A single cell atlas of *in vitro* multi-systems uncovers *in vivo* lineage trajectory and cell state in the human lung

Woochan Lee[1,15], Seyoon Lee[1,15], Jung-Ki Yoon[2,15], Dakyung Lee[1], Yuri Kim[3], Yeon Bi Han[4], Rokhyun Kim[1], Sungji Moon[5], Young Jun Park[6], Kyunghyuk Park[7], Bukyoung Cha[7], Jaeyong Choi[1], Juhyun Kim[1], Na-young Ha[3], Kwhanmien Kim[8], Sukki Cho[8], Nam-Hyuk Cho[3,9,10], Tushar J. Desai[2], Jin-Haeng Chung[4,*], Joo-Hyeon Lee[11,12,*], Jong-Il Kim[1,5,13,14,*]

[1]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea.
[2]Division of Pulmonary, Allergy and Critical Care, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA.
[3]Institute of Endemic Diseases, Medical Research Center, Seoul National University, Seoul, Korea.
[4]Department of Pathology and Translational Medicine, Seoul National University Bundang Hospital, Seongnam, Korea.
[5]Interdisciplinary Program in Cancer Biology, College of Medicine, Seoul National University, Seoul, Korea
[6]Department of Translational Medicine, Seoul National University College of Medicine, Seoul, Korea
[7]Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul, Korea.
[8]Department of Thoracic and Cardiovascular Surgery, Seoul National University Bundang Hospital, Seongnam, Korea.
[9]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea.
[10]Department of Microbiology and Immunology, Seoul National University College of Medicine, Seoul, Korea.
[11]Wellcome–MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge CB2 0AW, UK.
[12]Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EL, UK.
[13]Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, Korea.
[14]Cancer Research Institute, Seoul National University, Seoul, Korea.
[15]These authors contributed equally.

*Corresponding authors. Email: jongil@snu.ac.kr (J.-I.K.); jhl62@cam.ac.uk (J.-H.L.); chungjh@snu.ac.kr (J.-H.C.)

Abstract

We present an in-depth single cell atlas of *in vitro* multi-culture systems on human primary airway epithelium derived from normal and diseased lungs of 27 individual donors. Our large-scale single cell profiling identifies new cell states and differentiation trajectories of rare airway epithelial cell types in human distal lungs. By integrating single cell datasets of human lung tissues, we discover immune primed subsets enriched in lungs and organoids derived from patients with chronic respiratory disease. To demonstrate the full potential of our platform, we further illustrate transcriptomic responses to various respiratory virus infections in *in vitro* airway models. Our work constitutes a single-cell roadmap for the cellular and molecular characteristics of human primary lung cells *in vitro* and their relevance to human tissues *in vivo*.

INTRODUCTION

Single cell transcriptomics has increased knowledge of the cellular and molecular composition of human tissues, including the identification of previously unknown cell types and states, some of which arise specifically during human diseases [1-5]. However, due to the limited tools to grow and manipulate human cells *in vitro*, the functional roles for these populations have yet to be fully elucidated. Over the past decade, three-dimensional (3D) culture system known as organoids have been developed to grow 'mini-human organs' derived from stem cells that self-renew and differentiate into multiple cell lineages [6]. Given their ability to recapitulate key aspects of *in vivo* organs, organoid models opened up new avenues for studying human organ development. Significantly, the current coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) highlighted the importance of these biological platforms to better understand the pathogenesis of human diseases [7-13]. Before expanding their use in clinical applications and translational research, it is critical to determine whether the gene expression programs and lineage relationships in *in vitro* models faithfully replicate those of the respective tissues.

The respiratory epithelium is a primary site of environmental exposures, including viruses, which can lead to life-threatening respiratory failure. Human lung stem cell-derived *in vitro* culture systems have emerged as powerful tools for studying human lung development and modeling pulmonary diseases. The air-liquid interface (ALI) 2D culture system is well-established and involves culture of airway epithelial cells such that the basal surface of the cells is in contact with medium while the apical surface is exposed to air [14-16]. In this system, basal cells expand and undergo proper differentiation into a pseudostratified monolayer barrier, consisting of ciliated and secretory cells, providing a robust *in vitro* airway model. The airway 3D organoids derived from basal or club cells can be maintained in a long-term and generate differentiated cells, allowing for the study of cellular lineage hierarchy in the human airway epithelium [17-19]. Current COVID-19 research, in particular, has fueled the application of these platforms to study viral tropism and infection pathogenesis of SARS-CoV-2 in human airways [20-24]. However, the fidelity of these two model systems for prospective modeling of infection has not been thoroughly validated.

Here, we establish biobanks of 83 human primary airway organoid lines and present a single cell atlas of them derived from a subset of 27 individuals. We directly compare the transcriptomes of 2D and 3D human airway *in vitro* models against *in vivo* human lung tissues by incorporating public datasets from the Human Lung Cell Atlas [1]. Our deep single cell profiling enabled us to map the differentiation trajectories of rare airway cell types emerging from basal cells over pseudotime. We also discovered distinct immune primed subsets in *in vitro* models derived from lungs of patients with chronic respiratory diseases that are also enriched in *in vivo* lungs. We go on to demonstrate the utility of *in vitro* models and our organoid single cell atlas by comparing cellular states and immune responses following various virus infections at the single cell level. Our organoid single cell atlas datasets constitute a valuable resource for interactive exploration by the research community online (OSCA: http://osca.snu.ac.kr).

METHODS

Human tissues

For the establishment of human primary distal airway organoid cultures, human distal lung tissues were obtained from lung cancer patients undergoing lung resection surgery at Seoul National University Bundang Hospital (SNUH) with written informed consent from approval of the ethical committee (IRB No. 2008-065-1148). Organoids were derived from lung tissues far most from a tumor lesion without (normal lung tissues) or with a history of chronic respiratory diseases (diseased lung tissues) (table S1). Fresh lung tissues were stored in media (Advanced DMEM/F12 (Gibco), 1% HEPES (Gibco), 1% Glutamax (Gibco), and 0.125 ug/ml Amphotericin B (Gibco)) at 4°C less than 12 hours before processing.

Isolation and culture of human primary distal airway organoids

Lung tissues were cut into < 5 mm pieces and washed in ice cold PBS to remove residue mucus and blood cells. Before snap freezing for DNA extraction, 2-3 pieces of lung tissues were kept in RNAlater solution (Invitrogen) for overnight at 4°C. The remaining tissue pieces were dissociated into single cells using gentleMACS Octo Dissociator with Heater (Miltenyi Biotec) according to the manufacturer's protocol. The tissue suspensions were filtered through a 70 μm cell strainer (Miltenyi Biotec) and washed with 5 ml of Advanced DMEM/F12. Cells were centrifuged at 400g for 10 min at 4°C and supernatant was aspirated. Cell pellets were resuspended in 1 ml of red blood cell lysis buffer (Miltenyi Biotec) for 10 min at room temperature (RT). The reaction was quenched by adding 9 ml of Advanced DMEM/F12 with 1% HEPES, 1% Glutamax, and 1% Penincillin-Streptomycin (Gibco) (hereafter ADF+++), followed by centrifugation at 400 g for 10 min at 4°C. Cell pellets were then resuspended in growth factor reduced Matrigel® (GFR-Matrigel; Corning) and plated as 40 μl droplets in a pre-warmed 24-well tissue culture plate. Plates were incubated at 37°C for 15 mins, followed by submersion in 500 μl of pre-warmed airway organoid medium (AO medium) with 10 uM Y-27632 (Tocris) for first 3 days [17]. Medium was changed every 3 days. Established three-dimensional (3D) airway organoids (3D-AOs) were passaged at 1:4 ratios every 1-2 weeks. For passaging, TrypLE express (Gibco) was added to each well and GFR-Matrigel was mechanically disrupted. Organoids were incubated at 37°C for 5 min. The reaction was quenched by adding ice-cold ADF+++ and cells were centrifuged at 400 g for 5 min at 4°C. Cells were resuspended in GFR-Matrigel and plated as described above. For cytokine treatment, we added 1 ng/ml IL-13 (Biolegend), 10 ng/ml IL-1β (Biolegend) or 10 ng/ml IL-6 (R&D systems) in 3D-AOs for 3 weeks.

Air-liquid interface (ALI) culture of human primary distal airway organoids

Established 3D-AOs were dissociated into single cells as described above. Transparent 24-well transwell inserts (Corning) were precoated with 1% v/v GFR-Matrigel for at least 1 hour at 37°C. Dissociated 200K airway cells were then resuspended in 200 μL of AO medium or PneumaCult™-Ex Plus medium (STEMCELL Technologies) and seeded in each transwell insert for 2D ALI cultures (ALI-Om or ALI-Pm, respectively). Then, 500 μL of medium was added in the lower compartment of transwell plates. After 5-8 days when monolayers reached 100% confluency, growth medium was removed from inserts to induce differentiation for another 21 days. For ALI-Pm, PneumaCult™-ALI Medium (STEMMCELL Technologies) was used for inducing differentiation.

Single cell RNA sequencing of *in vitro* human primary airway cultures

Established 3D-AOs between passages 3 and 10 from individual donors were used for sequencing. 3D-AOs were dissociated into single cells and plated for establishing ALI-Om and ALI-Pm, which were used for sequencing after 3 week differentiation. Dissociated cells were

resuspended in 1 ml ADF+++ and filtered through a 30 μm SmartStrainer (Miltenyi Biotec). Each sample was then analyzed using ReadyCount Green/Red Viability Stain (Invitrogen) for counting the number of cells and measuring cell viability with automated cell counter Countess 3 FL (Thermo Fisher). Cell suspensions from *in vitro* cultures established from different donors were pooled to obtain cell suspension mixture with equal proportion of each sample. The pooled cell suspension was centrifuged at 400 g for 5 min at 4°C, and then resuspended in 0.04% bovine serum albumin (BSA) solution at appropriate volume for microfluidic chip loading. Typically, 6 to 8 samples were loaded per lane of a 10x microfluidic chip device, and demultiplexed based on single-nucleotide polymorphisms. For 10x multiome analysis, one sample was loaded on Chromium Next GEM Chip J. All libraries were sequenced in NovaSeq 6000 system (Illumina) in paired-end mode.

Immunofluorescent staining of 2D ALI and 3D cultures

For immunofluorescent (IF) staining of 3D-AOs, Matrigel was dissolved with ice-cold Cell Recovery Solution (Corning) at 60 rpm for 30 to 60 mins at 4°C on horizontal shaker (CRYSTE). The recovered organoids were fixed in 4% paraformaldehyde for 30 min at 4°C. For IF staining of ALI-Oms, the insert was displaced from the plate and placed on a clean petri dish upside down. Using a surgical blade, the membrane with cells were detached from the insert and fixed in 4% paraformaldehyde for 30 min at 4°C. The samples were then washed with 0.1% TritonX-100/PBS (PBS-T). They were next permeabilized in 0.2% Triton X-100/PBS for 5 min and incubated in 5% donkey serum (Jackson ImmunoResearch)/PBS-T blocking solution for 2 hours at RT. Primary antibodies were incubated for overnight at 4°C at the indicated dilutions: Chicken-anti-CK5 (1:500, Biolegend, 905904), Mouse-anti-TP63 (1:100, Thermo Fisher, MA1-21871), Rabbit-anti-MUC5AC (1:200, Cell Signaling Technology, 61193), Rat-anti-SCGB1A1 (1:200, R&D Systems, MAB4218), Rabbit-anti-Acetyl-α-Tubulin (1:200, Cell Signaling Technology, 5335), Mouse-anti-SARS-CoV NP (1:150, Sino Biological, 40143-MM05), Mouse-anti-Influenza A NP (1:100, Meridian, C87050M), and Rabbit-anti-SARS-CoV NP (1:1000, Sino biological, 40143-T62). Alexa Fluor-coupled secondary antibodies (1:500, Invitrogen) were incubated for overnight at 4°C. After antibody staining, nuclei were stained with DAPI (1:1000, Sigma) and sections were embedded in Vectashield Plus antifade mounting medium (H-1900, Vector laboratories). Fluorescence images were acquired using either Leica SP8 or LSM700 confocal microscopes. LAS X (Leica) or ZEN software (Zeiss) was used for processing fluorescent images.

Quantitative PCR

Total RNA was extracted from 3D-AO and ALI-Om using rNeasy mini kit (Qiagen). Complementary DNA (cDNA) was synthesized from extracted RNA using 1st strand cDNA synthesis kit (Takara) according to the manufacturer's protocols. Primers for *ACE2*, *TMPRSS2*, and *GAPDH* were listed in table S3. Quantitative PCR (q-PCR) was performed with Viia7 Real-time PCR system (Thermofisher) using TB green premix ex taq II (Takara) according to the manufacturer's instructions. Relative expression (2^{-ddCt}) was normalized with Ct of GAPDH and dCt of 3D-AOs.

Virus preparation

SARS-CoV-2 alpha or delta strain was obtained from Korean National Culture Collection for Pathogen (NCCP no.43326 and 43390, respectively). Viruses included in beta or omicron lineage were isolated from a nasopharyngeal swab taken from COVID-19 patients (GenBank accession no. OP349649.1 and 349650.1, respectively). Wild type MERS-CoV (GenBank accession no. KT029139.1) was provided by Korean Centers for Disease Control (KCDC). MERS-CoV and SARS-CoV-2 were propagated on VeroE6 cells (CRL-1586, ATCC) in

DMEM (Welgene) supplemented with 2% fetal bovine serum (FBS, Gibco) and 100 IU/ml penicillin-streptomycin at 37°C in a humidified CO_2 incubator for 3 days and titrated under overlay medium containing 0.8 % Methylcellulose (Sigma) and 2% FBS in DMEM. Influenza A (VR-95™, ATCC) was propagated in embryonated chicken eggs and titrated in MDCK cells (CCL-34™, ATCC) by plaque assay. The culture supernatant was cleared by centrifugation and stored in aliquots at −80°C until use.

Virus infection

All work with infectious viruses was performed in a Class II Biosafety Cabinet under BSL-3 (MERS-CoV, SARS-CoV-2) or BSL-2 (Influenza A) conditions at the Seoul National University. For all the virus (SARS-CoV-2, MERS-CoV, and Influenza A) infection of 2D differentiated cultures, samples were washed twice with 200 μL ADF+++ before infection from the apical side of the ALI culture at a MOI of 1.0. The number of cells in each culture was calculated by counting the cells from 1 well equally cultured well. The cultures were incubated at 37°C 5% CO_2 for 2 hours before washing three times in 200 μL ADF+++. The samples were then cultured under normal culture condition with AO medium in the lower compartment for 3 days before the generation of scRNA-seq library, IF staining, or transmission electron microscopy (TEM).

Transmission Electron Microscopy

To observe SARS-CoV-2 virus particles using the TEM, infected ALI-Om were fixed with a mixture of 2.5% glutaraldehyde in 0.1M phosphate buffer (pH 7.2) and 2% paraformaldehyde in 0.1M phosphate buffer (pH 7.2) for 3 hours [25-27]. The polyester membranes intact with fixed cell layers were carefully removed from the transwell inserts. The membrane disks intact with cell layers were fixed subsequently for 1.5 hours at RT with 2% osmium tetroxide in 0.1M phosphate buffer. The samples were dehydrated in graded ethanol and then infiltrated by propylene oxide and EPON epoxy resin. The samples were then embedded in EMbed-812 resin (EMS) and were polymerized at 80°C for overnight. The samples were cut on ultramicrotome (RMC MT-XL) at 65 nm. Ultrathin sections were stained with saturated 4% uranyl acetate and 4% lead citrate. TEM imaging and analysis were carried out using a transmission electron microscope (JEM-1400) at 80 kV in Department of Research & Experiment at Seoul National University Hospital.

K-chip genotype array

DNA was separated and extracted from frozen tissue or organoids. DNA genotyping was performed using Korea Biobank Array V1.1 (ThermoFisher) [28]. Korea Biobank Array Project analytic protocol was used to pre-process the data, and minor allele variations with low minor allele frequencies were included. Variants in linkage disequilibrium were excluded. Plink v1.0.9 was used to perform genotype principal component analysis and imputation was done using East Asian-specific WGS imputation reference panel (Northeast Asian Reference Database V2, unpublished) [29,30]. Beagle5 and Minimac4 was used to perform phasing and imputation respectively [31,32]. Variants were chosen using Bcftools and Plink 2.00a3 according to the following criteria: (i) Variants with imputation R2 higher than 0.8, (ii) Hardy Weinberg equilibrium p-value higher than 1e-6, (iii) Variant call rate higher than 0.9, and (iv) Minor allele frequency higher than 0.05 [33].

Computational methods

Constructing an organoid single cell atlas (OSCA) by analysis of scRNA-seq data

We used Cell Ranger (v6.0.1) [34] to align sequenced reads to GRCh38 human reference genome. To demultiplex individuals, we used souporcell (v2.0) [35] with donors' genotype array data from

K-chip. All samples undergo ambient RNA removal using cellbender (v0.2.1) [36] to remove background noise expression and doublets. Only cells assigned a single donor, passed doublet filter, express more than 200 genes, and not express more than 10% mitochondrial genes are used to further analysis. SCTransformation in Seurat package (v4.0.3) [37] regressing out mitochondrial gene percentage (percent.mt) was done for normalization. We performed canonical correlation analysis (CCA) with up to 50 PCs as integration method to correct batch effect between each sequencing libraries. RunPCA, RunUMAP, and FindNeighbors followed by FindClusters with Louvain algorithm in Seurat package were accomplished to visualize UMAP plot and cluster cells.

Cell cycle scoring

To compare proportions of each cell cycling phase between culture methods, we used CellCycleScoring in Seurat package [37]. ‘’MCM’’,‘’PCN’’,‘’TYM’’,‘’FEN’’,‘’MCM’’,‘’MCM’’,‘’RRM’’,‘’UN’’,‘’GINS’’,‘’MCM’’,‘’CDC A’’,‘’DT’’,‘’PRIM’’,‘’UHRF’’,‘’MLF1I’’,‘’HELL’’,‘’RFC’’,‘’RPA’’,‘’NAS’’,‘’RAD51AP’’,‘’GM N’’,‘’WDR7’’,‘’SLB’’,‘’CCNE’’,‘’UBR’’,‘’POLD’’,‘’MSH’’,‘’ATAD’’,‘’RAD5’’,‘’RRM’’,‘’CDC 4’’,‘’CDC’’,‘’EXO’’,‘’TIPI’’,‘’DSCC’’,‘’BL’’,‘’CASP8AP’’,‘’USP’’,‘’CLSP’’,‘’POLA’’,‘’CHAF1 ’’,‘’BRIP’’,‘’E2F’’ were used as marker genes of S phase and ‘’HMGB’’,‘’CDK’’,‘’NUSAP’’,‘’UBE2’’,‘’BIRC’’,‘’TPX’’,‘’TOP2’’,‘’NDC8’’,‘’CKS’’,‘’NUF’’,‘’ CKS1’’,‘’MKI6’’,‘’TMP’’,‘’CENP’’,‘’TACC’’,‘’FAM64’’,‘’SMC’’,‘’CCNB’’,‘’CKAP2’’,‘’CKAP ’’,‘’AURK’’,‘’BUB’’,‘’KIF1’’,‘’ANP32’’,‘’GTSE’’,‘’KIF20’’,‘’HJUR’’,‘’CDCA’’,‘’HN’’,‘’CDC2’ ’,‘’TT’’,‘’CDC25’’,‘’KIF2’’,‘’RANGAP’’,‘’NCAPD’’,‘’DLGAP’’,‘’CDCA’’,‘’CDCA’’,‘’ECT’’,‘’ KIF2’’,‘’HMM’’,‘’AURK’’,‘’PSRC’’,‘’ANL’’,‘’LB’’,‘’CKAP’’,‘’CENP’’,‘’CTC’’,‘’NEK’’,‘’G2E’’ ,‘’GAS2L’’,‘’CBX’’,‘’CENP’’ were used as marker genes of G2M phase. TUBB4B gene is removed from default Seurat G2M marker genes because TUBB4B is also a marker of ciliated cell. Ratio of cycling cells of each culture method of a cell type was calculated of average ratio of S or G2M phase cells of individuals.

Integration of OSCA and *in vivo* tissue scRNA-seq data

To compare transcriptomes between *in vitro* human lung cells with *in vivo* human lung tissues, we integrated our *in vitro* single cell datasets with *in vivo* tissue datasets of human lung cell atlases from European Genome-phenome Archive (EGA) under accession EGAS00001004344 [1]. We used 1~50 PCs and SCT to transfer anchors between *in vitro* cell and tissue data using FindTransferAnchors [37]. After MapQuery was done, using integrated PCs, we regenerated UMAP and Louvain clusters. Cell types were identified by annotation using scHCL (v0.1.1)[38] and Azimuth [37] and manual curation with known marker gene expressions.

Cell type annotation prediction using CellTypist

The Python package CellTypist (v.1.2.0) [39] was used to perform cell type annotation prediction with logistic regression models. All the models provided by CellTypist trained with lung single cell data were used on our data for cell type annotation prediction and validation. We also used lung single cell data and annotated metadata from P.K.L. Murphy et al. as training dataset to create CellTypist model. Default parameters were used for model building. All the predictions were made with majority voting.

Rare cell score

For cell type identification, scHCL was used to match each cell to annotated scRNA-seq data [38]. Within 23 clusters divided by Louvain clustering algorithm, cluster 22 exclusively expressed canonical marker genes of rare airway epithelial cell populations, such as tuft cells, ionocytes and pulmonary neuroendocrine cells (PNECs), and was primarily labeled as

'Basal.cell.Airway.Epithelium_Plasschaert' or 'Ionocyte.Airway.Epithelium_Plasschaert'. We defined scHCL_score of 'Basal.cell.Airway.Epithelium_Plasschaert' or 'Ionocyte.Airway.Epithelium_Plasschaert' as 'rare cell score'.

Rare cell population lineage trajectory

Single cell RNA-seq data across sample origins were re-integrated using Seurat's CCA algorithm [37]. Monocle (v3.1.0.0) was then used to determine the paths that cells can take as they develop along the UMAP via the learnGraph function [40]. Cells were then ordered along the identified trajectory of pseudotime using the orderCell function assuming that the cluster predominantly comprised of basal cells serve as the initial root group. Then we used the prepPseudotimePlotDatasets function adopted from Goldfarbmuren et al. to view the expression of key marker genes along the pseudotime on a shared scale [41]. This approach enabled to generate smoothed expression curves across pseudotime normalized between zero and one.

Single cell multiome data analysis

Single cell multiome data were aligned to GRCh38 reference genome with cellranger-arc (v2.0.0) [34,42]. Gene expression data from each single cell followed identical pipeline which was used in 3' scRNA-seq data with Seurat (v4.0.3) [37]. Signac (v1.6.0) package [43] was used to filtering, clustering, and visualizing Assay for Transposase-Accessible Chromatin (ATAC) data. Cells with more than 200 open peaks and peaks detected in more than 10 cells were used. Using NucelosomeSignal and TSSEnrichment, cells with less than 4 nucleosome_signal and more than 2 TSS.enrichment are normalized with RunTFIDF, FindTopFeatures, and RunSVD. To identify TAB cell specific overrepresented motifs, we used FindMotifs comparing TAB cell and basal-2 cell. Most significantly overrepresented motif profiles are represented using JASPAR [44] database. Motif activity scores are calculated with chromVAR with bSgenome.Hsapiens.UCSC.hg38 [45].

Identification of enriched genes, pathways, and immune activity

Normalized expression profiles for each culture method or subcluster were compared using Wilcoxon rank sum test with FindMarkers function [37]. Genes showing an adjusted p-value < 0.05 considered as differentially expressed genes (DEGs). Gene set enrichment analysis (GSEA) was completed using GSEAPreranked with DEGs ranked by $\log_2$FC. Hallmark and C5: Biological process were tested for enrichment. Gene sets with FDR q-value < 0.01 are considered as significant. Many scRNA-seq data with tissue of lung disease patients and organoids with cytokine treatment were reference mapped to OSCA integrated with tissue to predict cell types and subclusters of each cell. We used idiopathic pulmonary fibrosis (IPF), chronic obstructive pulmonary disease (COPD), cystic fibrosis (CF), healthy control, and fetal lung tissue data from GEO using GSE135893, GSE150674, and EMBL-EBI ArrayExpress database (E-MTAB-8221) respectively [4,46,47]. Before reference mapping, each query data is manipulated with SCTransform, RunPCA, and RunUMAP as same as reference OSCA data. After transfer anchors, ratio of predicted subclusters by MapQuery are calculated by (number of immune primed cells)/(number of all basal or ciliated cells) for each sample.

Single cell RNA-seq data analysis with virus infected samples

We used chimeric reference genome with GRCh38 and NC_045512.2 [48], NC_019843.3 [49], and NC_002016.1~NC_002023.1 [50] for SARS-CoV-2, MERS-CoV, and Influenza A respectively. All processes including align, normalization, and filtering are same with the pipeline to construct OSCA. Infected cells are defined as cells expressing viral genes. IFN and Entry score

were measured using AddModuleScore in Seurat package with genes related with interferon signaling and viral entry factors [51].

RESULTS

Biobanking of human primary distal airway organoids

We established biobanks of human primary airway organoids with 83 cryopreserved organoid lines, matched snap-frozen tissues, and genotyping data. Surgically resected human distal lung tissues were enzymatically dissociated into single cells used to generate 3D airway organoids (Fig. 1) [17]. Established organoids were maintained by passaging up to 3 times (83/118; 70% of success rate) and biobanked. DNA panel sequencing for genotyping was performed on matched frozen tissues.

To test the impact of culture conditions on cellular behaviors and gene expression programs *in vitro*, we directly compared the two most widely used *in vitro* human airway models, namely 2D ALI and 3D organoid cultures. Airway organoids were dissociated into single cells to establish a submerged 2D ALI culture (Fig. 1A) [52]. We used the same media for 3D airway organoid (referred to hereafter as 3D-AO) and 2D ALI culture (referred to hereafter as ALI-Om) to eliminate this as a variable (See Methods). We also grew the cells in a conventional ALI culture that results in mucociliary differentiation using commercially available PneumaCult medium (referred to hereafter as ALI-Pm) (Fig. 1A; See Methods) [53-55]. We used immunofluorescent (IF) staining for airway lineage markers to confirm the spatial distribution of differentiated cells within the pseudostratified epithelium of 3D-AO and ALI-Om, including basal, goblet, club, and ciliated cells (fig. S1A).

Single cell transcriptome atlas reveals preserved cellular diversity of human primary airway cells across *in vitro* multi-culture systems

We then generated a single cell transcriptional atlas of human primary distal airway organoids cultured in three different conditions described above. We used the 10X genomics platform to measure the single cell transcriptome of 3D-AO, ALI-Om, and ALI-Pm derived from 27 individuals: 19 normal and 8 diseased lungs (Fig. 2A and table S1, and S4). 95,170 cells passed quality control and after raw data quality control, each of the separately constructed datasets were integrated via Canonical Correlation Analysis (CCA) batch correction (See Methods). We annotated the major airway epithelial cell types using canonical cell type marker genes for identification. The results showed that all 27 organoid lines contained all of the major cell types, regardless of culture methods. The profiles of *in vitro* multi-culture systems faithfully matched their *in vivo* counterparts: basal cells (*TP63, KRT5, KRT15*), secretory club cells (*SCGB1A1, BPIFB1*), goblet cells (*MUC1, MUC5B, MSMB*), and ciliated cells (*FOXJ1, PIFO, TPPP3, SNTN*) (Fig. 2, B, C, G, H, fig. S1B, and fig. S2). We identified two subsets of basal cells (basal-1 and basal-2) with basal-2 expressing higher levels of *CXCL14, RNF43, LGR6,* and *LRP4* than basal-1 (Fig 2. B to E and fig. S1B). Notably, our large-scale of single cell profiling allowed us to detect SCGB3A2^{+} cells that were recently found in the human lung terminal bronchioles (Fig. 2F) [2,3,56]. Significantly, given the active growth of organoids, we were able to capture cellular transitional states that are readily apparent in the homeostatic lung. For example, a group of organoid cells co-expresses basal and secretory cell marker genes (*KRT5*, *SCGB1A1*), suggesting suprabasal cells in the process of differentiating from basal to luminal secretory cell states. We also found a distinct cluster of deuterosomal cells that expresses both secretory and ciliated cell marker genes (*SCGB1A1*, *FOXJ1*), indicating transient state cells from secretory to ciliated cells. As expected, a subset of basal cells was actively proliferating (which we named proliferating basal), whereas differentiated cells demonstrated low cycling activity (Fig. 2, C and I, and fig. S1B). Interestingly, deuterosomal cells, which are precursors of ciliated cells, displayed remarkable cycling activity both in organoids and lung tissues (Fig. 2I and fig. S5C). Significantly, despite the differences in their proportions, all 27 *in vitro* lines maintained relatively comparable cellular diversity and state independent of culture conditions

or passage numbers (Fig. 2, G to I). There were also no discernable differences correlating with age, sex, smoking status, and disease background (fig. S1, C to G). These datasets are available for interactive analysis at the organoid single cell atlas (OSCA: http://osca.snu.ac.kr).

Large-scale single cell multiomics of *in vitro* multi-culture models to track rare airway epithelial lineage differentiation

Recent single cell transcriptome studies of human primary lung tissues allowed for the mapping of cellular composition and lineage hierarchy at the single cell level [1-3,57]. However, epithelial cell types such as tuft cells, ionocytes, and pulmonary neuroendocrine cells (PNECs) are rare and therefore not amenable for either deep molecular characterization or inferring lineage trajectories from *in vivo* lung atlases [58]. Notably, our large-scale single cell transcriptome profiling enabled us to detect all of these rare airway epithelial cell types both in 2D ALI and 3D organoid models: tuft cells (*POU2F3*), ionocytes (*FOXI1*), and PNECs (*ASCL1*) (Fig. 2, B, C, G, H, and fig. S1B). To build a pipeline for tracking lineage trajectories of these rare epithelial cell types, we first tested the *in vitro* lineage relationship between basal cells and secretory cells, which is well-established *in vivo* lineage hierarchy of basal to secretory cells in the lung [18]. Clusters of proliferating basal, basal-1, suprabasal and secretory cells were extracted and reintegrated from our single cell datasets for further analysis. As expected, pseudotime trajectory analysis revealed the differentiation path of basal cells to secretory cells, ensuring the fidelity of *in vitro* differentiation programs against *in vivo* lungs (fig. S3, A to D). Reclustering secretory, deuterosomal, and ciliated cells also showed the differentiation trajectories of secretory cells to ciliated cells via deuterosomal cells, as previously reported (fig. S3, E to H).

We then investigated lineage differentiation of rare cell types by establishing an enrichment score based on transcriptome similarity between each cell and known reference rare cells (referred to hereafter as rare cell score; See Methods) [38]. Among 23 unbiased Louvain clusters, cluster 22, which retains rare epithelial cell types, had the highest rare cell score in *in vitro* airway cells (Fig. 3, A to C). Interestingly, among non-rare lung epithelial cells, basal-2 cells (cluster 13) had the highest rare cell score, leading us to hypothesize that basal-2 cells could be the precursor for rare cell differentiation (Fig. 3, A to C). We thus extracted and reclustered cells in cluster 13 and 22 (Fig. 3, D and E). Notably, basal cells with high rare cell scores were closely aligned with clusters of distinct rare cell populations. Furthermore, based on the expression level of basal cell marker genes (*KRT5*, *KRT15*), rare cell scores, and pseudotime trajectory analysis, we were able to dissect transitioning airway basal cells (referred to hereafter as TAB) which showed lower expression of basal cell marker genes and closely associated with basal-2 and rare cell types (Fig. 3, E to H, and fig. S4). TAB cells were specifically marked by the expression of *NOTUM*, *FGF3*, and *FGF19*, suggesting a distinct molecular program modulating TAB cells in this transition (Fig. 3, G and H, and fig. S4). To better understand the relationship between TAB cells and rare airway lineages, we applied the 10x single cell multiomics on 3D-AO. Significantly, analysis of single-cell Assay for Transposase Accessible Chromatin using sequencing (scATAC-seq) demonstrated that TAB annotated by gene expression profiles were also closely assigned to rare cell populations (Fig. 3, I and J). We further determined transcription factor (TF) binding sites that were specifically opened in TAB cells (Fig. 3K). Interestingly, TF binding sites for *ASCL1*, a marker gene for PNECs, were opened in TAB cells whereas its expression was restricted to PNECs, suggesting that TAB cells could represent primed cell states differentiating into PNECs from basal cells (Fig. 3, L and M). Additionally, tuft cells were further classified as tuft-1 (*ASCL2*) and tuft-2 (*TRPM5*) cells based on unique signatures (Fig. 3, E and G, and fig. S4) [59-61].

A distinct transcriptomic signature of immune response in *in vivo* human lung tissues versus *in vitro* organoid models

To determine how closely the cellular and molecular features of *in vitro* human airway cultures replicate those of *in vivo* human lung tissue, we integrated well-annotated data of epithelial cell clusters from the Human Lung Cell Atlas with our dataset (Fig. 4A and fig. S5, A and B; See Methods) [1]. As previously stated, the cellular composition of *in vitro* models is comparable to that of lung tissues whereas the proportions of each cell type differ (Fig. 4B, and fig. S5C). *In vitro* airway cells contained more basal and intermediate cells, whereas epithelial cells from lung tissues had more differentiated cells. We then compared the molecular programs of *in vitro* airway cells with those of lung tissues. Significantly, the Gene Set Enrichment Analysis (GSEA) of Differentially Expressed Genes (DEGs) in cells derived from lung tissues and cultures revealed enriched gene expression patterns involved in immune responses in *in vivo* lungs (fig. S5D). We attribute this result to the unique feature of *in vitro* airway models that, despite faithfully mimicking *in vivo* tissue characteristic, are entirely composed of epithelial cells and lack immune cells. Interestingly, only the subsets of basal and ciliated cells showed enriched expressions of immune response genes (*CXCL1/2/8*, *CCL2/20*, *SAA1/2*, *IFITM1/2/3*) (Fig. 4, C and D, and fig. S5, E and F). The GSEA using hallmark gene sets of DEGs also showed that these subsets were characterized by increased expressions of genes associated with immune and inflammatory responses, implying 'immune primed cell states' (Fig. 4, C and E). Interestingly, although basal and ciliated cells are molecularly and functionally distinct cell types, transcriptional differences in their respective immune primed cells revealed strong similarities (Fig. 4E).

We next investigated the clinical relevance of immune primed cells in chronic lung diseases. Using reference-mapping, we combined our single-cell RNA sequencing (scRNA-seq) data with those from previous studies on patients with idiopathic pulmonary fibrosis (IPF), chronic obstructive pulmonary disease (COPD), and cystic fibrosis (CF) (See Methods) [47,62,63]. We found higher proportions of immune primed subsets in basal cells of diseased lungs compared to their normal controls, whereas fetal lung tissues barely exhibited immune primed cell states (Fig. 4F) [4]. We then further analyzed immune primed cells in our airway organoids derived from normal and chronic respiratory diseased lung tissues (table S1; See Methods). Organoids derived from diseased lung tissues revealed a significant increase in immune primed basal cells compared to organoids derived from normal lung tissues (Fig. 4G). Further, organoids derived from lung tissues of aged (>50) or smoking donors also showed an increase in immune primed basal cells compared to organoids derived from lung tissues of young or non-smoking donors, respectively (Fig. 4G). However, we observed no discernable increase in immune primed ciliated cells despite their absolute proportion is greater than basal cells regardless of donors characteristics. We next investigated whether inflammatory cytokines could be used to mimic the *in vivo* immune microenvironment in organoid cultures. Treatment of pro-inflammatory cytokines, such as IL-1β and IL-6, induced immune primed basal and ciliated cells in 3D-AOs derived from healthy lungs, demonstrating that human primary airway organoids replicate *in vivo* immune responses by modulating microenvironmental culture conditions (Fig. 4H). Notably, ciliated cells were more responsive to inflammatory stimuli than basal cells. We also noticed a trend of gradually decreasing proportion of immune primed basal cells in 3D-AO over multiple passages, implying that cell states influenced by local microenvironmental signals may be lost during extended *in vitro* cultures (Fig. 4I).

Transcriptomic responses of human airway organoids to respiratory virus infections

Finally, we utilized our platform to model various respiratory virus infections and compare transcriptomic signatures of infected human primary airway epithelium (Fig. 1A). Established

ALI-Om derived from 20 individual tissue donors were used for infection with SARS-CoV-2 variants (B.1.1.7 Alpha, B.1.351 Beta, B.1.617.2 Delta, and B.1.1.529 Omicron), MERS-CoV, and Influenza A viruses (Fig. 5, A and B, and table S2, and S5). We first assessed the expression of *ACE2* and *TMPRSS2* genes, which are necessary for SARS-CoV-2 viral entry, in ALI-Om during the air-liquid differentiation by quantitative PCR (q-PCR) analysis. The expression levels of *ACE2* and *TMPRSS2* were upregulated at 3 weeks after the induction of differentiation, when virus was delivered (fig. S6, A and B). Analysis of transmission electron microscopy (TEM) showed the widespread of SARS-CoV-2 alpha viral particles in ALI-Om at 72 hours post infection (fig. S6C). IF staining for nucleocapsid protein (NP) also confirmed the efficient infection of SARS-CoV-2 alpha and Influenza A viruses in these cells (fig. S6D).

We conducted scRNA-seq analysis of ALI-Om that were infected with SARS-CoV-2 variants, MERS-CoV, and Influenza A viruses. Utilizing aligned reads on viral reference genome, we measured the infection rate. ALI-Om derived from diseased lung tissues showed a significant increase in infection rate compared to those from normal tissues (Fig. 5C). However, age and sex of lung tissue donors had little effect on the infection rate (Fig. 5D). We then used our single-cell transcriptome platform as a reference dataset for untreated controls for the single-cell data of infected samples (Fig. 5E). The proportion of cell types was barely influenced by virus infection with the exception of omicron, which reduced basal-1 cell populations (Fig. 5F). Importantly, ALI-Om infected with SARS-CoV-2 (Beta), Omicron and Influenza A, revealed significant increases in immune primed basal subsets (Fig. 5G). However, we found no significant effect of age, sex, and disease history of donor tissues on the immune primed cell states induced by virus infection in ALI-Om (fig. S6, E to G). As previously reported, we also demonstrated enriched interferon and viral entry scores in our infected cells, particularly with SARS-CoV-2 Omicron, MERS-CoV, and Influenza A viruses (Fig. 5H) [51].

DISCUSSION

Given the potential of human primary organoids as a platform to study human biology, it is critical to determine whether and to what degree organoids mirror the cellular and molecular events of complex biological processes in human tissues *in vivo*. We generated a massive single cell atlas of three representative *in vitro* human airway models and provided a pipeline for comparing *in vitro* datasets to those of *in vivo* human lungs.

We demonstrate that, despite differences in the proportion of cellular composition, the cellular diversity and transcriptional cell states are significantly maintained in all *in vitro* models with 27 lines and are comparable to those of *in vivo* human lungs. These data strongly support that none of culture methods are inferior and that there may be potential to perform cross-comparisons. Importantly, these cellular diversities are maintained throughout multiple passages up to 10, suggesting the stability of *in vitro* human primary airway models. Our large-scale single cell transcriptome profiling allowed us to detect a distinct subset of basal cells expressing enriched Wnt signaling pathway genes and closely connecting to rare cell differentiation, suggesting their unique stem cell characteristics, which require further validation. Notably, we also captured a small subset of secretory cells expressing *SCGB3A2*, which was recently identified as potential progenitor cells differentiating from/to alveolar lineages in human distal terminal bronchioles [2,3]. Further, intriguingly, despite decreasing cell cycle activity with differentiation trajectory, deuterosomal cells, precursors of ciliated cells, display high levels of cell cycling on an equal level with proliferating basal cells even in *in vivo* lungs, suggesting their active contribution to the maintenance of ciliated cells in homeostatic lungs *in vivo*.

Recent studies highlighted the emergence of intermediate cell states during lung regeneration and their clinical relevance in human lung diseases, including pulmonary fibrosis and lung cancer [64-66]. Given the feature of *in vitro* models which recapitulate active regenerative contexts rather than homeostatic conditions, our single cell datasets readily detected cell states that are closely associated to two different lineages and have the molecular characteristics of transitioning cell states. Significantly, our massive single cell profiling of *in vitro* multi-culture models allowed us to capture differentiation dynamics of rare epithelial cell types, such as PNECs, ionocytes, and tuft cells, from basal cells, as previously reported in scRNA-seq of primary airway tissues [58,67]. We also detected subsets of tuft cells (tuft-1 and tuft-2) in our *in vitro* models, indicating the power of our platform to replicate lineage relationships of human lung tissues *in vivo*. A distinct basal subset of TAB cells was identified with the molecular characteristics of intermediate cell states between basal and rare cell types. Interestingly, Wnt signaling pathway differentiates subsets of basal cells (Fig. 2D and E). The temporal increase of NOTUM expression, a negative regulator of Wnt signaling, in TAB suggests the dynamic regulation of Wnt activity during the differentiation of basal-2 to rare cell lineages via TAB (Fig. 3G and H). A recent study demonstrated direct differentiation of human airway basal cells to PNECs in a hypoxia context [68]. Notably, we found that the TF binding site for the PNEC marker gene ASCL1 is opened in TAB cells whereas its expression is restricted to PNECs. This finding suggests the differentiation path of basal cells to PNECs via TAB states. It will be interesting to investigate further if TAB cells are specific cell states that transition to PNECs or whether they can also generate other rare cell types.

Because the lung epithelium is constantly exposed to environmental challenges *in vivo*, immune cells play an important role. Current *in vitro* models lack the ability to fully convey such dynamics, as demonstrated by our datasets. We identified unique immune primed subsets of basal and ciliated cells enriched in human lung tissues relative to *in vitro* airway cultures. Notably, ciliated cells in *in vivo* lungs are largely composed of immune primed cell population, as previously suggested that ciliated cells are readily responsive to inflammatory signals [19]. Interestingly, we also observed greater immune primed cell states in 3D-AO compared to ALI-Om and ALI-Pm, suggesting the potential contribution of extracellular matrix dimensionality to maintaining immune responsive cells *in vitro*. Further analysis of scRNA-seq datasets from the lungs of chronic respiratory disease patients revealed that they contained more immune primed subsets than in healthy lungs. Notably, we discovered that smoking and lung disease histories, as well as the age of tissue donors, affected the status of immune primed cells in established organoids, implying that cellular features of primary lung tissues are reflected in *in vitro* models. We did, however, observe a reduction in these features in organoids over multiple passages, which can be induced by the addition of pro-inflammatory cytokines. These findings demonstrate that cellular states can be shaped by microenvironmental factors, which impact on their functional behaviors in *in vitro* models. Thus, organoids may prove useful for understanding epithelial responses to resolution of active pulmonary inflammation and/or facilitate screening of compounds to augment immune responsiveness. Our findings provide important aspects of *in vitro* models as a platform for studying cellular characteristics and modeling lung diseases.

Given the urgent need to elucidate the pathogenesis of SARS-CoV-2 virus infection in the current pandemic, *in vitro* organoids have been extensively used to study virus tropisms and infection responses. We utilized our single cell platform to compare transcriptomic responses of human primary airway epithelial cells to various respiratory virus infections *in vitro*. The age, sex, and history of lung disease of tissue donors had little effect on the infection rate of

different viruses in cultures. Furthermore, regardless of the donor characteristics, we discovered a slight increase in immune primed subsets in infected cells. It would be intriguing to check whether additional inflammatory stimulus, such as adding immune cells or cytokines, triggers the expansion or activation of these primed cells in infected cells. It could be also useful to check the differential responses of infected ALI-Om derived from aged or/and diseased tissues to these stimuli. As previously reported, we also observed higher interferon responses in infected cells [69]. Most notably, respiratory viruses known to cause milder symptoms exhibited more pronounced changes in the transcriptomic responses, including interferon related signals. A recent study reported a higher innate interferon response in COVID-19 pediatric patients, which may correlate with reduced viral replication and milder disease progression in this population [51]. Our findings also demonstrated higher interferon signaling in ALI-Om infected with SARS-CoV-2 omicron variant and influenza A viruses, correlating with the milder symptoms with these pathogens.

In summary, by combining single cell multiomics and *in vitro* human primary organoids derived from multiple individuals, we produce a powerful database that can be used as a platform to better understand human lung stem cell activities and lineage trajectories. Our approach can also serve as a pre-clinical model that guides optimized individualized treatments based on molecular profiling to avoid delays in the use of the most effective therapeutic regimens for existing and novel viruses.

Acknowledgments
The authors thank all members of the lab for discussion.

Data and materials availability: All data needed to evaluate the conclusions are present in the paper and/or the Supplementary Materials. All raw and processed sequencing data will be publicly available at Gene Expression Omnibus (GEO). Data can be visualized at http://www.osca.snu.ac.kr.

Competing interests: The authors declare that they have no competing interests.

Funding: This work was supported by:
National Research Foundation of Korea grant, Korea government 2020R1A2C3012524 (J.-I.K)
Basic Science Research Program through the National Research Foundation of Korea, Ministry of Education 2020R1A6A1A03047972 (J.-I.K)
Korea Health Technology R&D Project through the Korea Health Industry Development Institute, Ministry of Health & Welfare, Republic of Korea HI21C1113 (J.-K.Y)
Wellcome Trust Senior Research Fellowship 221857/Z/20/Z (J.-H.L)
Suh Kyungbae Foundation SUHF-20010033 (J.-H.L)
National Research Foundation of Korea grant, Korea government 2021M3A9H5020761 (N.-H.C)

Author contributions:
Conceptualization: WL, SL, JKY, JHL, JIK
Methodology: JHC, YBH, SC, KK, KP, BC
Investigation: WL, SL, JKY, DL, RK, SM, YJP, YK, DK, JC, JK
Supervision: JHL, JIK
Writing: WL, SL, JHL

REFERENCES

1 Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619-625 (2020). https://doi.org:10.1038/s41586-020-2922-4

2 Kadur Lakshminarasimha Murthy, P. *et al.* Human distal lung maps and lineage hierarchies reveal a bipotent progenitor. *Nature* **604**, 111-119 (2022). https://doi.org:10.1038/s41586-022-04541-3

3 Basil, M. C. *et al.* Human distal airways contain a multipotent secretory cell that can regenerate alveoli. *Nature* **604**, 120-126 (2022). https://doi.org:10.1038/s41586-022-04552-0

4 Miller, A. J. *et al.* In Vitro and In Vivo Development of the Human Airway at Single-Cell Resolution. *Dev Cell* **53**, 117-128 e116 (2020). https://doi.org:10.1016/j.devcel.2020.01.033

5 Vieira Braga, F. A. *et al.* A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat Med* **25**, 1153-1163 (2019). https://doi.org:10.1038/s41591-019-0468-5

6 Kim, J., Koo, B. K. & Knoblich, J. A. Human organoids: model systems for human biology and medicine. *Nat Rev Mol Cell Biol* **21**, 571-584 (2020). https://doi.org:10.1038/s41580-020-0259-3

7 Mallapaty, S. *et al.* How COVID vaccines shaped 2021 in eight powerful charts. *Nature* **600**, 580-583 (2021). https://doi.org:10.1038/d41586-021-03686-x

8 Youk, J. *et al.* Three-Dimensional Human Alveolar Stem Cell Culture Models Reveal Infection Response to SARS-CoV-2. *Cell Stem Cell* **27**, 905-919 e910 (2020). https://doi.org:10.1016/j.stem.2020.10.004

9 Katsura, H. *et al.* Human Lung Stem Cell-Based Alveolospheres Provide Insights into SARS-CoV-2-Mediated Interferon Responses and Pneumocyte Dysfunction. *Cell Stem Cell* **27**, 890-904 e898 (2020). https://doi.org:10.1016/j.stem.2020.10.005

10 Salahudeen, A. A. *et al.* Progenitor identification and SARS-CoV-2 infection in human distal lung organoids. *Nature* **588**, 670-675 (2020). https://doi.org:10.1038/s41586-020-3014-1

11 Huang, J. *et al.* SARS-CoV-2 Infection of Pluripotent Stem Cell-Derived Human Lung Alveolar Type 2 Cells Elicits a Rapid Epithelial-Intrinsic Inflammatory Response. *Cell Stem Cell* **27**, 962-973 e967 (2020). https://doi.org:10.1016/j.stem.2020.09.013

12 Lamers, M. M. *et al.* SARS-CoV-2 productively infects human gut enterocytes. *Science* **369**, 50-54 (2020). https://doi.org:10.1126/science.abc1669

13 Pellegrini, L. *et al.* SARS-CoV-2 Infects the Brain Choroid Plexus and Disrupts the Blood-CSF Barrier in Human Brain Organoids. *Cell Stem Cell* **27**, 951-961 e955 (2020). https://doi.org:10.1016/j.stem.2020.10.001

14 Lechner, J. F., Haugen, A., McClendon, I. A. & Pettis, E. W. Clonal growth of normal adult human bronchial epithelial cells in a serum-free medium. *In Vitro* **18**, 633-642 (1982). https://doi.org:10.1007/BF02796396

15 Gray, T. E., Guzman, K., Davis, C. W., Abdullah, L. H. & Nettesheim, P. Mucociliary differentiation of serially passaged normal human tracheobronchial epithelial cells. *Am J Respir Cell Mol Biol* **14**, 104-112 (1996). https://doi.org:10.1165/ajrcmb.14.1.8534481

16 Fulcher, M. L., Gabriel, S., Burns, K. A., Yankaskas, J. R. & Randell, S. H. Well-differentiated human airway epithelial cell cultures. *Methods Mol Med* **107**, 183-206 (2005). https://doi.org:10.1385/1-59259-861-7:183

17 Sachs, N. *et al.* Long-term expanding human airway organoids for disease modeling. *EMBO J* **38** (2019). https://doi.org:10.15252/embj.2018100300

18 Rock, J. R. *et al.* Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc Natl Acad Sci U S A* **106**, 12771-12775 (2009). https://doi.org:10.1073/pnas.0906850106
19 Choi, J. *et al.* Release of Notch activity coordinated by IL-1beta signalling confers differentiation plasticity of airway progenitors via Fosl2 during alveolar regeneration. *Nat Cell Biol* **23**, 953-966 (2021). https://doi.org:10.1038/s41556-021-00742-6
20 Mlcochova, P. *et al.* SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **599**, 114-119 (2021). https://doi.org:10.1038/s41586-021-03944-y
21 Meng, B. *et al.* Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature* **603**, 706-714 (2022). https://doi.org:10.1038/s41586-022-04474-x
22 Wang, R. *et al.* Human airway lineages derived from pluripotent stem cells reveal the epithelial responses to SARS-CoV-2 infection. *Am J Physiol Lung Cell Mol Physiol* **322**, L462-L478 (2022). https://doi.org:10.1152/ajplung.00397.2021
23 Purkayastha, A. *et al.* Direct Exposure to SARS-CoV-2 and Cigarette Smoke Increases Infection Severity and Alters the Stem Cell-Derived Airway Repair Response. *Cell Stem Cell* **27**, 869-875 e864 (2020). https://doi.org:10.1016/j.stem.2020.11.010
24 Mulay, A. *et al.* SARS-CoV-2 infection of primary human lung epithelium for COVID-19 modeling and drug discovery. *Cell Rep* **35**, 109055 (2021). https://doi.org:10.1016/j.celrep.2021.109055
25 Graham, L. & Orenstein, J. M. Processing tissue and cells for transmission electron microscopy in diagnostic pathology and research. *Nat Protoc* **2**, 2439-2450 (2007). https://doi.org:10.1038/nprot.2007.304
26 Kumar, R., Kolloli, A. & Subbian, S. Inactivation and Elimination of SARS-CoV-2 in Biosamples Using Simple Fixatives and Ultrafiltration. *Methods Protoc* **4** (2021). https://doi.org:10.3390/mps4010018
27 Widera, M. *et al.* Evaluation of stability and inactivation methods of SARS-CoV-2 in context of laboratory settings. *Med Microbiol Immunol* **210**, 235-244 (2021). https://doi.org:10.1007/s00430-021-00716-3
28 Moon, S. *et al.* The Korea Biobank Array: Design and Identification of Coding Variants Associated with Blood Biochemical Traits. *Sci Rep* **9**, 1382 (2019). https://doi.org:10.1038/s41598-018-37832-9
29 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007). https://doi.org:10.1086/519795
30 Yoo, S. K. *et al.* NARD: whole-genome reference panel of 1779 Northeast Asians improves imputation accuracy of rare and low-frequency variants. *Genome Med* **11**, 64 (2019). https://doi.org:10.1186/s13073-019-0677-z
31 Browning, B. L. & Browning, S. R. Genotype Imputation with Millions of Reference Samples. *Am J Hum Genet* **98**, 116-126 (2016). https://doi.org:10.1016/j.ajhg.2015.11.020
32 Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782-784 (2015). https://doi.org:10.1093/bioinformatics/btu704
33 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021). https://doi.org:10.1093/gigascience/giab008
34 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049 (2017). https://doi.org:10.1038/ncomms14049
35 Heaton, H. *et al.* Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat Methods* **17**, 615-620 (2020). https://doi.org:10.1038/s41592-020-0820-1

36 Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from droplet-based single-cell experiments using <kbd>CellBender</kbd>. *bioRxiv*, 791699 (2022). https://doi.org:10.1101/791699

37 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021). https://doi.org:10.1016/j.cell.2021.04.048

38 Han, X. *et al.* Construction of a human cell landscape at single-cell level. *Nature* **581**, 303-309 (2020). https://doi.org:10.1038/s41586-020-2157-4

39 Dominguez Conde, C. *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022). https://doi.org:10.1126/science.abl5197

40 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019). https://doi.org:10.1038/s41586-019-0969-x

41 Goldfarbmuren, K. C. *et al.* Dissecting the cellular specificity of smoking effects and reconstructing lineages in the human airway epithelium. *Nat Commun* **11**, 2485 (2020). https://doi.org:10.1038/s41467-020-16239-z

42 Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* **37**, 925-936 (2019). https://doi.org:10.1038/s41587-019-0206-z

43 Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-1341 (2021). https://doi.org:10.1038/s41592-021-01282-5

44 Stormo, G. D. Modeling the specificity of protein-DNA interactions. *Quant Biol* **1**, 115-130 (2013). https://doi.org:10.1007/s40484-013-0012-4

45 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978 (2017). https://doi.org:10.1038/nmeth.4401

46 Habermann, A. C. *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* **6**, eaba1972 (2020). https://doi.org:10.1126/sciadv.aba1972

47 Carraro, G. *et al.* Transcriptional analysis of cystic fibrosis airways at single-cell resolution reveals altered epithelial cell states and composition. *Nat Med* **27**, 806-814 (2021). https://doi.org:10.1038/s41591-021-01332-7

48 Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269 (2020). https://doi.org:10.1038/s41586-020-2008-3

49 Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* **367**, 1814-1820 (2012). https://doi.org:10.1056/NEJMoa1211721

50 Winter, G. & Fields, S. Cloning of influenza cDNA ino M13: the sequence of the RNA segment encoding the A/PR/8/34 matrix protein. *Nucleic Acids Res* **8**, 1965-1974 (1980). https://doi.org:10.1093/nar/8.9.1965

51 Yoshida, M. *et al.* Local and systemic responses to SARS-CoV-2 infection in children and adults. *Nature* **602**, 321-327 (2022). https://doi.org:10.1038/s41586-021-04345-x

52 Prytherch, Z. *et al.* Tissue-Specific stem cell differentiation in an in vitro airway model. *Macromol Biosci* **11**, 1467-1477 (2011). https://doi.org:10.1002/mabi.201100181

53 Tata, P. R. *et al.* Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature* **503**, 218-223 (2013). https://doi.org:10.1038/nature12777

54 Pardo-Saganta, A. *et al.* Parent stem cells can serve as niches for their daughter cells. *Nature* **523**, 597-601 (2015). https://doi.org:10.1038/nature14553

55 McCauley, K. B. *et al.* Efficient Derivation of Functional Human Airway Epithelium from Pluripotent Stem Cells via Temporal Regulation of Wnt Signaling. *Cell Stem Cell* **20**, 844-857 e846 (2017). https://doi.org:10.1016/j.stem.2017.03.001

56 He, P. *et al.* A human fetal lung cell atlas uncovers proximal-distal gradients of differentiation and key regulators of epithelial fates. *Cell* **185**, 4841-4860 e4825 (2022). https://doi.org:10.1016/j.cell.2022.11.005

57 Deprez, M. *et al.* A Single-Cell Atlas of the Human Healthy Airways. *Am J Respir Crit Care Med* **202**, 1636-1645 (2020). https://doi.org:10.1164/rccm.201911-2199OC

58 Plasschaert, L. W. *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377-381 (2018). https://doi.org:10.1038/s41586-018-0394-6

59 Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333-339 (2017). https://doi.org:10.1038/nature24489

60 Billipp, T. E., Nadjsombati, M. S. & von Moltke, J. Tuning tuft cells: new ligands and effector functions reveal tissue-specific function. *Curr Opin Immunol* **68**, 98-106 (2021). https://doi.org:10.1016/j.coi.2020.09.006

61 O'Leary, C. E., Ma, Z., Culpepper, T., Novak, S. W. & DelGiorno, K. E. New insights into tuft cell formation: Implications for structure-function relationships. *Curr Opin Cell Biol* **76**, 102082 (2022). https://doi.org:10.1016/j.ceb.2022.102082

62 Adams, T. S. *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* **6**, eaba1983 (2020). https://doi.org:10.1126/sciadv.aba1983

63 Sauler, M. *et al.* Characterization of the COPD alveolar niche using single-cell RNA sequencing. *Nat Commun* **13**, 494 (2022). https://doi.org:10.1038/s41467-022-28062-9

64 Choi, J. *et al.* Inflammatory Signals Induce AT2 Cell-Derived Damage-Associated Transient Progenitors that Mediate Alveolar Regeneration. *Cell Stem Cell* **27**, 366-382 e367 (2020). https://doi.org:10.1016/j.stem.2020.06.020

65 Kobayashi, Y. *et al.* Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat Cell Biol* **22**, 934-946 (2020). https://doi.org:10.1038/s41556-020-0542-8

66 Strunz, M. *et al.* Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat Commun* **11**, 3559 (2020). https://doi.org:10.1038/s41467-020-17358-3

67 Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319-324 (2018). https://doi.org:10.1038/s41586-018-0393-7

68 Shivaraju, M. *et al.* Airway stem cells sense hypoxia and differentiate into protective solitary neuroendocrine cells. *Science* **371**, 52-57 (2021). https://doi.org:10.1126/science.aba0629

69 Han, Y. *et al.* Identification of SARS-CoV-2 inhibitors using lung and colonic organoids. *Nature* **589**, 270-275 (2021). https://doi.org:10.1038/s41586-020-2901-9

Figure Legends

Fig. 1. Establishing organoid biobanks and a single cell atlas of human primary distal airway organoids. (**A**) Schematic workflow of generating organoid biobanks, single cell transcriptome analysis, and the use of organoids in respiratory virus infection and cytokine treatment. 3D-AO (3D airway organoid), ALI-Om (Air-liquid interface culture with organoid media), ALI-Pm (Air-liquid interface culture with Pneumacult media), OSCA (Organoid single cell atlas). n, number of sample profiling.

Fig. 2. Single cell transcriptomic profiling of *in vitro* human airway multi-culture systems. (**A**) The number of profiled samples and cells for each culture method. (**B**) UMAP representing the scRNA-seq data of *in vitro* human primary airway cells indicated in (A). Ten different airway epithelial cell types are visualized with colors. TAB, transitioning airway basal. (**C**) Heatmap showing differential marker gene expressions for epithelial cell types. Each cell type column: 3D-AO (left), ALI-Om (middle), ALI-Pm (right). (**D**) Violin plot showing the expression of the indicated genes in basal-1 and basal-2 cells. (**E**) Bar plot showing the pathways enriched in basal-2 cells than basal-1 cells. Normalized enrichment score calculated by GSEA of DEGs. $P < 0.01$. (**F**) UMAP showing the expression of *SCGB1A1* and *SCGB3A2* in *in vitro* airway cells. (**G**) Bar chart depicting the proportion of various airway cell types in each of the 38 samples, which are ordered with culture methods and passage numbers. (**H**) UMAP showing the distribution of each culture data from (B). (**I**) The ratios of cycling cells for each culture condition, calculated by (number of S + G2M phase cells)/(number of all cells) of different cell types. Box plots with 10-90 percentile whiskers contain each sample derived from different donors.

Fig. 3. Combination of scRNA-seq and scATAC-seq to trace the lineage trajectory of rare epithelial cell types. (**A**) UMAP visualization of a rare cell score calculated from an enrichment score based on transcriptome similarity with a rare cell. (**B**) UMAP labeled with Louvain clusters, colored by 23 separate clusters. (**C**) Mean scatter plot with 95% CI showing an average rare cell score of each Louvain cluster. (**D**) Re-clustered UMAP with Louvain cluster 13 and 22 in (B) and (C), colored by a rare cell score (n = 2,962 cells). (**E**) Re-clustered UMAP revealing all of the expected rare epithelial cell types such as ionocyte, tuft, and pulmonary neuroendocrine cells (PNEC), and a new transitional airway basal cell state (TAB). (**F**) Pseudotime trajectory analysis measured by Monocle3. (**G**) Dot plot representing differential marker gene expressions for each cell type and state. (**H**) Smoothed expression curves across pseudotime started from basal-2 (t=0) show relative expression change of indicated genes. (**I** and **J**) UMAP with data of gene expressions (I) and ATAC peaks (J) obtained from 10x multiome sequencing. (**K**) Top six motif matrices and transcription factors calculated by logistic repression with basal-2 and TAB-specific open regions. (**L** and **M**) Bar plot with 95% CI shows normalized expressions measured by SCTransform (L) and motif activity measured by ChromVar (M) for ASCL1.

Fig. 4. Comparative scRNA-seq analysis of *in vitro* airway models and *in vivo* lung tissues identifies immune primed cell states. (**A**) Integrated UMAP visualization with cells in *in vitro* airway models and *in vivo* human lung tissues (n = 104,577 cells total) [1]. Note, alveolar lineage cells such as alveolar type 1 (AT1) and type 2 (AT2) cells are labeled only in *in vivo* lungs but not in *in vitro* cultures. (**B**) Pie charts showing the proportion of different cell types in each culture method and tissue. Only airway cells are used to measure the proportion of cell types in *in vivo* lungs. (**C**) UMAP showing distinct immune primed cell states only in basal and ciliated cells. Pie charts show the greater proportion of immune primed basal and ciliated cells

in *in vivo* lungs against *in vitro* lung cells. (**D**) Volcano plot showing differentially expressed immune responsive genes in immune primed vs non-primed basal cells. Dot size represents a ratio of immune primed basal cells (pct) expressing each gene. (**E**) Bar plot showing the pathways enriched in immune primed basal and ciliated cells than non-primed cells. Normalized enrichment score of HALLMARK gene set using GSEA. P <0.001. (**F**) Relative proportions of immune primed basal cells in *in vivo* human lung single cell transcriptome data with chronic respiratory disease [4,47,62]. Each value is normalized by their healthy controls. (**G** and **H**) Relative proportions of immune primed basal and ciliated cells in 3D-AO with the donor characteristics (G) and cytokine treatments (H). Each group normalized with healthy (G) and vehicle treated (H) samples. (**I**) The proportion of immune primed basal and ciliated cells over multiple passages in 3D-AO. *P < 0.05, **P <0.01, ***P < 0.001, ****P < 0.0001, unpaired t test.

Fig. 5. Transcriptomic responses of airway cells to respiratory virus infections. (**A**) Schematic workflow of virus infection and scRNA-seq of ALI-Om. (**B**) UpSet plot showing intersections of sequenced cells in each infection group. (**C**) Relative infection rate of SARS-CoV-2 viruses (alpha, beta, delta, and omicron) in ALI-Om derived from normal and diseased lung tissues. Values are normalized with an average infection rate of normal samples in each virus separately. *P < 0.05, unpaired t test. (**D**) Relative infection rate of SARS-CoV-2 viruses with age and sex of each infected sample. Two-tailed Pearson correlation. (**E**) UMAPs highlighted with query cells representing reference mapped cells from each infection group. Reference represents OSCA integrated into lung tissue data and query represents each ALI-Om data infected with indicated viruses. (**F**) Stacked bar charts showing the proportion of cell types in each infection group. (**G**) Bar plot with 95% CI representing the ratio of immune primed basal cells in each infection group. *P < 0.05, ****P < 0.0001, unpaired t test between vehicle and each infection group. (**H**) Dot plot representing the expression levels and fraction of cells that express indicated immune responsive marker genes according to the cell type and infection groups.