Lung epithelial stem cells express SARS-CoV-2 entry factors: implications for COVID-19

Anna A. Valyaeva,[1] Anastasia A. Zharikova,[1,2], Artem S. Kasianov,[2] Yegor S. Vassetzky,[3,4] and Eugene V. Sheval[5,6]

[1]*Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, 119991 Moscow, Russia*

[2]*The Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), 127051 Moscow, Russia*

[3]*CNRS, UMR 9018, Université Paris-Saclay, Institut Gustave Roussy, Villejuif, France*

[4]*Koltzov Institute of Developmental Biology, 117334 Moscow, Russia*

[5]*Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119991 Moscow, Russia*

[6]*Department of Cell Biology and Histology, Faculty of Biology, Lomonosov Moscow State University, 119991 Moscow, Russia*

Correspondence: yegor.vassetzky@cnrs.fr (Y.S.V.), sheval_e@belozersky.msu.ru (E.V.S.)

Summary

SARS-CoV-2 can infiltrate the lower respiratory tract, resulting in severe respiratory failure and a high death rate. Normally, the airway and alveolar epithelium can be rapidly reconstituted by multipotent stem cells after episodes of infection. Here, we analyzed published RNA-seq datasets and demonstrated that cells of four different lung epithelial stem cell types express SARS-CoV-2 entry factors, including *Ace2*. Thus, stem cells can be infected by SARS-CoV-2, which can lead to defects in regeneration capacity and account for the severity of SARS-CoV-2 infection and its consequences.

Introduction

The emergence of the novel SARS coronavirus 2 (SARS-CoV-2), which causes the respiratory disease COVID-19, poses an important threat to global public health. In contrast to endemic human coronaviruses, SARS-CoV-2 can infiltrate the lower respiratory tract, resulting in severe respiratory failure and a high death rate (Zhou et al., 2020; Zhu et al., 2020). Infection of epithelial cells of the two major compartments in the lungs, the conducting airways and the gas-exchanging alveoli, seems to be a main cause of COVID-19 development (Mason, 2020). Both the airway and alveolar epithelia are characterized by relatively slow renewal but can be rapidly reconstituted by multipotent stem cells after episodes of infection, inflammation or injury, which are commonly observed in respiratory diseases (Basil et al., 2020; Chen, 2017; Hogan et al., 2014).

To date, several types of epithelial stem cells in the distal airways and pulmonary alveoli have been identified in mice. Basal cells in the proximal airways are the major stem cell population that self-renew and, when necessary, give rise to multiple cell types, such as secretory club cells and ciliated cells (Montoro et al., 2018; Pardo-Saganta et al., 2015; Rock et al., 2009; Watson et al., 2015). The pulmonary alveolar epithelium is composed of two types of differentiated epithelial cells: alveolar type 1 (AT1) cells, which mediate gas exchange, and alveolar type 2 (AT2) cells, which secrete surfactant. A subpopulation of AT2 cells serves as alveolar stem cells that can differentiate into both AT1 and AT2 cells (Nabhan et al., 2018; Zacharias et al., 2018). In addition, lung injury activates specialized epithelial stem cells residing in the distal airways to reestablish alveolar barrier function. A number of distinct, small populations of stem cells have been reported to contribute to regeneration after lung injury: variant club/secretory cells (v-club cells), bronchioalveolar stem cells (BASCs), H2-K1^{high} club cell-like stem cells and p63^{+} basal cells (for review see: Basil et al., 2020).

Infection of epithelial stem cells can potentially lead to defects in lung regeneration capacity. Analysis of the expression of viral entry factors helps to identify human cells that can be infected by SARS-CoV-2. Cellular entry of coronaviruses depends on binding of the viral spike (S) proteins to cellular receptors and on S protein priming by host cell proteases. SARS-CoV-2 uses angiotensin-converting enzyme 2 (ACE2) for entry (Hoffmann et al., 2020a; Letko et al., 2020; Zhou et al., 2020) and the TMPRSS2 and FURIN proteases for S protein priming (Hoffmann et al., 2020a, b). Thus, (co)expression of ACE2, TMPRSS2 and FURIN is a convenient marker of cells that can be potentially infected by SARS-CoV-2 (Lukassen et al., 2020). Additional proteases potentially involved in SARS-CoV-2 priming

include ANPEP, used by HCoV-229E (Yeager et al., 1992), and DPP4, used by MERS-CoV (Raj et al., 2013). However, no data have been published on this subject for SARS-CoV-2.

Expression of SARS-CoV-2 entry factors can be analyzed using publicly available RNA-seq datasets. These factors, including ACE2, are highly expressed in nasal epithelial cells, but ACE2 expression on the cells of conducting airways and lung parenchyma is substantially lower (Sungnak et al., 2020). SARS-CoV-2 entry factors are expressed in secretory and ciliated cells of the conducting airways (Lukassen et al., 2020; Sungnak et al., 2020). In the alveolar epithelium, ACE2 expression is found only in a small subset (1-7%) of AT2 cells (Bezara et al., 2020; Lukassen et al., 2020; Qi et al., 2020; Sungnak et al., 2020; Zhao et al., 2020; Ziegler et al., 2020), although the severity of the disease suggests a more widespread distribution. Published data on the expression of SARS-CoV-2 entry factors in stem cells are restricted to basal cells of the airway epithelium, but infection of epithelial stem cells can lead to defects in lung regeneration.

Here, to determine whether lung stem cells can be infected by SARS-CoV-2, we analyzed the expression of SARS-CoV-2 entry factors in different epithelial stem cells using publicly available RNA-seq data. Because of the limited data on lung stem cells in humans we conducted this study on data obtained from mice. We demonstrated that different lung epithelial stem cells expressed SARS-CoV-2 entry factors and thus could be infected by SARS-CoV-2. This ability may account for the slow reconstitution of the lung epithelium during and after SARS-CoV-2 infection and may partially explain the severity of the disease.

Results

Airway epithelial cell types include basal cells, secretory club cells and ciliated cells, as well as several rare cell types – neuroendocrine, goblet and tuft cells and ionocytes. Basal cells are a heterogeneous population of stem cells of the conducting airways that can self-renew and differentiate into both secretory and ciliated epithelial lineages.

Recent studies in which publicly available scRNA-seq data were reanalyzed to detect *ACE2*-expressing cells demonstrated that only a small subpopulation of basal cells express *ACE2* and other SARS-CoV-2 entry factors; *ACE2* expression increased upon differentiation to secretory club cells (Sungnak et al., 2020). These data are inconsistent with the high pathogenicity of SARS-CoV-2. We proposed that the expression of *ACE2* and other SARS-CoV-2 entry factors might be underestimated due to the presence of "dropout" events, i.e., a gene is detected in one cell but is not detected in another cell, usually due to extremely low mRNA input and/or the stochastic nature of gene expression (Grün et al., 2014; Kharchenko

et al., 2014; Mereu et al., 2020; Stegle et al., 2015). We therefore compared datasets obtained by Montoro and coauthors using two different methods, 3′ single-cell RNA-seq (scRNA-seq) and full-length scRNA-seq (Montoro et al., 2018). Full-length scRNA-seq allows a higher number of reads per cell for each gene than 3′ scRNA-seq (396,000 reads/cell and 23,000 reads/cell, respectively (Montoro et al., 2018)), which can lead to a much more precise estimation of gene expression.

We analyzed the expression of SARS-CoV-2 entry factors only in basal, ciliated and club cells because the number of other cells in the full-length scRNA-seq dataset was extremely low. The proportion of cells expressing *Ace2* and the priming proteases was substantially higher in the full-length scRNA-seq data (Figure 1A). For example, the proportions of *Ace2*$^+$ basal cells were 0.60% and 9.38% in the 3′ scRNA-seq and full-length scRNA-seq datasets, respectively. This indicates that the population of epithelial cells potentially sensitive to SARS-CoV-2 infection is underestimated by 3′ scRNA-seq. The expression patterns of different genes by 3′ scRNA-seq and full-length scRNA-seq were similar for highly expressed genes (e.g., *Tmprss2*), although the detected expression levels of genes with relatively low mean expression levels were substantially higher for 3′ scRNA-seq (Figure 1B). The proportions of cells with greater-than-zero expression of different genes were extremely variable between the two datasets, particularly for genes with low and moderate expression levels; thus the dropout effect was higher for these genes (Figure 1C). Thus, the number of cells expressing SARS-CoV-2 entry factors obtained using 3′ scRNA-seq and similar methods used in cell atlas projects could be underestimated and should be carefully interpreted.

Only one type of lung stem cell – the basal cells of the conducting airways – has been extensively studied in both mice and humans. For the other types of stem cells, quality RNA-seq data can only be found for mice. However, the patterns of SARS-CoV-2 entry factor expression on different epithelial cells of mouse and human airways (Plasschaert et al., 2018) are highly similar (Figure S1), indicating that datasets from mice can be used for analysis of SARS-CoV-2 entry factors.

AT2 cells serve as alveolar stem cells and can differentiate into AT1 cells during alveolar homeostasis and postinjury repair (Barkauskas et al., 2013; Desai et al., 2014), but a small subpopulation of AT1 cells retains cellular plasticity and can transdifferentiate into AT2 cells, maintaining tissue integrity during alveolar regeneration (Wang et al., 2018). These cells are characterized by the *Hopx*$^+$/*Igfbp2*$^-$ phenotype, but they do not form separated clusters in *t*-distributed stochastic neighbor embedding (*t*-SNE) plots (Wang et al., 2018). We

analyzed the expression of SARS-CoV-2 entry factors in *Igfbp2*$^-$ and *Igfbp2*$^+$ AT1 cells (Figure 2A). Neither terminally differentiated *Hopx*$^+$/*Igfbp2*$^+$ nor *Hopx*$^+$/*Igfbp2*$^-$ cells expressed *Ace2*, indicating that AT1 cells are probably resistant to SARS-CoV-2 infection. Although it is generally accepted that SARS-CoV preferentially infects AT2 cells (Mossel et al., 2008), SARS-CoV-2 can infect both AT1 and AT2 cells *ex vivo* (Chu et al., 2020) and in macaques (Rockx et al., 2020). These data may indicate the presence of alternative, yet-unidentified pathways of SARS-CoV-2 entry into AT1 cells.

The major contributor to alveolar epithelial renewal is a subpopulation of AT2 cells, which serve as alveolar stem cells (Nabhan et al., 2018; Zacharias et al., 2018). These cells, which express *Axin2*, are referred to as alveolar epithelial progenitors (AEPs) (Zacharias et al., 2018). The bulk RNA-seq data for *Axin2*$^+$ AEPs and *Axin2*$^-$ AT2 cells (Zacharias et al., 2018) were reanalyzed starting from raw reads, and transcript abundances were estimated by Kallisto (Bray et al., 2016). The expression of *Ace2, Tmprss2* and *Dpp4* was 2.68-fold, 3.79-fold and 2.31-fold, respectively, higher in AEPs than in *Axin2*$^-$ AT2 cells (Figure 2B). Additionally, the data on SARS-CoV-2 entry factor expression were extracted from another scRNA-seq dataset (Wang et al., 2018). The percentage of *Axin2*$^-$ AT2 cells with *Ace2* expression was higher than that of differentiated AT2 cells (4.20% and 2.23%, respectively), but the level of *Ace2* expression was similar in *Axin2*$^+$ and *Axin2*$^-$ cells (Figure 2C). Interestingly, the number of cells expressing *Tmprss2* was low. However, both datasets indicate that alveolar stem cells (i.e., AEPs) could be even more sensitive to SARS-CoV-2 infection than differentiated AT2 cells.

The alveolar epithelium has a relatively low regeneration potential; therefore, distal airway stem cells mobilize after lung injury to occupy alveolar surfaces. Alveolar regeneration in humans after SARS-CoV-2 infection has not yet been described, but epithelial regeneration in small bronchioles has been demonstrated in a nonhuman primate model (Rockx et al., 2020). Numerous distinct populations of stem cell types have been reported to contribute to regeneration after injury (Basil et al., 2020), but we found RNA-seq datasets for only two types of stem cells.

(i) BASCs are activated and respond distinctly to different lung injuries; they also differentiate into multiple cell lineages, including club cells and ciliated cells of the terminal bronchioles and AT1 and AT2 cells of the alveoli (Liu et al., 2019; Salwig et al., 2019). Our reanalysis of a bulk RNA-seq dataset (Salwig et al., 2019) indicated that BASCs express elevated levels of *Ace2* and *Tmprss2* compared to AT2 cells (Figure 2D, left panel). In contrast, club cells express higher levels of these two entry factors than BASCs (Figure 2D,

right panel). We also reanalyzed a scRNA-seq dataset (Liu et al., 2019) to achieve better separation of the BASC cluster from other cell clusters in the *t*-SNE plot and refine the cluster labels (Figure S2A). BASCs coexpressed markers of AT2 cells (*Sftpc*) and club cells (*Scgb1a1*) (Figure S2B). Both the number of *Ace2*-expressing cells and the expression levels of *Ace2* were higher in BASCs than in AT2 cells but lower in BASCs than in club cells (Figure 2E). Interestingly, the proportion of *Ace2*-expressing ciliated cells was roughly equal to that of BASCs, but the *Ace2* expression level in ciliated cells was substantially lower.

(ii) Recently, a rare population of H2-K1^{high} club cell-like stem cells has been described. These cells, which differentiate into AT1 and AT2 cells following bleomycin-induced lung injury, were identified in scRNA-seq datasets of murine distal airways (Kathiriya et al., 2020). Since the cell type annotation for H2-K1^{high} cells was unavailable, we reanalyzed this dataset (Figure 2F). Club-like cells identified by the expression of secretory cell marker genes were reclustered into six subpopulations, including a H2-K1^{high} cell population (Fig. S2C). A relatively large proportion of H2-K1^{high} cells expressed *Ace2* (~37.93%) and other viral entry factors, indicating that this type of stem cell is also sensitive to SARS-CoV-2 infection (Figure 2G).

Discussion

SARS-CoV-2 can infect lung cells and induce severe respiratory failure, but this distinctive characteristic of COVID-19 is inconsistent with recent published reports based on scRNA-seq, which indicate that only a minor fraction of lower respiratory tract epithelial cells express SARS-CoV-2 entry factors (Bezara et al., 2020; Lukassen et al., 2020; Sungnak et al., 2020; Ziegler et al., 2020). To resolve this apparent contradiction, we compared two scRNA-seq datasets obtained using 3′ scRNA-seq and full-length scRNA-seq and found that the number of cells expressing *Ace2* could be substantially underestimated due to the dropout effect. Additionally, we found that epithelial stem cells (basal cells, AEPs, BASCs and H2-K1^{high} cells) express *Ace*2 and other SARS-CoV-2 entry factors. The expression of these factors in different stem cells was relatively low, but importantly, in the gas-exchanging alveoli, AEPs exhibited higher expression of SARS-CoV-2 entry factors than differentiated AT1 and AT2 cells. Notably, RNA-seq data for only murine lung epithelial stem cells have been published; however, the expression of SARS-CoV-2 entry factors may be a general feature of lung stem cells, making these cells probable targets of SARS-CoV infection in humans. Infection of stem cells and their subsequent loss can result in a decreased capacity for lung epithelial regeneration, which could be a determinant of SARS-CoV-2 pathogenicity.

Methods

Bulk RNA-seq analysis

The data from GSE97055 and GSE129440 datasets was reprocessed from raw reads and raw gene counts, respectively. Raw pair-end reads were quality and adapter trimmed with BBDuk (minlen=31 qtrim=r trimq=20 ktrim=r k=25 mink=11 hdist=1) from BBTools suite, and FastQC was used for quality control. Then the reads were pseudo-aligned to the mouse transcriptome (obtained from GRCm38 primary genome assembly and GENCODE gene annotation version M24 (https://www.gencodegenes.org/mouse/release_M24.html) using kallisto (Bray et al., 2016) with default parameters. TPM (transcripts per million) values provided by kallisto were imported and summarized into a gene-level matrix using tximport R package (Soneson et al., 2015). Differential gene expression analysis was carried out using DESeq2 R package (Love et al., 2014). Genes with Benjamini-Hochberg adjusted p-values < 0.05 were declared differentially expressed. Volcano plots were generated using ggplot2 R package (Wickham, 2016).

scRNA-seq analysis

For GSE103354 dataset containing both 3'-droplet-based and full-length plate-based scRNA-seq data UMI (unique molecular identifier) counts and TPM values were available (Montoro et al., 2018). 3' scRNA-seq data was normalized using NormalizeData function from Seurat R package (Stuart et al., 2019), and log(TP10K+1) values hereafter referred to as log(TPM+1) were obtained. TPM values of full-length scRNA-seq experiment were rescaled to add up to 10000 and log-transformed for better comparability between experiments. The further analysis of this dataset included generating average expression estimates (log of mean TPM) and percent of gene expressing cells for clusters of basal, club, and ciliated cells based on cell labelling from the original paper (Montoro et al., 2018).

Filtered and normalized by library size (see Plasschaert et al., 2018 for details) expression data from GSE102580 dataset was log-transformed and analysed as described above.

Raw gene counts from GSE118891 dataset were converted to TPM values (gene lengths calculated as the union of all gene exons were obtained from Ensembl (v91) gene annotation) and $\log_2$-transformed. Cells with fewer than 2000 or more than 7500 detected genes and more than 10% of mitochondrial fraction were excluded from the dataset. 472 high-quality cells out of 480 were used for further analysis. Log_2(TPM+1) values were imported into Seurat without any further normalization procedure, and standard Seurat clustering pipeline was applied. Principal-component analysis (PCA) was performed based

on 2000 highly variable genes (HVGs) using RunPCA function. First 10 principal components (PCs) were used as the input to construction of Shared Nearest Neighbor (SNN) graph (FindNeighbors function, k.param = 20), finding clusters (FindClusters function, resolution = 0.5) and visualising identified clusters with t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm (RunTSNE function, perplexity = 30). Although 4 generated clusters were mostly consistent with cell type labels from the original paper (Liu et al., 2019), cell type identities were refined based on the expression of known marker genes (Sftpc, Scgb1a1, Ly6a, H2-Ab, etc.) and cluster markers identified as differentially expressed genes between group of cells by FindAllMarkers function with parameters test.use = “wilcox”, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25.

Log(TPM+1) expression values (scaling factor 10000) from GSE106960 dataset were imported into Seurat with no further normalization. Since the standard cell clustering approach failed at separating $Hopx^{+}$/$Igfbp2^{-}$ cell subpopulation from the bulk AT1 cells, as well as $Axin2^{+}$ cells from AT1 cells, subclusters of interest were identified by the absolute expression of marker genes: *Hopx* and *Igfbp2* for AT1 cells and *Axin2* and *Sftpc* for AT2 cells.

For GSE129937 dataset raw UMI counts were available. Low-quality cells were filtered out with a mitochondrial fraction threshold of 10% and number of detected genes threshold of 1000; 3278 high-quality cells were retained for further analysis. After the data was normalized by the means of Seurat package (NormalizeData function with default parameters), linear dimensional reduction (PCA) was conducted based on 2000 HVGs (identified by FindVariableFeatures function with default parameters). Scores from first 8 significant PCs (they were identified by inspecting the elbow plot) were used for constructing SNN graph (FindNeighbors function, k.param = 20), finding clusters (FindClusters function, resolution = 0.4) and subsequent t-SNE visualization (RunTSNE function, perplexity = 50). The described analysis resulted in the identification of 10 clusters of single cells. Based on cluster marker genes obtained by FindAllMarkers function (only.pos = TRUE, test.use = "MAST"; requires MAST R package(Finak et al., 2015)) cell type identities were assigned to clusters. Cells from 5 smaller clusters were pooled into one large cluster of secretory club-like cells based on the expression of known club cell's marker genes. The pooled dataset of club-like cells were then analysed independently. To identify H2-$K1^{high}$ cell population with a high progenitor gene-signature supervised clustering using club cells' HVGs and the list of Sox9-associated progenitor genes (Ostrin et al., 2018) was carried out. Sox9-based progenitor gene list combined with HVGs was used as feature input for PCA. First 10 PCs declared as

significant by inspecting the elbow plot were then used to find clusters (FindNeighbors followed by FindClusters function, k.param=20, resolution = 0.5) and visualize them by t-SNE algorithm (RunTSNE function, perplexity = 50). H2-K1^{high} specific genes from Kathiriya et al., 2020 were used to locate the cluster of interest.

For all scRNA-seq datasets the average expression values (log of mean TPM) of SEFs and percent of cells expressing SEFs were calculated for every cluster of interest using DotPlot function from Seurat.

Acknowledgments

We are grateful to A.A. Penin and A.A. Saidova for their helpful discussion and assistance. This work was supported by the Russian Science Foundation grant 18-14-00195 to E.V.S. Computational resources of the Makarich HPC cluster were provided by the Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University.

Author contributions

A.A.V., A.A.Z., A.S.K. conducted the experiments and wrote the paper, Y.S.V. analyzed the data and wrote the paper; E.V.S. conceived and designed the study, analyzed the data and wrote the paper; the manuscript was approved by all authors.

Declaration of interests

The authors declare no competing interests.

References

Barkauskas, C.E., Cronce, M.J., Rackley, C.R., Bowie, E.J., Keene, D.R., Stripp, B.R., Randell, S.H., Noble, P.W., and Hogan, B.L.M. (2013). Type 2 alveolar cells are stem cells in adult lung. J. Clin. Invest. *123*, 3025–3036.

Basil, M.C., Katzen, J., Engler, A.E., Guo, M., Herriges, M.J., Kathiriya, J.J., Windmueller, R., Ysasi, A.B., Zacharias, W.J., Chapman, H.A., et al. (2020). The cellular and physiological basis for lung repair and regeneration: past, present, and future. Cell Stem Cell *26*, 482–502.

Bezara, M.E., Thurman, A., Pezzulo, A., Leidinger, M.R., Klesney-Tait, J.A., Karp, P.H., Tan, P., Wohlford-Lenane, C., McCray, P.B., and Meyerholz, D.K. (2020). Heterogeneous expression of the SARS-Coronavirus-2 receptor ACE2 in the human respiratory tract. BioRxiv preprint https://doi.org/10.1101/2020.04.22.056127

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Chen, J. (2017). Origin and regulation of a lung repair kit. Nat. Cell Biol. *19*, 885–886.

Chu, H., Chan, J.F.-W., Wang, Y., Yuen, T.T.-T., Chai, Y., Hou, Y., Shuai, H., Yang, D., Hu, B., Huang, X., et al. (2020). Comparative replication and immune activation profiles of SARS-CoV-2 and SARS-CoV in human lungs: an *ex vivo* study with implications for the pathogenesis of COVID-19. Clin. Infect. Dis. https://doi.org/10.1093/cid/ciaa410

Desai, T.J., Brownfield, D.G., and Krasnow, M.A. (2014). Alveolar progenitor and stem cells in lung development, renewal and cancer. Nature *507*, 190–194.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. *16*, 278.

Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat. Methods *11*, 637–640.

Hoffmann, M., Kleine-Weber, H., and Pöhlmann, S. (2020b). A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. Mol. Cell https://doi.org/10.1016/j.molcel.2020.04.022

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., et al. (2020a). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell *181*, 271–280.

Hogan, B.L.M., Barkauskas, C.E., Chapman, H.A., Epstein, J.A., Jain, R., Hsia, C.C.W., Niklason, L., Calle, E., Le, A., Randell, S.H., et al. (2014). Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. Cell Stem Cell *15*, 123–138.

Kathiriya, J.J., Brumwell, A.N., Jackson, J.R., Tang, X., and Chapman, H.A. (2020). Distinct airway epithelial stem cells hide among club cells but mobilize to promote alveolar regeneration. Cell Stem Cell *26*, 346–358.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat. Methods *11*, 740–742.

Letko, M., Marzi, A., and Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol *5*, 562–569.

Liu, Q., Liu, K., Cui, G., Huang, X., Yao, S., Guo, W., Qin, Z., Li, Y., Yang, R., Pu, W., et al. (2019). Lung regeneration by multipotent stem cells residing at the bronchioalveolar-duct junction. Nat. Genet. *51*, 728–738.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lukassen, S., Chua, R.L., Trefzer, T., Kahn, N.C., Schneider, M.A., Muley, T., Winter, H., Meister, M., Veith, C., Boots, A.W., et al. (2020). SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. EMBO J. e105114.

Mason, R.J. (2020). Pathogenesis of COVID-19 from a cell biology perspective. Eur. Respir. J. *55*, 2000607.

Mereu, E., Lafzi, A., Moutinho, C., Ziegenhain, C., McCarthy, D.J., Álvarez-Varela, A., Batlle, E., Sagar, Grün, D., Lau, J.K., et al. (2020). Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. Nat. Biotechnol. https://doi.org/10.1038/s41587-020-0469-4

Montoro, D.T., Haber, A.L., Biton, M., Vinarsky, V., Lin, B., Birket, S.E., Yuan, F., Chen, S., Leung, H.M., Villoria, J., et al. (2018). A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature *560*, 319–324.

Mossel, E.C., Wang, J., Jeffers, S., Edeen, K.E., Wang, S., Cosgrove, G.P., Funk, C.J., Manzer, R., Miura, T.A., Pearson, L.D., et al. (2008). SARS-CoV replicates in primary human alveolar type II cell cultures but not in type I-like cells. Virology *372*, 127–135.

Nabhan, A.N., Brownfield, D.G., Harbury, P.B., Krasnow, M.A., and Desai, T.J. (2018). Single-cell Wnt signaling niches maintain stemness of alveolar type 2 cells. Science *359*, 1118–1123.

Ostrin, E.J., Little, D.R., Gerner-Mauro, K.N., Sumner, E.A., Ríos-Corzo, R., Ambrosio, E., Holt, S.E., Forcioli-Conti, N., Akiyama, H., Hanash, S.M., et al. (2018). β-Catenin maintains lung epithelial progenitors after lung specification. Development *145*, dev160788.

Pardo-Saganta, A., Law, B.M., Tata, P.R., Villoria, J., Saez, B., Mou, H., Zhao, R., and Rajagopal, J. (2015). Injury induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell subpopulations. Cell Stem Cell *16*, 184–197.

Plasschaert, L.W., Žilionis, R., Choo-Wing, R., Savova, V., Knehr, J., Roma, G., Klein, A.M., and Jaffe, A.B. (2018). A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature *560*, 377–381.

Qi, F., Qian, S., Zhang, S., and Zhang, Z. (2020). Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. Biochem. Biophys. Res. Commun. *526*, 135–140

Raj, V.S., Mou, H., Smits, S.L., Dekkers, D.H.W., Müller, M.A., Dijkman, R., Muth, D., Demmers, J.A.A., Zaki, A., Fouchier, R.A.M., et al. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature *495*, 251–254.

Rock, J.R., Onaitis, M.W., Rawlins, E.L., Lu, Y., Clark, C.P., Xue, Y., Randell, S.H., and Hogan, B.L.M. (2009). Basal cells as stem cells of the mouse trachea and human airway epithelium. Proc. Natl. Acad. Sci. U. S. A. *106*, 12771–12775.

Rockx, B., Kuiken, T., Herfst, S., Bestebroer, T., Lamers, M.M., Oude Munnink, B.B., de Meulder, D., van Amerongen, G., van den Brand, J., Okba, N.M.A., et al. (2020). Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. Science doi: 10.1126/science.abb7314

Salwig, I., Spitznagel, B., Vazquez-Armendariz, A.I., Khalooghi, K., Guenther, S., Herold, S., Szibor, M., and Braun, T. (2019). Bronchioalveolar stem cells are a main source for regeneration of distal lung epithelia *in vivo*. EMBO J. *38*, e102099.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. *16*, 133–145.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21.

Sungnak, W., Huang, N., Bécavin, C., Berg, M., Queen, R., Litvinukova, M., Talavera-López, C., Maatz, H., Reichart, D., Sampaziotis, F., et al. (2020). SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. Nat. Med. *26*, 681–687.

Wang, Y., Tang, Z., Huang, H., Li, J., Wang, Z., Yu, Y., Zhang, C., Li, J., Dai, H., Wang, F., et al. (2018). Pulmonary alveolar type I cell population consists of two distinct subtypes that differ in cell fate. Proc. Natl. Acad. Sci. U. S. A. *115*, 2407–2412.

Watson, J.K., Rulands, S., Wilkinson, A.C., Wuidart, A., Ousset, M., Van Keymeulen, A., Göttgens, B., Blanpain, C., Simons, B.D., and Rawlins, E.L. (2015). Clonal dynamics reveal two distinct populations of basal cells in sow-turnover airway epithelium. Cell Rep. *12*, 90–101.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag).

Yeager, C.L., Ashmun, R.A., Williams, R.K., Cardellichio, C.B., Shapiro, L.H., Look, A.T., and Holmes, K.V. (1992). Human aminopeptidase N is a receptor for human coronavirus 229E. Nature *357*, 420–422.

Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y.,and Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. BioRxiv preprint https://doi.org/10.1101/2020.01.26.919985

Zacharias, W.J., Frank, D.B., Zepp, J.A., Morley, M.P., Alkhaleel, F.A., Kong, J., Zhou, S., Cantu, E., and Morrisey, E.E. (2018). Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. Nature *555*, 251–255.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature *579*, 270–273.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. *382*, 727–733.

Ziegler, C., Allon, S.J., Nyquist, S.K., Mbano, I., Miao, V.N., Cao, Y., Yousif, A.S., Bals, J., Hauser, B.M., Feldman, J., et al. (2020). SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is enriched in specific cell subsets across tissues. Cell https://doi.org/10.1016/j.cell.2020.04.035

Figures

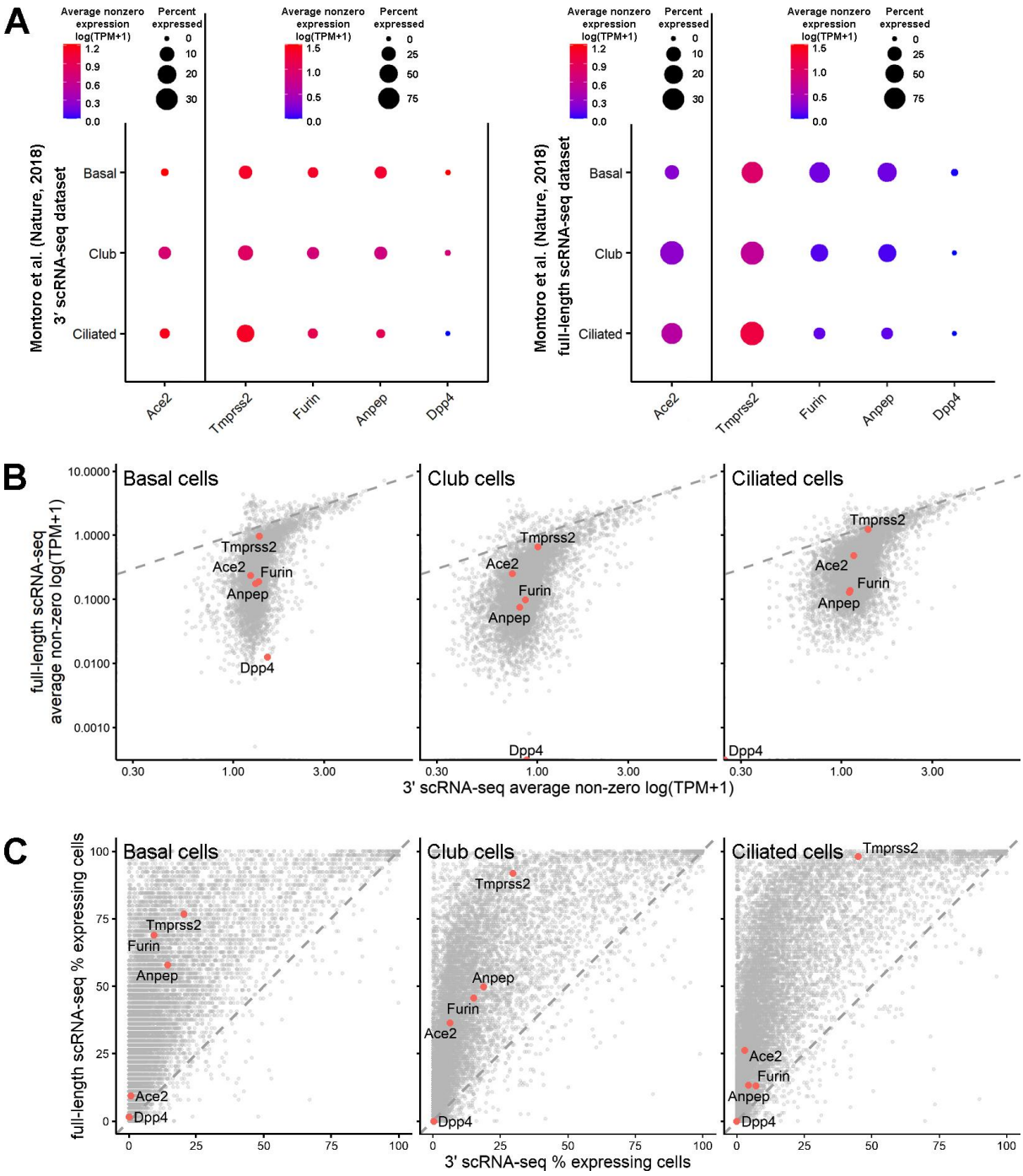

Figure 1. The proportions of cells expressing SARS-CoV-2 entry factors are underestimated in standard scRNA-seq datasets. (A) Expression of the SARS-CoV-2 entry factors *Ace2*, *Tmprss2, Furin*, *Anpep* and *Dpp4* in mouse trachea datasets from (Montoro et al., 2018): 3′ scRNA-seq dataset (left panel) and full-length scRNA dataset (right panel). For the 3′ scRNA-seq dataset, unique molecular identifier (UMI) counts were normalized to account for differences in coverage, multiplied by a scaling factor of 10000 to generate transcripts per kilobase million (TPM)-like values, and then log transformed. TPM

values from the full-length scRNA dataset were rescaled to sum to 10000 and were log transformed. Gene expression estimates were summarized in accordance with the cell type labels provided in the original paper. The dot size indicates the proportion of cells among the respective cell type population with greater-than-zero expression of the respective SARS-CoV-2 entry factor, while the dot color indicates the average nonzero expression value. (B) Correlation between gene expression in the 3′ scRNA-seq dataset and the full-length scRNA dataset for basal cells, club cells and ciliated cells. The *Ace2, Tmprss2, Furin, Anpep* and *Dpp4* expression levels are represented by colored dots. (C) Full-length scRNA-seq detects a substantially higher number of cells with greater-than-zero expression of genes, including SARS-CoV-2 entry factors, among basal cells, club cells and ciliated cells. The percentages of cells expressing *Ace2, Tmprss2, Furin, Anpep* and *Dpp4* are represented by the colored dots.

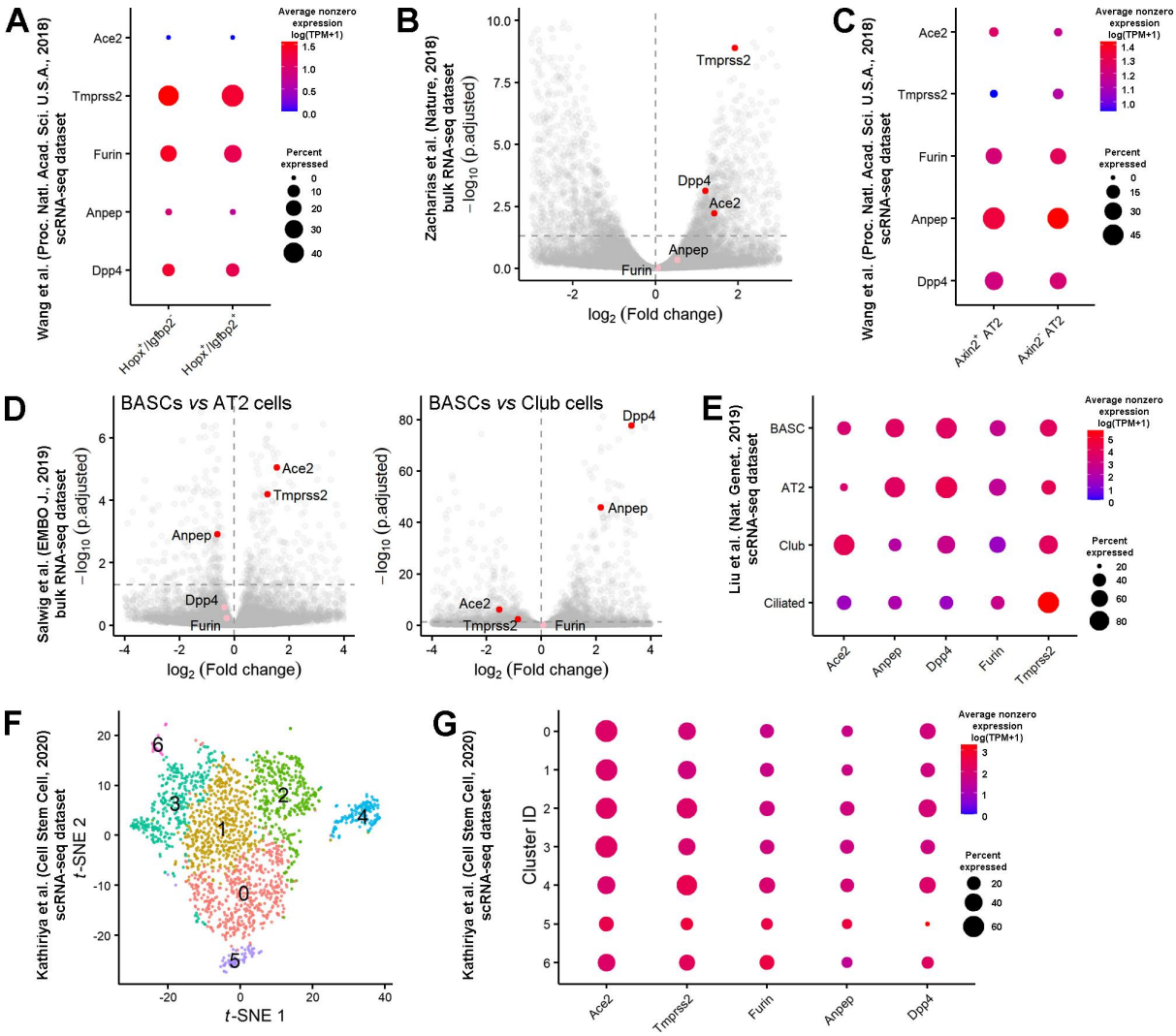

Figure 2. Lung epithelial stem cells express SARS-CoV-2 entry factors. (A) Expression of *Ace2, Tmprss2, Furin, Anpep* and *Dpp4* in the subpopulation of AT1 cells that maintain the ability to transdifferentiate into AT2 cells ($Hopx^+/Igfbp2^-$ cells) and in terminally differentiated AT1 cells ($Hopx^+/Igfbp2^+$ cells). The results for the scRNA-seq dataset from mice at postnatal day 60 (Wang et al., 2018) are presented; similar results were obtained for AT1 cells at postnatal days 3 and 15. The dot size indicates the proportion of cells in the respective cell type with greater-than-zero expression of the respective SARS-CoV-2 entry factor, while the dot color indicates the average nonzero expression value. (B) Volcano plot showing elevated expression of *Ace2, Tmprss2* and *Dpp4* in AEPs compared to differentiated AT2 cells (bulk RNA-seq dataset (Zacharias et al., 2018)). Each dot represents one gene. The $\log_2$ (fold change) in the expression levels of *Ace2, Tmprss2, Furin, Anpep* and *Dpp4* is indicated by the colored dots (red, differentially expressed genes (p.adjusted <0.05); pink, non-differentially expressed genes). (C) Expression of SARS-CoV-2 entry factors in AEP AT2 cells expressing *Axin2* in the dataset from (Wang et al., 2018). (D) Volcano plots

showing the expression levels of *Ace2* and *Tmprss2* in BASCs *versus* differentiated AT2 cells (left panel) or differentiated club cells (right panel). The $\log_2$ (fold change) in the expression levels of differentially expressed genes (p.adjusted < 0.05) is shown in red; of non-differentially expressed genes, in pink. (E) Expression of SARS-CoV-2 entry factors in BASCs, AT2 cells, secretory club cells and ciliated cells. The proportion of *Ace2*-expressing cells in the BASC population was higher than that in the alveolar AT2 cell population but lower than that in the bronchiolar club cell population. (F) *t*-SNE plot of 1952 scRNA-seq profiles from the scRNA-seq dataset (Kathiriya et al., 2020), colored by cluster assignment. Cluster 6 contains H2-K1^{high} club-like stem cells (Figure S2C and D). (G) SARS-CoV-2 entry factors are expressed in a large proportion of H2-K1^{high} club-like stem cells as well as in other clusters of club cells and AT2 cells.

Supplementary figures

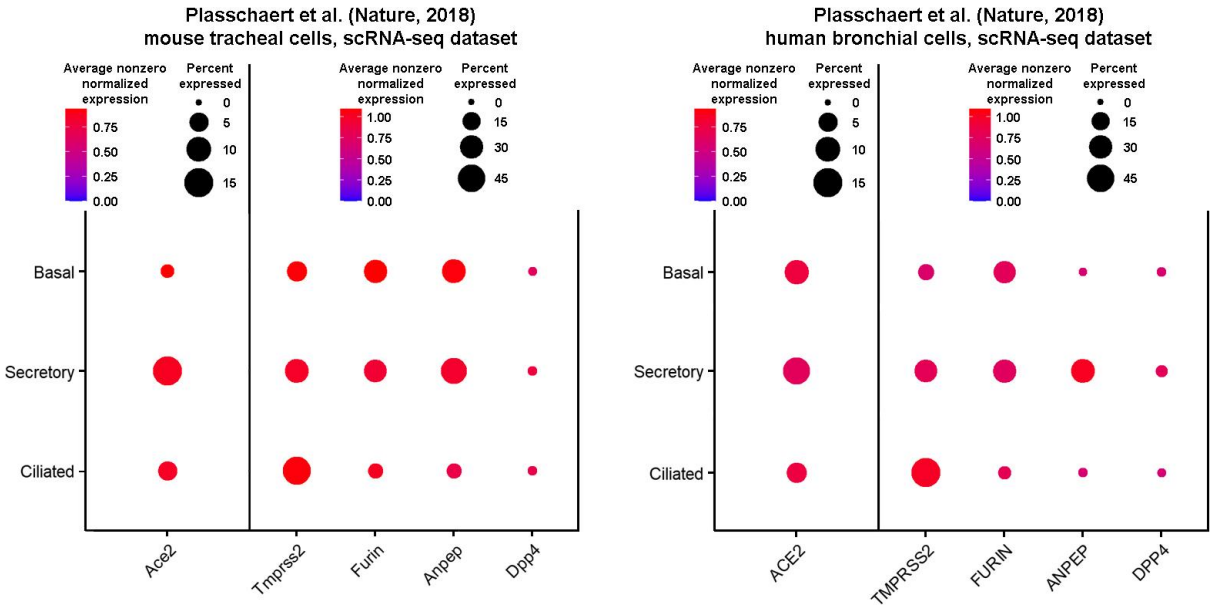

Figure S1. Expression of the SARS-CoV-2 entry factors in mouse and human proximal airway epithelial cells: mouse trachea scRNA-seq dataset (left panel) and human bronchiole scRNA dataset (right panel) (Plasschaert et al., 2018). Gene expression was estimated in accordance with the cell type labels provided in the original paper. The dot size indicates the proportion of cells among the respective cell type population with greater-than-zero expression of the respective SARS-CoV-2 entry factor, while the dot color indicates the average nonzero expression value.

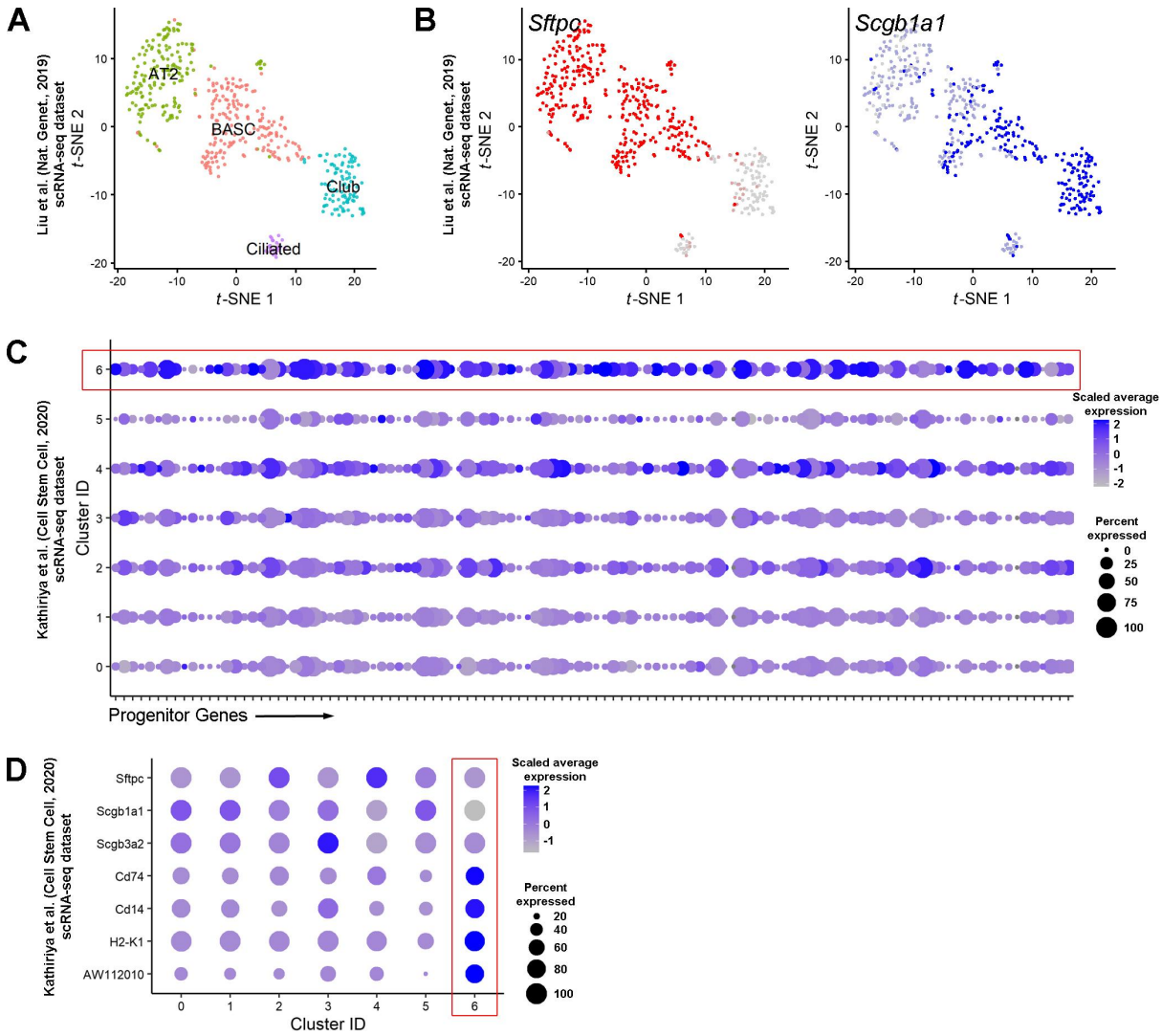

Figure S2. Validation of cell clusters in scRNA-seq datasets. (A) *t*-SNE visualization of 472 scRNA-seq profiles from the scRNA-seq dataset of FACS-sorted murine epithelial cells (Liu et al., 2019), colored by cluster assignment and annotated *post hoc*. (B) *t*-SNE of 472 scRNA-seq profiles (points) colored by expression of representative AT2 cell and club cell markers (*Sftpc* and *Scgb1a1*, respectively). (C) Expression of progenitor cell markers (Ostrin et al. 2018) in different subpopulations (1-6) of club cells (Kathiriya et al., 2020). Cells in Cluster 6 cells demonstrate an elevated expression of progenitor cell markers. (D) Expression of lineage markers of mature club cells (*Scgb1a1* and *Scgb3a2*) and AT2 cells (*Sftpc*). Note that Cluster 6 is negative or low for these markers. The cells in this cluster are characterized by enhanced expression of *Cd14, Cd74, H2-K1*, and the lncRNA AW112010s.