Material and Methods (extended)

Nkx2-1^{mCherry} mouse embryonic stem cell (ESC) line

The Nkx2-1^{mCherry} ESC line was generated as previously described (Bilodeau et al., 2014) by targeting an Internal Ribosomal Entry Site (IRES) coupled to the mCherry reporter cDNA into the Nkx2-1 locus 3'UTR in R1 mESCs. It was maintained in the undifferentiated state on mouse embryonic fibroblast feeder layers using serum containing media supplemented with LIF (ESGRO Chemicon ESG1106).

Mouse ESC directed differentiation

Definitive endoderm induction was performed in complete serum-free differentiation medium (cSFDM) consisting of 75% IMDM (Invitrogen 12440) and 25% Ham's Modified F12 medium (Cellgro 10-080-CV) supplemented with N2 and B27+RA (Invitrogen 17502-048 and 17504-44), 0.05% BSA (Invitrogen 15260-037), 2 mM L-glutamine (Invitrogen 25030-081), 0.05 mg/ml ascorbic acid (Sigma A4544) and $4.5x10^{-4}$M monothioglycerol (MTG, Sigma M6145) as previously described (Gouon-Evans et al., 2006; Longmire et al., 2012), summarized as follows: undifferentiated ESCs were trypsinized to form a single cell suspension and plated onto P100 Petri dishes (500,000 cells/ dish) resulting in the formation of embryoid bodies (EBs) over 60 hours. EBs were dissociated by trypsinization (0.05%, 1 min, 37°C) and plated as single cells into cSFDM supplemented with 50 ng/ml Activin A (R&D systems 338-AC) for an additional 60 hours. For anteriorization of endoderm, on day 5 (120 total hours of differentiation) EBs were plated onto P100 Petri dishes in Nog/SB media: cSFDM supplemented with 100 ng/ml mNoggin (R&D 1967-NG) and 10 µM SB431542 (Togris, Bristol, United Kingdom) as previously described (Longmire et al., 2012). For Nkx2-1+ endoderm induction, EBs were plated on gelatin coated 6 well plates at the equivalent density of 200,000 cells/well, or cells in single cell suspension obtained by trypsinization at 100,000 cells/well, in specification media: cSFDM supplemented with the factors as stated in the text at the following concentrations: 100 ng/ml mWnt3a (R&D 1324-WN), 10 ng/ml mBMP4 (R&D 5020-BP), 250 ng/ml mFGF2 (R&D 3139-FB) and 100 ng/ml Heparin Sodium Salt (Sigma H4784). When single cells were plated on day 6, 10µM Y-27632

Development • Supplementary information

(Tocris) were added to the media for the first 24 hours to enhance survival. Nkx2-1^{mCherry}+ cells or the populations otherwise indicated in the text were sorted using MoFlo or FACSARIA II SORP high speed cell sorters and replated onto gelatin-coated 24-well plates on day 12-14 at a density of $5x10^4$ cells/well.

Further differentiation and maturation of Nkx2-1+ sorted cells was performed in either 2D or 3D culture conditions, as indicated in the text. For 2D differentiation, cells sorted on day 14 were replated and grown for 8 more days in cSFDM supplemented with 250 ng/ml mFGF2, 100 ng/ml hFGF10 (R&D 345-FG), 100 ng/ml Heparin Sodium Salt and, where indicated in the text, Wnt3a at 200 ng/ml. On day 22, if stated in the text, the media was changed to DCI+K: Ham's F12 media, 15 mM HEPES (pH 7.4), 0.8 mM $CaCl_2$ (Sigma C1016), 0.25% BSA, ITS supplement (5 μg/ml insulin, 5 μg/ml transferrin, 5 ng/ml Sodium Selenite; BD 354352), 50 nM dexamethasone (Sigma D4902), 0.1 mM 8-Bromoadenosine 3',5'-cyclic monophosphate (8-Br-cAMP) sodium salt (Sigma B7880), 0.1 mM 3-Isobutyl-1-methylxanthine (IBMX) (Sigma I5879) and 10 ng/ml mKGF (R&D 5028-KG). For 3D culture, ESC-derived cells were sorted on day 14 of differentiation and purified Nkx2-1+ or the corresponding Nkx2-1- control cells were replated in pure growth factor reduced Matrigel drops (Corning 356230) at a density of $5x10^4$ cells in 80 μl Matrigel per well in a 24-well plate. From day 14 to the time indicated in the text, cells were grown in cSFDM supplemented with 250 ng/ml mFGF2, 100 ng/ml hFGF10 and 100 ng/ml Heparin Sodium Salt.

Human ESC and iPSC maintenance

The RUES2 human embryonic stem cell line was a kind gift from Dr. Ali H. Brivanlou of The Rockefeller University, New York, NY. The human iPSC line "C17" (aka "iPS17") was generated from a patient with cystic fibrosis as recently described (Crane et al., 2015). Human iPSCs and ESCs 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 354277) in mTeSR1 (Stem Cell Technologies) and passaged with Gentle Cell Dissociation Reagent (Stem Cell Technologies). For generating the NKX2-

Development • Supplementary information

1^{GFP} reporter iPS17 line, Transcription Activator Like Effector Nucleases (TALEN) technology was used to target a GFP cassette to the human NKX2-1 locus as published previously (Hawkins et al., 2017). All cell lines and donor targeting vectors are available from the corresponding authors and can be viewed and requested through www.kottonlab.com.

Human ESC and iPSC directed differentiation to lung epithelium

Human lung directed differentiation protocol was optimized by adapting published protocols (Longmire et al. 2012; Huang et al. 2014). We first induced definitive endoderm using STEMDiff definitive endoderm kit (STEMCELL Technologies) according to the manufacturer's protocol. After approximately 72 to 84 hours we harvested and analyzed cells by flow cytometry for efficiency of definitive endoderm induction by the co-expression of the surface markers C-kit [APC conjugated mouse monoclonal antibody (104D2), Life Technologies CD11705] and CXCR4 [PE-conjugated mouse monoclonal antibody (12G5), Life Technologies MHCXCR404] with appropriate APC (Life Technologies MG105) and PE (Life Technologies MG2A04) 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 approximately 150-200,000 cells/cm^2 on Matrigel-coated plates in complete serum-free differentiation media (cSFDM) supplemented with 2 μM Dorsomorphin (Stemgent, Lexington, MA) and 10μM SB431542 for 72 hours. 10 μM Y-27632 was added for the first 24 hours. To specify lung epithelium, differentiation media was changed on day 6 to specification media adding the following factors to cSFDM in the combinations stated in the text: 3 μM CHIR99021 (Tocris), 10 ng/ml rhFGF10, 10 ng/ml rhKGF, 10 ng/ml rhBMP4 (all from R&D Systems), 50-100 nM Retinoic acid (Sigma) (Huang et al. 2014). For experiments testing canonical Wnt signaling inhibition in Fig. 7, chemical inhibitors of beta-catenin interaction with p300 (IQ1, Tocris, 4713) or with CBP (ICG001/TCF Inhibitor VI, Calbiochem 504712) were used at a final concentration of 3 μM from day 6 to day 15.

Development • Supplementary information

Day 14-15 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 3 μM CHIR99021, 10 ng/ml rhFGF10, 10 ng/ml rhKGF, 50 nM dexamethasone, 0.1 mM 8-Br-cAMP sodium salt and 0.1 mM IBMX, named "CFK+DCI media", and plated on freshly-coated Matrigel plates. 10 μM Y-27632 was added to "CFK+DCI" media for the first 24 hours.

To generate 3D organoids, day 14 or 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 tube and centrifuged at 200×*g* for 5 min. The tube was then placed on ice, the supernatant aspirated and the cell pellet resuspended in Matrigel drops. 40-50 μl 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+DCI" media was then added to each well, supplemented with 10 μM Y-27632 media for the first 24 hours.

Human iPSC directed differentiation to thyroid epithelium

For differentiation, human iPSCs were passaged using Gentle Cell Dissociation Reagent and plated on 2D Matrigel-coated 6-well plates as single cells ($1x10^6$/well) in mTeSR1 with 10 μM Y-27632 for 24 hours. Definitive endoderm was induced using the StemDiff definitive endoderm kit for 48 hours. For endoderm anteriorization, definitive endoderm was re-plated in Matrigel drops as small clumps after dissociation using Gentle Cell Dissociation Reagent. Anteriorization was induced by maintaining cells in cSFDM containing 2 μM Dorsomorphin, 10 μM SB431542, and 10 μM Y-27632 for 48 hours. Thyroid lineage specification was induced by treating the cells with cSFDM supplemented with 250 ng/ml rhFGF2, 100 ng/ml rh BMP4, 100 ng/ml Heparin salt, and 10 μM Y-27632 for 12 days. For extended differentiation and maturation, cells were passaged and re-plated onto fresh Matrigel drops on day 16 and treated with thyroid maturation media [cSFDM supplemented with 100 ng/ml rhFGF2, 100 ng/ml rhFGF10, 100 ng/ml Heparin salt, 10 μM Y-27632, and 1mU/ml bovine TSH (Los Angeles

Biomedical Research Institute National Hormone & Peptide Program, AFP-8755B)]. On day 36, matured cells were purified by flow cytometry based on NKX2-1^{GFP} reporter gene expression as previously described (Hawkins et al., 2017) and RNA extracts from the sorted cells were analyzed by RT-qPCR.

Lentiviral production and mESC infection

A modified version of our lentiviral vector, previously published for SftpcdsRed reporter expression (Longmire et al., 2012) was cloned and deployed for transduction of mESC, following methods similar to our prior publication. Briefly, a 3.7kb human SPC promoter fragment (generous gift of Dr. Jeffrey A. Whitsett, University of Cincinnati) (Glasser et al., 1991) was cloned into the promoter position of the pHAGE lentiviral CMV-GFP-W plasmid replacing the CMV promoter (Wilson et al., 2010; Longmire et al. 2012; map and sequence available at www.kottonlab.com). VSV-G pseudotyped SftpcGFP lentiviral particles were packaged and concentrated as previously published (Wilson et al., 2010). Differentiating mESCs were infected on day 16 overnight in the presence of polybrene at 5 µg/ml and subsequent GFP gene expression was monitored by fluorescence microscopy and flow cytometry as indicated in the text, or sorted to purity for RNA isolation. A parallel cell sample was infected with the CMV-GFP-W lentivirus at the same MOI to confirm efficiency of transduction.

Immunofluorescence staining of paraffin sections

For immunofluorescent staining, sections were incubated with the following primary antibodies: anti-Nkx2-1 (Abcam ab76013; 1:100-200); anti-Tg (Abcam ab80783; 1:100); anti-EpCAM (anti-human CD326, Biolegend 324203,1:100), anti-Ecad phospho S838+S840 (Abcam ab76319, 1:100), anti-Scgb1a1 (Santa Cruz sc-9773, 1:300), anti-alpha tubulin tyrosinated (Millipore MAB1864, 1:100), anti-pro-Sftpc (Seven Hills WRAB-9337; 1:1000). Staining was detected with the corresponding Alexa Fluorophore-conjugated secondary antibodies (donkey anti-mouse; donkey anti-rabbit; donkey anti-rat; or donkey anti-goat 1:200, Molecular Probes, Invitrogen).

Intracellular staining of human NKX2-1 protein for flow cytometry

Cells were monodispersed with trypsin and fixed in suspension in 1.6% PFA at 37°C, washed twice in FACS buffer (1x PBS, 0.5% BSA and 0.02% NaN_3), and permeabilized with Permeabilization Wash Buffer (Biolegend 421002). Primary antibody rabbit anti-human Nkx2-1 (Abcam ab76013) was used at a 1:100 dilution in the same buffer for 30 min at RT. After washing twice, Alexa 647 conjugated goat anti-rabbit antibody (Invitrogen A21245) was used as secondary at 1:100 dilution, washed twice and resuspended in FACS buffer for flow cytometry analysis using a BD FACSCalibur (BD Bioscience).

Mouse foregut explants

Whole foreguts were dissected from mouse embryos (6-8 somite pairs) in Hank's balanced salt solution (HBSS) then explanted onto 8 µm pore size Whatman Nucleopore Track-Etch Membranes (Millipore). Explants were cultured for 2-3 days in a base medium [BGJb medium (Gibco) + 10% fetal bovine serum (FBS, Sigma) and 0.2 mg/ml ascorbic acid] containing either the BMPR inhibitor DMH-1 (1.5 µM, Tocris) or DMSO as a vehicle control. Whole mount immunostaining was performed using a modification of the method of Ahnfelt-Ronne et al. (Ahnfelt-Ronne et al., 2007). The primary antibodies used were: guinea pig anti-Nkx2-1 (Seven Hill Bioreagents; 1:500) and rat anti-Ecad (R&D; 1:2000). After staining, samples were cleared with Murray's clear (2:1 benzyl benzoate: benzyl alcohol) and imaged using a Nikon A1Rsi inverted laser confocal microscope. Imaris software was used to analyze the images.

Mouse whole embryo cultures

For mouse whole embryo cultures, E7.5 embryos were cultured in a 1:1 mixture of Ham's F12 medium and whole embryo culture rat serum (Harlan Labs) containing N-2 Supplement (Invitrogen). Vessels were placed on a roller culture apparatus (BTC Engineering, Cambridge, UK) and maintained for 2 days at 37°C and gassed with 20% O_2 and 5% CO_2. FGF signaling was inhibited by treatment with a 10 µM concentration of either PD173074 (Cayman Chemical), PD161570 (Tocris Bioscience), or BGJ398 (ApexBio), with DMSO serving as a vehicle control.

In vitro Endoderm recombination

Recombinations were performed essentially as previously described (Shannon et al. 1998). Briefly, cultured anterior endoderm rudiments were first recovered from Matrigel using Cell Recovery Solution (Corning), then recombined with 10-12 pieces of E12.5 LgM on the surface of a semisolid medium consisting of 0.5% agarose (Sigma) and 20% 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 mm pore size Whatman nucleopore filter and cultured in BGJb medium containing 20% FBS, 0.2 mg/ml vitamin C (Sigma) and 5 mg/ml recombinant mouse amino-terminal SHH (R&D Systems) to promote mesenchyme viability (Weaver et al., 2003). The recombinants were maintained for 6 days, with medium changes every other day.

Real Time Quantitative Polymerase Chain Reaction (RT-qPCR)

RNA was extracted from cells using the RNeasy kit (Qiagen 74104) following the manufacturer's protocol. 150 ng of RNA was reverse transcribed to cDNA using TaqMan reverse transcription reagents (Applied Biosystems N8080234) and the manufacturer's protocol in a final volume of 40 μl. cDNA template was diluted 1:4, and 1 μl of the diluted template was used per 13 μl real time PCR reaction using Taqman Fast Universal PCR Master Mix (Applied Biosystems 4352042) and ran in technical triplicates up to 40 cycles of PCR on the Applied Biosystems StepOne machine or the Applied Biosystems 7900HT Real Time PCR System. 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 ESC or iPSC levels, or if undetected, a cycle number of 40 was assigned to allow fold change calculations.

TaqMan gene expression assays for all mRNAs were purchased from Applied Biosystems:

Gene	TaqMan Probe Number
human 18S rRNA	4313413e
human 18s rRNA	Hs03928985_g1
Abca3	Mm00550501_m1
Aqp5	Mm00437579_m1
Axin2	Mm00443610_m1
Cd44	Mm01277163_m1
Cdh1	Mm01247357_m1
Col1a1	Mm00801666_g1
Epcam	Mm00493214_m1
Foxa2	Mm00839704_mH
Foxe1	Mm00845374_s1
FoxJ1	Mm00807215_m1
FoxP2	Mm00475030_m1
Hhex	Mm00433954_m1
Lamp3	Mm00616604_m1
Nkx2-1	Mm00447558_m1
p63	Mm00495788_m1
Pax8	Mm00440623_m1
Prlr	Mm04336676_m1
Scgb1a1	Mm00442046_m1
Scgb3a2	Mm00488144_m1
Sftpa	Mm00499170_m1
Sftpb	Mm00455681_m1
Sftpc human	Hs00161628_m1
Sftpc mouse	Mm00588144_m1
Sftpd	Mm00486060_m1
Shh	Mm00436528_m1
Slc5a5 (Nis)	Mm00475074_m1
Snai1	Mm00441533_g1
Tg	Mm00447525_m1
Tpo	Mm00456355_m1
Tsh-r	Mm00442027_m1
Twist1	Mm04208233_g1

Microarray analysis

Three independent biological replicates from each of the 2 differentiation conditions (either thyroid specification with BMP4+FGF2 or lung specification with BMP4+Wnt3a) were prepared, and cells on day 14 of differentiation were sorted: Nkx2-1$^{mCherry+}$ vs. Nkx2-1$^{mCherry-}$. Total RNA was extracted using miRNeasy Mini Kit (Qiagen). Quality-

Development • Supplementary information

assessed RNA samples were hybridized to Affymetrix GeneChip® Mouse Gene 2.0 ST arrays based on Affymetrix standard microarray operating procedure.

Mouse Gene 2.0 ST CEL files were normalized to produce gene-level expression values using the implementation of the Robust Multiarray Average (RMA) in the *affy* package (version 1.36.1) included in the Bioconductor software suite (version 2.12) and an Entrez Gene-specific probeset mapping (17.0.0) from the Molecular and Behavioral Neuroscience Institute (Brainarray) at the University of Michigan. Array quality was assessed by computing Relative Log Expression (RLE) and Normalized Unscaled Standard Error (NUSE) using the *affyPLM* package (version 1.34.0). 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 expression was assessed using the moderated (empirical Bayesian) t-test implemented in the *limma* package (version 3.14.4) (i.e., creating simple linear models with *lmFit*, followed by empirical Bayesian adjustment with *eBayes*). Correction for multiple hypothesis testing was accomplished using the Benjamini-Hochberg false discovery rate (FDR). Human homologs of mouse genes were identified using HomoloGene (version 68). All microarray analyses were performed using the R environment for statistical computing (version 2.15.1).

Normalization and quality assessment

The arrays were normalized together using the Robust Multiarray Average (RMA) algorithm and a CDF (Chip Definition File) that maps the probes on the array to unique Entrez Gene identifiers. The result is a matrix in which each row corresponds to an Entrez Gene ID and each column corresponds to a sample. The expression values were log2-transformed by default. Normalization was **not** performed using default Affymetrix probesets, as they do not all correspond 1:1 with individual genes or transcripts, and downstream analysis with such probeset identifiers usually requires an extra step to translate these probesets into commonly used identifier (Entrez Gene, Ensembl Gene, RefSeq, etc.) The technical quality of the arrays was first assessed by two relative quality metrics: Relative Log Expression (RLE) and Normalized Unscaled

Development • Supplementary information

Standard Error (NUSE). RLE is a measure of the relative quality of each sample (i.e., how much the signal was artificially boosted during normalization) compared to the other arrays in the batch. NUSE is a measure of the relative agreement between the probes for each gene, for each sample compared to the other arrays in the batch. For each sample, median RLE values > 0.1 or NUSE values > 1.05 are considered out of the usual limits. All arrays had median RLE and NUSE values well within these limits, indicating that all samples were of similar quality.

Differential expression analysis

t-tests *on a two-factor linear model of expression as a function of specification media and Nkx2-1 expression*

To identify genes whose expression changes coordinately with respect to specification media, Nkx2-1 expression, or the interaction of the two, a linear modeling approach was used, followed by t-tests. The main effects of specification media used and Nkx2-1 expression were assessed using a linear model of the form:

expression ~ specification media + Nkx2-1

and the interaction effect of specification media used and Nkx2-1 expression assessed with a linear model of the form:

expression ~ specification media + Nkx2-1 + specification media:Nkx2-1

In both models, '~' means 'is a function of' and ':' indicates an interaction between the two variables.

The *specification media* effect measures whether the expression of a given gene changes with respect to the specification media used, regardless of Nkx2-1 expression, and the *Nkx2-1* effect measures whether a given gene changes with respect to Nkx2-1 expression, regardless of the specification media. The *specification media:Nkx2-1* interaction effect measures whether a given gene changes more strongly with respect to specification media within only one Nkx2-1 expression group, or more strongly with respect to Nkx2-1 expression in only one specification media group.

For each effect (*specification media*, *Nkx2-1 expression*, and *specification media:Nkx2-*

Development • Supplementary information

1 expression), t-tests were performed on the coefficient of the linear model to obtain a *t* statistic and *p* value for each gene. A "moderated" t-test was used, which is a Bayesian analysis that does not test each gene independently, but rather, leverages information from all of the genes on the array to increase statistical power over Student's t-test.

Benjamini-Hochberg False Discovery Rate (FDR) correction was applied to obtain FDR-corrected *p* values (*q* values), which represent the probability that a given result is a false positive based on the distribution of all *p* values on the array. In addition, the FDR *q* value was also recomputed after removing genes that were not expressed above the array-wise median value of at least one array.

Using an FDR-adjusted p value cutoff <0.05, 7512 genes were differentially expressed between lung vs. thyroid specification media conditions, 3631 were differentially expressed between Nkx2-1+ and Nkx2-1- cell populations, and 4009 were differentially expressed by the interaction of both factors. As expected, when ranked by FDR-adjusted p value, Nkx2-1 was one of the top genes most significantly upregulated between Cherry- and Cherry+ cell populations, both in the lung (second hit) as well as in the thyroid media (third hit).

Unsupervised hierarchical clustering analysis across all samples based on interaction effect (*Specification media:Nkx2-1 effect)* identified 1315 genes (FDR<0.25 and fold change >2 between any group) in 9 clusters (Fig. 4 and supplemental Excel file). Observation of the heatmaps revealed clear differential patterns of gene expression including genes upregulated in lung Nkx2-1^{mCherry}- (cluster 1, 134 genes) or Nkx2-1^{mCherry}+ (cluster2, 180 genes), in thyroid Nkx2-1^{mCherry}- (cluster 3, 218 genes) or Nkx2-1^{mCherry}+ (cluster 5, 188 genes), or genes upregulated in both lung and thyroid Nkx2-1^{mCherry}+ cells (cluster 6, 214 genes).

<u>Single cell RNA-Sequencing</u>

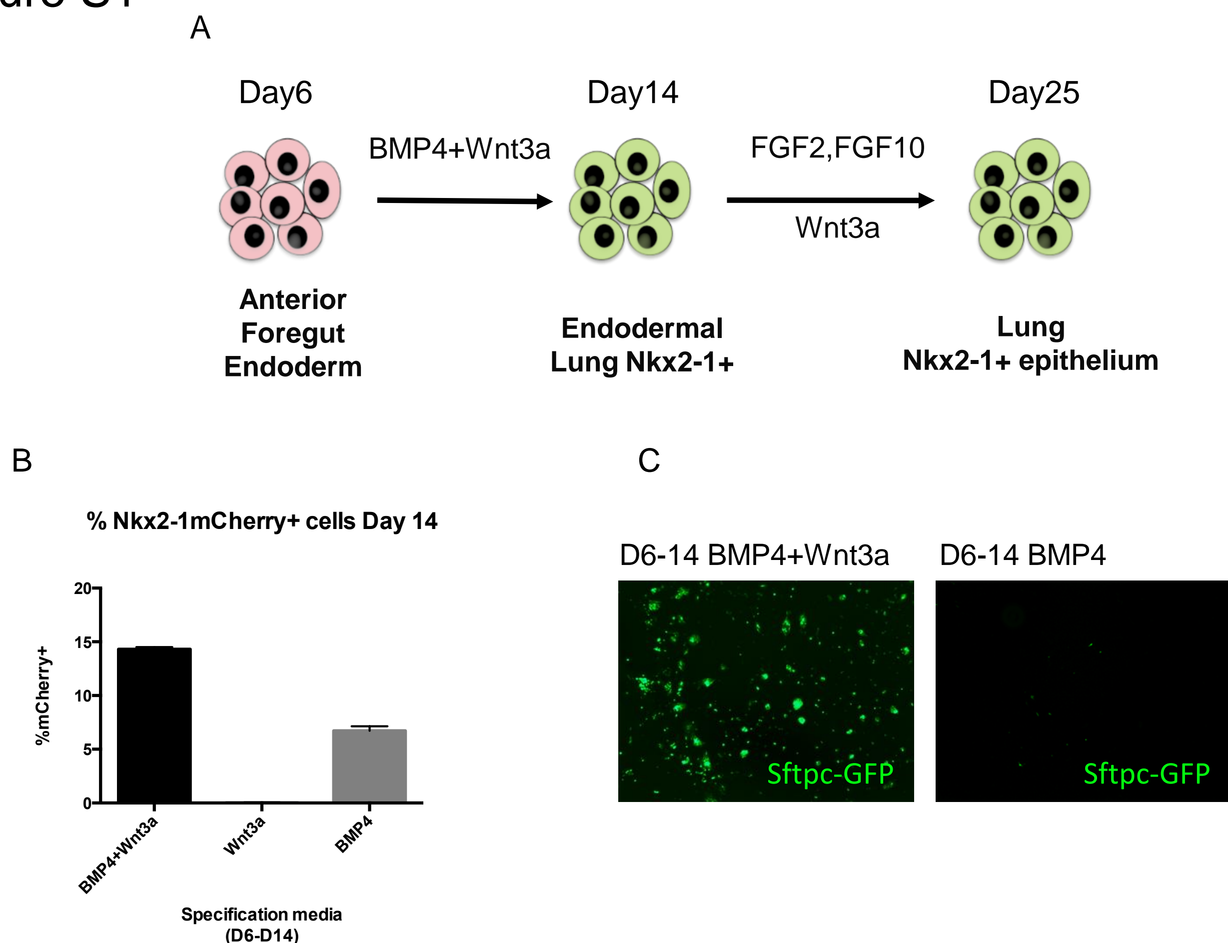
Counts were normalized using pool-based scaling factors and deconvolution (Lun et al., 2016) via the Scater Bioconductor package (McCarthy et al., 2017). A zero-inflated

negative binomial model (Risso et al., 2017 preprint) was applied to the normalized data, which resulted in a low-dimensional representation. This output was clustered using k-means with K set to 3. After the initial clusters were defined, an ANOVA was used to filter out genes with FDR-corrected p-value > 0.05 and variance < 3 and calculate pairwise fold-change coefficients across the clusters.

The heatmap was generated by applying Ward's hierarchical agglomerative clustering method (Murtagh and Legendre, 2014) on the row-scaled expression values of the top differentially expressed genes (absolute log2 fold-change > 1.5 and FDR-adjusted p-value < 0.05) for the pairwise contrasts across the three clusters.

Development • Supplementary information

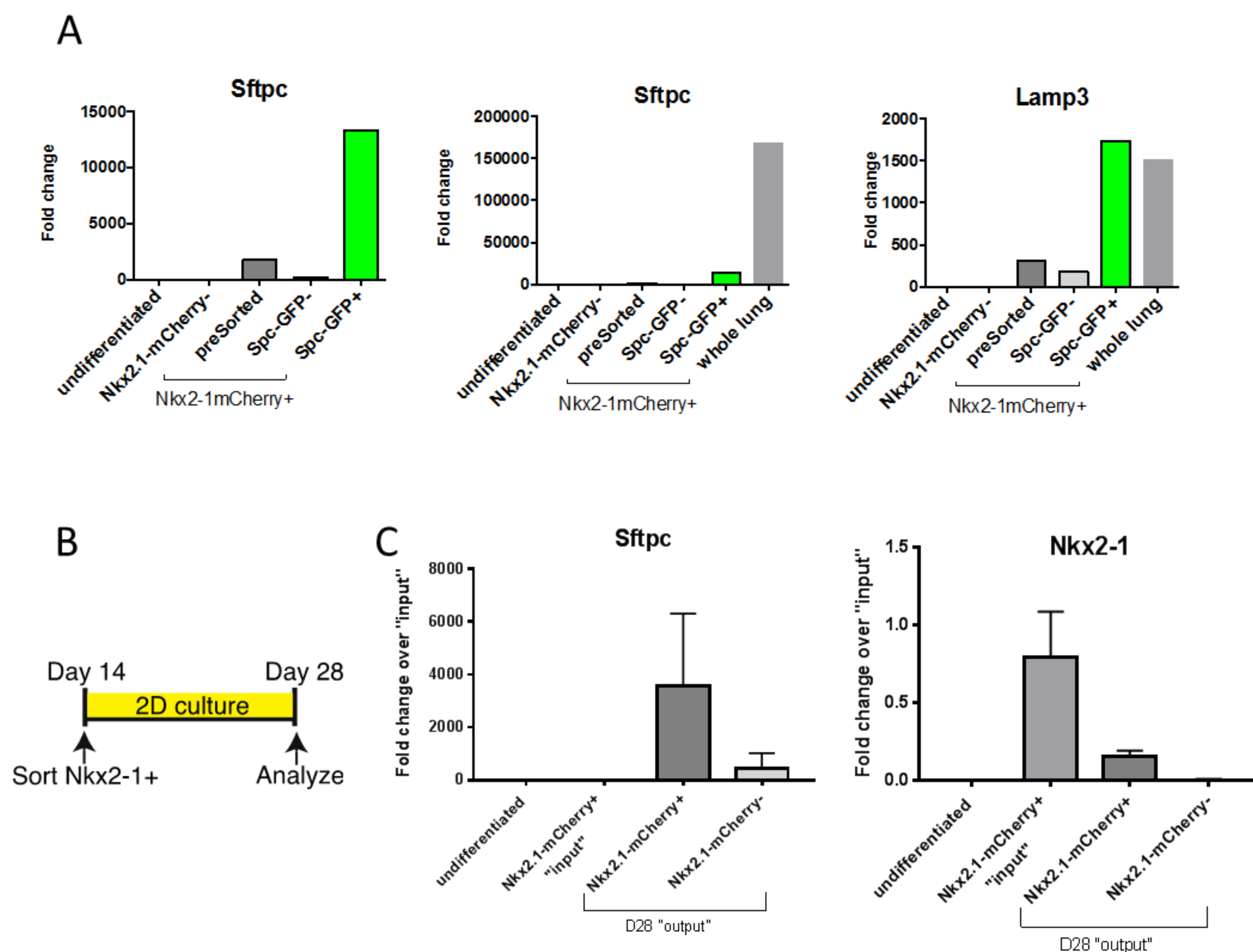
Figure S1



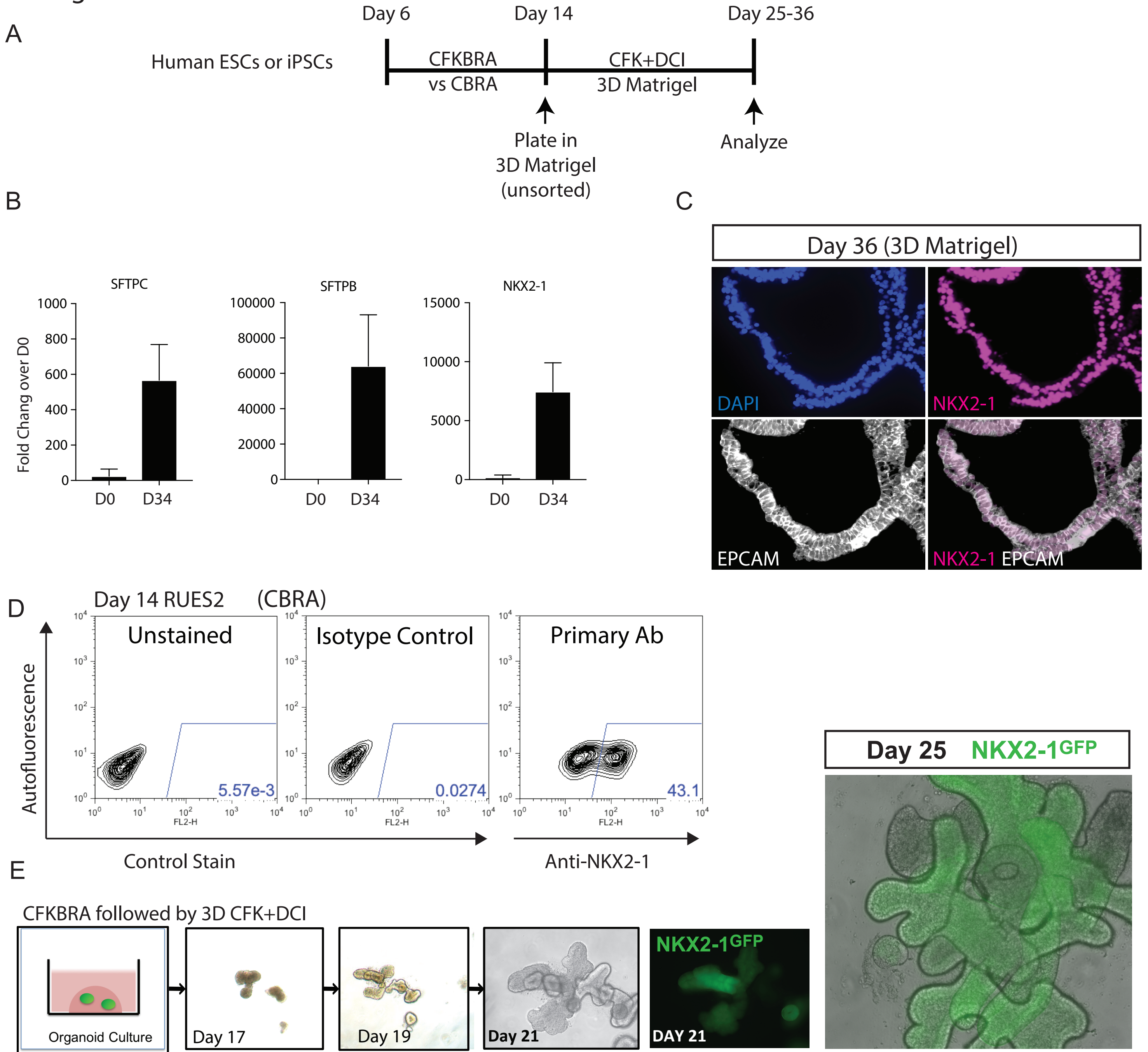
Supplemental Figure 1: Nkx2-1+ cells derived from mouse ESC-endoderm in the presence of BMP4 alone (days 6-14) do not exhibit lung competence. The mouse ESC line carrying the Nkx2-1^{mCherry} reporter was differentiated into anterior foregut-like endoderm and then exposed to either BMP4+Wnt3a or each factor alone from day 6 to 14 of directed differentiation in vitro. Only Nkx2-1+ cells specified in the presence of both BMP4+Wnt3a exhibited competence to proceed to differentiated lung epithelial cells, as evidenced by expression of the Sftpc^{GFP} lentiviral reporter by day 25. Nkx2-1+ cells induced in the presence of BMP4 alone did not display evidence of lung competence. No Nkx2-1+ cells emerged in the presence of Wnt alone. See also main figure 1. Faithfulness and specificity of the Sftpc reporter is presented in figure 3 and S2 as well as in Longmire et al, Cell Stem Cell 2012.

Development • Supplementary information

Figure S2



Supplemental Figure 2: Sorting Nkx2-1+ progenitors derived from mouse ESCs by day 14 identifies the entirety of distal-lung competent progenitors. (A) A subset of ESC-foregut endodermal-like cells exposed to Wnt3a+BMP4 displays induction of the Nkx2-1^{mCherry} reporter on days 6-14. Only Nkx2-1+ sorted cells display competence to express the lentiviral Sftpc^{GFP} reporter (labeled SPC-GFP+). After differentiation to day 25, the specificity of the lentiviral reporter is demonstrated since GFP+ sorting enriches for cells expressing Sftpc (top panel). These GFP+ cells also express the alveolar epithelial type 2 cell gene, Lamp3 (DC-Lamp) at levels higher than primary mouse lung tissue controls. Levels of Sftpc mRNA are approximately 25% of control lung levels. See also main figure 3. (B) Algorithm for repeat experiment of a; with cells analyzed on day 28 by RT-qPCR. (C) RT-qPCR expression of each indicated gene represented as fold change ($2^{-\Delta\Delta Ct}$; normalized for 18S) compared to day 14 sorted Nkx2-1$^{mCherry+}$ progenitors ("input").



Supplemental Figure S3: Supporting data on the characterization of human pluripotent stem cell (PSC)-derived NKX2-1+ cells.

(A) Schematic of protocol for differentiation of human PSC-derived day 14 cells in 3D Matrigel cultures. (B) RT-qPCR of gene expression in RUES2 cells on day 34 of differentiation (CFKBRA days 6-14). n=3. Cells were exposed to CFKBRA on days 6-14, followed by replating in 3D Matrigel in CFK+DCI thereafter. (C) Immunostaining of epithelial (EPCAM+/NKX2-1+) spheres on day 36 of outgrowth in the same culture conditions as B. Nuclei are counterstained with DAPI. (D) FACS gating and control stains employed on days 14-15 to accompany Figure 7 for quantitation of NKX2-1 protein positivity in cells generated in the presence of CBRA. Data shown are for the RUES2 human ESC line. (E) Matrigel 3D “organoid” culture of a single clump of human iPSC (C17) cells carrying the NKX2-1GFP reporter, followed over 10 days in culture and imaged for GFP fluorescence in the far right two panels. C=CHIR, F=FGF10, K=FGF7 (KGF), B=BMP4, RA=retinoic acid. DCI=dexamethasone, cyclic AMP, and IBMX. See also figure 7.

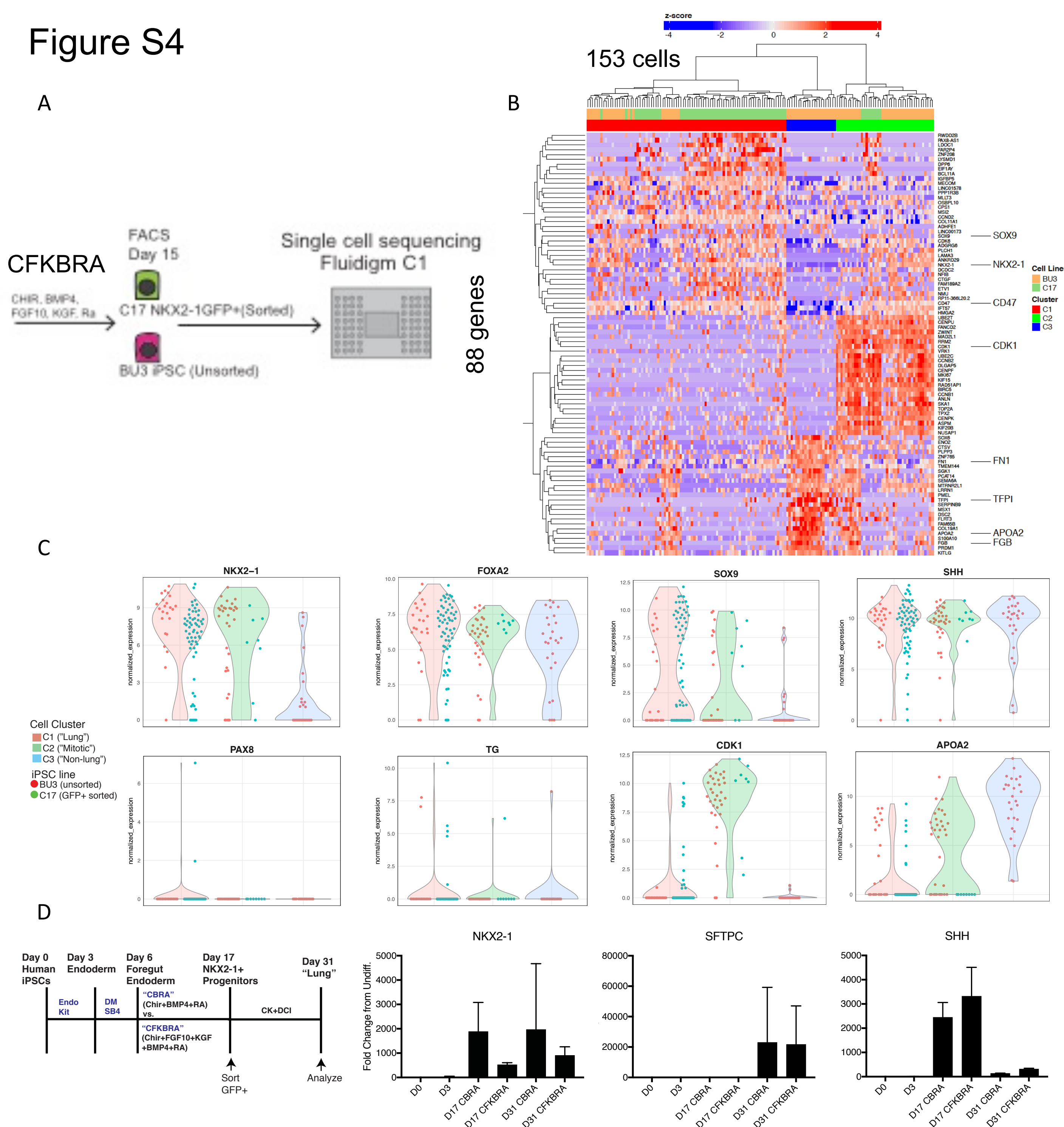


Figure S4: Human iPSC-derived NKX2-1+ endodermal progenitors generated with CFKBRA (without FGF2) are lung competent and do not express thyroid markers. (A) Experimental design outlining single cell RNA-Seq analysis profiling C17 NKX2-1^{GFP} sorted cells and BU3 unsorted cells on day 15 of differentiation of the indicated specification media, CFKBRA. (B) Hierarchical clustering of the most differentially expressed genes across the pairwise cluster comparisons. Each row represents the row-normalized expression of a gene with absolute log2 fold-change > 1.5 and FDR-adjusted p-value < 0.05. [(C1=NKX2-1+ cells, C2=proliferating cells (CDK1 high), C3=NKX2-1 negative non-lung endoderm (APOA2 high; liver-like.] (C) Violin plots and normalized expression levels of indicated genes in each cluster. Note NKX2-1+ cells are high in primordial lung/foregut markers, SHH, FOXA2, and SOX9, but low in expression of thyroid markers PAX8 and TG. (D) Head-to-head comparison of "lung" protocols where BU3 NKX2-1^{GFP}+ progenitors are specified with either CBRA vs. CFKBRA until day 17 and then further differentiated, as indicated, until day 31. Note the D17 samples are unsorted, whereas the day 31 samples are the outgrowth of sorted D17 GFP+ cells, replated from day 17-31. Gene expression of each indicated marker by RT-qPCR is shown ($2^{-\Delta\Delta Ct}$; 18S normalized.) C=Chir, F=FGF10, K=KGF, B=BMP4, RA=retinoic acid. DM=dorsomorphin. See also main figure 7.

Figure S5

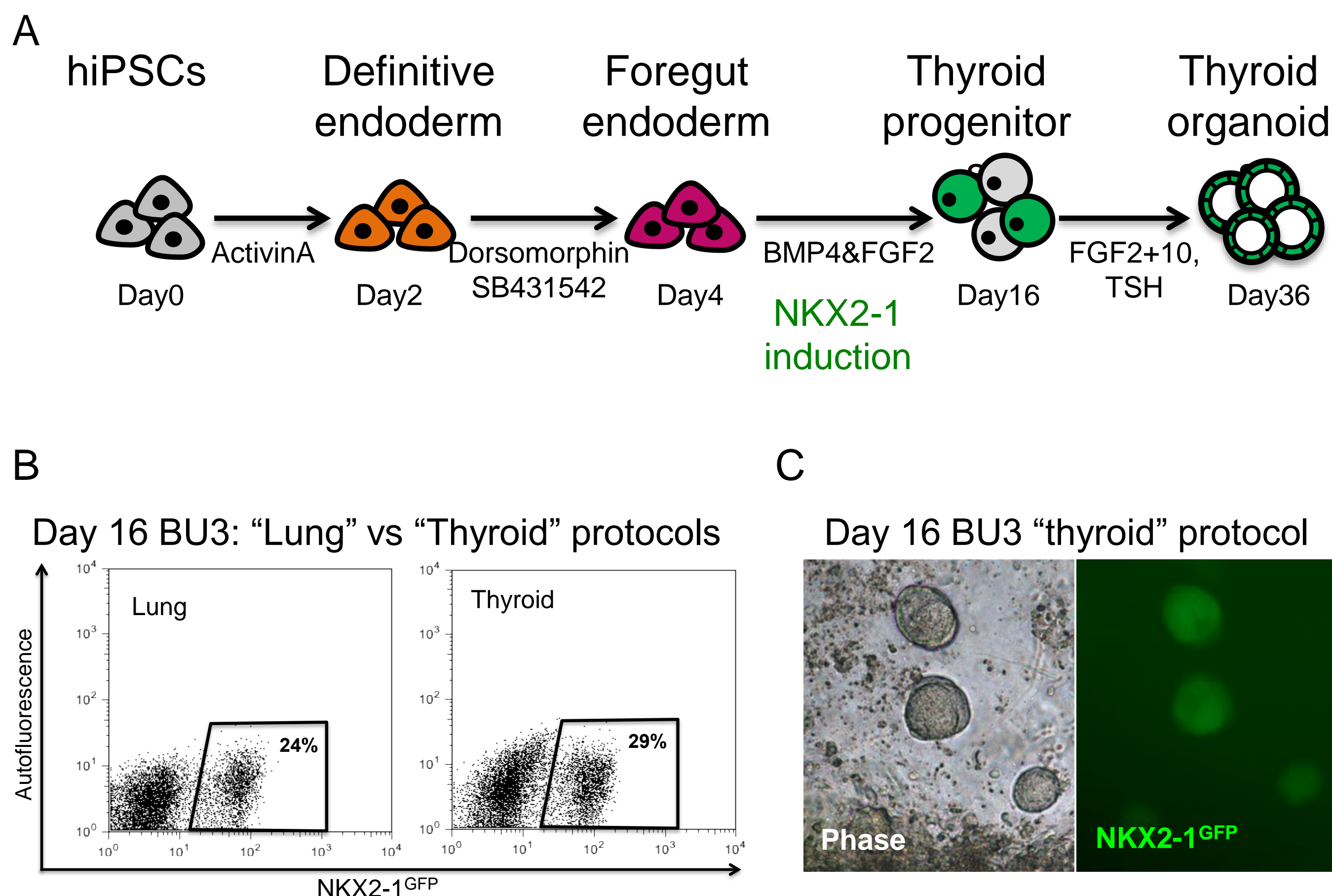
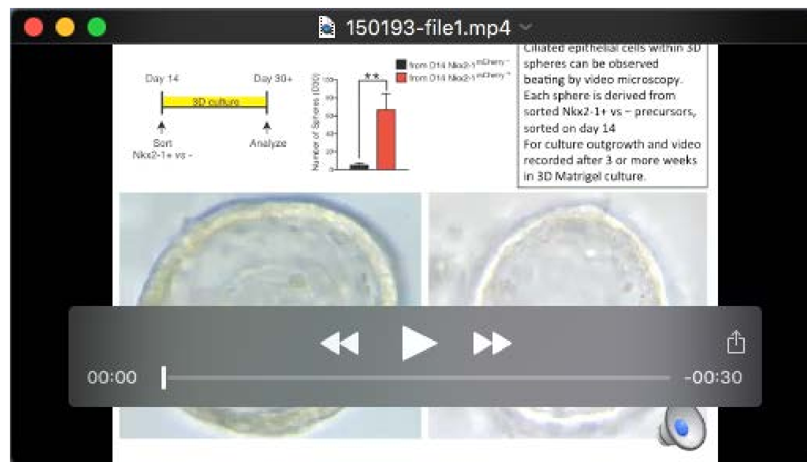


Figure S5: Human thyroid directed differentiation of iPSCs. (A) Schematic of protocol for thyroid differentiation. (B) Representative FACS dot plots indicating the day 16 efficiency of NKX2-1+ induction in BU3 iPSCs quantified based on expression of a NKX2-1^{GFP} reporter, after exposure to either the "lung" or "thyroid" differentiation recipes. (C) Phase and GFP fluorescence microscopy of cells in 3D matrigel culture demonstrating emergence of spherical clusters of GFP+ cells which are profiled in figure 7.

Supplemental Movie



Movie 1. Epithelial spheres in 3D culture contain beating multiciliated cells. Time-lapse microscopy of spherical colonies derived in 3D Matrigel from sorted Day14 Nkx2-1+ vs. – precursors, reveals beating multiciliated cells. The video was recorded after ≥3 weeks of 3D culture.

Table S1. Nine clusters of differentially expressed genes

Click here to Download Table S1

Development • Supplementary information