© 2017. Published by The Company of Biologists Ltd | Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

HUMAN DEVELOPMENT

RESEARCH ARTICLE

Pluripotent stem cell differentiation reveals distinct developmental pathways regulating lung- versus thyroid-lineage specification

Maria Serra[1,2], Konstantinos-Dionysios Alysandratos[1,2], Finn Hawkins[1,2], Katherine B. McCauley[1,2], Anjali Jacob[1,2], Jinyoung Choi[3], Ignacio S. Caballero[1], Marall Vedaie[1], Anita A. Kurmann[3], Laertis Ikonomou[1,2], Anthony N. Hollenberg[3], John M. Shannon[4,*,‡] and Darrell N. Kotton[1,2,*,‡]

ABSTRACT

The *in vitro*-directed differentiation of pluripotent stem cells (PSCs) through stimulation of developmental signaling pathways can generate mature somatic cell types for basic laboratory studies or regenerative therapies. However, there has been significant uncertainty regarding a method to separately derive lung versus thyroid epithelial lineages, as these two cell types each originate from Nkx2-1$^+$ foregut progenitors and the minimal pathways claimed to regulate their distinct lineage specification *in vivo* or *in vitro* have varied in previous reports. Here, we employ PSCs to identify the key minimal signaling pathways (Wnt+BMP versus BMP+FGF) that regulate distinct lung- versus thyroid-lineage specification, respectively, from foregut endoderm. In contrast to most previous reports, these minimal pathways appear to be evolutionarily conserved between mice and humans, and FGF signaling, although required for thyroid specification, unexpectedly appears to be dispensable for lung specification. Once specified, distinct Nkx2-1$^+$ lung or thyroid progenitor pools can now be independently derived for functional 3D culture maturation, basic developmental studies or future regenerative therapies.

KEY WORDS: Nkx2-1, Embryo, Endoderm, Lung, Pluripotent stem cells, Thyroid

INTRODUCTION

The *in vitro* differentiation of pluripotent stem cells (PSCs) into lineages that are otherwise difficult to access *in vivo* provides a novel source of cells for basic developmental studies, disease modeling and drug development. As with some primary cell types, transfer of these PSC-derived lineages into 3D culture systems at key developmental stages of differentiation has produced so-called 'organoids', 3D structures that begin to resemble the structural and cellular diversity of *in vivo* organs (Lancaster and Knoblich, 2014). Most published attempts to derive these differentiated cell types or structures from PSCs rely on *in vitro* recapitulation of known *in vivo* embryonic developmental signals; however, this approach can be problematic when the pathways regulating *in vivo* development of a particular tissue have not been established or appear to be poorly evolutionarily conserved across species. These hurdles are particularly apparent in prior attempts to generate lung epithelia from PSCs (Green et al., 2011; Hawkins and Kotton, 2015; Longmire et al., 2012; Mou et al., 2012). As the lung is an organ that emerged late in evolutionary time compared with other endodermally derived lineages, limited model systems based on embryos of lower species, most of which lack lungs, are available to study its developmental biology; therefore, reductionist mammalian *in vitro* model systems may help to examine the roles of individual germ layers or lineages in lung organogenesis. In particular, defining the minimal signaling pathways that specify a small group of progenitors in the anterior foregut endoderm into lung epithelial lineage, as marked by the onset of expression of Nkx2-1, has remained elusive.

In seminal *in vitro* work, Snoeck and colleagues used the Wnt signaling stimulator CHIR99021 (CHIR), together with FGF10, FGF7, BMP4, EGF and retinoic acid (RA), to direct the differentiation of PSCs into lung epithelial cells from anterior foregut endoderm (Green et al., 2011). This cocktail results in the acquisition of human lung cell fate and induction of NKX2-1 (Green et al., 2011; Huang et al., 2014). It differs significantly, however, from the growth factors employed in mouse models by us (Longmire et al., 2012) and others (Mou et al., 2012) to induce lung fate from mouse PSCs in culture, or from primary mouse foregut endoderm in explant models (Serls et al., 2005). A particularly dramatic and perplexing additional difference between species includes the observation that, in mouse PSC models, both lung and thyroid lineages, the two tissue types known to emerge via Nkx2-1$^+$ endodermal progenitors, tend to emerge together during *in vitro*-directed differentiation (Longmire et al., 2012), whereas in human PSC models lung epithelia without contaminating thyroid lineages can be generated (Dye et al., 2015; Huang et al., 2014). Thus, although stimulation of Wnt, BMP and FGF signaling pathways has been a unifying, common theme in most prior reports of lung differentiation (Green et al., 2011; Huang et al., 2014; Longmire et al., 2012; Mou et al., 2012), an unsettled controversy remains regarding whether human lung lineage specification fundamentally differs from mouse, requiring different developmental signals, such as different FGF ligands (e.g. ligands of FGFR2IIIb, such as FGF10 or FGF7), compared with lower species such as mice, which have been claimed to require the more broadly active FGF ligands FGF1 or FGF2 (Serls et al., 2005). For example, FGF1 or FGF2, via ligation of FGFR4, have been reported to be necessary and sufficient in mouse foregut endodermal explant models as an inducer of lung fate (Serls et al., 2005); however, mice deficient in FGF2 or FGFR4 develop lungs normally *in vivo* (Guzy et al., 2015; Weinstein et al., 1998; Zhou et al., 1998). Mice deficient in FGF10 or FGFR2IIIb display lung agenesis (De et al., 2000) and instead

[1]Center for Regenerative Medicine, Boston University and Boston Medical Center, Boston, MA 02118, USA. [2]The Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA. [3]Division of Endocrinology, Diabetes and Metabolism, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston MA 02215, USA. [4]Division of Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA.
*These authors contributed equally to this work

‡Authors for correspondence (john.shannon@cchmc.org; dkotton@bu.edu)

M.S., 0000-0002-9885-8093; D.N.K., 0000-0002-9604-8476

Received 6 February 2017; Accepted 15 September 2017

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

form a trachea-like structure. Specification of respiratory progenitors has occurred in FGF10-null embryos, however, as it has been shown that the mutant tracheal endoderm can be induced to form Sftpc-expressing organoids *in vitro* (Hyatt et al., 2004). This suggests that these FGF signals may act post-specification in branching morphogenesis and formation of primary lung buds. *In vivo* models of *Xenopus* and mouse lung development have also demonstrated the necessity of BMP signaling (Domyan et al., 2011; Rankin et al., 2016) and Wnt signaling (Goss et al., 2009; Harris-Johnson et al., 2009) for normal early lung development, causing further uncertainty as to whether these are the minimal signals required for lung specification or whether coincident FGF or other signaling is also necessary (Serls et al., 2005).

Further complicating matters are recent reports using the human PSC model system that employ widely varying multifactorial cocktails to induce lung fate (Dye et al., 2015; Green et al., 2011; Huang et al., 2014; Mou et al., 2012; Rankin et al., 2016; Wong et al., 2012), obscuring the possibility of distinguishing the minimal essential factors that act intrinsically on developing endoderm to specify lung cell fate. For example, combinations of Wnt/CHIR, BMP4, RA, SHH, FGF2, FGF4, FGF7, FGF10 or FGF18 have all been employed *in vitro* to induce lung fate in human PSC model systems in these varying reports. Only one previous report has addressed the key pathways required for lung specification across species, including frogs, mice and humans (Rankin et al., 2016).

Since the minimal pathways regulating *in vivo* lung lineage specification as well as their evolutionary conservation remain controversial, we employ a reverse approach, using PSC *in vitro* model systems to identify the key signaling pathways regulating lung lineage specification from foregut endoderm. In contrast to most previous claims, these minimal pathways appear to be evolutionarily conserved between murine and human species, and are similar to those recently found to regulate early lung specification in *Xenopus* and mice (Rankin et al., 2016). Our model systems suggest that FGF signaling, which was previously thought to be required for lung-lineage specification (Longmire et al., 2012; Roszell et al., 2009; Serls et al., 2005), appears to be dispensable, consistent with the findings reported for human PSC *in vitro* differentiation by Snoeck and colleagues (Huang et al., 2014). Of the many candidate signaling pathways previously proposed to regulate lung specification, Wnt+BMP signaling (in the presence of RA) appears to be necessary and sufficient to specify lung progenitors from foregut endoderm, whereas FGF +BMP signaling promotes specification to thyroid, the only other endodermal organ domain known to express *Nkx2-1*. Importantly, lung or thyroid progenitor pools can be isolated by cell sorting for functional 3D culture maturation once specified from PSCs, demonstrating that the progenitors specified with these minimal factors are competent to produce lung or thyroid epithelial spheres, respectively, for basic developmental studies or future regenerative medicine.

RESULTS

Wnt and BMP signals promote specification of lung-competent Nkx2-1^{+} endodermal progenitors from mouse ESCs, whereas FGF and BMP signals promote thyroid specification

To dissect the minimal factors needed to induce lung fate from anterior foregut endoderm, we used a previously developed *in vitro* model system in which mouse embryonic stem cells (ESCs) are differentiated into anterior foregut endoderm prior to differentiation into putative lung or thyroid cell fates (Kurmann et al., 2015; Longmire et al., 2012). The transcription factor *Nkx2-1* is selectively expressed in lung and thyroid epithelial lineages within definitive endoderm, and is the first known marker to be expressed upon lung- or thyroid-lineage specification, making it an ideal reporter for tracking the commitment of endodermal precursors into lung and thyroid epithelial cell fates (Kurmann et al., 2015; Longmire et al., 2012). We have previously used mouse ESCs carrying fluorochrome reporters (GFP or mCherry) targeted to the endogenous *Nkx2-1* locus to demonstrate that PSC-derived foregut endodermal cultures exposed to Wnt3a, BMP4 and FGF2 (in the presence of base media containing RA) upregulate both lung- and thyroid-lineage marker transcripts (Longmire et al., 2012). Cultures exposed to only BMP4 and FGF2, however, upregulate thyroid markers and show no robust expression of mature lung markers (Kurmann et al., 2015).

Using the mouse ESC line carrying an mCherry reporter targeted to the Nkx2-1 3′ untranslated region (hereafter Nkx2-1^{mCherry}) (Bilodeau et al., 2014; Kurmann et al., 2015), we employed our previously published differentiation protocol (Longmire et al., 2012; Kurmann et al., 2015) to derive anterior foregut endoderm over 6 days in serum-free culture, then tested the capacity of different combinations of Wnt3a, BMP4 and FGF2 to induce Nkx2-1^{+} from days 6-14 (Fig. 1A). Combinations of BMP4+FGF2, Wnt3a+BMP4, BMP4 alone or all three factors induced Nkx2-1$^{mCherry+}$ cells (Fig. 1B), whereas Wnt3a alone did not (Fig. S1). As previously published, FGF2 alone induced predominantly neuroectodermal Nkx2-1^{+} cells without evidence of lung-specific or thyroid-specific transcript expression (Longmire et al., 2012; Kurmann et al., 2015; and data not shown).

To test whether Nkx2-1^{+} cells produced under each condition contained lung or thyroid progenitors, we sorted each Nkx2-1$^{mCherry+}$ population on day 14 and replated Cherry^{+} cells in 2D cultures containing a serum-free FGF-supplemented medium (Fig. 1A,B) that we have previously shown to promote proliferation and outgrowth of Nkx2-1^{+} cells that co-express transcripts associated with lung and thyroid maturation (Longmire et al., 2012; Kurmann et al., 2015). Nkx2-1$^{mCherry+}$ cells specified with FGF2+BMP4 were competent to express markers of thyroid lineage differentiation by day 25 [*Tg*, *Tpo*, *Tshr* and *Slc5a5* (*Nis*); Fig. 1D] and did not upregulate lung differentiation markers (Fig. 1D). In contrast, the sorted Nkx2-1$^{mCherry+}$ population specified with Wnt3a+BMP4 upregulated markers of airway and alveolar lung epithelial differentiation by day 25 (*Sftpc*, *Sftpb*, *Scgb1a1* and *Aqp5*). Wnt3a+BMP4 induced Nkx2-1^{+} cells that were competent to differentiate into cells expressing ∼100,000-fold increased levels of *Sftpc* compared with our previously published Wnt3a+BMP4+FGF2 conditions. Although we have previously established that FGF2 is required for and promotes thyroid fate in developing foregut endoderm (Kurmann et al., 2015), these results unexpectedly implied that FGF2 may inhibit lung specification from endoderm in the mouse ESC model system during this narrow developmental window. Thus, we propose the foregut endodermal lineage specification model shown in Fig. 1E, in which lung and thyroid lineages specify distinctly in response to simple signals consisting of two exogenously induced signaling pathways (BMP4+Wnt3a versus BMP4+FGF2). We next sought to further test this model by assessing whether each distinct Nkx2-1^{+} endodermal population contained bona fide progenitors competent to form functional lung versus thyroid cells in 3D culture.

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

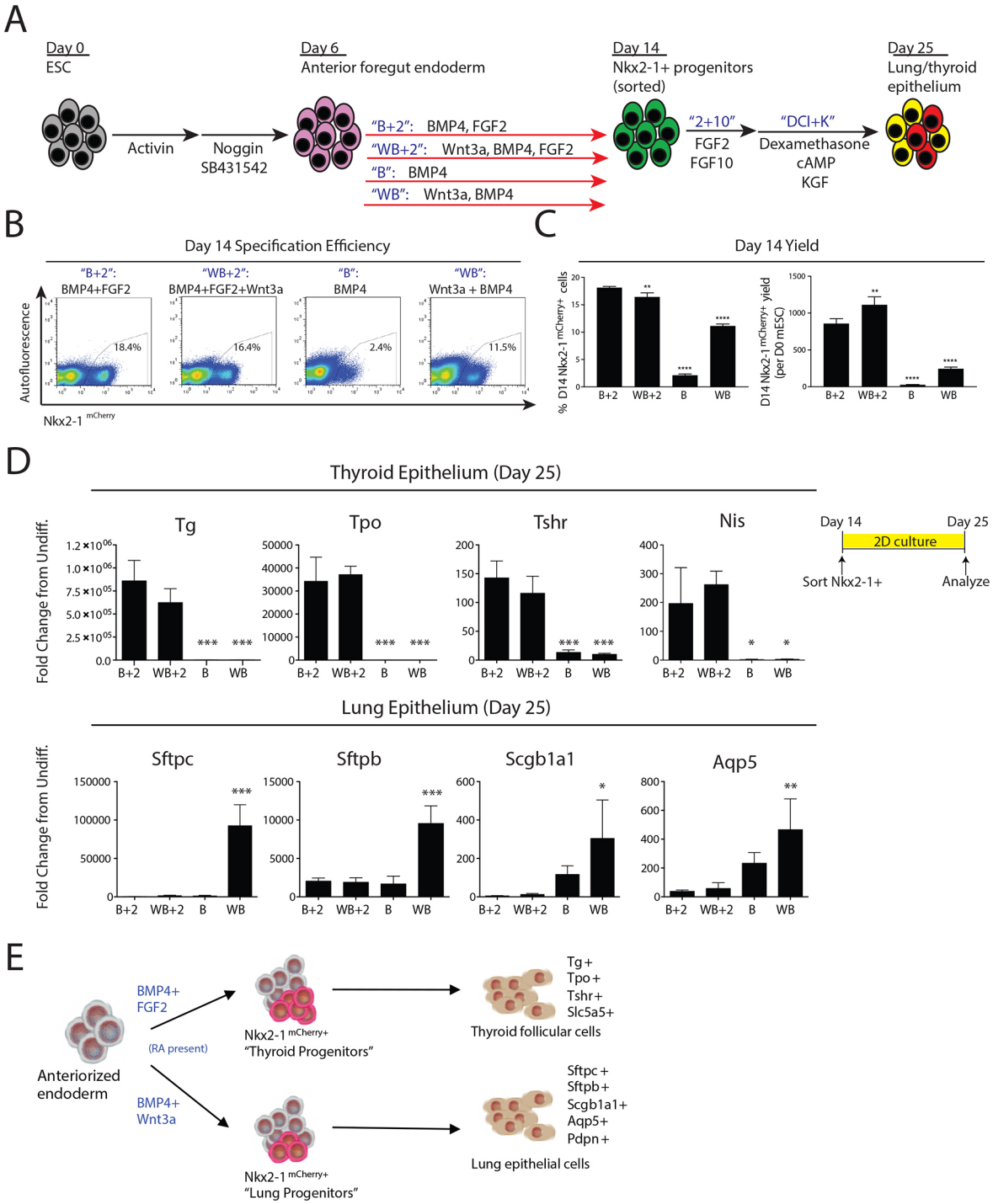


Fig. 1. Wnt and BMP promote lung specification of mouse ESC-derived Nkx2-1^{+} endodermal progenitors, whereas FGF and BMP signals promote thyroid specification. (A) Schematic depicting directed differentiation of mESCs into Nkx2-1^{+} endodermal cells, comparing various specification media. (B) Representative sort gates used to purify Nkx2-1^{mCherry} cells on day 14, showing efficiency of Nkx2-1^{+} reporter induction in each medium. (C) Day 14 Nkx2-1^{+} lung progenitor percentage and yield (per starting day 0 mESC) in each medium. (D) RT-qPCR on day 25 showing fold-change in gene expression over day 0 ($2^{-\Delta\Delta Ct}$). Data are mean±s.d. The schematic summarizes experimental design. (E) Schematic depicting proposed pathways for generation of thyroid versus lung lineages. See also Fig. S1. *P<0.05, **P<0.005 and *** or ****P<0.001 compared with B+2; one-way ANOVA with Tukey's multiple comparison test. n=3 biological replicates.

Nkx2-1^{+} cells specified with BMP4 and Wnt3a can form epithelial spheres in 3D culture that express markers of proximal and distal cells

To test whether the putative lung versus thyroid Nkx2-1^{+} endodermal progenitors specified in each condition displayed competence to form functional epithelial structures, we sorted each PSC-derived Nkx2-1$^{mCherry+}$ population to purity on day 14 for replating in 2D versus 3D cultures without any supporting mesenchyme (Fig. 2). Each condition, respectively, generated a day 14 yield of 239±22 versus 931±67 Nkx2-1^{+} cells per starting day 0

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

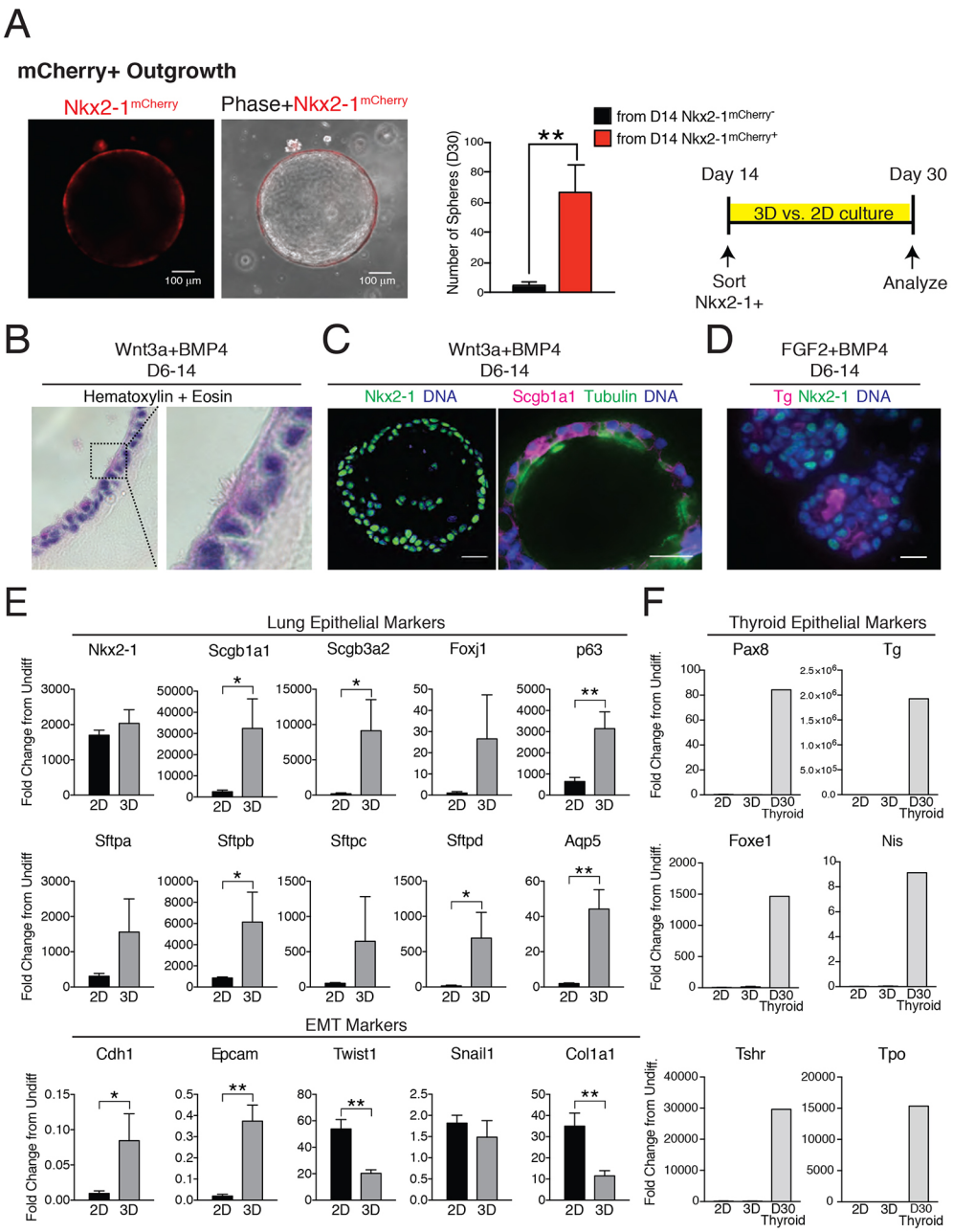


Fig. 2. Nkx2-1^{+} cells specified with BMP4 and Wnt3a can form epithelial spheres in 3D culture that express markers of proximal and distal cells. (A) Fluorescence microscopy of day 30 monolayered sphere derived in 3D Matrigel from Nkx2-1$^{mCherry+}$ cells. Quantitation of spheres formed from day 14 Nkx2-1^{+} versus Nkx2-1^{-} sorted populations (right panel). **P<0.05; unpaired t-test. (B) Bright-field photomicrograph of paraffin-embedded section showing an epithelial sphere derived from Nkx2-1^{+} cells stained with Hematoxylin and Eosin; enlarged view shows a multiciliated cell. (C) Immunostaining of day 30 epithelial spheres derived from Nkx2-1^{+} cells specified with BMP4+Wnt3a. Scgb1a1 (club cell marker), tubulin (cilia marker) and DNA staining (DAPI) are shown. Scale bars: 25 µm. (D) Immunostaining of day 30 thyroid follicle organoids derived in 3D Matrigel from Nkx2-1^{+} cells specified with BMP4+FGF2 for Nkx2-1 and thyroglobulin (Tg) with DNA stain (DAPI). Scale bar: 25 µm. (E) RT-qPCR on day 30 showing fold-change in gene expression over day 0 ($2^{-\Delta\Delta Ct}$), comparing 3D with 2D culture effects on expression of lung epithelial and epithelial mesenchymal transition (EMT) markers. Data are mean±s.d. n=3 biological replicates. *P<0.05, **P<0.01; unpaired t-test. (F) RT-qPCR on day 30 of 2D versus 3D cultures of Nkx2-1^{+} cells specified with Wnt3a+BMP4, showing fold-change in gene expression of thyroid epithelial markers over day 0 ($2^{-\Delta\Delta Ct}$) with day 30 mESC-derived thyroid outgrowths serving as positive controls. n=3 biological replicates.

'input' ESC (Fig. 1C), although these yields and specification efficiencies were not adjusted for day 6 definitive endoderm frequencies. We have previously demonstrated that, following thyroid-lineage specification, 3D culture outgrowth conditions augment subsequent thyroid epithelial gene expression and allow spontaneous self-organization of Nkx2-1^{+} cells into epithelial monolayered structures that resemble primary thyroid follicular epithelium. Importantly, these PSC-derived thyroid follicles are capable of sequestering thyroglobulin and initiating thyroid hormone biogenesis both *in vitro* and after transplantation *in vivo* (Kurmann et al., 2015; Dame et al., 2017). We verified that Nkx2-1^{+} cells specified with BMP4+FGF2 formed thyroid follicular epithelia (Fig. 2D), but lacked lung epithelial differentiation capacity in 3D culture. In marked contrast, sorted Nkx2-1$^{mCherry+}$ cells specified with Wnt3a+BMP4 lacked thyroid competence (Fig. 1D), but exhibited lung differentiation potential, including the capacity to organize into monolayered epithelial spheres expressing a broad diversity of airway and alveolar epithelial genes (*Nkx2-1*, *Scgb1a1*, *Scgb3a2*, *Foxj1*, *p63*, *Sftpa*, *Sftpb*, *Sftpc*, *Sftpd* and *Aqp5*; Fig. 2A-E). 3D culture of the Wnt3a+BMP4-specified sorted Nkx2-1^{+} cells significantly augmented lung (but not thyroid) epithelial gene expression, while suppressing mesenchymal marker expression, compared with 2D culture conditions (Fig. 2E,F). These results were specific to the mCherry^{+} population as Nkx2-1$^{mCherry-}$ cells sorted on day 14 for replating exhibited significantly lower sphere-forming capacity (Fig. 2A) and little, if any, detectable expression of lung marker genes (Fig. S2).

We sought to determine whether recognizable proximal and distal lung epithelial phenotypes were emerging in the 3D cultured outgrowths of sorted Nkx2-1^{+} cells. The day 14 sorted Nkx2-1^{+} cells replated for 3D outgrowth formed at least two morphologically distinguishable colonies after 3 to 5 days: larger spheroid-shaped colonies and smaller irregular colonies (Fig. 2A). Phase-contrast microscopy, immunofluorescence microscopy and Hematoxylin

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

and Eosin staining of the spheroid-type colony revealed the presence of secretory SCGB1A1^{+} cells, as well as beating, multiciliated, tubulin^{+} epithelial cells facing the inner lumina of these spheres, suggestive of differentiation of proximal airway cell types (Fig. 2B,C; Movie 1). In contrast, we observed pro-SFTPC protein expression in cells within irregular smaller clusters of cells, suggestive of distal epithelial differentiation (Fig. 3C,E).

The two distinct colony morphologies we observed arising from Nkx2-1^{+} sorted cells appeared to resemble the spherical airway versus irregular alveolar colony morphologies previously reported to arise from primary mouse airway and alveolar lung epithelia, respectively (Bilodeau et al., 2014; Lee et al., 2014). To further test the hypothesis that irregular-type PSC-derived colonies corresponded to distal lung cell types, we engineered a lentiviral reporter of distal lung epithelial gene expression composed of a previously published 3.7 kb human SFTPC proximal promoter element (Glasser et al., 1991) driving expression of a GFP reporter cDNA (hereafter SftpcGFP; Fig. 3A). We have previously published the faithfulness and specificity of a SftpcdsRed version of this lentiviral vector to identify murine cells that express *Sftpc* (Longmire et al., 2012). To test faithfulness of this lineage-specific lentiviral reporter in the mouse ESC *in vitro* differentiation system, we first tested the SftpcGFP vector in our 2D ESC system. We sorted Nkx2-1$^{mCherry+}$ versus Nkx2-1$^{mCherry-}$ cells on day 14 of ESC differentiation and on day 16-18 transduced each replated and sorted population with SftpcGFP lentivirus (Fig. 3A,B). GFP^{+} cells were observed arising from the Nkx2-1$^{mCherry+}$ sorted population as early as 24 h after transduction and appeared as distinct cell clusters that expanded over time in 2D culture (Fig. 3C). Few if any GFP^{+} cells emerged from the sorted Nkx2-1$^{mCherry-}$ population (Fig. S2). On day 30, we sorted SftpcGFP^{+} cells (representing 1-8% of the total outgrowth cells), and found these cells were enriched in

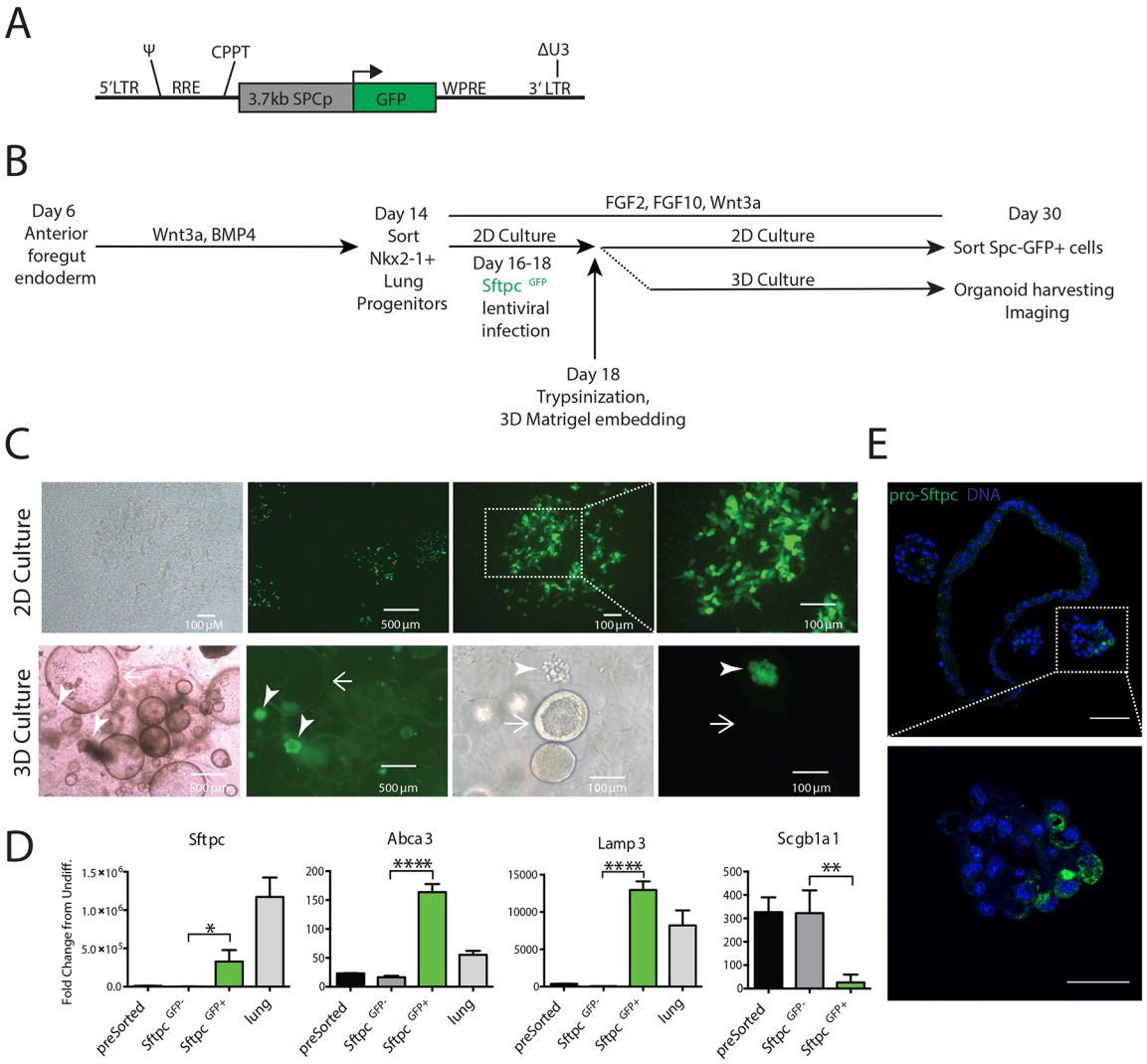


Fig. 3. SftpcGFP lentiviral reporter identifies alveolar epithelial differentiation of sorted Nkx2-1^{+} cells specified with BMP4+Wnt3a. (A) Schematic of SftpcGFP lentivirus. SPCp, human SPC promoter element; LTR, lentiviral long terminal repeats; RRE, rev responsive element; CPPT, central polypurine tract; WPRE, Woodchuck hepatitis virus post-transcriptional regulatory element; ΔU3, deleted U3 region for *in vivo* inactivation of the viral LTR promoter; Ψ, Psi lentiviral packaging sequence. (B) Schematic of mESC lung differentiation protocol and timing of SftpcGFP lentiviral infection and 3D Matrigel embedding. (C) Representative micrographs on day 30 showing SftpcGFP expression after 2D or 3D expansion of Nkx2-1^{+} sorted progenitors. Note that in 3D conditions, SftpcGFP^{+} clusters are irregular (arrowheads), whereas larger circular spheres (arrows) do not express the reporter. (D) RT-qPCR showing fold-change in gene expression over day 0 ($2^{-\Delta\Delta Ct}$) in sorted SftpcGFP^{+} and SftpcGFP^{-} cells. Data are mean±s.d. n=3 biological replicates. *P<0.05, **P<0.01, ****P<0.0001; unpaired t-test. See also Fig. S2. (E) Confocal microscopy of day 30 epithelial spheres derived from Nkx2-1^{+} cells specified with BMP4+Wnt3a for pro-SFTPC immunostaining with DNA stain (Hoechst). Scale bars: 50 µm (upper panel); 25 µm (lower panel).

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

expression of alveolar epithelial transcripts (*Sftpc*, *Abca3* and *Lamp3*) compared with presorted or SftpcGFP^{-} cells (Fig 3D, Fig. S2), whereas expression of the proximal secretory lung marker *Scgb1a1* was lower in the GFP^{+} compared with the GFP^{-} population (Fig. 3D). We therefore concluded the SftpcGFP lentiviral vector was a faithful reporter of endogenous distal alveolar epithelial gene expression. Next, we embedded the Nkx2-1^{+} cells infected with SftpcGFP lentivirus in Matrigel drops on differentiation day 18 to allow for 3D organoid formation (Fig. 3C). We detected SftpcGFP^{+} cells only in the smaller irregularly shaped colonies, whereas the larger spherical colonies containing beating cilia were rarely found to contain SftpcGFP^{+} cells. Taken together, these results suggested that the sorted day 14 Nkx2-1^{+} population specified with Wnt3a+BMP4 contains lung progenitors competent to form proximal airway and distal lung epithelia.

Transcriptional profiles of *in vitro*-derived putative lung and thyroid progenitors resemble those of *in vivo* Nkx2-1^{+} lung and thyroid progenitors

Having shown that distinct lung-competent versus thyroid-competent Nkx2-1^{+} progenitors could be generated from PSCs using distinct growth factor combinations, we next sought to profile the genetic programs of each progenitor population on day 14 of differentiation. Global transcriptomic profiles of the four sorted populations were obtained by microarray analysis (mCherry^{+} versus mCherry^{-} sorted samples prepared in 'lung' versus 'thyroid' media; Fig. 4A,B). *Nkx2-1* and the neighboring locus encoding the *Nkx2-1* associated non-coding intergenic RNA [NANCI/LL18 (Herriges et al., 2014), also known as E030019B13Rik] were each in the top three genes differentially expressed between mCherry^{+} and mCherry^{-} populations in either medium [ranked by either fold-change (FC) or *P*-value adjusted for false discovery rates (FDR)]. To identify differentially expressed genes that distinguish the four populations, we used moderated *t*-tests to establish the effect of specification media, the effect of Nkx2-1^{mCherry} expression status and an 'interaction effect' between Nkx2-1^{mCherry} status and medium condition (Fig. 4C, see Table S1 and supplementary Materials and Methods). We found that 1315 significantly differentially expressed genes, grouped into nine clusters (Fig. 4D), distinguished the four populations ($\log_2$ FC>2 between any two populations with interaction effect FDR<0.25). To assess whether this analysis revealed gene clusters known to be associated with either thyroid or lung, we searched for markers of lung or thyroid epithelia, as well as markers of lateral plate mesoderm or developing lung mesenchyme (Fig. 4C,D) (Grindley et al., 1997; Motoyama et al., 1998; Rankin et al., 2016; Sato et al., 2008; Zhang et al., 2013). We found that gene cluster 5 was uniquely enriched in the Nkx2-1$^{mCherry+}$ cells induced in 'thyroid media' (Fig. 4D). This cluster contained the core thyroid transcription factors *Pax8*, *Foxe1* and *Hhex*, as well as the thyroid specific marker *Tg*. Furthermore, cluster 5 contained a variety of additional transcripts previously shown to be enriched in developing or adult thyroid epithelial cells, including *Prlr* (Fagman et al., 2011), *Cd36* and genes known to be involved in thyroid hormone biogenesis, such as *Iyd*, *Slc16a2* (also known as *Mct8*; Schwartz and Stevenson, 2007) and *Duoxa2* (Grasberger et al., 2012) (Fig. 4E). *Sftpb* was also enriched in this cluster, a finding in keeping with the reported expression of this transcript in primary, as well as ESC-derived, thyroid epithelial cells (www.gtexportal.org/home/gene/SFTPB and Dame et al., 2017).

Genes known to be expressed in developing lung epithelial cells were found in cluster 2, which was enriched in the Nkx2-1$^{mCherry+}$ population induced in 'lung media' (Fig. 4D). This gene cluster included the transcription factors *Foxa1* (Fagman et al., 2011; Rossi et al., 1999), *Irx2* and *Lef1*, a known factor involved in Wnt signaling and lung development (Xie et al., 2014). Lung epithelial-selective differentiation markers *Sftpc* and *Cldn18* (Schlingmann et al., 2015), and the lung-specific cytokine *Cxcl15* (also known as 'lungkine'; Rossi et al., 1999) were found in this cluster, as were other genes previously described as expressed in the lung: *Gprc5a*, *Shh*, *Lama3*, *Lgi3* and *Wif1* (Fagman et al., 2011; Xu et al., 2011) (Fig. 4E). Interestingly, some genes observed in cluster 1 (unique to Nkx2-1$^{mCherry-}$ cells induced in 'lung medium') suggested potential enrichment of putative lung mesenchyme (*Foxf1*, *Gli1*, *Wnt2* and *Tbx4*; Motoyama et al., 1998; Rankin et al., 2016; Zhang et al., 2013) or non-lung foregut endodermal derivatives, such as esophagus and liver (*Pitx1/2* and *Ttr*). Cluster 6 comprised genes enriched in Nkx2-1$^{mCherry+}$ cells induced in both 'thyroid' and 'lung' media, and contained *Nkx2-1*, non-specific epithelial transcripts, such as *Cdh1* and *Itga6*, as well as semaphorins and the surfactant homeostasis regulator *Gpr116* (Yang et al., 2013). Taken together, our global transcriptomic analyses indicated that the genes found in each cluster were consistent with BMP4+Wnt3a and BMP4+FGF2 specifying early lung and early thyroid epithelial identity.

Confirming our microarray findings by RT-qPCR (Fig. 5), we found that day 14 putative primordial thyroid progenitors induced by FGF2+BMP4 expressed transcript markers of thyroid lineage specification, *Hhex*, *Pax8* and *Foxe1*, and both putative lung and thyroid lineages expressed *Cdh1*, *Nkx2-1* and *Epcam*. Additional transcripts that have been previously shown to distinguish early thyroid from early lung lineages *in vivo* in mice (Fagman et al., 2011) were also enriched in each respective PSC-derived lineage: *Cd44* and *Prlr* in the putative thyroid population; and *Slc15a2* and *Foxa2* in the putative lung population. *Shh* was enriched in the lung and absent in the thyroid progenitors, a finding in keeping with recent work demonstrating Shh signaling is required for lung but is dispensable for thyroid lineage specification *in vivo* (Rankin et al., 2016). Furthermore, the Wnt signaling target gene *Axin2* was enriched in PSC-derived lung progenitors and absent in thyroid progenitors. Even when thyroid progenitors were formed in the presence of Wnt3a, these cells appeared to be unresponsive to Wnt as *Axin2* remained suppressed (Fig. 5). These results are consistent with the necessity of Wnt signaling in primordial Nkx2-1^{+} lung endodermal cells but its dispensability in thyroid cells (Kurmann et al., 2015).

FGF signaling is dispensable for lung-lineage specification from endoderm, whereas Wnt stimulation is required

Our *in vitro* ESC model system unexpectedly suggested that FGF2 reduces expression of all lung markers while increasing early and mature thyroid markers (Fig. 1). This suggested FGF signaling might be required to specify Nkx2-1^{+} endodermal cells with thyroid, but not lung, competence. To test this hypothesis, we performed experiments in mouse foregut explants. We have previously shown that chemical inhibitors of BMP and FGF signaling blocked *Nkx2-1* and *Pax8* expression in the thyroid domain in developing *Xenopus* and mouse foregut explants (Kurmann et al., 2015). We used the same approach here to explore the effect of BMP or FGF chemical inhibition on *Nkx2-1* expression in lung and thyroid fields in mouse embryo foreguts and explant cultures (Fig. 6). Co-staining of FOXA2 and NKX2-1 revealed that NKX2-1 was first detected at 6-8 ss (somite stage) in forebrain (Fig. 6A) and was only weakly detected at 10 ss in the

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

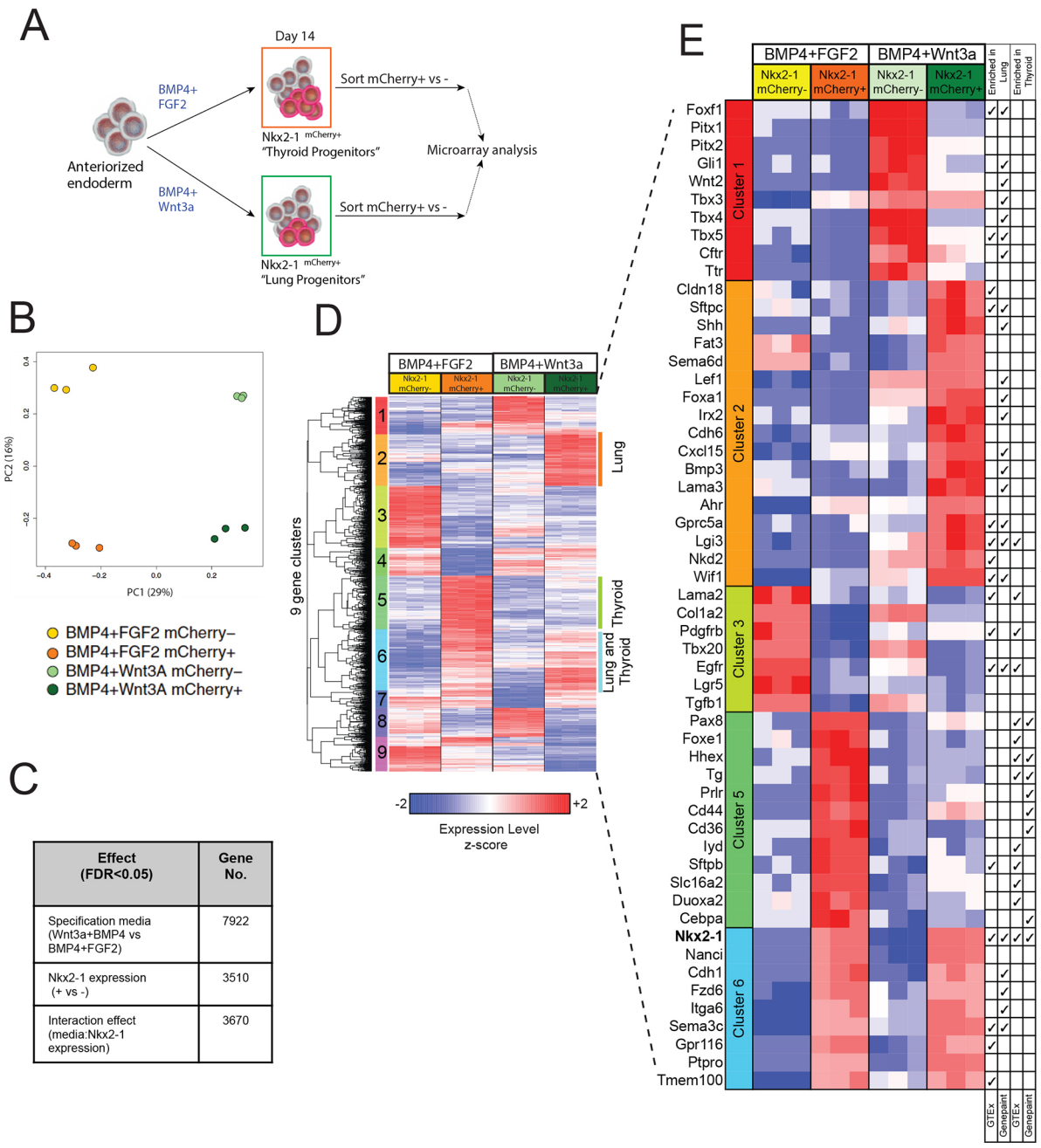


Fig. 4. Global transcriptomic profiles of *in vitro*-derived putative lung and thyroid progenitors. (A) Schematic of experimental design of microarray analysis of global transcriptomic profiles of thyroid versus lung progenitors. (B) Principal component analysis (PCA) across all genes and samples. (C) Table summarizing numbers of differentially expressed genes using a moderated *t*-test in each indicated comparison (medium effect, mCherry^{+} versus mCherry^{-} status, or interaction effect). (D) Heat map representing unsupervised hierarchical clustering of samples analyzed in the microarray, based on the 1315 differentially expressed transcripts by 'interaction effect' of specification medium and Nkx2-1 expression (FDR<0.25 and FC>2), as detailed in the Materials and Methods. Transcripts with similar patterns of gene expression were grouped into nine clusters. (E) Heatmaps of selected transcripts from five of the clusters shown in D. Clusters 2 and 5, differentially expressed lung- or thyroid-specific transcripts; cluster 6, transcripts expressed in both 'lung' and 'thyroid' conditions; clusters 1 and 3, selected mesenchymal or non-lung, non-thyroid endodermal transcripts differentially expressed in Nkx2-1$^{mCherry-}$ cells. Table attached to heatmap indicates genes expressed *in vivo* in the lung or thyroid in human adults or mouse embryos per GTEx Portal (http://www.gtexportal.org/) and GenePaint (www.genepaint.org/).

thyroid (not shown), with strong thyroid detection but no lung detection appreciated by 12 ss (~E8.5; Fig. 6A). By E9, both thyroid and lung Nkx2-1^{+} domains were detectable, with lung being FOXA2^{+} and thyroid FOXA2^{-}, consistent with previous reports (Fagman et al., 2011). To model the sequence of thyroid- and lung-lineage specification from endoderm, we developed a mouse foregut explant culture system, starting with foreguts harvested at 6-10 ss of development (prior to any endodermal *Nkx2-1* expression) followed by a 2-day explant culture period during which both thyroid and lung Nkx2-1^{+} domains have been specified, as detected by NKX2-1 immunostaining (Fig. 6B, left panel). In the presence of the BMP inhibitor DMH1, both lung and thyroid specification was completely blocked and no NKX2-1 signal was detected within the ECAD^{+}/SOX2^{+} developing foreguts (Fig. 6B, right panel). When

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

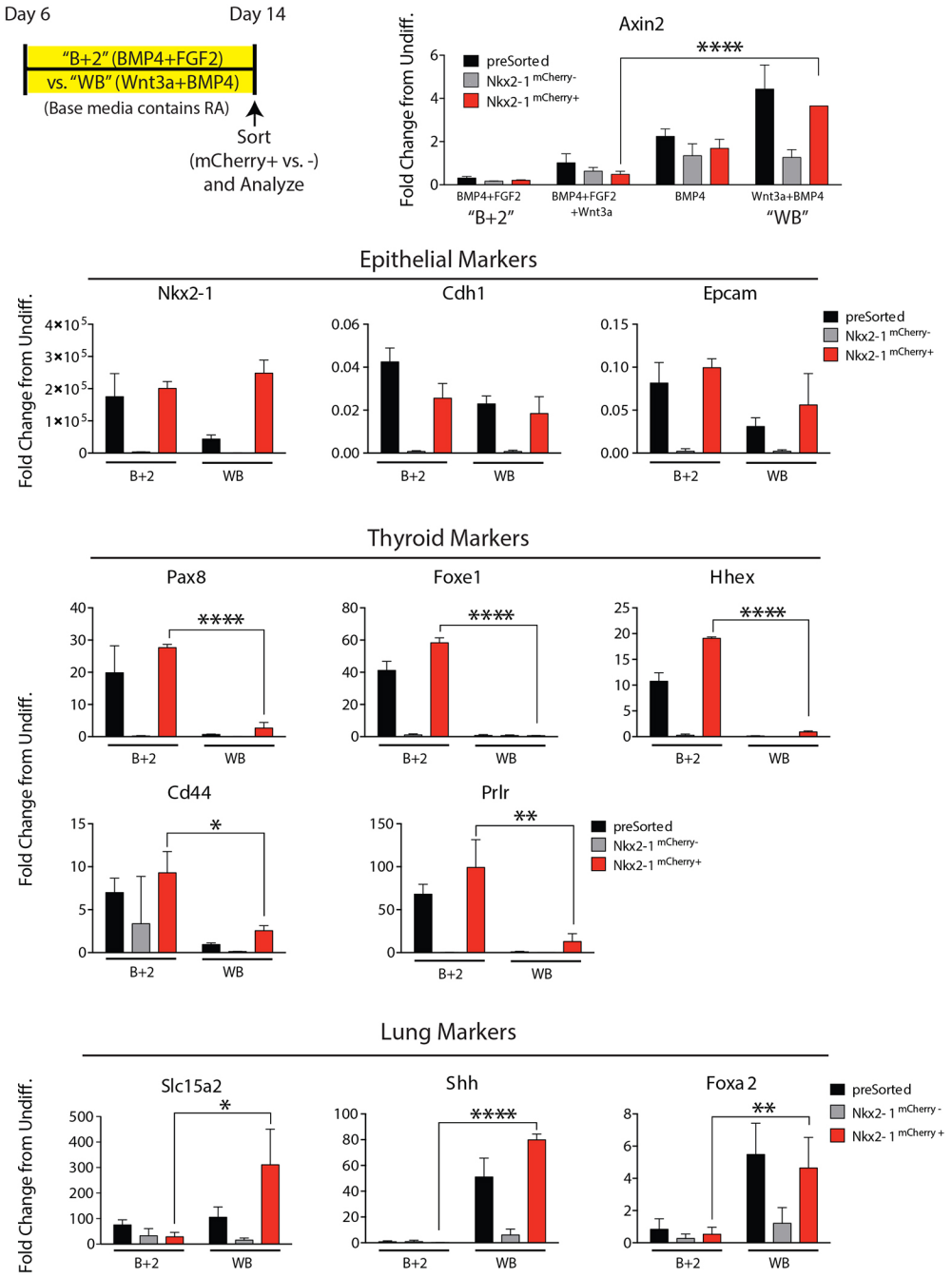


Fig. 5. Validation of microarray findings by RT-qPCR of key epithelial, early thyroid- or lung-specific markers. Schematic depicting experimental design. Data are mean±s.d. of fold-change in gene expression over day 0 ($2^{-\Delta\Delta Ct}$). n=3 biological replicates. *P<0.05, **P<0.01, ****P<0.0001; two-way ANOVA.

6-8 ss embryos were incubated for 48 h in three different inhibitors of FGF signaling (PD173074, BGJ398 or PD161570), we observed diminished NKX2-1 immunostaining in the thyroid but not the lung domain, in contrast to vehicle control-exposed embryos (Fig. 6C). Despite the presence of each FGF inhibitor, NKX2-1 expression in the lung domain was preserved; the thyroid domain, however, was smaller than in vehicle-exposed controls, and there was a reduced distance between the lung and thyroid domains. These results suggested FGF inhibition specifically perturbed thyroid, but not lung, primordium formation. We considered the possibility that FGF signaling to the endoderm at an earlier developmental stage might still be required for lung-lineage specification. Thus, we repeated our FGF blockade experiments, initiating blockade at the presomitic stage of endodermal development and continuing FGF inhibition through the stage of expected lung-lineage specification. In embryo explants cultured in the presence of each of the three FGF inhibitors, we observed no detectable NKX2-1 expression in the region of expected foregut thyroid, demonstrating that thyroid lineage specification was completely blocked; however, expression of NKX2-1 in the foregut endodermal lung domain was preserved (Fig. 6D). Taken together, these results suggest that FGF signaling is dispensable for lung lineage specification but required during a narrow developmental window for thyroid specification.

Next we sought to determine whether endodermal Nkx2-1^{+} foregut progenitors specified in the absence of FGF signaling, but in the presence of Wnt and BMP inducers, were bona fide lung progenitors competent to bud, branch and differentiate into lung epithelial cells. Our goal was to isolate mouse anterior endoderm,

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

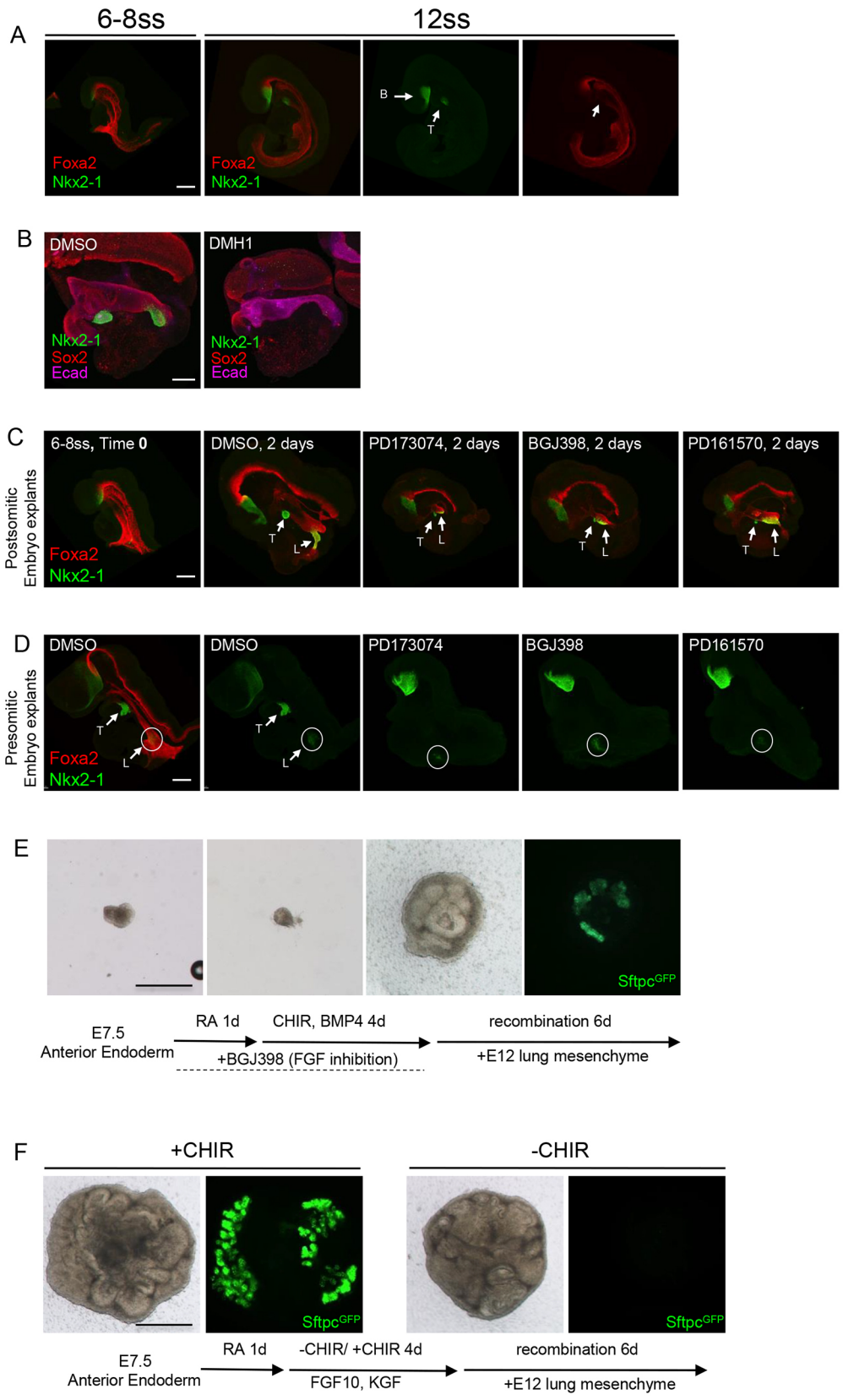


Fig. 6. Conserved pathways induce lung and thyroid cell fate in the developing mouse embryo. (A) Co-staining for Nkx2-1 and Foxa2 in mouse embryos at 6-8 ss and 12 ss. White arrows indicate forebrain (B) and thyroid (T) primordium. (B) Co-staining for Nkx2-1, Sox2 and Cdh1 (Ecad) in mouse foregut explants harvested at 6-10 ss and cultured for 2 days in medium supplemented with vehicle control (DMSO) or BMP signaling inhibitor DMH1. (C) Co-staining for Foxa2 and Nkx2-1 in 6-8 ss mouse foregut explants cultured for 2 days in control medium (DMSO) or media supplemented with FGF inhibitors PD173074, BGJ398 or PD161570. There are three domains of Nkx2-1 staining: brain, thyroid (T) and lung (L). (D) Co-staining for Foxa2 and Nkx2-1 in mouse embryos harvested at the presomitic stage and cultured for 2 days in control medium (DMSO) or medium supplemented with FGF inhibitors PD173074, BGJ398 or PD161570. (E) Mouse embryonic explant culture system where E7.5 anterior endoderm was isolated from a SftpcGFP transgenic mouse embryo and incubated with RA for 24 h, then with the BMP4+Wnt agonist CHIR99021 (CHIR) for 4 days in the presence of FGF chemical inhibitor BGJ398. SftpcGFP reporter expression and branching is induced after recombination with E12 embryonic mouse lung mesenchyme. (F) Mouse embryonic explant culture system in which E7.5 anterior endoderm from a SftpcGFP mouse embryo was incubated with either control medium or medium containing FGF10 and KGF with or without CHIR before recombination with E12 embryonic mouse mesenchyme. Scale bars: 150 µm in A,C,D; 200 µm in B; 500 µm in E,F. Embryos shown in B-F are representative of three repeated independent experiments consisting of a total of 10-12 embryos per condition shown.

free of the presence of contaminating mesoderm, in order to determine whether stimulation of Wnt and BMP signaling intrinsically in definitive endoderm, despite the presence of FGF blockade, is sufficient to induce lung progenitors competent to branch and differentiate. We developed an embryonic explant culture system where anterior endoderm isolated at the E7.5 early head-fold stage from SftpcGFP transgenic mice could be patterned first into lung-competent endoderm by exposure to RA, as recently

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

published (Rankin et al., 2016), followed by exposure to CHIR plus BMP4 during a 5-day culture period when the FGF inhibitor BGJ398 was always present (Fig. 6E). To determine whether lung-competent progenitors had been specified from endoderm by CHIR +BMP4 in the presence of FGF inhibition during this 5-day culture period, we then recombined the resulting endoderm for 6 days of culture with E12 mouse lung mesenchyme, a tissue we have previously demonstrated is able to induce branching and distal lung epithelial differentiation in developing lung epithelial cells (Shannon, 1994; Shannon et al., 1998). The endoderm rudiments cultured in the presence of CHIR, BMP4 and BGJ398 up to the time of recombinant culture, did not express the SftpcGFP reporter or exhibit any branching at the onset of recombinant culture, but subsequently underwent branching and upregulated expression of SftpcGFP after 6 days of recombination with E12 lung mesenchyme (Fig. 6E). These results indicate that primary definitive endoderm stimulated with RA followed by CHIR and BMP4 in the presence of FGF inhibition indeed contained lung progenitors competent to branch and differentiate into SftpcGFP^{+} lung epithelial cells. In order to confirm the requirement for canonical Wnt signaling in providing lung competence to the developing endoderm explant, we incubated the RA-treated endoderm with specification medium, either supplemented with CHIR or with control medium without CHIR, before recombining it with the lung mesenchyme. In comparison with CHIR-treated endoderm, which branched extensively and induced expression of the SftpcGFP reporter, endoderm never exposed to Wnt activation did not induce expression of the reporter. Although exogenous BMP addition or withdrawal in this model had no impact on lung specification (data not shown), the necessity of endogenous BMP signaling in the explant model was confirmed in separate experiments where addition of DMH1 blocked SftpcGFP expression (Fig. 6B; data not shown).

Conserved pathways induce lung cell fate in human ESC/iPSC-derived endoderm

Next we questioned whether the same minimal signaling pathways are also required to specify lung-competent NKX2-1-expressing endoderm in human PSCs differentiating *in vitro*. A combination of CHIR, FGF10, KGF, BMP4 and retinoic acid (hereafter CFKBRA) has been demonstrated to induce NKX2-1 expression in anterior foregut cells derived *in vitro* from human ESCs and induced PSCs (iPSCs) (Green et al., 2011; Huang et al., 2014; Kurmann et al., 2015; Hawkins et al., 2017) (Fig 7A, Fig. S3). First, using human PSC lines engineered to carry a GFP reporter targeted to the NKX2-1 locus (hereafter NKX2-1^{GFP}) (Hawkins et al., 2017), we verified that CFKBRA treatment of PSC-derived foregut endodermal cells generated NKX2-1^{+} cells by days 14-15, and further differentiation in 3D Matrigel cultures gave rise to EPCAM^{+}/NKX2-1^{+} spheres of epithelial cells with detectable *SFTPC* and *SFTPB* expression by RT-qPCR (Fig 7A-D, Fig. S3). We screened for the presence of any cells within the NKX2-1^{+} endodermal population generated with CFKBRA that might express thyroid markers (*PAX8*, *FOXE1*, *HHEX* or *TG*) by analyzing our published single-cell RNA-Seq dataset, profiling 153 individual iPSC-derived cells on differentiation day 15 (iPS17 and BU3 iPSC lines). Consistent with our findings in mouse Nkx2-1^{+} lung progenitors, we found the majority of human NKX2-1^{+} cells expressed high levels of *FOXA2* and *SHH*, but few (if any) cells expressed thyroid markers *PAX8*, *HHEX* or *TG*, and no cells expressed *FOXE1* (Fig 7E, Fig. S4), as expected in the absence of FGF2. Next, we tested the minimal exogenous factors required to induce lung-competent endoderm in human cells by withdrawing each of the five factors in CFKBRA one at a time from our lung-specification medium. Consistent with a prior report by Snoeck and colleagues (Huang et al., 2014), withdrawal of either KGF or FGF10 did not adversely impact the efficiency of specification of NKX2-1^{+} population by day 15 of differentiation (RUES2 or C17 PSC lines; Fig 7B, Figs S3, S4). However, withdrawal of CHIR produced virtually no NKX2-1^{+} cells (<1%) and withdrawal of BMP4 or RA had variable effects on the percentage of NKX2-1^{+}, depending on the ESC or iPSC clone tested (Fig. 7B and data not shown). These findings are consistent with our mouse ESC model, as our base medium in all mouse ESC experiments contains RA. They are also in keeping with recent work by Zorn and colleagues demonstrating that RA is required to pattern foregut endoderm to become Wnt and BMP responsive, and therefore lung competent, in *Xenopus*, mouse and human development (Rankin et al., 2016). We found Wnt3a could not substitute for CHIR in our human PSC system, consistent with the previously reported low-level response of human cells to *in vitro* treatment with recombinant Wnt (Fuerer and Nusse, 2010).

We next tested whether a minimal three-factor putative 'lung' medium containing only CHIR, BMP4 and RA (CBRA) versus a modified 'thyroid' protocol involving medium containing FGF2 +BMP4 but no CHIR could generate NKX2-1^{+} cells when added to human PSC-derived foregut-staged cells (Fig. S4D, Fig. S5). We observed that each specification medium (CBRA versus FGF2 +BMP4) generated NKX2-1^{+} cells by day 15 of differentiation (Fig. 7C-H, Fig. S5), but lung versus thyroid competence was distinct in each condition. For example, we observed the three-factor CBRA media on average induced 52% of cells (range 24-95%) to express NKX2-1^{GFP} by day 16 (Fig. 7C,D; equivalent to a yield of 7.2±2.2 NKX2-1^{+} cells per day 0 'input' iPSC); these sorted GFP^{+} cells expressed *FOXA2* and *SHH*, and exhibited the competence, upon replating in 3D matrigel cultures, to give rise to 3D epithelial spheres expressing the distal alveolar epithelial markers *SFTPB* and *SFTPC*, and lamellar bodies (Fig. 7F-H), without any detectable expression of thyroid marker genes. This experiment was repeated in four iPSC lines (C17, BU3, RC204 and T4; Fig. 7D) and one ESC line (RUES2) with varying NKX2-1 differentiation efficiencies but similar results, indicating lung but not thyroid competence in response to CBRA (Fig. 7H). Comparing sorted BU3 NKX2-1^{GFP+} progenitors induced head to head with either CFKBRA versus CBRA, we found no difference in their competence to subsequently upregulate *SFTPC* by day 31 of differentiation (Fig. S4D). In marked contrast, BU3 NKX2-1^{+} progenitors differentiated by day 16 in response to FGF2+BMP4 expressed low levels of *SHH* on day 16 and no detectable competence to upregulate *SFTPC*, but did express *PAX8* and did display competence to upregulate thyroid-selective markers such as *TSHR* (Fig. 7H; Kurmann et al., 2015; Hawkins et al., 2017). These findings are consistent with our previously published observations that markers of thyroid lineage (*PAX8*, *TG*, *NIS* or *TPO*; studied in RUES2, C17 and BU3 lines) are expressed in outgrowths of human PSC-derived foregut endoderm after exposure to FGF2+BMP4 (Kurmann et al., 2015).

We considered the possibility that CHIR, which stimulates canonical Wnt signaling by stabilizing β-catenin via GSK3β inhibition, can potentially have off-target effects. Hence, we sought to further test the hypothesis that canonical Wnt signaling via activated β-catenin is required for human lung lineage specification in our human PSC endoderm model system. To block β-catenin interactions with potential effector-binding partners, we repeated our human lung differentiation protocol

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

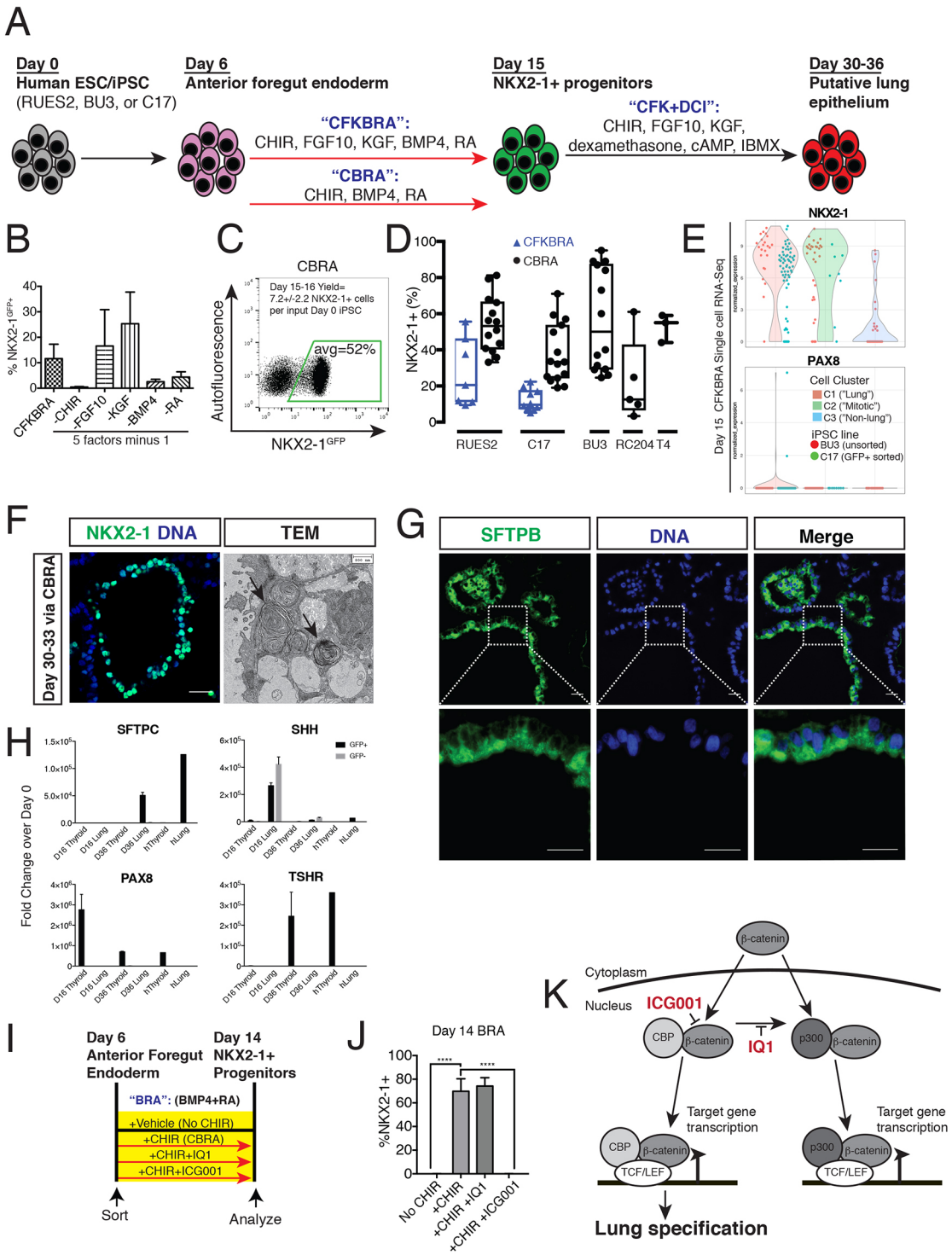


Fig. 7. See next page for legend.

using CHIR, BMP4 and RA to differentiate RUES2 cells during the putative lung specification stage, but additionally included inhibitors of β-catenin-mediated interactions with either p300 (IQ1) or CBP (CREB-binding protein; ICG001), two distinct downstream effector branches of canonical Wnt signaling (Eguchi et al., 2005; Moheimani et al., 2015; Sasaki and Kahn, 2014) (Fig. 7I-K). Although cells treated with CHIR during lung endoderm specification showed induction of NKX2-1 in more than 60% of cells, addition of the ICG001 inhibitor significantly blocked induction of NKX2-1, whereas treatment with IQ1 did not. Similar results were also observed in C17 iPSCs (data not shown). These results suggest that CHIR in the presence of BMP4 induces lung lineage specification in the human model system via canonical Wnt signaling that may be transduced, in part, through the interaction of β-catenin with CBP (Fig. 7K). Thus, our results indicate that, as in our mouse explant and mouse PSC model systems, Wnt and BMP signaling promote lung-lineage specification from human endoderm.

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

Fig. 7. Conserved pathways induce lung cell fate in human ESC/iPSC-derived endoderm. (A) Schematic of directed differentiation protocol for human ESCs or iPSCs, comparing various specification media (day 6-15). (B) Comparison of lineage specification (NKX2-1+ percentage) induced by the five growth factor standard cocktail (CFKBRA) versus conditions with one factor removed ('5 factors minus 1'): percentage of NKX2-1$^+$ cells induced in each condition is shown as quantified by flow cytometry on day 15 using the C17 iPSC line carrying a GFP reporter targeted to the NKX2-1 locus. Data are mean±s.d. of biological triplicates. (C) Day 15 NKX2-1 induction efficiency in human iPSCs. Representative flow cytometry dot plot showing NKX2-1^{GFP} reporter expression on day 15, along with average yield±s.d. calculated for the BU3 iPSC line. (D) Box and whiskers plot showing the range, median and quartiles of NKX2-1 induction efficiencies (% of all cells measured by FACS) for each indicated human ESC or iPSC clone after induction with either CFKBRA (blue triangles) or CBRA (black dots). Data represent day 13-16 analyses accumulated over a ~1 year period of experiments. (E) Violin plots of normalized gene expression measured for each indicated gene by single-cell RNA-Seq of 153 cells on day 15 of differentiation in CFKBRA. See also Fig. S4. (F) Confocal microscopy on day 33 showing NKX2-1 nuclear protein immunostaining of candidate lung progenitors derived with CBRA on days 6-15 followed by 3D culture outgrowth in conditions shown in A, and transmission electron microscopy (TEM) on day 30, indicating lamellar bodies (arrows). Scale bar: 25 µm. (G) Immunofluorescence microscopy of SFTPB protein expression in spherical epithelial cells derived in 3D Matrigel after specification in CBRA followed by differentiation in 3D Matrigel, according to the protocol in A. Scale bars: 25 µm. See also Fig. S3. (H) RT-qPCR showing fold-change in gene expression over day 0 ($2^{-\Delta\Delta Ct}$) of each indicated lung or thyroid marker gene in progenitors (day 16) or maturing cells (day 36) derived from BU3 iPSCs using 'lung' or 'thyroid' differentiation media, compared with control human fetal lung epithelium (hLung; 21 weeks gestation) or adult human thyroid tissue (hThyroid). Data are mean±s.d. (I) Experimental design for testing the effect of CHIR with/without chemical inhibition of β-catenin co-activator function (canonical Wnt signaling) on human lung progenitor specification. Inhibitors IQ1 or ICG001 were added to CBRA to block β-catenin interactions with either p300 or CBP transcription factors, respectively. (J) Day 14 NKX2-1 induction efficiency in RUES2 cells exposed to the conditions shown in I from days 6-14. Intracellular staining for NKX2-1 protein was quantified by FACS. Data are mean percentage of NKX2-1$^+$ cells±s.d. in biological triplicates. ****P<0.05; one-way ANOVA. (K) Schematic of the mechanism of action of IQ1 and ICG001 chemical inhibitors on canonical Wnt signaling.

DISCUSSION

In contrast to *in vivo* organogenesis, where differentiation proceeds in a highly orchestrated process, *in vitro*-directed differentiation of PSCs is often fraught with varying levels of contamination with undesired lineages and heterogeneity. Sorting strategies designed to isolate PSC-derived tissue-committed progenitor subsets can help to decrease this heterogeneity (Holtzinger et al., 2015). In the past, we have used *Nkx2-1* reporters to help isolate lung and thyroid lineages derived from PSCs in culture (Longmire et al., 2012; Kurmann et al., 2015); the lack of complete specificity of endodermal Nkx2-1, however, has resulted in the derivation of mixed populations of lung and thyroid lineages, an expected result given the known *in vivo* expression of Nkx2-1 at the time of lineage specification of these two tissues. In previous reports where human PSCs were differentiated into lung lineages without contaminating thyroid cells, it was unclear whether this was due to inherent differences in the responses of human cells compared with mouse cells or to subtle differences in the complex growth factor-supplemented media being used (Huang et al., 2014; Dye et al., 2015). Here we have used *in vitro* PSC model systems to dissect the minimal pathways required for lung- versus thyroid-lineage specification, in order to resolve these controversies and to provide progenitors capable of giving rise to functional epithelial spheres that are not hampered by mixed thyroid and lung populations. Our results demonstrate that surprisingly simple media supplemented with only two or three growth factors are required to produce each distinct lineage, and the signaling pathways that regulate their lineage specification is evolutionarily conserved across species from mice to humans.

Our results demonstrate that, in foregut endoderm, either in embryos or in PSC-based models, the combination of Wnt+BMP signaling (in the presence of RA) promotes lung- rather than thyroid-lineage specification, whereas the combination of FGF +BMP signaling promotes thyroid specification. RA, which is included in our base medium throughout endodermal and lung-lineage specification, is also required for successful lung NKX2-1$^+$ specification in our human model. However, as recently published, RA signaling appears to be dispensable at the moment of lung specification and is instead required at an earlier developmental stage during foregut endodermal patterning, in order to render the foregut competent to subsequently respond to Wnt and BMP signals (Rankin et al., 2016). Hence, we do not revisit this recently established role for RA signaling in lung development.

We were surprised to find that lung-lineage specification in our models did not require the addition of any FGFs. Indeed inhibition of FGF signaling did not appear to dampen lung-lineage specification in mouse foregut endoderm, implying that FGF signaling is dispensable at this developmental stage, although the inhibitor experiments were not repeated for the human ESC/iPSC model system. How can we reconcile our results with the previously published observation of Serls et al. (Serls et al., 2005) that FGF1 or FGF2 added to mouse foregut endoderm cultures appears to induce Nkx2-1 as well as Sftpc, whereas in our work FGF2 appears to inhibit lung specification in favor of thyroid specification? First, it should be emphasized that both Nkx2-1 and thyroglobulin were found to be induced by FGFs added to mouse endoderm in Serls et al.'s publication, raising the possibility that thyroid lineage specification was occurring, consistent with our results. In the years that have passed since the 2005 Serls et al. report, it has been established that canonical Wnt signaling (induced *in vivo* by Wnt2/2b secreted from lateral plate mesoderm) is required for lung-lineage specification (Goss et al., 2009). As Serls et al. only added FGFs to endoderm cultures that showed induction of Nkx2-1 and Sftpc, it appears likely that some mesodermal cells were likely present in their model. Contaminating mesenchymal cells might be able to respond to FGFs by secreting factors required for lung-lineage specification, such as Wnt. Our results suggesting that FGF signaling is dispensable for lung-lineage specification from foregut endoderm when Wnt signaling is present fit other published observations, such as: (1) genetic deletion of FGF ligands or FGF receptors in mice has not been shown to completely abrogate respiratory field specification (Weinstein et al., 1998; Zhou et al., 1998; De et al., 2000; Guzy et al., 2015); and (2) exogenous addition of FGF ligands to developing foregut *Xenopus* or mouse embryos is not required for lung-lineage specification (Rankin et al., 2016). FGF7 and FGF10, when added to stimulants of Wnt and BMP signaling in our model, did not significantly dampen lung-lineage specification or induce thyroid-lineage specification. This further emphasizes the different receptors that the various FGF ligands are known to bind to induce signaling, with broad ligation of FGF receptors occurring in response to FGF2, whereas FGF7 and FGF10 more specifically ligate the FGFR2IIIb form of the receptor.

Our finding that canonical Wnt signaling and BMP signaling are required for lung-lineage specification is consistent with genetic mouse models where deletion of Wnt 2/2b (Goss et al., 2009) or β-catenin (Harris-Johnson et al., 2009) results in lung agenesis, and deletion of BMP receptors in late foregut endoderm markedly reduces the Nkx2-1$^+$ respiratory field (Domyan et al., 2011). Our mouse explant model where BMP signaling could be inhibited at an

Development (2017) 144, 3879-3893 doi:10.1242/dev.150193

earlier endodermal developmental stage suggests that early BMP signaling is absolutely required for lung specification and is consistent with recently published findings in the *Xenopus* endodermal model system (Rankin et al., 2016).

Most importantly, our results suggest that the minimal signaling pathways that regulate cell fate decisions in lung and thyroid development are evolutionarily conserved between diverse species from mice to humans. Use of the mouse PSC model system, in concert with prior findings in *Xenopus*, was able to delineate the minimal signaling pathways (Wnt and BMP) required for lung specification; these pathways also appear to regulate human specification in the PSC model system. Although perturbations of human foregut *in vivo* at the time of lineage specification (3-4 weeks gestation) are not possible, our results imply that these same pathways may also be active in humans *in vivo* and serve as a strong argument in support of employing multi-species model systems, including the *in vitro* mouse PSC model for the discovery of human developmental mechanisms.

Our results do not necessarily indicate that lung and thyroid lineages are specified from the same foregut endodermal precursors. We cannot exclude the possibility that the various ligands being studied selectively enhance survival or proliferation of each distinct lung or thyroid progenitor pool rather than solely regulating their lineage specification. Single-cell profiling, limiting-dilution outgrowth cultures or clonal lineage-tracing studies in the future may help to define the heterogeneity of the Nkx2-1^{+} progenitor pool as well as the differentiation repertoire of individual foregut cells.

Overall, our results help to establish the minimal evolutionarily conserved pathways required for generating lung or thyroid progenitors from mouse or human PSCs, and thus provide precise control in regulating cell fate decisions of PSCs in cultures. These findings help to resolve long-standing controversies in the field of lung development and should now facilitate more reproducible derivation of an inexhaustible source of progenitors for basic developmental studies, lung or thyroid disease modeling, and testing of regenerative drug- or cell-based therapies.

MATERIALS AND METHODS

Mouse ESC directed differentiation

The Nkx2-1^{mCherry} ESC line was generated and employed for *in vitro* differentiation as previously described (Bilodeau et al., 2014; Kurmann et al., 2015) (see supplementary Materials and Methods). Definitive endoderm induction was performed for 5 days to generate embryoid bodies (EBs) in serum-free medium as previously published with 50 ng/ml activin A (R&D Systems, 338-AC) added to the base medium from days 2.5 to 5 (Longmire et al., 2012; Kurmann et al., 2015). 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 rmNoggin (R&D Systems, 1967-NG) and 10 µM SB431542 (Sigma, S4317), 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 stated in the text and detailed in the supplementary Materials and Methods. Nkx2-1$^{mCherry+}$ cells were sorted and replated onto gelatin-coated 24-well plates on day 12-14 at a density of 5×10^{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 and detailed in the supplementary Materials and Methods.

Human ESC and iPSC maintenance and differentiation

The RUES2 human embryonic stem cell line (a gift from Dr Ali H. Brivanlou, The Rockefeller University, New York, USA) and previously published human iPSC lines [BU3 (Kurmann et al., 2015) or C17 (Crane et al., 2015)] were maintained and differentiated as detailed in the supplementary Materials and Methods.

Lentiviral transduction of mESCs and immunofluorescence microscopy

The previously published SftpcdsRed lentiviral reporter vector (Longmire et al., 2012) was recloned as SftpcGFP as detailed in the supplementary Materials and Methods. Three-dimensional organoids were harvested by incubating with Cell Recovery Solution (Corning, 354253) for 1 h at 4°C, fixed with fresh 4% paraformaldehyde for 30 min at room temperature, and embedded in either 2% low melting gel agarose (Lonza SeaKem LE agarose, 50.000) or Richard-Allan Scientific HistoGel Specimen Processing Gel (ThermoFisher Scientific, HG-4000-012) according to the manufacturer's instructions, prior to embedding in paraffin. Sections were deparaffinized, rehydrated and stained for Hematoxylin and Eosin (AML Labs) using standard methods. For immunofluorescent staining, sections were incubated in antigen retrieval solution (Dako, S-1699) at 95°C for 20 min, cooled in same solution to room temperature for 30 min, and permeabilized/blocked with 0.25% Triton X-100 (Sigma, T-8787) and 4% normal donkey serum (NDS, Sigma, D9663) for 1 h at room temperature. Sections were incubated with primary antibodies in 4% NDS overnight at 4°C, washed and incubated with the corresponding Alexa Fluorophore-conjugated secondary antibodies for 30-60 min at room temperature. Nuclei were counterstained with DAPI (Invitrogen, 1:10,000) or Hoechst dye (ThermoFisher, 1:500). Fluorescent images were captured on a Nikon Eclipse Ni upright microscope using Nikon Elements D4.00 software or a Zeiss confocal microscope. Antibody information and sources, as well as methods of immunostaining for flow cytometry are detailed in the supplementary Materials and Methods.

Mouse foregut explants, whole-embryo cultures, or mouse 'recombinant' cultures

All studies involving mice were approved by the Institutional Animal Care and Use Committee of Cincinnati Children's Hospital. Foregut explant or whole-embryo cultures prepared from developing mice were all performed as we have previously published (Kurmann et al., 2015) and are detailed in the supplementary Materials and Methods.

Reverse transcriptase real-time quantitative PCR (RT-qPCR)

After reverse transcription, 40 cycles of PCR were performed to quantify gene expression, calculated based on 18S normalized expression levels represented as fold-change mRNA expression [$2^{-(\Delta\Delta Ct)}$] compared with undifferentiated (day 0) PSCs. Undetected genes after 40 cycles of PCR were arbitrarily assigned a Ct value of 40 to allow quantitative calculation of fold-change. Further details and primers used for all PCR reactions are provided in the supplementary Materials and Methods.

Microarray analysis

Three biological replicates for each indicated condition were processed on day 14 of differentiation for sorting of mESC-derived cells based on expression of the Nkx2-1^{mCherry} reporter by flow cytometry. RNA extracts from each sample were processed for analysis of gene expression by Affymetrix GeneChip Mouse Gene 2.0 ST arrays. Detailed analyses and statistical methods for normalization, principle component analysis, identification of differentially expressed genes by ANOVA, and unsupervised hierarchical clustering can be found in the supplementary Materials and Methods and Table S1, which lists all genes, samples, calculated fold-changes in expression and FDR-adjusted *P*-values. Raw data .cel files can be downloaded from the Gene Expression Omnibus database (accession number GSE92916).

Single-cell RNA-sequencing analysis

To screen for expression of thyroid markers in individual day 15 cells prepared without the use of FGF2, we re-analyzed our previously published dataset (Hawkins et al., 2017) of unsorted BU3 cells and NKX2-1^{GFP+} sorted C17 cells as follows: single-cell sequencing reads from 178 cells were aligned to the human genome (GRCh38) and quantified using STAR (Dobin

et al., 2013). From these, 25 cells were discarded owing to having either abnormally high mitochondrial gene counts or abnormally low aligned reads, leaving 153 cells for further analysis (82 BU3, 71 C17). Statistical methods applied for normalization, clustering and heatmap generation are detailed in the supplementary Materials and Methods.

Statistical analysis

Data are presented as sample means and standard deviations (s.d.) with sample numbers stated specifically within the text or figure legends. Differences between groups were analyzed using unpaired two-tailed Student's *t*-test, one-way or two-way analysis of variance (ANOVA) with Tukey's multiple-comparison post hoc test as stated in the text or figure legends; $P<0.05$ was used to indicate significant differences between groups.

Acknowledgements

We thank the members of the Kotton lab and KIWI group for insightful discussions. We thank Adam Gower and Eduard 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 and Dylan C. Thomas of the CReM for laboratory technical support. We thank Brian R. Tilton of the Boston University Flow Cytometry Core Facility for technical assistance.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: M.S., J.M.S., D.N.K.; Methodology: M.S., K.-D.A., F.H., K.B.M., A.J., J.C., I.S.C., M.V., A.A.K., L.I., A.N.H., J.M.S., D.N.K.; Software: M.S., K.-D.A., F.H., K.B.M., A.J., J.C., I.S.C., M.V., A.A.K., L.I., A.N.H., J.M.S., D.N.K.; Validation: M.S., K.-D.A., F.H., K.B.M., A.J., J.C., I.S.C., M.V., A.A.K., L.I., A.N.H., J.M.S., D.N.K.; Formal analysis: M.S., K.-D.A., F.H., K.B.M., A.J., J.C., I.S.C., M.V., A.A.K., L.I., A.N.H., J.M.S., D.N.K.; Investigation: M.S., K.-D.A., F.H., K.B.M., A.J., J.C., I.S.C., M.V., A.A.K., L.I., A.N.H., J.M.S., D.N.K.; Resources: L.I., A.N.H., J.M.S., D.N.K.; Data curation: M.S., K.-D.A., F.H., K.B.M., A.J., J.C., I.S.C., M.V., A.A.K., L.I., J.M.S., D.N.K.; Writing - original draft: M.S., J.M.S., D.N.K.; Writing - review & editing: M.S., K.-D.A., F.H., K.B.M., A.J., J.C., I.S.C., M.V., L.I., A.N.H., J.M.S., D.N.K.; Visualization: J.M.S., D.N.K.; Supervision: J.M.S., D.N.K.; Project administration: J.M.S., D.N.K.; Funding acquisition: J.M.S., D.N.K.

Funding

F.H. was supported by a grant from the Cystic Fibrosis Foundation (HAWKIN15XX0). This work was funded by the National Institutes of Health (R01DK105029 to A.N.H.; R01HL111574 to L.I.; R01HL098319 and U01HL122642 to J.M.S.; R01HL095993, R01HL108678, R01HL122442 and R01HL128172 to D.N.K.). Deposited in PMC for release after 12 months.

Data availability

Raw data .cel files can be downloaded from the Gene Expression Omnibus database (accession number GSE92916).

Supplementary information

Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.150193.supplemental

References

Ahnfelt-Ronne, J., Jørgensen, M. C., Hald, J., Madsen, O. D., Serup, P. and Hecksher-Sørensen, J. (2007). An improved method for three-dimensional reconstruction of protein expression patterns in intact mouse and chicken embryos and organs. *J. Histochem. Cytochem.* **55**, 925-930.

Bilodeau, M., Shojaie, S., Ackerley, C., Post, M. and Rossant, J. (2014). Identification of a proximal progenitor population from murine fetal lungs with clonogenic and multilineage differentiation potential. *Stem Cell Rep.* **3**, 634-649.

Crane, A. M., Kramer, P., Bui, J. H., Chung, W. J., Li, X. S., Gonzalez-Garay, M. L., Hawkins, F., Liao, W., Mora, D., Choi, S. et al. (2015). Targeted correction and restored function of the CFTR gene in cystic fibrosis induced pluripotent stem cells. *Stem Cell Rep.* **4**, 569-577.

Dame, K., Cincotta, S., Lang, A. H., Sanghrajka, R. M., Zhang, L., Choi, J., Kwok, L., Wilson, T., Kańdula, M. M., Monti, S. et al. (2017). Thyroid progenitors are robustly derived from embryonic stem cells through transient, developmental stage-specific overexpression of Nkx2-1. *Stem Cell Rep.* **8**, 216-225.

De, M. L., Spencer-Dene, B., Revest, J., Hajihosseini, M., Rosewell, I. and Dickson, C. (2000). An important role for the IIIb isoform of fibroblast growth factor receptor 2 (FGFR2) in mesenchymal-epithelial signalling during mouse organogenesis. *Development* **127**, 483-492.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21.

Domyan, E. T., Ferretti, E., Throckmorton, K., Mishina, Y., Nicolis, S. K. and Sun, X. (2011). Signaling through BMP receptors promotes respiratory identity in the foregut via repression of Sox2. *Development* **138**, 971-981.

Dye, B. R., Hill, D. R., Ferguson, M. A., Tsai, Y. H., Nagy, M. S., Dyal, R., Wells, J. M., Mayhew, C. N., Nattiv, R., Klein, O. D. et al. (2015). In vitro generation of human pluripotent stem cell derived lung organoids. *Elife* **4**, e05098.

Eguchi, M., Nguyen, C., Lee, S. C. and Kahn, M. (2005). ICG-001, a novel small molecule regulator of TCF/beta-catenin transcription. *Med. Chem.* **1**, 467-472.

Fagman, H., Amendola, E., Parrillo, L., Zoppoli, P., Marotta, P., Scarfò, M., De Luca, P., de Carvalho, D. P., Ceccarelli, M., De Felice, M. et al. (2011). Gene expression profiling at early organogenesis reveals both common and diverse mechanisms in foregut patterning. *Dev. Biol.* **359**, 163-175.

Fuerer, C. and Nusse, R. (2010). Lentiviral vectors to probe and manipulate the Wnt signaling pathway. *PLoS ONE* **5**, e9370.

Glasser, S. W., Korfhagen, T. R., Wert, S. E., Bruno, M. D., McWilliams, K. M., Vorbroker, D. K. and Whitsett, J. A. (1991). Genetic element from human surfactant protein SP-C gene confers bronchiolar-alveolar cell specificity in transgenic mice. *Am. J. Physiol.* **261**, L349-L356.

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

Gouon-Evans, V., Boussemart, L., Gadue, P., Nierhoff, D., Koehler, C. I., Kubo, A., Shafritz, D. A. and Keller, G. (2006). BMP-4 is required for hepatic specification of mouse embryonic stem cell-derived definitive endoderm. *Nat. Biotechnol.* **24**, 1402-1411.

Grasberger, H., De Deken, X., Mayo, O. B., Raad, H., Weiss, M., Liao, X.-H. and Refetoff, S. (2012). Mice deficient in dual oxidase maturation factors are severely hypothyroid. *Mol. Endocrinol.* **26**, 481-492.

Green, M. D., Chen, A., Nostro, M.-C., d'Souza, S. L., Schaniel, C., Lemischka, I. R., Gouon-Evans, V., Keller, G. and Snoeck, H. W. (2011). Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 267-272.

Grindley, J. C., Bellusci, S., Perkins, D. and Hogan, B. L. M. (1997). Evidence for the involvement of the Gli gene family in embryonic mouse lung development. *Dev. Biol.* **188**, 337-348.

Guzy, R. D., Stoilov, I., Elton, T. J., Mecham, R. P. and Ornitz, D. M. (2015). Fibroblast growth factor 2 is required for epithelial recovery, but not for pulmonary fibrosis, in response to bleomycin. *Am. J. Respir. Cell Mol. Biol.* **52**, 116-128.

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

Hawkins, F. and Kotton, D. N. (2015). Embryonic and induced pluripotent stem cells for lung regeneration. *Ann. Am. Thorac. Soc.* **12** Suppl 1, S50-S53.

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

Herriges, M. J., Swarr, D. T., Morley, M. P., Rathi, K. S., Peng, T., Stewart, K. M. and Morrisey, E. E. (2014). Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. *Genes Dev.* **28**, 1363-1379.

Holtzinger, A., Streeter, P. R., Sarangi, F., Hillborn, S., Niapour, M., Ogawa, S. and Keller, G. (2015). New markers for tracking endoderm induction and hepatocyte differentiation from human pluripotent stem cells. *Development* **142**, 4253-4265.

Huang, S. X., Islam, M. N., O'Neill, J., Hu, Z., Yang, Y. G., Chen, Y. W., Mumau, M., Green, M. D., Vunjak-Novakovic, G., Bhattacharya, J. et al. (2014). Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 84-91.

Hyatt, B. A., Shangguan, X. and Shannon, J. M. (2004). FGF-10 induces SP-C and Bmp4 and regulates proximal-distal patterning in embryonic tracheal epithelium. *Am. J. Physiol. Lung Cell Mol. Physiol.* **287**, L1116-L1126.

Kurmann, A. A., Serra, M., Hawkins, F., Rankin, S. A., Mori, M., Astapova, I., Ullas, S., Lin, S., Bilodeau, M., Rossant, J. et al. (2015). Regeneration of thyroid function by transplantation of differentiated pluripotent stem cells. *Cell Stem Cell* **17**, 527-542.

Lancaster, M. A. and Knoblich, J. A. (2014). Organogenesis in a dish: modeling development and disease using organoid technologies. *Science* **345**, 1247125.

Lee, J.-H., Bhang, D. H., Beede, A., Huang, T. L., Stripp, B. R., Bloch, K. D., Wagers, A. J., Tseng, Y.-H., Ryeom, S. and Kim, C. F. (2014). Lung stem cell differentiation in mice directed by endothelial cells via a BMP4-NFATc1-thrombospondin-1 axis. *Cell* **156**, 440-455.

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

Lun, A. T., Bach, K. and Marioni, J. C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75.

McCarthy, D. J., Campbell, K. R., Lun, A. T. and Wills, Q. F. (2017). Scater: pre-processing, quality control, nomalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186.

Moheimani, F., Roth, H. M., Cross, J., Reid, A. T., Shaheen, F., Warner, S. M., Hirota, J. A., Kicic, A., Hallstrand, T. S., Kahn, M. et al. (2015). Disruption of beta-catenin/CBP signaling inhibits human airway epithelial-mesenchymal transition and repair. *Int. J. Biochem. Cell Biol.* **68**, 59-69.

Motoyama, J., Liu, J., Mo, R., Ding, Q., Post, M. and Hui, C. C. (1998). Essential function of Gli2 and Gli3 in the formation of lung, trachea and oesophagus. *Nat. Genet.* **20**, 54-57.

Mou, H., Zhao, R., Sherwood, R., Ahfeldt, T., Lapey, A., Wain, J., Sicilian, L., Izvolsky, K., Musunuru, K., Cowan, C. et al. (2012). Generation of multipotent lung and airway progenitors from mouse ESCs and patient-specific cystic fibrosis iPSCs. *Cell Stem Cell* **10**, 385-397.

Murtagh, F. and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *J. Classif.* **31**, 274-295.

Rankin, S. A., Han, L., McCracken, K. W., Kenny, A. P., Anglin, C. T., Grigg, E. A., Crawford, C. M., Wells, J. M., Shannon, J. M. and Zorn, A. M. (2016). A retinoic acid-hedgehog cascade coordinates mesoderm-inducing signals and endoderm competence during lung specification. *Cell Rep.* **16**, 66-78.

Risso, D., Perraudeau, F., Gribkova, S. and Dudoit, S. (2017). ZINB-WaVE: A general and flexible method for signal extraction from single-cell RNA-seq data. *bioRxiv*, 125112.

Rossi, D. L., Hurst, S. D., Xu, Y., Wang, W., Menon, S., Coffman, R. L. and Zlotnik, A. (1999). Lungkine, a novel CXC chemokine, specifically expressed by lung bronchoepithelial cells. *J. Immunol.* **162**, 5490-5497.

Roszell, B. R., Mondrinos, M. J., Seaton, A., Simons, D. M., Koutzaki, S. H., Fong, G.-H., Lelkes, P. I. and Finck, C. M. (2009). Efficient derivation of alveolar type II cells from embryonic stem cells for in vivo application. *Tissue Eng. Part A.* **15**, 3351-3365.

Sasaki, T. and Kahn, M. (2014). Inhibition of beta-catenin/p300 interaction proximalizes mouse embryonic lung epithelium. *Transl. Respir. Med.* **2**, 8.

Sato, H., Murphy, P., Giles, S., Bannigan, J., Takayasu, H. and Puri, P. (2008). Visualizing expression patterns of Shh and Foxf1 genes in the foregut and lung buds by optical projection tomography. *Pediatr. Surg. Int.* **24**, 3-11.

Schlingmann, B., Molina, S. A. and Koval, M. (2015). Claudins: Gatekeepers of lung epithelial function. *Semin. Cell Dev. Biol.* **42**, 47-57.

Schwartz, C. E. and Stevenson, R. E. (2007). The MCT8 thyroid hormone transporter and Allan-Herndon-Dudley syndrome. *Best Pract. Res. Clin. Endocrinol. Metab.* **21**, 307-321.

Serls, A. E., Doherty, S., Parvatiyar, P., Wells, J. M. and Deutsch, G. H. (2005). Different thresholds of fibroblast growth factors pattern the ventral foregut into liver and lung. *Development* **132**, 35-47.

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

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

Weaver, M., Batts, L. and Hogan, B. L. M. (2003). Tissue interactions pattern the mesenchyme of the embryonic mouse lung. *Dev. Biol.* **258**, 169-184.

Weinstein, M., Xu, X., Ohyama, K. and Deng, C. X. (1998). FGFR-3 and FGFR-4 function cooperatively to direct alveogenesis in the murine lung. *Development* **125**, 3615-3623.

Wilson, A. A., Murphy, G. J., Hamakawa, H., Kwok, L. W., Srinivasan, S., Hovav, A. H., Mulligan, R. C., Amar, S., Suki, B. and Kotton, D. N. (2010). Amelioration of emphysema in mice through lentiviral transduction of long-lived pulmonary alveolar macrophages. *J. Clin. Invest.* **120**, 379-389.

Wong, A. P., Bear, C. E., Chin, S., Pasceri, P., Thompson, T. O., Huan, L.-J., Ratjen, F., Ellis, J. and Rossant, J. (2012). Directed differentiation of human pluripotent stem cells into mature airway epithelia expressing functional CFTR protein. *Nat. Biotechnol.* **30**, 876-882.

Xie, W., Lynch, T. J., Liu, X., Tyler, S. R., Yu, S., Zhou, X., Luo, M., Kusner, D. M., Sun, X., Yi, Y. et al. (2014). Sox2 modulates Lef-1 expression during airway submucosal gland development. *Am. J. Physiol. Lung Cell Mol. Physiol.* **306**, L645-L660.

Xu, B., Chen, C., Chen, H., Zheng, S.-G., Bringas, P., Jr., Xu, M., Zhou, X., Chen, D., Umans, L., Zwijsen, A. et al. (2011). Smad1 and its target gene Wif1 coordinate BMP and Wnt signaling activities to regulate fetal lung development. *Development* **138**, 925-935.

Yang, M. Y., Hilton, M. B., Seaman, S., Haines, D. C., Nagashima, K., Burks, C. M., Tessarollo, L., Ivanova, P. T., Brown, H. A., Umstead, T. M. et al. (2013). Essential regulation of lung surfactant homeostasis by the orphan G protein-coupled receptor GPR116. *Cell Rep.* **3**, 1457-1464.

Zhang, W., Menke, D. B., Jiang, M., Chen, H., Warburton, D., Turcatel, G., Lu, C.-H., Xu, W., Luo, Y. and Shi, W. (2013). Spatial-temporal targeting of lung-specific mesenchyme by a Tbx4 enhancer. *BMC Biol.* **11**, 111.

Zhou, M., Sutliff, R. L., Paul, R. J., Lorenz, J. N., Hoying, J. B., Haudenschild, C. C., Yin, M., Coffin, J. D., Kong, L., Kranias, E. G. et al. (1998). Fibroblast growth factor 2 control of vascular tone. *Nat. Med.* **4**, 201-207.

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

(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-

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.

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 37ºC, 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-

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

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-*

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).

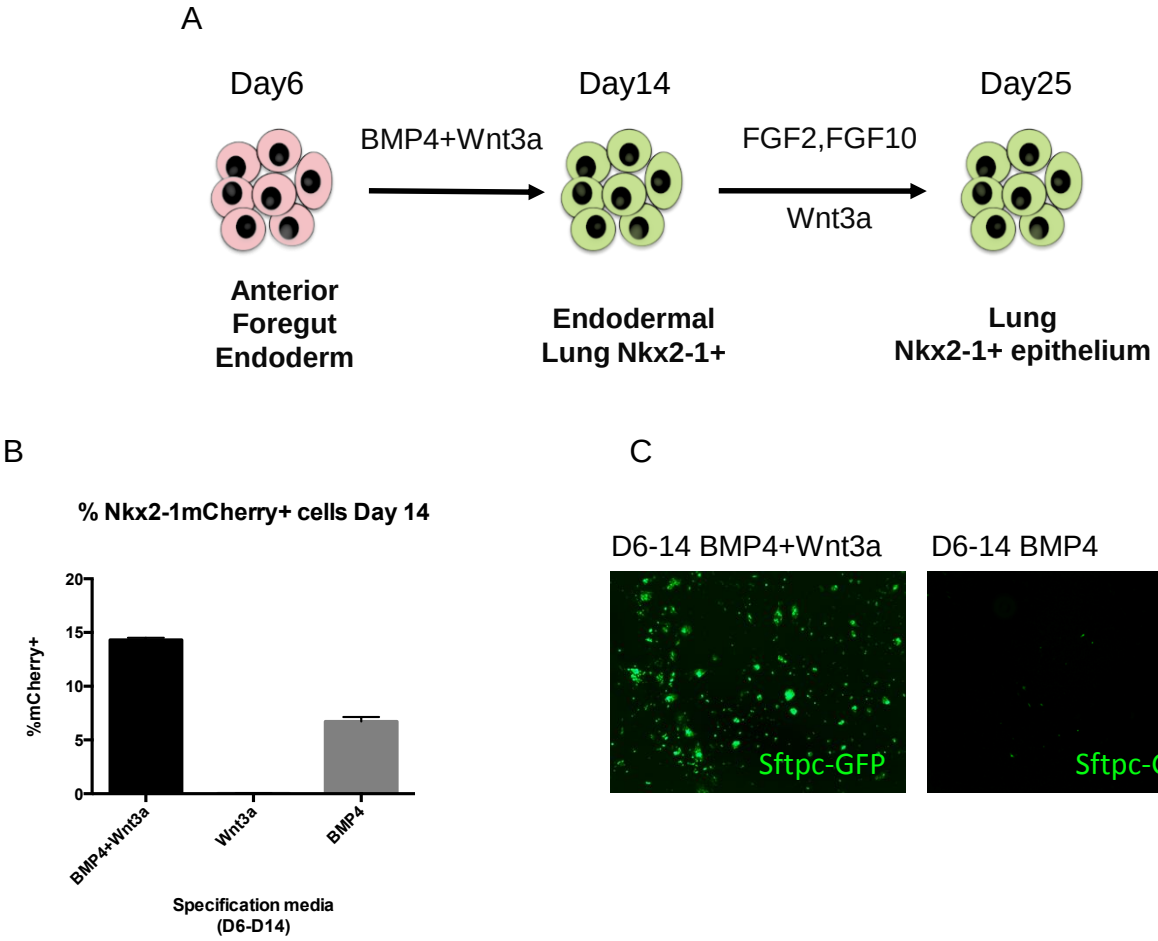
Single cell RNA-Sequencing

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.

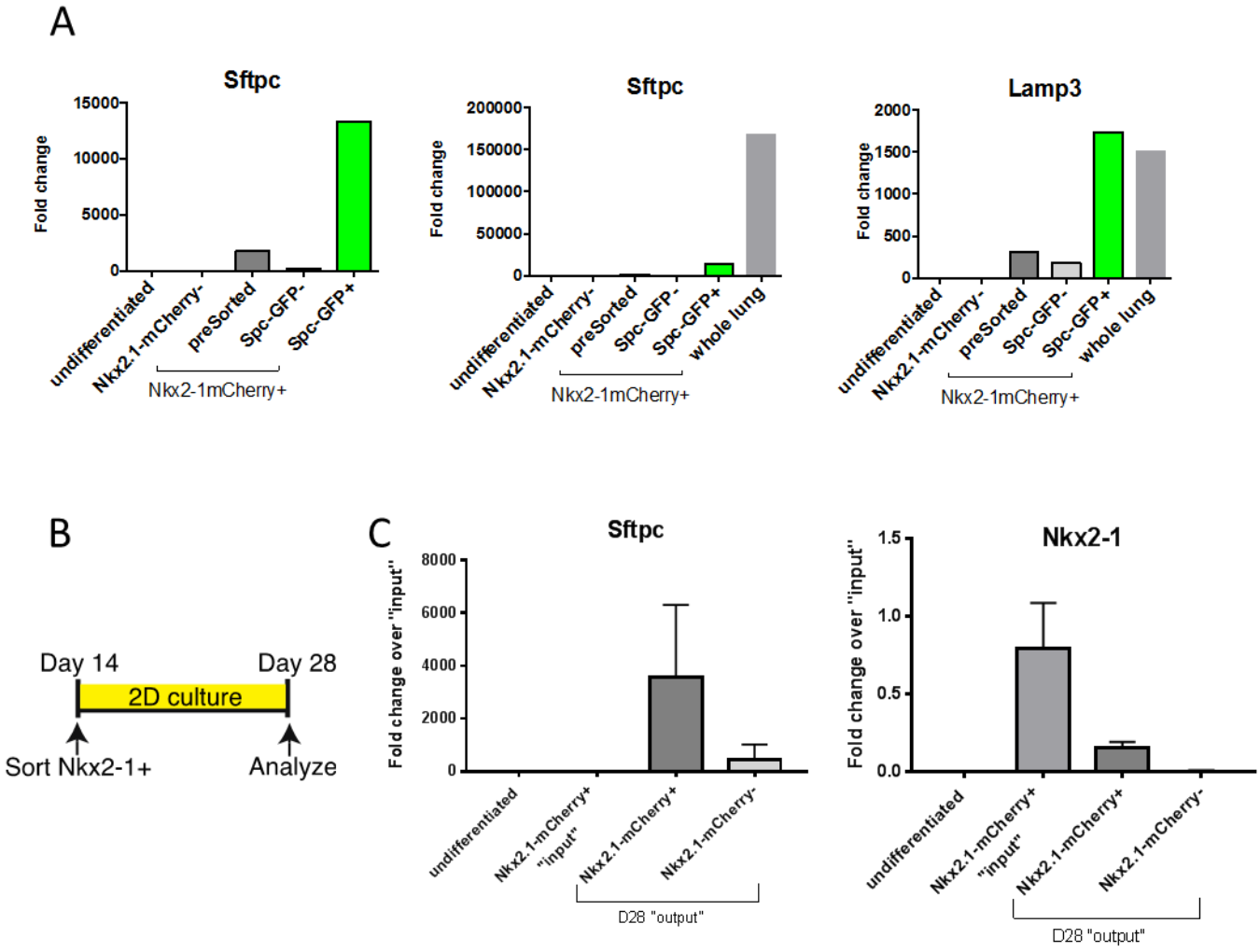
The heatmap was generated by applying Ward's hierarchical agglomerative clustering method (Murtagh and Legendre, 2014) on the row-scaled expression values of the top differentially expressed genes (absolute log2 fold-change > 1.5 and FDR-adjusted p-value < 0.05) for the pairwise contrasts across the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

Figure S1



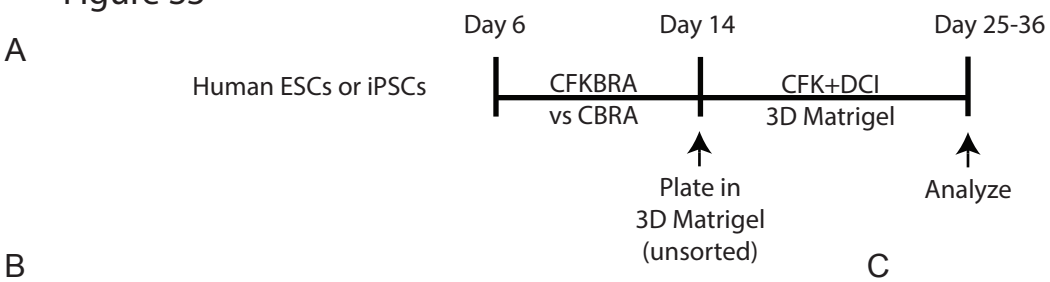
Supplemental Figure 1: Nkx2-1+ cells derived from mouse ESC-endoderm in the presence of BMP4 alone (days 6-14) do not exhibit lung competence. The mouse ESC line carrying the Nkx2-1^{mCherry} reporter was differentiated into anterior foregut-like endoderm and then exposed to either BMP4+Wnt3a or each 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.

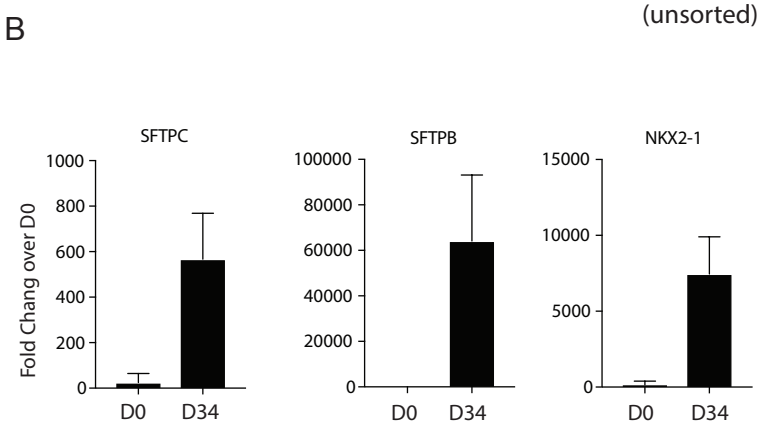
Figure S2

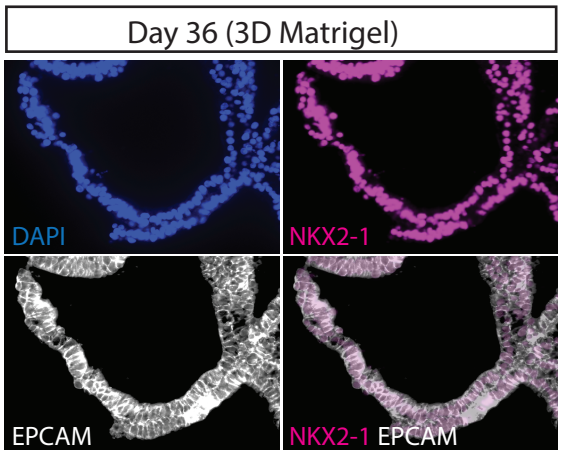


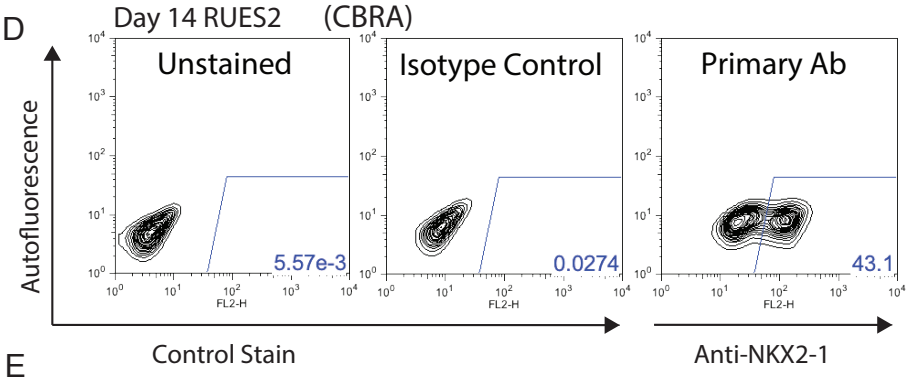
Supplemental Figure 2: Sorting Nkx2-1+ progenitors derived from mouse ESCs by day 14 identifies the entirety of distal-lung competent progenitors. (A) A subset of ESC-foregut endodermal-like cells exposed to Wnt3a+BMP4 displays induction of the Nkx2-1^{mCherry} reporter on days 6-14. Only Nkx2-1+ sorted 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").

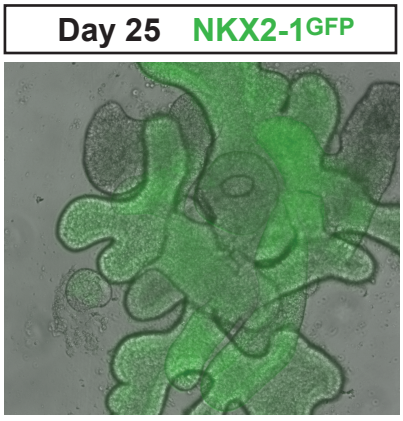
Figure S3

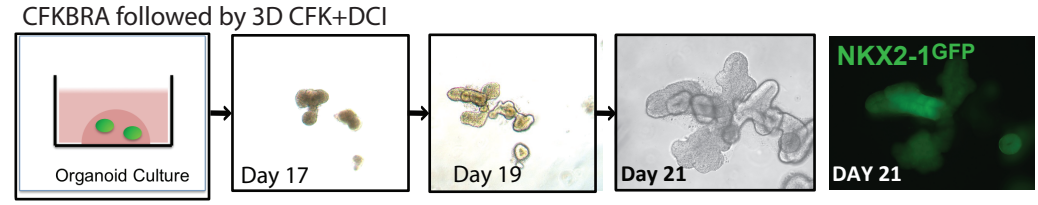












Supplemental Figure S3: Supporting data on the characterization of human pluripotent stem cell (PSC)-derived NKX2-1+ cells.

(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

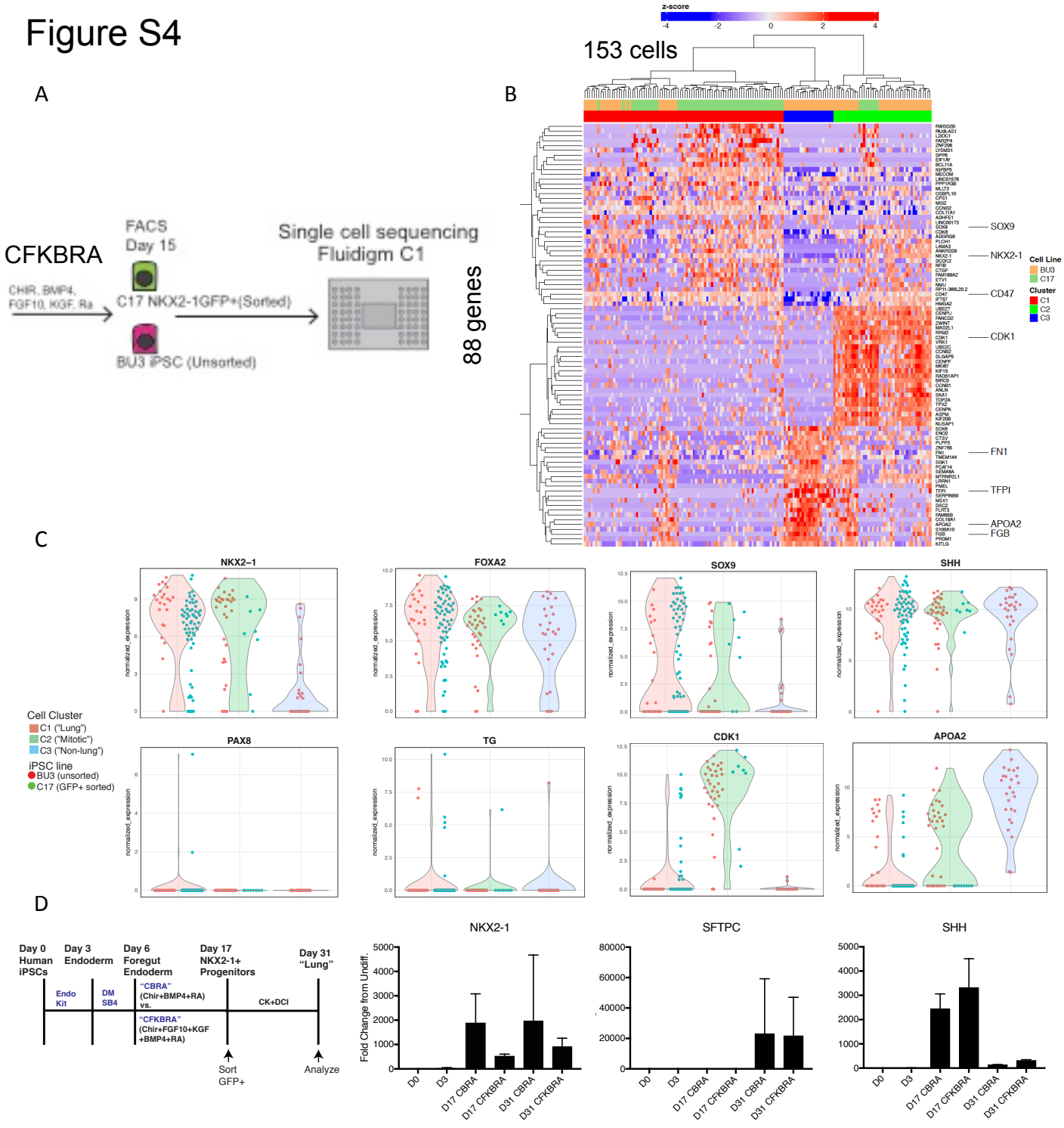


Figure S4: Human iPSC-derived NKX2-1+ endodermal progenitors generated with CFKBRA (without FGF2) are lung competent and do not express thyroid markers. (A) Experimental design outlining single cell RNA-Seq analysis profiling C17 NKX2-1^{GFP} sorted cells and BU3 unsorted cells on day 15 of differentiation 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

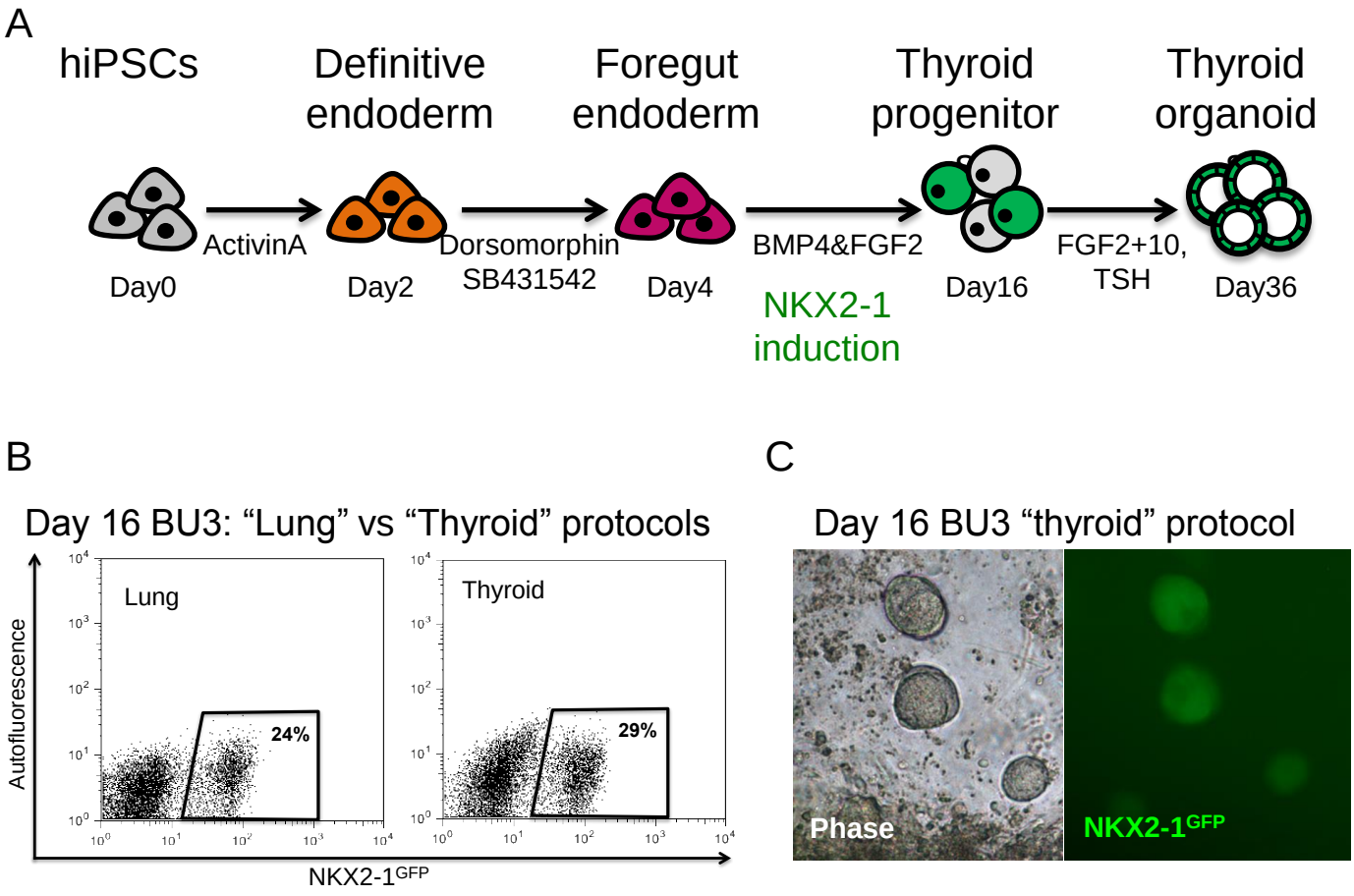
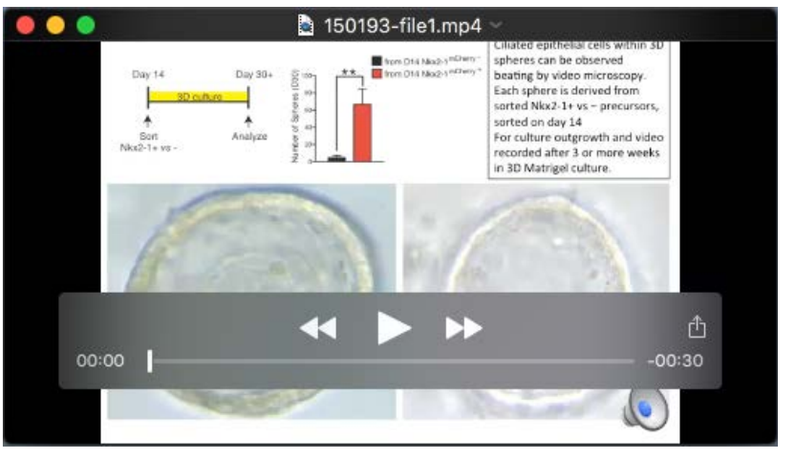


Figure S5: Human thyroid directed differentiation of iPSCs. (A) Schematic of protocol for thyroid differentiation. (B) Representative FACS dot plots indicating the day 16 efficiency of NKX2-1+ induction in BU3 iPSCs quantified based on expression of a NKX2-1^{GFP} reporter, after exposure to either the "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