *FOXA2*, *FOXA1*, *GATA6*, *SOX2*, *SOX9*, *GRHL2*, *IRX1*, *IRX2*, *NFIA*, *NFIB*, *FOXP2*, *HNFIB*, *ELF3*, and *ELF5*. Our developmental stage–dependent signatures also suggest novel genes requiring further study, and the ability to employ the human iPSC model system should now provide a tractable developmental human system to examine the roles of these and other genes in human lung lineage specification. Partnered future work focused on this early period of lung lineage specification in vivo in mice is likely to provide a further understanding of the phenotype and biology of primordial lung progenitors across mammalian species, and should enable isogenic head-to-head comparisons of iPSC-derived progenitors with their in vivo counterparts. It is important to point out that the differentiation kinetics of our human iPSC in vitro model is generally faster than that observed in vivo in developing human endoderm and lungs. Thus, we cannot exclude the possibility that cells in the in vitro developmental model take a slightly different or alternate path towards lung cell fates, and we propose that further comparisons of our cells to in vivo developing cells, both mouse and human, will be important to address this possibility.

Given the primordial nature of early NKX2-1$^+$ lung progenitors, whether derived from iPSCs/ESCs or emerging in vivo in embryos (2, 56), specific lung progenitor markers have not been previously identified with certainty to enable their prospective isolation. While FOXP2 in mice has been proposed as a lung primordial marker (56) and CPM has been proposed as a cell surface marker for sorting NKX2-1$^+$ cells derived from human iPSCs (16), the previous lack of any available tool for specifically tracking or purifying live NKX2-1$^+$ cells has left uncertainty regarding the specificity of those markers for prospective lung progenitor isolation. Our profiling of NKX2-1^{GFP+} primordial progenitors by microarrays, single-cell RNA-Seq, and FACS-based screens reveals a cell surface phenotype, CD47^{hi}CD26^{lo}, which can be used to prospectively isolate NKX2-1$^+$ progenitors when derived from iPSCs in culture. Our findings validate the utility of CPM, recently published by Gotoh et al., to also serve as a cell sorting marker (16), although CPM is also expressed in NKX2-1$^-$ hepatic cells which emerge in this lung-directed-differentiation protocol. While *FOXP2* is enriched in the NKX2-1$^+$ progenitor population, our results indicate that it lacks lung specificity and is also expressed elsewhere in the protocol, for example, in day 15 NKX2-1$^-$ cells (Supplemental Table 1).

Like other published markers for the purification of iPSC-derived endoderm (e.g., CKIT/CXCR4), CD47 is broadly expressed in many tissues. However, in the iPSC model system it has particular utility as a marker that allows sorting of NKX2-1$^+$ lung progenitors with ~90% purity based on its unexpectedly high levels of cell surface expression compared with other cells. CD47 is a broadly expressed cell surface glycoprotein with diverse roles in cellular processes including apoptosis, proliferation, and migration. The extracellular domain of CD47 acts as a thrombospondin-1 (TSP-1) receptor but also interacts with integrins and SIRPα (57). CD47 is expressed in lung epithelial cells in vivo and in vitro where it has a role ascribed to regulating leukocyte migration into the lung (58). Future work will be needed to determine if CD47 has a unique functional role during early lung development.

Just as our results provide an increased understanding of the level of heterogeneity present in iPSCs undergoing initial lung lineage specification in culture, further work is needed to interrogate the increasing heterogeneity that appears to emerge with each subsequent lung differentiation step. Similar to our NKX2-1^{GFP} tool, engineering multicolored reporters that become activated in more differentiated lung epithelial lineages should gradually facilitate this understanding and enable purification of subsets of increasingly mature cells in order to understand and overcome this obvious heterogeneity. These approaches should facilitate modulations of later lung differentiation stages in order to efficiently pattern cells into proximal versus distal lineages and their downstream progeny.

In summary, we have purified human lung progenitors derived from iPSCs, and these cells are reminiscent of early stages of lung developmental differentiation. Our profiling of these cells as well as their precursors and progeny during the time course of directed differentiation has resulted in an understanding of their global transcriptomic programs at the single-cell level and provides a validated set of cell surface markers and transcription factors selectively enriched in these cells. It should now be possible to test whether cells of similar primordial lung progenitor phenotype remain in the lung postnatally or can be rederived in patients during responses to injury. Given the broad differentiation repertoire of the primordial progenitors, we anticipate that access to pure populations of these cells should facilitate basic developmental studies as well as clinical applications focused on disease modeling, drug development, and potentially future regenerative therapies.

Methods

Human PSC maintenance and gene editing. Previously published PSC lines iPSC17 (18), BU3 (19), and WA09 (H9 ESC) were maintained in feeder-free conditions on Matrigel (Corning) in mTeSR1 (Stem Cell Technologies) and passaged with Gentle Cell Dissociation Reagent (Stem Cell Technologies). In order to generate NKX2-1^{GFP} reporter PSC lines, TALENs or CRISPR-based technologies were employed as detailed in the supplement to introduce a DNA double-stranded break into the second intron of *NKX2-1* (Supplemental Figure 1A). A donor matrix containing a splice acceptor, *NKX2-1* exon 3, 2A-eGFP, and loxP-flanked PGK-*puro*ΔTK selection cassette was integrated by homologous recombination and targeted PSC clones resistant to puromycin underwent Cre-mediated excision of the loxP-flanked *puro*ΔTK selection cassette (Supplemental Figure 1B), followed by confirmation of cassette excision, karyotyping, and characterization as detailed in the supplement.

Human iPSC directed differentiation. Neuroectodermal NKX2-1^{GFP+} cells were generated using STEMDdiff Neural Induction Medium (STEMCELL Technologies) according to the manufacturer's protocol. On day 6 of differentiation, 2 μM purmorphamine (Stemgent) was added to the base media. NKX2-1$^+$ cells were sorted on days 12–14. Thyroid NKX2-1^{GFP+} cells were derived using our recently published protocol (19) and lung NKX2-1^{GFP+} cells were generated by adapting published protocols (8, 9). For both lung and thyroid differentiations we first induced definitive endoderm using a STEMdiff Definitive Endoderm Kit (STEMCELL Technologies) according to the accompanying protocol. After approximately 72 to 84 hours, we harvested and analyzed cells by flow cytometry for efficiency of definitive endoderm induction by the coexpression of the surface markers c-KIT and CXCR4 (see supplemental experimental procedures). After definitive endoderm induction cells were plated in small clumps at approximately 50–150,000 cells/cm^2 on Matrigel-coated plates in complete serum-free differentiation media (CSFDM) supplemented with 2 μM dorsomorphin (Stemgent) and 10 μM SB431542 (Tocris) for 72 hours. Y-27632 (10 μM, Tocris) was added for the first 24 hours. To specify thyroid epithelium, differentiation medium was changed on day 6 to CSFDM supplemented with 250 ng/ml rhFGF2 (R&D Systems), 100 ng/ml rhBMP4 (R&D Systems), and 100 ng/ml heparin sodium salt (Sigma-Aldrich) according to our published methods (19). To specify lung epithelium, differentiation medium was changed on day 6 to CFKBRa: CSFDM supplemented with 3 μM CHIR99021 (Tocris), 10 ng/ml rhFGF10, 10 ng/ml rhKGF, 10 ng/ml rhBMP4 (all from R&D Systems), and 50–100 nM Ra (Sigma-Aldrich) (9). Day 15 cells were dissociated with 0.05% trypsin (Thermo Fisher Scientific) followed by resuspending as small clumps in CSFDM supplemented with 3 μM CHIR99021, 10 ng/ml rhFGF10, and 10 ng/ml rhKGF (CFK medium) and plated on freshly-coated Matrigel (Corning, 354277) plates. Y-27632 (10 μM) was added to CFK medium for the first 24 hours. On day 22 the medium was changed to CFK+DCI: CFK medium plus 50 nM dexamethasone (Sigma-Aldrich), 0.1 mM 8-bromoadenosine 3′,5′-cyclic monophosphate (8-Br-cAMP) sodium salt (Sigma-Aldrich), and 0.1 mM 3-isobutyl-1-methylxanthine (IBMX) (Sigma-Aldrich).

Sorting iPSC-derived lung progenitors and organoid generation. On day 15, live cells identified by propidium iodide (PI) exclusion were sorted by flow cytometry based on GFP expression for downstream applications including RNA analysis and generating organoids. To generate unsorted organoids, day 15 cells were dissociated with trypsin and pelleted cells were resuspended in Matrigel (Corning, 356230). CFK medium was then added to each well, supplemented with 10 μM Y-27632 medium for the first 24 hours. To generate sorted NKX2-1^{GFP+} or NKX2-1$^{GFP–}$ organoids, we resuspended the relevant sorted populations in Matrigel at a density of 50,000 cells per 50 to 100 μl Matrigel and allowed to gel as above. Further methods are detailed in the supplement.

Single-cell RNA-Seq analysis of day 15 iPSC-derived lung progenitors. Day 15 NKX2-1^{GFP+} and BU3 unsorted cells were generated using the lung protocol, dissociated, and sorted as described above. Fluidigm C1 integrated fluidics circuits (IFCs) were used to capture individual live cells, lyse, convert polyA$^+$ RNA into full-length cDNA, amplify and generate cDNA according to the manufacturer's protocol ("Using C1 to Generate Single-Cell cDNA Libraries for mRNA Sequencing", Fluidigm, PN 100-7168). Paired-end 50-bp reads were aligned with an average of 2.8 million reads per cell per end. Aligned, mapped reads were statistically analyzed using methods detailed in the supplement. The clustering, PCA, and significance testing were performed using SCICAST (details and a walkthrough can be found at https://github.com/iandriver/SCICAST) with additional hierarchical clustering linkage, Pearson's correlation coefficients, ANOVA, and FDR-adjusted *P* value calculation methods detailed in the supplement. Unbiased cell clustering was performed and cells ordered in pseudotime using Monocle 2 (47) (see supplemental experimental procedures). All raw data gene expression files can be downloaded from the NCBI Gene Expression Omnibus (GEO GSE96106).

Isolation of primary human fetal lung epithelium. Week 10 and 21 human lung tissues or their purified cell derivatives were isolated in the Guttentag lab using methods detailed in the supplement and previously published (59, 60).

Microarray analysis. Biological triplicates of all samples except human fetal lung were prepared. Biological duplicates from 1 embryo (uncultured naive lung epithelium) and a singlicate from a different embryo (differentiated AT2 cells) were prepared for the human fetal lung sample controls. Global gene expression in all 27 samples was analyzed by Affymetrix GeneChip Human Gene 2.0 ST arrays using methods and computational analyses detailed in the supplement. Differential gene expression with respect to experimental group across all samples was assessed by performing a 1-way ANOVA with correction for multiple hypothesis testing using the Benjamini-Hochberg FDR. All raw data gene expression files can be downloaded from the NCBI Gene Expression Omnibus (GEO GSE83310).

Recombination with mouse embryonic LgM. Recombinations were performed essentially as previously described (25). Briefly, small GFP$^+$ or GFP$^-$ fragments of day 21 human iPSC organoids were recombined with 10 to 12 pieces of mouse E12 LgM manually dissected free of any epithelia as described (25). The LgM rudiments were teased into close apposition to the human fragments with microsurgery knives (Fine Science Tools, Inc.). After overnight culture to promote tissue adherence, the recombinants were transferred to the surface of an 8-μm pore size Whatman nuclepore filter and cultured for 5 to 7 days in BGJb medium containing 20% FBS, 0.2 mg/ml vitamin C (Sigma-Aldrich), and 5 μg/ml recombinant mouse amino-terminal SHH (R&D Systems) to promote mesenchyme viability (61). The recombinants were maintained for 7 days, with medium changes every other day. Dexamethasone (50 nM) was added to the medium for the final 48 hours to promote lung epithelial differentiation.

RT-qPCR. RNA extracts were converted to cDNA and analyzed during 40 cycles of real-time PCR using Taqman probes (Applied Biosystems) according to methods detailed in the supplement. Relative gene expression, normalized to 18S control, was calculated as FC in 18S-normalized gene expression, over baseline, using the $2^{(-\Delta\Delta CT)}$ method. Unless otherwise specified in the text, baseline, defined as FC = 1, was set to undifferentiated stem cell levels, or if undetected, a cycle number of 40 was assigned to allow FC calculations.

Statistics. Statistical methods relevant to each figure are outlined in the accompanying figure legend. Unless otherwise indicated unpaired, 2-tailed Student's *t* tests were applied to 2 groups of $n = 3$ or more samples, where each replicate (n) represents either entirely separate differentiations from the PSC stage or replicates differentiated simultaneously and sorted into separate wells. A P value less than 0.05 was considered to indicate a significant difference between groups.

Study approval. The IRB of Boston University approved the generation and differentiation of human iPSCs with documented informed consent obtained from participants. Human lung tissues were obtained in the Guttentag laboratory under protocols originally reviewed by the IRB at the Children's Hospital of Philadelphia and subsequently reviewed by Vanderbilt University. Experiments involving mouse lung recombinant cultures were approved by the IACUC of Cincinnati Children's Medical Center.

For details of the methods used for immunostaining, mRNA in situ hybridization, and genomic DNA Southern blotting please see the supplemental procedures.

Author contributions

FH, PK, BRD, and DNK conceived the work, designed the experiments, and wrote the manuscript. PK, AMC, and BRD designed the targeting strategy and generated NKX2-1^{GFP} iPSCs/ESCs. FH, AJ, KBM, and DCT performed the lung-directed differentiation experiments. FH, AAK, ANH, and DNK developed the human thyroid differentiation. JMS performed the mouse LgM recombination experiments. SN, BGW, and ASK performed the time-lapse microscopy. SG provided human fetal lung samples. SXH provided critical technical assistance. NS assisted with the microarray analysis. ID and JRR performed the single-cell RNA-Seq analysis.

Acknowledgments

We wish to thank the members of the Kotton lab for insightful discussions. We thank Adam Gower and Eddy Drizik of the Boston University CTSI Microarray and Sequencing Resource Core for Affymetrix array processing and bioinformatics support (CTSA grant UL1-TR001430). We thank Anne Hinds of the Boston University Pulmonary Center for histology technical support. We thank Brian R. Tilton of the Boston University Flow Cytometry Core Facility for technical assistance. We are grateful to Greg Miller and Marianne James of the Boston University Center for Regenerative Medicine (CReM) for maintenance and characterization of patient-specific iPSCs, supported by NIH grants R24HL123828 and U01TR001810. We thank Linda Gonzales, formerly of the Division of Neonatology, Children's Hospital of Philadelphia, for sharing paraffin sections of fetal lung epithelium. F.H. was supported by a grant from the Cystic Fibrosis Foundation (HAWKIN15XX0), A.A.K. was supported by a grant from the Swiss National Science Foundation (PBBS P3-146612), B.R.D by grants from the Cystic Fibrosis Foundation (DAVISGO and DAVIS15XX1), and D.N.K. by NIH grants R01 HL095993, R01 HL108678, R01 HL122442, R01DK105029, R01HL128172, U01HL134766, and U01TR001810.

Address correspondence to: Brian R. Davis, Center for Stem Cell and Regenerative Medicine Brown Foundation, Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas 77030, USA. Phone: 713.500.3145; E-mail: brian.r.davis@uth.tmc.edu. Or to: Darrell N. Kotton, Center for Regenerative Medicine (CReM), Boston University and Boston Medical Center, 670 Albany Street, 2nd floor CReM, Boston, Massachusetts 02118, USA. Phone: 617.414.2969; E-mail:dkotton@bu.edu.

1. Kimura S, et al. The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev*. 1996;10(1):60–69.
2. Lazzaro D, Price M, de Felice M, Di Lauro R. The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. *Development*. 1991;113(4):1093–1104.
3. Minoo P, Su G, Drum H, Bringas P, Kimura S. Defects in tracheoesophageal and lung morphogenesis in Nkx2.1(-/-) mouse embryos. *Dev Biol*. 1999;209(1):60–71.
4. Coraux C, et al. Embryonic stem cells generate airway epithelial tissue. *Am J Respir Cell Mol Biol*. 2005;32(2):87–92.
5. Wang D, Haviland DL, Burns AR, Zsigmond E, Wetsel RA. A pure population of lung alveolar epithelial type II cells derived from human embryonic stem cells. *Proc Natl Acad Sci USA*. 2007;104(11):4449–4454.
6. Van Haute L, De Block G, Liebaers I, Sermon K, De Rycke M. Generation of lung epithelial-like tissue from human embryonic stem cells. *Respir Res*. 2009;10:105.
7. Green MD, et al. Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. *Nat Biotechnol*. 2011;29(3):267–272.
8. Longmire TA, et al. Efficient derivation of purified lung and thyroid progenitors from embryonic stem cells. *Cell Stem Cell*. 2012;10(4):398–411.
9. Huang SX, et al. Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. *Nat Biotechnol*. 2014;32(1):84–91.
10. Firth AL, et al. Generation of multiciliated cells in functional airway epithelia from human induced pluripotent stem cells. *Proc Natl Acad Sci U S A*. 2014;111(17):E1723–E1730.
11. Mou H, et al. Generation of multipotent lung and airway progenitors from mouse ESCs and patient-specific cystic fibrosis iPSCs. *Cell Stem Cell*. 2012;10(4):385–397.
12. Hawkins F, Rankin SA, Kotton DN, Zorn AM. The genetic programs regulating embryonic lung development and induced pluripotent stem cell differentiation. In: Jobe AH, Whitsett JA, Abman SH, eds. *Fetal and Neonatal Lung Development: Clinical Correlates and Technologies for the Future*. Cambridge: Cambridge University Press; 2016:ix–x.
13. Hawkins F, Kotton DN. Embryonic and induced pluripotent stem cells for lung regeneration. *Ann Am Thorac Soc*. 2015;12 Suppl 1:S50–S53.
14. Wong AP, et al. Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein. *Nat Biotechnol*. 2012;30(9):876–882.
15. Dye BR, et al. In vitro generation of human pluripotent stem cell derived lung organoids. *Elife*. 2015;4:e05098.
16. Gotoh S, et al. Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells. *Stem Cell Reports*. 2014;3(3):394–403.
17. Goulburn AL, et al. A targeted NKX2.1 human embryonic stem cell reporter line enables identification of human basal forebrain derivatives. *Stem Cells*. 2011;29(3):462–473.
18. Crane AM, et al. Targeted correction and restored function of the CFTR gene in cystic fibrosis induced pluripotent stem cells. *Stem Cell Reports*. 2015;4(4):569–577.
19. Kurmann AA, et al. Regeneration of thyroid function by transplantation of differentiated pluripotent stem cells. *Cell Stem Cell*. 2015;17(5):527–542.
20. Kim JE, et al. Investigating synapse formation and function using human pluripotent stem cell-derived neurons. *Proc Natl Acad Sci USA*. 2011;108(7):3005–3010.
21. Ma L, et al. Human embryonic stem cell-derived GABA neurons correct locomotion deficits in quinolinic acid-lesioned mice. *Cell Stem Cell*. 2012;10(4):455–464.
22. Rishniw M, et al. Molecular aspects of esophageal development. *Ann N Y Acad Sci*. 2011;1232:309–315.
23. Shannon JM, Hyatt BA. Epithelial-mesenchymal interactions in the developing lung. *Annu Rev Physiol*. 2004;66:625–645.
24. Shannon JM. Induction of alveolar type II cell differentiation in fetal tracheal epithelium by grafted distal lung mesenchyme. *Dev Biol*. 1994;166(2):600–614.
25. Shannon JM, Nielsen LD, Gebb SA, Randell SH. Mesenchyme specifies epithelial differentiation in reciprocal recombinants of embryonic lung and trachea. *Dev Dyn*. 1998;212(4):482–494.
26. Herriges JC, et al. Genome-scale study of transcription factor expression in the branching mouse lung. *Dev Dyn*. 2012;241(9):1432–1453.
27. Maeda Y, Davé V, Whitsett JA. Transcriptional control of lung morphogenesis. *Physiol Rev*. 2007;87(1):219–244.
28. Millien G, et al. Characterization of the mid-foregut transcriptome identifies genes regulated during lung bud induction. *Gene Expr Patterns*. 2008;8(2):124–139.
29. Vukicevic S, Helder MN, Luyten FP. Developing human lung and kidney are major sites for synthe-

sis of bone morphogenetic protein-3 (osteogenin). *J Histochem Cytochem*. 1994;42(7):869–875.

30. Takahashi H, Ikeda T. Transcripts for two members of the transforming growth factor-beta superfamily BMP-3 and BMP-7 are expressed in developing rat embryos. *Dev Dyn*. 1996;207(4):439–449.
31. Keegan CE, Herman JP, Karolyi IJ, O'Shea KS, Camper SA, Seasholtz AF. Differential expression of corticotropin-releasing hormone in developing mouse embryos and adult brain. *Endocrinology*. 1994;134(6):2547–2555.
32. Emanuel RL, Torday JS, Asokananthan N, Sunday ME. Direct effects of corticotropin-releasing hormone and thyrotropin-releasing hormone on fetal lung explants. *Peptides*. 2000;21(12):1819–1829.
33. Simard M, Côté M, Provost PR, Tremblay Y. Expression of genes related to the hypothalamic-pituitary-adrenal axis in murine fetal lungs in late gestation. *Reprod Biol Endocrinol*. 2010;8:134.
34. Provost PR, Tremblay Y. Genes involved in the adrenal pathway of glucocorticoid synthesis are transiently expressed in the developing lung. *Endocrinology*. 2005;146(5):2239–2245.
35. Herriges MJ, et al. Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. *Genes Dev*. 2014;28(12):1363–1379.
36. Herriges M, Morrisey EE. Lung development: orchestrating the generation and regeneration of a complex organ. *Development*. 2014;141(3):502–513.
37. Williams MC. Alveolar type I cells: molecular phenotype and development. *Annu Rev Physiol*. 2003;65:669–695.
38. Zorn AM, Wells JM. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol*. 2009;25:221–251.
39. Rankin SA, Kormish J, Kofron M, Jegga A, Zorn AM. A gene regulatory network controlling hhex transcription in the anterior endoderm of the organizer. *Dev Biol*. 2011;351(2):297–310.
40. Perkins AS, Mercer JA, Jenkins NA, Copeland NG. Patterns of Evi-1 expression in embryonic and adult tissues suggest that Evi-1 plays an important regulatory role in mouse development. *Development*. 1991;111(2):479–487.
41. Attar MA, Bailie MB, Christensen PJ, Brock TG, Wilcoxen SE, Paine R. Induction of ICAM-1 expression on alveolar epithelial cells during lung development in rats and humans. *Exp Lung Res*. 1999;25(3):245–259.
42. Li Y, Linnoila RI. Multidirectional differentiation of Achaete-Scute homologue-1-defined progenitors in lung development and injury repair. *Am J Respir Cell Mol Biol*. 2012;47(6):768–775.
43. Metzger DE, Xu Y, Shannon JM. Elf5 is an epithelium-specific, fibroblast growth factor-sensitive transcription factor in the embryonic lung. *Dev Dyn*. 2007;236(5):1175–1192.
44. Lu MM, Li S, Yang H, Morrisey EE. Foxp4: a novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues. *Mech Dev*. 2002;119 Suppl 1:S197–S202.
45. Yu Y, et al. Gene expression profiling in human fetal liver and identification of tissue- and developmental-stage-specific genes through compiled expression profiles and efficient cloning of full-length cDNAs. *Genome Res*. 2001;11(8):1392–1403.
46. Wilson AA, et al. Emergence of a stage-dependent human liver disease signature with directed differentiation of alpha-1 antitrypsin-deficient iPS cells. *Stem Cell Reports*. 2015;4(5):873–885.
47. Trapnell C, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;32(4):381–386.
48. Shalek AK, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*. 2013;498(7453):236–240.
49. Guo G, et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev Cell*. 2010;18(4):675–685.
50. Tang F, et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell*. 2010;6(5):468–478.
51. Buganim Y, et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*. 2012;150(6):1209–1222.
52. Lü J, Izvolsky KI, Qian J, Cardoso WV. Identification of FGF10 targets in the embryonic lung epithelium during bud morphogenesis. *J Biol Chem*. 2005;280(6):4834–4841.
53. Kho AT, et al. Transcriptomic analysis of human lung development. *Am J Respir Crit Care Med*. 2010;181(1):54–63.
54. Kido T, et al. CPM is a useful cell surface marker to isolate expandable bi-potential liver progenitor cells derived from human iPS cells. *Stem Cell Reports*. 2015;5(4):508–515.
55. Somers A, et al. Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. *Stem Cells*. 2010;28(10):1728–1740.
56. Sherwood RI, Chen TY, Melton DA. Transcriptional dynamics of endodermal organ formation. *Dev Dyn*. 2009;238(1):29–42.
57. Barclay AN, Van den Berg TK. The interaction between signal regulatory protein alpha (SIRPα) and CD47: structure, function, and therapeutic target. *Annu Rev Immunol*. 2014;32:25–50.
58. Herold S, et al. Alveolar epithelial cells direct monocyte transepithelial migration upon influenza virus infection: impact of chemokines and adhesion molecules. *J Immunol*. 2006;177(3):1817–1824.
59. Wade KC, et al. Gene induction during differentiation of human pulmonary type II cells in vitro. *Am J Respir Cell Mol Biol*. 2006;34(6):727–737.
60. Gonzales LW, Guttentag SH, Wade KC, Postle AD, Ballard PL. Differentiation of human pulmonary type II cells in vitro by glucocorticoid plus cAMP. *Am J Physiol Lung Cell Mol Physiol*. 2002;283(5):L940–L951.
61. Weaver M, Batts L, Hogan BL. Tissue interactions pattern the mesenchyme of the embryonic mouse lung. *Dev Biol*. 2003;258(1):169–184.