Generation and expansion of transitional lung organoids from human pluripotent stem cells.

Ivana Matkovic Leko,1,2,6 Nadine Schrode,3,4,6 Juan Torres,1,2 Mikael Pezet,1,2 Tania A. Thimraj,1,2 Kristin G. Beaumont,3,4,6 Hans-Willem Snoeck1,2,5,6

1Columbia Center for Human Development/Center for Stem Cell Therapies, Columbia University Irving Medical Center, New York, NY, USA
2Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA
3Department of Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, USA
4The Center for Advanced Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, USA
5Department of Microbiology and Immunology, Columbia University Irving Medical Center, New York, NY, USA
6These authors contributed equally

Corresponding authors/contacts

Hans-Willem Snoeck, MD, PhD
Byron M Thomashow Professor of Medicine
Director, Center for Stem Cell Therapies
Division of Pulmonary Medicine, Allergy and Critical Care
Departments of Medicine, and Microbiology & Immunology
Columbia University Vagelos College of Physicians and Surgeons
650 W. 168th Street, BB 801E
New York, NY 10032
(T) 212 342 0182
(F) 212 342 0550
hs2680@columbia.edu

Kristin G. Beaumont, PhD
Assistant Professor
Associate Director, Center for Advanced Genomic Technology
Department of Genetics and Genomic Sciences
Icahn School of Medicine at Mount Sinai
1425 Madison Ave.
New York, NY 10029
(T) 212 659 8317
kristin.beaumont@mssm.edu
Several novel distal lung populations were recently identified that may be involved in regeneration after injury.\textsuperscript{1–3} As these are absent in rodents, a deeper understanding of their roles and lineage relations requires availability of equivalent cells \textit{in vitro}. Here we report the generation of expandable, clonal spheres, called ‘transitional lung organoids’ (TLOs), from human pluripotent stem cells. TLOs consist mainly of previously identified type 0 alveolar epithelial (AT0) cells, terminal respiratory bronchiole stem cells and distal basal cells (BCs).\textsuperscript{3} Velocity analysis of single cell RNAseq data suggests that distal BCs are the most undifferentiated progenitors in the TLOs and give rise to AT0 cells. TLOs will be an important resource for studies in human lung regeneration, and potentially for regenerative approaches for human lung disease.

The respiratory epithelium contains basal (BC), ciliated, secretory, goblet and neuroendocrine cells in the airways, and alveolar type 1 (AT1) and surfactant-producing alveolar type 2 (AT2) cells in the alveoli, where gas exchange takes place.\textsuperscript{4,5} Additional complexity was discovered when transitional populations were identified in the terminal bronchioles (also called respiratory airways) of primates and ferrets, but not rodents, that may play a role in regeneration, as they were more abundant after lung injury.\textsuperscript{1–3} A feature of at least some of these populations is the expression of the secretory cell marker, SCGB3A2, first reported by Haberman et al. who found SCGB3A2\textsuperscript{+} cells co-expressing AT1 and AT2 markers in the distal lung.\textsuperscript{1} Basil et al. also reported SCGB3A2\textsuperscript{+} cells negative for the AT2 marker, SFPTC, in respiratory airways.\textsuperscript{2} Murthy et al. identified multiple subpopulations characterized by co-expression of another AT2 marker, SFTP\textsuperscript{B}.\textsuperscript{3} These included SCGB3A2\textsuperscript{+}SFTP\textsuperscript{B}SFTP\textsuperscript{C} ‘AT0’ cells, SFTP\textsuperscript{B}SCGB3A2\textsuperscript{+}SFTP\textsuperscript{C} terminal respiratory bronchiole secretory cells (TRB-SCs), SFTP\textsuperscript{B}SCGB3A2\textsuperscript{+}SCGB1A1\textsuperscript{+} pre-TRB-SCs, and SFTP\textsuperscript{B}SCGB3A2\textsuperscript{+} distal BCs, some of which are low to negative for the airway BC markers, KRT5 and p63. Based on airway organoids derived from human pluripotent stem cells (hPSCs), Basil et al.\textsuperscript{2} proposed unidirectional differentiation of SCGB3A2\textsuperscript{+} cells into AT2
cells, which was consistent with their Slingshot\textsuperscript{6} pseudotime analysis. On the other hand, Murthy et al.,\textsuperscript{3} using adult distal lung organoids, suggested that AT0 cells are transitional between AT2 and AT1 fates and can also give rise to TRB-SCs, a conclusion supporting their own Velocity trajectory analysis.\textsuperscript{7} As currently available organoid models yielded conflicting data and as these cell types are not present in mice,\textsuperscript{2,3} an \textit{in vitro} model that contains a large fraction of these cells would complement \textit{ex vivo} studies. Here we report the generation from hPSCs of expandable, clonal spheres containing these distal progenitor populations.

The GSK3 inhibitor CHIR, keratinocyte growth factor (KGF), dexamethasone, 3',5'-cyclic adenosine monophosphate (cAMP), and 3-isobutyl-1-methylxanthine (IBMX) (CK-DCI) promote outgrowth of spheres containing AT2 cells (alveolospheres) from hPSC-derived lung progenitors enriched by sequential cell sorting for reporter genes or cell surface markers.\textsuperscript{8–10} We previously reported the generation of fetal stage, branching distal lung organoids from hPSCs.\textsuperscript{11} After dissociation of ESC or iPSC-derived organoids and culture as single cells in Matrigel with CK-DCI (Fig. S1a), hollow spheres developed within 2 weeks (Fig. 1a, Fig. S1a), which we call 'transitional lung organoids' (TLOs). TLOs were clonal as mixing single cells from organoids expressing GFP or mScarlet yielded only spheres expressing either fluorescent reporter (Fig.S1b). TLOs highly expressed mRNAs encoding the AT2 markers, \textit{SFTPC}, \textit{SFTPB}, \textit{ABCA3}, \textit{LPCAT}, \textit{NAPSA}, \textit{SLC23A2}, and \textit{LAMP3} (Fig. 1b, Fig S1c). At variance with previously reported alveolospheres,\textsuperscript{8,9} however, strong induction of \textit{SCGB3A2} mRNA was observed (Fig. 1b, Fig. S1c). Confocal microscopy revealed uniform expression of EPCAM, NKX2.1 and SFTPБ with a subset of cells in each TLO also expressing pro-SFTPC and SCGB3A2 (Fig. 1c, Fig. S1d,e). In some TLOs the presumably more mature AT2 marker, HT2-280,\textsuperscript{12} (Fig. 1c), was detected, which was confirmed by flow cytometry (3-25%, n = 4 ESC-derived and n = 1 iPSC-derived TLOs) (Fig. 1d)). Western Blot showed presence of fully processed SFTPC and SFTPБ (Fig 1f, Fig. S2a). Transmission electron microscopy revealed apical presence of lamellar bodies (LBs), a feature of
AT2 cells,\(^{13}\) (Fig. 1g, Fig. S2b), consistent with the SFTPB staining pattern (Fig. 1c). However, these had low electron density and were often organized as multivesicular bodies (Fig. 1g, arrows), while the cytoplasm contained glycogen (Fig. 1g, stars), indicative of immaturity.\(^{14,15}\) Consistent with abundant presence of LBs, TLOs showed apical accumulation of the lysosomal dye, LysoTracker Red (Fig. S2c).

TLOs can be serially passaged every \(~3\) weeks for up to at least 7 passages with an average expansion of \(~10\)-fold between each passage. Over this time, TLOs were phenotypically stable (Fig. 1b pooled from 9 lines derived from organoids at days 40, 60 and 120 and between one and six passages). Furthermore, they can be cryopreserved and thawed while retaining their expression profile (Fig. S3a).

As removing WNT signaling has been reported to induce AT2 maturation in alveolospheres,\(^{9,10}\) we withdrew CHIR. CHIR withdrawal, however, led to disintegration (Fig. S3b), proliferation arrest (Fig. S3c), apoptosis (Fig. S3d) and loss of AT2 markers (Fig. S3e), indicating that GSK3 inhibition is required for TLO maintenance and propagation. Despite similar culture conditions, TLOs therefore differ from previously reported alveolospheres,\(^{9,10}\) which, in contrast to TLOs, do not contain SCGB3A2\(^+\) cells,\(^2\) undergo AT2 maturation after intermittent CHIR withdrawal,\(^{9,10}\) and require reporter lines or enrichment steps for their generation.\(^{8–10,16}\)

scRNAseq confirmed that almost all cells expressed SFTP\(B\), NKX2.1 and EPCAM and identified a distinct, large population of SCGB3A2\(^+\) cells that expressed the most SFTP\(B\). (Fig. 2a). Markers of mature airway cells were absent, and, in contrast to previously reported alveolospheres,\(^{9,10}\) contamination with gastric or intestinal cells was not detected (not shown). Clustering revealed 4 closely related clusters (0-3), and 3 small but more distinct clusters that each made up \(~1\)% of the cells (4-6) (Fig. 2b, heatmap in Fig. S4). Cluster 0 contained cells expressing AT2 markers (SFT\(P\)C, ABCA3, NAPSA) and the highest levels of SFTP\(B\), while clusters 0, 2 and 3 expressed SCGB3A2 (Fig. 2a). Cluster 1, at the other extreme, expressed
less SFTPB but did not express SCGB3A2 or AT2 markers (Fig. 2a). Cell type annotation using machine learning trained on the data of Murthy et al.\(^3\) matched cluster 0 with AT0 cells (SFTPB\(^+\)SCGB3A2\(^+\)SFTPC\(^-\)) and TRB-SCs (SFTPB\(^+\)SCGB3A2\(^+\)SFTPC\(^-\)). Cluster 1 corresponded to SFTPB\(^+\)KRT5\(^+\)P63\(^-\)SCGB3A2\(^-\) distal BCs, primarily characterized by elevated expression of IGFBP2 (Fig. 2a, Fig. S4), a marker of airway and in particular of basal cells according to LungMap.\(^17\) Clusters 2 and 3 were classified as a mixture of ‘differentiating BCs’ and rare ‘immature AT1 cells’, both likely representing differentiation intermediates (Fig. 2c). No matches with any other populations were found. Trajectory analysis using scVelo\(^7\) confirmed the notion that clusters 2 and 3 were intermediate between clusters 0 and 1, and showed differentiation pathways originating from distal BCs (cluster 1) and terminating in AT0 cells in cluster 0 (Fig. 2c). Given the clonal nature of the TLOs, these findings suggest a lineage from distal BCs over TRB-SCs to AT0 cells that was not recognized in recent scRNAseq studies of human lung.\(^1\)–\(^3,17,18\) Cluster 4 are rare ASCL1\(^+\) neuroendocrine cells, some of which, quite unusually, also expressed SFTPB and SCGB3A2 (Fig. 2a) and may represent precursors. The small cluster 5 contained proliferating cells that could not be further identified, while the equally small cluster 6 is distinguished by four IncRNAs (Fig. S4), and may also be a regenerative population.

Cell identity assignment based on the data of Haberman et al.,\(^1\) also showed SCGB3A2\(^-\) cells and transitional SCGB3A2\(^-\) AT2 cells mainly in cluster 0, and classified the remainder as BCs (they did not describe specific BC subsets) and some rare SCGB3A2\(^-\)SCGB1A1\(^+\) cells (likely closely related to pre-TRB-SCs from Murthy et al.\(^3\)) (Fig. 2e). Both training sets for cell identity assignment thus yielded consistent results.

The TLOs described here consist of a spectrum of recently identified potential distal lung progenitor cells.\(^1\)–\(^3\) Their clonality and scVelo analysis indicate close lineal hierarchy among the various cell types, with distal BCs probably the main progenitor population, at least in this in vitro
model. As these populations are absent in rodent models, TLOs will be an important resource to gain deeper insight into mechanism involved in normal and abnormal human lung regeneration, and may have applications in regenerative approaches for lung diseases.
Acknowledgments:

This work was supported by grants NIH HL120046 (HWS), NIH 1U01HL134760 (HWS), and NIH S10 OD032447 (HWS). This research was also funded by in part through the NIH/NCI Cancer Center Support Grant P30CA013696 and used the Tissue bank part of the Molecular Pathology Shared Resource.

Author Contributions

IML developed the generation of TLOs and performed most of the experiments, NS performed scRNAseq analysis, to which TAT contributed as well. KGB supervised scRNAseq work and analysis, JT performed Western Blots. MP provided critical conceptual input and hypothesized and showed that TLOs contain ATO cells. HWS provide concept, supervised and wrote the manuscript.

Competing interests

HWS and IML have file a provisional patent application on the generation of TLOs. The other authors declare no competing interests.
Methods

Human lung samples.

De-identified normal human lung samples were provided by the Herbert Irving Comprehensive Cancer Center Molecular Pathology Shared Resource Tissue Bank core under IRB-AAAT8682.

Maintenance of hPSCs

Before differentiation, Rockefeller University Embryonic Stem Cell Line 2 (RUES2, passage 24-32) or mRNAi PSC lines (generated using mRNA transfection and obtained from the Mount Sinai Stem Cell Core, NY) were maintained on mouse embryonic fibroblasts (MEFs) plated at 22,500 cells/cm². Cells were cultured in hESC maintenance media (DMEM/F12 (ThermoFisher, Carlsbad, CA), 20% Knock-out serum, (Stem Cell Technologies, Vancouver, BC), 0.1 mM β-mercaptoethanol (Sigma-Aldrich, Burlington, MA), 0.2% Primocin (InvivoGen, San Diego, CA), 20 ng/ml FGF2 (R&D Systems/Biotechne, Minneapolis, MN), and 1% Glutamax (ThermoFisher)) which was changed daily. hPSCs were passaged every 3-4 days with Accutase (Innovative Cell Technologies, San Diego, CA) at least two times before differentiation, washed and replated at a dilution of 1:24. Cultures were maintained in a humidified 5% CO₂ atmosphere at 37°C. Lines are karyotyped and verified for Mycoplasma contamination using PCR (InVivoGen) every 6 months.

Generation of hESC-derived lung organoids

The hESC-derived human lung organoids were generated as described previously. Briefly, MEFs were depleted by passaging 5-7x10⁶ hESCs onto Matrigel (Corning, Corning, NY) coated 10-cm dish. Cells were maintained in hESC media in a humidified 5% CO₂ atmosphere at 37°C. After 24 hours, cells were detached with 0.05% Trypsin/EDTA (ThermoFisher) and distributed to the 6-well low attachment plate containing primitive streak/embryoid body media (10 µM Y-27632 (Tocris/Biotechne, Minneapolis, MN), 3 ng/ml BMP4 (R&D/Biotechne) to allow embryo body
formation. Embryoid bodies were fed every day with fresh endoderm induction media (10 \( \mu \)M Y-27632, 0.5 ng/ml BMP4, 2.5 ng/ml FGF2 and 100 ng/ml ActivinA (R&D/Biotechne)) and maintained in a humidified 5% CO\(_2\)/5% O\(_2\) atmosphere at 37°C. Endoderm yield efficiency was determined by dissociating embryoid bodies and evaluating CXCR4 and c-KIT (Biolegend, San Diego, CA) co-expression by flow cytometry on day 4. Cells used in all experiments had > 90% endoderm yield and were plated on 0.2% fibronectin-coated wells (R&D/Biotechne) at a density of 80,000 cells/cm\(^2\). Cells were incubated in Anteriorization media-1 (100 ng/ml Noggin (R&D/Biotechne) and 10 \( \mu \)M SB431542 (Tocris)) for 24 hours, followed by Anteriorization media-2 (10 \( \mu \)M SB431542 (Tocris) and 1 \( \mu \)M IWP2 (Tocris)) for another 24 hours. At the end of anterior foregut endoderm induction, cells were switched to Ventralization/Branching media (3\( \mu \)M CHIR99021 (Tocris), 10 ng/ml FGF10 (R&D/Biotechne), 10 ng/ml rhKGF, (R&D/Biotechne), 10 ng/ml BMP4 and 50 nM all-trans Retinoic acid (Tocris)) for 48 hours and three-dimensional clump formation was observed. The adherent clumps were detached by gentle pipetting and transferred to the low-attachment plate, where they folded into lung bud organoids as early as d10-d12 (LBOs). Branching media was changed every other day until d20–d25 and LBOs were embedded in 100% Matrigel in 24-well transwell (BDFalcon, Franklin Lakes, NJ) inserts. Branching media was added after Matrigel solidified and changed every 2-3 days to facilitate proper growth into lung organoids. Culture of embedded organoids can be kept for more than 6 months.

**Generation of transitional lung organoids (TLOs)**

Matrigel embedded lung organoids can be used for TLO generation when they reach d42 of development. Media was removed from the transwell and 1 ml of 2 mg/ml dispase (Corning) was added to release lung organoid from the Matrigel for 30-45 minutes in normoxic incubator. The organoid was transferred to a 15 ml conical tube and washed with stop media (DMEM (Corning), 5% FBS (Atlanta Biologicals; Flowery Branch, Georgia), 1% Glutamax (ThermoFisher)) to neutralize Dispase, then centrifuged at 200 g for 5 minutes. The pellet was incubated with 1 ml of
0.05% Trypsin/EDTA in normoxic incubator for 10-12 minutes with occasional pipetting with a P1000 pipet tip. Single cell dissociation was verified using a bright field microscope. If a single cell suspension was not obtained after 12 minutes, cells were washed with stop media and incubated for additional 5 minutes with 0.05% Trypsin/EDTA. Cells were counted using a hemocytometer and 400 cells/µl of undiluted Matrigel were plated in a well of a 12-well non-tissue culture plate. The plate was placed in normoxic incubator for 30 minutes until Matrigel polymerized and 1 ml of CKDCI (3µM CHIR 99021, 10 ng/ml rhKGF, 50 ng/ml dexamethasone (ThermoFisher), 0.1 mM 8-Bromoadenosine 3',5'-cyclic monophosphate sodium salt (Tocris) and 0.1 mM 3-Isobutyl-1-methylxanthine (Sigma-Aldrich)) media was gently added using 10 ml serological pipette. Media was changed every 3 days. After 2-3 weeks, a TLO culture is established at that can be maintained by regular passaging for more than six months.

**RT-qPCR.**

Total RNA was extracted according to manufacturer’s instructions using the Direct-zol™ RNA Microprep (Zymo Research, Irvine, CA), and 500 ng of total RNA was reverse transcribed using the qScript™ XLT cDNA SuperMix (Quantabio, Beverly, MA). Technical triplicates of 15 ml reaction (for use in Applied Biosystems QuantStudio7 384-well System, Waltham, MA) were prepared with 3 µl of diluted cDNA and run for 40 cycles. Primers used are listed in Table S1. Relative gene expression was calculated based on the average cycle (Ct) value, normalized to GAPDH as the internal control and reported as fold change (2(-ddCT)).

**Immunofluorescence.**

After removal of media, a flat edge was used to carefully detach the Matrigel droplet with embedded TLOs from the bottom of the 12 well plate. The droplet was then transferred to an OCT histology mold and embedded in Tissue-Tek OCT Compound (Sakura Finetek, Torrance, CA). Frozen samples were cut on cryotome at the thickness of 10-12 mm and collected on adhesion
microscope slides, air-dried and fixed in 4% paraformaldehyde, then washed 2 times for 5 minutes in 50 mM glycine to inactivate PFA, followed by washing in PBS. Samples were permeabilized for 10 minutes in 0.2% PBST (PBS + 0.2% Triton X-100) and blocked by incubating in PBS containing 5% donkey serum, then incubated overnight in primary antibody (Table S2) in 0.2% Triton X-100 and 2% donkey serum. Next day, samples were washed three times in PBS and 1% donkey serum and incubated with secondary antibody (1:200, Table S2) for 1 hour at the room temperature. Nuclei were stained with DAPI (ThermoFisher) and sections were mounted with Mounting Reagent (DAKO, Santa Clara, CA) and coverslipped. Samples were imaged using a Leica TCS SP8 Stellaris Laser scanning confocal microscope, and Leica DMI1 Inverted Phase Contrast Microscope (Leica Microsystems, Deerfield, IL).

Flow cytometry

Spheres embedded in Matrigel were released by incubating with Dispase for 30-60 minutes, then washed and dissociated into single cells with 0.05% Trypsin/EDTA in normoxic incubator for 10-12 minutes with occasional pipetting with P1000. The single cell suspension was stained in polystyrene round-bottom 12 x 75 mm tubes (BD Falcon). Primary HTII-280 antibody (Terrace Biotech, San Francisco, CA, 1:150 dilution) was added to 150 μl of cell suspension and incubated for 1 hour at the room temperature. Cells were washed two times with FACS buffer (PBS, 10% FBS and 1% sodium azide) and centrifuged for 5 minutes at 1400 rpm. Fluorochrome-labeled secondary antibody (Alexa Fluor 488 goat anti-mouse IgM) in diluted in FACS buffer at 1:100 ratio was added for 30 minutes in the dark. Cells are washed two times by centrifugation. Conjugated human EPCAM antibody (Biolegend) was added for 30 minutes. Cells were washed and resuspended in FACS buffer for flow cytometric analysis.

Transmission Electron Microscopy
Transmission Electron Microscopy (TEM) was performed at the NYU Langone Medical Center Microscopy Core. TLOs were fixed with 2% paraformaldehyde and 2.5% glutaraldehyde in 0.1M sodium cacodylate buffer (pH7.2) for 2 hours and post-fixed with 1% osmium tetroxide for 1.5 hours at room temperature, then processed in a standard manner and embedded in EMbed 812 (Electron Microscopy Sciences, Hatfield, PA). Semi-thin sections were cut at 1 mm and stained with 1% Toluidine Blue to evaluate the quality of preservation and find the area of interest. Ultrathin sections (60 nm) were cut, mounted on copper grids and stained with uranyl acetate and lead citrate by standard methods. Stained grids were examined under Philips CM-12 electron microscope and photographed with a Gatan (4k ×2.7k) digital camera (Gatan, Inc., Pleasanton, CA).

Western Blot

Cultured cells embedded in Matrigel were released by incubating with Dispase (Corning) for 30 minutes. Cells were then washed, collected in PBS, and treated with lysis buffer (RIPA buffer, 1x Roche Complete Protease inhibitor cocktail, and Roche PhosSTOP). Buffer-treated cells were mechanically lysed using a 25-gauge needle and left on ice for 30 minutes. Human lung samples were homogenized with stainless steel beads and left to lyse on ice in lysis buffer for 4-6 hours. Samples were then centrifuged at 15,000 g for 10 minutes and the supernatant was collected and stored at -80°C until analysis. Protein concentration was measured using Pierce BCA Protein Assay Kit (ThermoFisher). A total of 18 μg of TLO, hPSC, and human lung lysate were loaded for mature SPC analysis and a total of 5 μg of each were loaded for mature SPB analysis. Lysates were resolved on pre-cast NuPage 4-12% Bis-Tris gels (Invitrogen) and transferred to nitrocellulose membranes (Bio-Rad, Hercules, CA). The following primary antibodies were used to probe the blots: mature SPC (1:1000, Seven Hills Bioreagents, Cincinatti, OH), mature SPB (1:1000, Seven Hills Bioreagents), GAPDH (1:5000, Cell Signaling Technology). Species-specific secondary antibodies with HRP conjugates (Cell Signaling Technology, Danvers, MA) were used.
at 1:15,000 dilution. Blots were then treated with Pierce ECL Western Blotting Substrate and visualized using the Amersham ImageQuant 800 (Cytiva, Marlborough, MA).

**Cryopreservation of TLOs**

Matrigel embedded spheres can be frozen and thawed for future use. 2-3 weeks after passaging of spheres, CK-DCI media was removed and 1 mL of Dispase (2 mg/mL) was used to break apart the Matrigel droplet. After 30-45 minutes in normoxic incubator, the spheres were transferred to a 15 mL conical and washed with wash media to neutralize the protease, then centrifuged at 200 g for 5 minutes. Wash media was aspirated, and the pellet was incubated with 1 mL of 0.05%Trypsin/EDTA in normoxic incubator for 10-12 minutes with occasional pipetting. Once cells are in single cell suspension, they were washed again with stop media, centrifuged at 200 g for 5 minutes, and the supernatant aspirated, followed by resuspension of the pellet in a small volume of CK-DCI and counted using a hemocytometer. To freeze down, a density of 500,000 to 1 million single cells is ideal. Cell are resuspended in equal volumes of CK-DCI media and 2x DMSO freezing medium (Quality Biological, Gaitersburg, MD) and transferred to a cryovial. Vials are immediately place in a container that allows freezing rate of -1°C/min and placed in a -80°C freezer. Next day, they are transferred to liquid nitrogen. To thaw, cryovial is places in a 37°C water bath until thawed. Cells are transferred to a 15 mL conical tube and washed with stop media, centrifuged at 200 g for 5 minutes, and aspirated. The ideal initial reseeding density post-thaw is 800 -1,600 cells per µL of undiluted Matrigel. Subsequent passages can be done at 400 cells per µL.

**Single-cell cDNA library preparation and scRNA-seq**

Viability of single cells was assessed using Trypan Blue staining, and debris-free suspensions of >80% viability were deemed suitable for single cell RNA Seq. Samples with lower viability were run with caution. Single-cell RNA-seq was performed on these samples using the Chromium
platform (10x Genomics, Pleasanton, CA) with the 3’ gene expression (3’ GEX) V3 kit, using an input of ~10,000 cells. Briefly, Gel-Bead in Emulsions (GEMs) were generated on the sample chip in the Chromium controller. Barcoded cDNA was extracted from the GEMs by Post-GEM RT-cleanup and amplified for 12 cycles. Amplified cDNA was fragmented and subjected to end-repair, poly A-tailing, adapter ligation, and 10X-specific sample indexing following the manufacturer’s protocol. Libraries were quantified using Bioanalyzer (Agilent) and QuBit (Thermofisher) analysis and were sequenced in paired end mode on a NovaSeq instrument (Illumina, San Diego, CA) targeting a depth of 50,000-100,000 reads per cell.

**scRNA-seq computational analysis**

Sequencing data were aligned and quantified using the Cell Ranger Single-Cell Software Suite (version 6.1.2, 10x Genomics) against the provided GRCh38 (Ensembl 98) human reference genome. All further computational analysis of scRNA-seq data was performed using R version 4.1.3 (https://www.R-project.org/) unless otherwise stated.

**QC and processing**

The aligned data was imported and processed using the R package Seurat v4.1.1. Quality control for doublets and low-quality cells was achieved through exclusion of cells with less than 500 or more than 9000 transcripts and those with a higher than 20% mitochondrial gene contribution, respectively. Additionally, transcripts were retained if they counted over 0 in more than 0.5 % of all cells, otherwise excluded. Count data was then log-normalized and transcripts were scaled and centered, using built-in Seurat functions. Variable transcripts were calculated based on standardized feature values using observed mean and expected variance of a local polynomial regression model. On the resulting variable transcripts 50 principal components were computed, which in turn were used as input for uniform manifold approximation and projection
(UMAP) dimensionality reduction. For clustering analysis, a shared nearest neighbor (SNN) graph was constructed and the modularity function optimized using the Leiden algorithm.

**Cell type annotation**

Potential cell types present in our dataset were predicted through machine learning. In brief, a random forest classifier (SingleCellNet R package)\textsuperscript{20} was trained on fully annotated published data by Murthy et al. and Haberman et al. and assessed on a withheld subset. This classifier was then applied to the current dataset and a matching cell type was predicted for each cell.

**RNA velocity**

RNA velocity was calculated in python using the packages Velocyto and scVelo, according to the developer's manual\textsuperscript{21,22} Briefly, ‘.loom’ files containing both exon and intron information were created from aligned raw data using Velocyto. scVelo was then used to normalize the data and compute moments. Subsequently, RNA velocity was estimated and projected onto Seurat-derived UMAP coordinates. The length and coherence of the velocity vectors, which indicate differentiation speed and directional confidence, respectively, were calculated. Finally, a dynamical model was applied to analyze transcriptional states and cell-internal latent time and subsequently recompute RNA velocities. The latent time, based solely on the transcriptional dynamics of a cell, was thereby determined.
References


Figure legends

**Figure 1. Generation and analysis of TLOs.** a. Representative bright field images of TLOs generated from RUES2 ESCs. b. RT-qPCR analysis of TLOs between 1 and 6 passages (84 days) (n=9) compared to lung bud organoids (LBOs) at day 21 (n=3) (*p<0.05, student’s unpaired t-test). c. Representative immunofluorescence images of expression of indicated markers between 1 and 6 passages after generation from RUES2 ESCs. d. Representative Western Blot analysis of the expression of fully processed SFPTB and SFTPC in undifferentiated RUES2 cells, in two distinct lines of TLOs derived from RUES2 ESCs, and in human adult lung. e. Flow cytometry for expression of HT2-280. f. Transmission electron microscopy of TLOs. g. Representative transmission electron microscopy images (at increasing magnification from left to right) of RUES2 ESC-derived TLOs. Stars: glycogen; arrows: LBs.

**Figure 2. Single cell RNAseq analysis.** a. UMAP feature plots for indicated markers. b. Clustering analysis. c. Cell identity assignment based on data of Murthy et al.. d. scVelo latent time velocity analysis. e. Cell identity assignment based on data of Haberman et al..
Figure 2

(a) NKX2-1, EPCAM, SFTP8

(b) 0, 1, 2, 3, 4, 5, 6

(c) unassigned, Neuroendocrine, Immature AT1, Differentiating Basal, Proliferating, SFTP8+, KRT5− Basal

(d) unassigned, Axonal Basal, SFTP8− SCGB3A2+, SCGB1A1−

(e) unassigned, Basal, SCGB3A2+, SCGB3A2+, SCGB1A1+, SFTP8− SCGB3A2+
Supplementary figures

Figure S1. a. Schematic representation of differentiation protocol (top) and representative bright field images of the various stages of the protocol (organoids: RUES2 ESCs, TLOs (right hand panels): iPSCs). b. TLOs generated from mixed single cells from organoids expressing either GFP or mScarlet, showing that each TLO expresses a single reporter, indicating clonality. c. RT-qPCR analysis of iPSC-derived TLOs (n=3) compared to lung bud organoids at day 21 (n=3, *P<0.05, students t-test). d. Representative immunofluorescence images of expression of indicated markers in TLOs generated from iPSCs. e. Lower magnification view of SCGB3A2 expression in several RUES2 ESC-derived TLOs.

Fig. S2. a. Representative Western Blot analysis of the expression of SFPTB and SFTPC, in two distinct lines of TLOs generated from RUES2 ESCs and one TLO line generated from iPSCs, and in human adult lung. b. Staining of ESC-derived TLOs with Lysotracker Red. c. Transmission election microscopy of RUES ESC-derived TLOs. Stars: glycogen; arrows: LBs.

Fig. S3. Effect of cryopreservation and of CHIR withdrawal. a. Expression of AT2 markers before and after cryopreservation of ESC-derived TLOs (n=3). b. Schematic representation of protocol (top) and bright field images of TLOs in CK-DCI and 5 days after withdrawal of CHIR (K-DCI). c. Expression of Ki67 and d. of cCASP3 in TLOs in CK-DCI and 5 days after withdrawal of CHIR. e. RT-PCR of select lung and midgut markers 1 to 5 days after withdrawal of CHIR (n=3).

Fig. S4. Heatmap of top differentially expressed genes in each cluster identified in scRNAseq of ESC-derived TLOs.
Figure S1

<table>
<thead>
<tr>
<th>NECs</th>
<th>Definitive endoderm</th>
<th>Anterior foregut endoderm</th>
<th>Ventralization</th>
<th>Lung bud organoids</th>
<th>Lung organoids</th>
</tr>
</thead>
<tbody>
<tr>
<td>1x media</td>
<td>N2, K17, B27, E8</td>
<td>E8, B27, K17</td>
<td>E8, B27, K17</td>
<td>E8, B27, K17</td>
<td>E8, B27, K17</td>
</tr>
</tbody>
</table>

(a) Images showing different stages of development.

(b) Fluorescence microscopy images highlighting different cell markers.

(c) Bar graph showing expression levels of different genes (SFTPB, SFTPC, LPCAT1, LAMP2, ABCC3, DSP, SLC34A2, HMOX1, SCGB3A2) in lung bud organoids and TLOs.

(d) Confocal images showing localization of different markers (NKK2.1, proSPC, EPCAM) in respiratory tract.

(e) Merged images showing the co-localization of DAPI and SCGB3A2 in respiratory tract.
Figure S2

(a) Western blot analysis showing expression levels of SFTPB, GAPDH, SFTPC, and GAPDH in human lung, spheres RUES2 line 3, pass. 1, spheres RUES2 line 1, pass. 3, spheres iPSC pass. 3.

(b) Electron microscopy images indicating the presence of structures marked by arrows.

(c) Immunofluorescence staining with Rhodamine Red showing cell morphology.
Figure S3

(a) Graph showing expression levels of various genes (SFTP6, SFTP, LPCAT1, LAMP3, ABCA3, MAPSA, SLC34A2) relative to GAPDH, with bars indicating fresh vs. freeze/thaw conditions.

(b) Diagram illustrating the process from Lung organoids to Spheres and K-DCl or CK-DCl.

(c) Immunofluorescence images of EPCAM and KI67 for CK-DCl and K-DCl conditions, with merged images.

(d) Immunofluorescence images of EPCAM and cCASP3 for CK-DCl and K-DCl conditions, with merged images.

(e) Bar graph showing log(fold change rel to LBOC) for various genes (SFTP6, SFTP, LPCAT1, LAMP3, ABCA3, MAPSA, SLC34A2, NKX2.1) with comparisons between CK-DCl and K-DCl.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCA3</td>
<td>CAGGCTGAGGGACAGGAG</td>
<td>TCCACAGAAGCCCTGAAGAT</td>
</tr>
<tr>
<td>GAPDH</td>
<td>AACTTTGGCATTGGAAGG</td>
<td>ACACATTGGGGGTAGGAACA</td>
</tr>
<tr>
<td>LPCAT1</td>
<td>GCCCTGTCTGCTGCTCTGC</td>
<td>TCTTGGCGCTCCGTACATCT</td>
</tr>
<tr>
<td>LAMP3</td>
<td>TGCTCATTTTTTATGGGATTGC</td>
<td>TGAGTTTATTTGATGCCTTCATCTT</td>
</tr>
<tr>
<td>NAPSA</td>
<td>GTAGATGGAATCCTGAGCGAG</td>
<td>AGACAGAATGGGAAAAACCGAG</td>
</tr>
<tr>
<td>NKX2.1</td>
<td>AGCACACGACTCCGGTTC</td>
<td>GCCCACTTTTCTGTAGCTTTCC</td>
</tr>
<tr>
<td>SCGB3A2</td>
<td>ACTCTGGGCATTTCGTTGAG</td>
<td>CATCCACCTCCGCTTTATCC</td>
</tr>
<tr>
<td>SFTPB</td>
<td>GAGCCGATGACCTATGCAAG</td>
<td>AGCAGCTTCAAGGGGAGGA</td>
</tr>
<tr>
<td>SFTPC</td>
<td>GCAAGAGGTCTGATGGGAG</td>
<td>ATCACCACGAGCATGAGGA</td>
</tr>
<tr>
<td>SLC34A2</td>
<td>CACCATCTTGCTCATACTCTCC</td>
<td>GTCAACCATGCAAAGGGAAAG</td>
</tr>
</tbody>
</table>
Table S2: Antibodies

<table>
<thead>
<tr>
<th>Primary antibodies for immunofluorescent staining</th>
<th>Name</th>
<th>Host species</th>
<th>Clone number</th>
<th>Manufacturer and Catalog number</th>
<th>RRID</th>
<th>URL</th>
<th>Dilution Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cCASP3</td>
<td>Rabbit</td>
<td>5A1E</td>
<td>Cell Signaling Cat #: 9664S</td>
<td>RRID:AB_20 70042</td>
<td><a href="https://scicrunch.org/res">https://scicrunch.org/res</a> olver/RRID:AB_2070042</td>
<td>1:400</td>
</tr>
<tr>
<td></td>
<td>NKX2.1</td>
<td>Rabbit</td>
<td>Polyclonal</td>
<td>Seven Hills Bioreagents Cat #: WRAB-1231</td>
<td>RRID:AB_28 32953</td>
<td><a href="https://scicrunch.org/res">https://scicrunch.org/res</a> olver/RRID:AB_2832953</td>
<td>1:200</td>
</tr>
<tr>
<td></td>
<td>proSPC</td>
<td>Rabbit</td>
<td>Polyclonal</td>
<td>Millipore Sigma Cat #: AB3786</td>
<td>RRID:AB_91 588</td>
<td><a href="https://scicrunch.org/res">https://scicrunch.org/res</a> olver/RRID:AB_91588</td>
<td>1:200</td>
</tr>
<tr>
<td></td>
<td>SCGB3A2</td>
<td>Rabbit</td>
<td>Monoclonal</td>
<td>Abcam Cat #: ab181853 (EPR11463)</td>
<td>-</td>
<td>-</td>
<td>1:250</td>
</tr>
<tr>
<td></td>
<td>SPB</td>
<td>Rabbit</td>
<td>Polyclonal</td>
<td>Seven Hills Bioreagents Cat #: WRAB-48604</td>
<td>-</td>
<td>-</td>
<td>1:500</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Secondary antibodies for immunofluorescent staining</th>
<th>Name</th>
<th>Host species</th>
<th>Conjugate</th>
<th>Manufacturer and Catalog number</th>
<th>RRID</th>
<th>URL</th>
<th>Dilution Factor</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Primary antibodies for Western blot</th>
<th>Name</th>
<th>Host species</th>
<th>Clone number</th>
<th>Manufacturer and Catalog number</th>
<th>RRID</th>
<th>URL</th>
<th>Dilution Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPDH (HRP Conjugate)</td>
<td>Rabbit</td>
<td>14C10</td>
<td></td>
<td>Cell Signaling Technologies; 36835</td>
<td>-</td>
<td>-</td>
<td>1:5000</td>
</tr>
<tr>
<td>SFTPB</td>
<td>Rabbit</td>
<td>Polyclonal</td>
<td></td>
<td>Seven Hills Bioreagents Cat #: WRAB-48604</td>
<td>-</td>
<td>-</td>
<td>1:1000</td>
</tr>
<tr>
<td>SFTPDC</td>
<td>Rabbit</td>
<td>Polyclonal</td>
<td></td>
<td>Seven Hills Bioreagents Cat #: WRAB-76694</td>
<td>-</td>
<td>-</td>
<td>1:1000</td>
</tr>
</tbody>
</table>

Secondary antibodies for Western blot

<table>
<thead>
<tr>
<th>Name</th>
<th>Host species</th>
<th>Conjugate</th>
<th>Manufacturer and Catalog number</th>
<th>RRID</th>
<th>URL</th>
<th>Dilution Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-Rabbit IgG (H+L)</td>
<td>Goat</td>
<td>HRP</td>
<td>Cell Signaling Technologies; 7074</td>
<td>RRID:AB_20 99233</td>
<td><a href="https://scicrunch.org/res">https://scicrunch.org/res</a> olver/RRID:AB_2099233</td>
<td>1:5000</td>
</tr>
</tbody>
</table>

Conjugated antibodies for flow cytometry

<table>
<thead>
<tr>
<th>Name</th>
<th>Isotype</th>
<th>Conjugate</th>
<th>Manufacturer and Catalog number</th>
<th>RRID</th>
<th>URL</th>
<th>Dilution Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>anti-human CD326 (Ep-CAM)</td>
<td>Mouse IgG2b</td>
<td>APC</td>
<td>Biolegend Cat #: 324208</td>
<td>RRID:AB_756081</td>
<td><a href="https://scicrunch.org/res">https://scicrunch.org/res</a> olver/AB_756081</td>
<td>1:100</td>
</tr>
</tbody>
</table>