Single-cell longitudinal analysis of SARS-CoV-2 infection in human bronchial epithelial cells

Neal G. Ravindra[1,2,*], Mia Madel Alfajaro[3,4,*], Victor Gasque[1,2,5,6,*], Jin Wei[3,4], Renata B. Filler[3,4], Nicholas C. Huston[7], Han Wan[8], Klara Szigeti-Buck[9,10,11], Bao Wang[3,4], Ruth R. Montgomery[12], Stephanie C. Eisenbarth[3,4], Adam Williams[13], Anna Marie Pyle[8,14], Akiko Iwasaki[4,8,14], Tamas L. Horvath[7,8,9], Ellen F. Foxman[3,4], David van Dijk[1,2,†], and Craig B. Wilen[3,4,†]

[1]Cardiovascular Research Center, Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
[2]Department of Computer Science, Yale University, New Haven, CT, USA
[3]Department of Laboratory Medicine, Yale School of Medicine, New Haven, CT, USA
[4]Department of Immunobiology, Yale School of Medicine, New Haven, CT, USA
[5]Université Claude Bernard Lyon 1, Faculté de Médecine Lyon Est, Lyon, France
[6]Département de Bioinformatique, Univ Evry, Université Paris-Saclay, Paris, France
[7]Department of Molecular Biophysics & Biochemistry, Yale School of Medicine, New Haven, CT, USA
[8]Department of Molecular, Cellular, and Developmental Biology, Yale School of Medicine, New Haven, CT, USA
[9]Department of Comparative Medicine, Yale School of Medicine, New Haven, CT, USA
[10]Department of Neuroscience, Yale School of Medicine, New Haven, CT, USA
[11]Department of Obstetrics, Gynecology, and Reproductive Sciences, Yale School of Medicine, New Haven, CT, USA
[12]Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
[13]The Jackson Laboratory, Farmington, CT, USA
[14]Howard Hughes Medical Institute, Chevy Chase, MD, USA

[*]Contributed equally
[†]Correspondence to David van Dijk (david.vandijk@yale.edu) and Craig Wilen (craig.wilen@yale.edu)

May 7, 2020

SARS-CoV-2, the causative agent of COVID-19, has resulted in more than 3,000,000 infections and 200,000 deaths. There are currently no approved drugs or vaccines for the treatment or prevention of COVID-19. Enhanced understanding of SARS-CoV-2 infection and pathogenesis is critical for the development of therapeutics. To reveal insight into viral replication, cell tropism, and host-viral interactions of SARS-CoV-2 we performed single-cell RNA sequencing of experimentally infected human bronchial epithelial cells (HBECs) in air-liquid interface cultures over a time-course. This revealed novel polyadenylated viral transcripts and highlighted ciliated cells as the major target of infection, which we confirmed by electron microscopy. Over the course of infection, cell tropism of SARS-CoV-2 expands to other epithelial cell types including basal and club cells. Infection induces cell intrinsic expression of type I and type III IFNs and IL6 but not IL1. This results in expression of interferon stimulated genes in both infected and bystander cells. Here, we have conducted an in-depth analysis of SARS-CoV-2 infection in HBECs and provide a detailed characterization of genes, cell types, and cell state changes associated with the infection.

1 Introduction

In December 2019, a novel viral pneumonia, now referred to as Coronavirus Disease 2019 (COVID-19), was observed in Wuhan, China [1]. Severe Acute Respiratory Syndrome (SARS)- Coronavirus (CoV)-2, the causative agent of COVID-19, has caused more than $3,000,000$ infections and $200,000$ deaths in 187 countries. There are currently no approved drugs or vaccines for the treatment or prevention of COVID-19. Enhanced understanding of viral pathogenesis at the cellular and molecular level is critical for enhanced prognostic tools and novel therapeutics. Presentation is highly variable ranging from asymptomatic infection to acute respiratory distress syndrome and death [2].

CoVs are enveloped viruses with positive-sense, single-stranded RNA genomes ranging from 26–30 kb [3]. Six human CoVs have been previously identified: HCoV-NL63 and HCoV-229E, which belong to the Alphacoronavirus genus; and HCoV-OC43, HCoV-HKU1, SARS-CoV, and Middle East Respiratory Syndrome CoV (MERS-CoV), which belong to the Betacoronavirus genus [4]. In the past two decades, CoVs have become a major public health concern due to potential zoonotic transmission, as revealed by the emergence of SARS-CoV in 2002, which infected $8,000$ people worldwide with a mortality rate of 10–15%, and MERS-CoV in 2012 and 2019, which infected $2,500$ people with a mortality rate of 35%, and now SARS-CoV-2 (WHO).

Tissue and cell tropism are key determinants of viral pathogenesis. SARS-CoV-2 entry into cells depends on the binding of the viral spike (S) protein to its cognate receptor angiotensin-converting enzyme II (ACE2) on the cell surface [2]. ACE2 is also the receptor for SARS-CoV and HCoV-NL63, yet these viruses induce profoundly different morbidity and mortality suggesting unknown determinants of coronavirus pathogenesis [5, 6]. Additionally, proteolytic priming of the S protein by host proteases is also critical for viral entry [7]. The cellular serine protease Type II transmembrane (TMPRSS2) is used by SARS-CoV-2 for S protein priming [8, 7, 9, 10]. This is also used by SARS-CoV alongside the endosomal cysteine proteases cathepsin B and L [11, 12]. Another host protease, furin, has been suggested to mediate SARS-CoV-2 pathogenesis; however, the precise role of host proteases in SARS-CoV-2 entry remains to be determined [13, 10].

SARS-CoV and MERS-CoV caused fatal pneumonia associated with rapid virus replication, elevation of proinflammatory cytokines, and immune cell infiltration [14]. These characteristics are similarly observed in SARS-CoV-2 infection. COVID-19 patients have increased levels of proinflammatory effector cytokines, such as TNFα, IL1B, and IL6, as well as chemokines, such as CCL2 and CXCL10, especially in those who are critically ill [15, 16, 17, 18]. These studies suggest that an over exuberant immune response characterized by cytokine storm rather than direct virus-induced damage may be responsible for COVID-19 pathogenesis. The cell types and mechanisms underlying this immune response are unclear for SARS-CoV-2.

Our knowledge of SARS-CoV-2 biology and pathogenesis is limited. To address this gap, we performed single-cell (sc) RNA sequencing (RNA-seq) on organotypic human bronchial epithelial cells (HBECs) infected with SARS-CoV-2. This culture system supports epithelial cell differentiation and mimics key aspects of the mucosal epithelium. By utilizing scRNA-seq and electron microscopy, we revealed that ciliated cells are a major target of SARS-CoV-2 infection. During the course of infection, cell tropism of SARS-CoV-2 extend to other epithelial cells including basal and club cells. Furthermore, SARS-CoV-2 infection elicited intrinsic expression of type I and type III interferons and IL6 but not IL1 transcripts. Interferon stimulated gene (ISG) expression was observed in both infected and bystander cell populations. Here, we provide a detailed analysis of SARS-CoV-2 infection in HBECs, characterizing SARS-CoV-2 transcription, cell tropism, host gene expression and cell state related to infection.

2 Results

2.1 Viral infection dynamics

To characterize SARS-CoV-2 interactions with the human airway, we performed single-cell RNA sequencing of SARS-CoV-2 infected airway epithelium. We cultured primary HBECs at an air-liquid interface (ALI) for 28 days and then challenged the apical surface of the epithelium with 10^4 plaque forming units (PFU) of SARS-CoV-2 (Fig 1A). Exponential viral replication over the course of the experiment was demonstrated by qRT-PCR of cell lysate for the SARS-CoV-2 nucleocapsid (N) gene (Fig 1B). At 1, 2, and 3 days post-infection (dpi), a single cell suspension was generated and 3' single-cell RNA sequencing was performed on 77,143 cells across four samples cells per sample with an average of 31,383 reads per cell (Fig 1C, S1A). To define SARS-CoV-2 infected cells, we mapped reads to the viral reference genome and quantified viral transcript abundance on a per cell basis (Fig 1D). We defined productively infected cells as those with at least ten viral transcripts per cell, which controls for background due to misaligned reads in the mock sample. Consistent with the viral genome replication (Fig 1B), we observed a time-dependent increase in the abundance of infected cells from 1 to 3 dpi (Fig 1E, S1B).

We first characterized the SARS-CoV-2 transcriptome at the single-cell level. In addition to the reads expected to align immediately upstream of the canonical SARS-CoV-2 poly-A tail, our results show additional

reads aligning elsewhere in the genome suggesting the existence of non-canonical, poly-adenylated subgenomic RNAs (sgRNAs) (Fig 1F, G). The distribution of polyadenylated viral transcripts shifts from 3' to 5' during the infection time-course (Fig 1F,G,S1C). Mapping of productive infected cells reveals multiple infected cell clusters that expand over time and are not present in the mock sample (Fig 1H). Using RT-PCR, we successfully validate two unique peaks, one that mapped in the middle of the Open Reading Frame (ORF)1ab region, and a second peak that mapped near the ORF6 boundary (Fig S1D, top panel). Our results confirm that RT-PCR products corresponding to each of the two peaks appears after 2 dpi (Fig S1D, bottom panel, red arrows). Importantly, the absence of these RT-PCR bands in the mock and 1 dpi samples suggests they are not the result of non-specific oligo-d(T) priming of cellular or viral RNAs. We included two positive controls, amplifying RT-PCR products of increasing length from the canonical SARS-CoV-2 poly-A tail (Fig S1D, bottom panel, green arrows). These RT-PCR bands appear as early as 1 dpi, are specific to infected cells and run at their expected lengths. These positive controls validate that we are able to capture known poly-adenylated viral transcripts with this RT-PCR priming strategy.

2.2 Identification of the cell tropism of SARS-CoV-2

The human airway is comprised of diverse epithelial cell types with critical functions in gas exchange, structure, and immunity. We sought to determine the cellular tropism of SARS-CoV-2 in the bronchial epithelium, as the airway is a critical target of viral pathogenesis. Clustering a BB-kNN graph by Louvain community detection resulted in identification of eight major clusters comprising canonical epithelial cell types: ciliated cells, basal cells, club cells, goblet cells, neuroendocrine cells, ionocytes, and tuft cells (Fig 2A, S2A). We also observed a cell population intermediate between basal cells and club cells (BC/Club) likely representing basal stem cells differentiating into club cells. Analysis of differentially expressed genes reveal these cell clusters express classical epithelial cell type-specific markers (Fig 2B). Mapping viral infected cells within specified epithelial cell types reveals that ciliated, basal, club, and BC/Club cells are susceptible to SARS-CoV-2 infection whereas goblet, neuroendocrine, tuft cells, and ionocytes are relatively resistant to infection (Fig 2C,D). At 1 dpi, ciliated cells represent 83% of infected cells and continue to comprise the majority of infected cells throughout infection (Fig 2E,F). However, during productive infection, the number of infected basal, club, and BC/Club cells also increases, suggesting that these cells are major secondary targets (Fig 2E,F). The distribution of polyadenylated viral transcripts along the length of the genome is similar across infected cell types (Fig S2C).

To independently verify SARS-CoV-2 cell tropism, HBECs cultured under identical conditions as for scRNA-seq were assessed by transmission electron microscopy. At 2 dpi, we observed numerous virus particles approximately 80 nm in size in ciliated cells (Fig 2G). This is consistent with the known size distribution of coronaviruses [19]. These particles were not observed in a mock control sample (Fig 2G). Together, this confirms that ciliated cells are a major target of SARS-CoV-2 infection in the human bronchus.

2.3 Determinants of cell tropism

Next, we sought to determine the host transcriptional correlates of SARS-CoV-2 cell tropism. As viral entry is the major determinant of cell tropism, we first investigated whether expression of the SARS-CoV-2 receptor ACE2 predicted infection. We observed ACE2 expression at low levels across ciliated, basal, club and BC/club cells in the mock condition (Fig 3). Surprisingly, ACE2 expression was poorly correlated with SARS-CoV-2 infection on a per cell basis (Spearman's r between viral genome and ACE2 in ciliated cells, -0.06, and between ACE2 and infection score in ciliated cells, 0.09). However, ACE2 expression was increased in the four susceptible cell populations: ciliated, basal, club, and BC/club relative to the non-susceptible cell types: neuroendocrine, ionocytes, goblet, and tuft cells (Fig S3). Notably, expression of ACE and CLTRN which are structural homologs of ACE2 and the aminopeptideases ANPEP and DPP4 (the MERS receptor) were also poorly correlated with SARS-CoV-2 susceptibility (Fig 3B-E).

ACE2 was recently demonstrated to be an ISG [20]. Here we observe a modest increase in ACE2 expression in both infected and bystander cells relative to the mock sample consistent with dynamic regulation of ACE2 expression by the host innate immune response to SARS-CoV-2 (Fig 3B-E). To examine whether expression of other potentially pro-viral genes explain SARS-CoV-2 cell tropism, we assessed the expression of the proteases that may potentiate SARS-CoV-2 infection. The transmembrane serine protease TMPRSS2

and cathepsin L have been implicated in SARS-CoV-2 entry [21]. We also examined the related protease TMPRSS4 which cleaves influenza hemaglutinin, similar to TMPRSS2, and may also play a role in SARS-CoV-2 entry [21, 22]. TMPRSS2 and CTSL were expressed predominantly in basal, club and ciliated cells while TMPRSS4 was broadly expressed in all epithelial cell types. The specific role of proteases in governing SARS-CoV-2 tropism in the human airway epithelium remains to be further elucidated.

2.4 Innate immune response to SARS-CoV-2 infection

We investigated the transcriptome to assess the host immune response to SARS-CoV-2 infection at single-cell resolution in the human airway epithelium. We observed robust induction of both type I interferon (IFNB1) and type III interferons (IFNL1, IFNL2, and IFNL3) in ciliated, basal, club, and BC/club cells co-expressing SARS-CoV-2 transcripts (Fig 4). Interestingly, the kinetics of IFNB1 induction were delayed relative to type III interferon. In contrast, there was scant IFN induction in uninfected ciliated, basal, club, and BC/club cells. We also did not observe IFN induction in neuroendocrine, ionocytes, goblet, or tuft cells consistent with these cell types not being major target cells of SARS-CoV-2 (Fig S4). This demonstrates direct viral infection of a given cell is critical for interferon induction. Type I and III interferons signal through IFNAR and IFNLR, respectively, resulting in expression of hundreds of ISGs. Consistent with this, we observed broad ISG induction (IFI27, IFITM3, IFI6, MX1, and ISG15) in both infected and bystander cells of all cell types (Fig 4, S4) suggesting IFN from infected cells is acting*in trans* on both infected cells and uninfected bystander cells.

The host anti-viral response also results in chemokine induction leading to recruitment of immune cells, a hallmark of severe COVID-19. Here, we observe induction of CXCL9, CXCL10, and CXCL11 which propagate signals through the cognate CXCR3 receptor to recruit activated T cells and NK cells (Fig 4). This induction was evident in infected but not bystander cells (Fig 4). In contrast, CCL2 and CXCL16 which recruit monocytes and T cells, respectively, were not dynamically regulated over the conditions evaluated (Fig 4 and S4). We also observed substantial induction of the pro-inflammatory cytokine IL-6 in infected ciliated, basal, club, and BC/club cells but not in uninfected bystander cells of these same populations. Interestingly, expression of pro-inflammatory IL-1 was modestly downregulated in all cell types after infection whereas IL-10 and TNFα expression were not significantly regulated by infection in this system (Fig 4).

2.5 Differentially expressed genes in response to SARS-CoV-2 infection

To determine how SARS-CoV-2 infection perturbed the cellular transcriptome, we computationally pooled the three infected samples and analyzed the top 100 differentially expressed genes between infected and uninfected bystander cells of a given cell type within the 1, 2, and 3 dpi samples (Fig 5A). PANTHER gene ontology analysis revealed infected ciliated cells had increased expression of genes involved in apoptosis (e.g. PMAIP1, SQSTM1, ATF3), translation initiation and viral gene expression (e.g. RPS12, RPL37A) and inflammation (e.g. NFKBIA and NFKBIZ) compared to bystander cells (Fig 5B,C and differentially expressed gene lists in Supplemental Files). Similar genes are enriched in other infected cell populations (Fig S5). In contrast, infected ciliated cells showed significant downregulation of genes included in biological processes involved in cilium function (e.g. DYNLL1), calcium signaling (e.g. CALM1, CALM2), and iron homeostasis (e.g. FTH1, FTL; Fig 5B,C and S5). Together this suggests that SARS-CoV-2 infection reprograms the cellular transcriptome, resulting in promotion of viral infection and potentially resulting in cell dysfunction and apoptosis that contributes to COVID-19 pathogenesis.

3 Discussion

To effectively treat COVID-19, we must first understand how SARS-CoV-2 causes disease and why the clinical presentation varies from asymptomatic infection to lethal disease. Here, we report the first longitudinal single-cell transcriptomic analysis of SARS-CoV-2 infected respiratory epithelium using an established organoid model that reproduces the orientation of airway epithelium. The transcriptional data generated is of high quality, with an average of between 2,400 to 3,600 unique genes detected per condition. Our data reveals several novel viral transcripts and our methodology differentiates infected from bystander cells. Further, we demonstrate that ciliated cells are the major target cell of SARS-CoV-2 infection in the bronchial

epithelium at the onset of infection and that cell tropism expands to basal, club, and BC/club cells over time. We also reveal that SARS-CoV-2 potently induces IFN in infected cells resulting in broad ISG expression in both infected and bystander cells. We also observe potent induction of the pro-inflammatory cytokine IL-6 and chemokines, which likely contribute to the inflammatory response *in vivo*.

Single-cell transcriptomics enabled us to elucidate the SARS-CoV-2 transcriptome at single-cell resolution in multiple primary cell types over time. We developed a novel method to differentiate productively infected cell types by the distribution of the sub-genomic and genomic viral transcripts. We also identified polyadenylated viral transcripts remote from the 3' end of the viral genome, which was unexpected given our sequencing method. Our RT-PCR validation experiments confirm the production of at least two unique, TRS-independent transcripts with poly-A tails that do not appear to result from non-specific oligo-d(T) priming. As the reported recombination rate for coronaviruses is high [23, 24] it is possible these short reads correspond to non-specific polymerase jumping. However, recent studies have identified TRS-independent chimeric RNAs produced during SARS-CoV-2 infection of Vero cells, a small portion (1.5%) of which are fused in frame [25]. Taken together with our results, this may suggest non-canonical sub-genomic RNAs with coding potential are produced during SARS-CoV-2 infection; however, this would require further validation. HCoV-299E non-structural protein 8 (nsp8) was recently shown to possess template-independent adenyltransferase activity [26]. Because poly-A tails play important roles in the stability and translation potential of canonical SARS-CoV-2 sgRNAs, it is interesting to speculate that coronaviruses might rely on the production of non-canonical, poly-adenylated sgRNAs to serve as decoys for cellular deadenylases. This would result in preservation of the poly-A tails of the genomic and subgenomic RNAs. Indeed, the production of sgRNAs during flaviviral infections is important for resistance to cellular exoribonucleases and innate immune evasion [27, 28].

Identification of the cell types infected by SARS-CoV-2 informs pathogenesis. We find that SARS-CoV-2 infects ciliated cells, basal, club, and BC/club cells. This may result in aberrant function of these critical cell types. Ciliated cells, which are abundant in the respiratory epithelium, propel mucus and associated foreign particles and microbes proximally away from the lower airway. Our finding that ciliated cells are the predominant target cell of SARS-CoV-2 infection in the bronchial epithelium has several important implications. First, dysfunction of ciliated cells by infection by SARS-CoV-2 may impair mucociliary clearance and increase the likelihood of secondary infection. Second, asthma, chronic obstructive pulmonary disease, and smoking are associated with both cilia dysfunction and increased severity of COVID-19. Whether these underlying conditions alter ciliated cells and thus increase their susceptibility to infection remains unclear. The cell tropism of SARS-CoV-2 in the nasal epithelium and lower airway remain important areas of future investigation which will further enhance our understanding of COVID-19 pathogenesis.

Disease in COVID-19 patients is characterized by a lag following transmission with symptom onset at day seven and disease severity peaking 14 days post infection [29, 30]. This is in contrast to seasonal human coronaviruses and implicates an important role for the host immune response in COVID-19 progression. Several recent studies have revealed seemingly contradictory roles regarding SARS-CoV-2 induced innate immunity [31, 32]. Here, we show that the innate response to SARS-CoV-2 is intact and rapid, as characterized by interferon, chemokine, and IL-6 induction. Interestingly, while we do not observe broad depletion of virus-susceptible cell populations, we detect increased expression of cell-death associated genes, which suggests the host anti-viral response is cytotoxic and may contribute to disease pathogenesis. Consistent with this, IL-6 is a potent pro-inflammatory cytokine and serum IL-6 levels predict respiratory failure [33]. Therapies targeting the IL-6 receptor are currently in clinical trials for the treatment of COVID-19. This work raises a number of important future directions including whether other airway and endothelial tissues similarly interact with SARS-CoV-2 and how these interactions vary *in vitro*.

4 Methods

4.1 Air-liquid interface culture of HBECs

HBECs, from Lonza, were cultured in suspension in PneumaCult-Ex Plus Medium according to manufacturer instructions (StemCell Technologies, Cambridge, MA, USA). To generate air-liquid interface cultures, HBECs were plated on collagen-coated transwell inserts with a 0.4-micron pore size (Costar, Corning, Tewksbury, MA, USA) at $5x10^4$ cells/ml per filter and inserted into 24 well culture plates. Cells were maintained for

the first 3 days in PneumaCult-Ex Plus Medium, then changed to PneumaCult-ALI Medium (StemCell Technologies) containing the ROCK inhibitor Y-27632 for 4 days. Fresh medium, 100 μl in the apical chamber and 500 μl in the basal chamber, was provided every day. At day 7, medium at the apical chambers were removed, while basal chambers were maintained with 500 μl of PneumaCult-ALI Medium. HBECs were maintained at air-liquid interface for 28 days allowing them to differentiate. Medium in the basal chamber was changed every 2-3 days (500 μl).

4.2 Viral infection

SARS-CoV-2 isolate USA-WA1/2020 was obtained from BEI reagent repository. All infection experiments were performed in a Biosafety Level 3 facility, licensed by the State of Connecticut and Yale University. Immediately prior to infection, the apical side of the HBEC ALI culture was gently rinsed three times with 200 μl of phosphate buffered saline without divalent cations (PBS-/-). 10^4 plaque forming units (PFU) of SARS-CoV-2 in 100 μl total volume of PBS was added to the apical compartment. Cells were incubated at $37°C$ and 5% CO_2 for 1 hour. Unbound virus was removed and cells were cultured with an air-liquid interface for up to three days. Infections were staggered by one day and all four samples were processed simultaneously for single-cell RNA sequencing, as described below.

4.3 Sample preparation for single-cell RNA sequencing

Inoculated HBECs were washed with 1X PBS-/- and trypsinized with TrypLE Express Enzyme (ThermoFisher, Waltham, MA, USA) to generate single-cell suspensions. $100\mu l$ of TrypLE was added on the apical chamber, incubated for 10 min at $37°C$ in a CO_2 incubator, and was gently pipetted up and down to dissociate cells. Harvested cells were transferred in a sterile 1.5 ml tube and neutralized with DMEM containing 10 percent FBS. An additional 100 μl of TrypLE was placed on the apical chamber repeating the same procedure as above for a total of 30 min to maximize collection of cells. Cells were centrifuged at 300 x g for 3 min and resuspended in 100 μl DMEM with 10 percent FBS. Cell count and viability was determined using trypan blue dye exclusion on a Countess II (ThermoFisher Scientific). The targeted cell input was 10,000 cells per condition. The Chromium Next GEM (Gel Bead-In Emulsion) Single Cell 3' Gel beads v3.1 kit (10X Genomics, Pleasanton, CA, USA) was used to create GEMs following manufacturer's instruction. All samples and reagents were prepared and loaded into the chip and ran in the Chromium Controller for GEM generation and barcoding. GEMs generated were used for cDNA synthesis and library preparation using the Chromium Single Cell 3' Library Kit v3.1 (10X Genomics) following the manufacturer's instruction. Generated libraries were sequenced on NovaSeq 6000 system using HiSeq 100 base pair reads and dual indexing. Cells were sequenced to an average depth of 31,383 reads per cell. The human genome, Ensembl GRCh38.98.gtf, and the SARS-CoV-2 genome, NCBI Genome database accession MT020880.1, were combined and used for alignment. We ran the standard 10x Genomics cellranger pipeline with a combined human and SARS-CoV-2 genome to obtain count matrices for each of the 4 growth conditions. Per condition, there were an average of between 10,000 to 15,000 counts per cell or an average of 2,400 to 3,600 unique genes detected per condition.

4.4 Quantitative RT-PCR of SARS-CoV2

Viral RNA from SARS-CoV-2 infected HBEC cell lysates was extracted using TRIzol (Life Technologies) and purified using Direct-zol RNA MiniPrep Plus according to manufacturer's instructions (Zymo Research, Irvine, CA, USA). A two-step cDNA synthesis with 5 μl RNA, random hexamer, and ImProm-II Reverse Transcriptase (Promega, Madison, WI, USA) was performed. The qPCR analysis was performed in duplicate for each of the samples and standard curves generated using SARS-CoV-2 nucleocapsid (N1) specific oligonucleotides from Integrated DNA Technologies (Coralville, IA, USA): Probe: 5' 6FAM ACCCCGCATTACGTTTGGTGGACC-BHQ1 3'; Forward primer: 5' GACCCCAAAATCAGCGAAAT 3'; Reverse primer: 5' TCTGGTTACTGCCAGTTGAATCTG 3'. The limit of detection was 10 SARS-CoV-2 genome copies/μl. The virus copy numbers were quantified using a control plasmid which contain the complete nucleocapsid gene from SARS-CoV-2.

4.5 Validation of polyadenylated SARS-CoV-2 transcripts

Huh7.5 cells grown in DMEM containing 10 percent FBS were infected with 10^4 PFU of SARS-CoV-2 and cell lysates were harvested at 0, 1, 2, and 3 dpi. Using 0.3 μg total RNA extracted from mock or SARS-CoV-2-infected Huh7.5 cells at different time points, reverse transcription was performed with oligo-d(T)20 (ThermoFisher) and MarathonRT, a highly processive group II intron-encoded RT. MarathonRT purification and RT reactions were performed as previously described [34]. PCR (NEBNext Ultra II Q5®Master Mix, NEB, Ipswich, MA, USA) was performed with a gene-specific forward primer designed 700-nt upstream of the apparent boundary between the SARS-CoV-2 genome body and the putative poly-A tail. Oligo-d(T)20 was used as a reverse primer. Touchdown PCR cycling was used to enhance specificity of the PCR reaction. RT-PCR products were resolved on a 1.3% agarose gel with ladder (100 bp DNA Ladder, 1 kb Plus DNA Ladder, Invitrogen). Forward PCR oligonucleotides used in this experiment are below, which includes two positive controls.

Primer Name	Position on Genome	5'-3' Sequence:
F_Val_1	7700	GAGAGACTTGTCACTACAGTTTAAA
F_Val_2	26650	AATTTGCCTATGCCAACAGGA
F_Val_(+)ve_1	28600	AGATCTCAGTCCAAGATGGTA
F_Val_(+)ve_2	29000	GGTAAAGGCCAACAACAACAA

4.6 Electron microscopy

The cells were fixed using 2.5% glutaraldehyde in 0.1M phosphate buffer, osmicated in 1% osmium tetroxide, and dehydrated in increasing ethanol concentrations. During dehydration, 1% uranyl acetate was added to the 70% ethanol to enhance ultrastructural membrane contrast. After dehydration the cells were embedded in Durcupan. 70 nm ultrathin sections were cut on a Leica ultramicrotome, collected on Formvar coated single-slot grids, and analyzed with a Tecnai 12 Biotwin electron microscope (FEI).

4.7 ScRNA-seq data analysis

4.7.1 Cell type annotation

We used the standard scRNA-seq analysis pipeline for clustering [35]. Briefly, to account for transcript dropout inherent to scRNA-seq, we removed genes that were expressed in fewer than 3 cells and removed cells that expressed fewer than 200 genes. Next, we filter out cells with more than 10 percent of mitochondrial genes. We did not find a correlation between viral copy number and mitochondrial expression. The resulting raw unique molecular identifier (UMI) counts in each cell were normalized to their library size. Then, normalized counts were square-root transformed, which is similar to a log transform but does not require addition of a pseudo count. Data pre-processing was performed in Python (version 3.7.4) using Scanpy (version 1.4.6) [36].

We visually observed batch effects between conditions in 2-dimensional cellular embeddings. To remove these batch effects for clustering, cell-type annotation, and visualization, we used an approximate batch-balanced kNN graph for manifold learning (BB-kNN batch-effect correction) using Scanpy's fast approximation implementation [37, 36]. BB-kNN assumes that at least some cell types are shared across batches and that differences between batches for a same cell type are lower than differences between cells of different types within a batch. For each cell, the 3-nearest neighboring cells in each condition were identified by Euclidean distance in 100-dimensional Principal Component Analysis (PCA) space. This kNN graph was used as the basis for downstream analysis.

To visualize the scRNA-seq data we implemented various non-linear dimension reduction methods and used the BB-kNN batch-corrected connectivity matrix as input for Uniform Manifold Approximation and Projection (UMAP) [38] and Potential of Heat-diffusion for Affinity-based Trajectory Embedding (PHATE) [39]. UMAP projections were generated using a minimum distance of 0.5. PHATE projections were generated with a gamma parameter of 1.

For cell clustering we used the Louvain community detection method [40] with the BB-kNN graph. We used high-resolution community detection and merged clusters based on expression of bronchial epithelium

cell-type markers in order to isolate some rare cell types, e.g. Tuft cells [41, 42]. To annotate the different cell types present in HBECs we analyzed expression of a range of marker genes that were reported in a molecular cell atlas from Travaglini et al. [41]. We focused on 8 cell types: (i) Basal cells (KRT5, DAPL1, TP63), (ii) Ciliated cells (FOXJ1, CCDC153, CCDC113, MLF1, LZTFL1), (iii) Club cells (SCGB1A1, KRT15, CYP2F2, LYPD2, CBR2), (iv) BC/club (KRT4, KRT13), (v) Neuroendocrine cells (CHG1, ASCL1), (vi) Tuft cells (POU2F3, AVIL, GNAT3, TRPM5), (vi) Ionocytes (FOXI1, CFTR, ASCL3) and (viii) goblet cells (MUC5AC, MUC5B, GP2, SPDEF).

4.7.2 Infection threshold and infection score

Counting a viral transcript in a cell does not mean the cell is infected, as this count can come from virus attached to the surface of the cell, ambient virus in the suspension, or from read misalignment. Given the reported shared 3' poly(A) tail in coronavirus transcripts [43], we were unsure whether we could correctly capture the different ORFs using the 10x Genomics 3' gene expression library. Therefore, we aligned the viral reads to a genome-wide single "exon," i.e., a count is given for a read mapped to SARS-CoV-2 ORFs and intergenic regions. These counts were used to infer individual cells' infectious state. To filter out cells with viral genome transcript counts that may result from viral-cell surface attachment, ambient virus in the droplet suspension, or read misalignment, we considered infected cells to have ≥ 10 viral transcripts counts. This value was determined by a threshold of viral counts in the mock condition. While the mock condition is not expected to have viral counts, we did observe a small number that we attribute to misalignment. We observed only 5 mock cells with full SARS-CoV-2 viral genome transcript counts ≥ 10 transcripts. These criteria allowed us to find 144 infected cells at 1 dpi, 1428 cells at 2 dpi and 3173 cells at 3 dpi.

To quantify the extent to which an individual cell is transcriptionally similar to an infected cell, we used a previously developed graph signal processing approach called Manifold Enhancement of Latent Dimensions (MELD) [44]. We encoded a raw experimental score for each cell in the dataset such that -1 represents a bystander or uninfected cell and +1 represents an infected cell. Using the kernel from the BB-kNN graph (described above), these raw scores were smoothed in the graph domain, yielding an "infection score" per cell that represents the extent to which an individual cell is transcriptionally similar to infected cells. For summary statistics, this score was stratified by cell type and condition.

4.7.3 Viral genome read coverage analysis

To visualize the viral read coverage along the viral genome we used the 10X Genomics cellranger barcoded binary alignment map (BAM) files for every sample. We filtered the BAM files to only retain reads mapping to the viral genome using the *bedtools intersect* tool. We converted the BAM files into sequence alignment map (SAM) files in order to filter out cells that were removed in our single cell data preprocessing pipeline. The sequencing depth for each base position was calculated using *samtools count*. To characterize read distribution along the viral genome we counted transcripts of 10 different ORFs: ORF1ab, Surface glycoprotein (S), ORF3a, Envelope protein (E), Membrane glycoprotein (M), ORF6, ORF7a, ORF8, Nucleocapsid phosphoprotein (N) and ORF10.

4.7.4 Differential Expression Analysis

To find differentially expressed genes across conditions, we used a combination of three metrics: the Wasserstein or Earth Mover's distance, an adjusted p-value from a two-sided Mann-Whitney U test with continuity and Benjamini-Hochberg correction, and the binary logarithm of fold change between mean counts. Significance was set to $p_{adjusted} \leq 0.01$. The Earth Mover's distance, or 1-dimensional Wasserstein distance can be defined as the minimal cost to transform of distribution to another, and was previously used to assess gene expression that significantly differ between conditions [45, 46]. We performed several binary comparisons for each timepoint and for pooling 1, 2, and 3 dpi: infected vs. bystander, infected vs. mock cells, and bystander vs. mock cells. The 30 most differentially expressed genes (up- or downregulated, ranked by Wasserstein distance) in each condition, cell type, and analysis were represented in heatmaps. To identify putative cellular functions changed across conditions, we performed PANTHER-GO [47] statistical over-representation tests for up-regulated genes in each cell type and condition separately, using the default Human PANTHER-GO reference list as a background.

5 Code and Data Availability

All differential gene expression analyses and their associated metrics are publicly available at the Van Dijk Lab GitHub: https://github.com/vandijklab. The scRNA-seq data will be made publicly available around two weeks after pre-print submission, to allow time for finalizing the manuscript. The data will be deposited in the NCBI Geo database.

6 Acknowledgments

This work was supported by NIH grants K08 AI128043 (CBW), R21AI133440 (AW), R01AI141609 (AW), and K08AI119139 (EEF). CBW was also supported by a Burroughs Wellcome Fund Career Award for Medical Scientists and Robert Leet and Clara Guthrie Patterson Trust Award. CBW and DvD were supported by a Fast Grant (Emergent Ventures). AI and AMP are Investigators of the Howard Hughes Medical Institute.

Author contributions: conceptualization, MMA, JW, NGR, VG, DvD, CBW; methodology, MMA, NGR, EFF, DvD, CBW; computational analysis NGR, VG, DvD; validation, MMA, CBW; formal analysis NGR, VG, DvD; investigation, MMA, RBF, JW, CBW; resources, BW, EFF; writing—original Draft, MMA, NGR, VG, DvD, CW; writing—review & editing, all authors; visualization, NGR, VG, DvD; supervision, DvD, CBW; funding acquisition, DvD, CBW.

Declaration of interests: The authors declare no competing interests.

Schematics were created with BioRender.com.

References

[1] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*, vol. 395, no. 10223, pp. 470–473, 2020.

[2] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, *et al.*, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, no. 7798, pp. 270–273, 2020.

[3] J. Cui, F. Li, and Z.-L. Shi, "Origin and evolution of pathogenic coronaviruses," *Nature reviews Microbiology*, vol. 17, no. 3, pp. 181–192, 2019.

[4] Q. Tang, Y. Song, M. Shi, Y. Cheng, W. Zhang, and X.-Q. Xia, "Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition," *Scientific reports*, vol. 5, p. 17155, 2015.

[5] W. Li, M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, *et al.*, "Angiotensin-converting enzyme 2 is a functional receptor for the sars coronavirus," *Nature*, vol. 426, no. 6965, pp. 450–454, 2003.

[6] H. Hofmann, K. Pyrc, L. van der Hoek, M. Geier, B. Berkhout, and S. Pöhlmann, "Human coronavirus nl63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry," *Proceedings of the National Academy of Sciences*, vol. 102, no. 22, pp. 7988–7993, 2005.

[7] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, *et al.*, "Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor," *Cell*, 2020.

[8] I. Glowacka, S. Bertram, M. A. Müller, P. Allen, E. Soilleux, S. Pfefferle, I. Steffen, T. S. Tsegaye, Y. He, K. Gnirss, *et al.*, "Evidence that tmprss2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response," *Journal of virology*, vol. 85, no. 9, pp. 4122–4134, 2011.

[9] N. Iwata-Yoshikawa, T. Okamura, Y. Shimizu, H. Hasegawa, M. Takeda, and N. Nagata, "Tmprss2 contributes to virus spread and immunopathology in the airways of murine models after coronavirus infection," *Journal of virology*, vol. 93, no. 6, pp. e01815–18, 2019.

[10] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, "Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein," *Cell*, 2020.

[11] S. Matsuyama, N. Nagata, K. Shirato, M. Kawase, M. Takeda, and F. Taguchi, "Efficient activation of the severe acute respiratory syndrome coronavirus spike protein by the transmembrane protease tmprss2," *Journal of virology*, vol. 84, no. 24, pp. 12658–12664, 2010.

[12] G. Simmons, D. N. Gosalia, A. J. Rennekamp, J. D. Reeves, S. L. Diamond, and P. Bates, "Inhibitors of cathepsin l prevent severe acute respiratory syndrome coronavirus entry," *Proceedings of the National Academy of Sciences*, vol. 102, no. 33, pp. 11876–11881, 2005.

[13] B. Coutard, C. Valle, X. de Lamballerie, B. Canard, N. Seidah, and E. Decroly, "The spike glycoprotein of the new coronavirus 2019-ncov contains a furin-like cleavage site absent in cov of the same clade," *Antiviral research*, vol. 176, p. 104742, 2020.

[14] R. Channappanavar and S. Perlman, "Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology," *Seminars in immunopathology*, vol. 39, no. 5, pp. 529–539, 2017.

[15] G. Chen, D. Wu, W. Guo, Y. Cao, D. Huang, H. Wang, T. Wang, X. Zhang, H. Chen, H. Yu, *et al.*, "Clinical and immunological features of severe and moderate coronavirus disease 2019," *The Journal of clinical investigation*, vol. 130, no. 5, 2020.

[16] B. Diao, C. Wang, Y. Tan, X. Chen, Y. Liu, L. Ning, L. Chen, M. Li, Y. Liu, G. Wang, *et al.*, "Reduction and functional exhaustion of t cells in patients with coronavirus disease 2019 (covid-19)," *Medrxiv*, 2020.

[17] M. Liao, Y. Liu, J. Yuan, Y. Wen, G. Xu, J. Zhao, L. Chen, J. Li, X. Wang, F. Wang, *et al.*, "The landscape of lung bronchoalveolar immune cells in covid-19 revealed by single-cell rna sequencing," *medRxiv*, 2020.

[18] S. F. Pedersen, Y.-C. Ho, *et al.*, "Sars-cov-2: a storm is raging," *The Journal of Clinical Investigation*, vol. 130, no. 5, 2020.

[19] B. W. Neuman, B. D. Adair, C. Yoshioka, J. D. Quispe, G. Orca, P. Kuhn, R. A. Milligan, M. Yeager, and M. J. Buchmeier, "Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy," *Journal of virology*, vol. 80, no. 16, pp. 7918–7928, 2006.

[20] C. G. Ziegler, S. J. Allon, S. K. Nyquist, I. M. Mbano, V. N. Miao, C. N. Tzouanas, Y. Cao, A. S. Yousif, J. Bals, B. M. Hauser, J. Feldman, C. Muus, M. H. Wadsworth, S. W. Kazer, T. K. Hughes, B. Doran, G. J. Gatter, M. Vukovic, F. Taliaferro, B. E. Mead, Z. Guo, J. P. Wang, D. Gras, M. Plaisant, M. Ansari, I. Angelidis, H. Adler, J. M. Sucre, C. J. Taylor, B. Lin, A. Waghray, V. Mitsialis, D. F. Dwyer, K. M. Buchheit, J. A. Boyce, N. A. Barrett, T. M. Laidlaw, S. L. Carroll, L. Colonna, V. Tkachev, C. W. Peterson, A. Yu, H. B. Zheng, H. P. Gideon, C. G. Winchell, P. L. Lin, C. D. Bingle, S. B. Snapper, J. A. Kropski, F. J. Theis, H. B. Schiller, L.-E. Zaragosi, P. Barbry, A. Leslie, H.-P. Kiem, J. L. Flynn, S. M. Fortune, B. Berger, R. W. Finberg, L. S. Kean, M. Garber, A. G. Schmidt, D. Lingwood, A. K. Shalek, and J. Ordovas-Montanes, "Sars-cov-2 receptor ace2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues," *Cell*, 2020.

[21] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, *et al.*, "Characterization of spike glycoprotein of sars-cov-2 on virus entry and its immune cross-reactivity with sars-cov," *Nature communications*, vol. 11, no. 1, pp. 1–12, 2020.

[22] R. Zang, M. F. Castro, B. T. McCune, Q. Zeng, P. W. Rothlauf, N. M. Sonnek, Z. Liu, K. F. Brulois, X. Wang, H. B. Greenberg, *et al.*, "Tmprss2 and tmprss4 mediate sars-cov-2 infection of human small intestinal enterocytes," *bioRxiv*, 2020.

[23] S. Su, G. Wong, W. Shi, J. Liu, A. C. Lai, J. Zhou, W. Liu, Y. Bi, and G. F. Gao, "Epidemiology, genetic recombination, and pathogenesis of coronaviruses," *Trends in Microbiology*, vol. 24, no. 6, pp. 490 – 502, 2016.

[24] C. L. Liao and M. M. Lai, "Rna recombination in a coronavirus: recombination between viral genomic rna and transfected rna fragments.," *Journal of Virology*, vol. 66, no. 10, pp. 6117–6124, 1992.

[25] D. Kim, J.-Y. Lee, J.-S. Yang, J. W. Kim, V. N. Kim, and H. Chang, "The architecture of sars-cov-2 transcriptome," *Cell*, 2020.

[26] J. Tvarogová, R. Madhugiri, G. Bylapudi, L. J. Ferguson, N. Karl, and J. Ziebuhr, "Identification and characterization of a human coronavirus 229e nonstructural protein 8-associated rna 3'-terminal adenylyltransferase activity," *Journal of Virology*, vol. 93, no. 12, 2019.

[27] G. P. Pijlman, A. Funk, N. Kondratieva, J. Leung, S. Torres, L. [van der Aa], W. J. Liu, A. C. Palmenberg, P.-Y. Shi, R. A. Hall, and A. A. Khromykh, "A highly structured, nuclease-resistant, noncoding rna produced by flaviviruses is required for pathogenicity," *Cell Host & Microbe*, vol. 4, no. 6, pp. 579 – 591, 2008.

[28] G. P. Göertz, J. J. Fros, P. Miesen, C. B. F. Vogels, M. L. van der Bent, C. Geertsema, C. J. M. Koenraadt, R. P. van Rij, M. M. van Oers, and G. P. Pijlman, "Noncoding subgenomic flavivirus rna is processed by the mosquito rna interference machinery and determines west nile virus transmission by culex pipiens mosquitoes," *Journal of Virology*, vol. 90, no. 22, pp. 10145–10159, 2016.

[29] N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, and H. Nishiura, "Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data," *Journal of clinical medicine*, vol. 9, no. 2, p. 538, 2020.

[30] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application," *Annals of Internal Medicine*, vol. 172, pp. 577–582, 05 2020.

[31] D. Blanco-Melo, B. Nilsson-Payant, W.-C. Liu, R. Møller, M. Panis, D. Sachs, R. Albrecht, *et al.*, "Sars-cov-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems," *BioRxiv*, 2020.

[32] P. V'kovski, M. Gultom, S. Steiner, J. Kelly, J. Russeil, B. Mangeat, E. Cora, J. Pezoldt, M. Holwerda, A. Kratzel, L. Laloli, M. Wider, J. Portmann, T. Tran, N. Ebert, H. Stalder, R. Hartmann, V. Gardeux, D. Alpern, B. Deplancke, V. Thiel, and R. Dijkman, "Disparate temperature-dependent virus – host dynamics for sars-cov-2 and sars-cov in the human respiratory epithelium," *bioRxiv*, 2020.

[33] T. Herold, V. Jurinovic, C. Arnreich, J. C. Hellmuth, M. von Bergwelt-Baildon, M. Klein, and T. Weinberger, "Level of il-6 predicts respiratory failure in hospitalized symptomatic covid-19 patients," *medRxiv*, 2020.

[34] L.-T. Guo, R. L. Adams, H. Wan, N. C. Huston, O. Potapova, S. Olson, C. M. Gallardo, B. R. Graveley, B. E. Torbett, and A. M. Pyle, "Sequencing and structure probing of long rnas using marathonrt: A next-generation reverse transcriptase," *Journal of Molecular Biology*, 2020.

[35] R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, and A. Regev, "Spatial reconstruction of single-cell gene expression data," *Nature Biotechnology*, vol. 33, no. 5, pp. 495–502, 2015.

[36] F. A. Wolf, P. Angerer, and F. J. Theis, "SCANPY: large-scale single-cell gene expression data analysis," *Genome Biology*, vol. 19, no. 1, p. 15, 2018.

[37] K. Polański, M. D. Young, Z. Miao, K. B. Meyer, S. A. Teichmann, and J.-E. Park, "Bbknn: fast batch alignment of single cell transcriptomes," *Bioinformatics*, vol. 36, no. 3, pp. 964–965, 2020.

[38] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

[39] K. R. Moon, D. van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. van den Elzen, M. J. Hirn, R. R. Coifman, *et al.*, "Visualizing structure and transitions in high-dimensional biological data," *Nature Biotechnology*, vol. 37, no. 12, pp. 1482–1492, 2019.

[40] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[41] K. J. Travaglini, A. N. Nabhan, L. Penland, R. Sinha, A. Gillich, R. V. Sit, S. Chang, S. D. Conley, Y. Mori, J. Seita, *et al.*, "A molecular cell atlas of the human lung from single cell rna sequencing," *bioRxiv*, p. 742320, 2019.

[42] D. T. Montoro, A. L. Haber, M. Biton, V. Vinarsky, B. Lin, S. E. Birket, F. Yuan, S. Chen, H. M. Leung, J. Villoria, *et al.*, "A revised airway epithelial hierarchy includes cftr-expressing ionocytes," *Nature*, vol. 560, no. 7718, pp. 319–324, 2018.

[43] S. Sawicki and D. Sawicki, "A new model for coronavirus transcription," in *Coronaviruses and Arteriviruses*, pp. 215–219, Springer, 1998.

[44] D. B. Burkhardt, J. S. Stanley, A. L. Perdigoto, S. A. Gigante, K. C. Herold, G. Wolf, A. J. Giraldez, D. van Dijk, and S. Krishnaswamy, "Quantifying the effect of experimental perturbations in single-cell rna-sequencing data using graph signal processing," *bioRxiv*, 2019.

[45] T. Wang and S. Nabavi, "Sigemd: a powerful method for differential gene expression analysis in single-cell rna sequencing data," *Methods*, vol. 145, pp. 25–32, 2018.

[46] D. Y. Orlova, N. Zimmerman, S. Meehan, C. Meehan, J. Waters, E. E. Ghosn, A. Filatenkov, G. A. Kolyagin, Y. Gernez, S. Tsuda, *et al.*, "Earth mover's distance (emd): a true metric for comparing biomarker expression levels in cell populations," *PloS one*, vol. 11, no. 3, 2016.

[47] H. Mi, A. Muruganujan, D. Ebert, X. Huang, and P. D. Thomas, "Panther version 14: more genomes, a new panther go-slim and improvements in enrichment analysis tools," *Nucleic acids research*, vol. 47, no. D1, pp. D419–D426, 2019.

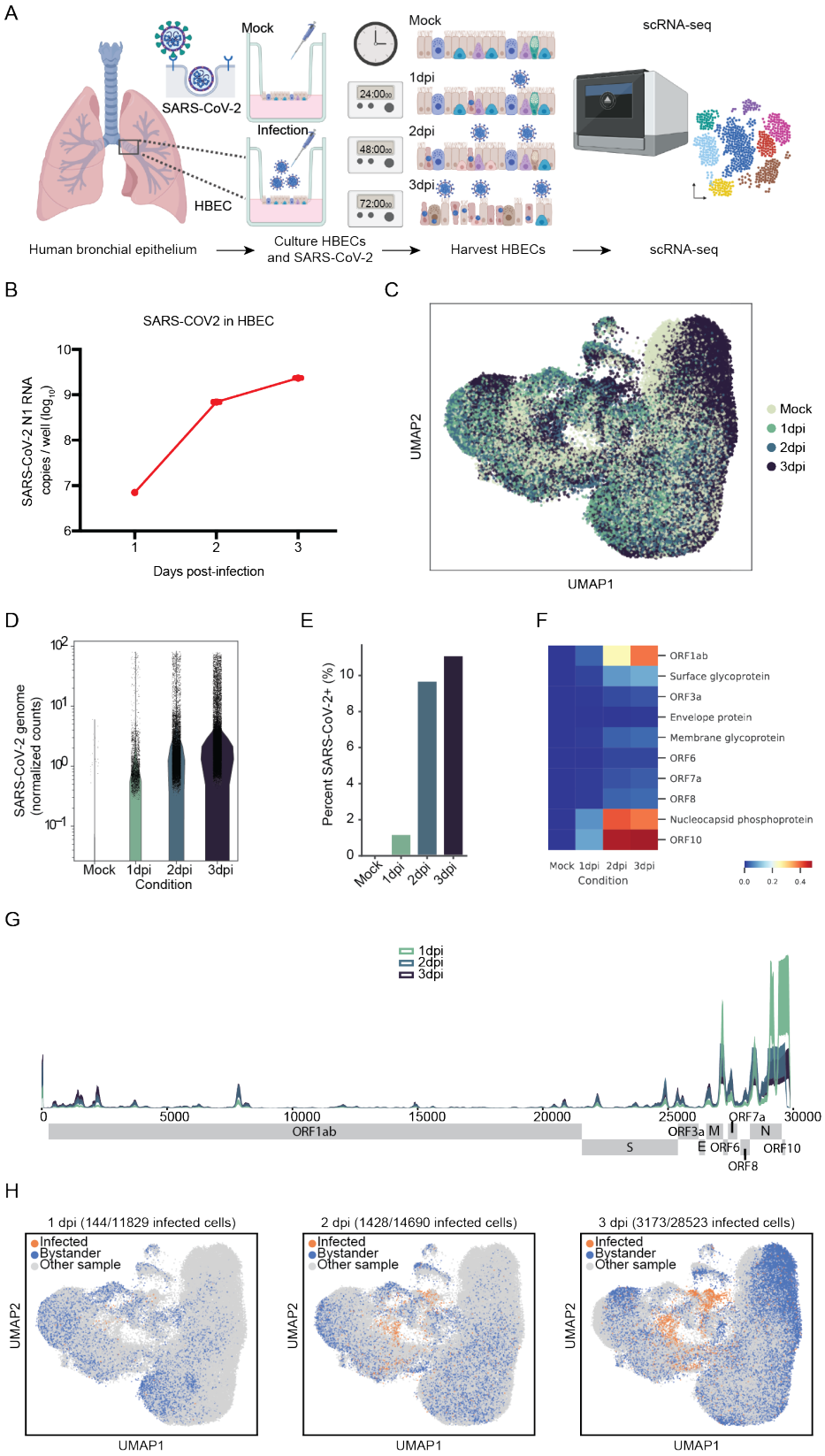

Figure 1: *Caption next page.*

Figure 1: scRNA-seq reveals SARS-CoV-2 infection of HBECs. **A.** Schematic of the experiment. Human bronchial epithelial cells (HBECs) were cultured and infected or not (mock) with SARS-CoV-2. Infected cultures were collected for scRNA-seq at 1, 2 and 3 days post infection (dpi). **B.** RT-qPCR in cultured HBEC to detect viral transcripts at each dpi (copies/well). **C.** UMAP visualization of the scRNA-seq gene counts after batch correction. Each point represents a cell, colored by sample. **D.** Normalized counts of viral counts in each condition. For each cell, viral counts were determined by aligning reads to a single, genome-wide reference. **E.** Percent of cells infected by SARS-CoV-2, based on a viral genes count threshold (see Materials and Methods) **F.** Normalized heatmap of the viral Open Reading Frame (ORF) counts in each condition. Reads were aligned to each 10 SARS-CoV-2 ORFs. **G.** Coverage plot of viral reads aligned to SARS-CoV-2 genome. The sequencing depth was computed for each genomic position for each condition. As infection progresses, coverage becomes more dispersed on the genome. **H.** UMAP visualizations of infected and bystander cells in each condition after batch correction. Bystander cells are defined as cells that remained uninfected in infected HBEC samples.

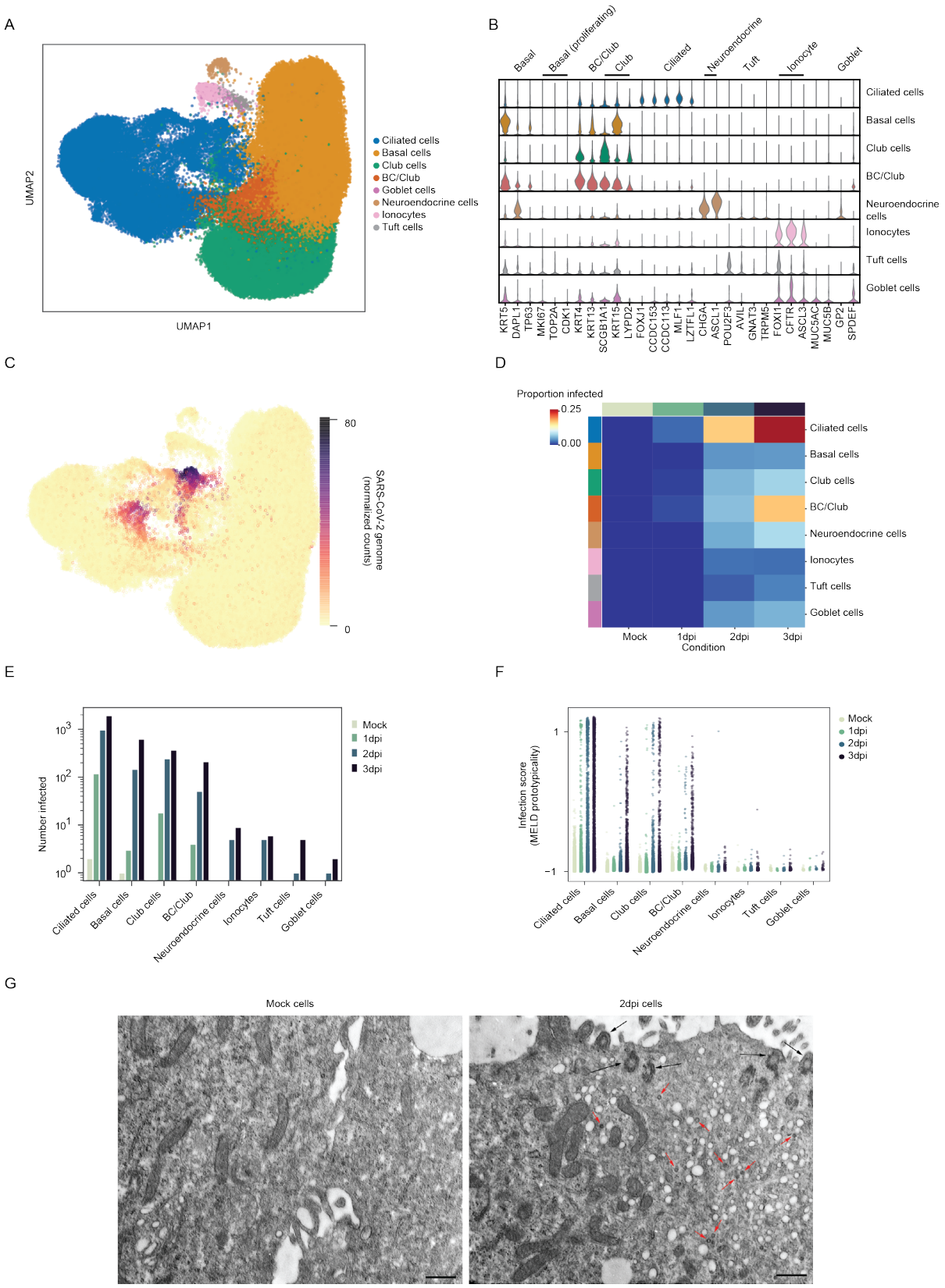

Figure 2: *Caption next page.*

Figure 2: SARS-CoV-2 cell tropism. **A.** UMAP visualization of the cell clusters manually annotated. Cells were first clustered with the Louvain algorithm, then annotated according to a panel of marker genes. **B.** Violin plot of annotation marker genes and SARS-CoV-2 putative relevant genes based on the recent literature. **C.** Uniform Manifold Approximation and Projection (UMAP) visualization of the normalized counts of SARS-CoV-2 reads. Reads were determined here as in Figure 1D.**D.** Proportion of infected cells across conditions and cell types. **E.** Histogram of the number of infected cells per cell type across conditions. **F.** Infection score inferred from Manifold Enhancement of Latent Dimensions (MELD) showing extent of infection per cell stratified by cell time. **G.** Transmission electron microscopy image of mock (left) and SARS-CoV-2 human bronchial epithelial cell (HBEC) at 2 days post infection (dpi) (right). Scale (bottom) corresponds to 500 nm. Red arrows denoted virus particles and black arrows cilia.

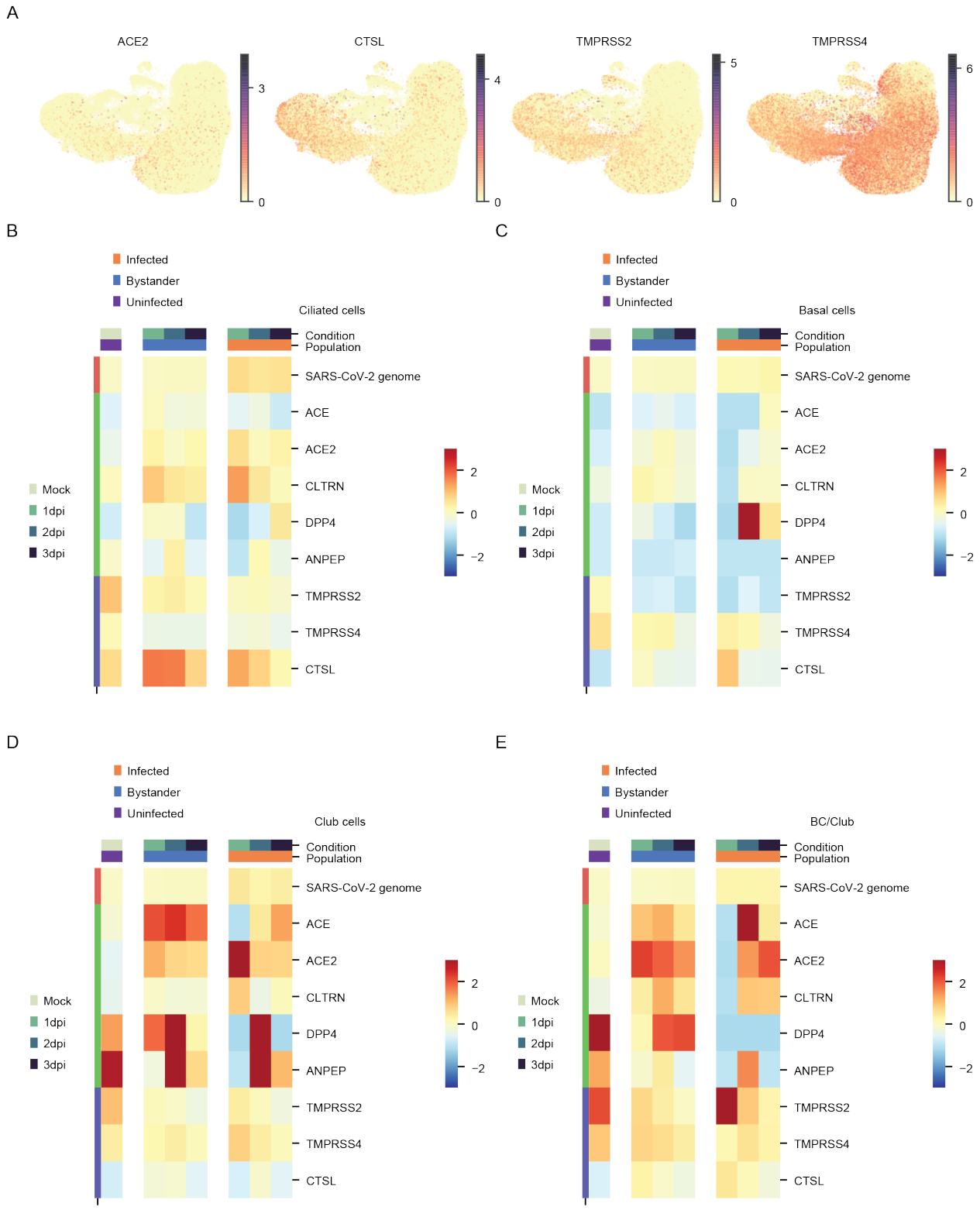

Figure 3: *Caption next page.*

Figure 3: Expression of known entry determinants across bronchial epithelial cell types. **A.** UMAP visualizations, colored by expression of four receptors and proteases expressed in human bronchial epithelial cells (HBECs): ACE2, TMPRSS2, TMPRSS4 and CTSL. **B-E.** Heatmaps of receptors and proteases in ciliated (**B.**), basal (**C.**), club (**D.**) and BC/Club cells (**E.**).

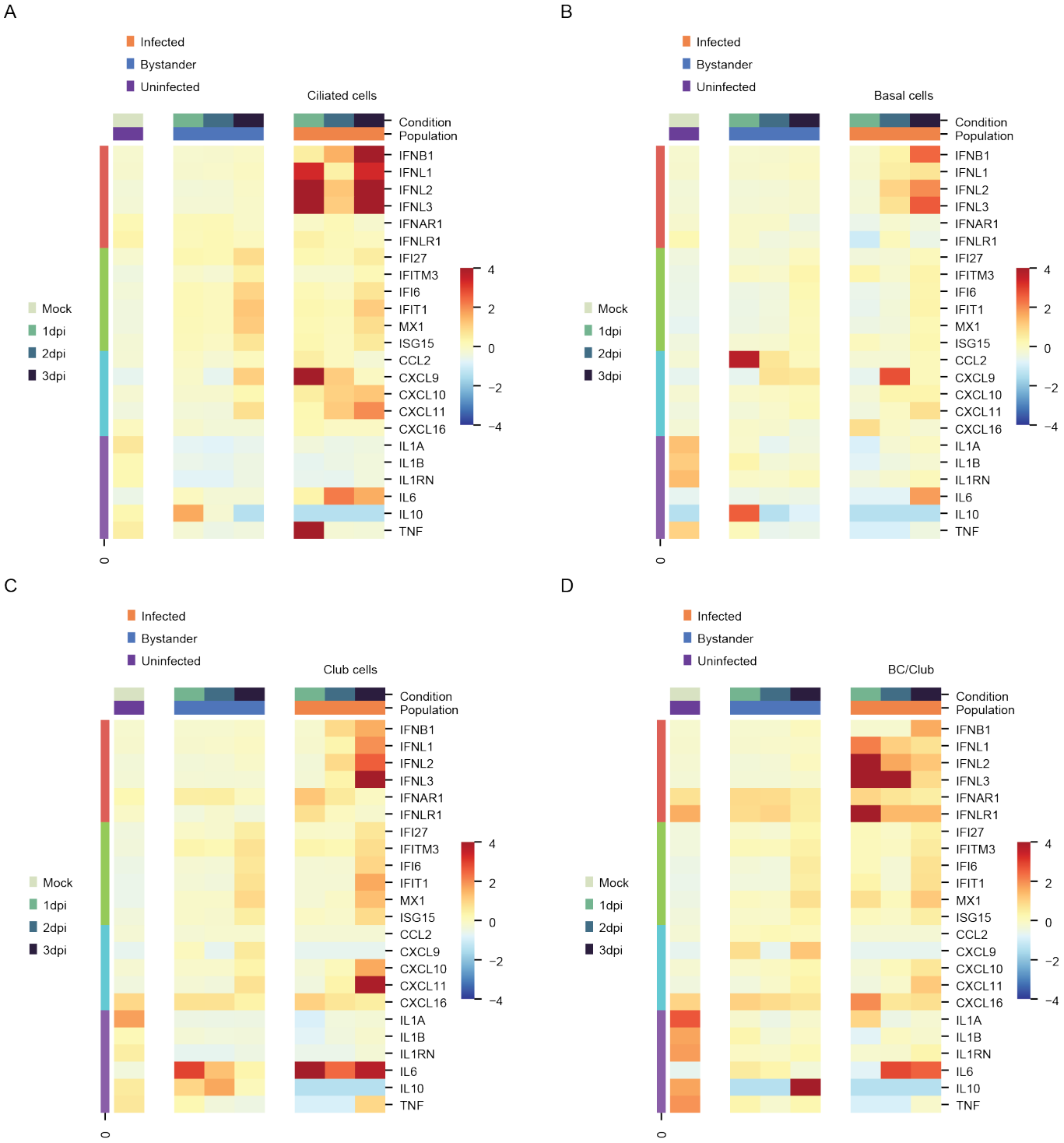

Figure 4: *Caption next page.*

Figure 4: Innate immunity markers in SARS-CoV-2 infection. **A-D.** Heatmaps of cytokines, chemokines, interferons and interferon-stimulated genes in ciliated (**A.**), basal (**B.**), club (**C.**) and BC/Club cells (**D.**)

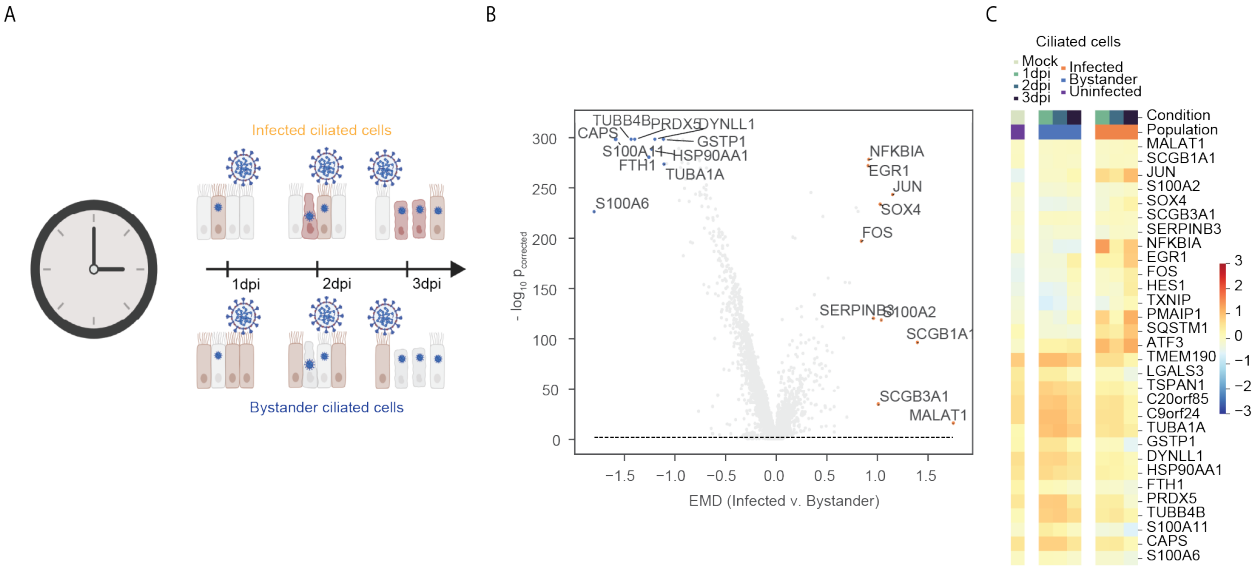

Figure 5: Expression of differentially expressed genes. **A.** Schematic of the differential expression analysis. Two main cell populations are observed : bystander cells that were not infected by the virus at 3 days post infection (dpi) and infected cells that contain active viral replication and transcription at 3 dpi. **B.** Volcano plots highlighting the most differentially expressed genes between infected and bystander cells in ciliated cells at 3 dpi as measured by earth mover's distance (EMD). **C.** Heatmap of the most differentially expressed genes between uninfected, infected and bystander cells in ciliated cells in all conditions.

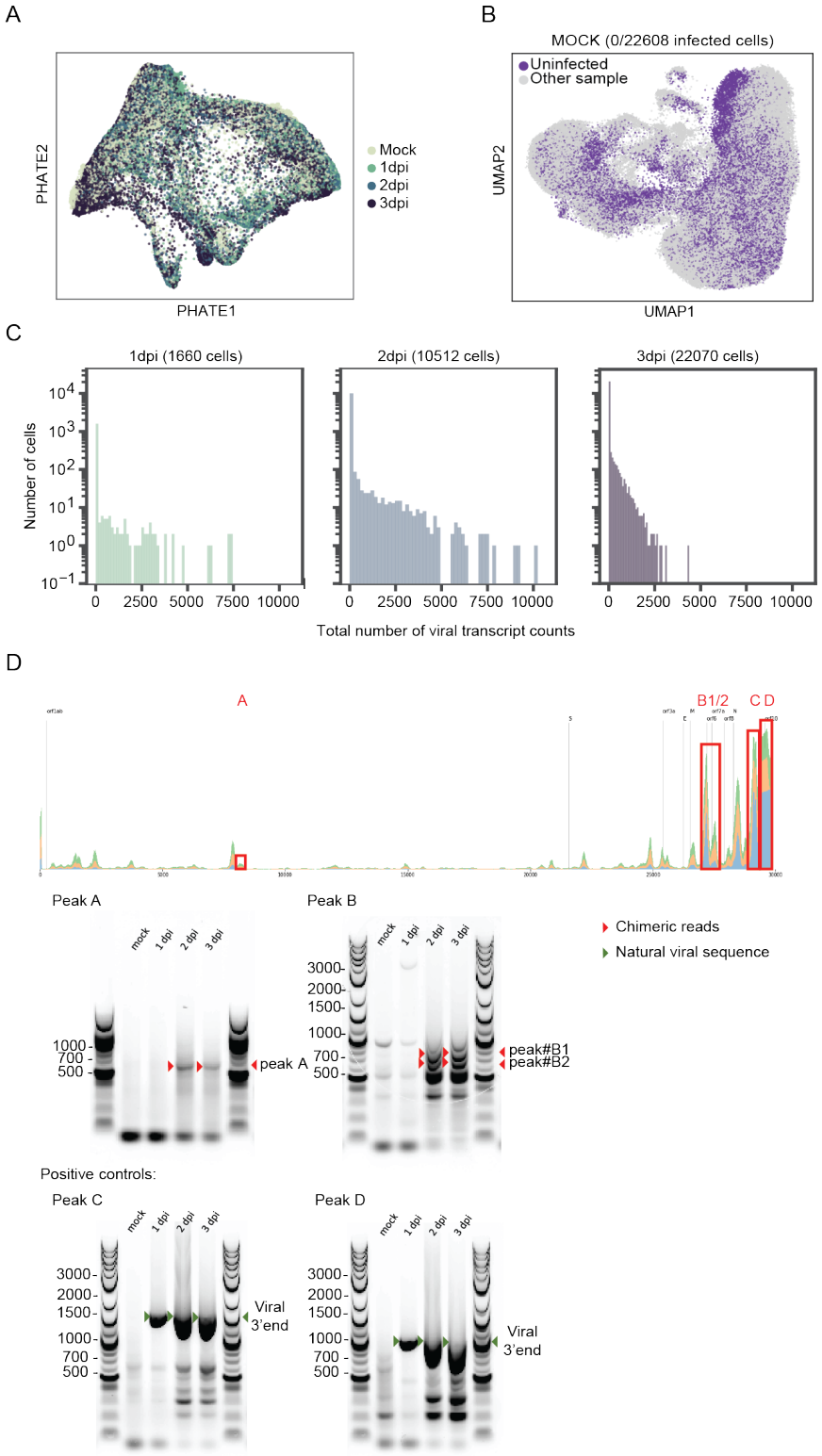

Figure S1: Caption next page.

Figure S1: SARS-CoV-2 viral genome transcript counting to determine infections state. **A.** PHATE visualization of the scRNA-seq gene counts after batch correction. Each point represents a cell. Cells were colored according to their samples. **B.** UMAP visualization of the mock sample, 1 dpi, 2 dpi and 3 dpi are represented in Figure 1. **C.** Histograms of viral transcript counts per cell on a logarithmic scale for each condition. **D. Top panel**: Schematic of reads aligning across the SARS-CoV-2 genome. Red boxes indicate mapped reads validated with RT-PCR, including two previously unreported poly-adenylated transcripts (A, B). **Bottom panel**: RT-PCR spanning the junctions between poly-A tails and SARS-CoV2 genome body for non-canonical transcripts (Peaks A, B) and two positive controls (Peaks C, D). The products were run on agarose gels. Red arrowheads denote the expected amplicons for novel transcripts, while green arrowheads denote amplicons for the natural viral 3'end.

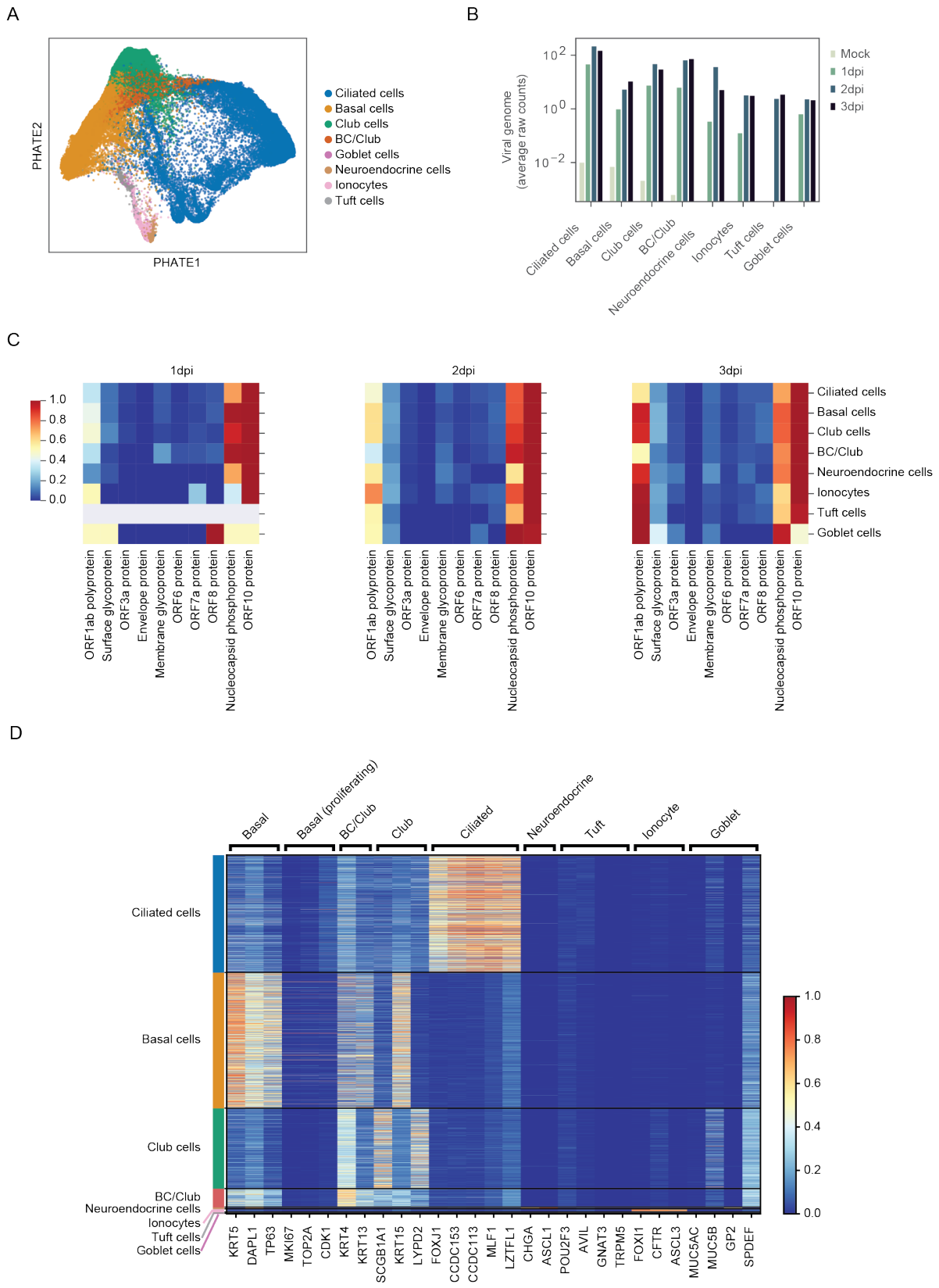

Figure S2: Caption next page.

Figure S2: SARS-CoV-2 cell tropism and SARS-CoV2 sub-genomic expression across bronchial epithelial cell types. **A.** PHATE visualization of the cell types. **B.** Histogram of the average raw counts of viral transcripts per cell type across conditions. **C.** Normalized heatmap of the viral Open Reading Frame (ORF) counts in each cell type across three conditions: 1 dpi, 2 dpi and 3 dpi. **D.** Heatmap displaying expression of marker genes for each cell types and SARS-CoV-2 putative relevant genes.

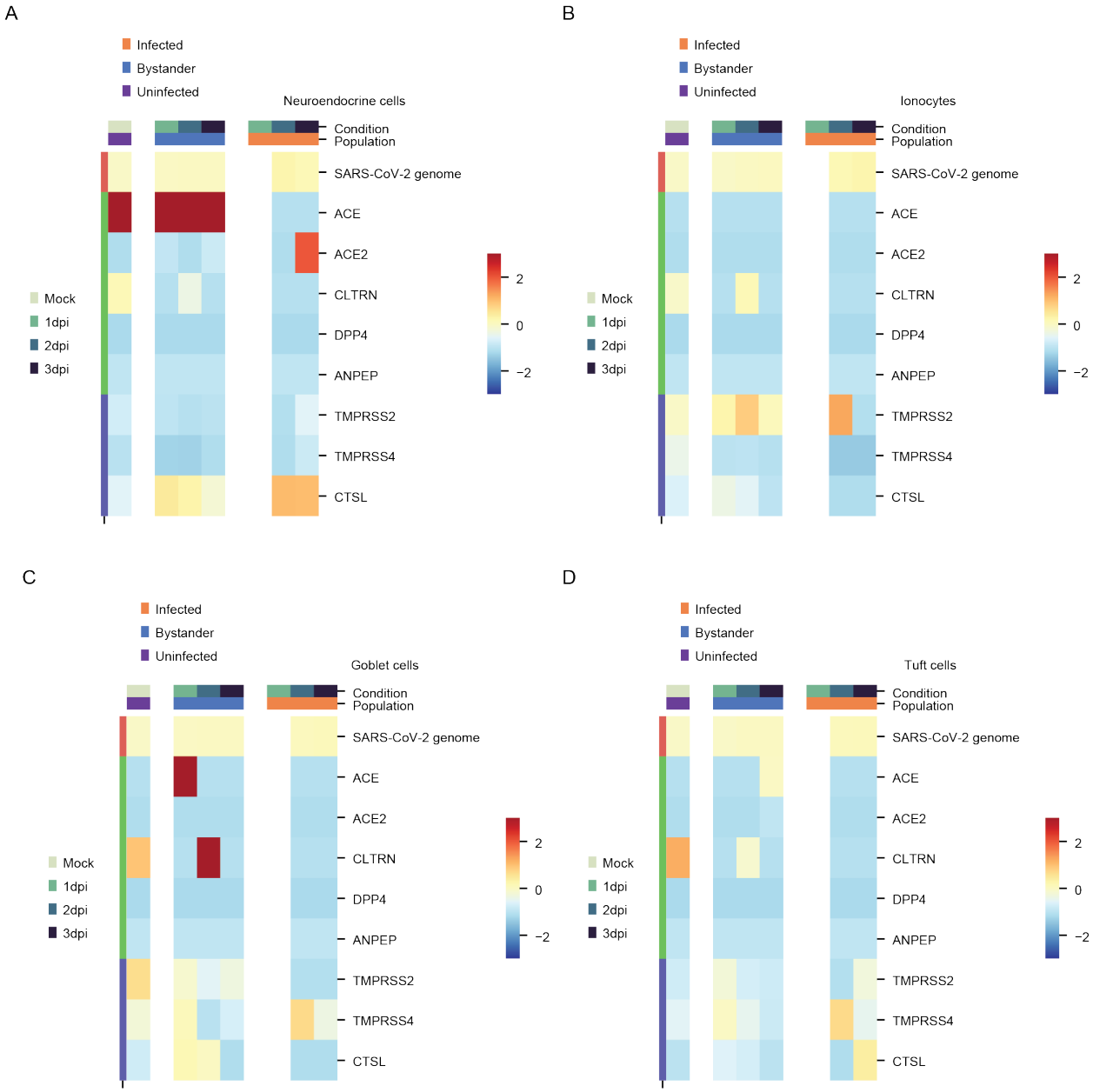

Figure S3: *Caption next page.*

Figure S3: Expression of known entry determinants across bronchial epithelial cell types. **A-D.** Heatmaps of receptors and proteases in neuroendocrine cells cells (**A.**), ionocytes (**B.**), goblet cells (**C.**) and tuft cells (**D.**).

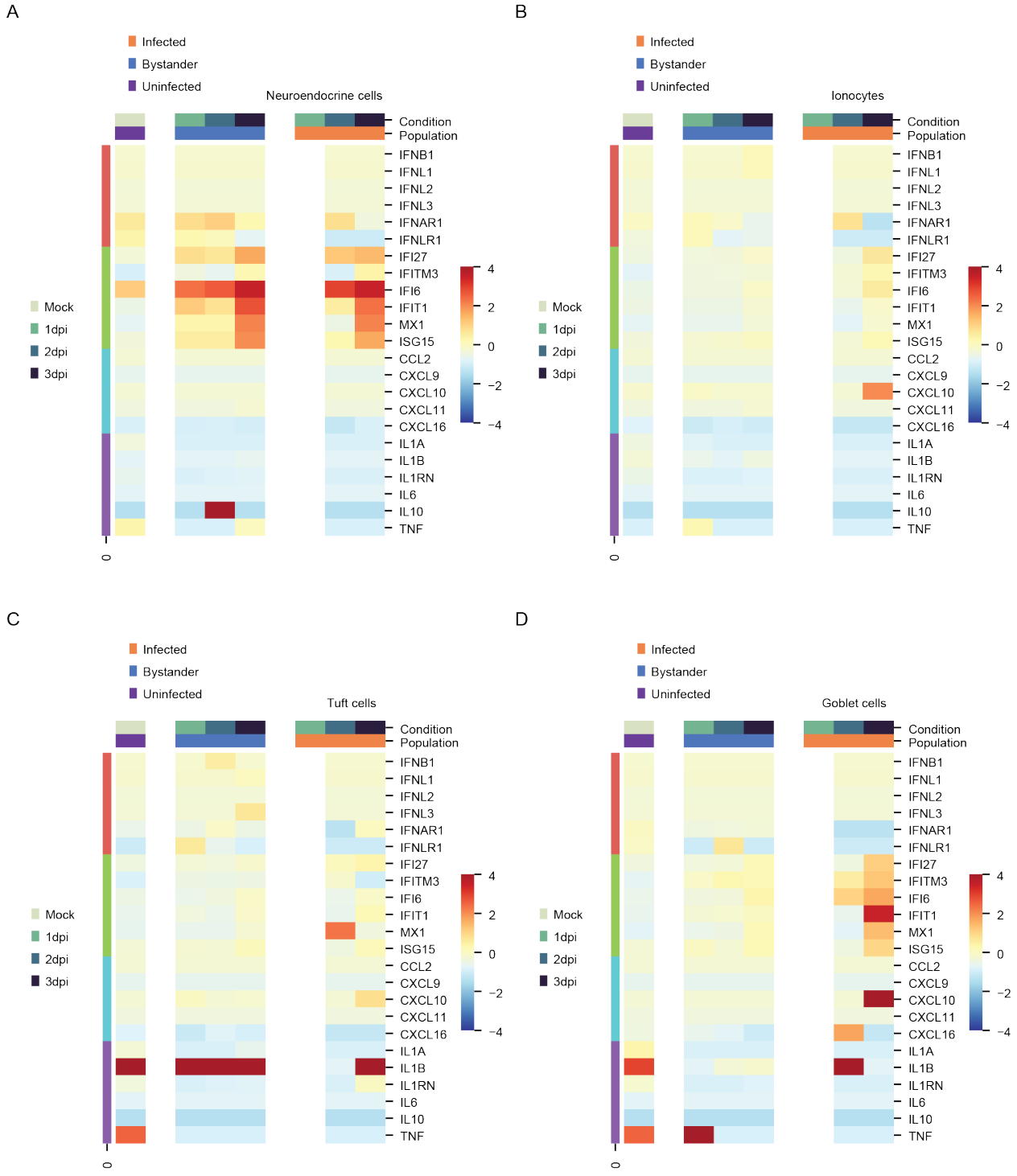

Figure S4: Innate immunity markers in SARS-CoV-2 infection. **A-D.** Heatmaps of cytokines, chemokines, interferons and interferon-stimulated genes in neuroendocrine (**A**), ionocytes (**B**), goblet (**C**) and tuft cells (**D**).

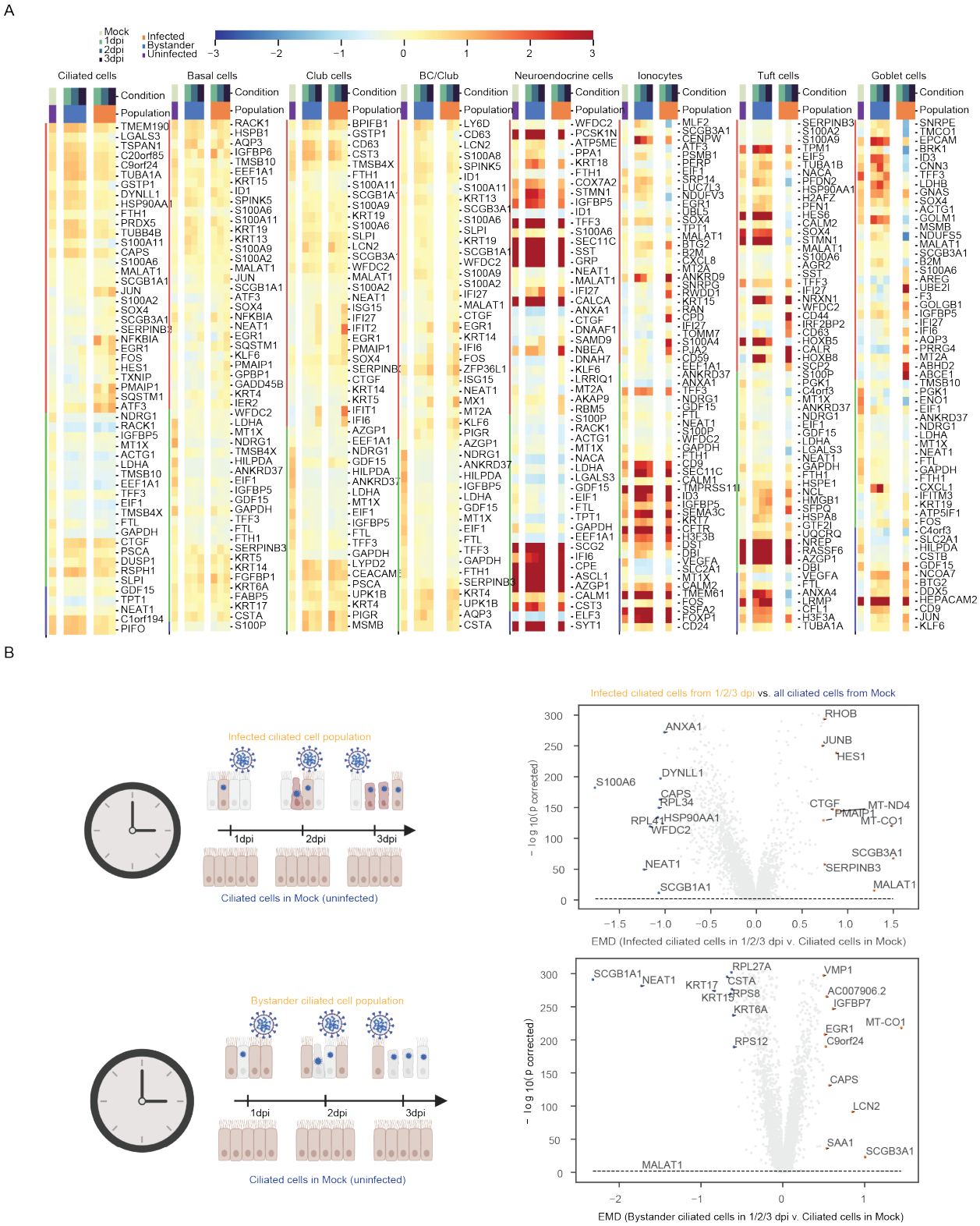

Figure S5: *Caption next page.*

Figure S5: Expression of differentially expressed genes. **A.** Normalized heatmaps of the 30 most differentially expressed genes between uninfected, bystander and uninfected cells across cell types and conditions. Left to right : basal cells, BC/Club cells, club cells, neuroendocrine cells, ionocytes, goblet cells and tuft cells. **B.** Volcano plots highlighting the most differentially expressed genes between bystander and uninfected cells in ciliated cells.