Identifying a lung stem cell subpopulation by combining single-cell morphometrics, organoid culture, and transcriptomics

藤村, 崇

(Degree)
博士（理学）

(Date of Degree)
2023-09-01

(Date of Publication)
2024-09-01

(Resource Type)
doctoral thesis

(Report Number)
乙第3432号

(URL)
https://hdl.handle.net/20.500.14094/0100485863

Doctoral Dissertation

Identifying a lung stem cell subpopulation by combining single-cell morphometrics, organoid culture, and transcriptomics

一細胞形態解析、オルガノイド培養、一細胞遺伝子発現解析の統合法による肺組織幹細胞亜集団の同定

July 2023

Graduate School of Science

Kobe University

Takashi Fujimura

藤村 崇

Table of Contents

ABSTRACT

Single-cell RNA sequencing is a valuable tool for dissecting cellular heterogeneity in complex systems. However, it is still challenging to estimate the proliferation and differentiation potentials of subpopulations within dormant tissue stem cells. Here, we established a new single-cell analysis method for profiling the organoid-forming capacity and differentiation potential of tissue stem cells to disclose stem cell subpopulations by integrating single-cell morphometrics, organoid-forming assay, and RNA sequencing, a method named scMORN. To explore lung epithelial stem cells, we initially developed feeder-free culture system, which could expand all major lung stem cells, including basal, club, and alveolar type 2 (AT2) cells and found that club cells contained a subpopulation, which showed better survival rate and high proliferation capacity and could differentiate into alveolar cells. Using the scMORN method, we discovered a club cell subpopulation named *Muc5b*$^{+}$ and large club (ML-club) cells that efficiently formed organoids than other club or AT2 cells in our feeder-free organoid culture and differentiated into alveolar cells *in vitro*. Single-cell transcriptome profiling and immunohistochemical analysis revealed that ML-club cells localized at the intrapulmonary proximal airway and distinct from known subpopulation of club cell such as BASCs. Furthermore, we identified CD14 as a cell surface antigen of ML-club cells and showed that purified CD14^{+} club cells engrafted into injured mouse lungs had better engraftment rate and expansion than other major lung stem cells, reflecting the observations in organoid culture systems. The scMORN method could be adapted to

different stem cell tissues to discover useful stem cell subpopulations.

CHAPTOR 1. INTRODUCTION

Organoid culture systems with primary tissues have been used for multiple aspects, including expanding tissue stem cells as a source of cell-based therapy, assessing the differentiation capacity of stem cells, recapitulating tissue regeneration, and modeling human diseases *in vitro*. The organoid culture facilitates the self-organization of stem cells into a 3D tissue structure resembling partial tissue structures of real organs, including the intestine, lung, stomach, liver, and prostate. [1-6]. Importantly, these organoids are clonally grown from an isolated single-stem cell, demonstrating a tremendous ability of individual stem cells to regenerate a multicellular system with a unique tissue structure. Simultaneously, the organoid-forming efficiency is generally low (e.g., 6% for $LGR5^+$ intestinal stem cells, 5% for alveolar type 2 (AT2) cells, 1.1% for $LGR5^+$ liver stem cells, 1.4% for prostate stem cells, and 1.9% in the esophageal basal cells [6-10]), and the generated organoids are inconsistent in quality [11], suggesting that the tissue stem cells are heterogeneous as the seeds of organoids. Dissonant organoid-formation efficiency and differentiation potential due to stem cell heterogeneity limit the contribution of organoids to basic research and translational approaches. Therefore, it is important to identify and isolate a stem cell subpopulation with high organoid-formation capacity and able to differentiate into the cell types of interest.

Breathing is essential for animals including human beings. The organ responsible for the respiratory system is the lung in humans and most other animals. The lung harbors various types of

epithelial tissue stem cells that maintain tissue homeostasis and repair acute damage caused by inhaled insults, such as chemical particles, viruses, and bacteria [12]. The new outbreak of sever acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from December 2019 has gotten attention over the world. A great amount of research and efforts have been made worldwide to prevent further outbreaks and cure the disease. The virus reminded us once again how important is understanding the process of not only injury by virus but also respiratory regeneration in animals [13]. The respiratory system harbors various types of epithelial tissue stem cells that maintain tissue homeostasis and repair acute damage caused by inhaled insults, such as chemical particles, viruses, and bacteria [12,14,15]. In the alveolar epithelium, which is located in the most distal area of the respiratory system, AT2 cells play a central role in both homeostasis and regeneration by proliferating and differentiating into alveolar type 1 (AT1) cells specialized for gas exchange [1,16-18]. Therefore, scholars have focused on the mechanisms regulating the self-renewal and differentiation of AT2 cells. These efforts have revealed heterogeneous features within the AT2 cell population [19,20]. In addition to AT2, bronchiolar club/club-like cells supply alveolar epithelial cells during regeneration following a severe injury, such as bleomycin treatment, suggesting that multiple sources of tissue stem cells provide alveolar epithelial cells [21-23]. Bronchoalveolar stem cells (BASCs), variant club/UPK3A$^+$, and H2-K1$^+$ cells are known as club-like cells, which are found at distal conducting airways and are capable of migration into alveoli after injury. These club-like cells then undergo transdifferentiation into AT1 and AT2 cells to cover the damaged alveolar epithelium [22-

[28]. These cells are a potential resource for cell-based therapy of acute alveolar injury, such as acute respiratory distress syndrome [29].

In this study, we sought to establish an *in vitro* method for identifying the population of lung epithelial stem cells, which can contribute to alveolar regeneration. Single-cell RNA sequencing (scRNA-seq) is a powerful tool for profiling diverse cell populations in a high throughput manner [30-32]. However, this new technology has still difficulty in directly linking the proliferation and differentiation capacity of dormant tissue stem cell to their expression profiles. To overcome this issue, in this study, we developed a new method that can identify stem cell subpopulations having the capacity to form organoids and differentiate into the cell populations of interest by combining single-cell morphometrics, organoid-formation assay, and RNA sequencing, a method named scMORN (Schematic overview in Figure 1). Using the scMORN method, we discovered Muc5b-positive and large club cells (ML-club cells) as a club cell subpopulation that shows higher organoid-formation efficiency than other club or AT2 cells in our feeder-free organoid culture and can generate alveolar epithelial cells. *In vivo* transplantation experiments with injured lungs confirmed that ML-club cells were efficiently expanded at the damaged area more than AT2 cells and other club cell subpopulation and differentiated into alveolar cells. We show that the scMORN method can identify a stem cell subpopulation that generates the desired organoids.

CHAPTOR 2. Identifying a lung stem cell subpopulation by combining single-cell morphometrics, organoid culture, and transcriptomics

MATERIALS AND METHODS

Mouse lines

The animals were housed in controlled environment rooms, and all the experimental procedures using animals were approved by the Institutional Animal Care and Use Committee of the RIKEN Kobe Branch. We handled the mice in accordance with the ethical guidelines of the institute. *SFTPC*-green fluorescent protein (*GFP*) (PMID: 18178827), *Sftpc-CreERT2* (B6.129S-*Sftpc*$^{tm1(cre/ERT2)Blh}$/J) (PMID: 22123957, JAX Stock No: 028054), *Scgb1a1-CreER* (B6N.129S6(Cg)-*Scgb1a1*$^{tm1(cre/ERT)Blh}$/J) (PMCID: PMC2730729, JAX Stock No: 016225), and *Rosa26-mTmG* (B6.129(Cg)-*Gt(ROSA)26Sor*$^{tm4(ACTB-tdTomato,-EGFP)Luo}$/J) (PMID: 17868096, JAX Stock No: 007676)[21,33-35] and have been described previously. Nude mice were purchased from Nihon SLC. *Scgb1a1-CreER*; *Rosa26-mTmG* and *Sftpc-CreERT2*; *Rosa26-mTmG* mice were created by crossing *Scgb1a1-CreER*, *Sftpc-CreERT2*, and *Rosa26-mTmG*. To label *Scgb1a1-CreER* and *Sftpc-CreERT2* lineage cells, these mice were injected with tamoxifen (0.25 mg/g body weight) in peanut oil five times for Scgb1a1+ and three times for Sftpc+ cells on alternate days, and 3 weeks after the last injection; these mice were used for

all experiments.

Human samples

HPSAEpiC (Human Small Airway Epithelial Cells, #3230) and HPAEpiC (Human Plumonary Alveolar Epithelial Cells, #3200) were obtained from ScienCell.

Preparation of primary cells and FACS sorting

Primary cells were prepared by the previously described protocol with some modifications[21]. Animals were killed by carbon dioxide, perfused with 10-mL saline solution through the right ventricle. The following was injected through the trachea, which was tied with a string: 1-mL protease solution (Collagenase type I (450 U/mL, Worthington), Elastase (1 U/mL, Worthington), Dispase (5 U/mL, Corning), DNaseI (0.05 mg/mL, SIGMA), trypsin (0.025%, Thermo Fisher Scientific) in DMEM/F12). The removed lung without the trachea was chopped into small pieces with a blade and incubated in 3–4 mL protease solution for 30 min at 37°C with a rotator. After the tissue was dissociated by pipetting, and a total of 15-mL DMEM/F12 containing 10% FBS was added. And then, cells were filtered through 100 and 40 μm strainers. After centrifugation, cells were incubated at room temperature for 3 min in 2-mL red blood cell lysis buffer, centrifuged and resuspended in DMEM/F12 containing 3% FBS. These cells were used for FACS sorting.

To sort GFP$^{neg/low/hi}$ cells from *SFTPC-GFP* mice or Sftpc+/AT2 cells from *Sftpc-CreERT2*; *Rosa-mTmG* mice, these cells were stained with PE-Cy7 anti-Epcam (1:100, #25-5791-80, Thermo Fisher Scientific), biotin anti-CD31 (1:250, #13-0311-82, Thermo Fisher Scientific) and biotin anti-CD45 (1:400, #13-0451-82, Thermo Fisher Scientific) for 20 min on ice. After being washed with DMEM/F12 containing 3% FBS, cells were incubated for 10 min in streptavidin APC-Cy7 solution (0.25 μL/10^6 cells, Biolegend). Next cells were washed, resuspended in DMEM/F12 containing 3% FBS, added 7-AAD was added (5 μL/10^6 cells, BD Biosciences). 7-AAD$^-$, CD31$^-$, CD45$^-$, Epcam$^+$ and GFP$^{neg/low/hi}$ or positive cells were sorted on a FACS AriaII (BD Biosciences).

To sort Scgb1a1+/CD24$^{neg/mid/hi}$ cells from *Scgb1a1-CreER*; *Rosa26-mTmG* mice, cells were stained with PE-Cy7 anti-Epcam (1:100, #25-5791-80, Thermo Fisher Scientific), biotin anti-CD31 (1:250, #13-0311-82, Thermo Fisher Scientific), biotin anti-CD45 (1:400, #13-0451-82, Thermo Fisher Scientific) and APC anti-CD24 (1:70, #17-0242-82, Thermo Fisher Scientific) for 20 min on ice. 7-AAD$^-$, CD31$^-$, CD45$^-$, Epcam$^+$, GFP$^+$ and CD24$^{neg/mid/hi}$ cells were sorted.

To sort Scgb1a1+/CD14$^{+/-}$ and proximal/distal CD14$^+$ cells from *Scgb1a1-CreER*; *Rosa26-mTmG* mice, cells were stained with, biotin anti-CD31 (1:250, #13-0311-82, Thermo Fisher Scientific), biotin anti-CD45 (1:400, #13-0451-82, Thermo Fisher Scientific), PE-Cy7 anti-CD24 (1:50, #25-0242-82, Thermo Fisher Scientific), eFluor450 anti-MHCII (1:150, #48-5321-82, Thermo Fisher Scientific) and APC anti-CD14 (1:60, #17-0141-82, Thermo Fisher Scientific) for 20 min on ice. 7-

AAD^-, $CD31^-$, $CD45^-$, $MHCII^-$, GFP^+, $CD24^{mid}$and $CD14^{+/-}$ cells were sorted. The proximal from 1/3 to 2/5 of each lobe was used for isolation of proximal CD14+ cells and the residue of distal side for distal CD14+ cells.

Organoid culture

Sorted cells were mixed in an equal volume of GFR Matrigel (#356230, Corning) and a 20-μL drop placed on the bottom of the plate, one drop/well in a 48-well plate or four drops/well in a 6-well plate. The cell number of seeded cells ranged from 150 to 5,000 cells per drop. A 250-μL medium was added to the 48-well plate or 2-mL to the 6-well plate. To establish a feeder-free culture system for $GFP^{neg/low/hi}$ cells, basic medium (DMEM/F12 containing 1× B/27, 5 % FBS, 15 mM HEPES, 0.03 % $NaHCO_3$, 250 ng/mL Amphotericin B, 1× Penicillin/Streptomycin) was supplemented with Y27632 (10 μM, #LCL-Y-5301-5, LC Laboratories), HGF(30 ng/mL, #2207-HG-025, R and D Systems), FGF10 (50 ng/mL, #100-26, PeproTech), KGF (50 ng/mL, #5028-KG-025, R and D Systems), NOGGIN (100 ng/mL, #250-38, PeproTech), SB431542 (10 μM, # 616461, Sigma-Aldrich) and/or CHIR99021 (3 μM, #SML1046, Sigma-Aldrich). Complete medium was basic medium supplemented with Y27632, HGF, FGF10, KGF, NOGGIN, and SB431542. The differentiation medium has the same contents as that of the basic medium. The medium was changed every 3 days. Y27632 was included in the medium for the first 3 days.

Retrospective trajectory analysis by long-term live-imaging system

Three hundred cells were resuspended with 250 μL complete medium containing 2.5% Matrigel and seeded in ultra-low-attachment 96-well plate (CELLSTAR Cell-repellent, #655970, Greiner Bio-One). The medium was a filtrated with 0.45-μm filter to remove the debris of Matrigel before mixing with cells. These cells were set in an imaging system equipped with an incubator (Celldiscovere7, Carl Zeiss) for at least 30 min to fall on the bottom of the plate. The conditions of the Incubator were kept at 37 ℃ and 5% humidified CO_2. At least 2 h pre-equilibration of the microscope incubator was important to avoid the plane of focus drifting during live-imaging acquisition. The medium was not changed for ten days. Bright-field cell culture images of the whole well (14220 × 15434 pixels for the X–Y plane and 2.65 μm for Z-axis step; 17 sections) were obtained at different 16 time points: day 0, day 0.17, day 0.33, day 0.5, day 0.67, day 0.83, day 1–10 by using 10× objective lens (Plan-APOCHROMAT 10×/0.35, Carl Zeiss). Collected images were analyzed by using ImageJ ver1.52n (https://imagej.net/NIH_Image, National Institutes of Health). Cells were manually tracked from day 0 to day 10 with live-imaging data (resized to 3555 × 3858 pixels for the X–Y plane). Cells that matched the following criteria were excluded from further analysis: cells were out of focus, were located on the edge of image tiles, were attached to other cells, or were not capable of being tracked until day 6. All spheres over 50 μm on the major axis at day 10 were classified as organoids. The cells

were classified into two categories: organoid-forming cells and non-organoid forming cells. The images of cells at day 0 were cropped from the original size image data, and z projection was applied to these cropped images by using StackReg (ImageJ plugin, http://bigwww.epfl.ch/thevenaz/stackreg/ , Unser Lab)[36]. These projection images were used for morphometric analysis.

Single-cell organoid culture system

FACS-sorted $CD24^{mid}$ cells (5,000–10,000) were resuspended with 2-mL complete medium without Matrigel including 10-μL PI (Propidium Iodide, #130-093-233, Miltenyi Biotec) and seeded in 6well-Elplasia plate (#4444, Corning). The plate was coated with Biosurfine (#AWP-MRH, TOYO GOSEI) to become a non-adhesive bottom before use. These cells were set in a cell picking and imaging system (CELLHNDLER, YAMAHA) for 30 min to fall at the bottom of the plate. Bright-field and red fluorescent images were obtained using 10× objective lens. We manually selected single live cells. That is, they were alone in a microwell and PI-negative. Single cells were automatically moved to an ultra-low-attachment 384 well plate (CELLSTAR Cell-repellent, #655970, Greiner Bio-One) fulfilled with 50 μL/well complete medium containing 2.5% Matrigel. The seeded cells were incubated for 10 days at 37°C and 5% CO_2. All spheres over 50-μm in the major axis at day 10 were classified as organoids. The cells were classified into two categories, namely organoid-forming cells and non-organoid-forming cells. Projection images automatically produced by CELLHANDLER were

used for morphometric analysis.

Single-cell images of FACS-sorted cells

Sorted $CD24^{mid}$ or $CD14^{+/-}$ cells were resuspended with 250-μL basic medium including 5-μL PI and added to 96-well glass-bottom plate (SensoPlate, #655892, Greiner Bio-One). The cells were set under a microscope for at least 30 min to fall on the bottom of the plate. Bright-field, green fluorescent, and red fluorescent images were obtained with IX83 fluorescent microscope (Olympus) with UCPLan FL N 20×/0.70. Cells that matched the following criteria were excluded from further analysis; cells were GFP-negative, dead (PI-positive), out of focus, located on the edge of image tiles, attached to other cells. Bright-field images were used for morphometric analysis.

Single-cell RNA-seq with image data

$CD24^{mid}$ cells were sorted as single cells to PCR plates using CELLHANDLER (YAMAHA). The procedure was similar to that of the abovementioned single-cell organoid culture system above mentioned. However, the resuspension medium was changed to a basic medium, including 10-μM Y27632 and 10-μL PI. Further, 96-well PCR plates (#0030129512, Eppendorf) used as a receiver plate were fulfilled with 2-μL/well RNA inhibitor solution (0.2-μL RNasin® Plus Ribonuclease Inhibitor (#N2615, Promega), 0.4-μL 5×Maxima H-Minus Reverse Transcriptase RT buffer (#EP0752, Thermo

Fisher Scientific), and 1.4-μL RNase free water. The detailed protocol of single-cell RNA sequencing described elsewhere (Jin et al., submitted). Briefly, for 286 single cells from one mouse and two wells with only medium (no cell) in abovementioned 96-well PCR plates, in-house library preparation including cell lysis, RNA fragmentation, cDNA generation, amplification, and purification was performed using the Bravo NGS workstation (Agilent Technologies) and thermal cyclers (Mastercycler X50s, Eppendorf). The libraries were mixed and sequenced on Miseq (Illumina) using custom primers. After the sequencing results were processed, the detected number of RNA molecules for each gene in each cell was counted based on molecular barcodes. Morphometric analysis was performed on the sequenced 286 cells.

Gene expression analysis

Single cell clustering based on gene expressions (the detected number of RNA molecules) and marker gene identification were performed using Seurat ver3.2.2[37] by following its instructions. The cells meeting following criteria were removed from further analyses: nCount_RNA < 1000 or > 12000 and nFeature_RNA < 2200. The remained cells were clustered using the FindNeighbors and FindClusters functions with the parameters “resolution = 0.8” and “dims = 1:10”. By principal components analysis using the RunPCA function, 8 dispersive cells in PC1-PC2 space which were apparently out of three main clusters were found. To focus on the main clusters, these dispersive cells

were removed from further analyses using the subset function with the parameter “subset = PC_2 >= -10”. The clusters were visualized using Uniform Manifold Approximation and Projection (UMAP) algorithm. For each cluster, marker genes were determined using the FindMarkers function, and were visualized using the FeaturePlot and DotPlot functions.

Computational integrated analysis of scRNA-seq datasets

The integrated analysis of this study’s dataset and GSE118891[25], which was obtained from Gene Expression Omnibus (GEO), was performed according to the standard integration workflow of Seurat v3[37]. Briefly, using FindintegrationAnchors and IntegrateData functions, we identified common anchors between two datasets and integrated two datasets with these anchors. Integrated two data sets were clustered using the FindNeighbors and FindClusters functions with the parameters “resolution = 0.5” and “dims = 1:7”.

Total RNA isolation, cDNA preparation and quantitative RT-PCR

Total RNA from FACS-sorted cells was purified with ISOGEN (#319-90211, NIPPON GENE) and that of organoid and whole lung was purified with RNeasy Mini Kit (#74104, QIAGEN), according to manufacturer’s instructions. Total RNA was reverse transcribed to cDNA by SuperScript III First-strand synthesis system (#18080-51, INVITROGEN) according to manufacturer’s

instructions. Quantitative RT-PCR was performed using primer sets (show in below table) with Thunderbird SYBR qPCR mix (#QPS-201, TOYOBO) on a QuantStudio5 Real-time PCR machine (Thermo Fisher Scientific). PCR cycling parameters were 95°C for 1 min (one cycle); 95°C for 15 sec, 62°C for 15 sec, 72°C for 35 sec (40 cycles). Gapdh (glyceraldehyde 3-phosphate dehydrogenase) was used to normalize gene expression as housekeeping gene.

Table Primer lists for quantitative RT-PCR

Gene name	Primer Sequence (Forward, 5' to 3')	Primer Sequence (Reverse, 5' to 3')
Ager	gctggcacttagatgggaaa	tgcaggagaaggtaggatgg
Foxj1	ctacttccgccatgcagacc	atccttctcccgaggcactt
Gapdh	aatgtgtccgtcgtggatctga	gatgcctgcttcaccaccttct
Hopx	acttcaacaaggtcaacaagcac	tggctccctagtccgtaaca
Ki67	gcgaagagagcatccatcag	tgtgtttgtttcacgccaag
Krt5	gttctttgatgcggagctgt	ctcgtactgggccttgacct
Muc5b	tgtgggtgtgtgcttgtcct	ggcagtgtccagttgggttc
Pdpn	gaccgtgccagtgttgttct	ccatgccgtctcctgtacct
Scgb1a1	caccaaagcctccaacctctac	gggatgccacataaccagactc
Scgb3a1	ttaagccacttggccatcct	ttcccaggttcccctcaac
Sftpc	accctgtgtggagagctacca	tttgcggagggtctttcct
Sox2	gaccagctcgcagacctaca	gcctcggacttgaccacaga
Tff2	gcagtggtcctggttttgg	tggtttggaagtgggtgga

Immunohistochemistry

Lungs and organoids were fixed with 4% paraformaldehyde at 4°C for 2 h to overnight (for lungs) or at room temperature for 20 min (for organoids), embedded in paraffin or OCT (for frozen sections) and sectioned at 6–8 μm. Sections were stained with primary antibodies listed in Table S4. Secondary

antibodies conjugated with Alexa Fluor 488/555/594/647 (Thermo Fisher Scientific and Jackson ImmunoResearch Labs) were used. DAPI (Nacalai tesque) was used as a nuclear counterstain. Images were obtained with LSM710 confocal laser scanning microscope (Carl Zeiss) with ×63/1.4 NA Oil lens or IX83 fluorescent microscope (Olympus) with UPLanSApo 4×/0.16.

Detailed immunostaining conditions listed in below table.

Table Antibody lists and conditions for immunohistochemistry.

Antibody dilution	Company, Catalog code.	Tissue preparation	Antigen retrieval
Chicken anti-GFP (1:200)	Thermo Fisher Scientific Cat# A10262	Paraffin	105°C, 15 min in AUShigh**
Goat anti-SCGB3A1 (1:500)	R and D Systems Cat#AF2954	Paraffin	105°C, 15 min in AUS*
Goat anti-SOX2 (1:200)	R and D Systems Cat#AF2018	Paraffin	105°C, 15 min in AUShigh**
Hamster anti-PDPN (1:250)	Abcam Cat#ab11936	Paraffin	105°C, 15 min in AUShigh**
Mouse anti-ACTUB (1:1000)	Sigma-Aldrich Cat#T7451	Paraffin	105°C, 15 min in AUShigh**
Mouse anti-SCGB1A1 (1:200-500)	Santa Cruz Cat#sc-390313	Paraffin	105°C, 15 min in AUShigh** or AUS*
Mouse anti-HT2-280 (1:50)	Terrace Biotech Cat#TB-27AHT2-280	Paraffin	105°C, 15 min in AUShigh**
Rabbit anti-HOPX (1:200)	Sigma-Aldrich Cat#HPA030180	Paraffin	5 min in 0.05% Trypsin
Rabbit anti-KRT5 (1:500)	Abcam Cat#ab52635	Paraffin	105°C, 15 min in AUShigh**
Rabbit anti-MUC5B (1:500)	Camille Ehre lab, UNC #UNC223 (Camille Ehre lab, UNC)	Paraffin	105°C, 15 min in AUS*
Rabbit anti-Prosurfactant Protein C (1:300)	Millipore Cat#AB3786	Paraffin	105°C, 15 min in AUShigh** or AUS*
Rabbit anti-TFF2 (1:200)	Proteintech Cat#13681-1-AP	Frozen	90°C, 5 min in AUS*
Rat anti-RAGE (1:200)	R and D Systems Cat#MAB1179	Paraffin	5 min in 0.05% Trypsin
Mouse anti-P63 (1:100)	Abcam Cat#ab735	Paraffin	105°C, 15 min in Hist VT One***
Rat anti-KI67 (1:200)	Thermo Fisher Scientific Cat# 14-5689-82	Paraffin	105°C, 15 min in Hist VT One***

***: Histo VT One (#06380-05, Nacalai), **: Antigen Unmasking Solution (high pH) (#H-3301, Vector laboratories), *: Antigen Unmasking Solution (#H-3300, Vector laboratories)

Immunocytochemistry

Isolated cells were fixed with with 4% paraformaldehyde at room temperature for 15 min and permeabilized with 0.5 % TritonX-100 at room temperature for 10 min. They were stained with anti-SCGB3A1 (R and D Systems, Cat#AF2954, 1:100). Secondary antibodies conjugated with Alexa Fluor 647 (Thermo Fisher Scientific and Jackson ImmunoResearch Labs) were used. DAPI (Nacalai tesque) was used as a nuclear counterstain.

Transplantation

One week before transplantation, nude mice (10 weeks, over 23g weight, male) were injured by intranasal instillation of a single dose of Bleomycin (3 mg/Kg, #170006109, Nippon Kayaku) under isoflurane anesthesia. Seven thousand cells of AT2 or $CD14^{+/-}$ cells were resuspended with 75-μL complete medium without Matrigel and intratracheally transplanted to injured mice under isoflurane anesthesia with three mice in each group. Two weeks after transplantation, the collected right lungs were embedded in paraffin for immunostaining and the left lungs in OCT for counting transplanted cells with frozen sections.

Organoid forming efficiency

Projection images of organoids were obtained using an IX83 fluorescent microscope (Olympus)

with UPLanSApo 4×/0.16 or PLAPON 1.25×/0.04. All spheres over 50 μm on major axis at day 6 were counted as organoid. The organoid number was divided by input cells to determine organoid-forming efficiency (OFE). OFE values were pooled from three to four replicate drops per mouse to calculate the average value per condition.

Morphometric analysis

We used ImageJ for morphometric analysis of the single-cell images obtained by the live-imaging system, single-cell organoid culture system, and fluorescent microscopy for FACS-sorted cells. First, we performed binarization using automated thresh holding based on Otsu's method. And then, we performed close, dilation, and erosion to fill the holes of binarized images. Finally, we set ROIs (Region of Interest) by using analyze particle (threshold = 30 pixels, ImageJ plugin) and quantified following items; Area, Major axis, Minor axis, Perimeter and so on (See Table S1).

Quantification of transplanted cells

To evaluate the proliferative capacity of Sftpc+, AT2 and Scgb1a1+, $CD14^{+/-}$ cells in regenerative conditions, the frozen blocks of transplanted left lobe were used. These blocks were sectioned at 8 μm. The sections were collected at each 100-μm distance. About 40 sections per mouse were evaluated, and images of GFP-positive cells were obtained using IX83 fluorescent microscopy with UPLanSApo

40×/0.95. When the distance from one GFP-positive cell to another GFP-positive cell was within 100 μm, these cells were determined as the same cluster. The number of clusters and grafted cells per cluster was counted. The cluster number was divided by transplanted cells to determine cluster formation efficiency (CFE).

Statistical analysis

Statistical analyses were conducted using GraphPad Prism9 (GraphPad) or R software version 3.6.3 (https://www.r-project.org/, R Development Core Team). Two-tailed unpaired Student's t-test, Wilcoxson rank-sum test, one-way analysis of variance followed by Tukey's multiple comparison test, and Spearman's rank correlation coefficient were performed as shown in figure legends. $P < 0.05$ was considered statistically significant.

RESULTS

Evaluation of various additives to establish feeder-free organoid culture system with various lung epithelial cells

Several groups have established mouse lung epithelial organoid culture systems, most of which require lung fibroblasts as feeder cells to support the growth of stem cells in addition to ingredient

growth factors [1,38-40]. Because these fibroblasts interfere with single-cell live-imaging approaches, we attempted to develop a feeder-free organoid culture system that allows various lung epithelial stem cells to proliferate.

To collect various mouse lung epithelial stem cells, we took advantage of a transgenic reporter mouse line, *SFTPC-GFP*, in which the level of GFP expression varies according to the airway region [38]. EpCAM^{+} lung epithelial cells were collected from *SFTPC-GFP* mice by fluorescence-activated cell sorting (FACS) and divided into three fractions according to the GFP intensity (Figure 2A). Using quantitative reverse transcription polymerase chain reaction (qRT-PCR), we confirmed that the GFP-negative subset (GFP^{neg}) included bronchial airway epithelial cells, such as *Krt5*$^{+}$ basal, *Scgb1a1*$^{+}$ club, and *Foxj1*$^{+}$ ciliated cells. In contrast, the GFP-low subset (GFP^{low}) had bronchiolar club and fewer ciliated cells. The GFP-high subset (GFP^{hi}) mainly contained Sftpc^{+} AT2 cells, as previously reported (Figure 2B) [38].

To develop a feeder-free lung organoid culture medium, we optimized growth factor cocktails that are sufficient enough to maintain and induce proliferation of the lung epithelial stem cells. We cultured mouse lung epithelial cells by adding KGF, FGF10, HGF, and inhibitors of TGFβ and BMP signaling (SB431542, NOGGIN). It was reported that these receptor tyrosine kinase signals and Wnt signal promote the proliferation and Tgfβ superfamily signals promote the differentiation to functional cells in lung development and/or regeneration [10,12,41-46]. We evaluated these combinations and

successfully generated organoids from each GFP^{hi}, GFP^{low}, and GFP^{neg} population (Figures 2C–G). Unlike a previous report [41], the addition of a Wnt agonist, CHIR99021, did not improve the organoid-forming efficiency (OFE) of any subset of lung stem cells (Figure 2F). Our culture system also revealed that the GFP^{low} cells have better OFE than GFP^{neg} and GFP^{hi} (10% compared to 3%–5%) (Figure 2H).

To determine the cellular contents of these organoids, we evaluated the mouse lung epithelial cell lineage markers by qRT-PCR and immunocytochemistry at culture day 9 (Figure 2I-L). As suggested by a previous report [38], GFP^{neg} organoids expressed bronchial airway epithelial markers, whereas GFP^{hi} organoids showed high expression of the alveolar markers. GFP^{low} organoids contained club, AT1, and AT2 cells in a sphere that exhibited a mixed phenotype of conducting airway and alveoli (Figures 2I and 2K). These organoids were further incubated with basic medium for accelerating differentiation and were examined on day 12 (Figure 2M). Each GFP^{neg} and GFP^{hi} organoids still showed exclusive expression of bronchial airway epithelial and alveolar markers, respectively (Figures 2J, 2L, 2Q, and 2S). In contrast, GFP^{low} organoids increased AT1 cell markers, AGER and HOPX, suggesting the progression of alveolar differentiation (Figures 2K and 2R). Supporting this result, it has been reported that some club cell subpopulations, such as BASCs, UPK3A^{+}, and H2-K1^{+}, can migrate to the alveolar and transdifferentiate into AT1 and AT2 cells [23-26,47]. These data sets prompted us to hypothesize that club cells in the GFP^{low} subset include a stem cell subpopulation that

forms organoids more efficiently than the other club or AT2 cells and can differentiate into AT1 and AT2 cells in our feeder-free culture condition.

Club cell subpopulation transdifferentiates into alveolar cells

To validate that a part of club cells provides the source of alveolar organoids, we employed *Scgb1a1-CreER*, *Rosa26-mTmG* mice, and the conducting airway epithelial marker CD24 to isolate club cells (Scgb1a1^{+}, CD24^{mid}) by separating them from ciliated (Scgb1a1^{+}, CD24^{hi}) and AT2 cells (Scgb1a1^{+}, CD24^{neg}) using FACS (Figure 3A). Like the GFP^{low} organoids, the Scgb1a1^{+}, CD24^{mid} cells generated organoids and showed the bronchoalveolar phenotype by day 9 and then further differentiated into alveolar AT1 and AT2 cells by day 12 (Figures 3B-E). To ask whether the contaminated AT2 cells expanded in the culture, we excluded AT2 cells by FACS with anti-MHCII antibody. Scgb1a1+, CD24^{mid}, MHCII^{neg} cells showed similar OFE and differentiation capacity with Scgb1a1+, CD24^{mid} cells, indicating that the influence from contaminated AT2 cells is negligible (Figures 3F-I). These data proved our idea that club cells or their subpopulation can transdifferentiate into alveolar cells with better OFE than AT2 cells.

Single-cell morphometry following long-term live-organoid imaging unveiled the morphological features of the club cell subpopulation

Studies have demonstrated that particular subpopulations but not all club cells contribute to

alveolar tissue regeneration [22-26]. We found isolated club cells showed morphological diversity (Figure 4A and B). To identify the subpopulation of club cells that can form organoids, we took advantage of long-term live-organoid imaging and retrospective analyses (Figure 1A, left). For long-term live imaging, the feeder-free culture system was optimized by reducing the concentration of Matrigel from 50% to 2.5% to culture stem cells on the surface of plate (see Methods). Using this optimized method, mouse Scgb1a1^{+}, CD24^{mid} club cells were cultured into a microscope equipped with an incubator. They were imaged for 10 days (see Methods), allowing us to examine the entire process of organoid formation from single cells to spheres (Figure 4C). Reflecting the OFE (12%), only a few cells became organoids, and the others remained as single-cell colonies or formed tiny aggregates (Figure 4D, red and white arrowheads). To predict subpopulations of club cells on day 0 and determine whether they efficiently form organoids and provide alveolar cells in the culture system, we attempted to correlate the OFE with the morphological characteristics of individual club cells by analyzing single-cell morphometry. The organoid-forming or nonforming were determined on day 0 by retrospective analysis of organoid-formation assay time-series images (Figure 4D). One thousand fifty-six single-cell images were collected and classified into organoid-forming (235 images) and non-organoid-forming cells (821 images; Figure 4E). We quantitatively profiled these 1,056 cells in 31 different measurements using ImageJ (Figure 4F). We found 11 measurements that can statistically distinguish between organoid-forming cells from non-organoid-forming cells (Table 1 and see Methods). Most of

these measurements are related to cell sizes such as area, length in major axis, and perimeters (Figure 4G). Of note, the organoid-forming club cells are slightly but significantly large (organoid-forming club cells; 134 $\mu m^2 \pm 37$ vs. nonforming club cells; 100 $\mu m^2 \pm 34$ (Mean ± SD), p = 2.2E–16). Thus, our single-cell morphometric analysis following long-term live-organoid imaging suggests that average-size of the organoid-forming club cells may be larger in comparison to non-organoid-forming cells.

Larger club cells efficiently form organoid confirmed by using cell picking and imaging robot

Next, we examined whether isolated solitary large club cells show better organoid-forming capacity than the other club cells. As FACS sorting could not separate those two populations due to subtle differences in cell size, we took advantage of an automation system by which individual cells can be imaged, picked up, and transferred to another culture plate (Figure 5A and Methods). First, isolated mouse Scgb1a1^{+}, CD24^{mid} cells were arrayed on the bottom of an Elplasia microsquare bottom plate. Then, each single-cell was imaged with an upright camera, picked by the robot (Figures 5B), and inoculated into each well of a 384-well plate to culture for organoid assay. The organoid formation was assessed on day 10 (Figure 5C), and similar to the above experiment, single-cell morphometry analysis using ImageJ was performed on each organoid-forming (307 imagas) and nonforming club cells (746 images, Figure 5D), determining that organoid-forming club cells are slightly but

statistically larger than nonforming club cells, even in an isolated condition (Figure 5E). We evaluated the correlation between OFE and cell size distribution and found that club cells showed better OFE as cell size increased (Figure 5F). These experiments determined a positive correlation between these OFEs and cell size ($R = 0.781$, $p = 0.000587$). If the cell size was more than 150 um^2, OFE was approximately 50%. Thus, the large club cells indeed show better survival rate and high proliferation capacity.

scRNA-seq revealed transcriptional features of the club cell subpopulations

These results prompted us to hypothesize that club cells are composed of subpopulations with different organoid-forming capacities, and especially large club cells have a better capacity. Next, we sought mouse club cell subpopulations with different cell sizes to test these ideas. To unveil transcriptional signatures of club cell subpopulations, we performed scRNA-seq analysis following single-cell morphometry. The Scgb1a1^{+}, CD24^{mid} cells were imaged and sorted into PCR plates by cell imaging and picking robot, respectively, and were then analyzed by a plate-based scRNA-seq method (Figures 1A (right) and 6A). We analyzed 286 cells and found that these cells were divided into three distinct groups, club #1, club #2, and AT2, based on gene expression patterns (Figures 6B and C). *Scgb1a1* was dominantly expressed in club #1 and #2, while *Sftpc* was expressed only in AT2 (Figures 6D and E). We linked each single-cell expression data to the results of morphometry analysis

of the same individual cells. We determined that club #2 was significantly larger than club #1, #1 vs. #2 = 117 ± 25 μm^2 vs. 125 ± 27 μm^2 (Mean ± SD, p = 0.02196) (Figure 3F). This observation suggested that mouse club cell subpopulations are distinguished by unique transcriptomes and different cell size. In club #2, we found 184 genes significantly and more highly expressed than in club#1 (Table 1). Because several upregulated genes are relative to the mucus layer and secretory proteins, including *Muc5b*, *Tff2*, *Reg3g*, *Bpifb1,* and *Scgb3a1*, we speculated that club #2 cells appear to be large due to the large amount of accumulated secretions. We conducted immunocytochemistry and confirmed that SCGB3A1$^+$ club cells were larger than SCGB3A1$^-$ club cells (Figures 6G and H). Because of its gene expression and cell morphological features, club #2 was named ML (Muc5b$^+$, Large)-club cells.

Isolated ML-club cells efficiently formed organoid and differentiated into alveolar lineage cells *in vitro*

To isolate a large number of ML-club cells, we sought cell surface antigens specific to ML-club cells and found nine distinct antigens (Figure 7A). We tried to isolate ML-club cells with commercially available antibodies using FACS. Anti-KCNE3 and anti-IL13RA1 antibodies did not work for FACS sorting method; however, anti-CD14 antibody fortunately worked and distinguished CD14-positive and negative (CD14$^+$ and CD14$^-$) with isotype control antibody (Figure 7B). CD14+ club cells could be isolated from Scgb1a1$^+$, CD24^{mid} cells except MHCII$^+$ AT2 cells. The expression of *Muc5b,*

Scgb3a1, and *Tff2* in $CD14^+$ club cells was confirmed by qRT-PCR (Figure 7C). Single-cell morphometry further confirmed that $CD14^+$ cells are statistically larger in comparison to $CD14^-$ cells ($CD14^+$ vs. $CD14^-$ = 107 ± 22 μm^2 vs. 85 ± 25 μm^2, Mean ± SD, p = 2.2e-16), suggesting that the CD14 antibody is useful to enrich ML-club cells from Scgb1a1$^+$, $CD24^{mid}$ cells by FACS (Figure 7D). The $CD14^+$ club cells indeed showed better OFE than $CD14^-$ and generated alveolar organoids (Figure 7E, $CD14^+$ vs. $CD14^-$ = 13.4 ± 2.6% vs. 4.5 ± 1.2%, Mean ± SD, p = 0.0008), which consist of AT1 and AT2 cells by day 12 in the feeder-free organoid culture system (Figures 7F and G). In addition, CD14+ cells were assessed for correlation between OFE and cell size. Likewise Scgb1a1+, $CD24^{mid}$ cells (Figure 7H), and CD14+ cells showed a positive correlation (R = 0.673, p = 0.00602). These large CD14+ cell-derived (more than 140 μm^2) organoids expressed all of alveolar cell markers (SFTPC, AGER, and HOPX), ML-club cells markers (MUC5B and SCGB3A1), and pan-club cell marker (SCGB1A1) in one organoid at day 10 (Figure 7I). Conclusively, we identified two mouse club cell subpopulations that exhibit different properties in organoid-forming capacity and cell size and gene expression profile. In particular, we discovered ML-club cells showing high OFE that can be isolated using the cell surface antigen CD14. However, H2-K1+ progenitors also express *Cd14* gene. We will describe this point in the next section.

Through a series of experiments, we established a method for identifying a stem cell subpopulation that has the potential to efficiently form organoids and generate the cell populations we

need by combining quantitative data from single-cell morphometrics, organoid-forming assay, and single-cell RNA transcriptome analysis. As a result, this new method was named scMORN (Single-Cell Morphometrics, Organoid-forming assay, and RNA sequencing).

Unique features of ML-club cells in localization and transcriptome

Previously reported club cell subpopulations have been shown to exhibit unique localizations on the conducting airways: BASCs on the bronchoalveolar duct junction, UPK3A$^+$ cells around the neuroepithelial bodies, and H2-K1 progenitors in distal bronchioles [23-26]. To evaluate the localization of ML-club cells on the airway of adult mouse, we performed immunostaining for the pan-club cells and ML-club cell markers (Figures 8A and B). While the pan-club cell marker SCGB1A1 detected all club cells throughout the airway epithelium, the ML-club cell markers TFF2, MUC5B, and SCGB3A1 mainly appeared at the main bronchi of the intrapulmonary region (Figures 8A-1, 2, and B-1, 2). Most SCGB1A1$^+$ club cells at the main bronchus expressed these three ML-club cell markers. In contrast, club cells at the distal bronchiole and bronchoalveolar duct junction did not express these ML-club cell markers, suggesting that ML-club cells are different from known club cell subpopulations, such as BASCs, UPK3A$^+$, and H2-K1$^+$ cells. We conducted immunohistochemistry for secreted proteins (SCGB1A1, SCGB3A1, and MUC5B) and observed that the proximal club cells accumulated these secreted proteins into vesicle-like structures, supporting the idea that ML-club cells appear to be large

due to the large amount of accumulated secretions (Figures 8C and D).

To confirm the unique genetic features of ML-club cells, we combined our data and previously published single-cell transcriptome data of BASCs [25] and analyzed them (Figure 8E). ML-club cells showed a distinct cluster from the two BASCs that include BASCs#1 and BASCs#2 expressing specific markers, *Mfsd2a* and *Lamb3*, respectively; AT2 cells, which express *Sftpc*; and ciliated cells, which express *Foxj1* (Figures 8F and G). BASCs and ciliated cells express lower level of *Scgb1a1* than pan-club cells including ML-club cells. In the ML-club cell cluster, neither *H2-K1* nor *Upk3a* was accumulated, whereas high *Muc5b* and *Scgb3a1* expression were detected (Figure 8G), demonstrating that ML-club cells have a distinct transcriptional feature from BASCs, UPK3A$^+$, and H2-K1$^+$ cells at least in *in silico* analysis.

Proximal dissection of lung enriched ML club cells

On the contrary, H2-K1$^+$ cells express *Cd14* [23], making it difficult to completely separate ML-club cells from H2-K1$^+$ cells by FACS. To overcome this issue, the proximal and distal conducting airways were surgically separated and CD14$^+$ club cells were isolated from each tissue (Figure 9A). Proximal CD14$^+$ cells accumulated ML-club cells confirmed by qPCR (Figure 9B). Interestingly, the OFE of the proximal CD14$^+$ cells were higher than that of distal CD14$^+$ cells (Figure 9C). We further confirmed that proximal and distal CD14$^+$ cells can differentiate into alveolar epithelial cells in

differentiation medium by qPCR and IHC (Figures 9D-I). The proximal and distal $CD14^+$ cells differentiated into alveolar cells but the proximal $CD14^+$ cell-derived organoids tended to express ML-club cells markers. Unexpectedly, these ML-club cell-derived organoids are positive for p63 but not Krt5 and basal cell markers (Figures 9J-M). This observation might reflect transitional state between conducting airway and alveolar epithelial states (see Discussion). These results demonstrate that ML-club cells are distinct from other club cell subpopulations and have relatively high viability.

Based on these results, we propose that ML-club cells are a new club cell subpopulation that can generate alveolar epithelial cells *in vitro* and shows better OFE than other club and AT2 cells, at least in our organoid culture system.

ML-club cells are engrafted and expanded *in vivo* more efficiently than other club and AT2 cells

To determine the stem cell capacity of ML-club cells *in vivo*, we performed a transplantation assay with the proximal $CD14^+$ club cells enriched for ML-club cells. We prepared 3400 GFP-expressing $CD14^+$ cells from the proximal airways that is the maximum number of cells that can be collected from one transgenic mouse, 7000 GFP-expressing $CD14^+$ and $CD14^-$ club cells from whole lungs of *Scgb1a1-CreER*; *Rosa26-mTmG* mouse lines, and $Sftpc^+$ AT2 cells from *Sftpc-creERT2*; and *Rosa26-mTmG* mouse lines to transplant into nude mice. Three mice in each group were injured by bleomycin inhalation to model an acute lung injury (Figure 10A). If ML-club cells have a better

proliferation capacity than other lung epithelial stem cells, the proximal $CD14^+$ cells should expand more at the engrafted region than other types of lung epitheliums ($CD14^+$, $CD14^-$, and AT2 cells). Two weeks after transplantation, GFP of transplanted lungs indicated engrafted cells at the injured region. GFP^+ cells appeared to form foci, reflecting a clonal expansion of engrafted cells. As we expected, the proximal $CD14^+$ cells formed larger foci than $CD14^-$ or AT2 cells (Figures 10B-F). We quantified the number of cells within a focus and found that the proximal $CD14^+$ foci contain significantly more cells than others, reflecting the better proliferation capacity, while there is no significant difference in cluster formation efficiency (Figures 10G and H). We further performed immunostainings for alveolar cell types to evaluate the alveolar differentiation of engrafted cells. PDPN and SFTPC were detected in the proximal $CD14^+$ cells-derived GFP^+ clusters (Figure 10I). Most of the transplanted cells were differentiated into $SFTPC^+$ alveolar or $SCGB1A1^+$ bronchiolar cells, whereas the minor $SFTPC^+$, $SCGB1A1^+$, double positive cells, and unstained cells were also observed (Figures 10J and K).

DISCUSSION

Investigating the heterogeneity of adult tissue stem cells is still challenging because these cells are often in a dormant state, and single-cell RNA sequencing does not directly link the stem cells' potential to their expression profiles. Here, we established the scMORN method, a new single-cell analysis method, for profiling the organoid-forming capacity of tissue stem cells by integrating single-cell morphometrics, organoid-forming assay, and single-cell RNA sequencing.

In this study, single-cell morphometry and long-term live-organoid imaging of isolated club cells were used to examine the relations between stem cell morphological features and organoid-formation potential. A morphological feature of organoid-forming club cells, especially the cell size, was found from a retrospective analysis. These club cells are relatively larger than other club cells in average and can differentiate into alveolar cells. We further revealed the single-cell transcriptome of the club cells by combining automated single-cell picking technology with scRNA-seq. Based on the genetic characteristics and morphometric features of this subpopulation, we named it "ML-club cells." An *in vivo* transplantation experiment with the injured lung confirmed that the proximal $CD14^+$ club cells, which enrich ML-club cells, are engraft and expand at the damaged area more efficiently than $CD14^-$ club cells or AT2 cells and differentiate into alveolar cells that are consistent with the situation in *in vitro* organoid culture. Higher engraftment efficiency of ML-club cells than AT2 cells, consistent with a recent study, shows the need for the higher number of mature AT2 cells for effective engraftment [23,48]. Although, our accumulation method using anti-CD14 antibody is not effective enough to purify ML-club cells due to contamination of several small-club cells (Figures 7D and E). Genetic labeling approach for ML-club cells using Muc5b-CreER or Scgb3a1-CreER would overcome this issue in the future study. In our organoid culture, organoids generated from the proximal and distal $CD14^+$ cells coexpressed P63 and SFTPC not KRT5 at day 9 (Figures 9J-M). The functions of p63 in alveolar regeneration were reported and were under discussion [49,50]. Our finding might indicate that p63

expression in airway epithelial cells reflect intermediated state of transdifferentiation into alveolar epithelial cells. In humans, *MUC5B* upregulation caused by the single-nucleotide polymorphism on the promoter region is associated with familial interstitial pneumonia and idiopathic pulmonary fibrosis [51]. Hence, in the future, we should investigate this population more to know whether ML-club cells contribute to proper tissue regeneration or pathological disorder in pulmonary disease.

The scMORN method could be improved by overcoming some technical limitations. We performed live-imaging and single-cell picking as different procedures, because a machine equipped with both options does not exist. Future improvements of machines would shorten the experimental procedures and achieve high-resolution imaging to distinguish subtle morphological features by machine learning. It has been attempted to predict cellular states from cell morphology by quantifying single-cell morphological information and connecting it with scRNA-seq data. For example, single-cell imaging and scRNA-seq of fission yeast revealed that gene expression patterns were tightly related to cell size [52]. Integrating analysis with human glioblastoma samples also unveiled that there is a clear correlation between the major gene expression and basic imaging features for the malignantly transformed cells in this tumor [53]. Our study successfully identified the lung stem cell subpopulation using the scMORN method. Because human lung cells are able to expand in our feeder-free organoid culture system, the scMORN method would be applied to study human lung cells to find therapeutic resources (Figure 11). Thus, such meta-analysis combining single-cell morphometry, functional assay, and scRNA-seq

analysis has a lot of potential for further improvement of stem cell science and medical applications.

CONCLUSION

We established *in vitro* scMORN (Single-Cell Morphometrics, Organoid-forming assay, and RNA sequencing) method to identify cells with stem cell potential. With this method, we identified a subpopulation of club cells called ML-club cells that show a high capacity for proliferation and differentiation into bronchoalveolar cells in both *in vitro* organoid culture and *in vivo* transplantation experiment. This method has potentials for application to human lung cells and other types of tissue cells.

TABLES

Table 1 Measurement items in which organoid-forming cells and non-forming cells showed significantly distinct values.

Categories	Measurement items
Cell shape	Area, Perimeter, Bounding rectangle (BX, BY, Width, Height), Fit ellipse (Major axis, Minor axis), Feret's diameter (Min Feret) Shape descriptors (Solidity).
Subcellular structure	Integrated density

Table 2 Cluster marker of ML-club.

No.	Gene symbol	Average log fold-change	Adjusted p values
1	*Bpifa1*	4.21	2.05E-25
2	*Tff2*	2.99	3.11E-39
3	*Reg3g*	3.71	4.05E-39
4	*Bpifb1*	3.01	1.34E-37
5	*Muc5b*	2.28	6.21E-37
6	*Sult1d1*	1.42	8.71E-35
7	*Fxyd3*	1.72	4.43E-34
8	*Scgb3a1*	3.76	1.18E-33
9	*Lypd2*	1.89	5.06E-30
10	*Chad*	1.84	1.87E-29
11	*Gsto1*	1.82	2.10E-24
12	*Pglyrp1*	1.28	1.61E-28
13	*S100a6*	1.65	1.89E-29
14	*Scgb3a2*	1.50	3.69E-22
15	*Lgals3*	1.47	7.96E-23
16	*Aldh3a1*	1.47	1.66E-25

No.	Gene symbol	Average log fold-change	Adjusted p values
17	*Pigr*	1.37	2.05E-22
18	*Ltf*	2.54	1.06E-21
19	*Qsox1*	1.33	1.01E-20
20	*Calml3*	1.13	4.34E-21
21	*Clca1*	1.28	1.57E-10
22	*Cyp2a5*	1.95	4.26E-20
23	*Adh7*	1.22	9.72E-30
24	*AW112010*	1.20	6.48E-17
25	*Pdpn*	1.18	5.66E-29
26	*Krt19*	1.17	4.32E-17
27	*Anxa8*	1.16	9.86E-26
28	*Ly6a*	1.13	1.62E-15
29	*Oit1*	0.78	5.66E-20
30	*Agr2*	1.25	1.87E-18
31	*Anxa1*	0.99	1.26E-09
32	*Krt8*	0.99	6.36E-14
33	*Epas1*	0.98	1.11E-19
34	*Mt2*	0.98	7.09E-06
35	*Cfh*	0.93	7.88E-18
36	*Cpd*	0.91	2.69E-19
37	*Fmo3*	1.07	2.00E-16
38	*Cxcl17*	0.90	3.46E-14
39	*Sgk1*	0.90	3.30E-16
40	*Cyp4a12b*	0.73	3.14E-16
41	*Tspan1*	0.88	5.12E-12
42	*Slc16a11*	0.87	2.39E-11
43	*Cxcl5*	0.87	2.49E-07
44	*P2rx4*	0.91	1.74E-15
45	*Serpinb1a*	0.83	8.39E-16
46	*Perp*	0.84	2.59E-14
47	*Klf4*	0.82	1.93E-10

No.	Gene symbol	Average log fold-change	Adjusted p values
48	*Il13ra1*	0.82	2.48E-12
49	*Tspan13*	0.82	1.78E-14
50	*Pam*	0.82	7.86E-13
51	*Ptp4a1*	0.81	8.66E-11
52	*Kcne3*	0.64	6.24E-14
53	*Golm1*	0.83	1.85E-13
54	*Cp*	0.75	9.20E-09
55	*Zfp703*	0.80	4.75E-13
56	*Gna14*	0.60	9.37E-13
57	*Gm13363*	0.71	6.44E-12
58	*C3*	0.71	1.34E-06
59	*Card10*	0.55	8.46E-12
60	*Mt1*	0.71	8.66E-03
61	*Mfge8*	0.70	3.91E-14
62	*Ceacam1*	0.68	9.17E-08
63	*Ano1*	0.68	6.02E-11
64	*Lif*	0.67	7.14E-06
65	*Timp2*	0.67	2.18E-07
66	*Hspb1*	0.66	3.62E-06
67	*Foxq1*	0.90	1.35E-11
68	*Bhlhe40*	0.66	5.03E-07
69	*Idh1*	0.65	2.38E-06
70	*Basp1*	0.65	2.80E-05
71	*Slc15a2*	0.74	1.72E-11
72	*Gp2*	0.64	7.11E-12
73	*Tmem176a*	0.64	9.37E-07
74	*Tmem176b*	0.62	1.84E-07
75	*F3*	0.61	6.69E-04
76	*Cfb*	0.60	7.01E-06
77	*Cd14*	0.72	2.81E-10
78	*St8sia6*	0.60	7.28E-10

No.	Gene symbol	Average log fold-change	Adjusted p values
79	*Cyp2a4*	0.60	3.29E-08
80	*Anxa2*	0.59	2.23E-06
81	*Mtus1*	0.58	1.38E-05
82	*Tstd1*	0.58	4.71E-07
83	*Aldh1a1*	0.58	5.68E-05
84	*Lrrc26*	0.57	7.15E-05
85	*Elf3*	0.57	1.39E-03
86	*Ptp4a2*	0.56	4.63E-05
87	*Tubb2a*	0.56	2.36E-08
88	*Endod1*	0.56	2.53E-06
89	*Mlph*	0.66	1.38E-08
90	*Msln*	0.55	1.40E-06
91	*Pon1*	0.55	6.04E-05
92	*Ifitm1*	0.55	1.41E-10
93	*Ctgf*	0.55	1.50E-07
94	*Fcgbp*	0.54	5.15E-06
95	*Cotl1*	0.54	1.88E-06
96	*Sfn*	0.54	6.53E-04
97	*S100a13*	0.54	4.17E-04
98	*Mia3*	0.53	1.20E-04
99	*Runx1*	0.53	1.43E-03
100	*Wfdc2*	0.52	3.65E-09
101	*Clic6*	0.52	7.10E-07
102	*Slc12a2*	0.51	1.19E-03
103	*Pax9*	0.51	1.57E-10
104	*Gipc2*	0.51	4.30E-07
105	*Ahnak*	0.51	1.08E-04
106	*St3gal4*	0.49	1.60E-04
107	*Fam46b*	0.49	2.52E-09
108	*Cyp4b1*	0.49	1.65E-03
109	*Psap*	0.48	1.29E-03

No.	Gene symbol	Average log fold-change	Adjusted p values
110	*D17H6S56E-5*	0.48	1.71E-04
111	*Isg20*	0.48	7.79E-04
112	*Gabrp*	0.48	5.41E-04
113	*Atp6v1b1*	0.48	1.02E-07
114	*Muc4*	0.47	6.94E-03
115	*Slc45a3*	0.47	1.92E-07
116	*Ece1*	0.47	5.29E-04
117	*Sox21*	0.46	3.33E-09
118	*Slc44a4*	0.46	5.71E-05
119	*Krt18*	0.46	1.50E-02
120	*Pgap2*	0.46	7.68E-03
121	*Gpd1l*	0.46	7.53E-04
122	*Mgst1*	0.45	3.73E-04
123	*Pllp*	0.45	1.15E-05
124	*Tst*	0.44	6.37E-03
125	*Arf4*	0.44	3.66E-02
126	*Adam28*	0.44	1.42E-05
127	*Rdh10*	0.44	1.22E-03
128	*1700047I17Rik2*	0.44	1.78E-05
129	*Fam177a*	0.44	1.78E-05
130	*Pof1b*	0.43	3.28E-06
131	*Adgre5*	0.43	2.51E-03
132	*Porcn*	0.43	1.93E-03
133	*Cd2ap*	0.43	1.09E-03
134	*Rell1*	0.43	3.29E-06
135	*Fam3c*	0.42	3.08E-02
136	*Pir*	0.42	3.98E-03
137	*Cbr3*	0.42	2.00E-02
138	*Nbl1*	0.42	9.57E-04
139	*Arrdc3*	0.41	5.26E-03
140	*Psmc5*	0.41	2.77E-03

No.	Gene symbol	Average log fold-change	Adjusted p values
141	*Sox2*	0.40	6.88E-03
142	*Shisa5*	0.40	4.10E-03
143	*Cd24a*	0.40	2.69E-02
144	*Nans*	0.40	9.06E-05
145	*Rhoc*	0.40	1.06E-03
146	*Morf4l1b*	0.40	4.61E-02
147	*Parva*	0.39	5.20E-03
148	*Iqgap2*	0.39	2.25E-03
149	*Rara*	0.39	1.37E-02
150	*Tubb4b*	0.39	1.27E-02
151	*Sec11c*	0.39	3.08E-02
152	*Ddrgk1*	0.39	2.08E-03
153	*Nek7*	0.39	5.96E-03
154	*S100a16*	0.39	7.50E-04
155	*Morf4l1*	0.38	3.00E-02
156	*S100a14*	0.38	2.90E-04
157	*Mgat3*	0.37	2.87E-02
158	*Emp1*	0.37	2.67E-02
159	*Misp*	0.37	9.79E-04
160	*Atp2c2*	0.37	5.16E-04
161	*Morf4l1-ps1*	0.37	5.19E-03
162	*Ppp1r1b*	0.36	2.51E-02
163	*S100a11*	0.36	1.51E-02
164	*Vasp*	0.36	4.27E-03
165	*Galnt7*	0.36	4.16E-06
166	*Tnfaip8*	0.36	8.07E-06
167	*Uqcc2*	0.36	6.27E-03
168	*Bmpr1b*	0.35	2.55E-02
169	*Aqp4*	0.35	6.80E-04
170	*Wnt7b*	0.35	3.94E-03
171	*Id1*	0.35	1.36E-02

No.	Gene symbol	Average log fold-change	Adjusted p values
172	*Plscr1*	0.35	2.73E-02
173	*Otx1*	0.35	6.53E-07
174	*Gramd3*	0.35	1.01E-04
175	*Ildr1*	0.35	1.06E-03
176	*Slk*	0.34	2.44E-02
177	*Tmem30b*	0.34	2.17E-02
178	*Sgms2*	0.32	1.27E-04
179	*Ptprs*	0.32	1.50E-02
180	*Wfdc1*	0.32	8.63E-03
181	*Fam20b*	0.31	3.10E-02
182	*Cxcl16*	0.30	1.13E-02
183	*Foxp2*	0.28	1.86E-02
184	*Vtcn1*	0.25	4.54E-02

FIGURES

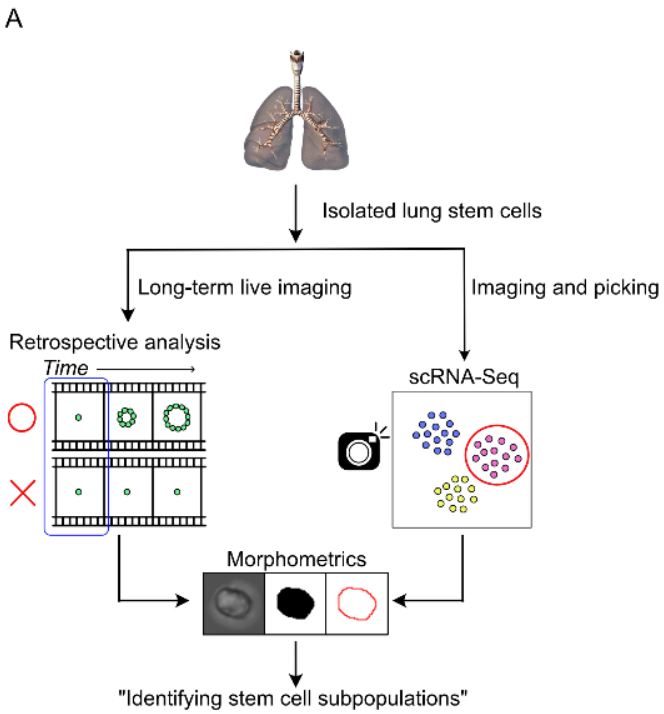

Figure 1. Overview of novel established *in vitro* identifying method of stem cells

(A) Overview of the scMORN method.

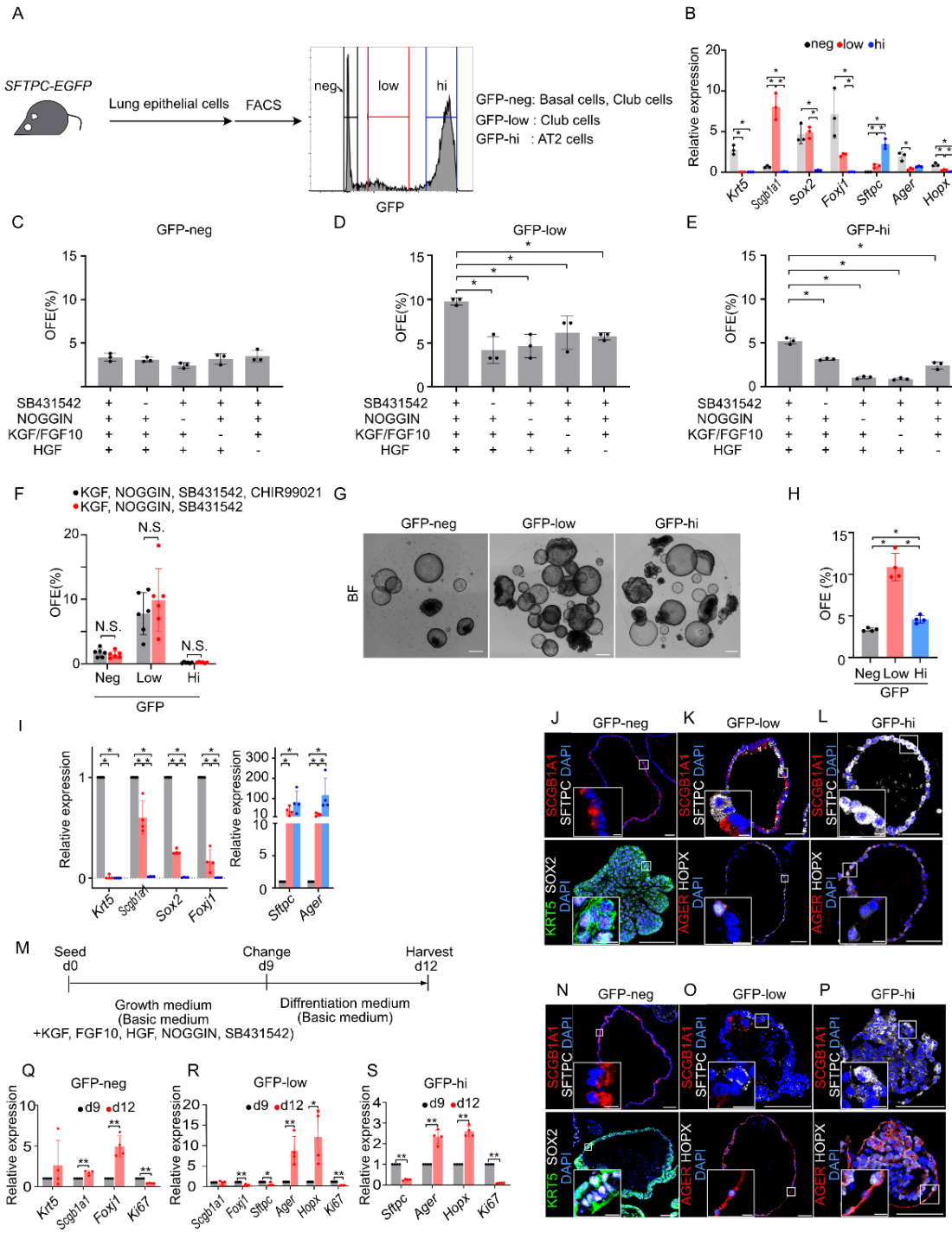

Figure 2. Evaluation of various additives to establish feeder-free organoid culture system with various lung epithelial cells

(A) Experimental scheme of sorting major lung epithelial stem cells from *SFTPC-GFP* mice.

(B) Quantification for lung cell lineage marker expression of GFP$^{neg/low/hi}$ cells using qRT-PCR. Data are presented as relative expression to whole lung cells (n = 3 mice). Airway genes: *Krt5*, *Scgb1a1*,

Sox2 and *Foxj1*. Alveolar genes: *Sftpc* and *Ager*.

(C)(D)(E) Organoid-forming efficiency of each *SFTPC-GFP* population (C, GFP^{neg}. D, GFP^{low}. E, GFP^{hi}.) under indicated conditions (n = 3 wells)

(F) Organoid-forming efficiency of each population under indicated conditions (n = 6 mice).

(G) Representative bright-field images of organoids at culture day 9.

(H) OFE of each $GFP^{neg/low/hi}$ cells under the feeder-free culture condition (n = 3 mice).

(I) Quantification for lung cell lineage marker expression of day 9 organoids derived from $GFP^{neg/low/hi}$ cells. Data are presented as relative expression to GFP^{neg} (n = 4 mice). Conducting airway genes: *Krt5*, *Scgb1a1*, *Sox2,* and *Foxj1*. Alveolar genes: *Sftpc* and *Ager*.

(J-L) (N-P) Immunostaining of organoids derived from $GFP^{neg/low/hi}$ cells for SCGB1A1, SFTPC, KRT5, ACTUB, SOX2, AGER, and HOPX, and DAPI staining at day 9 (J-L) and day 12 (N-P).

(M) Experimental scheme for lung stem cell amplification culture and differentiation.

(Q-S) Quantification for lung cell lineage marker expression of organoids at day 9 and day 12 by qRT-PCR (n = 4 mice). Data are presented as the relative expression of each sample at day 9.

All quantified data are presented as mean ± SD $**p < 0.01$, $*p < 0.05$ and N.S. (not significant) using one-way ANOVA followed by Tukey's multiple comparisons test (B, C, D, E, H) and two-tailed unpaired Student's t-test (F, I, Q, R, S). Scale bars in images 500 μm (G) and100 μm (J, K, L, N, O, P); in enlarged images of the white boxed region 10 μm (J, K, L, N, O, P).

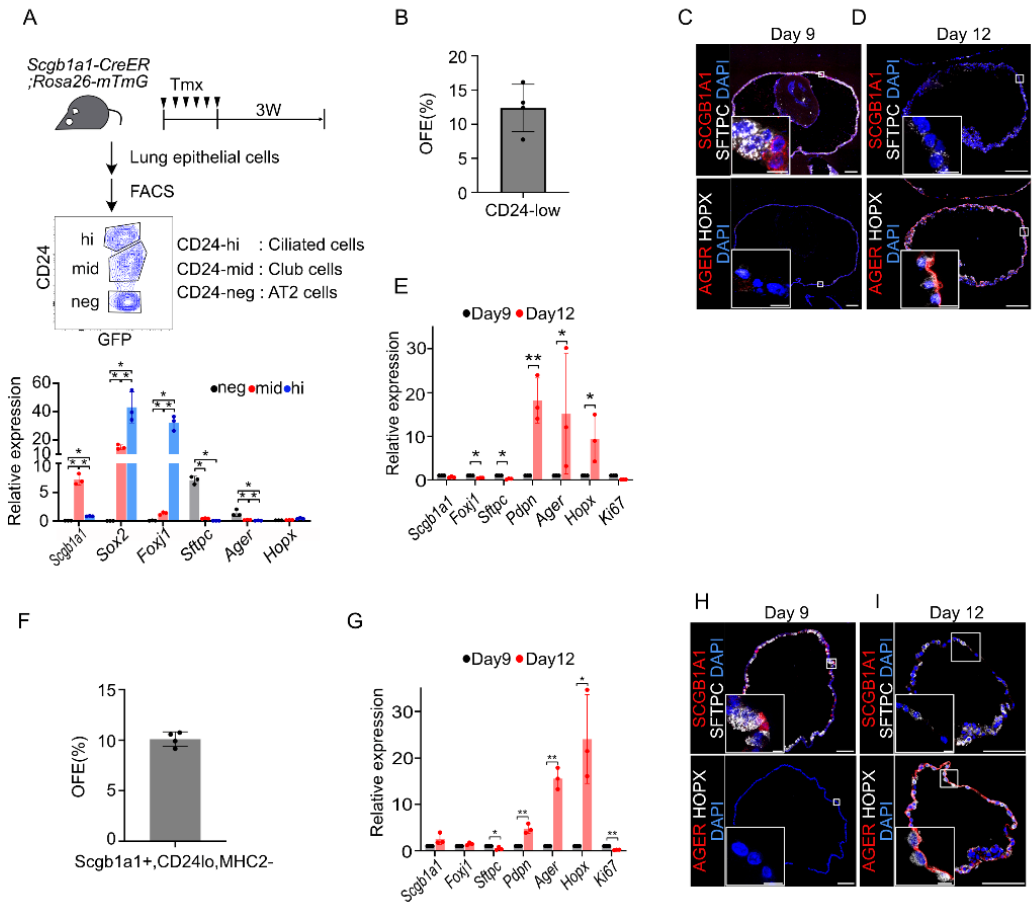

Figure 3. Club Cell Subpopulation Transdifferentiates into Alveolar Cell

(A) Experimental scheme of sorting Scgb1a1-lineage-positive cells from *Scgb1a1-creER; Rosa26-mTmG* mice. Quantification of lung cell lineage marker expression in the sorted populations ($CD24^{neg/mid/hi}$) by qRT-PCR. Data are presented as relative expression to whole lung cells (n = 3 mice).

(B) OFE of each $CD24^{mid}$, Scgb1a1-lineage cells (n = 4 mice).

(C and D) Immunostaining of organoids derived from $Scgb1a1^{+}$, $CD24^{mid}$ club cells for SCGB1A1, SFTPC AGER, and HOPX on day 9 (C) and day12 (D).

(E) qPCR analysis was performed using organoids derived from $Scgb1a1^{+}$, $CD24^{mid}$ club cells at day 9 and day 12 (n = 3 mice). Data are presented as relative expression on day 9.

(F) OFE of each $CD24^{mid}$, $MHC2^{neg}$, and Scgb1a1-lineage cells (n = 4 mice).

(G) qPCR analysis was performed using organoids derived from Scgb1a1+, $CD24^{mid}$, and $MHC2^{neg}$ club cells at day 9 and day 12 (n = 3 mice). Data are presented as relative expression on day 9.

(H and I) Immunostaining of organoids derived from Scgb1a1+, CD24mid, and MHC2neg club cells for SCGB1A1, SFTPC, AGER, and HOPX, and DAPI staining at day 9 (H) and day12 (I)

All quantification data are presented as mean ± SD **p < 0.01, *p<0.05 using one-way analysis of variance followed by Tukey's multiple comparisons test (A) and two-tailed unpaired Student's t-test

(E, G). Scale bars in images, 100 μm (C, D, H, I); in enlarged images of the white boxed region, 10 μm (C, D, H, I).

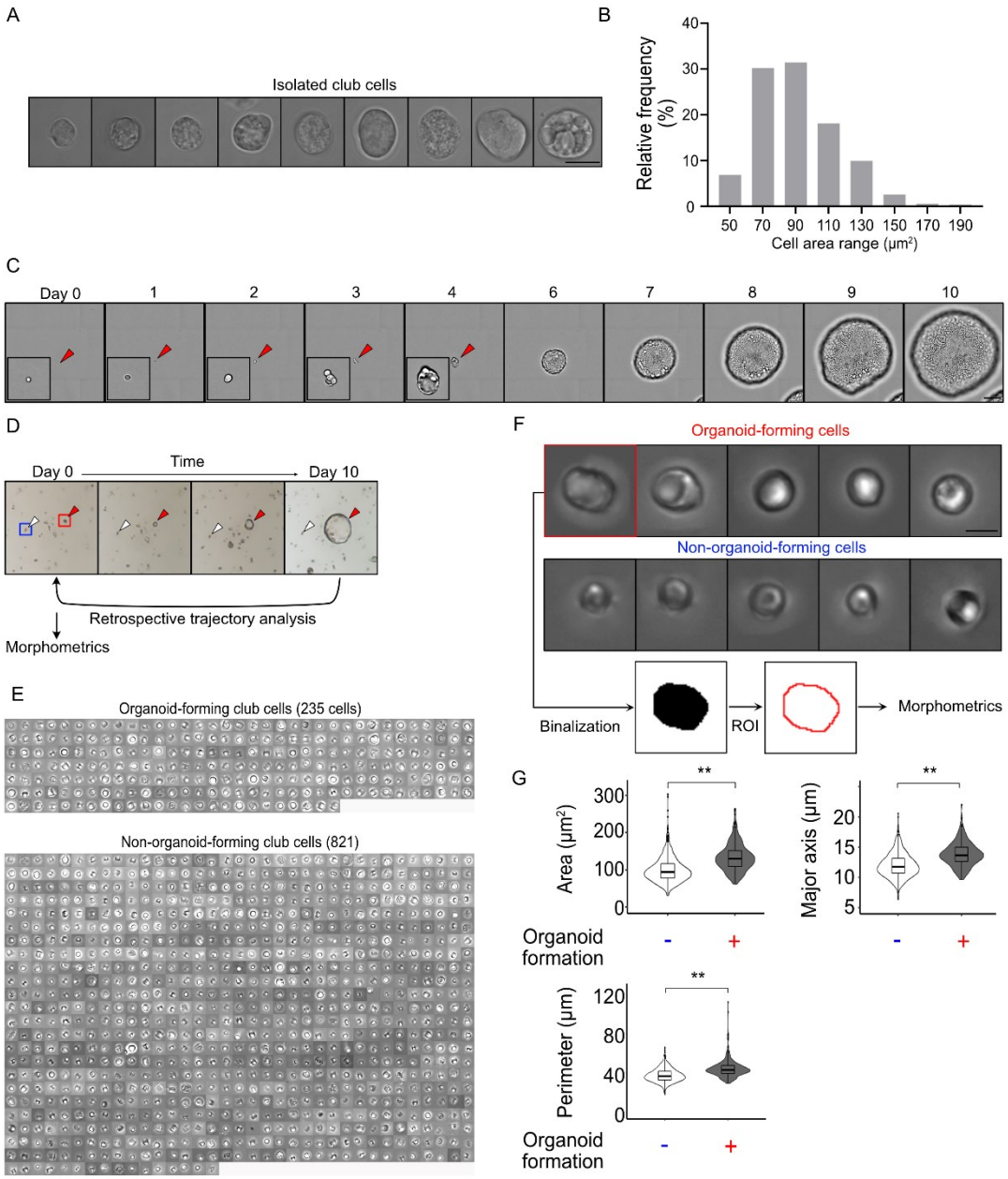

Figure 4. Single-cell Morphometry Following Long-term Live-Organoid Imaging Unveiled Morphological Features of the club Cell Subpopulation.

(A) Representative images of single isolated club cells derived from *Scgb1a1-CreER*; *Rosa26-mTmG* mice. Scale bar: 10 μm.

(B) The relative frequency of isolated club cells at different cell size.

(C) Long-term time-lapse images of developing lung organoid for ten days. Insets are high magnification images of red-arrowhead-indicated single cells and organoids. Scale bars: 200 μm.

(D) Representative images of retrospective trajectory analysis. A stem cell that forms a lung organoid

(red arrowheads) and does not form an organoid (white arrowheads) are indicated.
(E) Each single-cell images of organoid-forming cells (left, n=235) and non-organoid-forming cells (right, n=821) taken by Celldiscoverer7 (see Methods). Scale bar: 10 μm.
(F) Images of organoid-forming cells (upper) and non-organoid-forming cells (middle) taken by a microscope equipped with an incubator (see Methods). The ImageJ analyzed each single-cell image and lined by ROI. for morphometrics (lower, see Methods for details). Scale bar: 10 μm.
(G) Violin and box plots of the cell area, the length of the major axis, and perimeter of each single cells (Non-organoid-forming cells (n = 821), Organoid-forming cells (n = 235)). These experiments were independently performed three times. **$p < 0.01$ (Wilcoxon rank-sum test).

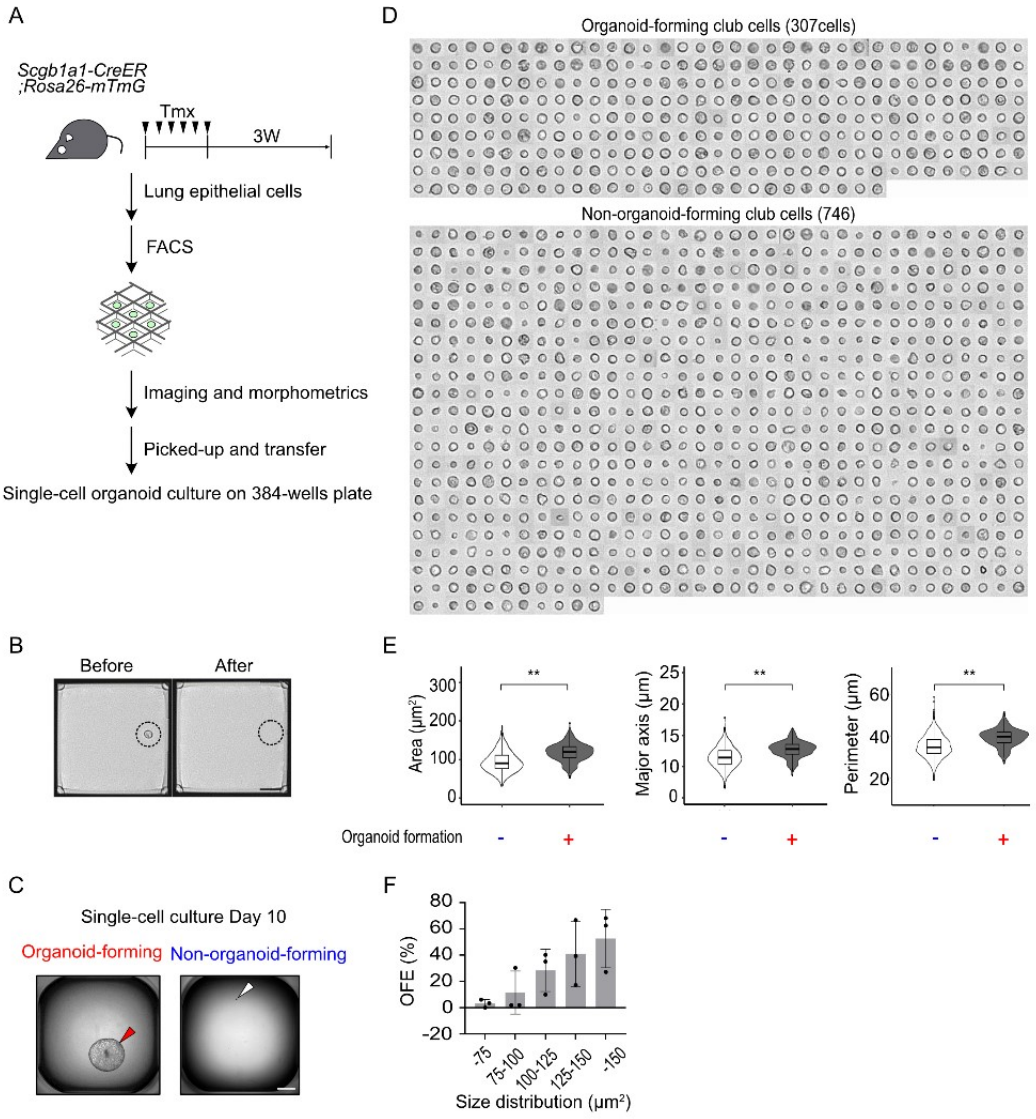

Figure 5. Larger club cells efficiently form organoid confirmed by using cell picking and imaging robot

(A) Experimental scheme of a series of experiments with single-cell imaging, picking, and organoid culture.

(B) Representative single-cell image in micro-well of Elplasia. Dashed line circles show the same region before and after picking. Scale bars: 50 μm.

(C) Organoid culture in 384-well plate at day 10. The red arrowhead indicates growing lung organoid, and the white arrowhead indicates non-organoid-forming cells. Scale bars: 500 μm.

(D) Each single-cell images of organoid-forming cells (upper, n=307) and non-organoid-forming cells (lower, n=746) taken by Cellhandler (see Methods). Scale bars: 10 μm.

(E) Violin and box plots of the cell area, the length of major axis and perimeter of each single cell (Non-organoid-forming cells (n = 746), Organoid-forming cells (n = 307)). These experiments were

independently performed three times. **p < 0.01 (Wilcoxon rank-sum test).

(F) The ratio of organoid forming cells of isolated Scgb1a1+, $CD24^{mid}$ cells at different cell size.

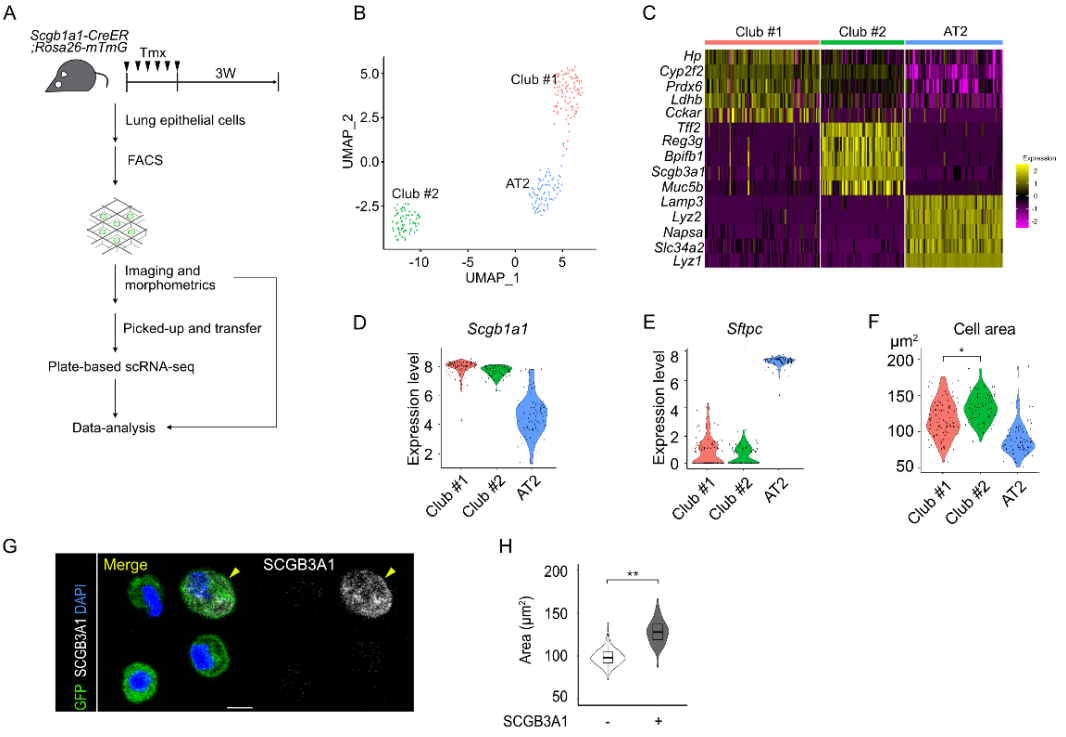

Figure 6. scRNA-seq Revealed Unique Transcriptional Features of ML-club cells.

(A) Experimental scheme of a series of experiments with single-cell imaging, picking, and a plate-based scRNA-seq analysis.

(B) Unsupervised UMAP clustering analysis revealed three distinct clusters in Scgb1a1+ and CD24^{mid} cells.

(C) Heatmap analysis of the five top-upregulated genes of each cluster.

(D-F) Violin plots of expression level of *Scgb1a1* (D), *Sftpc* (E), and cell area distribution (F) of each cluster. *$p < 0.05$ (Wilcoxon rank-sum test between club #1 and club #2).

(G) Representative Immunostaining image of Scb1a1+ club cells for SCGB3A1. Scale bars in images, 10 μm.

(H) Violin and box plots of the cell area of each single cell (SCGB3A1- cells (n = 504), Organoid-forming cells (n = 70)). **$p < 0.01$ (Wilcoxon rank-sum test).

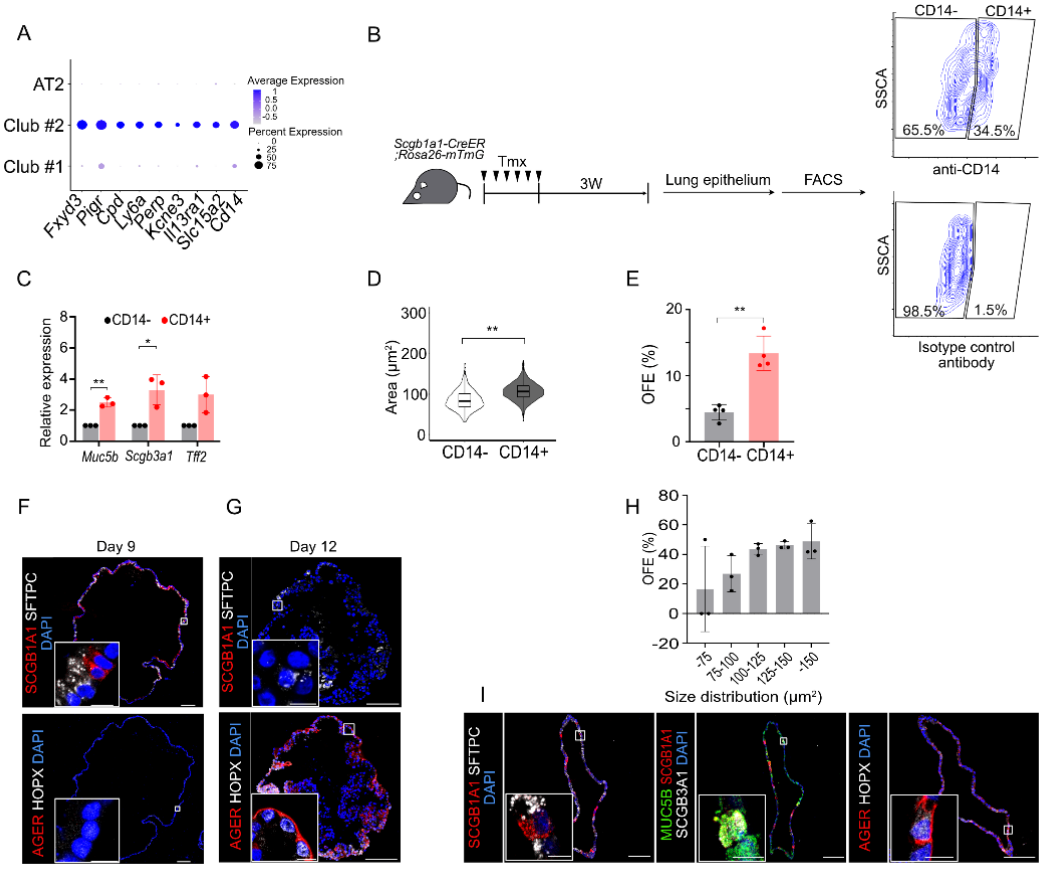

Figure 7. Isolated ML-club cells efficiently formed organoid and differentiated into alveolar lineage cells *in vitro*

(A) Dot plot analysis of nine cell surface antigens for club #2 (ML-club cells).

(B) Experimental scheme of sorting CD14+ and CD14− club cells using FACS.

(C) Quantification for ML-club cell marker expression of sorted CD14+ and CD14− club cells by qRT-PCR. Data are presented as relative expression to CD14− club cells and as mean ± SD (n = 3 mice). **p < 0.01, *p<0.05 (two-tailed unpaired Student's t-test).

(D) Violin and box plot of the cell area of each CD14+ (n =1353) and CD14− (n = 830) club cells. These experiments were repeated three times. **p < 0.01 (Wilcoxon rank-sum test).

(E) OFE of each CD14+ and CD14− club cells. Data are presented as mean ± SD (n = 4 mice). **p < 0.01 (two-tailed unpaired Student's t-test).

(F and G) Immunostaining of organoids derived from CD14+ club cells for lung cell lineage markers, SCGB1A1, SFTPC, AGER, and HOPX, and DAPI staining at culture day 9 (F) and day 12 (G).

(H) The ratio of organoid forming cells of isolated CD14+ cells at different cell size (n = 967). These experiments were independently performed three times.

(I) Immunostaining of organoids derived from more than 140 μm^2 CD14+ club cells for lung cell lineage markers, SCGB1A1, SFTPC, MUC5B, SCGB3A1, AGER, and HOPX, and DAPI staining at

culture day 10 with serial sections.

Scale bars in images (F, G, I), 100 μm; in enlarged images of the white boxed region, 10 μm (F, G, I).

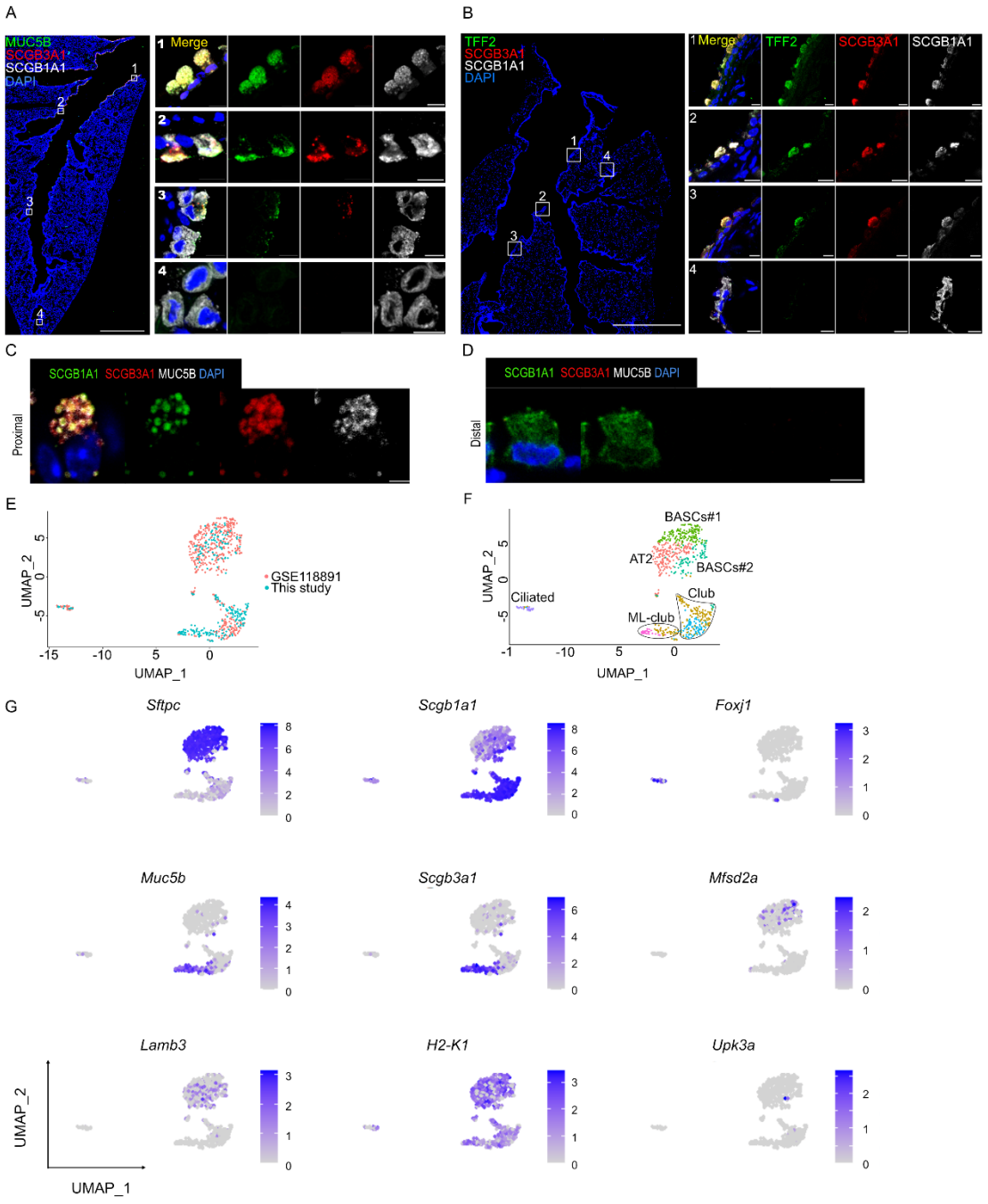

Figure 8. Unique features of ML-club cells in localization and transcriptome

(A) Distribution of club and ML-club cells in adult lung detected by immunostaining for MUC5B, SCGB3A1, and SCGB1A1, and DAPI staining. Scale bars in the left panel, 1 mm; in enlarged images,

10 μm.

(B) Immunostaining of a wild mouse lung for TFF2 (green), SCGB3A1 (red), SCGB1A1 (white) and DAPI (blue). Scale bars in left panel, 1 mm; in enlarged images, 10 μm.

(C and D) Representative proximal (C) and distal (D) airway cell image immunostained for SCGB1A1, SCGB3A1, and MUC5B. Scale bars in images, 5 μm.

(E and F) Dimension plot (E) and unsupervised UMAP clustering analysis (F) of integrated data of GSE118891 and this study's data.

(G) The expression patterns of lineage markers for cell types specific markers (*Sftpc*: AT2, *Scgb1a1*: Club, *Foxj*: Ciliated, *Muc5b* and *Scgb3a1*: ML-Club, *Mfsd2a*: BASCs#1, *Lamb3*: BASCs#2, *H2-K1*:H2K1 progenitor, *Upk3a*: UPK3A+ Club) with integrated data of GSE118891 and this study's data.

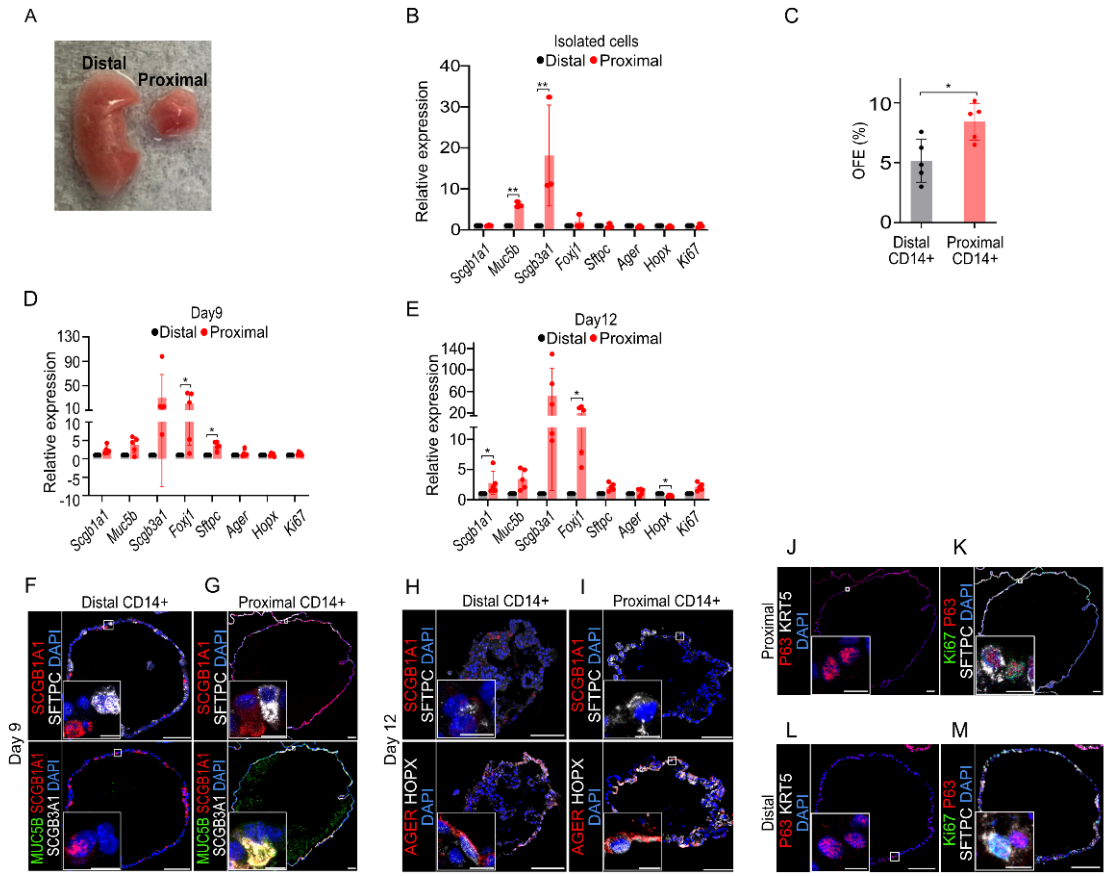

Figure 9. Proximal dissection of lung enriched ML-club cells

(A) Image of distal and proximal region of left lobe.

(B) qPCR analysis was performed using isolated distal and proximal CD14+ club cells. Data are presented as relative expression of distal CD14+ cells.

(C) OFE of each distal and proximal CD14+ club cells (n=5 mice)

(D and E) qPCR analysis was performed using organoids derived from distal and proximal CD14+ Scgb1a1-lineage cells at day9 (D) and day12 (E) (n = 5 mice). Data are presented as relative expression of distal CD14+ organoids. All quantification data are presented as mean ± SD **p < 0.01, *p<0.05 using one-way analysis of variance followed by two-tailed unpaired Student's t-test (B, C, D, E).

(F-I) Immunostaining of organoids derived from distal and proximal CD14+ club cells for lung cell lineage markers, SCGB1A1, SFTPC, MUC5B, SCGB3A1, AGER, and HOPX, and DAPI staining at culture day 9 (F, G) and day 12 (H, I).

(J-M) Immunostaining of organoids derived from proximal (J, K) and distal (L, M) CD14+ club cells for P63 (J, K, L, M), KRT5 (J, L), Ki67 (K, M), and SFTPC (K, M), and DAPI staining at day 9.

Scale bars in these images, 100 μm (F-M); in enlarged images of white boxed region, 10 μm (F-M)

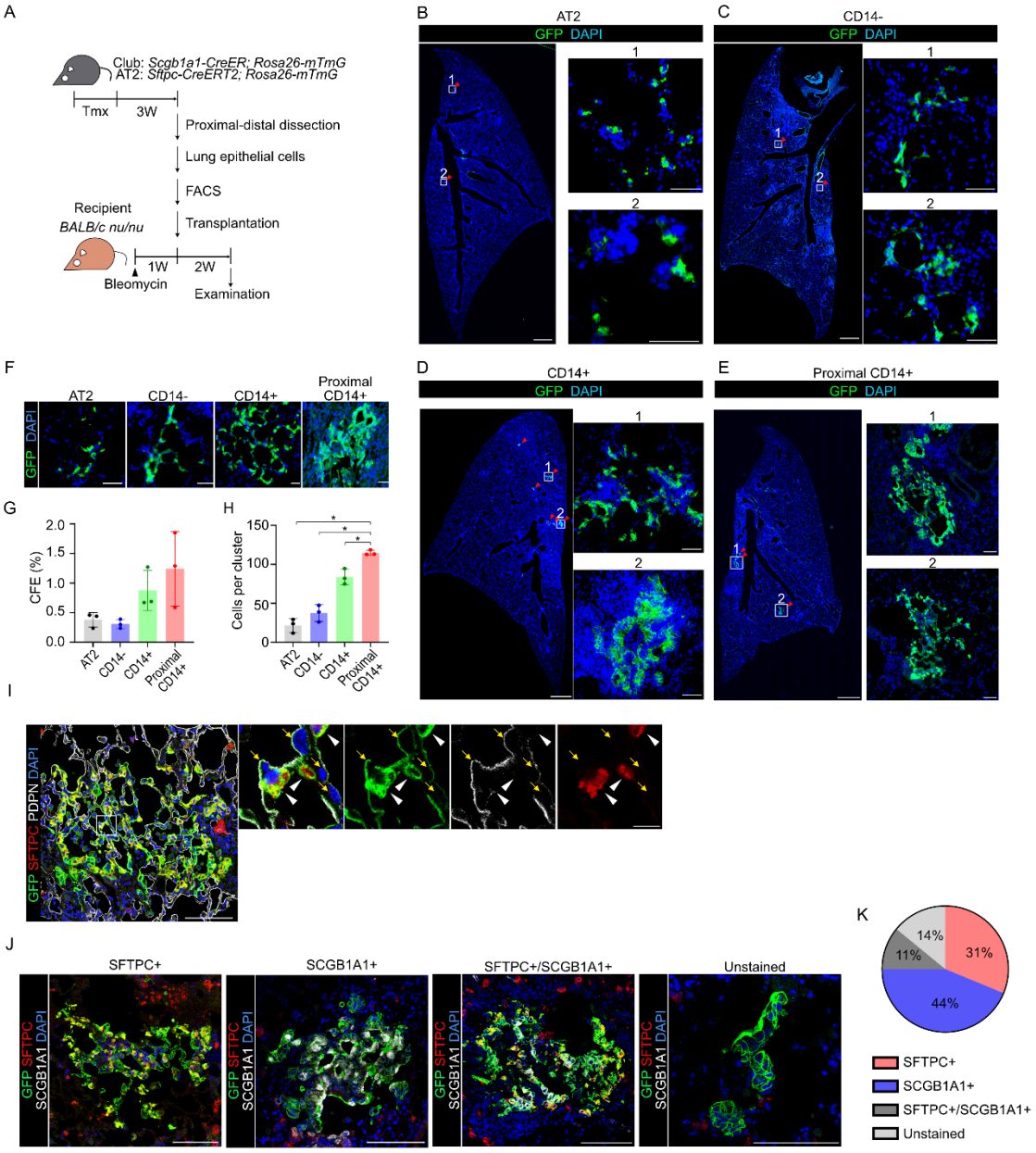

Figure 10. ML-club cells are engrafted and expand in vivo more than other club and AT2 cells. See also Figure S3.

(A) Experimental scheme of club and AT2 cells transplantation into bleomycin-injured nude mice.

(B-E) Representative images of transplanted cells (AT2 (B), CD14- (C), CD14+ (D), and proximal CD14+ (E) which express membrane-localized GFP. Counter stain with DAPI. Scale bars in left panels, 1 mm; in enlarged images of white boxed region, 50 μm.

(F) Detection of transplanted AT2, CD14− club, CD14+ club, and proximal CD14+ club cells with GFP fluorescence. Scale bars: 50 μm.

(G-H) Quantification of the CFE (cluster formation efficiency) (G) and the number of GFP+ cells per cluster (H) in the lungs transplanted with AT2, CD14−, and CD14+ club cells. Data are presented as mean ± SD (n = 3 mice). *p<0.05 (one-way ANOVA followed by Tukey's multiple comparisons test).
(I) Immunostaining to detect transplanted GFP-expressing proximal CD14+ club and lung lineage markers, Sftpc and Pdpn, and DAPI staining. Right panels show higher magnification of the white boxed region. White arrowheads for Sftpc-expressing GFP+ cells. Yellow arrows for PDPN-expressing GFP+ cells. Scale bars in the left panel, 50 μm; in enlarged images, 10 μm.
(J) Immunostaining to determine colocalization of GFP and lung lineage markers, SFTPC and SCGB1A1. Scale bars: 50 μm.
(K) The pie chart shows the ratio of lineage-positive GFP clusters (n=128 clusters from three different experiments).

A

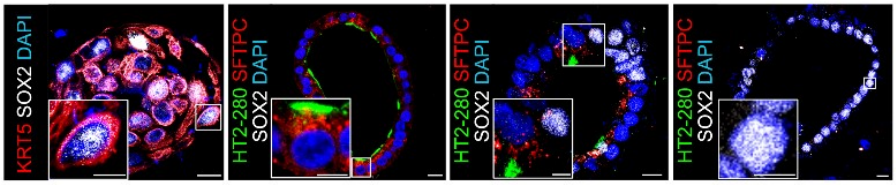

Figure 11. Organoids derived from human lung cells

(A) Immunostaining of organoids derived from human lung epitheliums for KRT5 (red), SOX2 (white) and DAPI (blue) in leftmost panels and HT2-280 (Green), SFTPC (red), SOX2 (white) and DAPI (blue) in other panels at day9. HPSAEpiC (Human Small Airway Epithelial Cells, #3230, ScienCell) and HPAEpiC (Human Plumonary Alveolar Epithelial Cells, #3200, ScienCell) were used for organoid culture.

Scale bars in images, 100 μm; in enlarged images of the white boxed region, 5 μm .

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to Prof. Dr. Mitsuru Morimoto for all supports and advices in my research for over 5 years. He encouraged me and led the way.

I thank collaborators and coauthors of the study presented in Chapter II; Dr. Yasunori Enomoto, Dr. Hiroaki Katsura, Dr. Taisaku Ogawa, Dr. Katsuyuki Shiroguchi, Ms. Saori Baba, Mr. Akira Ogata, and Ms. Akira Yamaoka. They give me many supports and advices, and encouraged me when I was on the fence or discouraged.

I thank all members and alumnus of Morimoto laboratory for giving good research life to me. Ms. Yuka Noda gave me constructive comments and warm encouragement. Dr. Keishi Kishimoto and Dr. Hirofumi Kiyokawa gave me useful advices and discussions.

Last but not the least, I would like to thank my family members, especially my wife Hitomi Fujimura for her understanding and all supports she has continuously given me to complete my research and dissertation.

REFERENCES

1 Barkauskas, C. E. *et al.* Type 2 alveolar cells are stem cells in adult lung. *J. Clin. Invest.* **123**, 3025-3036, doi:10.1172/jci68782 (2013).

2 Barker, N. *et al.* Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro. *Cell Stem Cell* **6**, 25-36, doi:10.1016/j.stem.2009.11.013 (2010).

3 Fatehullah, A., Tan, S. H. & Barker, N. Organoids as an in vitro model of human development and disease. *Nat. Cell Biol.* **18**, 246-254, doi:10.1038/ncb3312 (2016).

4 Huch, M. *et al.* In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature* **494**, 247-250, doi:10.1038/nature11826 (2013).

5 Karthaus, W. R. *et al.* Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* **159**, 163-175, doi:10.1016/j.cell.2014.08.017 (2014).

6 Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262-265, doi:10.1038/nature07935 (2009).

7 Chua, C. W. *et al.* Single luminal epithelial progenitors can generate prostate organoids in culture. *Nat. Cell Biol.* **16**, 951-961, 951-954, doi:10.1038/ncb3047 (2014).

8 DeWard, A. D., Cramer, J. & Lagasse, E. Cellular heterogeneity in the mouse esophagus implicates the presence of a nonquiescent epithelial stem cell population. *Cell Rep* **9**, 701-711, doi:10.1016/j.celrep.2014.09.027 (2014).

9 Huch, M. *et al.* Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* **160**, 299-312, doi:10.1016/j.cell.2014.11.050 (2015).

10 Zepp, J. A. *et al.* Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung. *Cell* **170**, 1134-1148.e1110, doi:10.1016/j.cell.2017.07.034 (2017).

11 Lukonin, I. *et al.* Phenotypic landscape of intestinal organoid regeneration. *Nature* **586**, 275-280, doi:10.1038/s41586-020-2776-9 (2020).

12 Hogan, B. L. *et al.* Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. *Cell Stem Cell* **15**, 123-138, doi:10.1016/j.stem.2014.07.012 (2014).

13 Zhao, F., Ma, Q., Yue, Q. & Chen, H. SARS-CoV-2 Infection and Lung Regeneration. *Clin. Microbiol. Rev.* **35**, e0018821, doi:10.1128/cmr.00188-21 (2022).

14 Kiyokawa, H. & Morimoto, M. Molecular crosstalk in tracheal development and its recurrence in adult tissue regeneration. *Dev. Dyn.*, doi:10.1002/dvdy.345 (2021).

15 Kotton, D. N. & Morrisey, E. E. Lung regeneration: mechanisms, applications and emerging stem cell populations. *Nat. Med.* **20**, 822-832, doi:10.1038/nm.3642 (2014).

16 Adamson, I. Y. & Bowden, D. H. The type 2 cell as progenitor of alveolar epithelial regeneration. A cytodynamic study in mice after exposure to oxygen. *Lab. Invest.* **30**, 35-42 (1974).

17 Desai, T. J., Brownfield, D. G. & Krasnow, M. A. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* **507**, 190-194, doi:10.1038/nature12930 (2014).

18 Rock, J. R. & Hogan, B. L. Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu. Rev. Cell. Dev. Biol.* **27**, 493-512, doi:10.1146/annurev-cellbio-100109-104040 (2011).

19 Nabhan, A. N., Brownfield, D. G., Harbury, P. B., Krasnow, M. A. & Desai, T. J. Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* **359**, 1118-1123, doi:10.1126/science.aam6603 (2018).

20 Zacharias, W. J. *et al.* Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* **555**, 251-255, doi:10.1038/nature25786 (2018).

21 Rock, J. R. *et al.* Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proc Natl Acad Sci U S A* **108**, E1475-1483, doi:10.1073/pnas.1117988108 (2011).

22 Kim, C. F. *et al.* Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121**, 823-835, doi:10.1016/j.cell.2005.03.032 (2005).

23 Kathiriya, J. J., Brumwell, A. N., Jackson, J. R., Tang, X. & Chapman, H. A. Distinct Airway Epithelial Stem Cells Hide among Club Cells but Mobilize to Promote Alveolar Regeneration. *Cell Stem Cell* **26**, 346-358.e344, doi:10.1016/j.stem.2019.12.014 (2020).

24 Guha, A., Deshpande, A., Jain, A., Sebastiani, P. & Cardoso, W. V. Uroplakin 3a(+) Cells Are a Distinctive Population of Epithelial Progenitors that Contribute to Airway Maintenance and Post-injury Repair. *Cell Rep* **19**, 246-254, doi:10.1016/j.celrep.2017.03.051 (2017).

25 Liu, Q. *et al.* Lung regeneration by multipotent stem cells residing at the bronchioalveolar-duct junction. *Nat. Genet.* **51**, 728-738, doi:10.1038/s41588-019-0346-6 (2019).

26 Salwig, I. *et al.* Bronchioalveolar stem cells are a main source for regeneration of distal lung epithelia in vivo. *EMBO J.* **38**, doi:10.15252/embj.2019102099 (2019).

27 Vaughan, A. E. *et al.* Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature* **517**, 621-625, doi:10.1038/nature14112 (2015).

28 Hong, K. U., Reynolds, S. D., Giangreco, A., Hurley, C. M. & Stripp, B. R. Clara cell secretory protein-expressing cells of the airway neuroepithelial body microenvironment include a label-retaining subset and are critical for epithelial renewal after progenitor cell depletion. *Am J Respir Cell Mol Biol* **24**, 671-681, doi:10.1165/ajrcmb.24.6.4498 (2001).

29 Sueblinvong, V. & Weiss, D. J. Stem cells and cell therapy approaches in lung biology and diseases. *Transl Res* **156**, 188-205, doi:10.1016/j.trsl.2010.06.007 (2010).

30 Fan, H. C., Fu, G. K. & Fodor, S. P. Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science* **347**, 1258367, doi:10.1126/science.1258367 (2015).

31 Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201, doi:10.1016/j.cell.2015.04.044 (2015).

32 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

33 Lo, B., Hansen, S., Evans, K., Heath, J. K. & Wright, J. R. Alveolar epithelial type II cells induce T cell tolerance to specific antigen. *J. Immunol.* **180**, 881-888, doi:10.4049/jimmunol.180.2.881 (2008).

34 Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593-605, doi:10.1002/dvg.20335 (2007).

35 Rawlins, E. L. *et al.* The role of Scgb1a1+ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell* **4**, 525-534, doi:10.1016/j.stem.2009.04.002 (2009).

36 Thévenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans Image Process* **7**, 27-41, doi:10.1109/83.650848 (1998).

37 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821, doi:10.1016/j.cell.2019.05.031 (2019).

38 Chen, H. *et al.* Airway epithelial progenitors are region specific and show differential responses to bleomycin-induced lung injury. *Stem Cells* **30**, 1948-1960, doi:10.1002/stem.1150 (2012).

39 Lee, J. H. *et al.* Anatomically and Functionally Distinct Lung Mesenchymal Populations Marked by Lgr5 and Lgr6. *Cell* **170**, 1149-1163.e1112, doi:10.1016/j.cell.2017.07.028 (2017).

40 McQualter, J. L., Yuen, K., Williams, B. & Bertoncello, I. Evidence of an epithelial stem/progenitor cell hierarchy in the adult mouse lung. *Proc Natl Acad Sci U S A* **107**, 1414-1419, doi:10.1073/pnas.0909207107 (2010).

41 Aros, C. J., Pantoja, C. J. & Gomperts, B. N. Wnt signaling in lung development, regeneration, and disease progression. *Commun Biol* **4**, 601, doi:10.1038/s42003-021-02118-w (2021).

42 Chung, M. I., Bujnis, M., Barkauskas, C. E., Kobayashi, Y. & Hogan, B. L. M. Niche-mediated BMP/SMAD signaling regulates lung alveolar stem cell proliferation and differentiation. *Development* **145**, doi:10.1242/dev.163014 (2018).

43 Panos, R. J., Patel, R. & Bak, P. M. Intratracheal administration of hepatocyte growth factor/scatter factor stimulates rat alveolar type II cell proliferation in vivo. *Am J Respir Cell Mol Biol* **15**, 574-581, doi:10.1165/ajrcmb.15.5.8918364 (1996).

44 Volckaert, T. *et al.* Parabronchial smooth muscle constitutes an airway epithelial stem cell niche in the mouse lung after injury. *J. Clin. Invest.* **121**, 4409-4419, doi:10.1172/jci58097 (2011).

45 Xie, Y. *et al.* FGF/FGFR signaling in health and disease. *Signal Transduct Target Ther* **5**, 181, doi:10.1038/s41392-020-00222-7 (2020).

46 Yanagita, K. *et al.* Hepatocyte growth factor may act as a pulmotrophic factor on lung regeneration after acute lung injury. *J. Biol. Chem.* **268**, 21212-21217 (1993).

47 Choi, J. *et al.* Release of Notch activity coordinated by IL-1β signalling confers differentiation plasticity of airway progenitors via Fosl2 during alveolar regeneration. *Nat. Cell Biol.* **23**, 953-966, doi:10.1038/s41556-021-00742-6 (2021).

48 Weiner, A. I. *et al.* Mesenchyme-free expansion and transplantation of adult alveolar progenitor cells: steps toward cell-based regenerative therapies. *NPJ Regen Med* **4**, 17, doi:10.1038/s41536-019-0080-9 (2019).

49 Lv, Z. *et al.* Airway secretory cell-derived p63⁺ progenitors contribute to alveolar regeneration after sterile lung injury. *bioRxiv*, 2023.2002.2027.530122, doi:10.1101/2023.02.27.530122 (2023).

50 Weiner, A. I. *et al.* ΔNp63 drives dysplastic alveolar remodeling and restricts epithelial plasticity upon severe lung injury. *Cell Rep* **41**, 111805, doi:10.1016/j.celrep.2022.111805 (2022).

51 Seibold, M. A. *et al.* A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med* **364**, 1503-1512, doi:10.1056/NEJMoa1013660 (2011).

52 Saint, M. *et al.* Single-cell imaging and RNA sequencing reveal patterns of gene expression heterogeneity during fission yeast growth and adaptation. *Nat Microbiol* **4**, 480-491, doi:10.1038/s41564-018-0330-4 (2019).

53 Liu, Z. *et al.* Integrating single-cell RNA-seq and imaging with SCOPE-seq2. *Sci Rep* **10**, 19482, doi:10.1038/s41598-020-76599-w (2020).