

# Stable iPSC-derived NKX2-1<sup>+</sup> lung bud tip progenitor organoids give rise to airway and alveolar cell types

Renee F. C. Hein<sup>1,2,\*</sup>, Ansley S. Conchola<sup>3,\*</sup>, Alexis S. Fine<sup>4</sup>, Zhiwei Xiao<sup>4</sup>, Tristan Frum<sup>4</sup>, Lindy K. Brastrom<sup>4</sup>, Mayowa A. Akinwale<sup>1</sup>, Charlie J. Childs<sup>1,2</sup>, Yu-Hwai Tsai<sup>4</sup>, Emily M. Holloway<sup>1</sup>, Sha Huang<sup>4</sup>, John Mahoney<sup>5</sup>, Idse Heemskerk<sup>1,2</sup> and Jason R. Spence<sup>1,3,4,6,‡</sup>

## ABSTRACT

Bud tip progenitors (BTPs) in the developing lung give rise to all epithelial cell types found in the airways and alveoli. This work aimed to develop an iPSC organoid model enriched with NKX2-1<sup>+</sup> BTP-like cells. Building on previous studies, we optimized a directed differentiation paradigm to generate spheroids with more robust NKX2-1 expression. Spheroids were expanded into organoids that possessed NKX2-1<sup>+</sup>/CPM<sup>+</sup> BTP-like cells, which increased in number over time. Single cell RNA-sequencing analysis revealed a high degree of transcriptional similarity between induced BTPs (iBTPs) and *in vivo* BTPs. Using FACS, iBTPs were purified and expanded as induced bud tip progenitor organoids (iBTOs), which maintained an enriched population of bud tip progenitors. When iBTOs were directed to differentiate into airway or alveolar cell types using well-established methods, they gave rise to organoids composed of organized airway or alveolar epithelium, respectively. Collectively, iBTOs are transcriptionally and functionally similar to *in vivo* BTPs, providing an important model for studying human lung development and differentiation.

**KEY WORDS:** Lung organoid, Bud tip progenitor, Human pluripotent stem cell, Single cell RNA sequencing

## INTRODUCTION

Advances in directed differentiation methods have led to the development of numerous embryonic or induced pluripotent stem cell (iPSC)-derived cell and organoid models of the airway and alveoli, which have enhanced our ability to model human lung development and disease (Wang et al., 2007; Mou et al., 2012; Gotoh et al., 2014; Huang et al., 2014; Dye et al., 2015, 2016;

Konishi et al., 2016; Chen et al., 2017; McCauley et al., 2017; Yamamoto et al., 2017, 2020; Hawkins et al., 2017; Miller et al., 2018, 2019; Tamò et al., 2018; de Carvalho et al., 2019; Jacob et al., 2019; Leibel et al., 2020). Airway and alveolar cell types in mice and humans are derived from a common, developmentally transient progenitor population, called bud tip progenitors, which reside at the tips of the branching tree-like network of tubes that make up the lung epithelium (Serra et al., 1994; Perl et al., 2005; Abler et al., 2008; Goss et al., 2009; Rawlins et al., 2009; Rockich et al., 2013; Alanis et al., 2014; Nikolić et al., 2017; Miller et al., 2018).

Bud tip progenitors obtained from the human fetal lung and grown as organoids serve as a useful tool for studying the mechanisms responsible for bud tip progenitor cell maintenance and differentiation into airway and alveolar cell types (Nikolić et al., 2017; Miller et al., 2018, 2020; Conway et al., 2020; Sun et al., 2020). Despite this progress, organoids derived from fetal tissue are not broadly accessible to the research community and are associated with ethical and regulatory challenges, emphasizing the importance of iPSC-derived lung models. Although we and others have made some progress in developing bud tip progenitor cell-like models from iPSCs (Chen et al., 2017; Miller et al., 2018), new technologies such as single cell RNA-sequencing (scRNA-seq) and single cell lineage tracing have highlighted off-target cell types and unexpected plasticity in iPSC-derived cultures, where cells appear to be committed to a specific cell type or lineage but subsequently change fate (Little et al., 2019; Hurley et al., 2020). This concept of cellular plasticity within the lung has also been demonstrated *in vivo*, where hyperactive WNT signaling in lung progenitors was shown to cause differentiation of intestinal cells in transgenic mouse embryos (Okubo and Hogan, 2004). Therefore, a challenge in the field, addressed in the current work, is to develop a long-lived and transcriptionally stable bud tip progenitor-like model from iPSCs.

Single cell RNA-sequencing technologies have also made it possible to benchmark iPSC-derived cultures against primary tissue to compare transcriptional similarity and to accurately catalogue the diversity of on-target or off-target cell types observed *in vitro* (Hawkins et al., 2017, 2021; McCauley et al., 2018; Holloway et al., 2020; Hor et al., 2020; Yu et al., 2020). As iPSC-derived cultures are known to be plastic and iPSC differentiation is not 100% efficient, benchmarking has become an important step towards understanding the full complement of cells present in a culture. This study therefore also sought to benchmark iPSC-derived bud tip progenitor organoids to interrogate the diversity of cell types in culture and the similarity to primary bud tip progenitor organoids from the fetal lung.

Here, we have optimized an iPSC directed differentiation paradigm to generate self-organizing 3D spheroids with robust NKX2-1 expression. Expansion of NKX2-1<sup>+</sup> cells in bud tip progenitor medium over 3-17 weeks gave rise to heterogeneous

<sup>1</sup>Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA. <sup>2</sup>Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109, USA. <sup>3</sup>Program in Cell and Molecular Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA. <sup>4</sup>Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109, USA. <sup>5</sup>Therapeutics Lab, Cystic Fibrosis Foundation, Lexington, MA 02421, USA. <sup>6</sup>Department of Biomedical Engineering Medicine, University of Michigan Medical School, Ann Arbor, MI 48109, USA.

\*These authors contributed equally to this work

‡Author for correspondence (spencejr@umich.edu)

 R.F.C.H., 0000-0002-1614-2420; A.S.C., 0000-0002-2385-6712; T.F., 0000-0003-3452-8762; C.J.C., 0000-0002-4480-2625; J.M., 0000-0001-9414-9110; I.H., 0000-0002-8861-7712; J.R.S., 0000-0001-7869-3992

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Handling Editor: Matthias Lutolf

Received 28 February 2022; Accepted 28 July 2022

organoids that contained NKX2-1<sup>+</sup> bud tip progenitor-like cells co-expressing markers of human bud tip progenitors, including SOX9, SOX2 and the cell surface marker CPM (Yamamoto et al., 2020). Using an NKX2-1 reporter iPSC line along with CPM to quantitatively assess cultures via flow cytometry, we observed that bud tip progenitor-like cells expanded over subsequent weeks in culture. FACS isolation and further culture allowed for the expansion of NKX2-1<sup>+</sup>/CPM<sup>+</sup> cells as bud tip progenitor-like organoids (iBTOs) that maintained ~80% NKX2-1<sup>+</sup>/CPM<sup>+</sup> cells for at least 8 weeks. scRNA-seq analysis of bud tip progenitor cells from unsorted organoids or after FACS-enrichment revealed a high degree of transcriptional similarity to primary bud tip progenitor organoids, as well as a shared transcriptional signature with *in vivo* bud tip progenitors. In addition, scRNA-seq from iBTOs that have spent less (3 weeks) or more (10 weeks) time in culture suggest that induced bud tip progenitors become more transcriptionally similar to native bud tip progenitors as they age. Finally, we used well-established methods to direct differentiation of iBTOs into organoids composed of airway epithelium (including basal, secretory, ciliated, goblet and neuroendocrine cells) or alveolar type II (AT2) cells (Miller et al., 2020; Jacob et al., 2017). Collectively, this study describes a robust method to generate bud tip progenitor-like cells from iPSCs that closely resemble organoids derived from primary tissue. This model can be readily used to study lung development and illustrates a proof-of-concept for cellular engineering and cell therapy.

## RESULTS

### Lung spheroids are optimized for NKX2-1 expression but remain heterogenous

NKX2-1 is the earliest marker during lung epithelial specification (Lazzaro et al., 1991) and loss of NKX2-1 leads to lung agenesis (Minoo et al., 1995, 1999; Little et al., 2019; Kuwahara et al., 2020). We therefore sought to build on a previously published method to generate iPSC-derived foregut spheroids (Dye et al., 2015) by optimizing endoderm induction efficiency and foregut spheroid development in order to improve NKX2-1 expression.

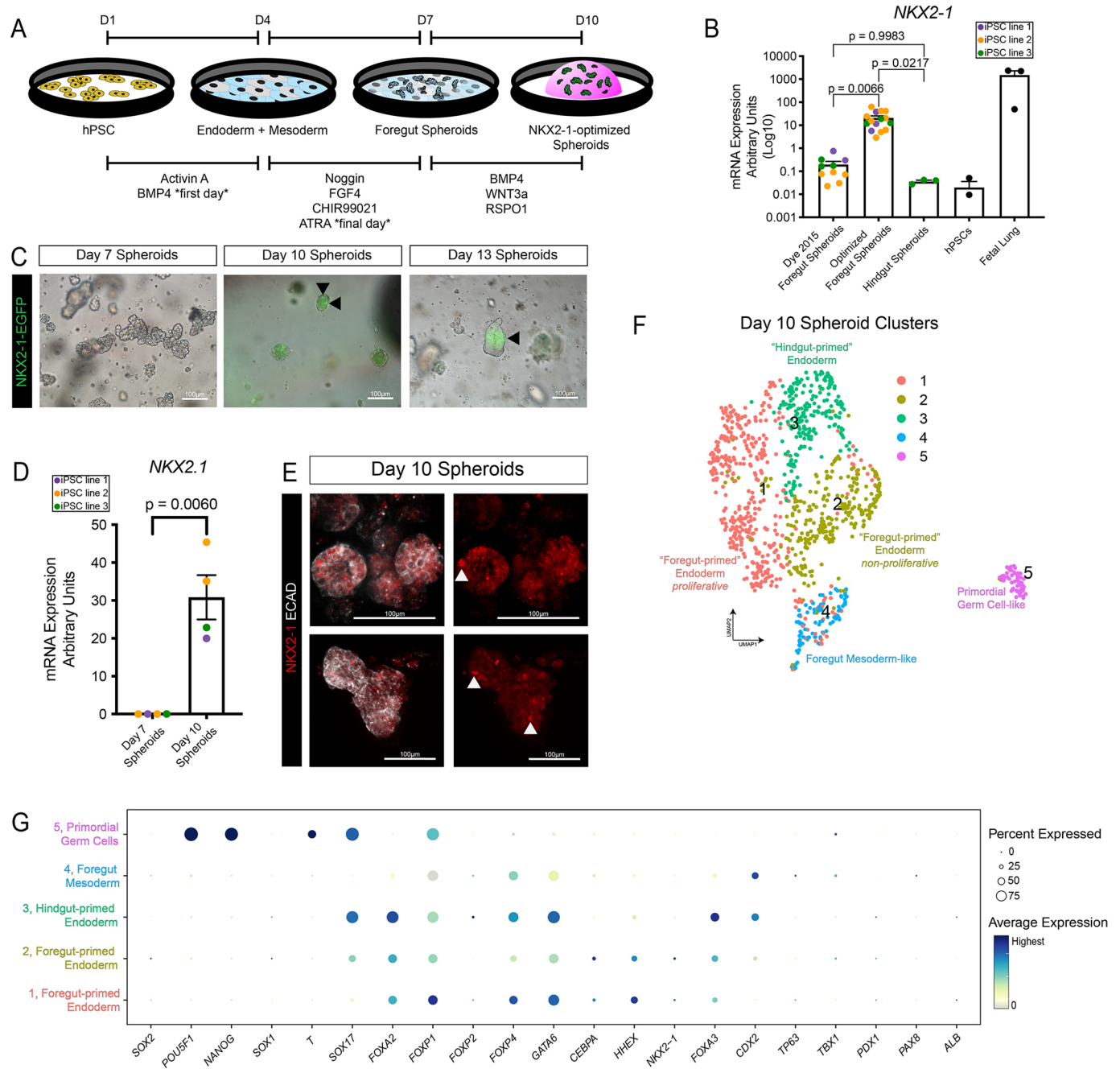
We began by testing conditions to improve definitive endoderm (DE) differentiation efficiency and reproducibility. DE induction of varying efficiencies is achieved by using activin A (ACTA) ligand (Kubo et al., 2004; D'Amour et al., 2006; Spence et al., 2011) with WNT and BMP signaling playing a synergistic role during the initial stages of DE specification (Gadue et al., 2006; Green et al., 2011; Loh et al., 2014; Matsuno et al., 2016; Rankin et al., 2018; Heemskerk et al., 2019). Therefore, we tested combinations of ACTA alongside the small molecule WNT activator CHIR99021 (CHIR) or BMP4 on the first day of a 3-day ACTA differentiation culture. Using flow cytometry with a SOX17-tdTomato and SOX2-mCitrine hESC reporter line (Martyn et al., 2018) to quantitate cell composition, we observed that ACTA alone induced 48% SOX17<sup>+</sup> cells, while the addition of CHIR or BMP4 both enhanced DE differentiation, leading to 96% SOX17<sup>+</sup> cells or 87% SOX17<sup>+</sup> cells, respectively (Fig. S1A). Addition of CHIR and BMP4 together led to 71% SOX17<sup>+</sup> cell induction (Fig. S1A). DE cultures from an additional cell line were co-stained with SOX17 and FOXA2 to confirm definitive endoderm cell identity (Fig. S1B). We observed that near-pure SOX17<sup>+</sup> cultures obtained via ACTA and CHIR failed to give rise to self-organizing 3D spheroids, consistent with published data showing that self-organizing foregut and hindgut organoids consist of both epithelium and mesenchymal lineages (Spence et al., 2011; Dye et al., 2015, 2016). Therefore, ACTA and BMP4 were used for subsequent experiments.

After DE specification, monolayers were directed into 3D foregut endoderm spheroids by combining a method that efficiently induces ventral foregut endoderm competent to be specified as lung (Rankin et al., 2016) alongside methods that induce 3D self-organization (Spence et al., 2011; Dye et al., 2015). This included BMP inhibition via noggin (NOG), FGF4 and CHIR (required for 3D spheroid formation) for 3 days plus all-trans retinoic acid (ATRA) on the last day (Fig. S1C,D). After 3 days, spheroids were collected and suspended in Matrigel and treated for 3 additional days with low BMP4 as well as WNT3A and RSPO1, which together stimulate WNT signaling (Shu et al., 2005; Goss et al., 2009; Harris-Johnson et al., 2009; Domyan et al., 2011; Jacobs et al., 2012; Miller et al., 2012; Serra et al., 2017). WNT activation by combined RSPO and WNT or by CHIR resulted in comparable expression levels of foregut and hindgut markers (Fig. S1E); however, CHIR is a GSK3 $\beta$  inhibitor and can have non-WNT mediated effects, so RSPO and WNT were used as more specific activators of WNT signaling.

At the end of the 9-day directed differentiation (Fig. 1A), spheroids were analyzed for NKX2-1 expression via qRT-PCR (on day 10) (Fig. 1B). When this optimized method was directly compared with spheroids generated using previously published foregut/lung spheroid protocols (Dye et al., 2015; Miller et al., 2019), NKX2-1-optimized foregut spheroids expressed ~100-fold more NKX2-1 (Fig. 1B). Undifferentiated iPSCs and hindgut spheroids were included as controls, both of which had very low NKX2-1, and fetal lung was used as a positive control. By analyzing EGFP expression using a NKX2-1-EGFP reporter cell line, EGFP was not detected at day 7, whereas a low level of ubiquitous expression with scattered NKX2-1<sup>HI</sup> cells could be detected starting on day 10, and expression was localized to specific regions by day 13 (Fig. 1C). NKX2-1 induction between day 7 and day 10 was confirmed by qRT-PCR (Fig. 1D), and whole-mount immunofluorescence of day 10 spheroids correlated with reporter expression, with individual cells expressing high levels of NKX2-1 protein (Fig. 1E, arrowheads).

When day 10 spheroids were analyzed by scRNA-seq, NKX2-1 was also observed in a subset of cells (Fig. 1F,G), revealing heterogeneity within the foregut spheroids and supporting EGFP reporter expression. Spheroids contained two clusters (clusters 1 and 2) of lung-fated ventral foregut endoderm-like cells expressing NKX2-1, FOXA2, FOXP1, FOXP4 and HHEX (Bogue et al., 1998; Shu et al., 2007; Spence et al., 2009; Kearns et al., 2013; Davenport et al., 2016; Li et al., 2016). Clusters and cells were largely negative for other foregut lineage markers, including TP63, TBX1, PDX1, PAX8 and ALB; however, there was a clear population of cells that express FOXA2, FOXA3, SOX17 and CDX2 (cluster 3), which we refer to as 'hindgut-primed' endoderm. Additionally, we observed a cluster of CDX2<sup>+</sup>/HAND1<sup>+</sup>/ISLI<sup>+</sup>/BMP4<sup>+</sup>/FOXF1<sup>+</sup>/LEF1<sup>+</sup> (cluster 4) foregut mesoderm-like cells (Han et al., 2020) and a small population of cells (cluster 5) expressing markers indicative of primordial germ cells, including POU5F1, NANOG, T (TBXT), SOX17, NANOS3 and TFAP2C (Davenport et al., 2016; Jo et al., 2022) (Fig. 1G, Fig. S1G). Comparing optimized foregut spheroids to previously published methods (Dye et al., 2015; Miller et al., 2019) by qRT-PCR, we observed that the hindgut endoderm marker CDX2 was not statistically different, but the early foregut and dorsal foregut endoderm marker SOX2 (Que et al., 2007) was reduced with the optimized method (Fig. S1F), which suggests previous methods are competent to generate foregut that is relatively immature or more dorsal rather than ventral. Taken together, our data show that





**Fig. 1. Optimization of lung spheroids for NKX2-1 expression.** (A) Schematic displaying the directed differentiation protocol from hPSCs to NKX2-1-optimized spheroids. (B) qRT-PCR data comparing *NKX2-1* expression in previously published foregut spheroids (Dye et al., 2015) with optimized foregut spheroids (this protocol) and hindgut spheroids (Spence et al., 2011). hPSCs and whole fetal lung are also included as references. Each colored dot represents a technical replicate from one or more experiments with a unique iPSC line (purple, WTC11; orange, iPSC17 WT 7B2; green, iPSC line 72.3). Data are mean  $\pm$  s.e.m. Statistical tests used were an ordinary one-way ANOVA followed by Tukey's multiple comparison test. (C) Representative reporter expression for NKX2-1-EGFP on day 7, 10 or 13 spheroids. (D) qRT-PCR data comparing *NKX2-1* expression in spheroids collected on day 7 or day 10 (see A). Each colored dot represents the result of an independent experiment with a unique iPSC line (purple, WTC11; orange, iPSC17 WT 7B2; green, iPSC line 72.3). Data are mean  $\pm$  s.e.m. The statistical test used was an unpaired Welch's one-tailed *t*-test. (E) Maximum intensity immunofluorescence confocal z-series stained for the pan-epithelial marker ECAD and the lung epithelial marker NKX2-1 on day 10 spheroids. (F) UMAP cluster plot of scRNA-seq data from day 10 spheroids ( $n=1$  batch,  $\sim 100$  spheroids). Each dot represents a single cell and cells were computationally clustered based on transcriptional similarities. The plot is colored and numbered by cluster. Cell-type labels for each cluster are based on expression of canonical cell-type markers displayed in the dot plot in G or Fig. S1G. (G) Dot plot of cell lineage genes in each cluster of the UMAP plot in F. The dot size represents the percentage of cells expressing the gene in the corresponding cluster; the dot color indicates log-normalized expression level of the gene.

optimized foregut spheroids have much higher levels of NKX2-1 when compared with previous methods, but they are still heterogeneous with distinct populations of lung-fated cells and hindgut-fated cells.

#### iPSC-derived bud tip progenitors emerge over time

Once NKX2-1<sup>+</sup> spheroids were formed, we asked how efficiently these spheroids would give rise to bud tip progenitor (BTP)-like cells as spheroids expanded into larger organoid structures. Our

previous studies have shown that ‘3 Factor (3F) medium’ possessing FGF7, CHIR99021 and ATRA expands primary BTPs derived from the fetal lung and SOX2<sup>+</sup>/SOX9<sup>+</sup> iPSC-derived BTP-like cells; however, efficiency and culture over prolonged periods of time were not assessed (Miller et al., 2018). After inducing NKX2-1-optimized spheroids on day 10, medium was switched to 3F BTP medium (Fig. 2A). Spheroids were maintained in 3F for several weeks, where they expanded into complex, branching structures termed ‘lung progenitor organoids (LPOs)’, similar to what we have observed previously (Miller et al., 2018) (Fig. 2B). LPOs were passaged every 2-3 weeks as either intact organoids with minimum fragmentation (whole passaged) or were sheared by being drawn through a hypodermic needle or pipette, which is a standard method for passaging BTP organoids derived from primary tissue (Miller et al., 2018, 2020; Hein et al., 2022). Using an NKX2-1-EGFP reporter iPSC line to compare retention of lung identity after various passaging methods, whole passaged organoids maintained robust EGFP<sup>+</sup> reporter expression while fragmenting organoids ultimately led to a loss of NKX2-1-EGFP<sup>+</sup> cells (Fig. 2C, quantified in Fig. 2F). Based on the maintenance of NKX2-1-EGFP expression, we therefore chose to use whole passaged LPOs for our remaining experiments.

We evaluated the presence of induced bud tip progenitor (iBTP) cells within whole passaged LPOs by immunofluorescence of paraffin wax-embedded sections and whole-mount organoids (Fig. 2D, Fig. S2A). LPOs possessed regions of NKX2-1-expressing cells, which co-expressed the BTP markers CPM, SOX9 and SOX2 (Nikolić et al., 2017; Miller et al., 2018; Yamamoto et al., 2020) (Fig. 2D, Fig. S2A, Movie 1). LPOs also contained sub-regions of distinct NKX2-1<sup>-</sup>/CDX2<sup>+</sup> cells, with many expressing MUC2, indicative of goblet-like intestinal cells, and smaller regions of NKX2-1<sup>-</sup>/CDX2<sup>-</sup> cells (Fig. 2E, Fig. S2B, Movie 2).

We quantitatively assessed EGFP<sup>+</sup>/CPM<sup>+</sup> cells in LPOs grown for 3-17 weeks in 3F medium using fluorescence-activated cell sorting (FACS) (Fig. 2F, Fig. S2C). We observed an increase in EGFP<sup>+</sup>/CPM<sup>+</sup> cells in culture over time (Fig. 2F, purple bars). At 3.5 weeks, 17% of cells expressed CPM and EGFP, in contrast to 17 weeks, where 51% of cells in culture were EGFP<sup>+</sup>/CPM<sup>+</sup> (Fig. 2F). A small proportion (<10%) of cells were singly positive for EGFP (green) or CPM (blue) at any timepoint, and a population of double-negative cells was observed at all times (Fig. 2F), suggesting that some heterogeneity is maintained in LPOs. We also investigated LPOs derived from two additional non-reporter iPSC lines, using only CPM to quantify iBTPs. We observed a similar increase of iBTPs in culture over time, with 19% CPM<sup>+</sup> cells at 3.5 weeks and 42% CPM<sup>+</sup> cells by 17 weeks (Fig. S2D, left). Finally, organoids derived from previously published lung organoid protocols (Dye et al., 2015; Miller et al., 2019) contained approximately 0.5-6% CPM<sup>+</sup> cells at any timepoint examined (Fig. S2D, right). The increase in CPM<sup>+</sup> cells over time suggests that optimized culture conditions promote the emergence and selection of iBTPs.

To further interrogate the heterogeneity and complexity of the LPOs, we performed scRNA-seq on whole passaged LPOs at 3, 6 and 10 weeks (Fig. 2G). We identified one cluster with robust levels of NKX2-1 and BTP gene expression (cluster 2), several clusters enriched for hindgut markers (clusters 0, 5 and 6), a mesenchymal cluster (cluster 3), a neuroendocrine-like cluster (cluster 4) and a cluster of unknown/uncommitted cells (cluster 1) (Fig. 2G, Figs S2G, S3). The proportion of cells in the BTP cluster (cluster 2) increased over time, while mesenchyme (cluster 3) was depleted over time and hindgut cells (clusters 0, 5, 6) were persistent

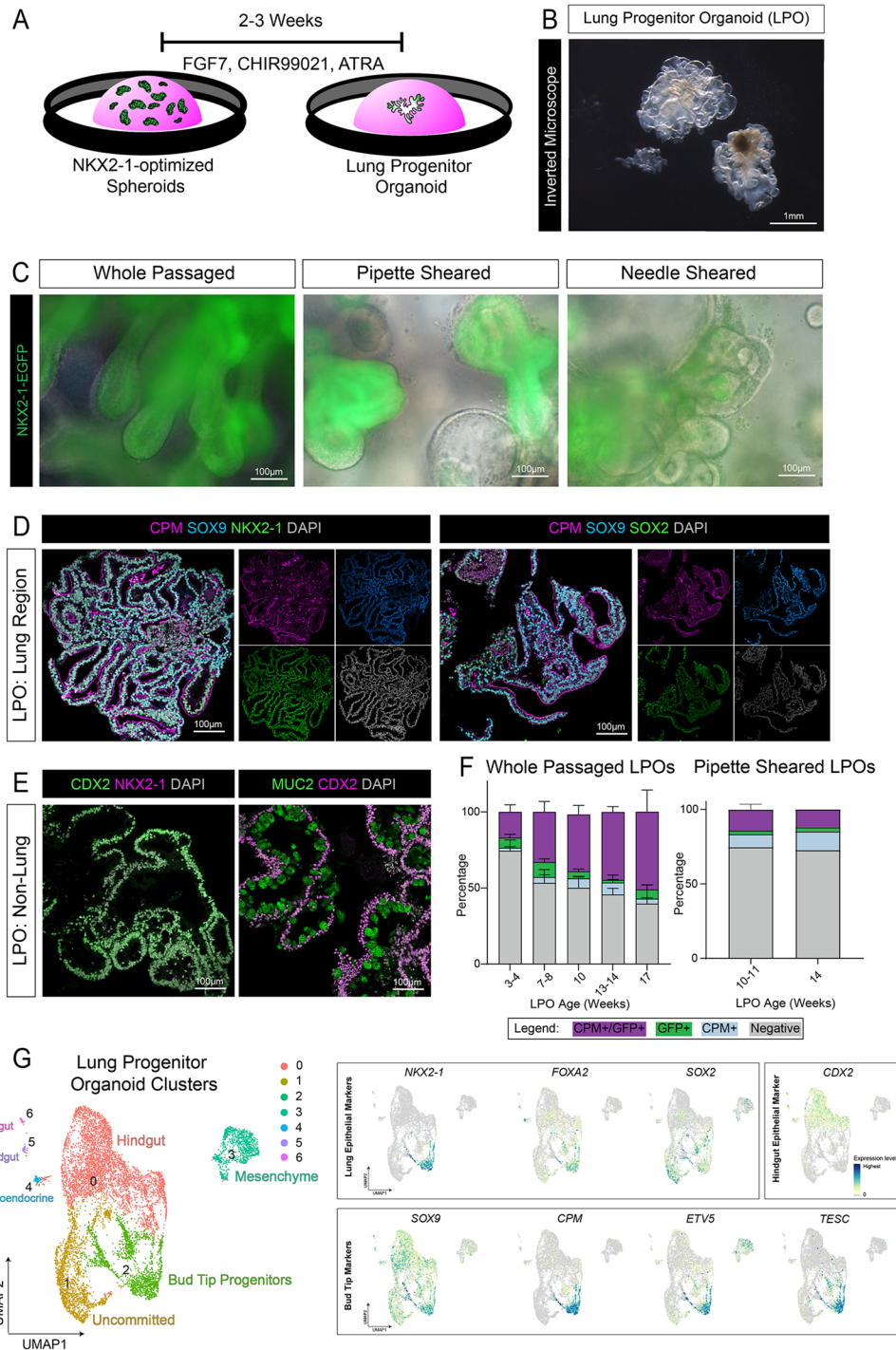
(Fig. S2E,F). Together these data support FACS data suggesting that iBTPs continue to expand within LPOs over time and identify contaminating lineages that persist.

To assess the proliferation of the cultures, we quantified KI67 expression within each cluster and sample from the LPO scRNA-seq data (Fig. S2H). KI67<sup>+</sup> cells were present in every cluster; however, the hindgut (clusters 0, 5 and 6), BTP (cluster 2) and mesenchymal (cluster 3) clusters accounted for 92.24% of KI67<sup>+</sup> cells within the culture compared with 7.76% from the uncommitted (cluster 1) and neuroendocrine (cluster 4) clusters (Fig. S2H, top). Contribution of KI67<sup>+</sup> cells to each cluster by sample was also calculated and normalized to the number of cells within each sample (Fig. S2H, bottom). The contribution of samples to the KI67<sup>+</sup> cells within clusters was varied. Between the 3-week samples, there were proliferating cells that contributed to all clusters; however, there was a significant contribution of this early timepoint to the uncommitted cluster (cluster 1), which comprised 87.11% KI67<sup>+</sup> cells. The 6-week samples also contributed KI67<sup>+</sup> cells to all the clusters, with the most significant contributions to hindgut (clusters 0, 5 and 6) and the mesenchyme (cluster 3). The data suggest that by 6 weeks any remaining contaminating cell types (i.e. CDX2<sup>+</sup> intestinal cells) are most proliferative, while the earlier cultures are altogether generally proliferative. Finally, the 10-week sample contributed a significant proportion of proliferating cells to the BTP cluster (cluster 2; 39.14% of KI67<sup>+</sup> cells), the neuroendocrine cluster (cluster 4) and the two smaller hindgut clusters (clusters 5 and 6) (Fig. S2H, top). Importantly, all samples contributed to the BTP cluster, with a robust contribution from the 10-week sample (Fig. S2H, top).

Cell death in LPOs, evaluated by Hematoxylin and Eosin, and cleaved caspase 3 (CCAS3) stains, was evident at all stages of growth and did not appear to increase with time or localize to contaminating CDX2<sup>+</sup> regions; however, localization of CCAS3 staining transitioned from individual cells at 3 weeks to luminal regions by 17 weeks in culture (Fig. S2I). These data suggest that the increasing number of iBTP cells observed is not likely due to death of contaminating cell types.

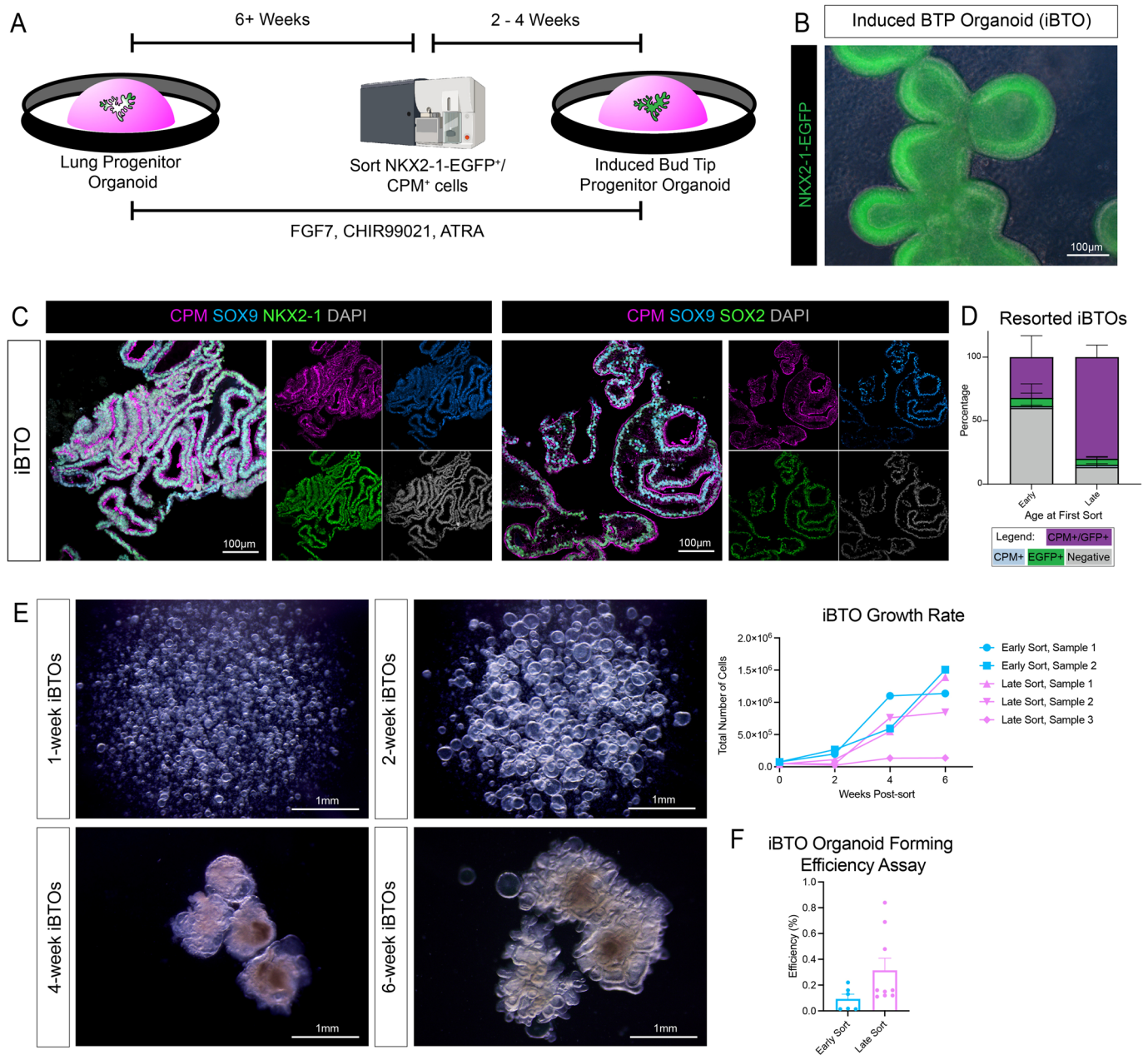
### iPSC-derived bud tip progenitors can be isolated, expanded and maintained long-term

Although LPOs at every time point contain a proliferating population of iBTPs, given the large population and persistence of hindgut lineages, we aimed to isolate NKX2-1-EGFP<sup>+</sup>/CPM<sup>+</sup> cells to generate higher purities of induced bud tip progenitor organoid (iBTO) cultures. iBTPs were isolated via FACS with CPM and NKX2-1-EGFP (or CPM only for non-reporter cell lines) and were replated in Matrigel at 5000 cells/μl in 3F medium to form iBTOs (Fig. 3A). After sorted cells formed into organoids, iBTOs were passaged by whole passaging approximately every 2 weeks. The NKX2-1-EGFP reporter line showed uniform expression of NKX2-1-EGFP in iBTOs (Fig. 3B), and immunofluorescence of paraffin wax-embedded sections showed iBTOs contain a near-homogenous population of cells expressing BTP markers CPM, SOX9 and SOX2 (Fig. 3C). Interestingly, the age of LPOs at the time of sorting was a crucial determinant of the ability of iBTOs to maintain NKX2-1 and CPM expression. For example, iBTOs generated from <6-week LPOs maintained 32% NKX2-1-EGFP<sup>+</sup>/CPM<sup>+</sup> cells when re-sorted 7 weeks later (Fig. 3D). In contrast, iBTOs from >6-week LPOs maintained 80% NKX2-1-EGFP<sup>+</sup>/CPM<sup>+</sup> cells when re-sorted 8 weeks later (Fig. 3D). This suggests that iBTPs undergo increasing commitment to a bud tip progenitor identity as they are maintained in 3F medium at the LPO stage.



**Fig. 2. iPSC-derived bud tip progenitors emerge over time in LPOs.** (A) Schematic displaying the lung progenitor organoid (LPO) expansion protocol from NKX2-1-optimized spheroids. LPOs form after 2-3 weeks in culture. (B) Bright-field image of 6-week LPOs on an inverted microscope. (C) Representative NKX2-1-EGFP reporter images of 10- to 11-week LPOs, passaged whole or sheared (pipette and needle). (D) Immunofluorescence staining of paraffin wax-embedded sections of the lung-like regions of 12-week LPOs for bud tip progenitor markers CPM and SOX9, and the lung epithelial markers NKX2-1 (left) or SOX2 (right). (E) Immunofluorescence staining on paraffin wax-embedded sections for the intestinal epithelial marker CDX2, the intestinal goblet cell marker MUC2 and the lung epithelial marker NKX2-1 on non-lung regions of 12-week LPOs. (F) FACS quantification of NKX2-1-EGFP<sup>+</sup> CPM<sup>+</sup> cells in whole passed or pipette-sheared LPOs in aggregate time course (3-17 weeks) from the NKX2-1-EGFP reporter cell line (iPSC17 WT 7B2). Percentages of live cells expressing neither marker (negative, gray) or each separate marker (CPM<sup>+</sup> only, blue; EGFP<sup>+</sup> only, green), or dual-expressing cells (CPM<sup>+</sup>/EGFP<sup>+</sup>, purple) are reported as mean±s.e.m. for 3-7 replicates per time point. (G) UMAP cluster plot of scRNA-seq data from LPOs ( $n=2$  biological replicates for 3- and 6-week timepoints,  $n=1$  for 10-week timepoint). Each dot represents a single cell and cells were computationally clustered based on transcriptional similarities. The plot is colored and numbered by cluster. Cell-type labels for each cluster are based on expression of canonical cell-type markers displayed in the heat map in Fig. S3 and the dot plot and feature plots in G and Fig. S2G. Feature plots corresponding to the LPO cluster plot and displaying canonical bud tip progenitor markers (*SOX9*, *CPM*, *ETV5* and *TESC*) (Miller et al., 2018, 2020; Yamamoto et al., 2020), lung epithelial markers (*NKX2-1*, *FOXA2* and *SOX2*) and a hindgut epithelial marker (*CDX2*). The color of each dot in the feature plot indicates log-normalized expression level of the labeled gene in the represented cell.





**Fig. 3. iPSC-derived bud tip progenitors can be isolated and expanded in the long term.** (A) Schematic displaying isolation and expansion of induced bud tip progenitor organoids (iBTOs). LPOs are maintained in 3F medium and whole passaged for at least 6 weeks, then dissociated for FACS. iBTPs are isolated using CPM<sup>+</sup> expression with or without NKX2-1-EGFP<sup>+</sup> reporter expression and replated as isolated iBTPs. iBTPs re-form to iBTOs over 2-4 weeks, are maintained in 3F medium and whole passaged. Schematic created using Biorender.com. (B) Representative NKX2-1-EGFP reporter image of 3-week iBTOs. (C) Immunofluorescence staining on paraffin wax-embedded sections of nearly homogenous 4-week iBTOs for bud tip progenitor markers CPM and SOX9, and the lung epithelial markers NKX2-1 (left panels) or SOX2 (right panels). (D) FACS quantification of NKX2-1-EGFP<sup>+</sup>/CPM<sup>+</sup> cells in iBTOs from 3-week sorted LPOs (early) or 8-17 week sorted LPOs (late) from the NKX2-1-EGFP reporter cell line (iPSC17 WT 7B2). iBTOs were analyzed 7-8 weeks after iBTP purification from LPOs. Percentages of live cells expressing neither marker (negative, gray) and each separate marker (CPM<sup>+</sup> only, blue; EGFP<sup>+</sup> only, green), or dual-expressing cells (CPM<sup>+</sup>/EGFP<sup>+</sup>, purple) are reported as mean ± s.e.m. for three replicates per time point. (E) (Left) Bright-field images of iBTOs 1-, 2-, 4- and 6-weeks post-sort from 11-week LPOs on an inverted microscope. (Right) Quantification of iBTO growth from the day of sorting to 6 weeks post-sorting as the number of cells in organoids generated from the same number of starting cells from iBTOs sorted from 4-6 week LPOs (early sorts) or 10-11 week LPOs (late sorts). (F) Organoid-forming efficiency assay of iBTOs sorted from 4-6 week LPOs (early sort) or 10-11 week LPOs (late sort). Organoid-forming efficiency was measured as the number of cysts formed during the 2 weeks after iBTOs were digested to a single cell suspension and re-plated in Matrigel at 1000 or 2500 cells per μl.

To assess iBTO expansion, we evaluated the growth rate of iBTOs from the day iBTPs were isolated from LPOs to 6 weeks of growth as iBTOs. Small cysts could be detected by 1 week of culture, which grew in circumference by 2 weeks (Fig. 3E). By 4 weeks of culture, cysts took on a more complex, branched

phenotype, which increased in complexity by 6 weeks (Fig. 3E). Growth of iBTOs sorted from 4- to 6-week LPOs (early timepoints) or 10- to 11-week LPOs (late timepoints) was quantified by counting the number of cells in organoids generated from 40,000 to 75,000 original cells (depending on organoid batch) at 2, 4 and 6

weeks in culture. Overall, a steady increase in cell number was observed for iBTOs sorted from early or late LPOs, with rare batches showing low growth efficiency (Fig. 3E, right). We also measured organoid-forming efficiency of iBTOs by dissociating iBTOs into single cells and replating iBTPs to determine how many cysts formed after 2 weeks. iBTOs from later sorts (10- to 11-week LPOs) had higher organoid-forming efficiency than iBTOs from earlier sorts (4- to 6-week LPOs) (Fig. 3F).

### iPSC-derived bud tip progenitors are transcriptionally similar to primary bud tip progenitor cultures

To further interrogate iBTO cellular composition, and to directly compare cells within iBTOs with primary (*in vivo*) bud tip progenitors and primary bud tip progenitor organoids derived from fetal tissue (Miller et al., 2018), we performed scRNA-seq on iBTOs derived from 4- and 10-week LPOs. iBTOs were expanded for 4 weeks in culture post-sorting before scRNA-seq was carried out. Integrated analysis of both datasets resulted in four epithelial clusters (Fig. 4A). Cells in cluster 4 expressed hindgut/intestinal markers (i.e. *CDX2*), which could also be identified by immunofluorescence (Fig. S4A) and represented a small fraction of cells that were predominantly derived from the 4-week sample (Fig. 4B,C). The remaining three clusters (clusters 1, 2 and 3) have enriched expression of *NKX2-1* as well as BTP markers (*SOX9*, *CPM*, *ETV5*, *TESC*, *FGF20*, *SOX11*, *HMGB2*, *NPC2*, *LGR5* and *ETV4*) (Miller et al., 2018, 2020; Hein et al., 2022); however, expression of proliferation genes was variable, suggesting there is some heterogeneity representing different iBTP cell states that is likely driven by the expression of proliferation genes (Fig. 4D, Fig. S4B). Similar to the analysis carried out on LPOs, we quantified *KI67*-expressing cells in each cluster and sample in the iBTO scRNA-seq data (Fig. S4C). All clusters had similar normalized percentages of *KI67*<sup>+</sup> cells (between 15 and 34%) (Fig. S4C, left panel); however, the contribution of *KI67*<sup>+</sup> cells to each cluster varied by sample (i.e. timepoint). iBTOs from the 10-week sort contributed more *KI67*<sup>+</sup> cells to BTP clusters 0 and 1, while iBTOs from the 4-week sort contributed more to the hindgut cluster (cluster 4). Both samples contributed similarly to BTP cluster 3 (Fig. S4C, right panel). Together with the analysis on LPOs (Fig. S2H), this suggests that, although iBTPs are proliferative, the highly proliferative contaminating lineages (i.e. hindgut) can persist in culture despite the sorting strategy if the cultures are sorted too early. This could be attributed to plasticity within cultures during early time points.

To further benchmark transcriptional similarities of iBTPs to *in vivo* BTPs, we re-analyzed published scRNA-seq data from the primary human fetal lung (Miller et al., 2020; Hein et al., 2022) and identified a panel of the top 100 enriched genes expressed in *in vivo* BTPs relative to all other epithelial cell types in the developing lung (see Materials and Methods, Table S1). We used this gene list as a reference to assign an '*in vivo* bud tip progenitor cell score' to each individual cell from the LPOs and iBTOs using published computational methods (Holloway et al., 2020; Hein et al., 2022). For the purposes of this comparison, both whole LPOs and extracted *NKX2-1*<sup>+</sup> cells from the LPO datasets were included in the analysis. As additional control comparisons, we analyzed published data and included scores for primary BTOs, *in vivo* basal cells and *in vivo* neuroendocrine cells (Miller et al., 2018, 2020; Hein et al., 2022) (Fig. 4E). As expected, *in vivo* BTPs had the highest score (median score=0.80), while *in vivo* neuroendocrine cells had the lowest score (0.37), followed by whole LPOs (0.47). A noteworthy observation is that human primary BTOs had a mean BTP score of 0.62,

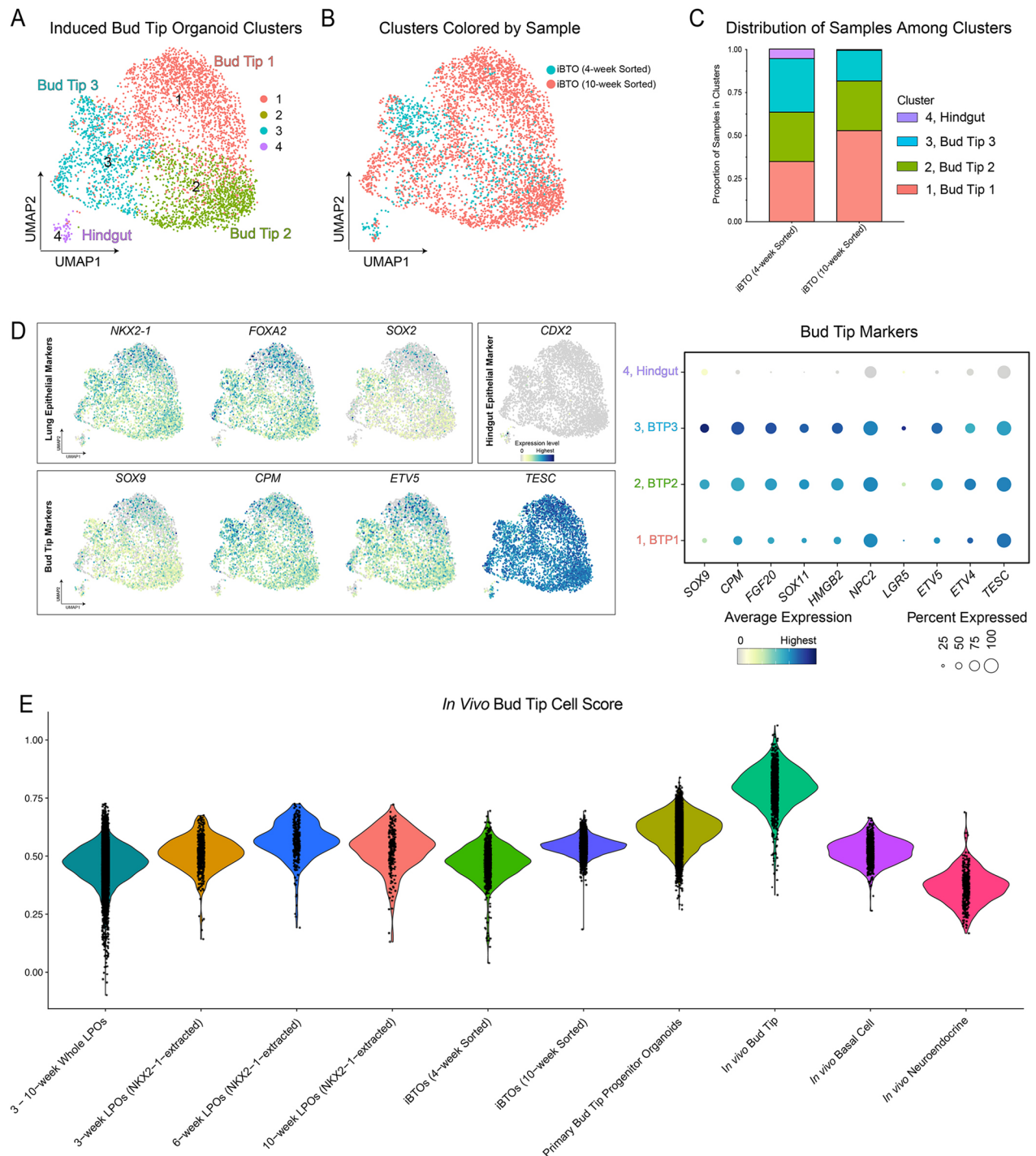
which was much lower than the *in vivo* BTP score of 0.80, indicating that the *in vitro* culture conditions significantly influence gene expression, as has been observed previously (Miller et al., 2020). Relative to primary BTOs, both *NKX2-1*<sup>+</sup> cells from LPOs and iBTOs showed high transcriptional similarity; their scores improved with longer time in culture at the LPO stage (0.51 to 0.57 for *NKX2-1*<sup>+</sup> cells from LPOs; 0.48 to 0.54 for iBTOs) (Fig. 4E). To provide additional confidence in this comparison method, we generated a similar *in vivo* basal cell score and *in vivo* neuroendocrine cell score (Table S1) and observed that the expected populations (i.e. *in vivo* basal cells and *in vivo* neuroendocrine cells, respectively) scored the highest relative to all the other data sets analyzed (Fig. S4D).

As an additional analysis, we used published scRNA-seq data from primary human fetal lung epithelium and performed label transfer, which mapped iBTOs to the most similar cell type within the human fetal lung reference dataset (Stuart et al., 2019; Miller et al., 2020; Hein et al., 2022). After clustering the primary epithelial cells (Fig. S4E), iBTOs from LPOs sorted at 4 weeks or 10 weeks largely mapped to four different clusters (Fig. S4F,G). Upon closer investigation into these clusters, three of the clusters (clusters 2, 7 and 9) scored highly for BTP genes (using the BTP cell score) and expressed known markers of BTP cells; the fourth cluster (cluster 6) resembled neuroendocrine progenitor cells (Fig. S4F). Quantification of this mapping revealed that the majority of iBTO cells mapped to two of the three BTP clusters (clusters 2 and 7) whereas only a small subset mapped to the third BTP cluster (cluster 9) and the neuroendocrine cluster (cluster 6) (Fig. S4G). Taken together, these data support the conclusion that iBTOs contain a stable bud tip progenitor cell that shares a high degree of transcriptional similarity with primary bud tip progenitors.

### iBTOs can give rise to airway or alveolar fates

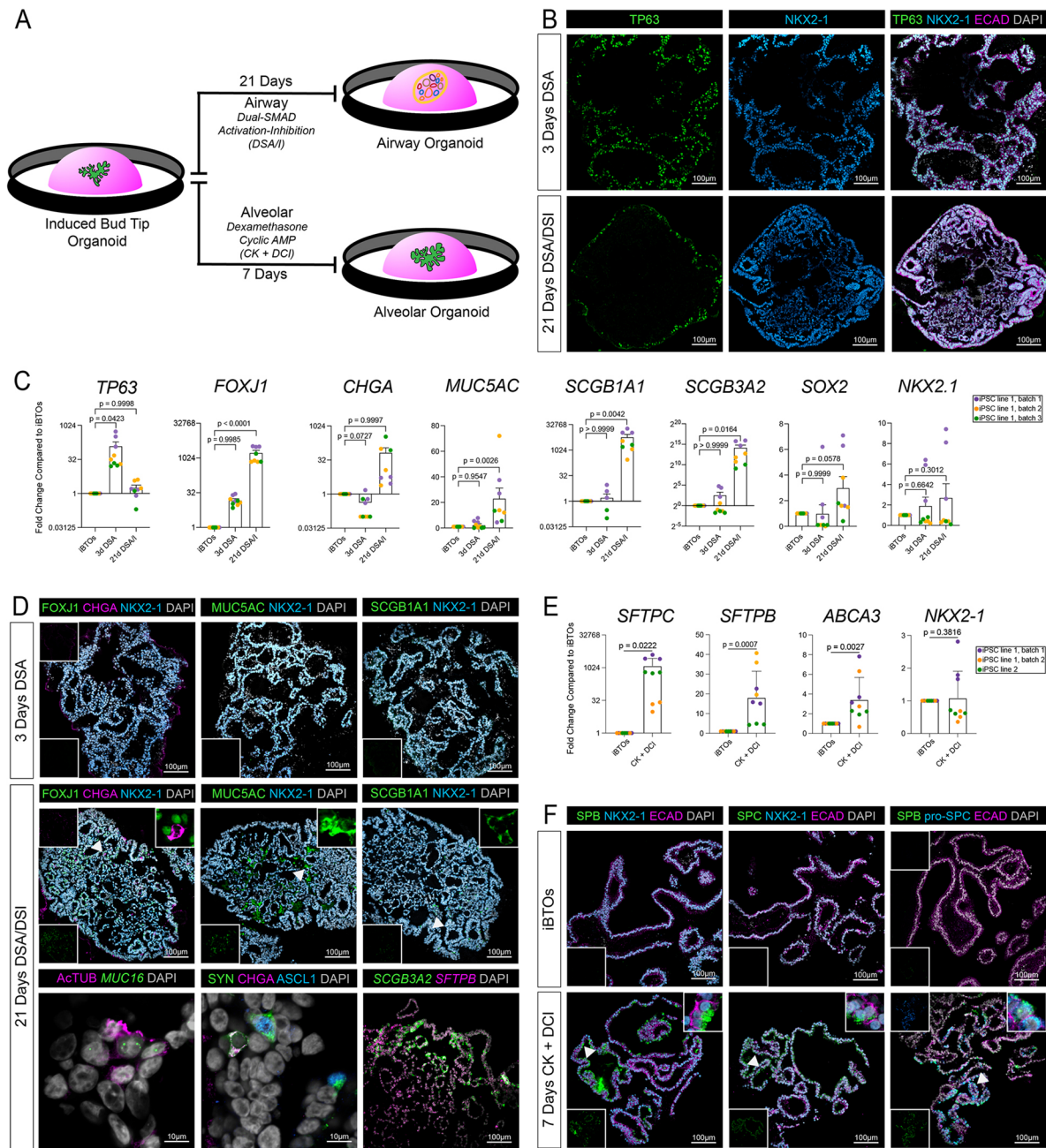
Given that BTPs in the developing lung give rise to airway and alveolar fates, we hypothesized that iBTOs could be guided into airway and alveolar lineages. To test this possibility, we used methods previously developed to efficiently induce lung progenitors into airway (Miller et al., 2020) or alveolar (Jacob et al., 2017) lineages (Fig. 5A). Airway induction involved 3 days of dual-SMAD activation (DSA) followed by 18 days of dual-SMAD inhibition (DSI) and resulted in condensed structures that maintained *NKX2-1*-EGFP expression and expressed mCherry driven by the TP63 promoter (TP63-mCherry) (Fig. S5A, left column). By immunofluorescence, we observed that TP63 expression was highly induced after 3 days of DSA, as expected; after 18 days of DSI, TP63<sup>+</sup> cells organized around the perimeter of the organoids (Fig. 5B,C).

We also performed airway induction on unsorted LPOs, hypothesizing that airway differentiation may select for expansion of lung lineages within the LPOs, while suppressing non-lung lineages. *NKX2-1*-EGFP<sup>+</sup> and TP63-mCherry<sup>+</sup> cells were quantified using flow cytometry on both LPOs and iBTOs following the 21-day DSA/DSI protocol or untreated controls (Fig. S5B). Both untreated LPOs and iBTOs contained <1% *NKX2-1*-EGFP<sup>+</sup>/TP63-mCherry<sup>+</sup> cells. Only upon treatment with the 21-day DSA/DSI protocol were significant numbers of EGFP<sup>+</sup>/mCherry<sup>+</sup> cells detected (7% in LPOs, 5% in iBTOs) (Fig. S5B). DSA/DSI-treated iBTOs also maintained a large proportion (71%) of *NKX2-1*-EGFP<sup>+</sup>/TP63-mCherry<sup>-</sup> cells, which represented the spectrum of differentiated airway epithelial cell types identified by immunofluorescence and qRT-PCR, including multiciliated, neuroendocrine, goblet and secretory cells (Fig. 5C,D, Fig. S5B). Airway organoids contained AcTUB<sup>+</sup>/*MUC16*<sup>+</sup> ciliated cells



**Fig. 4. iPSC-derived bud tip progenitors are transcriptionally similar to human fetal bud tip progenitor cultures.** (A) UMAP cluster plot of scRNA-seq data from iBTOs ( $n=1$  biological replicates for iBTOs from 4- and 10-week LPOs). Each dot represents a single cell and cells were computationally clustered based on transcriptional similarities. The plot is colored and numbered by cluster. Cell-type labels for each cluster are based on expression of canonical cell-type markers displayed in the dot plots and feature plots in D and Fig. S4B. (B) UMAP plot corresponding to the iBTO cluster plot in A. Each dot represents a single cell and dots/cell are colored according to the sample from which they came. (C) Stacked bar graph displaying the proportion of cells from each sample in each cluster of the iBTO cluster plot in A. (D) Feature plots and dot plot corresponding to the iBTO cluster plot in A and displaying canonical bud tip progenitor markers (*SOX9*, *CPM*, *ETV5*, *TESC*, *FGF20*, *SOX11*, *HGMB2*, *NPC2*, *LGR5* and *ETV4*) (Miller et al., 2018, 2020; Yamamoto et al., 2020), lung epithelial markers (*NKX2-1*, *FOXA2* and *SOX2*) and hindgut epithelial marker (*CDX2*). The color of each dot in the feature plot indicates log-normalized expression level of the labeled gene in the represented cell. For the dot plot, the dot size represents the percentage of cells expressing the gene in the corresponding cluster, and the dot color indicates log-normalized expression level of the gene. (E) Violin plot displaying an *in vivo* bud tip progenitor cell score, calculated as the average expression of the top 100 enriched genes in *in vivo* bud tip progenitor cells for each sample. Samples include whole LPOs (two 3-week LPOs, two 6-week LPOs and one 10-week LPOs), *NKX2-1*-extracted cells from 3-, 6- and 10-week LPOs ( $n=2$  for 3- and 6-week LPOs,  $n=1$  for 10-week LPOs), whole iBTOs (derived from LPOs sorted for *NKX2-1*<sup>+</sup>/*CPM*<sup>+</sup> cells at 4 or 10 weeks,  $n=1$  of each), human fetal-derived (primary) bud tip progenitor organoids (14 weeks post-conception) (Miller et al., 2020) and primary *in vivo* tissue (8.5–19 weeks post-conception), including computationally extracted bud tip, basal and neuroendocrine cells (Miller et al., 2020; Hein et al., 2022).





**Fig. 5. iBTOs are competent for proximal airway and distal alveolar differentiation.** (A) Schematic displaying the airway (Miller et al., 2020) and alveolar (Jacob et al., 2017) induction protocols from iBTOs. (B) Immunofluorescence staining on paraffin wax-embedded sections for the airway progenitor marker TP63, lung epithelial marker NKX2-1 and general epithelial marker ECAD on iBTOs that have undergone 3 days of dual-SMAD activation (DSA) or 3 days of DSA followed by 18 days of dual-SMAD inactivation (DSI) in the airway induction protocol. (C) qRT-PCR data comparing expression of airway markers *TP63*, *FOXJ1*, *CHGA*, *MUC5AC*, *SCGB1A1*, *SCGB3A2* and *SOX2*, and lung epithelial marker *NKX2-1* in untreated iBTOs or iBTOs that have undergone 3 days of DSA or 3 days of DSA followed by 18 days of DSI. Each colored dot represents a result from an independent experiment using the iPSC17 WT 7B2 line. Data are mean $\pm$ s.e.m. Statistical tests used were one-way ANOVA followed by Dunnett's multiple comparison test. (D) Fluorescence *in situ* hybridization and/or immunofluorescence staining on paraffin wax-embedded sections for differentiated airway epithelial markers (multiciliated, FOXJ1, ActTUB and MUC16; neuroendocrine, CHGA, SYN and ASCL1; goblet, MUC5AC; secretory, SCGB1A1, SCGB3A2 and SFTPB) and the lung epithelial marker NKX2-1 on iBTOs that have undergone 3 days DSA or 3 days DSA followed by 18 days of DSI. Insets in the top right corners are zooms of the areas indicated by the arrowheads (DAPI and NKX2-1 channels removed for clarity). Insets in the bottom left or top left corners are single-channel images of non-DAPI or non-NKX2-1 channels, respectively. (E) qRT-PCR data comparing expression of alveolar markers *SFTPC*, *SFTPB* and *ABCA3*, and the lung epithelial marker *NKX2-1* in untreated iBTOs (in 3F medium) or after the alveolar differentiation protocol (7 days CK+DCI). Each colored dot represents a result from an independent experiment with a unique iPSC line or an independent experiment (purple and orange, iPSC17 WT 7B2 LPOs sorted by CPM and NKX2-1-EGFP; green, iPSC line 72.3 LPOs sorted by CPM only). Data are mean $\pm$ s.e.m. Statistical test used was an unpaired Welch's one-tailed *t*-test. (F) Immunofluorescence staining on paraffin wax-embedded sections for differentiated type II alveolar epithelial markers SFTPB, pro-SFTPC and SFTPC, and the lung epithelial marker NKX2-1 or the general epithelial marker ECAD on iBTOs and iBTOs that have undergone the alveolar differentiation protocol (7 days CK+DCI). Insets in the top right corner are zooms of the area indicated by the arrows. Insets in the bottom left or top left corners are single channel images of non-DAPI and non-NKX2-1 or non-ECAD channels, respectively.

(Carraro et al., 2021) (Fig. 5D), and SYN<sup>+</sup> cells were double positive for ASCL1 (early neuroendocrine marker) or CHGA (late neuroendocrine marker) (Fig. 5D). Treated LPOs contained a small proportion of cells expressing only TP63-mCherry (5%) and a large proportion were negative for both markers (58%) (Fig. S5B,D), suggesting non-lung lineages continue to expand in these conditions.

Alveolar induction consisted of 7 days of treatment with cyclic AMP and dexamethasone (CK+DCI), as has been previously described with the exception of an alternate composition of the base medium (see Materials and Methods) (Jacob et al., 2017). Treatment of iBTOs with CK+DCI resulted in expanded budded structures that maintained NKX2.1-EGFP expression (Fig. S5A, right column). Alveolar type II markers *SFTPB*, *SFTPC* and *ABCA3* were low to undetectable in iBTOs but were highly expressed upon treatment with CK+DCI, as determined by qRT-PCR, and were co-expressed within the same cells, as shown by immunofluorescence for pro-SFTPC, SFTPC, SFTPB and NKX2-1 (Fig. 5E,F). The lung-specific epithelial marker NKX2-1 was robust in both airway and alveolar organoids, while expression of the intestinal epithelial marker CDX2 was not regularly detected (Fig. 5C-F, Fig. S5C). Taken together, differentiation of iBTOs into airway and alveolar cell types with minimal contaminating non-lung cell types supports the observation that iBTOs have similar developmental potential to *in vivo* bud tip progenitors.

## DISCUSSION

In the current work, we demonstrate an optimized *in vitro* model system for differentiating iPSCs into an NKX2-1-expressing bud tip progenitor lineage, which we show can give rise to both airway and alveolar cell types. Several groups, including ours (Dye et al., 2015, 2016; Miller et al., 2018, 2019), have previously characterized protocols that yield lung-like cell types from iPSCs (Wang et al., 2007; Mou et al., 2012; Gotoh et al., 2014; Huang et al., 2014; Konishi et al., 2016; Nikolić et al., 2017; Chen et al., 2017; Hawkins et al., 2017; McCauley et al., 2017; Tamò et al., 2018; de Carvalho et al., 2019; Jacob et al., 2019; Yamamoto et al., 2020; Leibel et al., 2020). Although many of these protocols capture a transient bud tip progenitor-like stage, they focus primarily on deriving more mature lung epithelial lineages, and the induction and maintenance of NKX2-1<sup>+</sup> bud tip progenitor-like cells from iPSCs has not been previously reported. Furthermore, benchmarking of induced cells using scRNA-seq in order to compare cell types with a human reference atlas has only recently become commonplace (Holloway et al., 2020; Hor et al., 2020; Yu et al., 2020; Hawkins et al., 2021). The method reported here expands upon an improved understanding of crucial signaling events that regulate bud tip progenitor maintenance in the human lung to better mirror this process in a dish (Conway et al., 2020).

Although we demonstrate that optimizing NKX2-1 expression leads to a 100-fold increase when compared with previously reported methods (Dye et al., 2015; Miller et al., 2019), scRNA-seq and whole-mount immunofluorescence analysis of day 10 spheroids revealed that low levels of NKX2-1 are expressed in many cells, whereas a small number of cells expressed high levels of NKX2-1, highlighting the heterogeneity that still exists at this early timepoint. We also observed non-lung lineages that co-emerge in spheroids, possibly owing to the fact that the same signaling pathways play a role during lineage commitment for multiple lineages, while also reflecting the inherent plasticity of newly committed cells. We observed that non-lung lineages, and particularly hindgut lineages, persisted over time in culture. Based on this observation, there is an

opportunity to further refine the signaling pathways that are manipulated to expand and maintain NKX2-1<sup>+</sup> lung lineages; however, this may prove challenging given the redundant use of some pathways (i.e. WNT) in maintaining stem/progenitor cells from both lineages (Pinto and Clevers, 2005; Kapoor et al., 2007; Volckaert and De Langhe, 2015; Chin et al., 2016; Nikolić and Rawlins, 2017; Yan et al., 2017; Miller et al., 2018; Rabata et al., 2020; Aros et al., 2021). Additionally, as mesenchyme was observed by scRNA-seq at the spheroid and LPO stages, how these cells influence epithelial cell fate should be a topic of further exploration.

We observed that hindgut lineages expand after passaging organoids using the standard approach of mechanically shearing them into small fragments. It is currently unclear whether this phenomenon occurs because shearing increases selection for highly proliferative CDX2<sup>+</sup> cells, if the mechanical stress results in higher levels of apoptosis in one lineage (i.e. lung) compared with the other, or if shearing disrupts the bud tip niche. Thus, although we have optimized the current methods to generate a high proportion of NKX2-1<sup>+</sup> cells, LPO cultures still exhibited high plasticity that could be triggered through perturbations such as organoid dissociation.

One important observation from our current study is that induced bud tip progenitors exhibited increased lineage commitment with longer time in culture. Isolating NKX2-1-EGFP<sup>+</sup>/CPM<sup>+</sup> cells prior to 6 weeks resulted in cultures replete with hindgut cells. Although follow-up experiments are required to determine the origin of contaminating hindgut cells, the fact that we still see these cells in purified iBTO cultures suggests a fate switch from a lung bud tip progenitor to a hindgut identity. On the other hand, purifying cells at 6-10 weeks reliably led to the establishment of NKX2-1-EGFP<sup>+</sup>/CPM<sup>+</sup> iBTOs that expanded and maintained their fate. Whether or not the mechanism underlying the loss of bud tip progenitors caused by sorting LPO cultures too early is the same as that underlying bud tip progenitor loss caused by shearing cultures is an avenue for future research.

We attempted airway differentiation on unsorted LPOs, hypothesizing that the process of differentiation itself may select against gut lineages and expand only lung lineages; unfortunately, non-lung lineages persisted. Given this result, sorting remains an important step for establishing iBTO cultures that can be differentiated into airway or alveolar cell types in the absence of non-lung cell types. Finally, the experiments carried out here relied heavily on an NKX2-1-EGFP reporter; however, our results suggest that this should not limit studies with non-reporter iPSC lines, as we also observed that sorting with CPM alone highly enriched for bud tip progenitors. As we detected some NKX2-1-EGFP<sup>-</sup>/CPM<sup>+</sup> cells, identifying additional bud tip progenitor cell surface markers to further improve the purity of cultures may be valuable for enhancing purity in non-reporter iPSC cultures.

The use of emerging technologies such as scRNA-seq has provided crucial insight into the heterogeneity and complexity of human tissues that were once difficult to study (Miller et al., 2020; Yu et al., 2020). New information from human tissue has provided both an atlas against which *in vitro* model systems can be benchmarked and a roadmap that can be used to infer transcriptional and signaling mechanisms that control cellular transitions. Here, we used scRNA-seq to benchmark *in vitro* cultures at several stages of differentiation to catalog the induction, emergence and maintenance of lung-fated cells as they acquire a bud tip progenitor fate. One interesting observation from the benchmarking carried out in this study is a significant shift in the



transcriptome when comparing primary *in vivo* tissue with *in vitro* grown primary organoids, even when the source of the cells is the same. For example, here, we compared iBTOs with both primary BTOs and *in vivo* fetal bud tip progenitors. We observed a significant shift when comparing fetal lung bud tip progenitors with primary BTOs, indicating that the *in vitro* environment significantly changes the transcriptome of the cell, as has been recently reported (Miller et al., 2020; Alysandratos et al., 2022preprint). With this caveat in mind, iBTOs shared a very similar transcriptome to primary BTOs and a high degree of similarity to *in vivo* primary bud tip progenitors. The cell scoring metric used to compare similarity across samples also suggested that iBTOs became more transcriptionally similar to the *in vivo* bud tip progenitors as they spent more time in culture, supporting the idea that induced bud tip progenitors undergo continued differentiation towards a bud tip progenitor identity as they are maintained in culture.

Functionally, iBTOs also behave like bud tip progenitors. Using methods to differentiate lung progenitor cells into airway or alveolar cell types (Miller et al., 2020; Jacob et al., 2017), we demonstrated that iBTOs robustly differentiate into airway and alveolar cell types. Taken together, the current work describes a robust iPSC-derived bud tip progenitor model for studying human lung epithelial development and differentiation, and uses scRNA-seq to benchmark both on- and off-target cell types present in the cultures. Practically, this model presents a novel opportunity to expand the downstream efforts of lung organoid studies, where iBTOs can be shared to reduce the expertise needed for lung cell generation and ultimately facilitate faster turnaround through iBTO banking. Overall, this study enhances the utility of iPSC-derived lung organoids to interrogate lung development and to study disease and regeneration.

## MATERIALS AND METHODS

### Tissue processing, staining and quantification

All sectioned fluorescence images were taken using a Nikon A1 confocal microscope, an Olympus IX83 inverted fluorescence microscope or an Olympus IX71 inverted fluorescence microscope. Whole-mount fluorescent images were taken on a Nikon X1 Yokogawa Spinning Disk Microscope. Acquisition parameters were kept consistent for images in the same experiment and all post-image processing was performed equally on all images in the same experiment. Images were assembled in Adobe Photoshop CC 2022.

### Tissue processing

Tissue was immediately fixed in 10% neutral buffered formalin (NBF) for 24 h at room temperature on a rocker. Tissue was then washed three times in UltraPure DNase/RNase-Free Distilled Water (Thermo Fisher, 10977015) for 15 min each and then dehydrated in an alcohol series of concentrations diluted in UltraPure DNase/RNase-Free distilled water for 1 h per solution: 25% methanol, 50% methanol, 75% methanol, 100% methanol, 100% ethanol, 70% ethanol. Dehydrated tissue was then processed into paraffin wax blocks using an automated tissue processor (Leica ASP300) with 1 h solution changes. Sections at 4  $\mu\text{m}$  (fluorescence *in situ* hybridization) or 7  $\mu\text{m}$  (immunofluorescence) were cut from paraffin blocks onto charged glass slides (within 1 week of performing fluorescence *in situ* hybridization). For fluorescence *in situ* hybridization, microtome and slides were sprayed with RNase Away (Thermo Fisher, 700511) before sectioning. Slides were baked for 1 h in 60°C dry oven (within 24 h of performing fluorescence *in situ* hybridization). Slides were stored at room temperature in a slide box containing a silica desiccator packet and the seams sealed with Parafilm wrap.

### Immunofluorescence protein staining on paraffin wax sections

Tissue slides were rehydrated in Histo-Clear II (National Diagnostics, HS-202) twice for 5 min each, followed by serial rinses through the following solutions twice for 2-3 min each: 100% ethanol, 95% ethanol, 70% ethanol

and 30% ethanol, and finally in double-distilled water (ddH<sub>2</sub>O) twice for 5 min each. Antigen retrieval was performed by steaming slides in 1×sodium citrate buffer [100 mM trisodium citrate (Sigma, S1804), 0.5% Tween 20 (Thermo Fisher, BP337), pH 6.0] for 20 min, followed by cooling and washing quickly twice in ddH<sub>2</sub>O and twice in 1×PBS. Slides were incubated in a humidified chamber at room temperature for 1 h with blocking solution (5% normal donkey serum (Sigma, D9663) in PBS with 0.1% Tween 20). Slides were then incubated in primary antibody diluted in blocking solution at 4°C overnight in a humidified chamber. Next, slides were washed three times in 1×PBS for 5 min each and incubated with secondary antibody with DAPI (1  $\mu\text{g}/\text{ml}$ ) diluted in blocking solution for 1 h at room temperature in a humidified chamber. Slides were then washed three times in 1×PBS for 5 min each and mounted with ProLong Gold (Thermo Fisher, P36930) and imaged within 2 weeks. Stained slides were stored in the dark at 4°C. All primary and secondary antibody concentrations are listed in Table S2. Immunofluorescence and fluorescence *in situ* hybridization stains were repeated on at least three independent differentiations and representative images are shown.

### Fluorescence *in situ* hybridization

The fluorescence *in situ* hybridization protocol was performed according to the manufacturer's instructions (ACDbio, RNAscope multiplex fluorescent manual) with a 5-min protease treatment and 15-min antigen retrieval. For immunofluorescence co-staining with antibodies, the last step of the fluorescence *in situ* hybridization protocol was omitted and instead the slides were washed once in PBS followed by the immunofluorescence protocol above, starting at the blocking step. A list of probes and reagents can be found in Table S2. Immunofluorescence and fluorescence *in situ* hybridization stains were repeated on at least three independent differentiations and representative images are shown.

### Whole-mount immunofluorescence protein staining

All tips and tubes were coated with 1% BSA in PBS to prevent tissue sticking. 3D cultures (spheroids, LPOs) were dislodged from Matrigel using a P1000 cut tip and transferred to a 1.5 ml microcentrifuge tube. 500  $\mu\text{l}$  of Cell Recovery Solution (Corning, 354253) was added to the tube and the tube was placed on a rocker at 4°C for 45 min to completely dissolve Matrigel. Tube was spun at 100  $g$  for 5 min, and solution with remaining Matrigel was then removed. Tissue was fixed in 10% NBF overnight at room temperature on a rocker. Tissue then was washed three times for 2 h with 1 ml organoid wash buffer (OWB) (0.1% Triton, 0.2% BSA in 1×PBS) at room temperature on a rocker. Wash times varied (30 min–2 h) depending on tissue size. 1 ml CUBIC-L (TCI Chemicals, T3740) was added to the tube, and the tube was placed on a rocker for 24 h at 37°C. Tissue was then permeabilized for 24 h at 4°C on a rocker with 1 ml permeabilization solution (5% normal donkey serum, 0.5% Triton in 1×PBS). After 24 h, permeabilization solution was removed and 500  $\mu\text{l}$  primary antibody (diluted in OWB) was added overnight at 4°C on a rocker. The next day, tissue was washed three times with 1 ml of OWB, for 2 h each at room temperature. Secondary antibody (500  $\mu\text{l}$  diluted in OWB) was added and incubated overnight at 4°C, wrapped in foil. Tissue was washed again three times with 1 ml OWB at room temperature: first wash for 2 h then 30 min for the remaining washes. Samples were transferred to imaging plate (ThermoFisher, 12-566-70) and then cleared and mounted with 50  $\mu\text{l}$  CUBIC-R (only enough to cover tissue) (TCI Chemicals, T3741). Immunofluorescence and fluorescence *in situ* hybridization stains were repeated on at least three independent differentiations and representative images are shown.

### Cell lines and culture conditions

#### hPSC lines and culture conditions

##### hPSC lines

LPOs were generated from three human induced pluripotent stem cell (iPSC) lines: WTC11 (RRID: CVCL\_Y803) was obtained from Bruce Conklin at the University of California San Francisco (USA) (Kreitzer et al., 2013), human iPSC line 72.3 was obtained from Cincinnati Children's Hospital Medical Center (USA) (McCracken et al., 2014) and iPSC17 WT 7B2, which expresses NKX2-1-EGFP and TP63-mCherry, was obtained from the Cystic Fibrosis Foundation (Therapeutics Lab, Lexington, MA, USA) (Crane et al.,



2015). In addition to iPSC lines, human embryonic stem cell (hESC) lines were used for definitive endoderm and spheroid optimizations: hESC line H9 (NIH registry no. 0062) was obtained from the WiCell Research Institute and hESC line RUES2-GLR (NIH registry number 0013), expressing SOX2-mCitrine, BRA-mCerulean and SOX17-tdTomato, was obtained from The Rockefeller University (New York, USA) (Martyn et al., 2018). The University of Michigan Human Pluripotent Stem Cell Research Oversight Committee approved all experiments using hESC and iPSC lines. Stem cells were maintained as previously described (Spence et al., 2011) and grown in mTeSR Plus medium (Stem Cell Technologies, 100-0276).

#### NKX2-1-optimized spheroid differentiation protocol

Generation of definitive endoderm (DE) from hPSCs (differentiation days 1–3) was carried out as previously described with slight modifications (D'Amour et al., 2006; McCracken et al., 2011; Spence et al., 2011; Dye et al., 2015). Briefly, 100 ng/ml activin A (R&D Systems, 338-AC) was added in RPMI 1640 medium (Thermo Fisher, 21875034) with increasing concentrations of HyClone defined FBS (dFBS) (Thermo Fisher, SH3007002) on subsequent days (0% day 1, 0.2% day 2 and 2% day 3). 50 ng/ml BMP4 (R&D Systems, 314-BP) was added on day 1. After DE specification, anterior foregut spheroids were generated (differentiation days 4–6) by a 3-day treatment with 500 ng/ml FGF4 (lab purified – see supplementary Materials and Methods), 200 ng/ml noggin (R&D Systems, 6057-NG) and 2  $\mu$ M CHIR99021 (APEX-BIO, A3011). All trans retinoic acid (2  $\mu$ M; Sigma, R2625) was added on the final day of anterior foregut spheroid generation (day 6). On day 7, self-organizing 3D spheroids that had detached from the tissue culture dish were collected with a P200 pipette and transferred into Matrigel (Corning, 354234) as previously described (McCracken et al., 2011; Miller et al., 2019). After the Matrigel had solidified, encapsulating spheroids, they were cultured in medium containing 250 ng/ml WNT3a (R&D Systems, 5036-WN), 500 ng/ml RSP01 (lab purified – see supplementary Materials and Methods) and 10 ng/ml BMP4 for 3 days to induce NKX2-1<sup>+</sup> cells (differentiation days 7–9). After DE specification (after day 3), RPMI 1640 medium+2% dFBS was used on all days to dilute growth factors. Fresh media and growth factors were added each day.

#### Growth and maintenance of LPOs

NKX2-1-optimized spheroids were transferred to '3 Factor' (3F) bud tip maintenance medium as previously described (Miller et al., 2018, 2019), including 50 nM all-trans retinoic acid (Sigma, R2625), 10 ng/ml FGF7 (R&D Systems, 251-KG) and 3  $\mu$ M CHIR99021 (APEX-BIO, A3011) in serum-free basal medium. Serum-free basal medium consists of DMEM/F12 containing HEPES and L-Glutamine (Corning, 10-092-CV), 100 U/ml penicillin-streptomycin (Thermo Fisher, 15140122), 1 $\times$  B-27 supplement (Thermo Fisher, 17504044), 1 $\times$  N-2 supplement (Thermo Fisher, 17502048), 0.05% BSA (Sigma, A9647), 50  $\mu$ g/ml L-ascorbic acid (Sigma, A4544) and 0.4  $\mu$ M 1-thioglycerol (Sigma, M1753). LPOs were grown for 3 weeks, then whole passaged or needle or pipette sheared every 2–4 weeks. Whole passaging was achieved by collecting LPOs into a 1.5 ml microcentrifuge tube and gently releasing them from Matrigel using a P1000 cut pipette tip (P200 tip was used for young/small LPOs). LPOs were spun in a microcentrifuge tube, residual media and Matrigel were removed, then LPOs were re-suspended in Matrigel (Corning, 354234) with a P1000 cut pipette tip. Approximately 35  $\mu$ l droplets of Matrigel were placed into the center of wells of a 24-well tissue culture plate (Thermo Fisher, 12565163), and the plate was inverted and placed in an incubator at 37°C for 20 min. Whole passaging was performed approximately every 2–3 weeks at a ratio of 1:2 (individual LPOs were kept whole but each were given more space, i.e. *n* spheroids were given twice the space to grow) for up to 17 weeks. For LPOs passaged by needle or pipette shearing, LPOs were passed through a 27-gauge needle or P200 pipette, respectively, and embedded in fresh Matrigel as previously described (Miller et al., 2018, 2019, 2020). LPOs were fed with 3F medium every 2–4 days.

#### Growth and maintenance of iBTOs

iBTPs in LPOs grown in 3F medium for 3–17 weeks were isolated by FACS (see below). After collection, iBTPs were centrifuged at 300 *g* for 5 min at

4°C and medium was removed. Matrigel (Corning, 354234) was added to the cells at a concentration of 5000 cells/ $\mu$ l Matrigel and 3–20  $\mu$ l droplets of Matrigel were placed into the center of wells of a 24-well tissue culture plate (Thermo Fisher, 12565163). The plate was inverted and placed in an incubator at 37°C for 20 min. iBTPs were fed every 2–4 days with 3F medium and allowed to reform organoids (iBTOs) for up to 4 weeks. iBTOs were whole passaged as described above every 2–3 weeks after organoid formation.

#### Airway and alveolar differentiations

Airway differentiation was carried out as previously described (Miller et al., 2020). Briefly, iBTOs were exposed to dual-SMAD activation via 100 ng/ml BMP4 (R&D Systems, 314-BP-050) and 100 ng/ml TGF $\beta$ 1 (R&D Systems, 240-B-002) in 3F medium (described above) for 3 days. On the fourth day, iBTOs were exposed to dual-SMAD inactivation via 1  $\mu$ M A8301 (APEX-BIO, 3133), 100 ng/ml noggin (R&D Systems, 6057), 10  $\mu$ M Y-27632 (APEX-BIO, B1293) and 500 ng/ml FGF10 (lab purified – see supplementary Materials and Methods) in serum-free basal medium for 18 days (media changed every 3–4 days) with whole passaging as necessary.

Alveolar differentiation was carried out as previously described (Jacob et al., 2017). Briefly, iBTOs were transitioned to alveolar differentiation medium for 7 days (media changed on days 3 and 6). Alveolar differentiation medium consists of a modified serum-free basal medium (SFB-VA) with DMEM/F12 containing HEPES and L-glutamine (Corning, 10-092-CV) supplemented with 1 $\times$  N-2 supplement (Thermo Fisher, 17502048), 1 $\times$  B-27 supplement without vitamin A (Thermo Fisher, 12587010), 0.05% BSA (Sigma, A9647), 100 U/ml penicillin-streptomycin (Thermo Fisher, 15140122), 50  $\mu$ g/ml L-ascorbic acid (Sigma, A9647) and 0.4  $\mu$ M 1-thioglycerol (Sigma, M6145). On the first day of alveolar differentiation, SFB-VA was supplemented with 10 ng/ml FGF7 (R&D Systems, 251-KG/CF), 3  $\mu$ M CHIR99021 (APEX-BIO, A3011), 100  $\mu$ M 3-Isobutyl-1-methylxanthine (Sigma, I5879), 100  $\mu$ M 8-bromoadenosine 3', 5' -cyclic monophosphate sodium salt (Sigma, B7780) and 50 nM dexamethasone (Sigma, D4802).

#### iBTO growth rate experiments

iBTPs in LPOs grown in 3F medium for 4–6 weeks (early timepoints) or 10–11 weeks (late timepoints) were isolated by FACS (see below). After collection, iBTPs were centrifuged at 300 *g* for 5 min at 4°C and media was removed. Matrigel (Corning, 354234) was added to the cells at a concentration of 5000 cells/ $\mu$ l Matrigel and 8–15  $\mu$ l droplets (droplet size was kept consistent per batch) of Matrigel were placed into the center of wells of a 24-well tissue culture plate (Thermo Fisher, 12565163). The plate was inverted and placed in an incubator at 37°C for 20 min. iBTPs were fed every 2–4 days with 3F medium. At timepoints 2, 4 and 6 weeks post-FACS, iBTOs from one original well of plated iBTPs were collected. iBTOs were removed from Matrigel using a P1000 and/or a P200 pipette tip and vigorously pipetted in a 1.5 ml tube to remove as much Matrigel as possible. Tissue was centrifuged at 300 *g* for 5 min at 4°C, then excess media and Matrigel were removed. Tissue was digested to single cells using 250–500  $\mu$ l TrypLE (Invitrogen, 12605010), depending on pellet size, and incubated at 37°C for 5–15 min, adding mechanical digestion with a pipette every 5 min, until a single cell suspension was reached. Trypsinization was quenched with DMEM/F-12 (Corning, 10-092-CV)+10  $\mu$ M Y-27632 (APEX-BIO, B1293). Cells were centrifuged at 300 *g* for 5 min at 4°C, then liquid was removed. Cells were resuspended in 100–500  $\mu$ l 1 $\times$  PBS+10  $\mu$ M Y-27632 and counted using a hemocytometer.

#### Organoid forming efficiency assay

After dissociation for iBTO growth rate experiments described above, at the 4-week timepoint, iBTPs were centrifuged at 300 *g* for 5 min at 4°C, then media were removed. Matrigel (Corning, 354234) was added to the cells at a concentration of 1000 cells/ $\mu$ l for late timepoints or 25,000 cells/ $\mu$ l for early timepoints (organoids did not form at 1000 cells/ $\mu$ l for early timepoints), and 20  $\mu$ l droplets of Matrigel were placed into the center of wells of a 24-well tissue culture plate (Thermo Fisher, 12565163). The plate was inverted and placed in an incubator at 37°C for 20 min. iBTPs were fed every 2–

4 days with 3F medium. Two weeks later, cultures were imaged for counting the number of cysts. Counting was carried out using the count tool in Adobe Photoshop CC 2022.

#### Culture media, growth factors and small molecules

Additional information for media, growth factors and small molecules, and the in-house expression and purification of human recombinant proteins (FGF4, RSPO1 and FGF10), can be found in the supplementary Materials and Methods.

#### Tissue prep for scRNA-seq

All tubes and pipette tips were pre-washed in 1× HBSS with 1% BSA to prevent cell adhesion to the plastic. 3D cultures (spheroids, LPOs, iBTOs) were removed from Matrigel using a P1000 pipette tip and vigorously pipetted in a 1.5 ml microcentrifuge tube to remove as much Matrigel as possible. Tissue was centrifuged at 300 g for 3 min at 4°C, then excess media and Matrigel were removed. Tissue was digested to single cells using 0.5 ml TrypLE (Invitrogen, 12605010) and incubated at 37°C for 30 min with mechanical digestion with a pipette applied every 10 min. After 30 min, trypsinization was quenched with 1× HBSS+1% BSA. Cells were passed through a 40 µm filter (Bel-Art Flowmi, 136800040) and centrifuged at 300 g for 3 min at 4°C. Cells were resuspended in 1 ml 1× HBSS+1% BSA and counted using a hemocytometer, centrifuged at 300 g for 3 min at 4°C, and resuspended to a final concentration of 1100 cells/µl. If samples were planned for combined submission, cells would be cryopreserved in CryoStor CS10 (Biolife Solutions, 210102). If cryopreservation was used, ice-cold CryoStor CS10 solution was added to the cells and mixed thoroughly then transferred to a cryovial. The cells were incubated at 2–8°C for 10 min, cryopreserved in an isopropanol freezing container for 24 h and then transferred to liquid nitrogen. Cells were thawed in a 37°C water bath, washed in DMEM/F12 (Corning, 10-092-CV) with 10% FBS, centrifuged at 300 g for 5 min, and supernatant was removed. The pellet was then washed in HBSS+1% BSA, then resuspended in 1 ml fresh HBSS+1% BSA, passed through a 40 µm filter (Bel-Art Flowmi, 136800040) and counted using a hemocytometer. Cells were centrifuged at 300 g for 3 min at 4°C and resuspended to a final concentration of 1100 cells/µl. Approximately 100,000 fresh or thawed cells were put on ice and single cell libraries were immediately prepared at the 10X Chromium at the University of Michigan Sequencing Core with a target of 10,000 cells per sample.

#### RNA extraction, cDNA and qRT-PCR

Each analysis includes three biological replicates from three separate differentiation attempts, as well as three technical replicates. mRNA was isolated using the MagMAX-96 Total RNA Isolation Kit (Thermo Fisher, AM1830) (airway and alveolar differentiations) or the PicoPure RNA Isolation Kit (Thermo Fisher, KIT0204) (spheroids). RNA quality and yield were measured on a Nanodrop 2000 spectrophotometer immediately before cDNA synthesis. cDNA synthesis was performed using 100 ng RNA per sample with the SuperScript VILO cDNA Kit (Thermo Fisher, 11754250). qRT-PCR was performed on a Step One Plus Real-Time PCR System (Thermo Fisher, 42765592R) using QuantiTect SYBR Green PCR Kit (Qiagen, 204145). Primer sequences can be found in Table S2. Gene expression as a measure of arbitrary units was calculated relative to GAPDH using the following equation:

$$2^{-(GAPDH_{ct} - Gene_{ct})} \times 10,000.$$

#### Bioinformatics/scRNA-seq

##### Overview

To visualize distinct cell populations within the single cell RNA sequencing dataset, we employed the recommended workflow outlined by the Seurat 4.0 R package (Hao et al., 2021). This pipeline includes the following steps: filtering cells for quality control by applying the SCTransform technique (Hafemeister and Satija, 2019) in place of traditional log normalization; variable gene selection and scaling; identifying anchors and integrating if multiple single cell RNA samples are involved (Stuart et al., 2019); reducing dimensionality with principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) (McInnes et al., 2018

preprint; Becht et al., 2019); clustering by either the Louvain algorithm (Blondel et al., 2008) or the Leiden algorithm (Traag et al., 2019); and log normalization on RNA assay for final visualization and for differential gene expression analysis.

#### Sequencing data and processing FASTQ reads into gene expression matrices

All single cell RNA sequencing was performed at the University of Michigan Advanced Genomics Core with an Illumina Novaseq 6000. The 10X Genomics Cell Ranger pipeline was used to process raw Illumina base calls (BCLs) into gene expression matrices. BCL files were demultiplexed to trim adaptor sequences and unique molecular identifiers (UMIs) from reads. Each sample was then aligned to the human reference genome (hg19) to create a filtered feature bar code matrix that contains only the detectable genes for each sample.

#### Quality control

To ensure quality of the data, all samples were filtered to remove cells expressing too few or too many genes (Fig. 1F,G, Fig. S1G and Fig. S4E-G, <500 and >8000; Fig. 2G, Fig. S2E-H and Fig. S3A, <500 and >7000; Fig. 4A-D and Fig. S4B,C, <200 and >9500), with too low or too high UMI counts (Fig. 1F,G and Fig. S1G, <500, >50,000; Fig. 2G, Fig. S2E-H and Fig. S3A, <500 and >50,000; Fig. 4A-D and Fig. S4B-C, <200 and >50,000; Fig. S4E-G, <500 and >60,000) or a fraction of mitochondrial genes greater than 0.1. Following the above steps, a total of 1067 cells and 36,601 genes (Fig. 1F,G and Fig. S1G), 9133 cells and 36,601 genes (Fig. 2G, Fig. S2E-H and Fig. S3A), 4334 cells and 36,602 genes (Fig. 4A-D and Fig. S4B,C); and 10,888 cells and 34,738 genes (Fig. S4E-G) were kept for downstream analysis and visualization.

#### SCTransform and integration

Seurat's SCTransform method allows efficient pre-processing, normalization and variance stabilization of molecular count data from scRNA-seq samples. Running this algorithm will reveal a model of technical noise in the scRNA-seq data through 'regularized negative binomial regression', the residuals of which are returned as the SCTransform-normalized values that can be used for further downstream analysis such as dimension reduction. During the SCTransform process, we also chose to regress out a confounding source of variation: mitochondrial mapping percentage. When dealing with one sample (Fig. 1F,G and Fig. S1G), there is no batch effect. But when multiple samples are present (Fig. 2G, Fig. S2E-H, Fig. S3A, Fig. 4A-D, Fig. S4B,C and Fig. S4E-G), we have noticed certain amounts of batch effects when clustering data, owing to technical artifacts such as timing of data acquisition or differences in dissociation protocol. To mitigate these effects, we chose to follow Seurat's integration workflow due to its optimal efficiency in harmonizing large datasets. The two methods used are integration on SCTransform-normalized datasets (Fig. 2G, Fig. S2E-H, Fig. S3A, Fig. 4A-D and Fig. S4B,C) and integration on Log-normalized datasets (Fig. S4E-G). After completion of such batch correction, cell clustering should no longer be driven by technical artifacts.

#### Dimension reduction and clustering

Principal component analysis (PCA) was conducted on the corrected expression matrix as follows. Using the top principal components, a neighborhood graph was calculated for the 20 nearest neighbors (Fig. 1F,G and Fig. S1G, 20 principal components) or the 30 nearest neighbors (Fig. 2G, Fig. S2E-H, Fig. S3A, Fig. 4A-D, Fig. S4B,C and Fig. S4E-G, 30 principal components). The UMAP algorithm was then applied for visualization in two dimensions. Using the Leiden algorithm, clusters were identified with a resolution of 0.2 (Fig. 1F,G, Fig. S1G, Fig. 4A-D, Fig. S4B,C and Fig. S4E-G). Using the Louvain algorithm, clusters were identified with a resolution of 0.08 (Fig. 2G, Fig. S2E-H and Fig. S3A).

#### Cluster annotation

Using canonically expressed gene markers, the general cell identity of each cluster was annotated. Cell identities (with markers) include epithelial (*EPCAM*, *KRT18*, *KRT8* and *CLDN6*), mesenchymal (*POSTN*, *DCN*, *COL1A2* and *COL3A1*), neuronal (*S100B*, *STMN2*, *ELAVL4* and *ASCL1*),

endothelial (*ESAM*, *CDH5*, *CLDN5* and *KDR*), proliferative (*MKI67*, *TOP2A* and *CDKI*), primordial germ cell (*POU5F1*, *NANOG*, *TBXT*, *NANOS3* and *TFAP2C*), and foregut mesodermal (*ISL1*, *HAND1*, *BMP4*, *FOXF1* and *LEF1*).

### Cell scoring

Gene lists for cell scoring (Fig. 4E/ Fig. S4D) are found in Table S1 and application of cell scoring strategy is as previously described (Holloway et al., 2020; Hein et al., 2022). Briefly, cells were scored based on expression of a set of 100 marker genes per cell type. Gene lists were compiled by analyzing previously published data from human fetal lung (Miller et al., 2020; Hein et al., 2022). Clusters were first identified by major cell classes (epithelium, mesenchyme, neuronal, endothelium and immune) then the epithelium was sub-clustered to identify bud tip progenitor, basal and neuroendocrine cell clusters by visualizing canonical marker gene expression for each respective cell type. In the case of bud tip progenitor and basal cells, clusters were again sub-clustered to identify the clusters with enriched bud tip progenitor or basal cell marker expression, respectively. Setting the rest of the epithelial cell population as the comparison group, the top 100 differentially expressed genes from bud tip progenitor, basal or neuroendocrine cell clusters were defined as the gene sets for cell scoring. See Table S1 for gene lists. After obtaining the scaled expression values for the dataset, scores for each cell were calculated with the *AddModuleScore* function of Seurat. Cell scores were visualized by violin plots or feature plots.

### Normalization for visualization and differential gene expression

As recommended by Seurat developers, we employed the method of log normalization on the standard RNA assay for graphing dot plots and feature plots, and conducting DGEs. Expression matrix read counts per cell were normalized by the total expression, multiplied by a scale factor of 10,000, and finally log transformed. For the differential gene expression testing, we tested only features that are, first, detected in a minimum fraction of 0.25 in either of the two cell populations and, second, show at least 0.25-fold difference in log-scale between the two cell populations on average.

### Quantification of *KI67*<sup>+</sup> cells

First, the cells in our LPO and iBTO data were grouped by their sample origin or cluster assignment. Then, within each group, we established the frequency table of *KI67* expression based on *KI67* RNA count of an individual cell, and all cells with expression > 0 were quantified as *KI67*<sup>+</sup>. To calculate the contribution of each cluster to *KI67*<sup>+</sup> cells in the data, the number of *KI67*<sup>+</sup> cells in each cluster was divided by the total number of cells in the given cluster ('normalized value') and then calculated as a percentage of the total normalized values of all the clusters. To calculate the contribution of each sample to *KI67*<sup>+</sup> cells per cluster, a 'normalization ratio' for each cluster per sample was created by dividing the number of cells each sample contributed to each cluster by the total number of cells in the given sample. The normalization ratio was then used to recalculate a 'normalized' number of *KI67*<sup>+</sup> cells in each cluster by multiplying the ratio by the number of *KI67*<sup>+</sup> cells in each cluster by sample. The percentage of contribution of each sample to *KI67*<sup>+</sup> cells of a given cluster was then calculated by dividing the number of normalized *KI67*<sup>+</sup> cells from each sample in each cluster by the total number of normalized *KI67*<sup>+</sup> cells in the given cluster.

### Label transfer and UMAP projection

Seurat's single cell reference mapping pipeline allows efficient projection between scRNA-seq datasets (Stuart et al., 2019). The log-normalized integration-prepared primary fetal lung epithelial data (Fig. S4E) were set as reference, while the log-normalized PCA-prepared iBTOs samples were the query data in our analysis. Subsequently, a set of anchors between the reference and query objects were obtained based on PCA projection and were later used to transfer the reference UMAP structure and cluster annotation. Finally, visualization of query data was achieved based on its predicted UMAP representation and cluster labels (Fig. S4F,G).

### Quantification and statistical analysis

Graphs and statistical analysis for qRT-PCR and FACS quantification were performed in GraphPad Prism Software. See figure legends for the number of replicates used, statistical test performed and the *P*-values used to determine the significance for each analysis.

### Fluorescence-activated cell sorting and flow cytometry

#### hPSC flow cytometry

1 ml Accutase (Sigma, A6964) was added to each well of hPSC cultures in a 6-well plate and incubated at 37°C for 5-10 min, until cells detached. An equal volume of mTeSR Plus medium (StemCell Technologies, 100-0276) with 10 μM Y-27632 (APExBIO, B1293) was added to cells and cells were dissociated mechanically by pipetting with a P1000 pipette twice then centrifuged at 300 *g* for 5 min at 4°C. Excess medium was removed and cells were resuspended in fluorescence-activated cell sorting (FACS) buffer (1× PBS containing 2% BSA, 10 μM Y-27632 and 100 U/ml penicillin-streptomycin). Cells were passed through a 70 μm cell strainer, pre-coated with FACS buffer, and centrifuged at 300 *g* for 5 min at 4°C. Cells were resuspended in 1 ml FACS buffer and transferred to 5 ml FACS tubes (Corning, 352063). DAPI (0.2 μg/ml) was added to respective tubes. Flow cytometry was performed using a Bio Rad Ze5#3 and accompanying software.

#### 3D culture sorting

3D cultures (for LPOs, iBTOs and airway organoids) were removed from Matrigel using a P1000 pipette tip and vigorously pipetted in a 15 ml conical tube to remove as much Matrigel as possible. Tissue was centrifuged at 300 *g* for 3 min at 4°C, then excess media and Matrigel were removed. Tissue was digested to single cells using 2-4 ml TrypLE (Invitrogen, 12605010), depending on pellet size and incubated at 37°C for 30 min, adding mechanical digestion with a pipette every 10 min. After 15 min, DNase I (Qiagen, 79254) was added to the digestion at 7.5 μl/ml TrypLE. After 30 min, trypsinization was quenched with DMEM/F-12 (Corning, 10-092-CV)+10 μM Y-27632 (APExBIO, B1293). Cells were passed through a 70 μm cell strainer, pre-coated with DMEM/F-12+10 μM Y-27632 and centrifuged at 500 *g* for 5 min at 4°C. Cells were resuspended in 4 ml FACS buffer (2% BSA, 10 μM Y-27632, 100 U/ml penicillin-streptomycin) and transferred into 5 ml FACS tubes (Corning, 352063). Cells were centrifuged again at 300 *g* for 3 min at 4°C, then resuspended in 1 ml FACS buffer and counted. 10<sup>5</sup> cells were placed into new FACS tubes for all controls (no antibody, DAPI only and individual antibodies/fluorophores) and all remaining cells were centrifuged and resuspended in FACS buffer at a concentration of 10<sup>6</sup> cells/100 μl. Primary antibodies were incubated for 30 min on ice. FACS buffer (3 ml) was added to each tube after 30 min and tubes were centrifuged at 300 *g* for 3 min at 4°C. Cells were washed again with 3 ml FACS buffer and centrifuged at 300 *g* for 3 min at 4°C. Secondary antibodies were incubated for 30 min on ice. FACS buffer (3 ml) was added to each tube after 30 min and tubes were centrifuged at 300 *g* for 3 min at 4°C. Cells were washed again with 3 ml FACS buffer and centrifuged at 300 *g* for 3 min at 4°C. Cells were resuspended in FACS buffer and 0.2 μg/ml DAPI was added to respective tubes. FACS was performed using a Sony MA900 cell sorter and accompanying software. Cells were collected in 1 ml 3F medium+10 μM Y-27632. All primary and secondary antibody concentrations are listed in Table S2.

### Acknowledgements

We thank the University of Michigan Advanced Genomics core for the operation of the 10X Chromium single cell capture platform, the University of Michigan Microscopy core for providing access to confocal microscopes and the Flow Cytometry core for providing access to flow cytometers.

### Competing interests

J.R.S. holds intellectual property rights pertaining to lung organoid technologies.

### Author contributions

Conceptualization: R.F.C.H., A.S.C., J.R.S.; Methodology: R.F.C.H., A.S.C., A.S.F., C.J.C., Y.-H.T., E.M.H.; Software: Z.X.; Validation: R.F.C.H., A.S.C., A.S.F., T.F.; Formal analysis: R.F.C.H., A.S.C., Z.X., T.F., Y.-H.T.; Investigation: R.F.C.H.,



A.S.C., A.S.F., T.F., L.K.B., M.A.A., C.J.C., Y.-H.T.; Resources: S.H., J.M., I.H.; Data curation: Z.X.; Writing - original draft: R.F.C.H., A.S.C., Z.X., S.H., J.R.S.; Writing - review & editing: R.F.C.H., A.S.C., A.S.F., T.F., C.J.C., Y.-H.T., J.M., I.H., J.R.S.; Visualization: R.F.C.H., A.S.C., T.F., C.J.C., Y.-H.T.; Supervision: J.R.S.; Funding acquisition: R.F.C.H., A.S.C., J.R.S.

### Funding

J.R.S. is supported by the Cystic Fibrosis Foundation Epithelial Stem Cell Consortium, by the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation (CZF2019-002440), and by the National Heart, Lung, and Blood Institute (NHLBI; R01HL119215). R.F.C.H. was supported by a National Institutes of Health Tissue Engineering and Regenerative Medicine Training Grant (NIH-NIDCR; T32DE007057) and by a Ruth L. Kirschstein Predoctoral Individual National Research Service Award (NIH-NHLBI; F31HL152531). A.S.C. was supported by the T32 Michigan Medical Scientist Training Program (5T32GM007863-40) and by a Ruth L. Kirschstein Predoctoral Individual National Research Service Award from the National Institutes of Health (NIH-NHLBI; F30HL156474). T.F. was supported by a National Institutes of Health Tissue Engineering and Regenerative Medicine Training Grant (NIH-NIDCR; T32DE007057). E.M.H. was supported by a Ruth L. Kirschstein Predoctoral Individual National Research Service Award from the National Institutes of Health (NIH-NHLBI; F31HL146162). Open Access funding provided by the University of Michigan Medical School. Deposited in PMC for immediate release.

### Data availability

Sequencing data used and generated in this study are available in ArrayExpress. Single cell RNA sequencing of human fetal lung and human fetal bud tip progenitor organoids: human fetal lung (E-MTAB-8221 and E-MTAB-10662) (Miller et al., 2020; Hein et al., 2022); and spheroids, LPOs and iBTOs (E-MTAB-11953) (this study). Code used to process our data can be found at [https://github.com/jason-spence-lab/Hein\\_Conchola\\_2022](https://github.com/jason-spence-lab/Hein_Conchola_2022)

### Peer review history

The peer review history is available online at <https://journals.biologists.com/dev/lookup/doi/10.1242/dev.200693.reviewer-comments.pdf>

### References

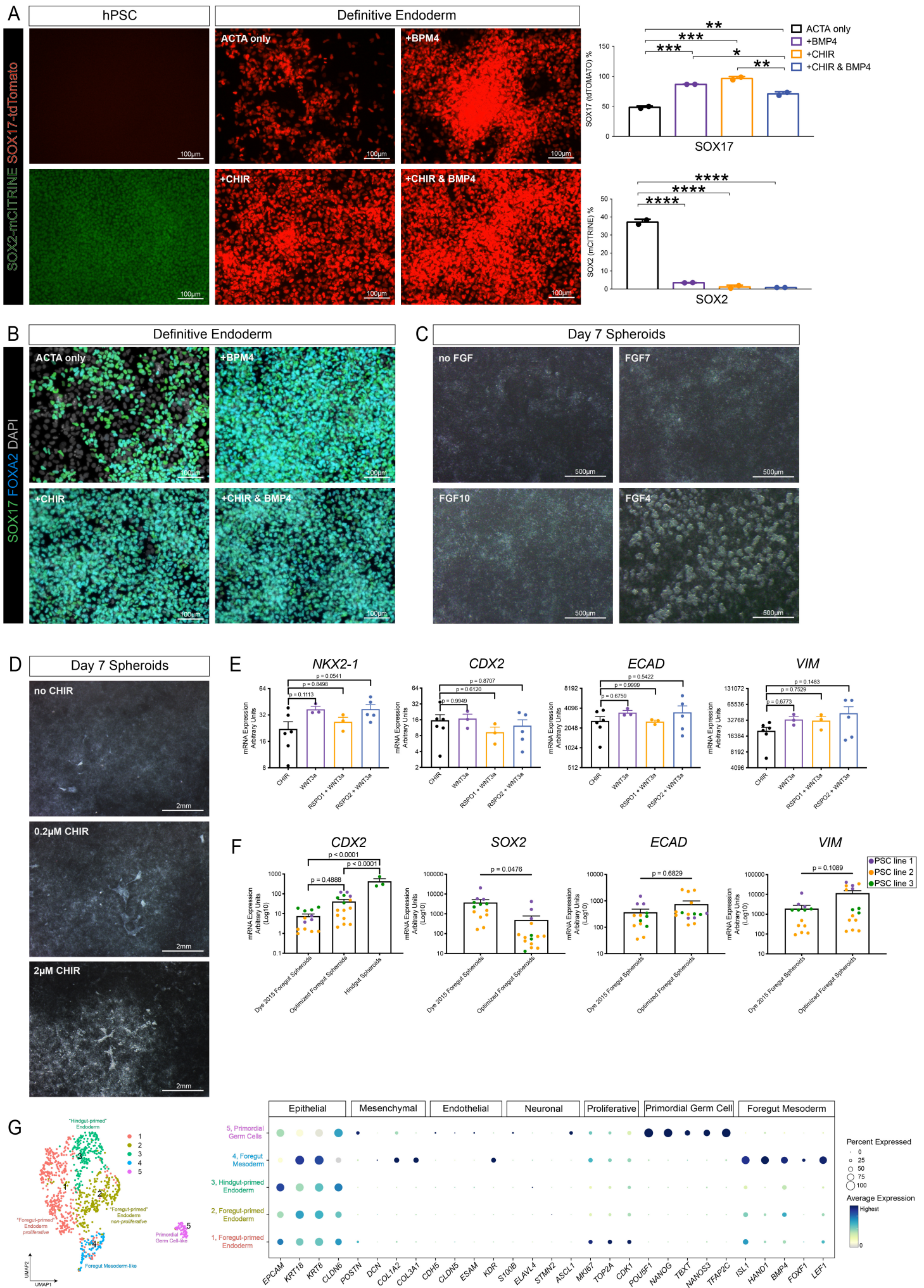
- Abler, L. L., Mansour, S. L. and Sun, X.** (2008). Conditional gene inactivation reveals roles for Fgf10 and Fgfr2 in establishing a normal pattern of epithelial branching in the mouse lung. *Dev. Dyn.* **238**, 1999-2013. doi:10.1002/dvdy.22032
- Alanis, D. M., Chang, D. R., Akiyama, H., Krasnow, M. A. and Chen, J.** (2014). Two nested developmental waves demarcate a compartment boundary in the mouse lung. *Nat. Commun.* **5**, 3923. doi:10.1038/ncomms4923
- Alysandratos, K.-D., De Alba Rivas, C. G., Yao, C., Pessina, P., Villacorta-Martin, C., Huang, J., Hix, O. T., Minakin, K., Konda, B., Stripp, B. R. et al.** (2022). Impact of cell culture on the transcriptomic programs of primary and iPSC-derived human alveolar type 2 cells. *bioRxiv*, 2022.02.08.479591. doi:10.1101/2022.02.08.479591
- Aros, C. J., Pantoja, C. J. and Gomperts, B. N.** (2021). Wnt signaling in lung development, regeneration, and disease progression. *Commun. Biol.* **4**, 601. doi:10.1038/s42003-021-02118-w
- Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W. H., Ng, L. G., Ginhoux, F. and Newell, E. W.** (2019). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38-44. doi:10.1038/nbt.4314
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E.** (2008). Fast unfolding of communities in large networks. *J. Stat. Mech., Theory Exp.* **2008**, P10008. doi:10.1088/1742-5468/2008/10/P10008
- Bogue, C. W., Ganea, G., Sturm, E., Zhao, F. and Jacobs, H. C.** (1998). Hex expression during development suggests an important role in both gastrulation and organogenesis. *Pediatr. Res.* **43**, 45-45. doi:10.1203/00006450-199804001-00272
- Carraro, G., Langerman, J., Sabri, S., Lorenzana, Z., Purkayastha, A., Zhang, G., Konda, B., Aros, C. J., Calvert, B. A., Szymaniak, A. et al.** (2021). Transcriptional analysis of cystic fibrosis airways at single-cell resolution reveals altered epithelial cell states and composition. *Nat. Med.* **27**, 806-814. doi:10.1038/s41591-021-01332-7
- Chen, Y.-W., Huang, S. X., De Carvalho, A. L. R. T., Ho, S.-H., Islam, M. N., Volpi, S., Notarangelo, L. D., Ciancanelli, M., Casanova, J.-L., Bhattacharya, J. et al.** (2017). A three-dimensional model of human lung development and disease from pluripotent stem cells. *Nat. Cell Biol.* **19**, 542-549. doi:10.1038/ncb3510
- Chin, A. M., Tsai, Y.-H., Finkbeiner, S. R., Nagy, M. S., Walker, E. M., Ethen, N. J., Williams, B. O., Battle, M. A. and Spence, J. R.** (2016). A dynamic WNT/ $\beta$ -CATENIN signaling environment leads to WNT-Independent and WNT-dependent proliferation of embryonic intestinal progenitor cells. *Stem Cell Rep.* **7**, 826-839. doi:10.1016/j.stemcr.2016.09.004
- Conway, R. F., Frum, T., Conchola, A. S. and Spence, J. R.** (2020). Understanding human lung development through in vitro model systems. *BioEssays* **42**, 2000006. doi:10.1002/bies.202000006
- 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. doi:10.1016/j.stemcr.2015.02.005
- D'Amour, K. A., Bang, A. G., Eliazer, S., Kelly, O. G., Agulnick, A. D., Smart, N. G., Moorman, M. A., Kroon, E., Carpenter, M. K. and Baetge, E. E.** (2006). Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat. Biotechnol.* **24**, 1392-1401. doi:10.1038/nbt1259
- Davenport, C., Diekmann, U., Budde, I., Detering, N. and Naujok, O.** (2016). Anterior-posterior patterning of definitive endoderm generated from human embryonic stem cells depends on the differential signaling of retinoic acid, Wnt-, and BMP-signaling. *Stem Cells* **34**, 2635-2647. doi:10.1002/stem.2428
- De Carvalho, A. L. R. T., Strikoudis, A., Liu, H.-Y., Chen, Y.-W., Dantas, T. J., Vallee, R. B., Correia-Pinto, J. and Snoeck, H.-W.** (2019). Glycogen synthase kinase 3 induces multilineage maturation of human pluripotent stem cell-derived lung progenitors in 3D culture. *Development (Cambridge, England)* **146**, dev171652. doi:10.1242/dev.171652
- 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 (Cambridge, England)* **138**, 971-981. doi:10.1242/dev.053694
- Dye, B. R., Hill, D. R., Ferguson, M. A. H., 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. doi:10.7554/eLife.05098
- Dye, B. R., Dedhia, P. H., Miller, A. J., Nagy, M. S., White, E. S., Shea, L. D. and Spence, J. R.** (2016). A bioengineered niche promotes in vivo engraftment and maturation of pluripotent stem cell derived human lung organoids. *eLife* **5**, e19732. doi:10.7554/eLife.19732
- Gadue, P., Huber, T. L., Paddison, P. J. and Keller, G. M.** (2006). Wnt and TGF- $\beta$  signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **103**, 16806-16811. doi:10.1073/pnas.0603916103
- 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. doi:10.1016/j.devcel.2009.06.005
- Gotoh, S., Ito, I., Nagasaki, T., Yamamoto, Y., Konishi, S., Korogi, Y., Matsumoto, H., Muro, S., Hirai, T., Funato, M. et al.** (2014). Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells. *Stem Cell Rep.* **3**, 394-403. doi:10.1016/j.stemcr.2014.07.005
- 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. doi:10.1038/nbt.1788
- Hafemeister, C. and Satija, R.** (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296. doi:10.1186/s13059-019-1874-1
- Han, L., Chaturvedi, P., Kishimoto, K., Koike, H., Nasr, T., Iwasawa, K., Giesbrecht, K., Witcher, P. C., Eicher, A., Haines, L. et al.** (2020). Single cell transcriptomics identifies a signaling network coordinating endoderm and mesoderm diversification during foregut organogenesis. *Nat. Commun.* **11**, 4158. doi:10.1038/s41467-019-13993-7
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M. et al.** (2021). Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29. doi:10.1016/j.cell.2021.04.048
- 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. doi:10.1073/pnas.0902274106
- 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. doi:10.1172/JCI89950
- Hawkins, F. J., Suzuki, S., Beermann, M. L., Barillà, C., Wang, R., Villacorta-Martin, C., Berical, A., Jean, J. C., Le Suer, J., Matte, T. et al.** (2021). Derivation of airway basal stem cells from human pluripotent stem cells. *Cell Stem Cell* **28**, 79-95.e8. doi:10.1016/j.stem.2020.09.017
- Heemskerk, I., Burt, K., Miller, M., Chhabra, S., Guerra, M. C., Liu, L. and Warmflash, A.** (2019). Rapid changes in morphogen concentration control self-organized patterning in human embryonic stem cells. *eLife* **8**, e40526. doi:10.7554/ELIFE.40526
- Hein, R. F. C., Wu, J. H., Holloway, E. M., Frum, T., Conchola, A. S., Tsai, Y.-H., Wu, A., Fine, A. S., Miller, A. J., Szenker-Ravi, E. et al.** (2022). R-SPONDIN2+ mesenchymal cells form the bud tip progenitor niche during human lung development. *Dev. Cell* **57**, 1598-1614.e8. doi:10.1016/j.devcel.2022.05.010

- Holloway, E. M., Wu, J. H., Czerwinski, M., Sweet, C. W., Wu, A., Tsai, Y.-H., Huang, S., Stoddard, A. E., Capeling, M. M., Glass, I. et al. (2020). Differentiation of human intestinal organoids with endogenous vascular endothelial cells. *Dev. Cell* **54**, 516-528.e7. doi:10.1016/j.devcel.2020.07.023
- Hor, P., Punj, V., Calvert, B. A., Castaldi, A., Miller, A. J., Carraro, G., Stripp, B. R., Brody, S. L., Spence, J. R., Ichida, J. K. et al. (2020). Efficient generation and transcriptomic profiling of human iPSC-derived pulmonary neuroendocrine cells. *iScience* **23**, 101083. doi:10.1016/j.isci.2020.101083
- Huang, S. X. L., 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. doi:10.1038/nbt.2754
- Hurley, K., Ding, J., Villacorta-Martin, C., Herriges, M. J., Jacob, A., Vedaie, M., Alysandratos, K. D., Sun, Y. L., Lin, C., Werder, R. B. et al. (2020). Reconstructed single-cell fate trajectories define lineage plasticity windows during differentiation of human PSC-derived distal lung progenitors. *Cell Stem Cell* **26**, 593-608.e8. doi:10.1016/j.stem.2019.12.009
- Jacob, A., Morley, M., Hawkins, F., Mccauley, K. B., Jean, J. C., Heins, H., Na, C.-L., Weaver, T. E., Vedaie, M., Hurley, K. et al. (2017). Differentiation of human pluripotent stem cells into functional lung alveolar epithelial cells. *Cell Stem Cell* **21**, 472-488.e10. doi:10.1016/j.stem.2017.08.014
- Jacob, A., Vedaie, M., Roberts, D. A., Thomas, D. C., Villacorta-Martin, C., Alysandratos, K.-D., Hawkins, F. and Kotton, D. N. (2019). Derivation of self-renewing lung alveolar epithelial type II cells from human pluripotent stem cells. *Nat. Protoc.* **14**, 3303-3332. doi:10.1038/s41596-019-0220-0
- Jacobs, I. J., Ku, W.-Y. and Que, J. (2012). Genetic and cellular mechanisms regulating anterior foregut and esophageal development. *Dev. Biol.* **369**, 54-64. doi:10.1016/j.ydbio.2012.06.016
- Jo, K., Teague, S., Chen, B., Khan, H. A., Freeburne, E., Li, H., Li, B., Ran, R., Spence, J. R. and Heemskerk, I. (2022). Efficient differentiation of human primordial germ cells through geometric control reveals a key role for Nodal signaling. *eLife* **11**, e72811. doi:10.7554/eLife.72811
- Kapoor, A., Li, H. J. and Leiter, A. B. (2007). Intestinal development: the many faces of Wnt signaling. *Gastroenterology* **133**, 710-712. doi:10.1053/j.gastro.2007.06.029
- Kearns, N. A., Genga, R. M. J., Ziller, M., Kapinas, K., Peters, H., Brehm, M. A., Meissner, A. and Maehr, R. (2013). Generation of organized anterior foregut epithelia from pluripotent stem cells using small molecules. *Stem Cell Res.* **11**, 1003-1012. doi:10.1016/j.scr.2013.06.007
- Konishi, S., Gotoh, S., Tateishi, K., Yamamoto, Y., Korogi, Y., Nagasaki, T., Matsumoto, H., Muro, S., Hirai, T., Ito, I. et al. (2016). Directed induction of functional multi-ciliated cells in proximal airway epithelial spheroids from human pluripotent stem cells. *Stem Cell Rep.* **6**, 18-25. doi:10.1016/j.stemcr.2015.11.010
- Kreitzer, F. R., Salomonis, N., Sheehan, A., Huang, M., Park, J. S., Spindler, M. J., Lizarraga, P., Weiss, W. A., So, P.-L. and Conklin, B. R. (2013). A robust method to derive functional neural crest cells from human pluripotent stem cells. *Am. J. Stem Cell* **2**, 119-131. Available at: /pmc/articles/PMC3708511/ (Accessed: 15 February 2022).
- Kubo, A., Shinozaki, K., Shannon, J. M., Kouskoff, V., Kennedy, M., Woo, S., Fehling, H. J. and Keller, G. (2004). Development of definitive endoderm from embryonic stem cells in culture. *Development (Cambridge, England)* **131**, 1651-1662. doi:10.1242/dev.01044
- Kuwahara, A., Lewis, A. E., Coombes, C., Leung, F.-S., Percharde, M. and Bush, J. O. (2020). Delineating the early transcriptional specification of the mammalian trachea and esophagus. *eLife* **9**, e55526. doi:10.7554/eLife.55526
- Lazzaro, D., Price, M., De Felice, M. and Di Lauro, R. (1991). The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. *Development (Cambridge, England)* **113**, 1093-1104. Available at: doi:10.1242/dev.113.4.1093
- Leibel, S. L., Mccicar, R. N., Winkquist, A. M., Niles, W. D. and Snyder, E. Y. (2020). Generation of complete multi-cell type lung organoids from human embryonic and patient-specific induced pluripotent stem cells for infectious disease modeling and therapeutics validation. *Curr. Protoc. Stem Cell Biol.* **54**, e118. doi:10.1002/cpsc.118
- Li, S., Morley, M., Lu, M. M., Zhou, S., Stewart, K., French, C. A., Tucker, H. O., Fisher, S. E. and Morrisey, E. E. (2016). Foxp transcription factors suppress a non-pulmonary gene expression program to permit proper lung development. *Dev. Biol.* **416**, 338-346. doi:10.1016/j.ydbio.2016.06.020
- Little, D. R., Gerner-Mauro, K. N., Flodby, P., Crandall, E. D., Borok, Z., Akiyama, H., Kimura, S., Ostrin, E. J. and Chen, J. (2019). Transcriptional control of lung alveolar type 1 cell development and maintenance by NK homeobox 2-1. *Proc. Natl Acad. Sci. USA* **116**, 20545-20555. doi:10.1073/pnas.1906663116
- Loh, K. M., Ang, L. T., Zhang, J., Kumar, V., Ang, J., Auyeong, J. Q., Lee, K. L., Choo, S. H., Lim, C. Y. Y., Nichane, M. et al. (2014). Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell* **14**, 237-252. doi:10.1016/j.stem.2013.12.007
- Martyn, I., Kanno, T. Y., Ruzo, A., Siggia, E. D. and Brivanlou, A. H. (2018). Self-organization of a human organizer by combined Wnt and Nodal signalling. *Nature* **558**, 132-135. doi:10.1038/s41586-018-0150-y
- Matsuno, K., Mae, S.-I., Okada, C., Nakamura, M., Watanabe, A., Toyoda, T., Uchida, E. and Osafune, K. (2016). Redefining definitive endoderm subtypes by robust induction of human induced pluripotent stem cells. *Differentiation* **92**, 281-290. doi:10.1016/j.diff.2016.04.002
- Mccauley, K. B., Hawkins, F., Serra, M., Thomas, D. C., Jacob, A. and Kotton, D. N. (2017). Efficient derivation of functional human airway epithelium from pluripotent stem cells via temporal regulation of Wnt signaling. *Cell Stem Cell* **20**, 844-857.e6. doi:10.1016/j.stem.2017.03.001
- Mccauley, K. B., Alysandratos, K.-D., Jacob, A., Hawkins, F., Caballero, I. S., Vedaie, M., Yang, W., Slovik, K. J., Morley, M., Carraro, G. et al. (2018). Single-cell transcriptomic profiling of pluripotent stem cell-derived SCGB3A2+ airway epithelium. *Stem Cell Rep.* **10**, 1579-1595. doi:10.1016/j.stemcr.2018.03.013
- McCracken, K. W., Howell, J. C., Wells, J. M. and Spence, J. R. (2011). Generating human intestinal tissue from pluripotent stem cells in vitro. *Nat. Protoc.* **6**, 1920-1928. doi:10.1038/nprot.2011.410
- McCracken, K. W., Catá, E. M., Crawford, C. M., Sinagoga, K. L., Schumacher, M., Rockich, B. E., Tsai, Y.-H., Mayhew, C. N., Spence, J. R., Zavros, Y. et al. (2014). Modeling human development and disease in pluripotent stem cell-derived gastric organoids. *Nature* **516**, 400. doi:10.1038/nature13863
- McInnes, L., Healy, J. and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*. arXiv. Available at: <http://arxiv.org/abs/1802.03426> (Accessed: 13 January 2021).
- Miller, M. F., Cohen, E. D., Baggs, J. E., Lu, M. M., Hogenesch, J. B. and Morrisey, E. E. (2012). Wnt ligands signal in a cooperative manner to promote foregut organogenesis. *Proc. Natl. Acad. Sci. USA* **109**, 15348-15353. doi:10.1073/pnas.1201583109
- Miller, A. J., Hill, D. R., Nagy, M. S., Aoki, Y., Dye, B. R., Chin, A. M., Huang, S., Zhu, F., White, E. S., Lama, V. et al. (2018). In vitro induction and in vivo engraftment of lung bud tip progenitor cells derived from human pluripotent stem cells. *Stem Cell Rep.* **10**, 101-119. doi:10.1016/j.stemcr.2017.11.012
- Miller, A. J., Dye, B. R., Ferrer-Torres, D., Hill, D. R., Overeem, A. W., Shea, L. D. and Spence, J. R. (2019). Generation of lung organoids from human pluripotent stem cells in vitro. *Nat. Protoc.* **14**, 518-540. doi:10.1038/s41596-018-0104-8
- Miller, A. J., Yu, Q., Czerwinski, M., Tsai, Y.-H., Conway, R. F., Wu, A., Holloway, E. M., Walker, T., Glass, I. A., Treutlein, B. et al. (2020). In vitro and in vivo development of the human airway at single-cell resolution. *Dev. Cell* **53**, 117-128.e6. doi:10.1016/j.devcel.2020.01.033
- Minoo, P., Hamdan, H., Bu, D., Warburton, D., Stepanik, P. and Delemos, R. (1995). TTF-1 regulates lung epithelial morphogenesis. *Dev. Biol.* **172**, 694-698. doi:10.1006/dbio.1995.8080
- Minoo, P., Su, G., Drum, H., Bringas, P. and Kimura, S. (1999). Defects in tracheoesophageal and lung morphogenesis in Nkx2.1(-/-) mouse embryos. *Dev. Biol.* **209**, 60-71. doi:10.1006/dbio.1999.9234
- Mou, H., Zhao, R., Sherwood, R., Ahfeldt, T., Lapey, A., Wain, J., Sicilian, L., Izvolsky, K., Lau, F. H., Musunuru, K. 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. doi:10.1016/j.stem.2012.01.018
- Nikolić, M. Z. and Rawlins, E. L. (2017). Lung organoids and their use to study cell-cell interaction. *Curr. Pathobiol. Rep.* **5**, 223-231. doi:10.1007/s40139-017-0137-7
- Nikolić, M. Z., Carigt, O., Jeng, Q., Johnson, J.-A., Sun, D., Howell, K. J., Brady, J. L., Laresgoiti, U., Allen, G., Butler, R. et al. (2017). Human embryonic lung epithelial tips are multipotent progenitors that can be expanded in vitro as long-term self-renewing organoids. *eLife* **6**, e26575. doi:10.7554/eLife.26575
- Okubo, T. and Hogan, B. L. M. (2004). Hyperactive Wnt signaling changes the developmental potential of embryonic lung endoderm. *J. Biol.* **3**, 11. doi:10.1186/JBIOL3
- Perl, A.-K. T., Kist, R., Shan, Z., Scherer, G. and Whitsett, J. A. (2005). Normal lung development and function after Sox9 inactivation in the respiratory epithelium. *Genesis (New York, N.Y. : 2000)* **41**, 23-32. doi:10.1002/gene.20093
- Pinto, D. and Clevers, H. (2005). Wnt control of stem cells and differentiation in the intestinal epithelium. *Exp. Cell Res.* **306**, 357-363. doi:10.1016/j.yexcr.2005.02.022
- Que, J., Okubo, T., Goldenring, J. R., Nam, K.-T., Kurotani, R., Morrisey, E. E., Taranova, O., Pevny, L. H. and Hogan, B. L. M. (2007). Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. *Development (Cambridge, England)* **134**, 2521-2531. doi:10.1242/dev.003855
- Rabata, A., Fedr, R., Soucek, K., Hamp, A. and Koledova, Z. (2020). 3D cell culture models demonstrate a role for FGF and WNT signaling in regulation of lung epithelial cell fate and morphogenesis. *Front. Cell Dev. Biol.* **8**, 574. doi:10.3389/fcell.2020.00574
- 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 reports* **16**, 66-78. doi:10.1016/j.celrep.2016.05.060
- Rankin, S. A., Mccracken, K. W., Luedeke, D. M., Han, L., Wells, J. M., Shannon, J. M. and Zorn, A. M. (2018). Timing is everything: Reiterative Wnt, BMP and RA signaling regulate developmental competence during endoderm organogenesis. *Dev. Biol.* **434**, 121-132. doi:10.1016/j.ydbio.2017.11.018

- Rawlins, E. L., Clark, C. P., Xue, Y. and Hogan, B. L. M.** (2009). The Id2+ distal tip lung epithelium contains individual multipotent embryonic progenitor cells. *Development* **136**, 3741-3745. doi:10.1242/dev.037317
- Rockich, B. E., Hrycaj, S. M., Shih, H. P., Nagy, M. S., Ferguson, M. A. H., Kopp, J. L., Sander, M., Wellik, D. M. and Spence, J. R.** (2013). Sox9 plays multiple roles in the lung epithelium during branching morphogenesis. *Proc. Natl Acad. Sci. USA* **110**, E4456-E4464. doi:10.1073/pnas.1311847110
- Serra, R., Pelton, R. W. and Moses, H. L.** (1994). TGF beta 1 inhibits branching morphogenesis and N-myc expression in lung bud organ cultures. *Development* **120**, 2153-2161. doi:10.1242/dev.120.8.2153
- Serra, M., Alysandratos, K.-D., Hawkins, F., Mccauley, K. B., Jacob, A., Choi, J., Caballero, I. S., Vedaie, M., Kurmann, A. A., Ikonou, L. et al.** (2017). Pluripotent stem cell differentiation reveals distinct developmental pathways regulating lung- versus thyroid-lineage specification. *Development (Cambridge, England)* **144**, 3879-3893. doi:10.1242/dev.150193
- Shu, W., Guttentag, S., Wang, Z., Andl, T., Ballard, P., Lu, M. M., Piccolo, S., Birchmeier, W., Whitsett, J. A., Millar, S. E. et al.** (2005). Wnt/ $\beta$ -catenin signaling acts upstream of N-myc, BMP4, and FGF signaling to regulate proximal-distal patterning in the lung. *Dev. Biol.* **283**, 226-239. doi:10.1016/j.ydbio.2005.04.014
- Shu, W., Lu, M. M., Zhang, Y., Tucker, P. W., Zhou, D. and Morrisey, E. E.** (2007). Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development* **134**, 1991-2000. doi:10.1242/dev.02846
- Spence, J. R., Lange, A. W., Lin, S.-C. J., Kaestner, K. H., Lowy, A. M., Kim, I., Whitsett, J. A. and Wells, J. M.** (2009). Sox17 regulates organ lineage segregation of ventral foregut progenitor cells. *Dev. Cell* **17**, 62-74. doi:10.1016/j.devcel.2009.05.012
- Spence, J. R., Mayhew, C. N., Rankin, S. A., Kuhar, M. F., Vallance, J. E., Tolle, K., Hoskins, E. E., Kalinichenko, V. V., Wells, S. I., Zorn, A. M. et al.** (2011). Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature* **470**, 105-109. doi:10.1038/nature09691
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P. and Satija, R.** (2019). Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21. doi:10.1016/j.cell.2019.05.031
- Sun, D., Evans, L., Perrone, F., Sokleva, V., Lim, K., Rezakhani, S., Lutolf, M., Zilbauer, M. and Rawlins, E. L.** (2021). A functional genetic toolbox for human tissue-derived organoids. *eLife* **10**, e67886. doi:10.7554/eLife.67886
- Tamò, L., Hibaoui, Y., Kallol, S., Alves, M. P., Albrecht, C., Hostettler, K. E., Feki, A., Rougier, J.-S., Abriel, H., Knudsen, L. et al.** (2018). Generation of an alveolar epithelial type II cell line from induced pluripotent stem cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **315**, L921-L932. doi:10.1152/ajplung.00357.2017
- Traag, V. A., Waltman, L. and Van Eck, N. J.** (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233. doi:10.1038/s41598-019-41695-z
- Volckaert, T. and De Langhe, S. P.** (2015). Wnt and FGF mediated epithelial-mesenchymal crosstalk during lung development. *Dev. Dyn.* **244**, 342-366. doi:10.1002/dvdy.24234
- Wang, D., Haviland, D. L., Burns, A. R., Zsigmond, E. and Wetsel, R. A.** (2007). A pure population of lung alveolar epithelial type II cells derived from human embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **104**, 4449-4454. doi:10.1073/pnas.0700052104
- Yamamoto, Y., Gotoh, S., Korogi, Y., Seki, M., Konishi, S., Ikeo, S., Sone, N., Nagasaki, T., Matsumoto, H., Muro, S. et al.** (2017). Long-term expansion of alveolar stem cells derived from human iPS cells in organoids. *Nat. Methods* **14**, 1097-1106. doi:10.1038/nmeth.4448
- Yamamoto, Y., Korogi, Y., Hirai, T. and Gotoh, S.** (2020). A method of generating alveolar organoids using human pluripotent stem cells. *Methods Cell Biol.* **159**, 115-141. doi:10.1016/bs.mcb.2020.02.004
- Yan, K. S., Janda, C. Y., Chang, J., Zheng, G. X. Y., Larkin, K. A., Luca, V. C., Chia, L. A., Mah, A. T., Han, A., Terry, J. M. et al.** (2017). Non-equivalence of Wnt and R-spondin ligands during Lgr5+ intestinal stem-cell self-renewal. *Nature* **545**, 238-242. doi:10.1038/nature22313
- Yu, Q., Kilik, U., Holloway, E. M., Tsai, Y.-H., Harmel, C., Wu, A., Wu, J. H., Czerwinski, M., Childs, C., He, Z., Glass, I. A. et al.** (2020). Charting human development using a multi-endodermal organ atlas and organoid models. *Cell*, **184**, 3281-3298.e22. doi:10.1016/j.cell.2021.04.028



Supplemental Figure 1

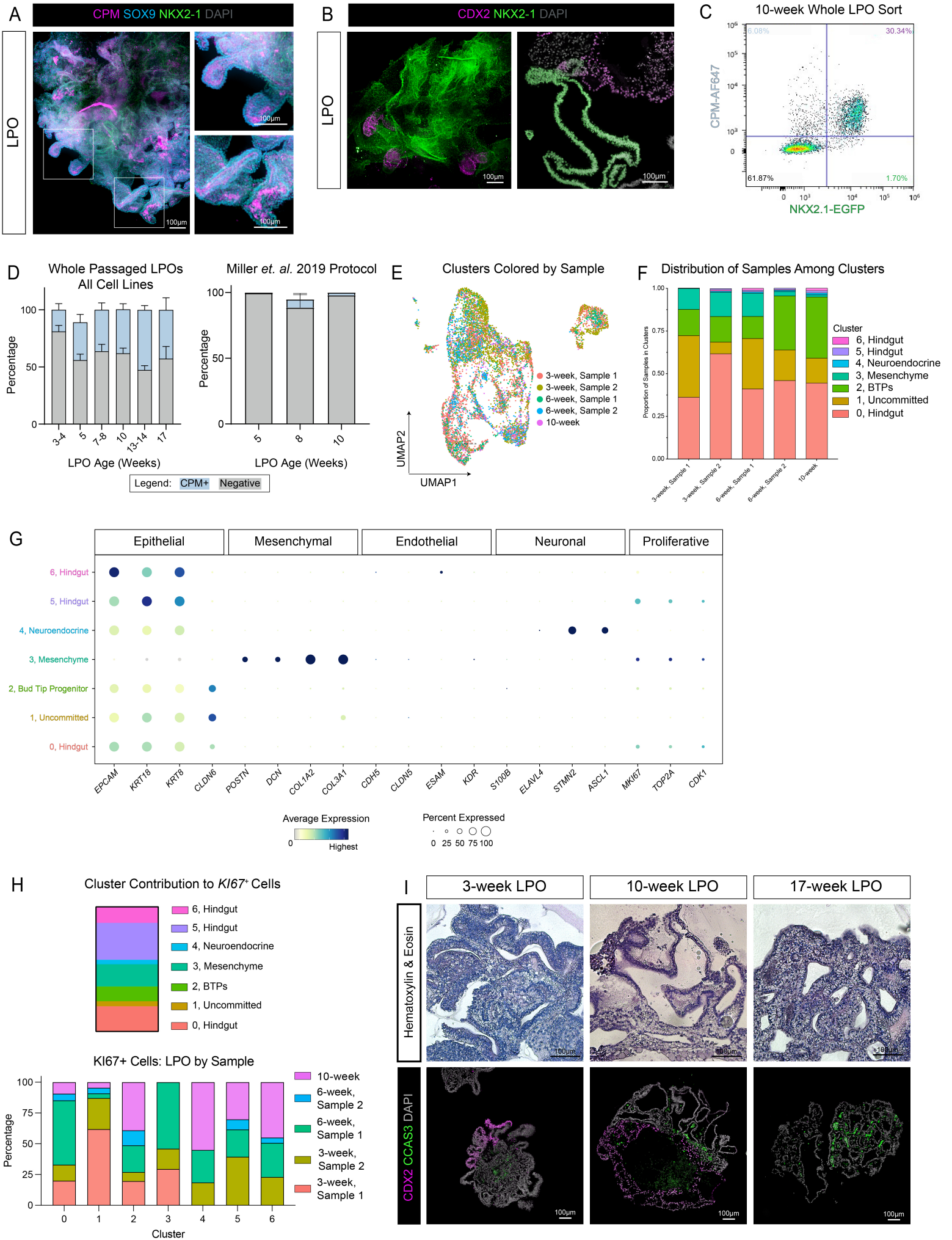


**Fig. S1. Optimization of Directed Differentiation of hPSCs at Definitive Endoderm, Foregut Spheroid and NKX2-1<sup>+</sup> Spheroid Stages, Related to Figure 1**

- (A) (Left) SOX2-mCitrine (pluripotent stem cell marker) and SOX17-tdTomato (definitive endoderm marker) reporter images of SOX2-mCITRINE and SOX17-tdTomato on hPSCs and hPSC-derived day 4 definitive endoderm (DE) where 50ng/mL BMP4 and/or 2 $\mu$ M CHIR were added on day 1 of a 3-day 100ng/mL Activin A (ACTA) treatment to induce DE (see Fig. 1A) in a hPSC line expressing SOX2-mCitrine/SOX17-tdTomato hPSC line (Martyn *et al.*, 2018). (Right) Flow cytometry quantifications of day 4 DE cultures. Two technical replicates from the same experiment were performed for flow cytometry quantifications. Statistical tests were performed by ordinary one-way ANOVA.
- (B) Immunofluorescence staining for the DE markers SOX17 and FOXA2 on day 4 DE testing conditions as noted in S1A. This experiment was performed on two independent hPSC lines different than the reporter line in Fig. S1A. Representative images are shown from H9 ESCs.
- (C) Brightfield images of day 7 spheroids evaluated for FGF-dependent spheroid formation. Conditions include no FGF, 10ng/mL FGF7, 500ng/mL FGF10, or 500ng/mL FGF4 on days 4 – 6.
- (D) Brightfield images of day 7 spheroids evaluated for WNT-dependent spheroid formation. Conditions include no CHIR99021, 0.2 $\mu$ M CHIR99021, or 2 $\mu$ M CHIR99021 on days 4 – 6.
- (E) RT-qPCR data comparing expression of the lung epithelial marker *NKX2-1*, the hindgut epithelial marker *CDX2*, the pan-epithelial marker *ECAD*, and the pan-mesenchymal marker *VIM* when CHIR99021, WNT3a, RSPO1 & WNT3a, or RSPO2 & WNT3a was used to activate WNT signaling on days 7 – 9 of the NKX2-1-optimized spheroid directed differentiation protocol. Data points for CHIR99021 and WNT3a + RSPO2 represent 2 independent experiments with 2 – 3 technical replicates and data points from other conditions represent one independent experiment with 3 technical replicates. All experiments were performed on H9 ESCs. Statistical tests compared all conditions to CHIR99021 and were performed by ordinary one-way ANOVA followed by Dunnett's multiple comparison test.
- (F) RT-qPCR data comparing expression of the hindgut epithelial marker *CDX2*, the posterior foregut marker *SOX2*, the pan-epithelial marker *ECAD*, and pan-mesenchymal marker *VIM* from previously published foregut spheroids (Dye *et al.*, 2015) to optimized foregut spheroids, and in the case of *CDX2*, hindgut spheroids. Each color represents an independent experiment with a unique iPSC line (purple: WTC11, orange: iPSC17 WT 7B2, green: iPSC line 72.3). Each data point of the same color represents a technical replicate from the same iPSC line in one or more independent experiments. Error bars represent standard error of the mean. Statistical test for *CDX2* was performed by ordinary one-way ANOVA followed by Tukey's multiple comparison test and statistical tests for the remaining markers was performed by unpaired Welch's one-tailed t-tests.
- (G) Cluster plot of scRNA-seq data from day 10 spheroids and dot plot of epithelial, mesenchymal, endothelial, and neuronal lineage genes as well as proliferation markers in each cluster of the UMAP. The dot size represents the percentage of cells expressing the gene in the corresponding cluster, and the dot color indicates log-normalized expression level of the gene.



Supplemental Figure 2

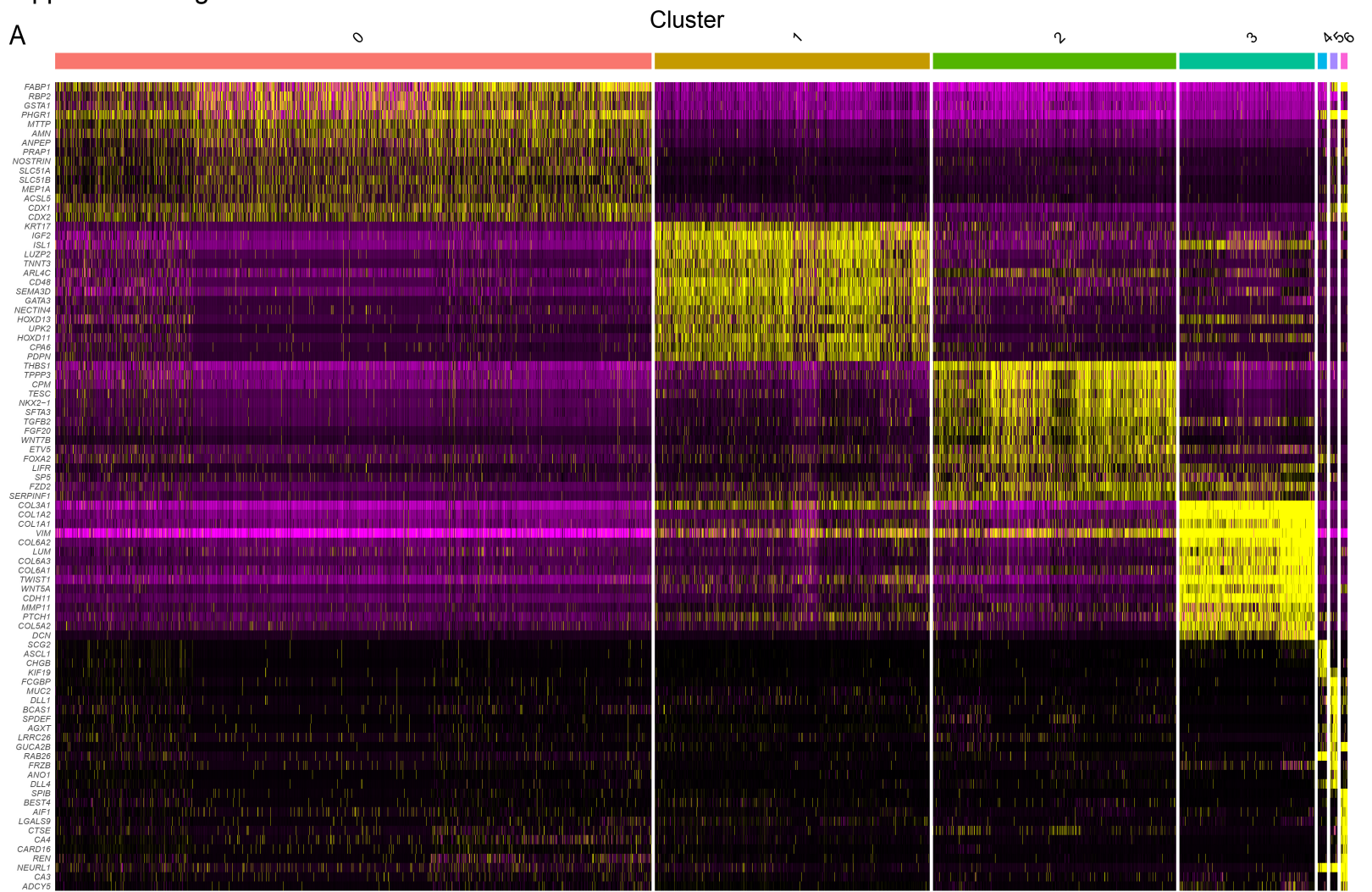




**Fig. S2. Bud Tip Progenitors and Contaminating Lineages Emerge and Expand in LPOs Over Time, Related to Figure 2**

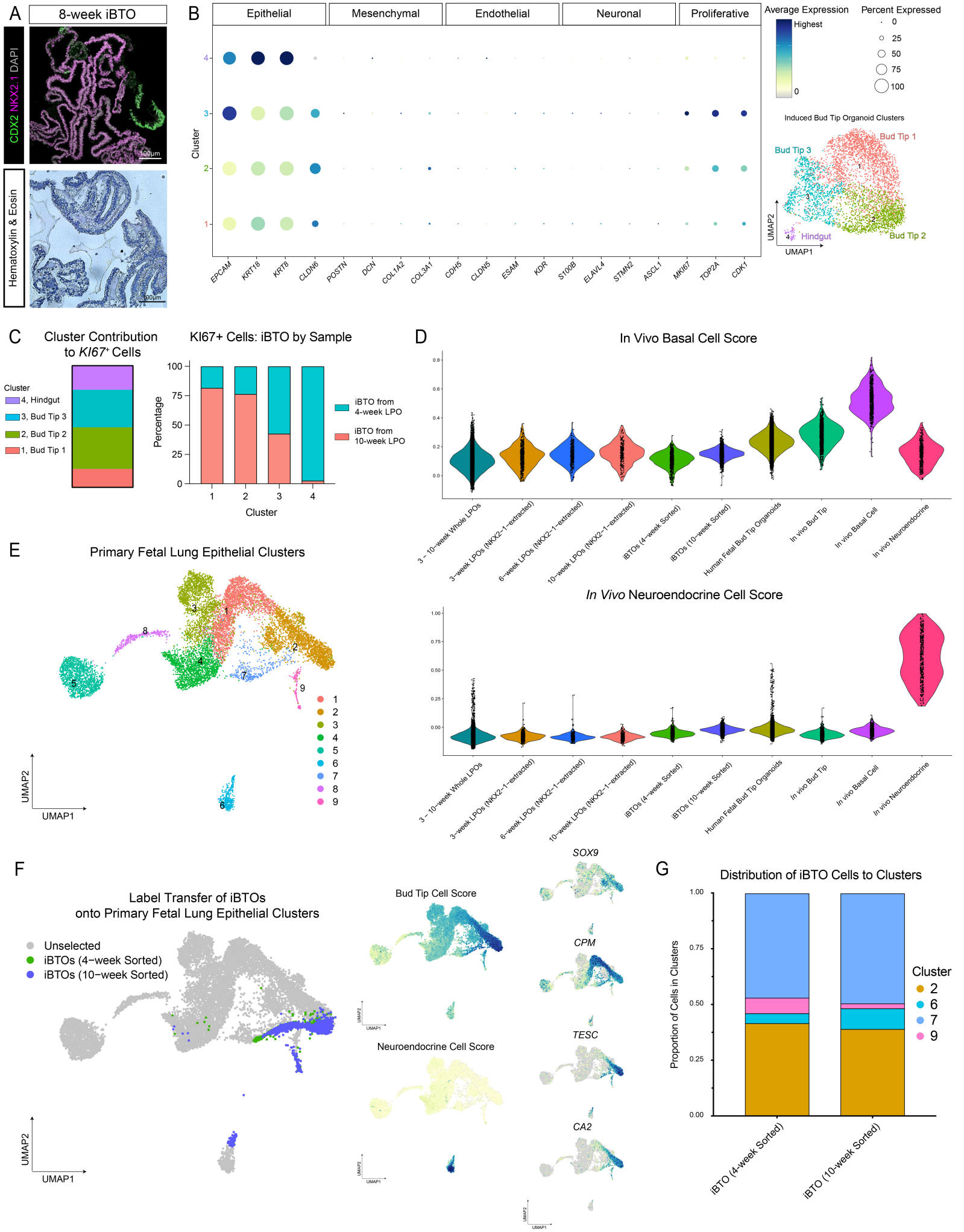
- (A) Maximum intensity projection of a whole mount immunofluorescence confocal Z-series staining of a 10-week LPO for bud tip progenitor markers CPM and SOX9 and lung epithelial marker NKX2-1.
- (B) Maximum intensity projection of a whole mount immunofluorescence confocal Z-series staining of a 10-week LPO (left) or immunofluorescence staining on a paraffin section of 4-week LPOs (right) for intestinal epithelial marker CDX2 and lung epithelial marker NKX2-1.
- (C) Representative flow cytometry plot for iBTP selection based on positive CPM and NKX2-1-EGFP selection on 10-week LPOs.
- (D) (Left) FACS quantification of CPM<sup>+</sup> cells in LPOs in aggregate time course from three separate cell lines, including the NKX2-1-EGFP reporter line (iPSC17 WT 7B2), iPSC line 72.3 and WTC11. LPOs were sorted using CPM from 3 – 17 weeks in 3F media. Percentages of live cells expressing CPM (blue) or negative (grey) are reported as mean ± SEM for 3 – 9 replicates per time point. (Right) FACS quantification of CPM<sup>+</sup> cells in LPOs generated from spheroids using a previously published protocol (Miller *et al.*, 2019) from two separate cell lines, including the NKX2-1-EGFP reporter line and iPSC line 72.3 (WTC11-derived LPOs did not survive to 10 weeks) (Miller *et al.*, 2019).
- (E) UMAP plot corresponding to the LPO cluster plot in Fig. 2G. Each dot represents a single cell and dots/cells are colored by the sample from which they came from.
- (F) Stacked bar graph displaying the proportion of cells from each sample in each cluster of the LPO cluster plot in Fig. 2G.
- (G) Dot plot of epithelial, mesenchymal, endothelial, and neuronal lineage genes as well as proliferation markers in each cluster of the LPO cluster plot in Fig. 2G. The dot size represents the percentage of cells expressing the gene in the corresponding cluster, and the dot color indicates log-normalized expression level of the gene.
- (H) (Top) Stacked bar graph showing each cluster's normalized contribution to KI67<sup>+</sup> cells, calculated from the scRNA-seq data. (Bottom) Stacked bar graph showing each sample's normalized contribution to KI67<sup>+</sup> cells per cluster, calculated from the scRNA-seq data.
- (I) (Top) Hematoxylin & eosin or (Bottom) immunofluorescence staining on paraffin sections of 3-, 10- and 17-week LPOs. IF stains for cell death marker cleaved caspase 3 (CCAS3) and intestinal epithelial marker CDX2.

## Supplemental Figure 3

**Fig. S3. LPOs are Transcriptionally Heterogenous**

(A) Heat map including 12 genes per cluster picked from the top 50 highly enriched genes in each cluster corresponding to the UMAP cluster plot in Fig. 2G. Top-50 genes were based on log-fold change of each gene.

Supplemental Figure 4

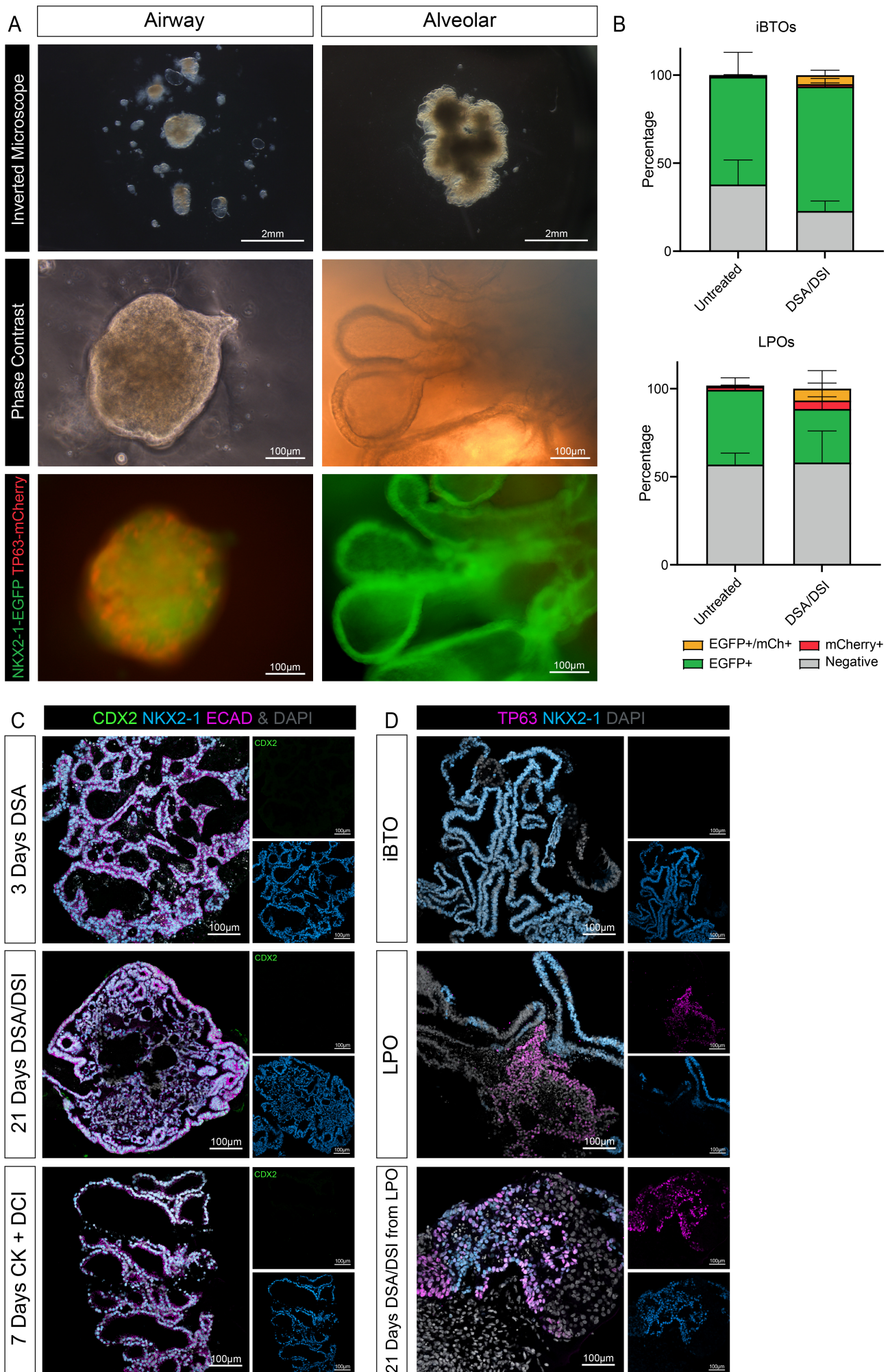




**Fig. S4. Induced Bud Tip Organoids are Enriched for Bud Tip Progenitor Cells, Related to Figures 3 & 4**

- (A) (Top) Immunofluorescence staining on paraffin sections for lung epithelial marker NKX2-1 and intestinal epithelial marker CDX2 or (Bottom) hematoxylin & eosin stain on 8-week iBTOs.
- (B) Dot plot of epithelial, mesenchymal, endothelial, and neuronal lineage genes as well as proliferation markers in each cluster of the iBTO cluster plot in Fig. 4A. The dot size represents the percentage of cells expressing the gene in the corresponding cluster, and the dot color indicates log-normalized expression level of the gene.
- (C) (Left) Stacked bar graph showing each cluster's normalized contribution to Ki67<sup>+</sup> cells, calculated from the scRNA-seq data. (Right) Stacked bar graph showing each sample's normalized contribution to Ki67<sup>+</sup> cells per cluster, calculated from the scRNA-seq data.
- (D) Violin plots displaying an *in vivo* basal (top) or neuroendocrine (bottom) cell score, calculated as the average expression of the top 100 enriched genes in *in vivo* basal or neuroendocrine cells, for each sample. Samples include whole LPOs (n = 2 for 3- and 6-week LPOs, n = 1 for 10-week LPOs), NKX2-1-extracted cells from 3-, 6-, and 10-week LPOs whole iBTOs (derived from LPOs sorted for NKX2-1<sup>+</sup>/CPM<sup>+</sup> cells at 4 or 10 weeks, n=1 of each), human fetal-derived (primary) bud tip organoids, and primary *in vivo* tissue including computationally-extracted bud tip, basal, and neuroendocrine cells (Miller *et al.*, 2020; Hein *et al.*, 2022).
- (E) UMAP cluster plot of previously-published scRNA-seq data from primary fetal lung epithelium spanning 8.5 – 21 weeks post-conception (n = 11 samples from trachea, airway, and distal bud tips) (Miller *et al.*, 2020; Hein *et al.*, 2022). Each dot represents a single cell and cells were computationally clustered based on transcriptional similarities. The plot is colored and numbered by cluster.
- (F) Label transfer (Stuart *et al.*, 2019) of iBTO cells (derived from 4- or 10-week) sequenced via scRNA-seq onto the primary fetal lung epithelial UMAP embedding shown in Figure S4E. Feature plots display either an *in vivo* bud tip progenitor or *in vivo* neuroendocrine cell score, calculated as the average expression of the top 100 enriched genes in *in vivo* bud tip progenitor or neuroendocrine cells respectively, or single bud tip progenitor cell markers (*SOX9*, *CPM*, *TESC*, *CA2*). The color of each dot in the feature plot indicates log-normalized expression level of the individual genes or set of genes in the represented cell.
- (G) Stacked bar graph displaying the percentage of iBTO cells from each timepoint that mapped to clusters 2, 6, 7, or 9 in the UMAP embedding shown in Figure S4E.

Supplemental Figure 5



**Fig. S5. iBTOs Maintain NKX2-1 Expression after Differentiation into Proximal Airway and Distal Alveolar Organoids, Related to Figure 5**

- (A) Representative brightfield, phase contrast, and reporter expression for NKX2-1-EGFP and TP63-mCherry images of iBTOs undergone the airway or alveolar induction protocol shown in Fig. 5A. Phase contrast and reporter images are from the same field of view.
- (B) FACS quantification of NKX2-1-EGFP<sup>+</sup> and mCherry-TP63<sup>+</sup> cells from iBTOs and LPOs from the NKX2-1-EGFP/TP63-mCherry reporter cell line, (iPSC 17 WT 7B2) either in 3F media (untreated) or after 21 days of the airway differentiation protocol (DSA/DSI treatment). Percentages of live cells expressing neither reporter, each separate reporter, or dual-expressing are reported as mean  $\pm$  SEM for 3 – 10 replicates per time point.
- (C) Immunofluorescence staining on paraffin sections for the intestinal epithelial marker CDX2, lung epithelial marker NKX2-1, and general epithelial cell-type marker ECAD on 12-week iBTOs undergone the airway differentiation protocol (DSA/DSI treatment) or 3-week iBTOs undergone the alveolar differentiation protocol (7 days CK + DCI) (see Fig. 5A).
- (D) Immunofluorescence staining on paraffin sections for the airway progenitor/basal cell marker TP63 and lung epithelial marker NKX2-1 on 12-week iBTOs, 4-week LPOs, and 3-week LPOs after 21 days of the airway differentiation protocol (DSA/DSI treatment).



**Table S1. Cell Scoring Gene Lists**

[Click here to download Table S1](#)

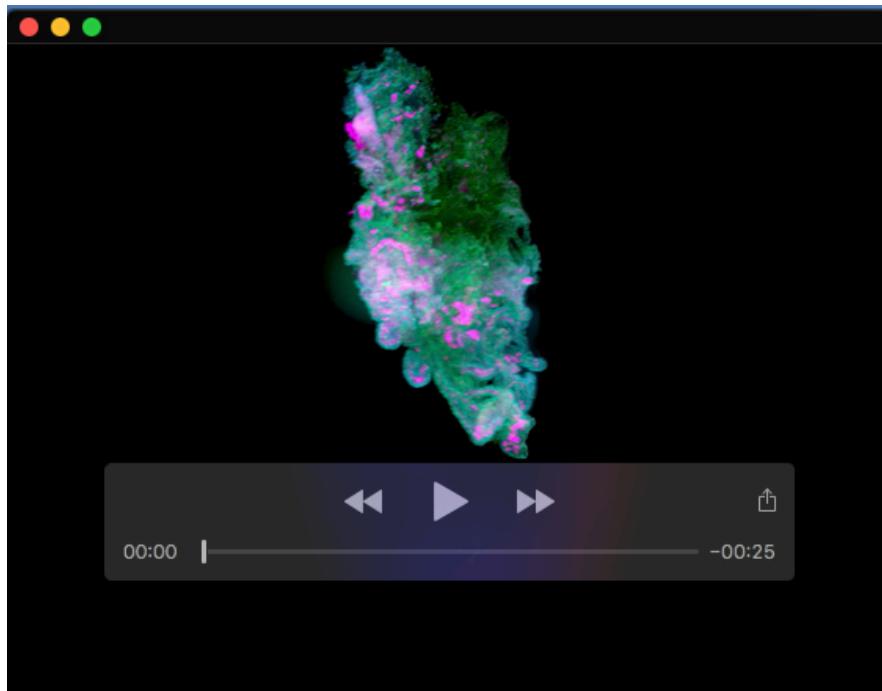
**Table S2. Antibody and TSA Dilutions and Primer Sequences**

Antibody	Dilution
FOXA2 (Seven Hills Bioreagents, Cat#WRAB-1200)	1:500
SOX17 (R&D Systems, Cat#AF1924)	1:500
VIM (R&D Systems, Cat#AFf2105)	1:500
TTF1/NKX2.1 (Abcam, Cat#ab76013)	1:500
SOX9 (Millipore, Cat#AB5535)	1:500
CPM (Wako/FujiFilm, Cat#014-27501)	1:500 for IF; 1:300 for FACS
CDX2 (BioGenex, Cat#MU392A-UC)	1:500
CDX2 (Cell Signaling Technology, Cat#12306) (Used only for co-stain with MUC2)	1:50
MUC2 (Santa Cruz Biotechnology, Cat#sc-515032)	1:300
Cleaved Caspase 3 (CCAS3) (Cell Signaling Technology, Cat#9664)	1:500
ECAD (R&D, Cat#AF748)	1:500
ECAD (BD Biosciences, Cat#610181)	1:500
SOX2 (R&D Systems, Cat#AF2018)	1:500
TP63 (R&D Systems, Cat#BAF1916)	1:500
CHGA (Santa Cruz Biotechnology, Cat#sc-1488)	1:100
FOXJ1 (Seven Hills Bioreagents, Cat#WMAB-319)	1:250
MUC5AC (Abcam, Cat#ab79082)	1:500

CC10/SCGB1A1 (Santa Cruz Biotechnology, Cat#sc-365992)	1:250
Acetylated Tubulin (AcTUB) (Sigma-Aldrich, Cat#T7451)	1:1000
ASCL1 (Santa Cruz Biotechnology, Cat#sc-374104)	1:50
Synaptophysin (SYN) (Abcam, Cat#ab32127)	1:500
SP-C (Santa Cruz Biotechnology, Cat#sc-518029)	1:200
Pro-SP-C (Seven Hills Bioreagents)	1:500
SP-B (Seven Hills Bioreagents, Cat#WMAB-1B9)	1:500
AffiniPure Donkey anti-Mouse IgG-AF488/Cy3/AF647 (Jackson ImmunoResearch, Cat#115-545-003, 115-165-003, 115-605-003)	1:500
AffiniPure Donkey anti-Goat IgG-AF488/Cy3/AF647 (Jackson ImmunoResearch, Cat#705-545-003, 705-165-003, 705-605-003)	1:500
AffiniPure Donkey anti-Rabbit IgG-AF488/Cy3/AF647 (Jackson ImmunoResearch, Cat#111-545-003, 111-165-003, 111-605-003)	1:500
<b>Probe</b>	<b>TSA dilution</b>
Hs-SCGB3A2-C2 (ACDbio RNAscope, Cat#549961-C2)	1:5000
Hs-SFTPb (ACDbio RNAscope, Cat#544251)	1:2500
Hs-MUC16-C2 (ACDbio RNAscope, Cat#40591-C2)	1:4000
<b>Primer</b>	<b>Sequence</b>
<i>NKX2.1</i>	F: CTCATGTTTCATGCCGCTC R: GACACCATGAGGAACAGCG
<i>CDX2</i>	F: GGGCTCTCTGAGAGGCAGTT R: GGTGACGGTGGGGTTTAGCA

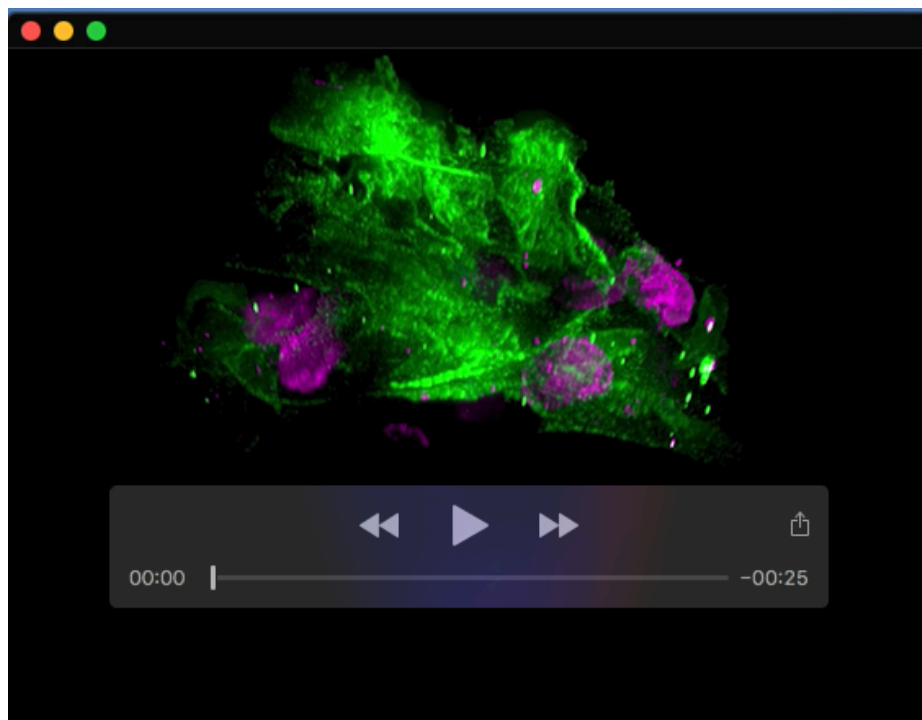
<i>ECAD</i>	F: TTGACGCCGAGAGCTACAC R: GACCGGTGCAATCTTCAAA
<i>VIM</i>	F: CTTCAGAGAGAGGAAGCCGA R: ATTCCACTTTGCGTTCAAGG
<i>SOX2</i>	F: GCTTAGCCTCGTCGATGAAC R: AACCCCAAGATGCACAACCTC
<i>TP63</i>	F: CCACAGTACACGAACCTGGG R: CCGTTCTGAATCTGCTGGTCC
<i>FOXJ1</i>	F: CAACTTCTGCTACTTCCGCC R: CGAGGCACTTTGATGAAGC
<i>CHGA</i>	F: CTGTCCTGGCTCTTCTGCTC R: TGACCTCAACGATGCATTTT
<i>MUC5AC</i>	F: GCACCAACGACAGGAAGGATGAG R: CACGTTCCAGAGCCGGACAT
<i>SCGB1A1</i>	F: ATGAAACTCGCTGTCACCCT R: GTTTCGATGACACGCTGAAA
<i>SCGB3A2</i>	F: GGGGCTAAGGAAGTGTGTAATG R: CACCAAGTGTGATAGCGCCTC
<i>SFTPC</i>	F: AGCAAAGAGGTCCTGATGGA R: CGATAAGAAGGCGTTTCAGG
<i>SFTPB</i>	F: GGGTGTGTGGGACCATGT R: CAGCACTTTAAAGGACGGTGT
<i>ABCA3</i>	F: TGCAGCGCCTACTTGAACCT R: CTGAGCACAGCCATCGTCT





### Movie 1. Bud Tip Progenitors in LPOs

Whole mount immunofluorescence confocal z-series staining of 10-week LPO for bud tip progenitor markers CPM (pink) and SOX9 (blue) and lung epithelial marker NKX2-1 (green).



### Movie 2. Lung and Intestine Regions in LPOs

Whole mount immunofluorescence confocal z-series staining of a 10-week LPO for lung epithelial marker NKX2-1 (green) and intestinal epithelial marker CDX2 (pink).

## Supplementary Materials and Methods

### Expression and Purification of Human Recombinant FGF10

The recombinant human FGF10 (rhFGF10) expression plasmid pET21d-FGF10 was a gift from James A. Bassuk (Bagai *et al.*, 2002) at the University of Washington School of Medicine. This plasmid was transformed to Novagen's Rosetta™ 2(DE3)pLysS competent cells (Millipore Sigma, Cat#71403-3) for rhFGF10 expression. In brief, *E.coli* strain Rosetta™ 2(DE3)pLysS bearing pET21d-FGF10 was grown in 2x YT medium (BD Biosciences, Cat#244020) with Carbenicillin (50 µg/mL) and Chloramphenicol (17 µg/mL). rhFGF10 expression was induced by addition of isopropyl-1-thio-β-D-galactopyranoside (IPTG). rhFGF10 was purified by using a HiTrap-Heparin HP column (GE Healthcare, Cat#17040601) with step gradients of 0.2 M to 0.92 M NaCl. From a 200 mL culture, 3 – 4 mg of at least 98% pure rhFGF-10 (evaluated by SDS-PAGE stained with Coomassie Blue R-250) was purified. In-house purified rhFGF10 was confirmed by western blot analysis using anti-FGF10 antibody and compared to commercially purchased rhFGF10 (R&D Systems, Cat#345-FG) to test/validate activity based on the efficiency to phosphorylate ERK1/2 in an A549 alveolar epithelial cell line (ATCC, Cat#CCL-185) as assessed by western blot analysis.

### Expression and Purification of Human Recombinant FGF4

*E.coli* strain Rosetta™(DE3)pLysS bearing pET28/MhFGF4L was a gift from Masayuki Kobayashi (Sugawara *et al.*, 2014) at Akita Prefectural University, Japan. Culture of the *E.coli* strain was in 2X YT medium with Kanamycin (50 µg/mL) and Chloramphenicol (17 µg/mL). Expression and purification of rhFGF4 was performed as described above for rhFGF10. In-house purified rhFGF4 activity was compared to commercially purchased rhFGF4 (R&D Systems, Cat#235-F4) based on the efficiency to differentiate definitive endoderm cells from hPSCs into CDX2<sup>+</sup> hindgut cells by qRT-PCR and immunofluorescence (McCracken *et al.*, 2011).

### Expression and Purification of Human Recombinant R-Spondin1

The stable 293T cell line expressing HA-R-Spondin1-Fc (Cultrex, Cat#3710-001-01) was cultured and passaged according to the manufacturer's protocol. For expression and purification, after at least five days in selection growth medium (0.3 mg/mL Zeocine), the cells were passaged to Triple Flasks (Nunc, Cat#132913) with 150-200 mL growth medium per flask without selection. The cells were cultured until they reached 90% confluency (~5 days), then medium was changed to CD293 medium (Gibco, Cat#11913-019) containing L-glutamine and Pen/Strep. Confluent cells were maintained in CD293 medium for 7 – 14 days, then the supernatant was collected and filtered for HA-R-Spondin1-Fc purification. The activity of the conditioned medium was determined by performing a TopFlash Luciferase Assay (Biotium, Cat#30085-2) by its ability to induce activation of Wnt/β-catenin. HA-R-Spondin1-Fc was purified using the Fc tag via protein A agarose purification kit (KPL, Cat#553-50-00) per manufacturer's instructions. From 1 L of conditioned medium, ~1.5 mg of at least 95% pure HA-R-Spondin1-Fc was purified. In-house purified HA-R-Spondin1-Fc activity was compared to commercially purchased rhR-Spondin1 (R&D Systems, Cat#4645-RS) based on the efficiency to differentiate CDX2<sup>+</sup> hindgut spheroids from hPSCs to human intestinal organoids (HIO) by qRT-PCR and HIO morphology (McCracken *et al.*, 2011).

### SUPPLEMENTAL-SPECIFIC REFERENCES

Bagai, S. *et al.* (2002) 'Fibroblast growth factor-10 is a mitogen for urothelial cells', *The Journal of biological chemistry*. J Biol Chem, 277(26), pp. 23828–23837. doi: 10.1074/JBC.M201658200.

Sugawara, S. *et al.* (2014) 'Production of an aminotermally truncated, stable type of bioactive mouse fibroblast growth factor 4 in Escherichia coli', *Journal of bioscience and bioengineering*. J Biosci Bioeng, 117(5), pp. 525–530. doi: 10.1016/J.JBIOSEC.2013.10.009.