bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Age-related expression of SARS-CoV-2 priming protease TMPRSS2 in the developing lung

Bryce A. Schuler[1], A. Christian Habermann[2], Erin J. Plosa[3], Chase J. Taylor[2], Christopher Jetter[3], Meghan E. Kapp[4], John T. Benjamin[3], Peter Gulleman[3], David S. Nichols[2], Lior Z. Braunstein[5], Michael Koval[6,7], Susan H. Guttentag[3], Timothy S. Blackwell[2,8,9], Vanderbilt COVID-19 Consortium Cohort[10], Steven A. Webber[1], Nicholas E. Banovich[11], Jonathan A. Kropski[2,8,9]*, Jennifer M. S. Sucre [3,8] *, HCA Lung Biological Network

[1]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN
[2]Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN
[3]Division of Neonatology, Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN
[4]Department Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN
[5]Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York
[6]Division of Pulmonary, Allergy, Critical Care, and Sleep Medicine, Emory University, Atlanta, GA
[7]Department of Cell Biology, Emory University, Atlanta, GA
[8]Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN
[9]Department of Veterans Affairs Medical Center, Nashville, TN
[10]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN
[11]Translational Genomics Research Institute, Phoenix, AZ

*Equal contribution

Address correspondence to:

Jennifer M. S. Sucre (jennifer.sucre@vumc.org) or
Jonathan A. Kropski (jon.kropski@vumc.org)

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Summary

The emergence of the SARS-CoV-2 novel coronavirus has led to a global pandemic (COVID-19), with more than 5 million cases as of May 2020[1]. Available data suggest that severe illness and death from COVID-19 are rare in the pediatric population[2]. Integrating single-cell RNA sequencing of the developing mouse lung with temporally-resolved RNA-in-situ hybridization (ISH) in mouse and human lung tissue, we found expression of SARS-CoV-2 Spike protein primer *TMPRSS2* was highest in ciliated cells and type I alveolar epithelial cells (AT1) and increased with aging in mice and humans. SARS-CoV-2 RNA colocalized with *TMPRSS2* mRNA in lung cells from a patient who died of SARS-CoV-2. Together, these data suggest developmental regulation of *TMPRSS2* may underlie the relative protection of infants and children from severe respiratory illness.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

One aspect of the COVID-19 pandemic that has eluded explanation is the striking diversity of clinical phenotypes accompanying SARS-CoV-2 infection, ranging from asymptomatic carriage to life threatening multi-organ failure[3–8]. Morbidity and mortality appear most severe among the elderly[9,10], while infection rates and hospitalizations among infants and children are substantially lower[1]. In a recently published study from China, 90% of children infected with SARS-CoV-2 exhibited mild symptoms or were asymptomatic[2]. The Human Cell Atlas Lung Biological Network recently reported an integrated analysis of single-cell transcriptomic data from multiple organs and datasets which demonstrated age, sex and smoking-associations of SARS-CoV-2 entry-related genes in several cell types, although this dataset was weighted heavily toward adult samples[11]. This led us to hypothesize that factors determining SARS-CoV-2 cellular infectivity in the respiratory epithelium may be developmentally regulated.

To investigate how SARS-CoV-2 susceptibility changes during lung development, we analyzed a previously unpublished scRNA-seq dataset profiling the epithelial and stromal cells in the developing mouse lung at five timepoints ranging from embryonic day 18 (E18) to postnatal day 64 (P64) (Fig. 1a-b, Fig. S1a-b). We interrogated expression profiles of genes linked to SARS-CoV-2 infectivity by analyzing a total of 67,850 cells across these 5 timepoints (Fig. 1c-d). Previous work has suggested that SARS-CoV-2 gains cellular entry by binding ACE2 on the cell surface[12–14], then the spike-protein undergoes a protease-mediated cleavage facilitating fusion with the cell membrane [15]. TMPRSS2 is the canonical protease mediating cellular entry for coronaviruses including SARS-CoV-2[15], although it should be noted that there are reports the SARS-CoV-2 spike protein may be cleaved by other proteases[14]. Consistent with recent reports analyzing single-cell transcriptomic data of the lung and other organs [11,16–19], we observed that during lung development, expression of *Ace2* was generally low, was largely restricted to epithelial cells, was most frequently and most highly expressed in secretory cells (Fig. 1e), and was expressed in a subset of AT2 cells. In contrast, *Tmprss2* was expressed broadly in the epithelium and was most highly expressed in ciliated cells and AT1 cells (Fig. 1e). A small proportion of fibroblasts and pericytes expressed *Ace2*, but there was minimal *Ace2* or *Tmprss2* in endothelial or other stromal cells (Fig. 1e). Examining the relative expression of *Tmprss2* and other putative priming proteases across developmental time (including furin and cathepsin B), we observed that specifically in ciliated airway epithelial cells, *Tmprss2* expression was significantly higher at P64 compared to all earlier developmental timepoints ($p=0.028$); similar pattern was observed for *Ctsb*. In AT1 cells, *Tmprss2* expression generally increased during alveolarization into adulthood.

To provide tissue validation and to spatially and temporally localize expression of *Tmprss2,* we performed RNA-ISH using formalin fixed paraffin-embedded (FFPE) lung tissue from the same developmental time points as well as lung tissue from 1yr and 2yr old mice. Colocalizing *Tmprss2* with *Scgb1a1* (secretory cells*), Foxj1* (ciliated cells), *Sftpc* (AT2 cells), and *Hopx* (AT1 cells) (Fig. 2a), we observed an age-dependent progressive increase the proportion of *Tmprss2*+ cells among all cell types (Fig. 2b). Further, in AT1 cells and ciliated cells, there was a marked increase in relative *Tmprss2* expression across developmental time that was most striking at 1 and 2 years of age (Fig. 2b). *Tmprss2* exhibited relatively low expression in secretory cells although levels increased in adult and aged mice. AT2 cells showed even lower

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

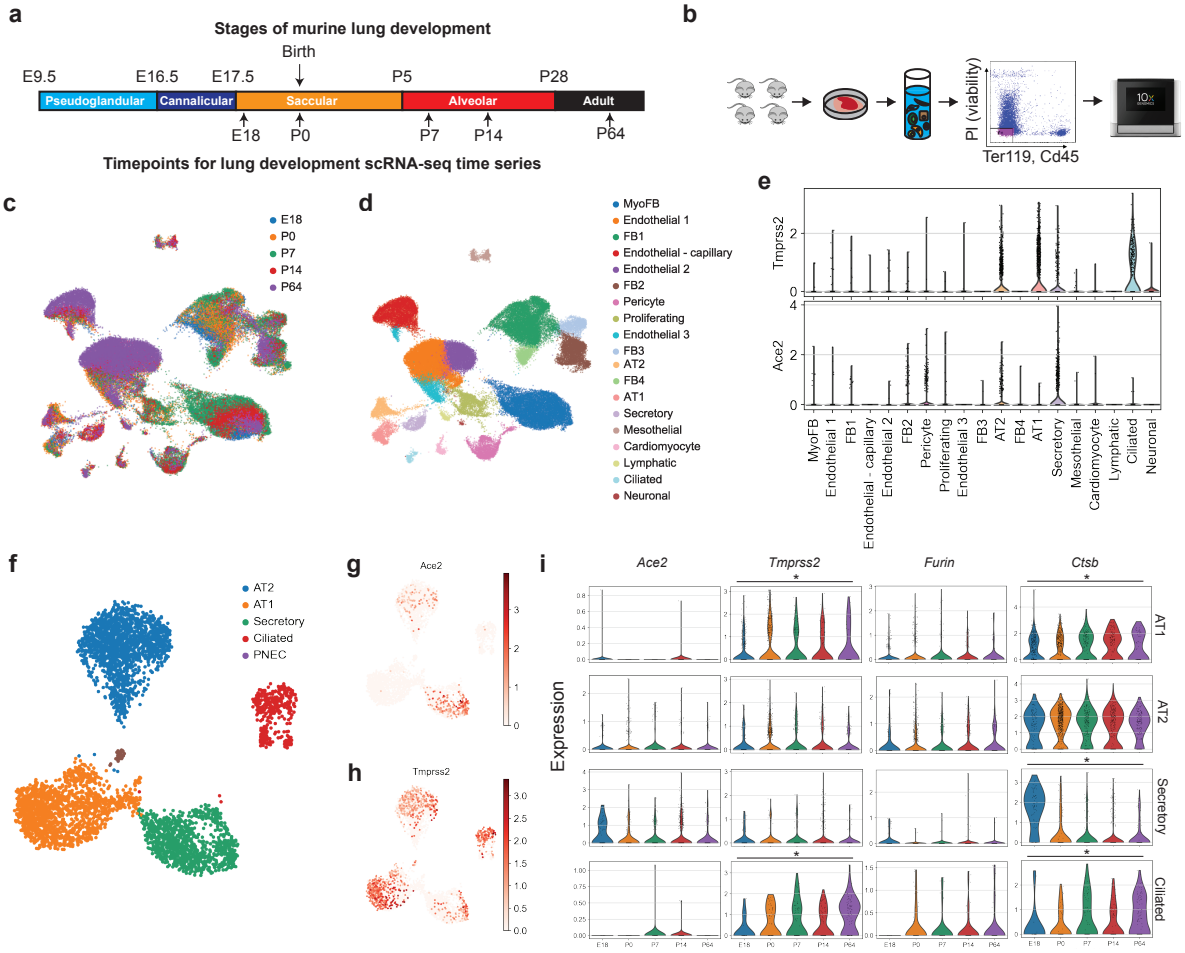

Figure 1. Time-series scRNA-seq of the developing mouse lung. a) Overview of mouse lung developmental stages and timepoints sampled for scRNA-seq time-series. b) Workflow of scRNA-seq time series. Single cell suspensions were generated from at least 4 mice at each timepoint. Viable, Cd45-, Ter119- cells were selected and underwent scRNA-seq library preparation using the 10X Genomics Chromium 5' platform. c-d) UMAP embedding of 67,850 cells annotated by c) developmental timepoint and d) cell-type. e) Violin plot depicting expression of key SARS-CoV2 receptor *Ace2* and coreceptor *Tmprss2* across cell types in the jointly analyzed dataset. f) UMAP embedding of 4,427 epithelial cells after subsetting and reclustering. g,h) UMAP embeddings depicting relative expression of g) *Ace2* and h) *Tmprss2* in epithelial cells. i) Relative expression of putative SARS-CoV-2 priming proteases across developmental time. *p<0.05 by ANOVA across developmental timepoints after Bonferroni correction for multiple testing.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

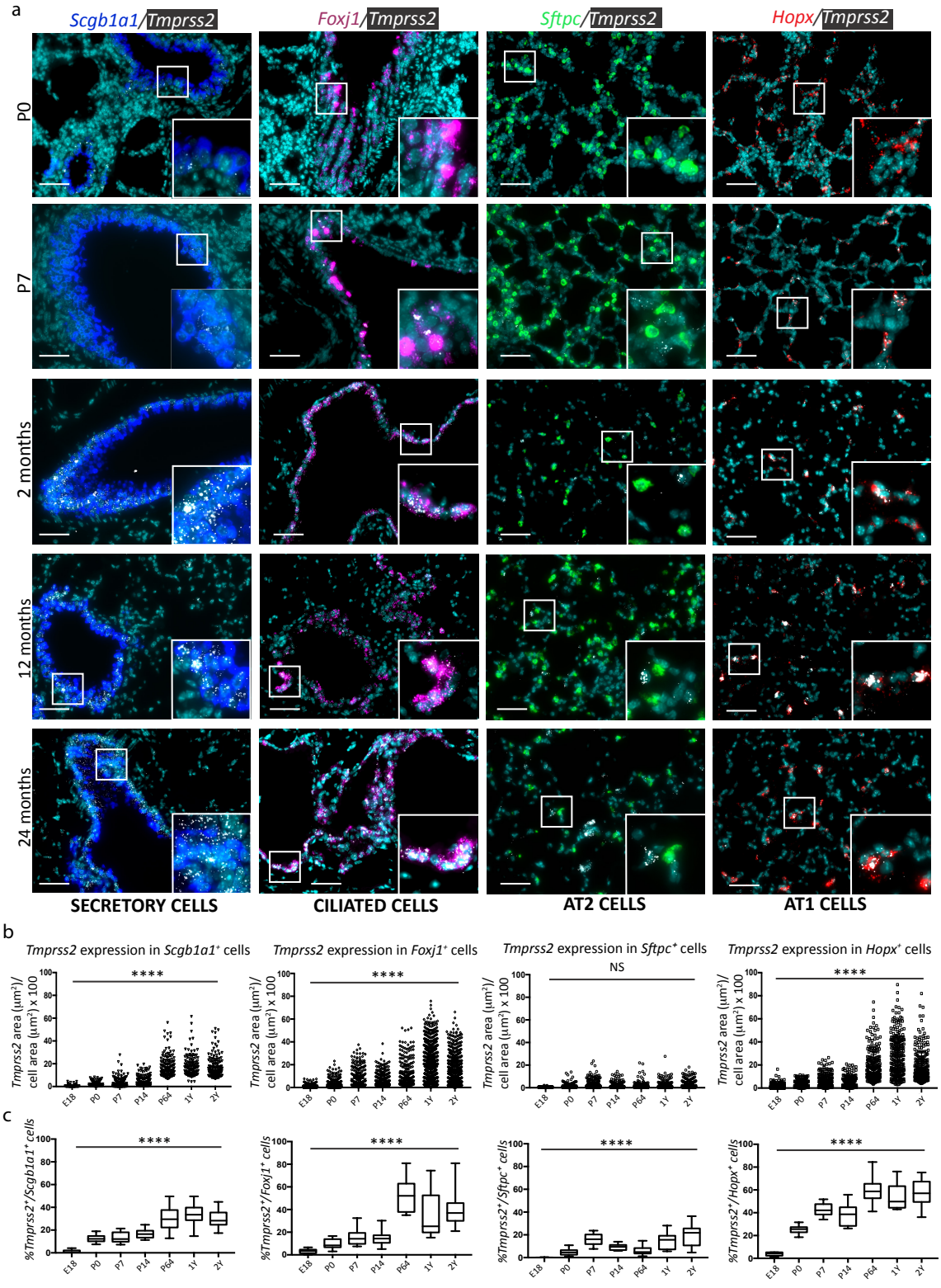

Figure 2. Spatial and temporal localization of *Tmprss2* expression across lung development. A) RNA in situ hybridization (ISH) of *Tmprss2* expression (white) with epithelial cell markers *Scgb1a1* (secretory cells, blue), *Foxj1* (ciliated cells, magenta), *Sftpc* (surfactant protein C, AT2 cells, green), *Hopx* (type 1 alveolar epithelial cells,

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

red). Formalin fixed paraffin embedded tissue from lungs at timepoints E18, P0, P7, P14, P64, 12 months, and 24 months was used, with image data from timepoints P0, P7, P64, 12 months, and 24 months shown in the figure. Lungs from 3 mice at each time point were used, with ten 40X images obtained per slide. Scale bar = 100um. B, C) Quantification of *Tmprss2* expression in each epithelial subtype across development measured as: B) a percentage of cellular area covered by *Tmprss2* probe for each cell expressing both *Tmprss2* and the epithelial cell marker and C) percentage of cells expressing the epithelial cell marker that also express *Tmprss2,* with positive *Tmprss2* expression defined has having 5 or more copies of *Tmprss2* probe. Greater than 1000 cells were counted at each timepoint. ****$p<0.0001$ by one-way ANOVA.

expression of *Tmprss2* when compared with other epithelial cells, without a detectable increase in expression across lung development . Notably, there was minimal detection of *Tmprss2* prenatally at E18, and relatively low levels in the saccular stage at P0 in all epithelial cell types. Consistent with our transcriptomic data, basal *Ace2* expression detected by RNA-ISH was low with little change during development and aging (Fig. S2).

To determine whether *TMPRSS2* is also developmentally regulated in humans, we used a similar approach to examine *TMPRSS2* expression by RNA-ISH during human lung development. We defined infants as individuals up to 2 years of age (n=7), children between the ages of 3 and 17 (n=9), and adult specimens were from subjects aged 54-69 (never smokers, n=4). Infants expressed *TMPRSS2* at very low levels across all four epithelial lineages evaluated, while children exhibited similarly low levels of *TMPRSS2* in secretory and alveolar epithelial cells with a significant increase in *FOXJ1*+ ciliated cells (Fig. 3a-b). Adult subjects had higher *TMPRSS2* expression in secretory, ciliated, and AT1 cells relative to both pediatric groups with very little *TMPRSS2* expression in AT2 cells (Fig. 3b). Human lung tissue demonstrated low levels of *ACE2* expression across in infants, children, and adult samples (Fig S3). These data are broadly consistent with results of a single-nucleus RNA-seq (snRNA-seq) study of 3 infants, children and adults up to age 30 that was reported while this manuscript was being finalized[20].

We next determined the cellular localization of SARS-CoV-2 by analyzing autopsy specimens from a SARS-CoV-2 patient who died 12 days after admission to the hospital with severe respiratory symptoms. Using RNA-ISH, we identified the presence of SARS-CoV-2 in both the large airway and lung parenchyma. SARS-CoV-2 was localized to epithelial cells expressing *SCGB1A1* (secretory), *FOXJ1* (ciliated), and *AGER* (AT1). Surprisingly, despite analyzing > 100 *SFTPC* positive cells (corresponding to AT2), none contained detectable SARS-CoV-2 RNA by this assay (Fig 4a-b). In large airways (mainstem bronchus), SARS-CoV-2 co-localized with *TMPRSS2 in SCGB1A1*- (presumably ciliated) cells (Fig 4c-d). The surprising absence of SARS-CoV-2 in AT2 cells could be explained by changes in viral replication during the course of the illness and/or increased cytotoxicity of infected AT2 cells. These data indicate that secretory, ciliated, and AT1 cells had the highest levels of *TMPRSS2* expression and that these cell types all harbored SARS-CoV-2 RNA in lung tissue from this patient, similar to data observed in an *ex-vivo* infection model[21]. Together, these data suggest that priming protease expression may be a crucial determinant of SARS-CoV-2 infectivity in the lower respiratory tract.

The epithelial cell tropism of SARS-CoV-2 is dependent upon viral attachment and activation by cell-surface proteins, e.g. ACE2 and TMPRSS2. While ACE2 has been well studied in *in vitro* platforms of differentiated lung epithelium, our findings and prior studies [11,16,17] suggest that ACE2 is expressed at low levels in respiratory epithelium (Figure 1, Supplemental Figure 2). Conversely, the rising expression of *TMPRSS2* with age in both the conducting airways and alveolar epithelium may explain the increased susceptibility to

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

symptomatic infection and severe ARDS observed in adults relative to children. To date, there have been very few cases of vertical transmission of SARS-CoV-2 to newborns[22], despite viral particles being identified in the placentas of women who were infected with SARS-CoV-2[23]. In light of the very low expression of *ACE2* in ciliated epithelial cells, our data raise the possibility

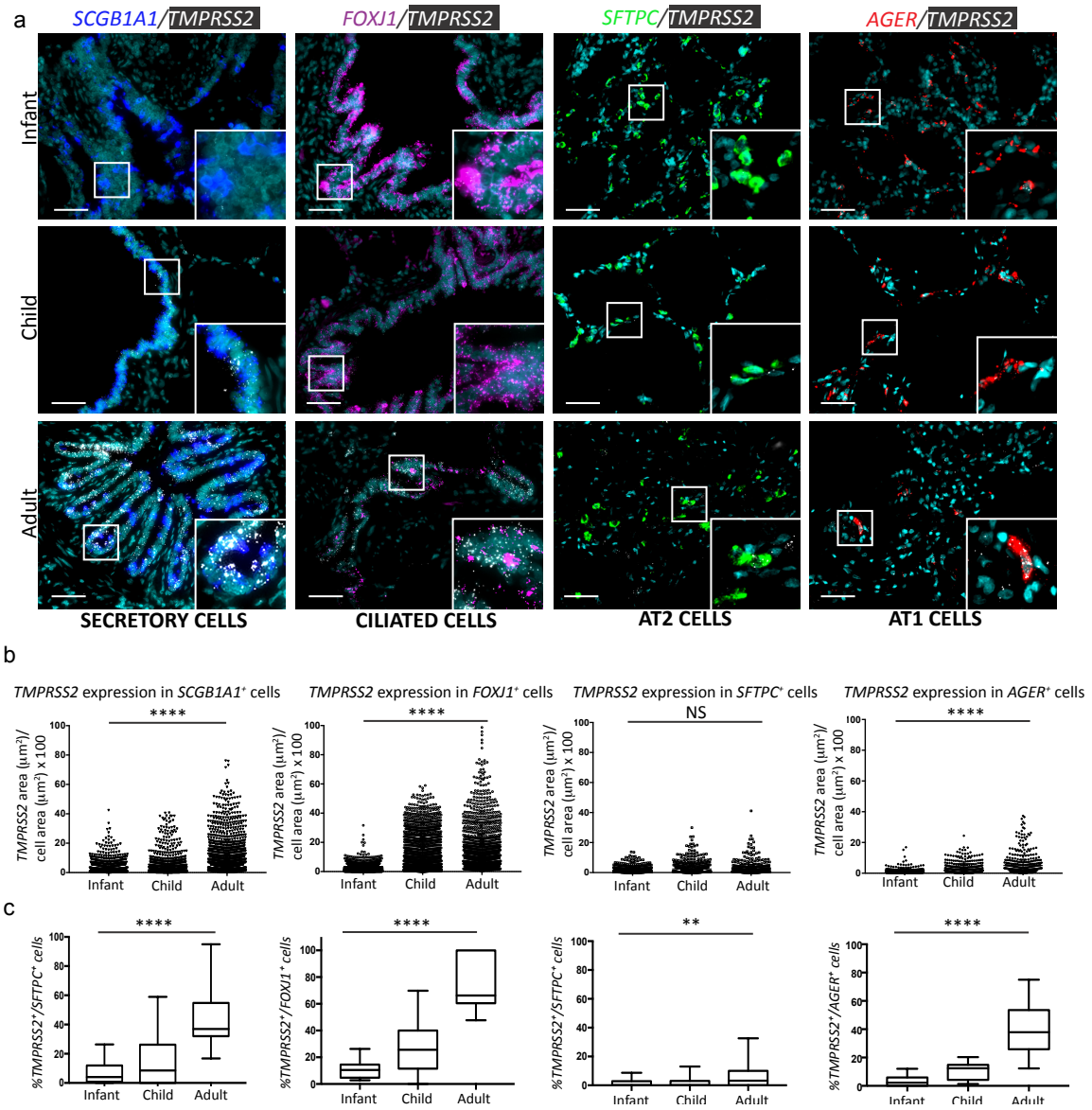

Figure 3. Spatial and temporal localization of *TMPRSS2* expression in the lung across the human lifespan. A) RNA *in-situ* hybridization (ISH) of *TMPRSS2* expression (white) with epithelial cell markers *SCGB1A1* (secretory cells, blue), *FOXJ1* (ciliated cells, magenta), *SFTPC* (surfactant protein C, AT2 cells, green), *AGER* (AT1 cells, red). Formalin fixed paraffin embedded tissue from 25 human lungs between the ages of birth and 69 years was analyzed. We defined infants as birth-2 years, children 3years-17 years, and adult specimens were between 53-69 years of age. Ten 40X images obtained per slide for analysis. Scale bar = 100 µm. B, C) Quantification of *TMPRSS2* expression in each epithelial subtype across development as measured as: B) a percentage of cellular area covered by *TMPRSS2* probe for each cell expressing both *TMPRSS2* and the epithelial cell marker and C) percentage of cells expressing the epithelial cell marker that also express *TMPRSS2,* with positive *TMPRSS2* expression defined has having 5 or more copies of *TMPRSS2* probe. Greater than 1000 cells were counted at each timepoint. ****p<0.0001, **p<0.01 by one-way ANOVA.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that ACE2-independent attachment may play a role in SARS-CoV-2 cellular entry and pathogenesis. CD147 (basigin, *BSG*) has also been proposed as an alternative SARS-CoV-2 receptor[24] and is expressed broadly including in ciliated cells in which its expression increased across developmental time; in contrast, Bsg expression decreased in other epithelial cell types (Fig. S4a-b). Alternatively, recent data indicate that ACE2 may be transiently induced by interferon signaling to facilitate cellular entry[17]. Across development time in the lung epithelium, we observed a progressively increasing interferon signature with increasing age, most prominent in AT2 and AT1 cells (Fig. S5a-b). This raises the possibility that the developing lung may have relatively less interferon "priming" and therefore have less efficient ACE2 induction to serve as the primary entry receptor to facilitate alveolar infectivity. In contrast, we did not observe developmental differences in basal expression of innate inflammatory genes linked to coronaviruses response including: *Irf3, Tnf, Mapk1/3, Nfkb1/2, and Cxcl15* [25,17,26,27] (Fig. S5c).

The very low levels of *TMPRSS2* expression in human infants, and nearly absent levels of expression of *Tmprss2* in prenatal mice, suggest a mechanism by which neonates may be protected against severe forms of COVID-19. After the SARS outbreak in 2003, several groups[28–31] reported that proteolytic cleavage of the SARS-CoV spike (S) protein is a prerequisite for viral activation and host cell entry, with SARS-CoV tropism for TMPRSS2-expressing cells in primates[32]. In addition, TMPRSS2 expression has been linked to the infectivity of other respiratory viruses, with TMPRSS2-knockout mice demonstrating resistance to influenza[33]. In humans, variants in TMPRSS2 that result in increased expression were associated with increased susceptibility to influenza infection during the H1N1 epidemic I n2009[34]. Prior work studying SARS-CoV has shown that inhibition of TMPRSS2 with the serine protease inhibitor Camostat, partially prevented SARS-CoV infection[35], and addition of a cathepsin inhibitor (Aloxistatin) to camostat potentiated the antiviral effect[35].

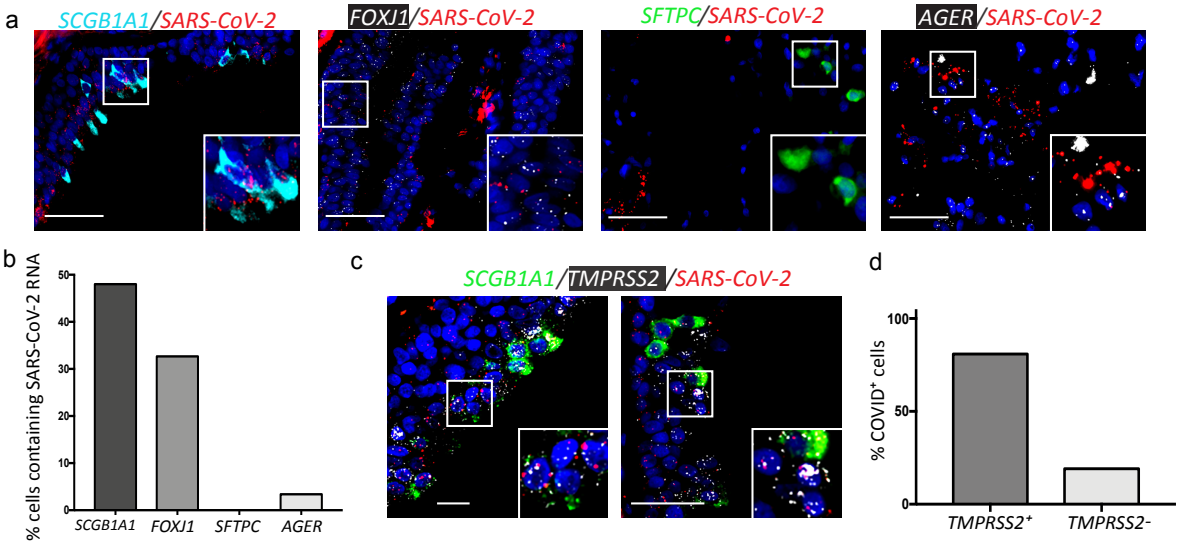

Figure 4. Spatial localization of SARS-CoV-2 RNA in the lungs of a patient who died from complications of COVID-19. A) RNA-ISH of SARS-CoV-2 RNA (red) with epithelial cell markers *SCGB1A1 (*secretory cells, cyan), *FOXJ1* (ciliated cells, white), *SFTPC* (AT2 cells, green), *AGER* (tAT1 cells, white); scale bar = 100µm. B) Quantification of cells containing SARS-CoV-2 RNA by epithelial subtype, as a percentage of total number of cells counted; a minimum of 100 cells were counted in each group. C)RNA ISH of large airway from same patient demonstrating RNA transcripts for *TMPRSS2* (white) in the same cells containing SARS-CoV-2 RNA (red), with secretory cells labeled in green (*SCGB1A1)* for context; scale bar = 100um. D) Quantification of *TMPRSS2* expression in *COVID+* cells, a total of 467 cells were counted from the large airway of the patient.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In summary, these data suggest that developmental regulation of viral entry-factors may be the primary determinant of the age-related differences in SARS-CoV-2 susceptibility and severity. The identification of changes in *TMPRSS2* expression associated with age presents a biological rationale for the observed rarity of severe lower respiratory tract SARS-CoV-2 disease in children, and underscores the opportunity to consider *TMPRSS2* inhibition as a potential therapeutic target for SARS-CoV-2.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

METHODS

Animal Care and Tissue Fixation: C57BL/6 mice were used for all experiments. Timed matings were performed as previously described[36] and mice were sacrificed at P0, P7, P14, or P64 for single cell RNA sequencing or lung block fixation. E18 lungs were isolated by removing pups from the mouse uterus and isolating lung tissue. E18 and P0 lungs were fixed in formalin. P7, P14, P64, 12 month, and 24 month old mouse lungs were inflation-fixed by gravity filling with 10% buffered formalin and paraffin embedded. This protocol was approved by the Institutional Animal Care and Use Committee of Vanderbilt University (Nashville, TN) and was in compliance with the Public Health Services policy on humane care and use of laboratory animals.

Single cell isolation and flow cytometry: At the indicated timepoints, lung lobes were harvested, minced, and incubated for 30 minutes at 37°C in dissociation media (RPMI-1640 with 0.7 mg/ml collagenase XI and 30 mg/ml type IV bovine pancreatic DNase). After incubation, lobes were passed through a wide bore pipet tip and filter through a 40 mm filter. Single cell lung suspension was then counted, aliquoted, and blocked with CD-32 Fc block (BD cat #553142) for 20 minutes on ice. After 2% FBS staining buffer wash, cells were incubated with the conjugated primary antibodies anti-CD45 (BD cat # 559864) and anti-Ter119 (BD cat# 116211) as indicated below. In the same manner, fluorescence minus one controls were blocked and stained with the appropriate antibody controls. Cells from individual mice were then incubated with identifiable hashtags, resuspended in staining buffer, and treated with PI viability dye. CD45 negative, Ter119 negative, viable cells were collected by fluorescence associated cell sorting using a 70 mm nozzle on a 4-laser FACSAria III Cell Sorter. Both single and fluorescence-minus-one controls were used for compensation.

scRNA-seq library preparation and next-generation sequencing: ScRNA-seq libraries were generated using the 10X Chromium platform 5' library preparation kits (10X Genomics) following the manufacturer's recommendations and targeting 10,000 - 20,000 cells per sample. Next generation sequencing was performed on an Illumina Novaseq 6000. Reads with read quality less than 30 were filtered out and CellRanger Count v3.1 (10X Genomics) was used to align reads onto mm10 reference genome.

Ambient RNA filtering : Ambient background RNA were cleaned from the scRNA-seq data with "SoupX" (version 1.2.2, Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK) ; [37]) in RStudio (version 1.2.5001, RStudio, Inc., Boston, Massachusetts, USA). Matrix files from CellRanger were read into the global environment and combined into a "SoupChannel." Data from the SoupChannel were passed to Seurat (version 3.1.4, New York Genome Center, New York City, New York, USA)[38,39] for data normalization, identification of variable features, data scaling, principal component analysis, uniform manifold approximation and projection (UMAP) for dimensionality reduction, calculation of nearest neighbors, and cluster identification. The SoupX pipeline was used for each time point to determine which genes were most likely contributing to the ambient background RNA. We utilized the following genes to estimate the non-expressing cells, calculate the contamination fraction, and adjust the gene expression counts: *Dcn, Bgn, Aspn, Ecm2, Fos, Hbb-bs, Hbb-bt, Hba-a1, Hba-a2, Lyz1, Lyz2, Mgp, Postn, Scgb1a1*. For time points that were sorted to select for epithelial cells, the following genes were

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

added to the SoupX pipeline: *Sftpc, Hopx, Ager, Krt19, Cldn4, Foxj1, Krt5, Sfn, Pecam1*. New matrix files were created by SoupX and used for subsequent analyses in Scanpy.

Data integration and clustering. Data integration and clustering was performed using a standard Scanpy workflow (Scanpy v1.46)[40]. Individual SoupX "cleaned" libraries were concatenated and analyzed jointly. After quality filtering (removal of cells with <500 or >5000 genes, or >10% mitochondrial gene expression), data were normalized, log-transformed, highly-variable genes were identified, and percent mitochondrial gene expression and and read depth were regressed. Following data scaling, and principal components analysis, Leiden clustering was performed followed by visualization of canonical marker genes. Red blood cells and doublet clusters (containing non-physiologic marker combinations, i.e. *Epcam*+/*Pecam1*+) were filtered. Batch correction for dataset integration was performed using batch-balanced K-nearest neighbors [41]. Principal components 1:30 were used for clustering and Uniform Manifold Approximation and Projection (UMAP) embedding[42]. Leiden clustering (resolution 0.8) was then performed on the integrated dataset and cell-types assigned based on marker expression profiles.

Subjects and samples: Lung tissue from 20 healthy human subjects was obtained at the time of surgical biopsy or autopsy, with death occurring from non-respiratory causes. Subjects were classified as infants (between birth and age 2 years), children (3-17 years), and adult (54-69 years old). Adult tissue included male and female donors, and all were lifetime nonsmokers. SARS-CoV-2 lung tissue was obtained at the time of autopsy from a patient who died from hypoxic respiratory failure on day 12 of illness (nonsmoking female, age >85). All studies were approved by the Vanderbilt Institutional Review Board (VanderbiltIRB #'s 060165, 171657, 200490)

RNA *in situ* hybridization: RNAScope technology (ACDBio) was used to perform all RNA *in situ* hybridization (RNA ISH) experiments(24) according to manufacturer's instructions. RNAScope probes to the following human genes were used: *TMPRSS2, ACE2, SCGB1A1, FOXJ1, SFTPC,* and *AGER*. Probes to the following mouse genes were used: *Tmprss2, Ace2, Scgb1a1, Foxj1, Sftpc, Hopx.* RNAScope probe to the SARS-CoV-2 virus (sense) was also used. Our validation of the SARS-CoV-2 probe included application of this probe to lung tissue from healthy adult lungs from patients who died in 2018, which showed no amplification or fluorescence (Figure S6).

Image acquisition and analysis: Fluorescent images were acquired using a Keyence BZ-X710 with BZ-X Viewer software with 40X objective. Lasers utilized for excitation included 405, 488, 561, and 647nm lines. Automated image analysis was performed with Halo software; an example of cellular segmentation and determination of co-localization is demonstrated in Figure S7. Cell area of *Tmprss2/TMPRSS2* probe was calculated as a percentage of total cell area for each epithelial subtype. Number of *Tmprss2/TMPRSS2* positive cells (defined as having 5 or more copies of *Tmprss2/TMPRSS2* by Halo analysis) as a percentage of total epithelial cells was also calculated. Presence of SARS-CoV-2 in epithelial cells was determined by Halo segmentation and analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data availability. Raw data and counts matrices used in generating this dataset are available through the Gene Expression Omnibus (GEO) (pending). Code used for dataset integration and analyses in this manuscript are available at https://github.com/KropskiLab.

Statistical approach: Statistical tests used for specific comparisons are described in the relevant figure legends. Comparisons of gene expression from scRNA-seq were performed by analysis of variance (anova) using the linear model (lm) function in R version 3.6.3.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

This work was supported by NIH K08143051 (JMS), K08HL130595 (JAK), R01HL145372(JAK/NEB), P01HL092470(TSB), K08HL127102(EJP), K08HL133484(JTB), R01AI077505 (DWH), P30AI110527 (SAM), R01AI142095(SAK/SAM), TR002243. Flow Cytometry experiments were performed in the VMC Flow Cytometry Shared Resource. The VMC Flow Cytometry Shared Resource is supported by the Vanderbilt Ingram Cancer Center (P30 CA68485) and the Vanderbilt Digestive Disease Research Center(DK058404).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Consortia

Vanderbilt COVID-19 Cohort Consortium (VC3):

Justin L. Balko, Suman Das, David Haas, Spyros A. Kalams, Jonathan A. Kropski, Christine Lovly, Simon A. Mallal, Elizabeth J. Phillips, Wei Zheng

Human Cell Atlas Lung Biological Network

The members of HCA Lung Biological Network are Nicholas E. Banovich, Pascal Barbry, Alvis Brazma, Tushar Desai, Thu Elizabeth Duong, Oliver Eickelberg, Christine Falk, Michael Farzan, Ian Glass, Muzlifah Haniffa, Peter Horvath, Deborah Hung, Naftali Kaminski, Mark Krasnow, Jonathan A. Kropski, Malte Kuhnemund, Robert Lafyatis, Haeock Lee, Sylvie Leroy, Sten Linnarson, Joakim Lundeberg, Kerstin B. Meyer, Alexander Misharin, Martijn Nawijn, Marko Z. Nikolic, Jose Ordovas-Montanes, Dana Pe'er, Joseph Powell, Stephen Quake, Jay Rajagopal, Purushothama Rao Tata, Emma L. Rawlins, Aviv Regev, Paul A. Reyfman, Mauricio Rojas, Orit Rosen, Kourosh Saeb-Parsy, Christos Samakovlis, Herbert Schiller, Joachim L. Schultze, Max A. Seibold, Alex K. Shalek, Douglas Shepherd, Jason Spence, Avrum Spira, Xin Sun, Sarah Teichmann, Fabian Theis, Alexander Tsankov, Maarten van den Berge, Michael von Papen, Jeffrey Whitsett, Ramnik Xavier, Yan Xu, Laure-Emmanuelle Zaragosi, and Kun Zhang. Pascal Barbry, Alexander Misharin, Martijn Nawijn, and Jay Rajagopal serve as the coordinators.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author Contributions

JMS and JAK conceived study. EJP, JB, CJT, CJ, PG, DSN performed experiments. BAS, ACH, NEB, JMS, JAK analyzed transcriptomic data. MK, SHG, VC3, TSB provided samples, BAS, ACH, JAK and JMS interpreted data with assistance from TSB, SHG, SHG, SAW, NEB, LZB, MHK, VC3 and the HCA Lung Biological Network. BAS, LTB, JAK and JMS wrote the manuscript and all authors provided critical feedback on the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Competing Interests

JAK has received advisory board fees from Boehringer Ingelheim, Inc, and has research contracts with Genentech. TSB has received advisory board fees from Boehringer Ingelheim, Inc, Orinove, GRI Bio, Morphic, and Novelstar, and has research contracts with Genentech and Celgene.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Home. *Johns Hopkins Coronavirus Resource Center* https://coronavirus.jhu.edu/.
2. Dong, Y. *et al.* Epidemiology of COVID-19 Among Children in China. *Pediatrics* (2020) doi:10.1542/peds.2020-0702.
3. Grasselli, G. *et al.* Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA* (2020) doi:10.1001/jama.2020.5394.
4. Richardson, S. *et al.* Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* (2020) doi:10.1001/jama.2020.6775.
5. Shi, P. *et al.* Clinical characteristics of imported and second-generation COVID-19 cases outside Wuhan, China: A multicenter retrospective study. *Infectious Diseases (except HIV/AIDS)* (2020) doi:10.1101/2020.04.19.20071472.
6. Guan, W.-J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2002032.
7. Moriarty, L. F. Public Health Responses to COVID-19 Outbreaks on Cruise Ships — Worldwide, February–March 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, (2020).
8. Bai, Y. *et al.* Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* (2020) doi:10.1001/jama.2020.2565.
9. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 — United States, February 12–March 28, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, (2020).
10. Wu, C. *et al.* Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern. Med.* (2020) doi:10.1001/jamainternmed.2020.0994.
11. Muus, C. *et al.* Integrated analyses of single-cell atlases reveal age, gender, and smoking status associations with cell type-specific expression of mediators of SARS-CoV-2 viral entry and highlights inflammatory programs in putative target cells. *bioRxiv* 2020.04.19.049254 (2020) doi:10.1101/2020.04.19.049254.
12. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* **5**, 562–569 (2020).
13. Luan, J., Lu, Y., Jin, X. & Zhang, L. Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochem. Biophys. Res. Commun.* **526**, 165–169 (2020).
14. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281–292.e6 (2020).
15. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* (2020) doi:10.1016/j.cell.2020.02.052.
16. Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* (2020) doi:10.1038/s41591-020-0868-6.
17. Ziegler, C. *et al.* SARS-CoV-2 Receptor ACE2 is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Enriched in Specific Cell Subsets Across Tissues. (2020) doi:10.2139/ssrn.3555145.
18. Zou, X. *et al.* Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. *Front. Med.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

14, 185–192 (2020).
19. Qi, F., Qian, S., Zhang, S. & Zhang, Z. Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. *Biochem. Biophys. Res. Commun.* **526**, 135–140 (2020).
20. Wang, A. *et al.* Single Nucleus Multiomic Profiling Reveals Age-Dynamic Regulation of Host Genes Associated with SARS-CoV-2 Infection. *bioRxiv* 2020.04.12.037580 (2020) doi:10.1101/2020.04.12.037580.
21. Hui, K. P. Y. *et al.* Tropism, replication competence, and innate immune responses of the coronavirus SARS-CoV-2 in human respiratory tract and conjunctiva: an analysis in ex-vivo and in-vitro cultures. *Lancet Respir Med* (2020) doi:10.1016/S2213-2600(20)30193-4.
22. Chen, H. *et al.* Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. *Lancet* **395**, 809–815 (2020).
23. Algarroba, G. N. *et al.* Visualization of SARS-CoV-2 virus invading the human placenta using electron microscopy. *Am. J. Obstet. Gynecol.* (2020) doi:10.1016/j.ajog.2020.05.023.
24. Wang, K. *et al.* SARS-CoV-2 invades host cells via a novel route: CD147-spike protein. *bioRxiv* 2020.03.14.988345 (2020) doi:10.1101/2020.03.14.988345.
25. Huang, I.-C. *et al.* Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. *PLoS Pathog.* **7**, e1001258 (2011).
26. Fung, T. S. & Liu, D. X. Human Coronavirus: Host-Pathogen Interaction. *Annu. Rev. Microbiol.* **73**, 529–557 (2019).
27. Zhao, X. *et al.* Interferon induction of IFITM proteins promotes infection by human coronavirus OC43. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6756–6761 (2014).
28. Simmons, G. *et al.* Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4240–4245 (2004).
29. Simmons, G. *et al.* Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 11876–11881 (2005).
30. Matsuyama, S., Ujike, M., Morikawa, S., Tashiro, M. & Taguchi, F. Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 12543–12547 (2005).
31. Huang, I.-C. *et al.* SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells. *J. Biol. Chem.* **281**, 3198–3203 (2006).
32. Matsuyama, S. *et al.* Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease TMPRSS2. *J. Virol.* **84**, 12658–12664 (2010).
33. Lambertz, R. L. O. *et al.* Tmprss2 knock-out mice are resistant to H10 influenza A virus pathogenesis. *J. Gen. Virol.* **100**, 1073–1078 (2019).
34. Cheng, Z. *et al.* Identification of TMPRSS2 as a Susceptibility Gene for Severe 2009 Pandemic A(H1N1) Influenza and A(H7N9) Influenza. *J. Infect. Dis.* **212**, 1214–1221 (2015).
35. Kawase, M., Shirato, K., van der Hoek, L., Taguchi, F. & Matsuyama, S. Simultaneous treatment of human bronchial epithelial cells with serine and cysteine protease inhibitors prevents severe acute respiratory syndrome coronavirus entry. *J. Virol.* **86**, 6537–6545 (2012).
36. Plosa, E. J. *et al.* Epithelial β1 integrin is required for lung branching morphogenesis and

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

alveolarization. *Development* **141**, 4751–4762 (2014).
37. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data. *bioRxiv* 303727 (2020) doi:10.1101/303727.
38. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
39. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
40. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
41. Polański, K. *et al.* BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).
42. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figures

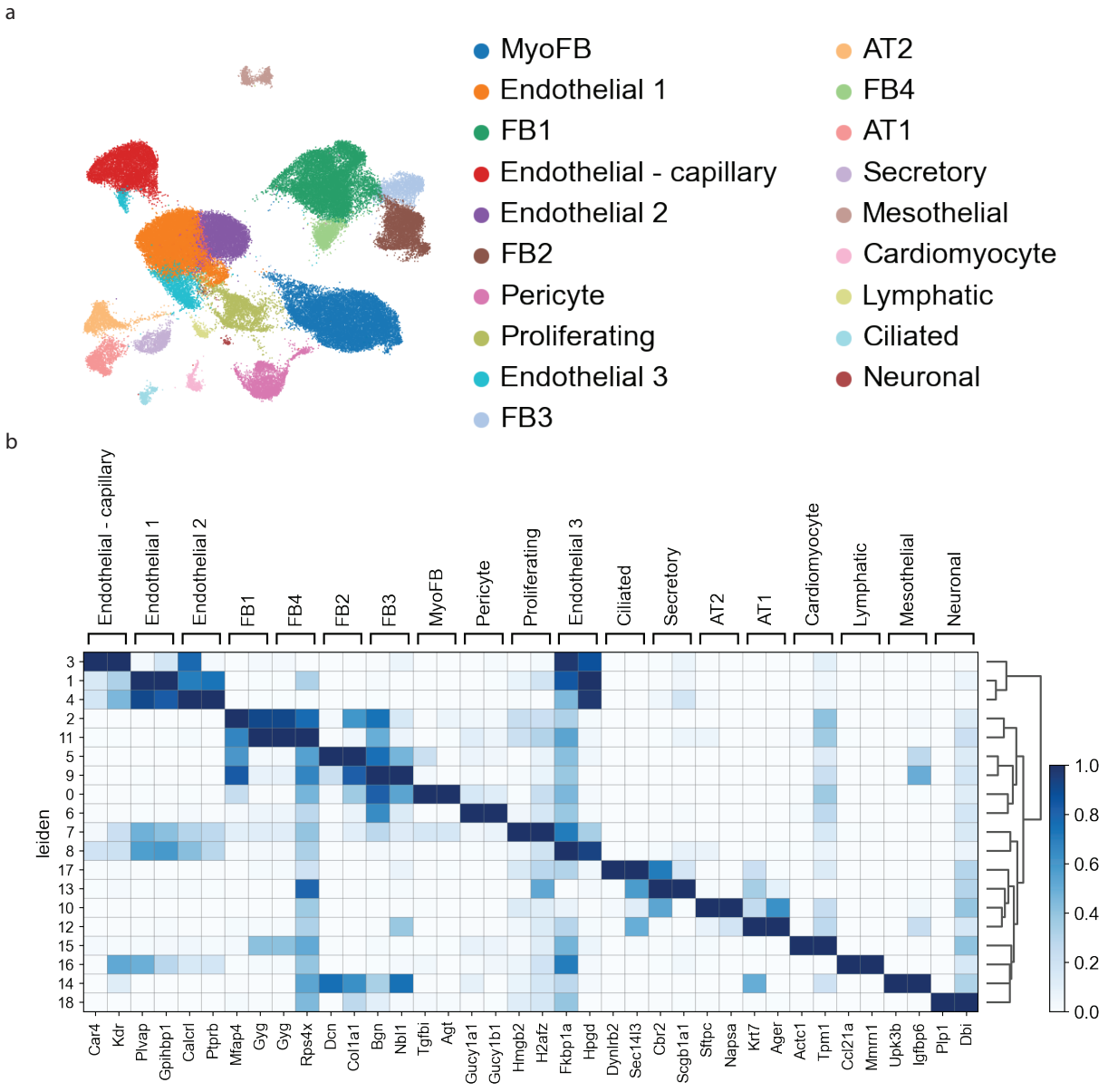

Figure S1. Cell types identified by time-series scRNA-seq. A) Enlarged UMAP embedded of 67,850 cells from across developmental time annotated by cell-type. B) Matrixplot depicting the top 2 markers for each cell-type determined by Wilcoxon-test comparing a given cluster to all others.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

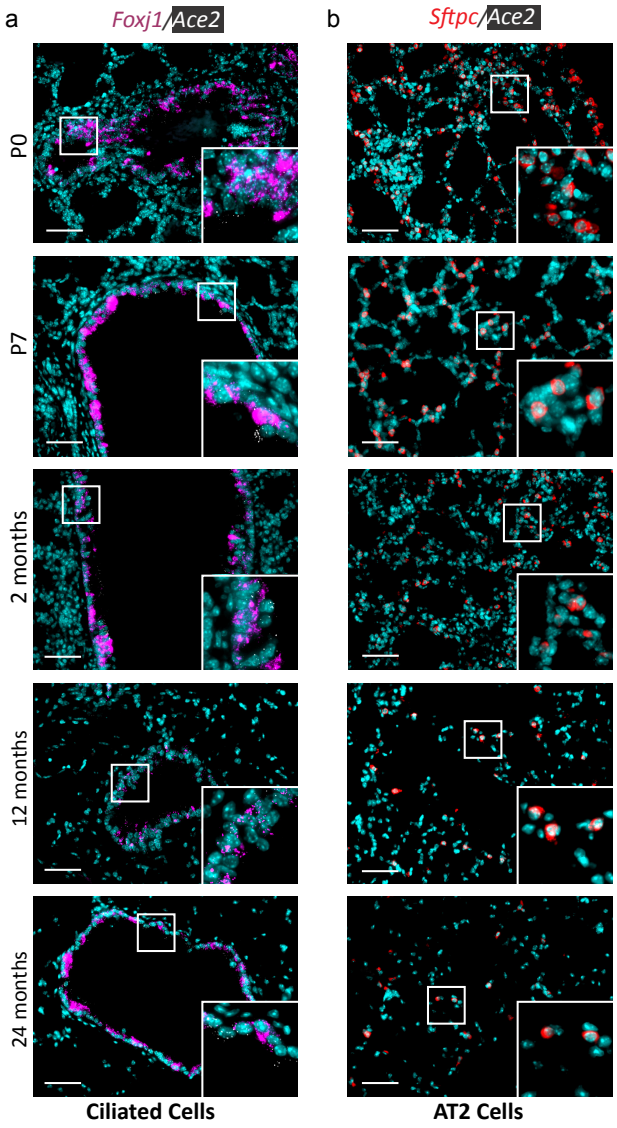

Figure S2: Spatial and temporal localization of *Ace2* expression across lung development. A) RNA in situ hybridization (ISH) of *Ace2* expression (white) with epithelial cell markers *Foxj1* (ciliated cells, magenta) and *Sftpc* (surfactant protein C, AT2 cells, red). Formalin fixed paraffin embedded tissue from lungs at timepoints P0, P3, P5, P7, P14, P64. Lungs from 3 mice at each time point were used, with five 40X images obtained per slide. Scale bar = 100um.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

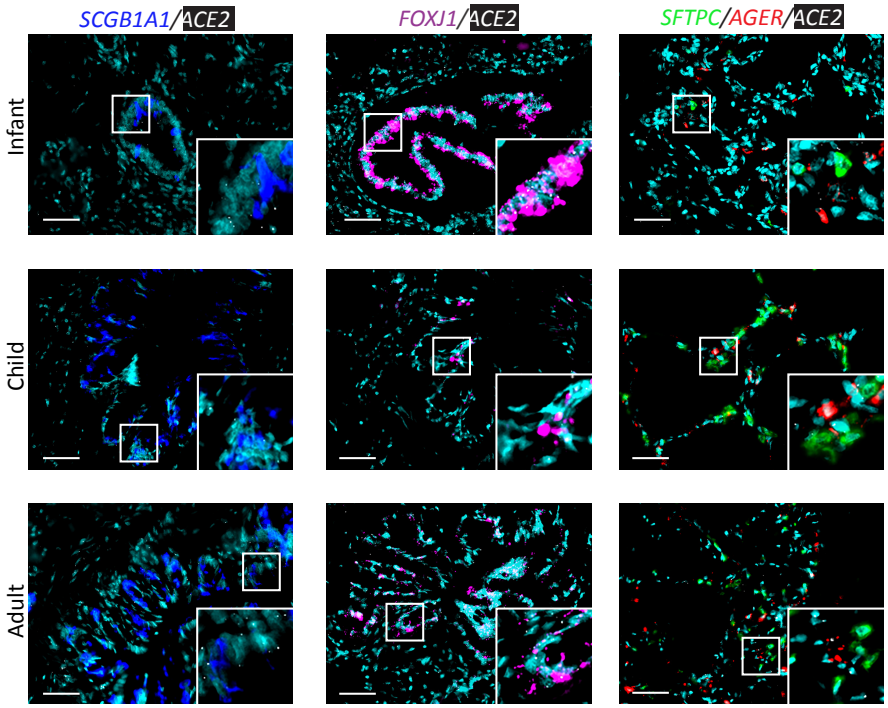

Figure S3: Spatial and temporal localization of *ACE2* expression in the lung across the human lifespan. A) RNA *in situ* hybridization (ISH) of *ACE2* expression (white) with epithelial cell markers *SCGB1A1* (secretory cells, blue), *FOXJ1* (ciliated cells, magenta), *SFTPC* (surfactant protein C, AT2 cells, green), *AGER* (AT1 cells, red). Formalin fixed paraffin embedded tissue from 25 human lungs between the ages of birth and 69 years was analyzed. We defined infants as birth-2 years, children 3years-17 years, and adult specimens were between 53-69 years of age. Five 40X images obtained per slide for analysis. Scale bar = 100 μm.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

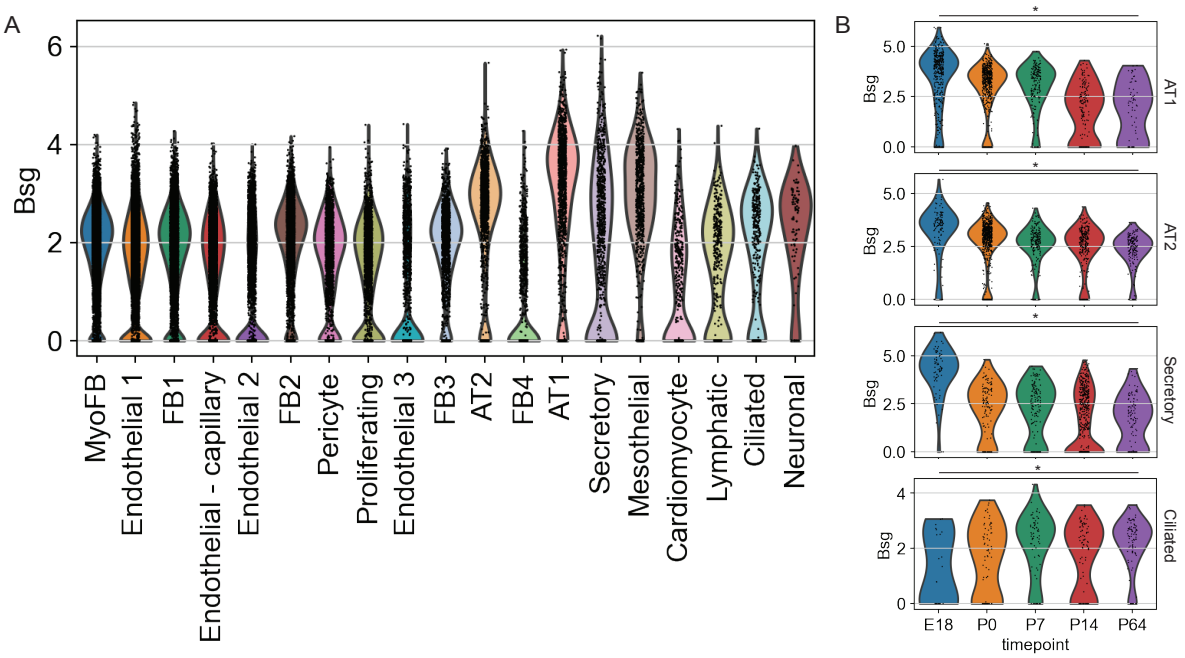

Figure S4. Expression of proposed alternative SARS-CoV-2 receptor CD147 (Bsg) across lung development. A) Violin plot depicting *Bsg* expression across cell types. B) Violin plots depicting *Bsg* expression in epithelial cell types across developmental time. * p <0.05 by ANOVA with Bonferroni correction for multiple testing.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

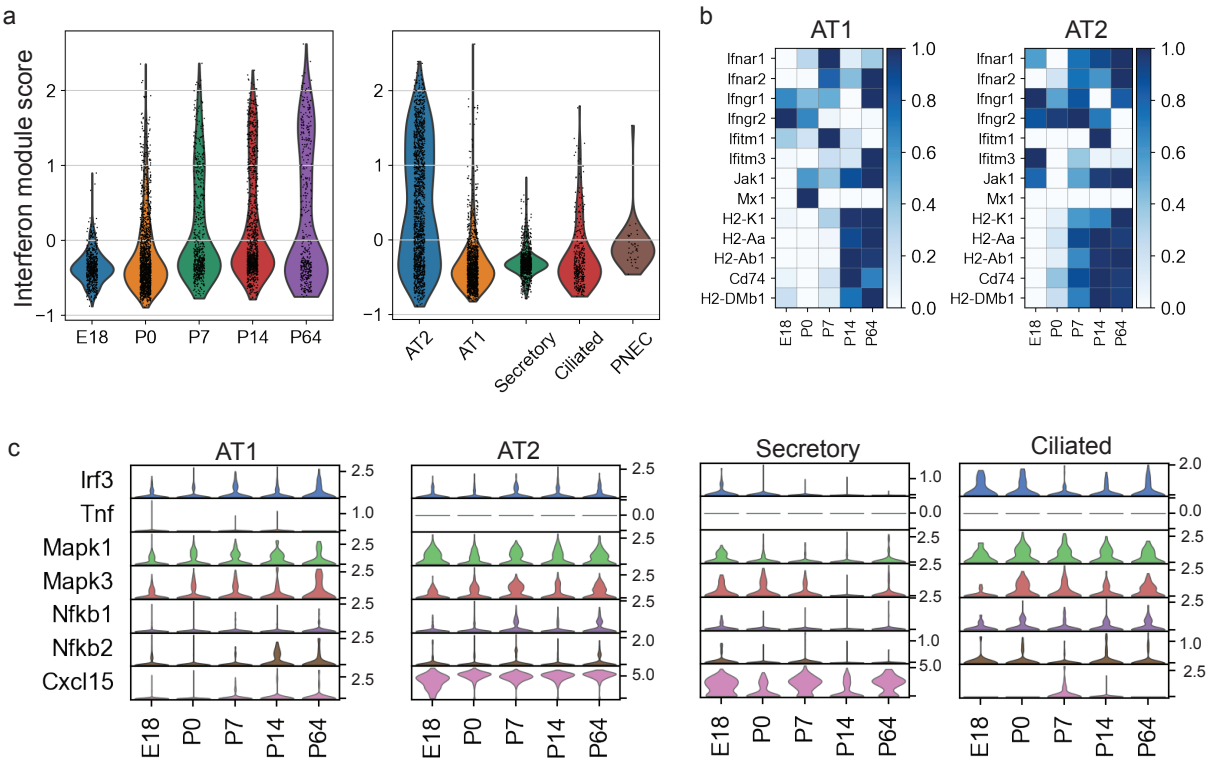

Figure S5. Immune expression programs across developmental time. Interferon gene module score was calculated using the score_genes tool in Scanpy v1.46 using interferon receptors and canonical interferon-response genes plotted A) in all epithelial cells by developmental timepoint and B) by epithelial cell type. C) Heatmap depicting relative expression of interferon-related genes in AT2 and AT1 cells across developmental time. D) Expression of innate immune gene across developmental time.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

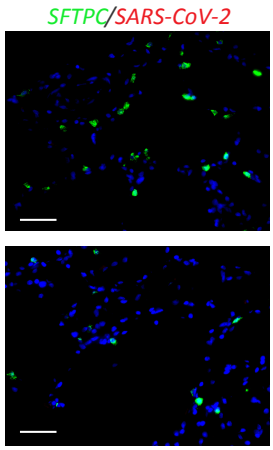

Figure S6: RNA *in situ* hybridization with probes for SARS-CoV-2 RNA (red) and *SFTPC* (green) in lungs from a patient who died in 2018, which demonstrates no amplification or fluorescence of SARS-CoV-2 RNA. Scale bar = 100 μm.

bioRxiv preprint doi: https://doi.org/10.1101/2020.05.22.111187; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

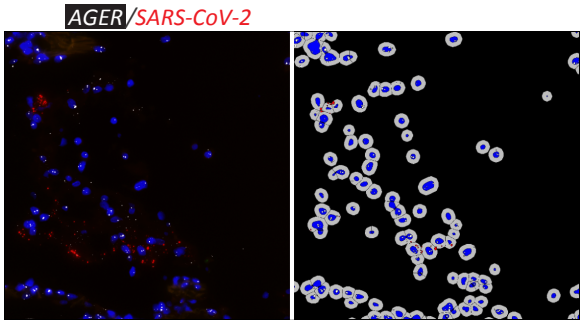

Figure S7: Example of cellular segmentation and automated labeling of epithelial cells as containing SARS-CoV-2 virus, shown here with *AGER* positive cells from lung parenchyma of a patient who died of COVID-19.