Supplemental Figures and Procedures

Fig. S1 – Generation, characterization and validation of NKX2-1-GFP 'knock-in' reporter line

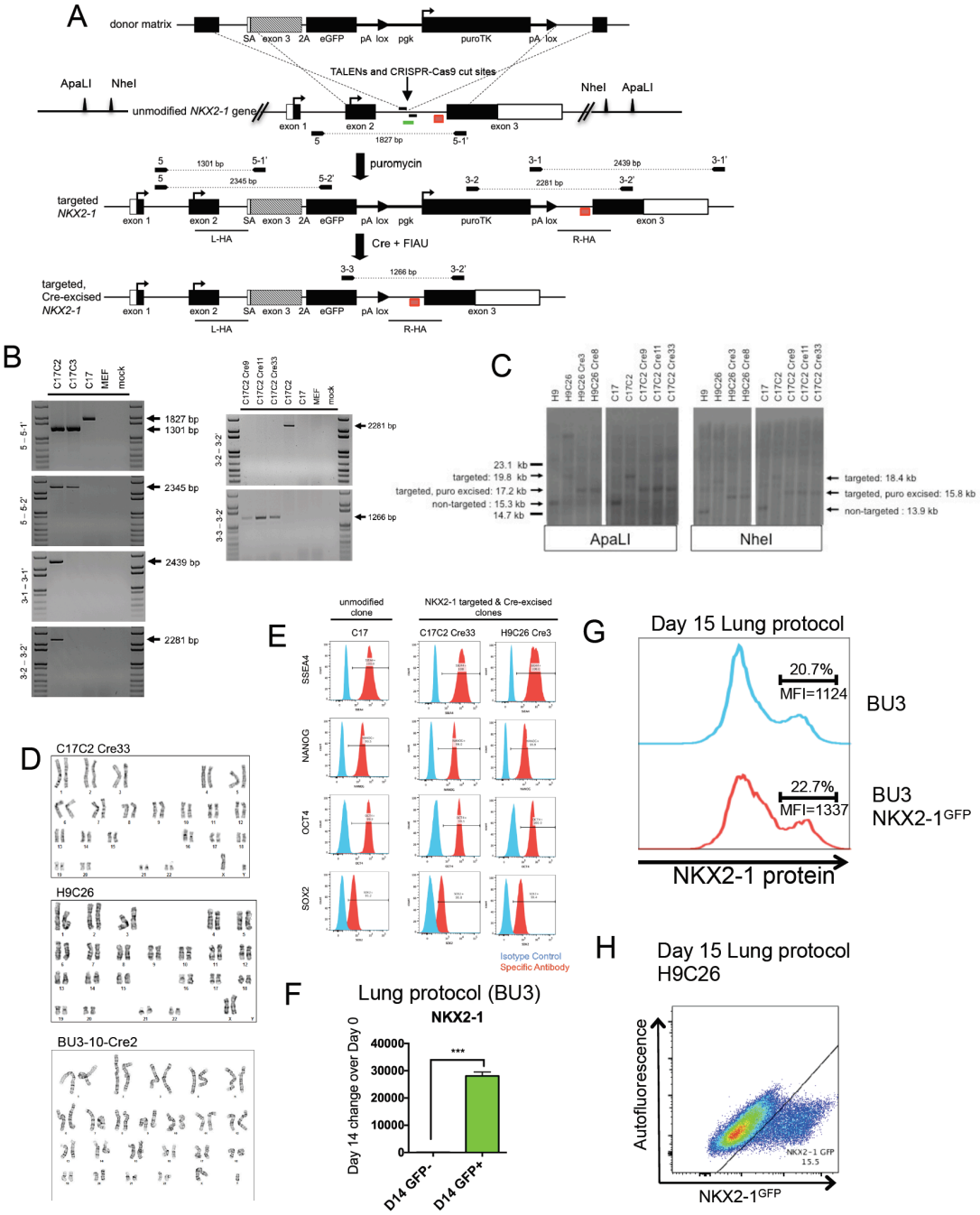

Fig. S1: Gene editing of the human NKX2-1 locus to engineer NKX2-1^{GFP} reporter iPSC lines.

A) Schematic of the targeting strategy used to introduce a 2A-GFP cassette at the end of exon 3 in order to preserve expression of each allele, in an effort to avoid haploinsufficiency. The donor vector includes a floxed PGK promoter-driven antibiotic selection cassette (puroTK, consisting of a fused Puro resistance-thymidine kinase [TK] cassette) which is excised following transient Cre recombinase exposure. FIAU exposure is used to kill any cells carrying the thymidine kinase (TK) cassette or to confirm successful PuroR-TK cassette excision. The TALENs cut site (in the case of C17 or WA09 PSCs,

black bars = left and right TALENs) or CRISPR-Cas9 cut site (in the case of BU3 iPSCs, green bar = guide RNA), primer binding sites (black arrows), or southern blot probe binding sites (red box) are indicated, to be used in panels B and C. Left and right arms of homology (L-HA and R-HA) as well as restriction endonuclease sites ApaLI and NheI are indicated. B) NKX2-1 locus targeting screening by PCR of gDNA from iPSCs, using the primer pairs indicated in A. C) Southern blot of gDNA extracted from each indicated iPSC clone after restriction enzyme digest with ApaLI or NheI and probing of the gel with the probe indicated as a red box in panel A. D) Karyotyping of each indicated iPSC and ESC clone after gene editing and antibiotic selection cassette excision. Normal 46XY and 46XX karyotypes are shown. E) Characterization by flow cytometry of pluripotency marker expression in each indicated clone before vs after gene editing. F) *NKX2-1* mRNA in Day 14 GFP+ vs GFP- sorted cells using BU3NKX2-1^{GFP}. (G) Expression of NKX2-1 protein by intracellular FACS staining on day 15 of the lung differentiation in pre-targeted BU3 vs. homozygous targeted BU3NKX2-1^{GFP}. The percent positive for NKX2-1 protein as well as the mean fluorescence intensity (MFI) of the NKX2-1 staining is indicated for each clone. (H) Percentage H9C26NKX2-1^{GFP+} cells on day 15 of lung directed differentiation.

Fig. S2 – Stage-specific optimization of lung differentiation protocol

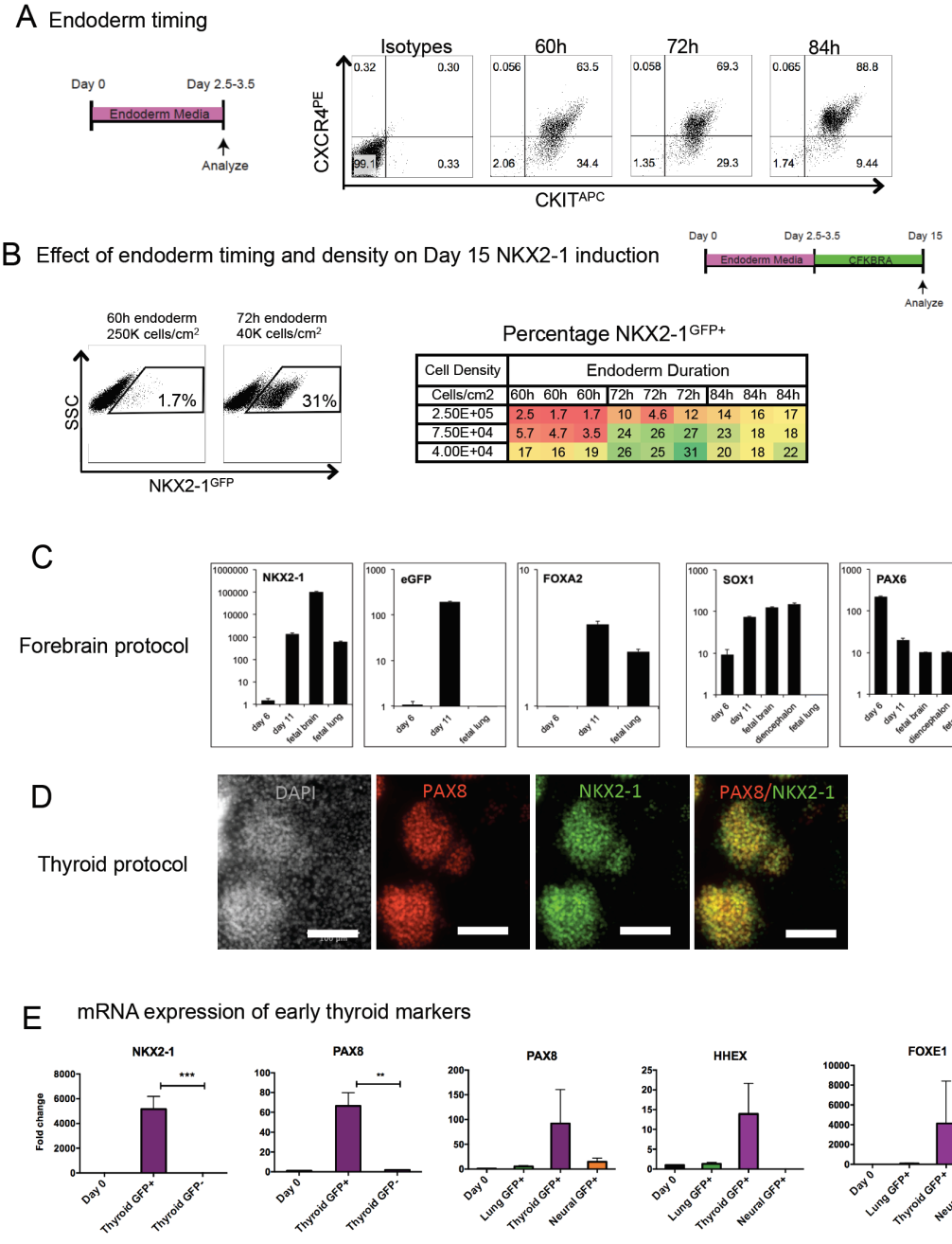

Fig. S2: Stage-specific optimization of lung directed differentiation protocol and comparison to thyroid and forebrain directed differentiation protocols.

A) Schematic of endoderm induction and efficiency of endoderm induction at 60, 72 and 84 hours of differentiation based on co-expression of CKIT and CXCR4 measured by flow cytometry (BU3). B) Schematic of experiment and representative flow cytometry plots of the effect of endoderm timing and density of replating on Day 15 NKX2-1^{GFP+} induction. Heatmap of Day 15 NKX2-1^{GFP+} percentage from 60, 72 or 84 hours endoderm induction at different cell plating densities (250,000, 75000 and 40000 cells per cm^2) (BU3). C) RT-qPCR of day 6 and sorted NKX2-1^{GFP+} cells on day 11 from the forebrain protocol compared to primary fetal brain, diencephalon and fetal

lung controls. Data indicate individual biological replicates (n=3) with mean±SD. Fold changes (RT-qPCR; $2^{(-\Delta\Delta CT)}$) in mRNA expression are compared to day 0 (C17). D) Immunostaining of day 17 cells from the thyroid directed differentiation protocol for NKX2-1 and PAX8 proteins (BU3). Cell nuclei are counterstained with DAPI. Scale bar = 100µm**.** E) mRNA expression by RT-qPCR in day 18 C17 NKX2-1^{GFP+} vs NKX2-1^{GFP-} cells from thyroid protocol (two left panels) and in sorted NKX2-1^{GFP+} cells from lung, thyroid and forebrain protocols (three right panels). Data indicate individual biological replicates (n=3) with mean±SD. Fold change [RT-qPCR; $2^{(-\Delta\Delta CT)}$] in mRNA expression are compared to day 0.

Fig. S3

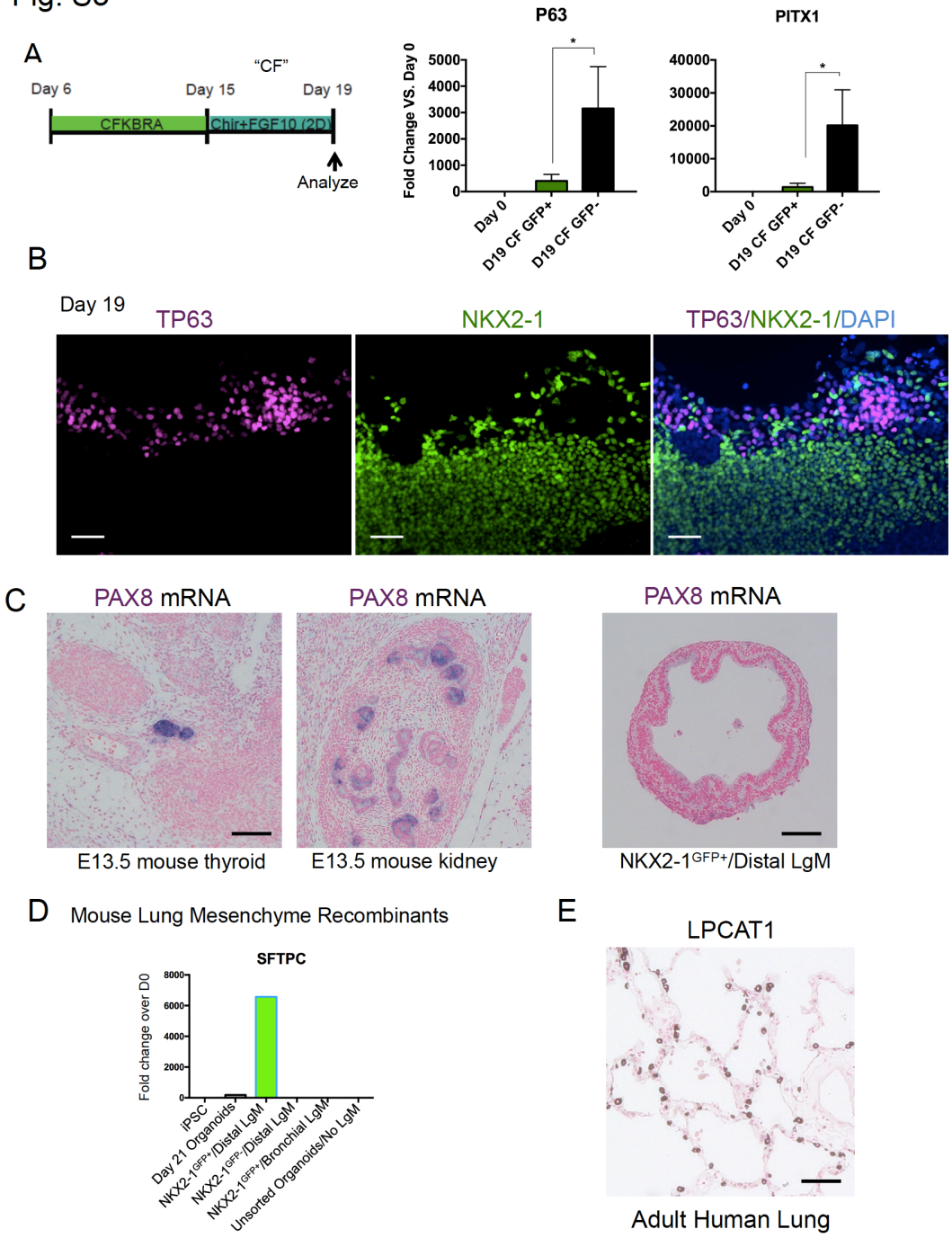

Fig. S3: Further characterization of iPSC-derived cells on days 15-19 of differentiation and after recombination with E12 mouse lung mesenchyme.

A) Schematic of experiment. On day 15 of the lung differentiation the media is changed from CFKBRA to Chir+FGF10 (CF) until day 19. GFP+ vs GFP- cells are sorted on day 19 for analysis. Fold change of *TP63* and *PITX1* mRNA expression in Day 15 and Day 19 GFP+ vs GFP- compared to day 0 by RT-qPCR; $2^{(-\Delta\Delta CT)}$, n=3 replicates for day 19 B) Immunostaining of TP63 (magenta), NKX2-1 (green) on day 19 demonstrates minimal colocalization (C17). Nuclei labeled with DAPI. Scale bar: 50 μm. C) *PAX8* mRNA

expression by in situ hybridization in embryonic mouse thyroid, kidney and in recombinant of C17 NKX2-1^{GFP+} microdissected organoid with E12.5 mouse distal lung mesenchyme (see also Figure 3). Scale bar: 200 µm. D) Fold change of human *SFTPC* mRNA expression compared to day 0 by RT-qPCR; $2^{(-\Delta\Delta CT)}$ indicates that distal rather than bronchial mouse LgM induces expression of *SFTPC* (C17). E) Adult human lung control for LPCAT1 immunostains (see also Figure 3D). Scale bar: 100 µm.

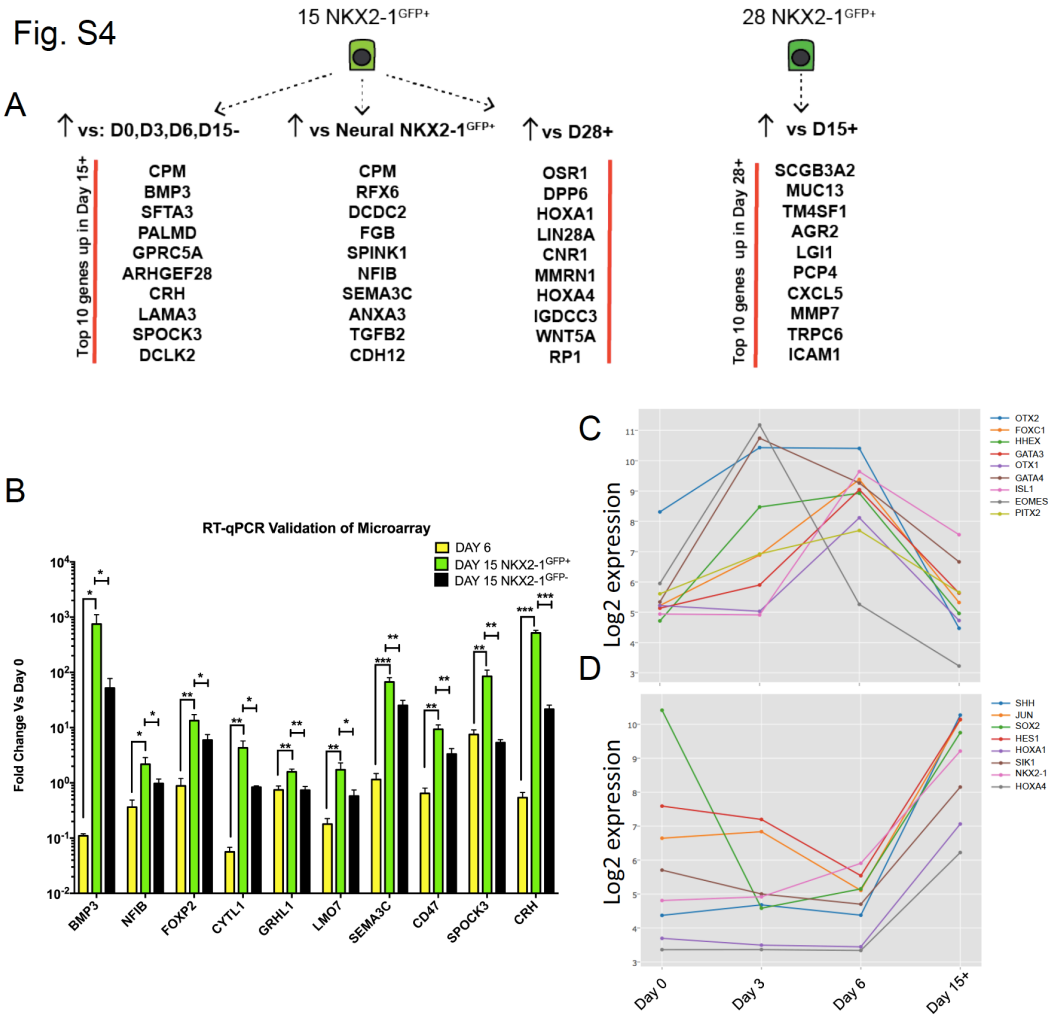

Fig. S4: Gene signature of iPSC-derived lung progenitors and a kinetic of foregut and early lung transcription factors.

A) From the microarray data of the top 100 genes differently expressed in day 15 NKX2-1^{GFP+} vs day 0, day 3, day 6 and day 15 NKX2-1^{GFP-} cells (see supplementary table 1), genes were ranked by fold change and filtered by FDR<0.01) in order to identify the top 10 differentially expressed genes in common across each comparison. Also listed are the top 10 genes differentially expressed in day 15 NKX2-1^{GFP+} cells vs neural NKX2-1^{GFP+} and day 15 NKX2-1^{GFP+} cells vs day 28 NKX2-1^{GFP+} cells. Top 10 genes differentially expressed by day 28 NKX2-1^{GFP+} compared to day 15 NKX2-1^{GFP+} cells (right panel). B) Validation of mRNA expression levels of key genes identified by time series microarray trancriptomic profiling. Fold change in expression levels for each indicated transcript is shown for Day 6, Day 15 GFP+ and Day 15 GFP- cells, compared to day 0 by RT-qPCR; $2^{(-\Delta\Delta CT)}$. Data indicate mean±SD, *p ≤0.05, **p ≤0.01, ****p ≤0.0001 (Student's t-test); n=3 biological replicates. C and D) Graph of log2 expression (y-axis) by microarrays of the indicated differentially expressed transcription factors on days 0, 3, 6, and 15+. Transcription factors with FC>4, FDR<0.01 and with known roles in foregut endoderm and developing lung development are included. All time points were from C17 iPSCs.

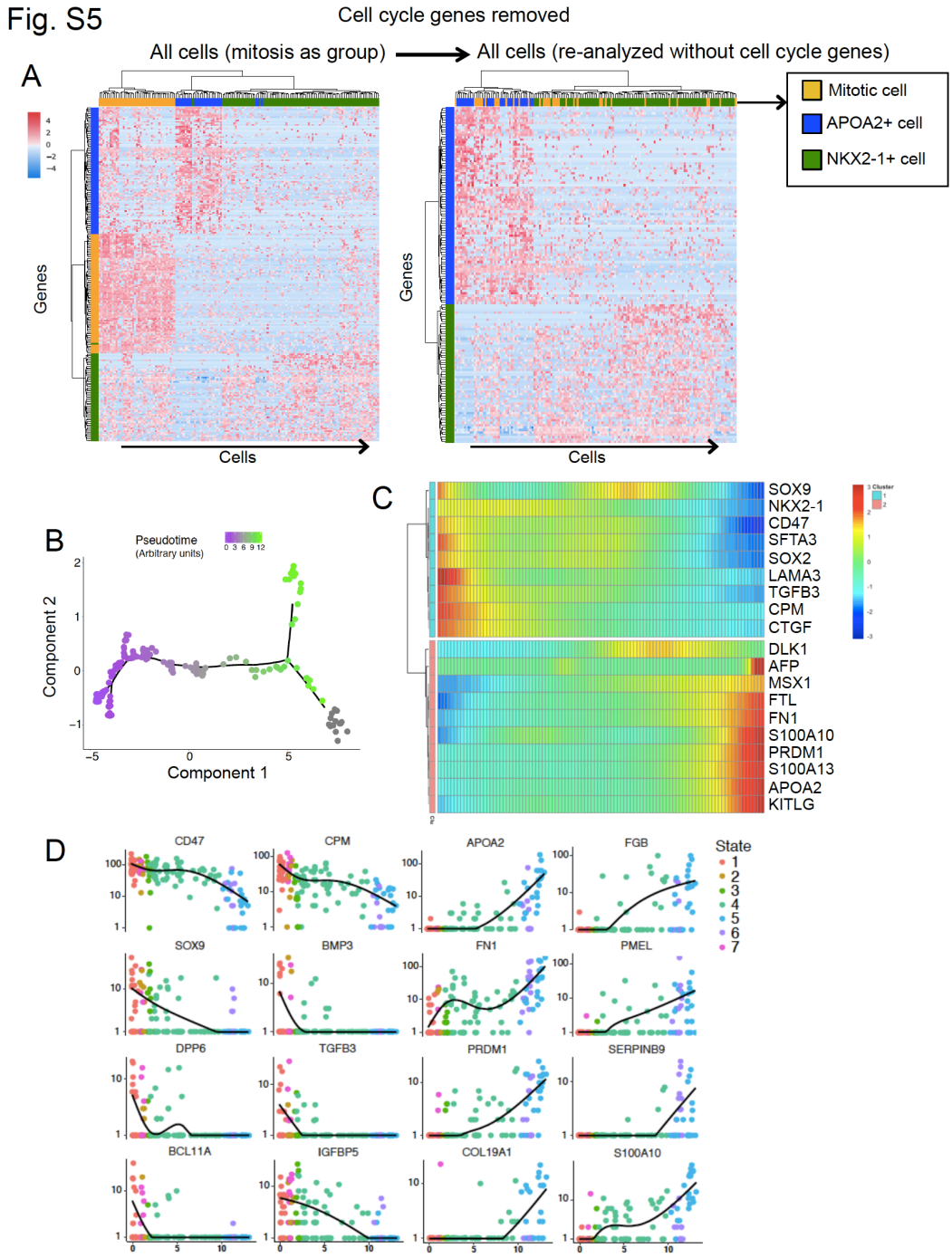

Fig. S5: Single cell RNA-Seq analysis of 153 iPSC-derived cells on day 15 of differentiation.

A) Unbiased hierarchical clustering and significance testing was performed using SCICAST. Top significant differentially expressed genes from the three cell clusters: Mitotic (Yellow), APOA2+ (Blue), and NKX2-1+ (Green) were selected and hierarchical clustering re-run with the selected genes (left panel). Mitotic genes were removed and hierarchical clustering was re-run with the same cell group assignments as in (A) revealing how the underlying identity of the mitotic cells falls among the two cell identities; APOA2 (Blue) or NKX2-1 (Green) (right panel). B) Pseudotime plot of all cells by applying the unbiased Monocle clustering algorithm, used in Figure 6. Units are arbitrary. C)

Heatmap of top genes that follow a similar kinetic in pseudotime. (D) 15 of the top 30 genes (by FDR-adjusted p value) that follow similar kinetics in pseudotime. Day 15 C17 NKX2-1^{GFP+} and BU3 (untargeted) cells were used in this single cell sequencing experiment.

Fig. S6

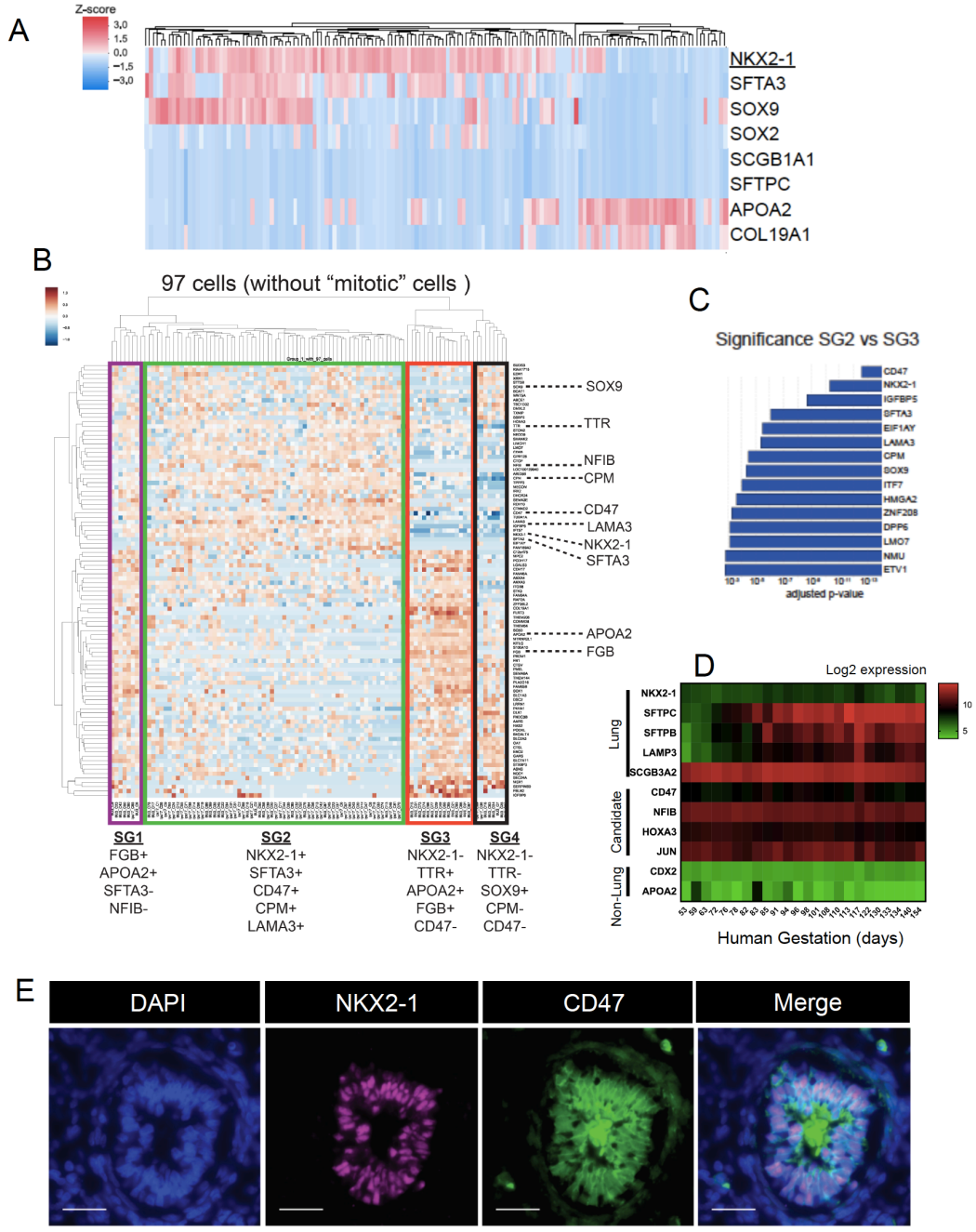

Fig. S6: Further bioinformatic analysis of cell clusters and differentially expressed genes identified by single cell RNA-seq.
A) Heatmap illustrating the expression of key lung genes including primordial markers (NKX2-1, SFTA3), early proximal vs distal markers (SOX2, SOX9), and more differentiated proximal vs distal markers (SCGB1A1, SFTPC) in the single-cell RNA-Seq analysis of Day 15 iPSC-derived cells (C17 NKX2-1^{GFP+} and BU3). B) Heatmap of 97 cells reanalyzed after removing the mitotic cell clusters (CC1 and 2) from Figure 6C. Unsupervised hierarchical clustering reveals four cell subgroups (SG1-4). Key genes are indicated and highly statistically significant, differentially expressed genes that identify each SG are summarized below the x-axis. C) Top 15 genes differentially expressed between NKX2-1+ subgroup SG2 vs NKX2-1- subgroup SG3, ranked by

adjusted p-value. D) Heatmap of expression of markers from a database of 24 human fetal lung samples ranging in gestation from 53 to 154 days (53). E) Immunostain of human fetal lung (week 10) for CD47 (green) and NKX2-1 proteins (pink). Nuclei counterstained with DAPI, scale bar = 25 μm.

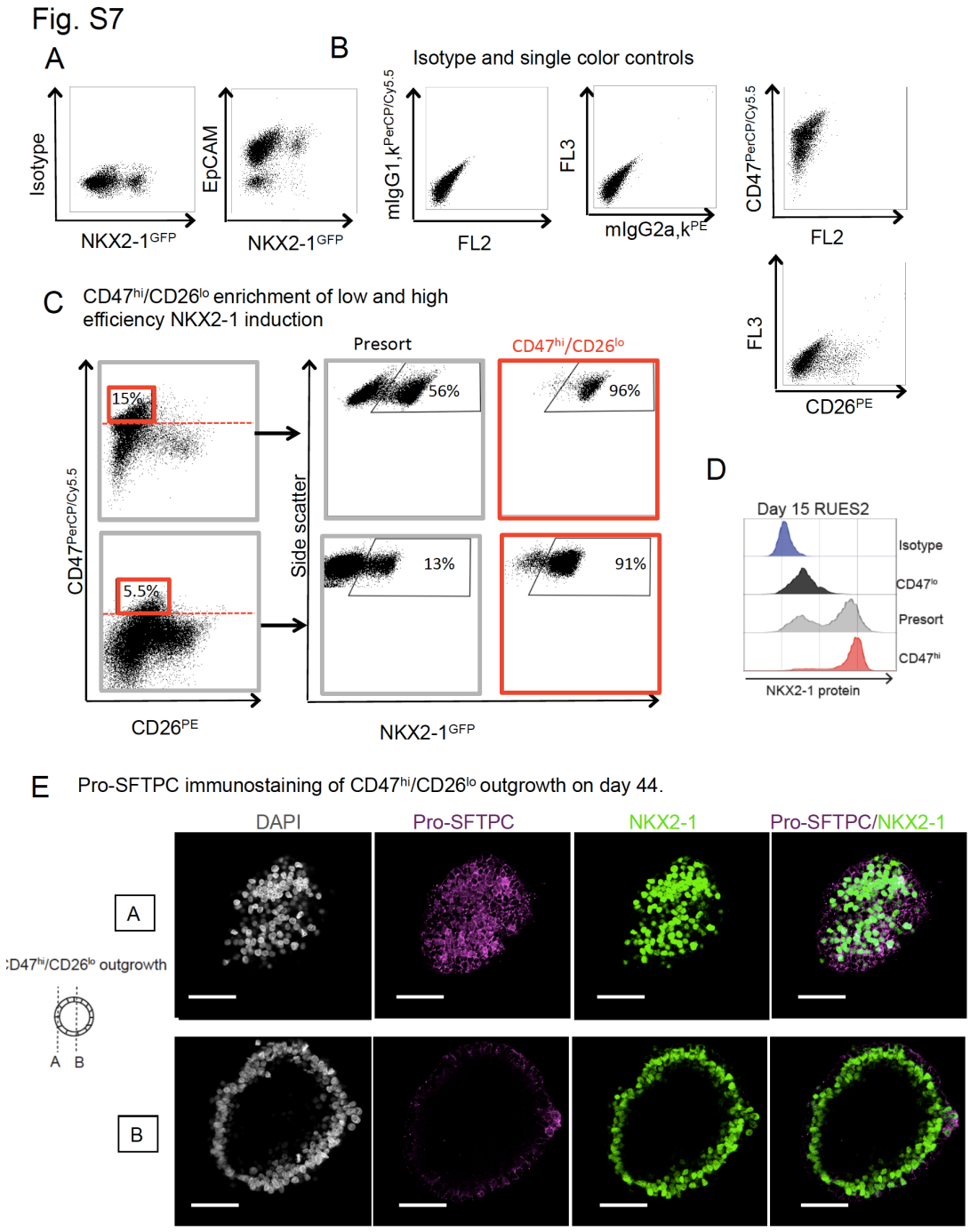

Fig. S7: Further cell surface profiling and prospective isolation of iPSC-derived NKX2-1+ primordial lung progenitors by CD47^{hi}/CD26^{lo} cell sorting.

A) Representative flow cytometry dot plot of EpCAM and NKX2-1^{GFP} on day 15 of differentiation vs isotype control (C17). B) Flow cytometry dot plots of isotype (mIgG1,k$^{PerCP/Cy5.}$5 and mIgG2a,k^{PE}), or single color (CD47$^{PerCP/Cy5.5}$ and CD26^{PE}) controls on day 15 of lung differentiation. C) Representative flow cytometry dot plots of CD47^{hi}/CD26^{lo} enrichment for NKX2-1^{GFP} on day 15 of a high efficiency (top row, presort=56%) and low efficiency (lower row, presort=13%) lung differentiation in independent experiments. D) Expression of intracellular NKX2-1 protein analyzed by FACS of RUES2 cells on day 15

of lung directed differentiation. Levels of NKX2-1 on day 15 are shown for presort, $CD47^{hi}/CD26^{lo}$ and $CD47^{lo}$ populations, compared to isotype antibody stained control day 15 cells. (E) Confocal microscopy of outgrowth organoids from progenitors sorted on $CD47^{hi}/CD26^{lo}$ on day 13 and analyzed on day 44 by immunostaining for NKX2-1 protein (green) and pro-SFTPC (purple (C17)). Cartoon depicts the two focal planes (panel A and B) of a single spherical organoid; the merge image in panel A is also shown in Figure 7G. Nuclei are counterstained with DAPI. Scale bar =25um. Each row is a different section through the same organoid.

Table S1: Summary analysis file of the microarray profiling of major stages of the lung directed differentiation protocol compared to human fetal lung and neural NKX2-1^{GFP+} controls (See also Figures 4-5, S4) can be found in the bioinformatics portal at kottonlab.com.. The file includes GO terms, KEGG pathways, nominal p values and FDR q values for the one way ANOVA, signed linear fold changes for each pairwise comparison between groups (e.g. for Day 15 NKX2-1^{GFP+} vs Day 15 NKX2-1^{GFP-}, +2 = 2-fold higher in Day 15 NKX2-1^{GFP+} than in Day 15 NKX2-1^{GFP}, and -2 = 2-fold lower in Day 15 NKX2-1^{GFP+} than in Day 15 NKX2-1^{GFP-}). Also included is the log2 (expression) across all samples used in the analysis, laid over a colored representation (heatmap), scaled so that red and blue indicate expression values ≥ 2 standard deviations above and below, respectively, the row-wise mean (white) computed across all samples in the row. Differentiations were performed with C17 iPSCs.

Table S2: The top 100 differentially expressed genes (ranked by fold change, filtered by FDR<0.01) of day 15 NKX2-1^{GFP+}, day 28 NKX2-1^{GFP+} and day 28 NKX2-1^{GFP-} samples (top row indicates sample being analyzed) across multiple comparisons (second row indicates sample being compared to). Differentiations were performed with C17 iPSCs.

Table S3: The most highly expressed transcription factors or genes with transcription factor activity between anterior foregut endoderm, day 15 NKX2-1^{GFP+} and day 28 NKX2-1^{GFP+} samples were identified in Fig. 5. Genes with an asterisk in Fig. 5 are detailed in this table with corresponding supporting reference for a previously identified role in foregut or lung development. Differentiations were performed with C17 iPSCs.

Movie 1: Time-lapse microscopy over 25 hours of unsorted organoids on day 19-20 of the lung directed differentiation protocol generated with the C17 NKX2-1^{GFP} iPSC line (merge of phase and GFP fluorescence images taken every 30 min).

Supplemental Experimental Procedures

Human iPSC Derivation and Maintenance

The iPSC line "C17" was generated from a patient with cystic fibrosis as recently described (1). "BU3" iPSC line (2) was generated from the peripheral blood of a normal donor according to our published methods. WA09 (H9) ESC line was obtained from WiCell (Madison, WI). RUES2 was a gift of the Brivanlou laboratory (The Rockefeller University, New York City, NY). iPSCs were initially maintained on a feeder layer of mitomycin-inactivated mouse embryonic fibroblasts in human iPSC media (WICELL feeder-dependent protocol) and subsequently transitioned to feeder free conditions on Matrigel (Corning) in mTeSR1 (Stem Cell Technologies) and passaged with Gentle Cell Dissociation Reagent (Stem Cell Technologies).

Generating a Fluorescent Reporter by CRISPR- or TALEN-targeting the NKX2-1 locus.

NXK2-1^{GFP+} iPSC: Transcription Activator Like Effector Nucleases (TALENs), with left TAL effector DNA-binding sequence TCGAGCGCCCGGCCCGG and right TAL effector DNA-binding domain AGTCTGGGCAGGTGGGA, were designed by Cellectis to introduce a DNA double stranded break into the second intron of NKX2-1 at a site at least 50bp distant from any known SNPs in either human iPSC (C17) or ES line (WA09) (Figure S1A). CRISPR-mediated introduction of a DNA double strand break in this same region in BU3 iPSCs was engineered by designing a guide RNA of sequence GCCCTCTCTCCAGCGGAGTC delivered through transfection of a plasmid encoding a Pol II-driven Cas9 (pSpCas9(BB)-2A-GFP; Addgene, PX458) as well as the Pol III-driven gRNA (pU6-gRNA; Sigma). Detection of error-prone non-homologous end joining (NHEJ) after TALEN-introduced or CRISPR introduced DNA break resulted in cleavage efficiencies of up to 24% in C17 iPSCs and 27% in WA09 determined by Surveyor nuclease assay (Integrated DNA Technologies)(data not shown). A donor matrix containing a splice acceptor, NKX2-1 exon 3, 2A-eGFP and loxP-flanked PGK-puroΔTK selection cassette was integrated by homologous recombination (in the presence of either TALEN or CRISPR based targeting). Targeted, puromycin-resistant clones were obtained after co-electroporation of TALEN expression plasmids and delivery of the targeting donor vector using Amaxas Nucleofector (Lonza). Isolated clones were confirmed by a total of four different PCRs spanning 5'-3' and 3'- 5' junctions between genomic DNA and donor matrix sequences or vice versa (Figure S1B, left panel). A unique PCR using primers 5 and 5-1'

was capable of binding to both the targeted and unmodified allele, amplifying a smaller and larger DNA fragment, respectively. Successful targeting was achieved in 1 out of 5 screened C17 iPSC, 10 out of 176 WA09 ESC, and 2 out 3 BU iPSC clones, including donor integration into both alleles and the generation of homozygous NKX2-1^{GFP} clones in both ESC/iPSC lines. Cre-mediated excision of the loxP-flanked puroΔTK selection cassette was confirmed by negative selection with FIAU, puromycin selection and PCR analysis (Figure S1B, right panel). Three homozygous targeted NKX2-1^{GFP+} ESC/iPSC reporter lines showed normal karyotype when analyzed at the Clinical and Research Cytogenetic Laboratory at the Texas Children's Hospital, Houston, TX (Figure S1D). Both targeted lines expressed pluripotency markers SSEA4, OCT-4, NANOG and SOX2 (Stemflow Human Pluripotent Stem Cell Transcription Factor Analysis Kit) (Figure S1E). Appropriate FITC mouse antihuman SSEA-4 (clone MC813-70) and FITC Mouse IgG3, kappa isotype (all BD Biosciences) were used as controls (Figure S1E). All cell lines, donor targeting vectors, and CRISPR constructs are available from the corresponding authors and can be viewed at www.kottonlab.com.

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, 2μM of purmorphamine (Stemgent) was added to the base media. NKX2-1+ cells were sorted on Day 12-14. Thyroid NKX2-1^{GFP+} cells were derived using our recently published protocol (2) lung NKX2-1^{GFP+} cells were generated by adapting published protocols (3, 4). For both lung and thyroid differentiations we first induced definitive endoderm using STEMDiff definitive endoderm kit (STEMCELL Technologies) according to the accompanying protocol. After 72 to 84 hours, as indicated in the results text, we harvested and analyzed cells by flow cytometry for efficiency of definitive endoderm induction by the co-expression of the surface markers CKIT (APC-conjugated mouse monoclonal antibody, Life Technologies CD11705) and CXCR4 (PE-conjugated mouse monoclonal antibody Life Technologies MHCXCR404) with appropriate APC (Life Technologies MG105) and PE (Life Technologies MG204) isotype controls and for intracellular endodermal markers SOX17 (APC goat anti-human SOX17, R&D Systems IC1924A) and FOXA2 (Alexa Fluor 488 goat anti-human FOXA2, R&D Systems IC2400G) with appropriate APC (R&D Systems IC108A) and Alexa Fluor 488 (R&D Systems, IC108G) isotype controls, respectively. After definitive endoderm induction cells were plated in small clumps at 50,000-150,000 cells/cm2 on Matrigel-coated plates in complete serum-free differentiation media (CSFDM) supplemented with 2μM Dorsomorphin (Stemgent), 10μM SB431542 (Tocris) for 72 hours. 10μM Y-27632 (Tocris) was added for the first 24 hours. CSFDM was composed of 375ml IMDM (ThermoFisher Scientific, 12440-053), 125ml Ham's F12 (Corning Cellgro, 10-080-CV), 50μg/ml Ascorbic acid (Sigma, A4544), 5ml B27 supplement (ThermoFisher Scientific, 12587-044), 2.5ml N2 supplement (ThermoFisher Scientific, 17502-048), 3.75ml bovine serum albumin (ThermoFisher Scientific, 15260-037), 5ml Glutamax (ThermoFisher Scientific, 35050061) 0.02μl monothioglycerol (Sigma, M6145) and 100μg/ml

Primocin (Invivogen). To specify thyroid epithelium, differentiation media was changed on day 6 to CSFDM supplemented with 250ng/ml rhFGF2 (R&D Systems), 100ng/ml of rhBMP4 (R&D) and 100ng/ml Heparin Salt (Sigma) according to our recently published methods (2). To specify lung epithelium, differentiation media was changed on day 6 to "CFKBRa"; CSFDM supplemented with 3μM CHIR99021 (Tocris), 10ng/ml rhFGF10, 10ng/ml rhKGF, 10ng/ml rhBMP4 (all from R&D Systems), 50-100nM Retinoic acid (Sigma) (4). Day 15 cells were dissociated by incubating in 0.05% trypsin (ThermoFisher Scientific) at 37°C for 2-4 minutes, aspirating trypsin, washing once with DMEM (ThermoFisher Scientific)+10% FBS (ThermoFisher Scientific), resuspending as small clumps in CSFDM supplemented with 3μM CHIR99021, 10ng/ml rhFGF10 and 10ng/ml rhKGF ("CFK" media) and plated on freshly-coated Matrigel (Corning 354277) plates. 10μM Y-27632 was added to "CFK" media for the first 24 hours. On day 22 media was changed to "CFK+DCI": "CFK" media plus 50nM dexamethasone (Sigma), 0.1mM 8-Bromoadenosine 3',5'-cyclic monophosphate (8-Br-cAMP) sodium salt (Sigma) and 0.1mM 3-Isobutyl-1-methylxanthine (IBMX) (Sigma).

Sorting iPSC-derived Lung Progenitors

On day 15, cells were sorted for downstream applications including RNA analysis and generating organoids. Day 15 cells were washed with CSFDM and incubated in 0.05% trypsin at 37°C for 12 to 18 minutes then gently triturated and inactivated with 10% FBS in DMEM. The resulting cell suspension was centrifuged at 200x*g* for 5 min and re-suspended in FACS Buffer: Hanks Balanced Salt Solution (Life Technologies 14175-079), 2% FBS, 25 mM HEPES (Life Technologies 15630), 2mM EDTA (Sigma E7889-100ml), 100μg/ml Primocin (Invivogen), 10μM y-27632 (Tocris 1254). The suspension was filtered twice through 40μm filters (Falcon 352340). For CD47/CD26 staining 1 x 10^6 cells in 100μL were incubated in CD47-PerCP/Cy5.5 (mouse monoclonal, Biolegend, B191878, 1:200) and CD26-PE (mouse monoclonal, Biolegend, 302706, 1:200) or isotype controls (PE mouse IgG1 isotype, Biolegend, 400113 and PerCP/Cy5-5 mouse IgG1 isotype, Biolegend, 400149) for 30min on ice. Cells were then washed in FACS buffer, centrifuged at 200x*g* for 5 min and re-suspended in FACS Buffer. Live cells were sorted by staining with propidium iodide (PI) (Life Technologies p3566) and excluding PI+ events or in the case of CD47/CD26 staining with calcein blue (ThermoFisher Scientific, C1429). Cells were sorted into FACS buffer. Sorting was performed on a Mo-Flo Legacy in the BU Flow Cytometry Core.

Generating Organoids

To generate unsorted organoids day 15 cells were dissociated with 0.05% trypsin for 2 to 4 min. Trypsin was aspirated and the cells were washed with DMEM+10% FBS, re-suspended as clumps in CSFDM in a 1.5ml Eppendorf tube and centrifuged at 200G for 5min. The Eppendorf was then placed on ice, the supernatant aspirated and the cell pellet resuspended in Matrigel (Corning 356230). 40-50μL of Matrigel was then pipetted into the center of each well of a 12 well tissue culture plate and allowed to gel in the incubator for 15 to 20 min. "CFK" media was then added to each well, supplemented with 10μM Y-27632 media 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 50µL Matrigel and allowed to gel as above.

Fluidigm Single Cell 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 and C1 integrated fluidics circuits (IFCs) were used to capture live cells, lyse, convert polyA+RNA into full length cDNA, amplify cDNA and generate cDNA according to their detailed protocol ("Using C1 to Generate Single-Cell cDNA Libraries for mRNA Sequencing", Fluidigm, PN 100-7168). 69/96 NKX2-1^{GFP+} cells were captured on IFC #1 and 84/96 BU3 cells were captured on IFC #2. Library preparation for sequencing was performed following the modified Illumina Nextera XT DNA library preparation protocol. The concentration of cDNA was determined using Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies). Sequencing was performed by Elim BioPharm (Oakland, CA) on 2 lanes of an Illumina HiSeq Flow cell. In total 570 million 50 bp reads were sequenced for each end.

Bioinformatics and Statistics:
For the single-cell Seq analysis sequenced reads were aligned and mapped using Tophat (v2.1.0) and Bowtie2 (v2.2.6) and Cufflinks (v2.2.1) software (5, 6). 515 million reads were aligned with an average of 2.8 million reads per cell per end. All of the accepted hits in bam files output from cufflinks were processed using Picard tools (http://broadinstitute.github.io/picard/) FixMateInformation and then counts were compiled using HTSeq-count[¶] using the UCSC hg19 human assembly. All of the gene counts for each cell were compiled into a single file. Filtering was then performed to remove any cell that did not have at least a 50% alignment rate. Genes that did not have a least one read aligned in at least 3 cells were removed. The resulting cell/gene matrix file was then normalized using DESeq2[¶¶]. The clustering, PCA and significance testing were performed using SCICAST (details and a walkthrough can be found at https://github.com/iandriver/scicast). Hierarchical clustering linkage method was performed using the Ward variance minimization algorithm[¶¶¶] and the distance was computed using the standardized Euclidean distance[¶¶¶¶]. All correlation coefficients are Pearson's correlation coefficients, 2-tailed p-values. P-values were computed using one way ANOVA and adjusted p-values were calculated using the R package p.adjust(https://stat.ethz.ch/R-manual/R-devel/library/stats/html/p.adjust.html) using the Benjamini & Hochberg (1995) "FDR" method. All raw data gene expression files can be downloaded from the NCBI Gene Expression Omnibus (GEO GSE96106).

[¶] http://wwwhuber.embl.de/users/anders/HTSeq/doc/count.html
[¶¶] https://bioconductor.org/packages/release/bioc/html/DESeq2.html
[¶¶¶] http://docs.scipy.org/doc/scipy-0.16.0/reference/generated/scipy.cluster.hierarchy.linkage.html
[¶¶¶¶] http://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.distance.pdist.html

Monocle Analysis:
Using the software package Monocle (v2.2)* cell data was tabulated as follows: CellOrigin was assigned to the respective cells by cell line (C17 or

BU3). NKX2-1 level was assigned as “High” or “No/low” expression based on a simple cutoff where all cells with an NKX2-1 expression value of 10 or more were assigned NKX2-1 “High” and all others were assigned NKX2-1 “No/low”. The three mitosis categories were assigned based on the unbiased clustering of all cells and all genes (see Figure S6C). The full code and data for recreating the analysis can be found at https://github.com/iandriver/ips17-BU3-single-cell. Cells were filtered on number of mRNAs expressed and genes were filtered on both the sum expression of that gene (>80 across all 145 cells) and the number of cells that expressed a given gene at any level above 0 (>6 cells with non-zero expression). 145 cells and 12837 genes were present in the final Monocle analysis. Ordering genes were selected in an unbiased fashion using scicast k-means clustering and significance testing functions. The PCA plot of cells served as input to the scikit-learn Kmeans function** for two clusters. All of the genes in the two clusters were then compared in using a one-way ANOVA test. The p-values were then adjusted using the Benjamini & Hochberg “FDR” method. Genes were then ranked by FDR-adjusted p-value and the top 200 genes were used as ordering genes for the Monocle package. Details and all of the associated files and code can be found at https://github.com/iandriver/ips17-BU3-single-cell.

*http://cole-trapnell-lab.github.io/monocle-release/articles/v2.2.0/)
**(http://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html

Isolation of Human Fetal Lung Epithelium

Week 10 or 21 human lung tissues were obtained in the Guttentag laboratory under protocols originally reviewed by the Institutional Review Board at the Children’s Hospital of Philadelphia and subsequently reviewed by Vanderbilt University. The cell stocks used in the present studies were donated to the Kotton laboratory for the purpose of providing reference data for this project. “Uncultured naïve lung epithelial cells” were isolated by the overnight culture of lung explants in Waymouth media; a technique that generally yields 86 ± 2% epithelial cells with the remaining cells consisting of fibroblasts with <1% endothelial cells (7). “Differentiated alveolar type 2 (AT2) cells” were prepared in a similar manner except that the lung explants were first cultured for 6 days in Waymouth media supplemented with 10 nM Dexamethasone, 0.1mM 8-Br-cAMP, and 0.1mM 3-isobutyl-1-1methylxanthine (8). The method for isolating the AT2 cells was modified in the following way. After initial digestion using Collagenase and DNase as previously described, cells were filtered, rinsed in PBS, and centrifuged at 1200 RPM for 4 minutes at room temperature. Cells were then digested in 15 mL of PBS supplemented with 2 mL Dispase with 160µl DNase for 30 min at room temperature with magnetic stirring, and the cells were then filtered through a 40µm filter. Adherence to plastic for removal of fibroblasts was conducted as previously described (7, 8). This resulted in a more enriched population of epithelial cells (90-95%) with the remaining cells being exclusively fibroblasts. Assessment of cell purity by immunostaining of plated cells has been described previously (7, 8). For long-term storage in liquid nitrogen cells were frozen in DMEM supplemented with 10% by volume of DMSO, 20% by volume of fetal calf serum, 2 mM glutamine, 1X Penicillin with Streptomycin (final 1 U/ml and 1 µg/ml, respectively). For immunophenotyping of intact fetal lung tissue, week 10 lungs were fixed in

paraformaldehyde and paraffin tissue sections were prepared for NKX2-1 and CD47 immunostainings.

Microarray analysis.
Biological triplicates of all samples except human fetal lung were prepared. Biological duplicates from one embryo ("uncultured naïve lung epithelium") and a singlicate from a different embryo ("differentiated AT2 cells") were prepared for the human fetal lung sample controls. All 27 samples for microarray analysis were lysed and stored in Qiazol (Qiagen). RNA extraction was performed using RNeasy Plus Mini kit (Qiagen). Both Nanodrop and Agilent 2100 Bioanalyzer determined RNA concentration and quality. All samples had an RNA Integrity Number (RIN) score of >7.0. Affymetrix GeneChip Human Gene 2.0 ST arrays were used for gene expression profiling. Technical quality of the arrays was assessed by two quality metrics: Relative Log Expression and Normalized Unscaled Standard Error. Analyzing X and Y-lined genes established adequate dynamic range of gene expression across samples. Principal Component Analysis (PCA) was performed using the prcomp R function with expression values that had been normalized across all samples to a mean of zero and a standard deviation of one. Differential gene expression with respect to experimental group across all samples was assessed by performing a one-way ANOVA computed using the f.pvalue function in the sva package (version 3.4.0), and the significance of each pairwise comparison between groups (corrected for multiple hypothesis testing) was obtained using Tukey's Honest Significant Difference post-hoc test. Correction for multiple hypothesis testing was accomplished using the Benjamini-Hochberg false discovery rate (FDR). All statistical analyses were performed using the R environment for statistical computing (version 2.15.1). Figure 4-F were generated using GENE-E (https://software.broadinstitute.org/GENE-E/index.html). The heatmap scale was determined by the row minimum and maximum log2 expression values in each row to convert values to colors. Heatmaps Sparkline plots (Figure S4C-D) were generated using Morpheus (https://software.broadinstitute.org/morpheus/). For the comparison of candidate lung markers from our experiments to published human fetal lung samples (Figure S6D) we normalized data matrix files from the NCBI GEO repository (http://www.ncbi.nlm.nih.gov/geo/, GSE14334) (9). We generated a heatmap of the log-transformed data for selected markers. For samples with biological replicated or triplicates we calculated the mean expression.

Recombination with Mouse Embryonic Lung Mesenchyme
Recombinations were performed essentially as previously described (10). Briefly, small GFP-positive or GFP-negative fragments of day 21 hiPSC organoids were recombined with 10-12 pieces of LgM on the surface of a semisolid medium consisting of 0.5% agarose (Sigma) and 10% FBS in DMEM. The LgM rudiments were teased into close apposition to the endoderm with microsurgery knives (Fine Science Tools, Inc) and excess liquid medium was removed with a flame-drawn Pasteur pipet. After overnight culture to promote tissue adherence, the recombinants were transferred to the surface of a 8 μm pore size Whatman nucleopore filter and cultured for 5-7 days in BGJb medium containing 20% FBS, 0.2 mg/ml vitamin C (Sigma), and

5 µg/ml recombinant mouse amino-terminal SHH (R&D Systems) to promote mesenchyme viability (11). The recombinants were maintained for 7 days, with medium changes every other day. Dexamethasone (50nM) was added to the medium for the final 48 hours to promote lung epithelial differentiation.

Reverse Transcription Quantitative Polymerase Chain Reaction (RT-qPCR)
RNA was extracted by first lysing cells in Qiazol (Qiagen) and subsequently using RNeasy Plus Minikit (Qiagen) according to the manufacturer's protocol. TaqMan reverse transcription reagents (Applied Biosystems) were used to reverse transcribe RNA to cDNA. 2.5µl of cDNA was added to a final volume of 25µl of PCR reaction using Taqman Fast Universal PCR Master Mix (Applied Biosystems) in technical triplicate wells of a 96 well PCR plate and analyzed on a StepOne Real-Time PCR machine (Applied Biosystems) for 40 cycles. In some experiments, 1µl of cDNA was used in a 20µl PCR reaction. Relative gene expression, normalized to 18S control, was calculated as fold change in 18S-normalized gene expression, over baseline, using the $2^{(-\Delta\Delta CT)}$ method. Baseline, defined as fold change = 1, was set to undifferentiated pluripotent stem cell levels, or if undetected, a cycle number of 40 was assigned to allow fold change calculations.

Gene	Vendor	Probe ID
NFIB	ThermoFischer Scientific	Hs01029175_m1
CYTL1	ThermoFischer Scientific	Hs01573280_m1
MUC1	ThermoFischer Scientific	Hs00159357_m1
SPOCK3	ThermoFischer Scientific	Hs00213568_m1
CRH	ThermoFischer Scientific	Hs01921237_s1
BMP3	ThermoFischer Scientific	Hs00609638_m1
FOXP2	ThermoFischer Scientific	Hs0036_m12818
LAMA3	ThermoFischer Scientific	Hs00165042_m1
GRHL2	ThermoFischer Scientific	Hs00227745_m1
SEMA3C	ThermoFischer Scientific	Hs00989373_m1
CD47	ThermoFischer Scientific	Hs00179953_m1
LMO7	ThermoFischer Scientific	Hs00245600_m1
ELF3	ThermoFischer Scientific	Hs00963881_m1
NKX2-1	ThermoFischer Scientific	Hs00968940_m1
SFTPC	ThermoFischer Scientific	Hs00161628_m1
SFTPB	ThermoFischer Scientific	Hs01090667_m1
TP63	ThermoFischer Scientific	Hs00978343_m1
PITX1	ThermoFischer Scientific	Hs00267528_m1

Surface Marker Screen
iPSC C17 was differentiated in the lung differentiation protocol and dissociated into a single cell suspension on day 15 of differentiation, as previously described. The BD Lyoplate Human Cell Surface Marker Screening Panel (BD Biosciences 560747) was used to screen, by flow cytometry, for the expression of 242 cell surface markers on day 15 cells according the manufacturer's instruction. The panel contains primary monoclonal mouse and rat antibodies against the 242 surface markers, appropriate isotype controls and secondary AlexaFluor 647 conjugated goat anti-mouse IgG and goat anti-rat IgG antibodies. Samples were analyzed on an BD LSR II flow

cytometer. Surface marker expression (AlexaFluor 647) was plotted against NKX2-1^{GFP+} to identify surface markers differentially expressed in either the NKX2-1^{GFP+} or NKX2-1^{GFP-} population.

Immunostaining

Day 15 cultures (Figure 1D) and day 36 organoids (Figure 2A,D) were fixed by incubating in 4% paraformaldehyde at room temperature for 10 minutes and 30 minutes respectively. Organoids were combined in low melting agarose (SeaPrep) to form a pellet and then paraffin embedded. For immunocytochemistry sections were rehydrated and antigen retrieval was performed at 95°C for 20 minutes in Dako Target Retrieval Solution (DAKO, S-1699). For fixed cells blocking was performed with 2.5% normal donkey serum (NDS) and 0.25% Triton X-100 (Sigma, T-8787) for 30 minutes followed by 2.5% NDS for 20 minutes. Paraffin sections were blocked with 4% NDS for 30 min. Samples were incubated in primary antibody in 4-5% NDS overnight at 4°C. The staining was detected with secondary antibodies purchased from Jackson Immunoresearch (donkey anti-mouse, donkey anti-rabbit, donkey anti-chicken at 1:300 to 1:500 dilution for 2 hours at room temperature). Nuclear counterstaining was performed with DAPI (Invitrogen, 1:10,000) or SlowFade Gold Antifade reagent containing DAPI (Life Technologies, S36938). Antibodies include NKX2-1 (rabbit monoclonal, Abcam, Ab76013, 1:250), NKX2-1 (mouse monoclonal, Abcam, Ab72876, 1:100), EPCAM (mouse monoclonal, Abcam, GR224588-1, 1:250), GFP (polyclonal chicken IgY, AVES, GFP 1020, 1:10,000), Pro-SPC (polyclonal rabbit, Seven Hills, WRAB-9337,1:100), SFTPC (polyclonal rabbit, Seven Hills, R460, 1:2000), LPCAT1 (polyclonal rabbit, Proteintech, 16112-1-AP, 1:2000) and PAX8 (polyclonal rabbit, Abcam, Ab122944 1:50-1:100), CD47 (mouse monoclonal Abcam, Ab3283, 1:100), CD47 (mouse monoclonal FITC conjugated, Biolegend, 323106, 1:100) and SOX9 (rabbit monoclonal, Ab185230, 1:500).

In Situ hybridization and Southern Blot

A full-length cDNA encoding human SFTPC was isolated by RT-PCR and cloned into vector pcDNA3, which was then used to transcribe digoxigenin-labeled antisense riboprobe. Pax8 cDNA (GE Dharmacon, Cat#MMM1013-202765959) was also cloned into vector pcDNA3. Whole mount in situ hybridization on tissue recombinants was performed according to the protocol described by Wilkinson (12). Southern blot using digoxigenin (DIG)-labelled hybridization probes: PCR DIG Probe Synthesis Kit (Roche), Southern blot probe primers (5'-GACTCTAAGGGTCCGAGCAG-3' and 5'-GAGACCGGTAAGCGACAAAC-3') and 10pg of NKX2-1 donor DNA were used to generate incorporated DIG-dUTP hybridization probes, by PCR labeling and following the manufacturer's instructions. Primer annealing temperature, primer and template concentrations were optimized prior to DIG-dUTP incorporation. 3.75 µg of genomic DNA samples from various unmodified, NKX2-1 targeted and Cre-excised clones were separated on a 0.7% agarose gel, transferred onto a positively charged nylon membrane (Roche) and UV crosslinked. The Southern blot was first prehybridized for 40 min at 42 degree Celsius using DIG Easy Hyb Granules (Roche) and afterwards hybridized with DIG-labeled NKX2-1 DNA probe according the

manufacturer's instructions. After the hybridization and wash procedure the NKX2-1 probe on the Southern blot was detected by chemiluminescent alkaline phosphatase substrate using ready-to-use CDP Star (Roche).

Time-lapse microscopy

To generate time-lapse movies of differentiation in culture, NKX2-1^{GFP} human C17 iPSCs (between days 19 and 21) were imaged in an 6-well plate at 37°C and 5% CO2, conditions maintained with a Controlled Environment Microscope Incubator (Nikon Instruments, Inc.) designed for live-cell imaging. Throughout, bright field images were collected every 30 minutes and fluorescent images were collected every 2 hours at 10x magnification (Plan Fluor 10x Ph1 Dll) using an Eclipse Ti-E inverted microscope (Nikon Instruments, Inc.) equipped with the "Perfect Focus System", a motorized stage, and a Clara-Echarge-coupled device (CCD) camera (Andor Technology). Each image in the time-lapse was constructed by stitching together two by two imaging fields (four total XY positions, 15% overlap) using the supplier's image acquisition software (NIS-Elements Advanced Research). Images were acquired in phase contrast configuration and in fluorescent (GFP and mCherry) channels. Filters and light sources (Nikon LED and Lumencor SPECTRA X Light Engine) were automatically controlled by the supplier's software (NIS-Elements Advanced Research).

References:

1. 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.

2. Kurmann AA et al. Regeneration of Thyroid Function by Transplantation of Differentiated Pluripotent Stem Cells. *Cell Stem Cell* 2015;17(5):527–542.

3. Longmire TA et al. Efficient derivation of purified lung and thyroid progenitors from embryonic stem cells. *Cell Stem Cell* 2012;10(4):398–411.

4. Huang SXL et al. Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. *Nat. Biotechnol.* 2014;32(1):84–91.

5. Kim D et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14(4):R36.

6. Trapnell C et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7(3):562–578.

7. 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.

8. 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):L940–51.

9. Kho AT et al. Transcriptomic analysis of human lung development. *Am. J. Respir. Crit. Care Med.* 2010;181(1):54–63.

10. 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.

11. Weaver M, Batts L, Hogan BLM. Tissue interactions pattern the mesenchyme of the embryonic mouse lung. *Dev Biol* 2003;258(1):169–184.

12. Wilkinson DG. In situ Hybridization. A Practical Approach. Edited by D.G. Wilkinson. IRL Press, Oxford. 1992. 163 pages. £18.50 Softback. ISBN 0 19 963327 4. In: Wilkinson DG ed. *Whole mount in situ hybridization of vertebrate embryo*. Oxford University Press; 1992:75–83