Longitudinal host transcriptional responses to SARS-CoV-2 infection in adults with extremely high viral load

Vasanthi Avadhanula[1†], Chad J. Creighton [2,4,5†] Laura Ferlic-Stark[1], Richard Sucgang[7], Yiqun Zhang[2], Divya Nagaraj[1], Erin G. Nicholson[1,6], Anubama Rajan[1], Vipin Kumar Menon[3,5], Harshavardhan Doddapaneni[3,5], Donna Marie Muzny[3,5], Ginger Metcalf[3], Sara Joan Javornik Cregeen[1], Kristi Louise Hoffman[1], Richard A Gibbs[3,5], Joseph Petrosino[1], Pedro A Piedra[1,6]

1. Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, 77030, USA.
2. Dan L. Duncan Comprehensive Cancer Center Division of Biostatistics, Baylor College of Medicine, Houston, TX 77030, USA.
3. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA.
4. Department of Medicine, Baylor College of Medicine, Houston, TX 77030, USA.
5. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA.
6. Department of Pediatrics, Baylor College of Medicine, Houston, TX, 77030, USA.
7. Center for Health Data Science and Analytics, Houston Methodist Research Institute, Houston, TX 77030, USA

† co-first authors

Correspondence to: Pedro A Piedra @ ppiedra@bcm.edu and Chad Creighton at creighto@bcm.edu

Abbreviations: COVID-19, Coronavirus disease 19; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2,

Abstract:

Current understanding of viral dynamics of SARS-CoV-2 and host responses driving the pathogenic mechanisms in COVID-19 is rapidly evolving. Here, we conducted a longitudinal study to investigate gene expression patterns during acute SARS-CoV-2 illness. Cases included SARS-CoV-2 infected individuals with extremely high viral loads early in their illness, individuals having low SARS-CoV-2 viral loads early in their infection, and individuals testing negative for SARS-CoV-2. We could identify widespread transcriptional host responses to SARS-CoV-2 infection that were initially most strongly manifested in patients with extremely high initial viral loads, then attenuating within the patient over time as viral loads decreased. Genes correlated with SARS-CoV-2 viral load over time were similarly differentially expressed across independent datasets of SARS-CoV-2 infected lung and upper airway cells, from both in vitro systems and patient samples. We also generated expression data on the human nose organoid model during SARS-CoV-2 infection. The human nose organoid-generated host transcriptional response captured many aspects of responses observed in the above patient samples, while suggesting the existence of distinct host responses to SARS-CoV-2 depending on the cellular context, involving both epithelial and cellular immune responses. Our findings provide a catalog of SARS-CoV-2 host response genes changing over time.

Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is the etiologic agent of the coronavirus disease 2019 (COVID-19) pandemic. The clinical spectrum of COVID-19, caused by SARS-CoV-2, is wide, ranging from asymptomatic infection to fatal disease. Risk factors for severe illness and death include age, sex, smoking, and comorbidities, such as obesity, hypertension, diabetes, and cardiovascular disease. Studies suggested that SARS-CoV-2 viral load can predict the likelihood of disease spread and severity [1–3]. A higher detectable SARS-CoV-2 plasma viral load was associated with worse respiratory disease severity [4]. Conversely, robust immune responses putatively mediate non-severe illness, in part, by controlling the replication of SARS-CoV-2 [5,6]. Emerging evidence indicates that age and sex differences in the innate and adaptive immune response can explain the higher risks observed in older adults and male cases [7,8].

Initial site of SARS-CoV-2 replication is the upper respiratory tract, and replication usually peaks within the first week of infection [6]. The amount of virus produced at the respiratory epithelium is considered to be a critical element in determining SARS-CoV-2 transmissibility, duration of illness or severity, although it is not the only factor [9,10]. Higher viral loads have been observed in hospitalized patients with severe disease, have been attributed to high transmission and superspreading events, and have resulted in prolonged viral RNA shedding [1,11–15].

Specific anatomic site or host cell type where viral replication occurs, can also determine the course of infection. For example, angiotensin-converting enzyme 2 (ACE-2) and transmembrane serine protease 2 (TMPRSS2) receptors expression is highest in the upper respiratory tract and decreases in the distal or lower respiratory tract, incidentally SARS-CoV-2 infection mirrored this pattern, with high replication in proximal (nasal) versus distal pulmonary (alveolar) epithelial cells [16]. Control of viral replication and resolution of the inflammatory

response is believed to be dependent, in part, on viral load and route of infection as well as the host immune response [17]. The early host immune response is regulated closely by the epithelial cell cytokine signaling in response to active viral replication [18]. Rapid and robust activation of the antiviral innate immune response at the site of viral replication is required to control and clear the virus. A delayed cytokine response can result in prolonged viral replication and worst clinical outcome as seen for other respiratory viruses [19]

Our understanding of the viral dynamics of SARS-CoV-2 and host responses driving the pathogenic mechanisms in COVID-19 is evolving rapidly. Multiple studies have reported various characteristics of immune/inflammatory responses to SARS-CoV-2. Cytokine or chemokines-related host inflammatory responses such as CCL2/MCP-1, CXCL10/IP-10, CCL3/MIP-1A, and CCL4/MIP-1B were detected in bronchoalveolar lavage samples of SARS-CoV-2 infected adults while activation of apoptosis and the P53 signaling pathway were observed in lymphocytes [20]. Inflammatory cytokine such as IL-1, IL-18, and IL-33 were enriched in the airways of COVID-19 patients [21]. In addition, a shotgun host transcriptomic analysis on nasopharyngeal samples revealed a wide range of antiviral responses. These included gamma and alpha interferon responses, elevated levels of ACE-2, interferon stimulated genes (ISGs), and interferon inducible (IFI) genes [22]. Very few studies have demonstrated the temporal correlation between viral load and host gene expression. Variation in viral load was associated with the SARS-CoV-2 disease and the host response dynamics via innate and adaptive immunity (To et al., 2020). Another study revealed that expression of interferon-responsive genes, including ACE-2, increased as a function of viral load, while transcripts for B cell–specific proteins and neutrophil chemokines were elevated in patients with lower viral load [23]. Rouchka et.al. reported that cellular antiviral responses strongly correlated with viral loads. However, COVID-19 patients who experienced mild symptoms had a higher viral load than those with severe complications[6].

We previously reported on a small group of adults with extremely high SARS-CoV-2 viral load, who had the potential to be super spreaders and a large group of adults with low SARS-CoV-2 viral load, both groups had mild illness[14] . Here, we wanted to determine the host response in relation to the viral load early during infection. We conducted a longitudinal study to investigate gene expression patterns detected in the secretion of the nasal epithelium during the acute phase of SARS-CoV-2 infection. The cases included SARS-CoV-2 infected individuals with an extremely high viral load early in their illness matched to individuals who either had a low SARS-CoV-2 viral load early in their infection or were otherwise stable patients who tested negative for SARS-CoV-2 prior to their outpatient surgical or aerosol generating procedure. We also determined the transcriptional response of a human nose organoid (HNO) line infected with SARS-CoV-2 and compared it to transcriptomic profiles generated from the upper respiratory tract secretion collected by nasal swabs from SARS-CoV-2 infected individuals.

Results

Study cohort

Ten SARS-CoV-2 cases were randomly selected from our population of adults with extremely high viral load (Ct <16 to N1 target) at the time of their first RT-PCR positive test. For each high viral load case, two additional human subjects were matched based on gender, week of first SARS-CoV-2 RT-PCR test, age, and home zip code. These additional subjects consisted of either 1) SARS-CoV-2 infected adults with low viral load (Ct 31-<40) (SARS-CoV-2 low viral load case) or 2) stable adults who were SARS-CoV-2 RT-PCR negative (SARS-CoV-2 negative control) for their out-patient surgical or aerosol generating procedure. Each high viral load case had two to three subsequent SARS-CoV-2 RT-PCR positive mid-turbinate (MT) swab samples collected over a 4-week period. Each SARS-CoV-2 low viral load case had similarly spaced SARS-CoV-2 positive MT swab samples matched to its respective extremely high viral load

case. On the other hand, the SARS-CoV-2 negative control only had one MT-swab sample collected with no longitudinal follow-up and was used to establish the transcriptomic baseline in the respiratory epithelium during the time the extremely high and low viral load matched cases were identified. The demographic and visit characteristics for the cohort is presented in Table 1. In general, age, gender, race, ethnicity, and zip code were comparable between the extremely high viral load, low viral load, and SARS-CoV-2 negative adults. The adults in the SARS-CoV-2 negative group were mostly asymptomatic at the time of testing, although their demographic information was not significantly different compared to that of both the SARS-CoV-2 extremely high and low viral load groups. The median Ct value difference between the extremely high and low viral load groups were 794,672-fold (19.6 Ct difference) and 724-fold (9.5 Ct difference) different at Visit 1 and Visit 2, respectively. At Visit 3, approximately 14 to 17 days after their Visit 1, the Ct values were comparable between the two groups.

RNA sequencing of serially collected specimens.

Of the 73 MT swab samples from the extremely high and low viral load SARS-CoV-2 groups with longitudinal follow-up and SARS-CoV-2 negative controls, only 44 (60.3%) MT swab samples from 20 (66.7%) individuals were of good quality to generate RNA-sequence data to study the host response to SARS-CoV-2 infection over time (Table 2). Demographic factors such as age, gender, race, ethnicity, zip code, disease severity and co-morbid conditions were comparable between the extremely high viral load, low viral load groups, and SARS-CoV-2 negative control group. Host response data were available on eight cases (extremely high viral load) with 23 samples. Six of the 8 extremely high viral load cases had gene expression data for Visits 1, 2, and 3, and two others for Visit 1 and 3. On the other hand, eight low viral load cases had 17 samples with gene expression data. Only two of the low viral load cases had gene

expression data for Visit 1, 2, and 3. Another two low viral load cases had gene expression data at Visit 1, Visit 2 or 3, and Visit 4. The remaining four low viral load cases had gene expression data at Visit 1 only (n=1), Visit 2 only (n=2) or Visit 1 and 2 (n=1). Only 4 of the 10 SARS-CoV-2 negative control adults had gene expression data. All together 44 MT swab samples were sequenced for RNA to observe gene expression changes in the host response of the cases with extremely high viral load over time, as compared to the SARS-CoV-2 low viral load matched cases and the negative controls.

Gene expression changes by viral load

From our RNA-seq dataset, we could identify widespread gene expression changes from the nasal epithelium attributable to transcriptional host responses to SARS-CoV-2 infection. By comparing the expression levels of each gene with the sample viral load (representing the inverse correlation with Ct value) across the 44 MT swab samples, 425 genes were statistically correlated at $p<0.01$ significance level and 112 genes at $p<0.001$ (Figure 1a, Pearson's correlation). A stricter statistical cutoff would involve fewer expected false positive genes from multiple testing. However, the above 425 genes with $p<0.01$ would still be highly enriched for true positives, as revealed by integrating these genes with information from external databases, as described below. We also compared the expression levels of genes at individual time points during infection of both the extremely high viral load and low viral load groups with the SARS-CoV-2 negative control group (Figure 1b). Comparing Visit 1 MT swab samples from the extremely high viral load cases (n=8 samples from eight subjects) with the MT swab samples in the SARS-CoV-2 negative control group (n=4) yielded the highest number of genes with statistically significant correlated expression, as opposed to comparisons involving later times for the extremely high viral load group or involving the low viral load group. The gene expression from the extremely high viral load cases at Visit 1 highly overlapped with the differentially expressed genes of the low viral load group at Visit 1 (Figure 1c) and remained highly correlated

throughout their last visit. Interestingly, genes from the extremely high viral load group that did not overlap with the low viral load group did not show significant overlap with information from external databases.

To further delineate the differences in host gene expression between extremely high and low SARS-CoV-2 viral load groups, we performed an upset plot analysis to identify unique and common intersecting genes between the samples (Figure S1). Among all the differentially expressed genes (DEGs) in the samples, 614 DEGs were unique to the subjects in the extremely high viral load group at visit 1 (first visit) and 226 genes were unique for the extremely high viral load at the last visit. The low viral load subjects on the first and last visit showed 157 and 93 unique DEGs respectively. There were 31 DEGS that were common between all the groups. We performed the Gene ontology (GO) analysis of the unique and overlapping DEGs sets, and we found significant enrichment (FDR <0.05, count =3) of the biological processes including defense response to virus, negative regulation of viral genome replication, innate immune response, response to virus (Figure S2) that were uniquely expressed in the extremely high viral load group at visit 1. SARS-CoV-2 infection in the low viral load group at either the early or later phase of the infection and the extremely high viral load group at the last visit did not show statistically significant enrichment of GO biological process. These findings indicate that subjects with extremely high viral load at their initial visit were responding to the infection with increased immune responses, and thus preventing prolonged viral infection with a poor prognosis.

Differentially expressed gene in respiratory samples from extremely high viral load adults

Focusing on the 112 top gene expression correlates of viral load across the 44 MT swab samples ($p<0.001$, Pearson's), 108 of these genes were higher in the SARS-CoV-2 infected adults with extremely high viral load. When visualizing the differential expression patterns of

these 108 genes by heat map (Figure 2a), the genes were highest at Visit 1 of the extremely high viral load group, then decreased in expression with subsequent time points, tracking with the decrease in viral load (i.e., increase in Ct value). The 108 genes showed intermediate relative expression levels in the low viral load group and low expression in the SARS-CoV-2 negative control group. The 367 genes increased with extremely high viral load at $p<0.01$ were highly enriched for functional gene categories, as defined by GO annotation terms. Enriched GO terms (Figure 2b, $p<=3E-5$, one-sided Fisher's exact test) included 'immune system process', 'response to virus', 'type I interferon signaling pathway', 'cytokine-mediated signaling pathway', 'response to stress', 'regulation of viral life cycle', 'immune response', 'response to cytokine', 'innate immune response', 'response to interferon-gamma', 'regulation of I-kappaB kinase/NF-kappaB signaling', 'JAK-STAT cascade', 'protein ubiquitination', 'regulation of cell death', 'T cell activation', 'vesicle-mediated transport', and 'complement activation'. Of the 17 functional gene categories, there were five gene categories - 'response to virus', 'type 1 interferon signaling', 'regulation of viral life cycle', 'response to interferon-gamma', and 'JAK-STAT cascade' – where approximately 20% or higher of the genes were over expressed for that pathway. Overall, the above gene categories were highly indicative as representing a host immune response to an acute viral infection. Some of the genes that were upregulated were *EIF2AK2* (eukaryotic translation initiation factor 2 alpha kinase 2), and *ZC3HAV1* (zinc finger CCCH-type containing, antiviral 1), which have anti-viral activity. Other genes like *IFIT2* and *IFIT3* (interferon induced protein with tetratricopeptide repeats) aid in apoptosis. Chemokine genes like *CXCL9* and *CXCL10* that are involved in T-cell trafficking were also highly expressed. In contrast, the 62 genes decreased with high viral load at $p<0.01$ were not highly enriched for GO terms. Some of the genes that were downregulated included *OR4A16* and *OR10X1*, involved in olfactory responses; *SALL3* and *MAGB6*, which aid in downregulation of transcription; and *TUBA3E*, *MLN*, and *ISTN1*, affecting tubulin functions.

Comparison of the top over expressed genes in respiratory samples from the extremely high and low viral load groups to other published data sets

Genes correlated with SARS-CoV-2 viral load over time were similarly differentially expressed across independent datasets of SARS-CoV-2 infected lung and upper airway cells (Figure 3). We examined differential expression patterns for the top 112 genes, at $p<0.001$ significance level, correlated with SARS-CoV-2 viral load across our serial sampling cohort (by Pearson's) in two independent RNA-seq datasets of SARS-CoV-2 infection: one of lung cancer cell lines A549 and Calu-3 infected with SARS-CoV-2 for 24 hours from Blanco-Melo et al [24]. and one of nasopharyngeal/oropharyngeal samples in 238 patients with COVID-19, other viral, or non-viral acute respiratory illnesses from Mick et al [25]. As a group, the genes that positively correlated with SARS-CoV-2 viral load were increased in SARS-CoV-2-infected Calu-3 cells and were high in samples of human subjects infected with SARS-CoV-2 or other viruses (Figure 3a). For the Mick et al. dataset, SARS-CoV-2 viral load data was available. Of the 112 genes correlated with viral load in our dataset, 105 were in common with the Mick et al. dataset, and 99 (94%) of these genes were positively correlated (Pearson's $p<0.05$) with viral load across the 94 SARS-CoV-2 infected patients. In contrast to Calu-3, A549 infected cells did not show as strong a correspondence to our 112-gene signature pattern. Taking the top genes that correlated positively with SARS-CoV-2 viral load across the Mick et al. patient samples ($p<0.01$, Pearson's correlation) and the top genes over-expressed in SARS-CoV-2-infected Calu-3 cells ($p<0.01$, t-test), these significantly overlapped with the genes that positively correlated ($p<0.01$, Pearson's) with SARS-CoV-2 viral load across our serially collected MT swab samples with a high overlap among the respective dataset results (Figure 3b). The 136 genes overlapping among all three datasets involved cytokines and inflammatory response pathways. In contrast, there was limited overlap among the datasets involving genes under-expressed with SARS-CoV-2 infection between our data set to either Mick *et al.* or Calu-3 cells (Figure 3c). This in part may reflect the

potential differences in the respective SARS-CoV-2 variants causing the infection or differences in the illness severity of the host.

Comparison of our respiratory sample gene sets to the transcriptional response of the human nose organoid infected with SARS-CoV-2

As another means to identify host transcriptional responses to SARS-CoV-2 infection, we generated RNA-seq data on the human nose organoid model HNO [26]. We sampled HNO cells infected with SARS-CoV-2 and mock control cells at 6hrs, 72hrs, and 6 days post-infection, and we profiled these samples for gene expression. In the HNO204 RNA-seq dataset, 1760 genes were statistically significant at $p<0.05$ significance level and 341 genes, at $p<0.01$, exceeding chance expected. The top 867 genes over-expressed in HNO with SARS-CoV-2 infection ($p<0.05$, t-test) showed significant overlapping patterns with the above-mentioned independent RNA-seq datasets of SARS-CoV-2 infection (Figure 4a). Only a small, albeit statistically significant, fraction of the HNO204 over-expressed genes overlapped with the top 367 genes that correlated positively with SARS-CoV-2 viral load in our serial MT swab dataset (Figure 4b). Of the 867 overexpressed genes in SARS-CoV-2 infected HNO, 35 overlapped with the 367 over expressed genes in the respiratory samples of extremely high and low viral load groups ($p=1E-5$, one-sided Fisher's exact test). At the same time, a substantial fraction of the 867 HNO genes overlapped with the genes high with SARS-CoV-2 infection in both A549 and Calu-3 lung cancer cell lines (Figures 4a and 4b), with 178 Calu-3 genes overlapping ($p<1E-20$, one-sided Fisher's exact test). In contrast, little overlap was observed between the genes under-expressed with SARS-CoV-2 infection in HNO and genes similarly under-expressed with SARS-CoV-2 in the other datasets (Figure 4c).

The 867 genes over-expressed in HNO at $p<0.05$ were significantly enriched for functional GO gene categories. Enriched GO terms (Figure 4d, $p<=0.0001$, one-sided Fisher's exact test)

included 'vesicle', 'extracellular vesicle', 'intracellular vesicles', 'MAP kinase phosphatase activity', 'regulation of locomotion', 'peptidase activator activity', 'endosome membrane', 'regulation of smooth muscle cell proliferation', 'regulation of cell motility', 'proteasomal protein catabolic process', 'negative regulation of signaling', 'programmed cell death' , proteolysis involved in cellular protein catabolic process', and 'inactivation of MAPK pathway'. Overall, the over expressed genes are representative of the regulation of extracellular signaling from virus infection on a wide range of cellular responses and function. The above findings of the HNO transcriptional response to SARS-CoV-2 in relation to transcriptional responses observed in other models and patient samples would suggest the existence of distinct host responses to SARS-CoV-2 depending on cellular context, such as we previously observed between A549 and Calu-3 lung cancer cell lines. The host response observed in HNO is reflective of a complex epithelial cell population responding to a SARS-CoV-2 infection. On the other hand, the host response genes detected in the upper respiratory tract secretion of our prospective longitudinal cohort and those of Mick *et al*. patient samples are a composite of the epithelial and cellular immune responses to the viral infection.

Discussion

The primary site for SARS-CoV-2 replication is thought to be the ciliated cells in the nasopharynx or nasal olfactory mucosa. The viral replication initiates a signaling cascade to promote the production of interferons and chemokines by epithelial cells and thereby promote immune cell activation to control the virus. SARS-CoV-2 infection causes upregulation of cytokines including IL-2, IL-6, IL-10, IL-12 and MCP-1 detected in tissues and serum, as well as infiltration of infected tissues by inflammatory cells such as macrophages [27]. In the present study, RNA seq analysis of MT swabs from SARS-CoV-2 infected individuals identified robust induction of interferon inducible, cytokine, stress response, and immune-related genes. A variety of genes such as *OAS2*, *PARP9*, *OASL*, *IFIT2*, *IFI3*, *CCL8*, *CXCL10*, etc., were highly

upregulated and correlated with high viral load, suggesting that innate immune response genes were activated in a viral load dose response manner to control the viral infection. These results are very consistent with recent studies from upper respiratory tract samples, which reported upregulation of anti-viral factors and interferon response pathways [22,23,28] .

In our study samples, the numbers of genes that were upregulated were much higher compared to down regulated genes (367 vs 62). Some of genes that were downregulated included those which operate olfactory functions (*OR4A16* and *OR10X1*), downregulation of transcription (*SALL3* and *MAGB6*), and tubulin functions (*TUBA3E* and *MLN*, and *ISTN1*). Previous studies have reported larger numbers of down regulated host response genes especially involving olfactory receptor pathway, neutrophil degranulation, and vesicle formation—indicating the role of these genes in loss of olfactory function in SARS-CoV-2 infections as well as the viral control of host-cell machinery [20,22,23]. One other study also showed very low number of downregulated genes with SARS-CoV-2 infection [29], one reason for the low number of downregulated genes observed in our longitudinal study could conceivably relate to the mild illness experienced by both the extremely high and low viral load groups, in addition to the timing of sample collection as compared to other studies as well as the SARS-CoV-2 variants respectively involved.

Remarkably, the highest number of significant expressed genes were driven by the extremely high viral load group at Visit 1 (first visit). Also, all the genes that were upregulated with the low viral load group at Visit 1 completely overlapped with the extremely high viral load group at Visit 1 except for one gene, -*CNN2*, which plays a role in cell adhesion and muscle contraction. The predominant sets of genes involved in defense response to virus, type I interferon signaling pathway, cytokine-mediated signaling pathway—such as *CXCL10*, *TGFB*, *IFIT2*, *IFIT3*, *OAS1*, and *IRF1*—were not found significantly upregulated in the low viral load group. Consonant with this, Rouchka *et. al*. also observed that subjects with high viral loads had robust interferon and cellular anti-viral response and even exhibited strong inverse correlation with disease severity [6].

We previously noted that some SARS-CoV-2 infected adults with low viral load experienced prolonged viral shedding and low fluctuation in viral load over time [14]. Absence or low expression of the anti-viral response in the low viral load group strengthens our observation of prolonged shedding in adults with a low viral load early in infection.

In our longitudinal study, the up-regulated host response genes that correlated with SARS-CoV-2 viral load over time in the respiratory secretion collected by the MT swabs were similarly differentially expressed across independent data sets of SARS-CoV-2 infected lung and upper airway cells [24]. About 170 of the differentially expressed genes observed in our study overlapped with SARS-COV-2 infected Calu-3 lung adenocarcinoma cell line but not with A549 cells. The observed difference across the cell lines could possibly be attributed to A549 cells not supporting robust replication of SARS-CoV-2 due to the low expression of ACE-2 [30]. Similarly, 207 up-regulated genes from our longitudinal study overlapped with nasopharyngeal swabs from SARS-CoV-2 infected patients (3). Genes involved in cytokines and inflammatory response pathways were the ones that overlapped the most, demonstrating that anti-viral innate immune responses are common with SARS-CoV-2 infections. In addition, the up-regulation of differentially expressed genes related to an inflammatory response in COVID-19 patients can result in the induction of interleukin-6 (IL-6), CXCL10 (IP-10), and TNF-α with hyperactivation of Th1/Th17 responses that results the recruitment and activation of pro-inflammatory neutrophils and macrophages into the airways [31]. This has been proposed as the prime reason for failure to resolve inflammation in severely symptomatic patients [31,32].

To better understand the contribution of epithelial cellular responses to SARS-CoV-2, we compared differentially expressed genes in the respiratory secretion of adults infected with SARS-CoV-2 to those that were expressed in HNO infected with SARS-CoV-2. A small, albeit statistically significant, fraction (35 of 867) of the HNO up-regulated genes overlapped with the 367 differentially expressed up-regulated genes detected from the SARS-CoV-2 cases from our

longitudinal cohorts. These included functional genes involved with intrinsic antiviral immunity and interferon signaling representing the epithelial cellular responses to SARS-CoV-2 infection. A greater number of up-regulated genes overlap between our longitudinal cohorts [170 (46.3%) of 367 genes] and the SARS-CoV-2 infected Calu-3 cell line [170 (9.3%) of 1836 genes] compared to our SARS-CoV-2 infected HNO204 line [35 (4.0%) of 867 genes]. This could reflect the difference in cellular complexity between the cell lines and greater diversity of the HNO epithelium resulting in fewer overlapping up-regulated genes. HNO204 is a complex pseudostratified epithelium composed of at least 9 different cell types including ciliated, goblet, secretory and basal cells [33,34]. In contrast, the Calu-3 cell line, was generated from a bronchial adenocarcinoma, a submucosal gland cell line of a single cell type [35]

Previous studies have demonstrated high expression of ACE2 in SARS-CoV-2 infected nasopharyngeal samples and these were greatly elevated in high viral load subjects, suggesting that higher replication occurs with increased receptor expression [22]. In our cohort we did not observe a statistically significant increase in *ACE2* expression in both extremely high and low viral load groups. However, the expression of *ACE2* was elevated in our HNO infected with SARS-CoV-2 but not *TMPRSS2*, which has increased expression in nasal airway epithelial brushings [36].

In summary, our longitudinal study investigated gene expression patterns in SARS-CoV-2 infected individuals with an extremely high viral load displayed strong immune responses that decreased over time, and eventually became comparable to those with low viral loads. We detected hundreds of up-regulated genes that were highly correlated to the SARS-CoV-2 viral load. Enriched cellular pathways involved in the innate immune response, antiviral interferon responses were observed in other cohorts of SARS-CoV-2 infected adults. A limited but highly significant up-regulated gene response overlapped with our human nose organoid line, a complex pseudostratified ciliated epithelium, suggesting that the gene expression profile

371 detected in SARS-CoV-2 infected adults is generated from both the epithelial and cellular
372 immune responses. In conclusion, high SARS-CoV-2 viral loads primarily elicit a heightened
373 host immune response for the control of viral replication and clearance.

374

Materials and Methods

Study cohort

Ten extremely high, viral load SARS-CoV-2 positive cases were matched to 10 low viral load SARS-CoV-2 positive adults, and 10 stable adults (SARS-CoV-2 negative controls) who were cleared for having an out-patient surgical or aerosol generating procedure. The cases and controls were selected from our population of 17,644 adults (24,822 samples) evaluated in the outpatient clinics at Baylor College of Medicine (BCM) and their affiliate institutions from March 18, 2020, through January 16, 2021, as previously described[14]. Three distinct adult populations were tested: 1) symptomatic employees utilizing occupational health services, 2) patients evaluated at medical and surgical clinics, and 3) patients who required clearance for an out-patient surgical or aerosol generating procedure. Serial samples were obtained from individuals who came back to be tested for evidence that the virus was cleared or were enrolled as sub-study to determine the viral shedding kinetics. Testing for SARS-CoV-2 was performed in our Clinical Laboratory Improvement Amendments (CLIA) Certified Respiratory Virus Diagnostic Laboratory (ID#: 45D0919666). Although RT-PCR testing was performed as a service to BCM, the collection of metadata was performed under an Institutional Review Board approved protocol with waiver of consent.

The extremely high viral load cases consisted of adults with an extremely high viral load (Ct <16) for the N1 target on their first mid-turbinate (MT) sample and had at least two subsequent positive MT samples [14]. Of the 104 individuals with an extremely high viral load in their first test, 30 individuals met the criteria for multiple positive samples over the ensuing 4 weeks. Adults from two other groups were matched to each extremely high viral load case: a low viral load (Ct 31-<40) SARS-CoV-2 positive adult (SARS-CoV-2 low viral load) and an otherwise stable control who tested negative for SARS-CoV-2 (SARS-CoV-2 negative control) and was cleared

for an out-patient surgical or aerosol generating procedure. Of the 453 individuals with a low viral load in their first test, 126 individuals met the criteria for multiple positive samples over the ensuing 4 weeks. The extremely high viral load cases were matched to the other two groups by gender, week of first test (+ 1 week), age (+ 1 year) and zip code (5 digits). If a match could not be found the range of the factors were expanded to + 3 weeks of first test, + 10 years and 3 digits for the zip code. The ten extremely high viral load cases were randomly selected from our pool of 30 individuals with an extremely high viral load with multiple positive MT samples. The best matched SARS-CoV-2 low viral load case and negative control were then selected for each extremely high viral load case.

SARS-CoV-2 RT- PCR

Viral RNA extraction and RT-PCR testing was performed as previously described (Avadhanula et al., 2021). In brief, viral RNA was extracted using the Qiagen Viral RNA Mini Kit (QIAGEN Sciences, Maryland, USA) with an automated extraction platform QIAcube (QIAGEN, Hilden, Germany). The extracted RNA samples were tested by CDC 2019-novel coronavirus (2019-ncoV) Real-Time RT-PCR Diagnostic panel [CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel for Emergency Use Only Instructions for Use]. RT-PCR reaction was set up using TaqPath™ 1-Step RT-qPCR Master Mix, CG (Applied Biosystems, CA) and run on 7500 Fast Dx Real-Time PCR Instrument with SDS 1.4 software. Respiratory samples with cycle threshold (Ct) values <40 for both N1 and N2 primers were considered RT-PCR positive for SARS-CoV-2.

Human Nose organoid model.

The differentiated human nose organoid derived air liquid interface (HNO-ALI) cells were apically infected with SARS-CoV-2 [Isolate USA-WA1/2020, obtained from Biodefense and Emerging Infectious resources (BEI)] at a multiplicity of infection of 0.01 or mock infected with

423 airway organoid differentiation media, as previously described [26] . At the respective time points,
424 the apical side of the transwells was washed twice and the cells were lysed using lysis buffer of
425 RNeasy mini kit and RNA extracted.

426 *RNA extraction, library preparation and sequencing*

427 Samples were extracted using the Qiagen RNeasy mini kit (#74104 rev. 10/19) following the
428 manufacturer's protocol for samples <5e6 cells. Samples were eluted in 50ul RNase-free water.
429 RNA quality and quantity were estimated using Agilent Bioanalyzer OR Caliper GX. To monitor
430 sample and process consistency, 1 µl of the 1:50 diluted synthetic RNA designed by External
431 RNA Controls Consortium (ERCC) (4456740, ThermoFisher) was added. Whole transcriptome
432 sequencing (total RNAseq) data was generated using the Illumina TruSeq Stranded Total RNA
433 with Ribo-Zero Globin kit (20020612, Illumina Inc.) cDNA was prepared following rRNA and
434 Globin mRNA depletion, and paired-end libraries were prepared on Beckman BioMek FXp liquid
435 handlers. For this, cDNA was A-tailed followed by ligation of the TruSeq UD Indexes (Cat #
436 20022370) and amplified for 15 PCR cycles following manufacturer’s recommendation. AMPure
437 XP beads (A63882, Beckman Coulter) were used for library purification. Libraries were
438 quantified using a Fragment Analyzer (Agilent Technologies, Inc) electrophoresis system and
439 pooled in equimolar ratios. This pool was quantified using qPCR to determine loading
440 concentration for sequencing. Sequencing was performed on the NovaSeq 6000 instrument
441 using the S4 reagent kit (300 cycles) to generate 2x150bp paired end reads.

442 *Primary Analysis for Total RNASeq*

443 The RNA-Seq analysis pipeline cleans and processes raw RNA sequencing data (FASTQs),
444 providing robust QC metrics and has the flexibility to map the reads to GRCh38 reference
445 genome (after excluding the alternate contigs). The latest versions of software for sequence
446 alignment (STAR v2.7.3a), for marking of duplicate reads (Picard v2.22.5) and for conversion of
447 BAM files to FASTQ files (Samtools v.1.9) are part of this pipeline. In addition to these

448 components, the pipeline uses RSEM (v.1.3.3) for measuring gene expression and RNA-SeQC
449 (v.1.1.9), Qualimap2 (v2.2.1) and ERCCQC (v.1.0) to generate quality control metrics on the
450 RNA-Seq data. The pipeline also produces the raw gene features counts by using
451 featureCounts (v2.0.1).

452 *Gene expression analysis*

453 For our serial MT swab dataset, where RNA-seq data were generated in different batches
454 involving time and differences in extraction and processing methods, Combat algorithm [37] was
455 used to correct for any observed batch effects. Fragments Per Kilobase Million (FPKM) values
456 were quantile normalized [38] and log2-transformed. The differential expression analyses to define
457 the host transcriptional response focus on 20000 genes for which an Entrez identifier could be
458 associated with the transcript feature. To identify expression patterns associated with the host
459 transcriptional response to SARS-CoV-2 infection in the serial MT swab dataset, Log2
460 expression values correlated with SARS-CoV-2 Ct values across all 44 samples in the RNA-seq
461 dataset. Additional two-group comparisons were carried using t-test on log2 expression values.
462 For the HNO204 nose organoid dataset, Log2 expression values were compared between
463 SARS-CoV-2 and Mock control by t-test, combining time points of 6hrs, 72hrs, and 6 days for
464 each group.

465 *Analysis of external transcriptome datasets*

466 To define transcriptional signatures of the host cell response to SARS-CoV-2 infection in lung
467 cancer cells, we referred to the GSE147507 RNA-seq dataset [39]. In this dataset, A549 and
468 Calu-3 were mock-treated or infected with SARS-CoV-2 and then profiled for gene expression.
469 We used data from the SARS-CoV-2 profiling experiments involving multiplicity-of-Infection
470 (MOI) of 2. We converted raw gene-level sequencing read counts to reads per million Mapped
471 (RPM) values and then log2-transformed them [37]. For the Mick et al. RNA-seq dataset of

nasopharyngeal/oropharyngeal samples in 238 patients with COVID-19, other viral, or non-viral acute respiratory illnesses [25], RPM values were quantile normalized before the analysis.

Statistical analysis

All p-values were two-sided unless otherwise specified. We performed all tests using log2-transformed gene expression values. False Discovery Rates (FDRs) due to multiple testing of genes were estimated using the method of Storey and Tibshirini [40]. Even in instances of nominally significant genes only moderately exceeding chance expectations by FDR, the nominally significant genes were found in downstream enrichment analyses (involving functional gene sets and results of external SARS-CoV-2-related RNA-seq datasets) to contain molecular information representing real biological differences. We evaluated enrichment of GO annotation terms [41] within sets of differentially expressed genes using SigTerms software [42] and one-sided Fisher's exact tests. Visualization using heat maps was performed using JavaTreeview (version 1.1.6r4) [43,44]. Gene ontology (GO) analysis of DEGs used in the upset plot (Figure S1) was performed using the web-based Database for Annotation, Visualization, and Integrated Discovery (DAVID; version - v2023q1) [45,46]

Data Availability

The RNA-seq dataset of serially collected samples and of nose organoids will be deposited at Gene Expression Omnibus (GEO) (GEO accession number pending). In terms of previously published data, we obtained RNA-seq expression data from experimental models of SARS-CoV-2 viral infection or other treatments from GEO (GSE147507). The Mick et al. RNA-seq dataset is available at GEO (GSE156063).

Figure Legends

Figure 1. Differential gene sets associated with the transcriptional host response to SARS-CoV-2 infection across serially collected samples. (a) For each of 20000 genes, expression (log2 FPKM) was correlated with viral load (inverse correlation with Ct value) across 44 samples from 20 subjects. Numbers of statistically significant genes (by Pearson's) at both p<0.01 and p<0.001 significance levels are represented, as compared to the chance expected by multiple testing. **(b)** Numbers of differential genes (p<0.01, t-test) when comparing: 1) Visit 1 samples from the extremely high viral load group (n=8 samples from eight subjects) with the samples in the negative group (n=4); samples at the latest time points for each of the subjects from the extremely high viral load group (n=8 samples) with the samples in the negative group; samples from the low viral load group (n=8 samples from eight subjects, using earliest time point) with the samples in the negative group. Chance expected genes at p<0.01 due to multiple testing would be on the order of 200 [40]. **(c)** Heat map comparing differential patterns across the three comparisons from part b, for the 1357 genes significant (p<0.01) for any comparison. Columns off to the side indicated which genes were correlated with viral load (p<0.01) across all 44 samples (from part a), and which genes have Gene Ontology (GO) annotation [41] 'response to virus'.

Figure 2. Differential expression patterns and functional gene groups associated with SARS-CoV-2 viral load across serially collected samples. (a) Across 44 MT swab samples representing 20 subjects, differential gene expression patterns for the set of 112 genes significantly correlated with SARS-CoV-2 viral load (i.e., inversely correlated with Ct value) at p<0.001 (Pearson's) are represented. Heat map contrast (bright yellow/blue) is 3-fold change from the average of the samples from the low viral load group. Genes listed off to the right have GO annotation 'response to virus'. Extremely high viral load, Ct<20. **(b)** Selected significantly enriched GO terms [41] within the genes over-expressed with SARS-CoV-2 viral load (p<0.01,

519 Pearson's). For each GO term, enrichment p-values and numbers of genes in the SARS-CoV-2-
520 associated gene set are indicated. Enrichment p-values by one-sided Fisher's exact test.

521 **Figure 3. Genes correlated with SARS-CoV-2 viral load over time are similarly expressed**
522 **in independent datasets of SARS-CoV-2 infected lung and upper airway cells. (a)**
523 Differential expression patterns for the 112 genes correlated with SARS-CoV-2 viral load across
524 our serial sampling cohort (p<0.001, from Figure 2a) were examined in two independent RNA-
525 seq datasets of SARS-CoV-2 infection: one of lung cancer cell lines (A549 and Calu-3) infected
526 with SARS-CoV-2 at multiplicity-of-Infection (MOI) of 2 for 24 hours [39], and one of
527 nasopharyngeal/oropharyngeal samples in 238 patients with COVID-19, other viral, or non-viral
528 acute respiratory illnesses [25]. Gene order is the same across all datasets. Heat map contrast
529 (bright yellow/blue) is 3-fold change from the corresponding comparison group (serial sampling
530 dataset, average of the samples from the low viral load group; lung cancer cell line dataset,
531 average of corresponding mock control group; Mick et al. dataset, average of "no virus"
532 samples). **(b)** Venn diagram representing the gene set overlaps among the genes increased
533 with SARS-CoV-2 infection in each of the three RNA-seq datasets from part a (with Calu-3 lung
534 cancer cell line being considered here over A549). A p-value cutoff of p<0.01 was used to
535 define top genes for each dataset (serial MT swab and Mick et al.
536 nasopharyngeal/oropharyngeal datasets, Pearson's correlation with viral load; Calu-3 dataset, t-
537 test). Gene set enrichment p-values by one-sided Fisher's exact test. Genes overlapping
538 between all three datasets are listed. **(c)** Similar to part b, but for genes decreased with SARS-
539 CoV-2 infection.

540 **Figure 4. Differential expression patterns and functional gene groups associated with**
541 **SARS-CoV-2 infection of nose organoids. (a)** HNO204 human nose organoids were infected
542 with SARS-CoV-2 at an MOI of 0.01, and samples at 6hrs, 72hrs, and 6 days post infection
543 were profiled for gene expression. Differential expression patterns for the top 867 genes over-

expressed in HNO204 with SARS-CoV-2 infection ($p<0.05$, t-test) are represented here. Next to the HNO204 dataset are the corresponding patterns for independent RNA-seq datasets of SARS-CoV-2 infection: lung cancer cell lines (A549 and Calu-3)[37], our serially collected MT swab samples from patients, and nasopharyngeal/oropharyngeal samples from Mick et al [25]. Gene order is the same across all datasets. Heat map contrast (bright yellow/blue) is 3-fold change from the corresponding comparison group. **(b)** Venn diagram representing the gene set overlaps among the genes increased with SARS-CoV-2 infection in each of the following RNA-seq datasets: HNO204, serial MT swab, and Calu-3 lung cancer cell line. Gene set enrichment p-values by one-sided Fisher's exact test. Genes overlapping between HNO204 and serial MT swab datasets are listed. **(c)** Similar to part b, but for genes decreased with SARS-CoV-2 infection. **(d)** Selected significantly enriched GO terms [41] within the genes over-expressed with SARS-CoV-2 infection in HNO204 ($p<0.05$, t-test). For each GO term, enrichment p-values and numbers of genes in the SARS-CoV-2-associated gene set are indicated. Enrichment p-values by one-sided Fisher's exact test.

References

1. Fajnzylber J, Regan J, Coxen K, et al. SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nat Commun*. 2020;11(1):5493. doi:10.1038/s41467-020-19057-5

2. He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med*. 2020;26(5):672-675. doi:10.1038/s41591-020-0869-5

3. Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review. *Ann Intern Med*. 2020;173(5):362-367. doi:10.7326/M20-3012

4. Pujadas E, Chaudhry F, McBride R, et al. SARS-CoV-2 viral load predicts COVID-19 mortality. *Lancet Respir Med*. 2020;8(9):e70. doi:10.1016/S2213-2600(20)30354-4

5. Lucas C, Klein J, Sundaram ME, et al. Delayed production of neutralizing antibodies correlates with fatal COVID-19. *Nat Med*. 2021;27(7):1178-1186. doi:10.1038/s41591-021-01355-0

6. Rouchka EC, Chariker JH, Alejandro B, et al. Induction of interferon response by high viral loads at early stage infection may protect against severe outcomes in COVID-19 patients. *Sci Rep*. 2021;11(1):15715. doi:10.1038/s41598-021-95197-y

7. Rydyznski Moderbacher C, Ramirez SI, Dan JM, et al. Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. *Cell*. 2020;183(4):996-1012.e19. doi:10.1016/j.cell.2020.09.038

8. Takahashi T, Ellingson MK, Wong P, et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature*. 2020;588(7837):315-320. doi:10.1038/s41586-020-2700-3

9. Puhach O, Adea K, Hulo N, et al. Infectious viral load in unvaccinated and vaccinated individuals infected with ancestral, Delta or Omicron SARS-CoV-2. *Nat Med*. 2022;28(7):1491-1500. doi:10.1038/s41591-022-01816-0

10. Liu J, Liao X, Qian S, et al. Community transmission of severe acute respiratory syndrome Coronavirus 2, Shenzhen, China, 2020. *Emerg Infect Dis*. 2020;26(6):1320-1323. doi:10.3201/eid2606.200239

11. Jones TC, Biele G, Mühlemann B, et al. Estimating infectiousness throughout SARS-CoV-2 infection course. *Science (80-)*. 2021;373(6551):eabi5273. doi:10.1126/science.abi5273

12. van Kampen JJA, van de Vijver DAMC, Fraaij PLA, et al. Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). *Nat Commun*. 2021;12(1):267. doi:10.1038/s41467-020-20568-4

13. Puhach O, Meyer B, Eckerle I. SARS-CoV-2 viral load and shedding kinetics. *Nat Rev Microbiol*. Published online 2022. doi:10.1038/s41579-022-00822-w

14. Avadhanula V, Nicholson EG, Ferlic-Stark L, et al. Viral Load of Severe Acute Respiratory Syndrome Coronavirus 2 in Adults During the First and Second Wave of Coronavirus Disease 2019 Pandemic in Houston, Texas: The Potential of the Superspreader. *J Infect Dis*. Published online February 15, 2021. doi:10.1093/infdis/jiab097

15. Goyal A, Reeves DB, Cardozo-Ojeda EF, Schiffer JT, Mayer BT. Viral load and contact heterogeneity predict SARS-CoV-2 transmission and super-spreading events. Walczak AM, Childs L, Forde J, eds. *Elife*. 2021;10:e63537. doi:10.7554/eLife.63537

16. Hou YJ, Okuda K, Edwards CE, et al. SARS-CoV-2 Reverse Genetics Reveals a Variable

Infection Gradient in the Respiratory Tract. *Cell*. 2020;182(2):429-446.e14. doi:10.1016/j.cell.2020.05.042

17. Schultze JL, Aschenbrenner AC. COVID-19 and the human innate immune system. *Cell*. 2021;184(7):1671-1692. doi:10.1016/j.cell.2021.02.029

18. Chua RL, Lukassen S, Trump S, et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol*. 2020;38(8):970-979. doi:10.1038/s41587-020-0602-4

19. Nicholson EG, Schlegel C, Garofalo RP, et al. Robust cytokine and chemokine response in nasopharyngeal secretions: Association with decreased severity in children with physician diagnosed bronchiolitis. *J Infect Dis*. 2016;214(4). doi:10.1093/infdis/jiw191

20. Xiong Y, Liu Y, Cao L, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg Microbes Infect*. 2020;9(1):761-770. doi:10.1080/22221751.2020.1747363

21. Daamen AR, Bachali P, Owen KA, et al. Comprehensive transcriptomic analysis of COVID-19 blood, lung, and airway. *Sci Rep*. 2021;11(1):7052. doi:10.1038/s41598-021-86002-x

22. Butler D, Mozsary C, Meydan C, et al. Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. *Nat Commun*. 2021;12(1):1660. doi:10.1038/s41467-021-21361-7

23. Lieberman NAP, Peddu V, Xie H, et al. In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age. *PLOS Biol*. 2020;18(9):e3000849. https://doi.org/10.1371/journal.pbio.3000849

24. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*. 2020;181(5):1036-1045.e9. doi:10.1016/j.cell.2020.04.026

25. Mick E, Kamm J, Pisco AO, et al. Upper airway gene expression reveals suppressed immune responses to SARS-CoV-2 compared with other respiratory viruses. *Nat Commun*. 2020;11(1):5854. doi:10.1038/s41467-020-19587-y

26. Rajan A, Weaver AM, Aloisio GM, et al. The Human Nose Organoid Respiratory Virus Model: an Ex Vivo Human Challenge Model To Study Respiratory Syncytial Virus (RSV) and Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Pathogenesis and Evaluate Therapeutics. *MBio*. 2022;13(1):e0351121. doi:10.1128/mbio.03511-21

27. Tjan LH, Furukawa K, Nagano T, et al. Early Differences in Cytokine Production by Severity of Coronavirus Disease 2019. *J Infect Dis*. 2021;223(7):1145-1149. doi:10.1093/infdis/jiab005

28. Zhang C, Feng Y-G, Tam C, Wang N, Feng Y. Transcriptional Profiling and Machine Learning Unveil a Concordant Biosignature of Type I Interferon-Inducible Host Response Across Nasal Swab and Pulmonary Tissue for COVID-19 Diagnosis. *Front Immunol*. 2021;12:733171. doi:10.3389/fimmu.2021.733171

29. Saravia-Butler AM, Schisler JC, Taylor D, et al. Host transcriptional responses in nasal swabs identify potential SARS-CoV-2 infection in PCR negative patients. *iScience*. 2022;25(11):105310. doi:10.1016/j.isci.2022.105310

30. Chang C-W, Parsi KM, Somasundaran M, et al. A Newly Engineered A549 Cell Line Expressing ACE2 and TMPRSS2 Is Highly Permissive to SARS-CoV-2, Including the Delta and Omicron Variants. *Viruses*. 2022;14(7). doi:10.3390/v14071369

649 31. Diamond MS, Kanneganti T-D. Innate immunity: the first line of defense against SARS-
650 CoV-2. *Nat Immunol*. 2022;23(2):165-176. doi:10.1038/s41590-021-01091-0

651 32. Darif D, Hammi I, Kihel A, El Idrissi Saik I, Guessous F, Akarid K. The pro-inflammatory
652 cytokines in COVID-19 pathogenesis: What goes wrong? *Microb Pathog*.
653 2021;153:104799. doi:10.1016/j.micpath.2021.104799

654 33. Rajan A, Weaver AM, Aloisio GM, et al. The Human Nose Organoid Respiratory Virus
655 Model: an Ex Vivo Human Challenge Model To Study Respiratory Syncytial Virus (RSV)
656 and Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Pathogenesis
657 and Evaluate Therapeutics. *MBio*. 2021;13(1):e0351121. doi:10.1128/mbio.03511-21

658 34. Travaglini KJ, Nabhan AN, Penland L, et al. A molecular cell atlas of the human lung from
659 single-cell RNA sequencing. *Nature*. 2020;587(7835):619-625. doi:10.1038/s41586-020-
660 2922-4

661 35. Fogh J, Fogh JM, Orfeo T. One hundred and twenty-seven cultured human tumor cell
662 lines producing tumors in nude mice. *J Natl Cancer Inst*. 1977;59(1):221-226.
663 doi:10.1093/jnci/59.1.221

664 36. Sajuthi SP, DeFord P, Li Y, et al. Type 2 and interferon inflammation regulate SARS-
665 CoV-2 entry factor expression in the airway epithelium. *Nat Commun*. 2020;11(1):5139.
666 doi:10.1038/s41467-020-18781-2

667 37. Chen F, Zhang Y, Sucgang R, et al. Meta-analysis of host transcriptional responses to
668 SARS-CoV-2 infection reveals their manifestation in human tumors. *Sci Rep*.
669 2021;11(1):2459. doi:10.1038/s41598-021-82221-4

670 38. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data
671 using empirical Bayes methods. *Biostatistics*. 2007;8(1):118-127.

672 doi:10.1093/biostatistics/kxj037

673 39. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology.
674 The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556

675 40. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods
676 for high density oligonucleotide array data based on variance and bias. *Bioinformatics*.
677 2003;19(2):185-193. doi:10.1093/bioinformatics/19.2.185

678 41. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad*
679 *Sci U S A*. 2003;100(16):9440-9445. doi:10.1073/pnas.1530509100

680 42. Creighton CJ, Nagaraja AK, Hanash SM, Matzuk MM, Gunaratne PH. A bioinformatics
681 tool for linking gene expression profiling results with public databases of microRNA
682 target predictions. *RNA*. 2008;14(11):2290-2296. doi:10.1261/rna.1188208

683 43. Saldanha AJ. Java Treeview--extensible visualization of microarray data. *Bioinformatics*.
684 2004;20(17):3246-3248. doi:10.1093/bioinformatics/bth349

685 44. Pavlidis P, Noble WS. Matrix2png: a utility for visualizing matrix data. *Bioinformatics*.
686 2003;19(2):295-296. doi:10.1093/bioinformatics/19.2.295

687 45. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene
688 lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.
689 doi:10.1038/nprot.2008.211

690 46. Sherman BT, Hao M, Qiu J, et al. DAVID: a web server for functional enrichment analysis
691 and functional annotation of gene lists (2021 update). *Nucleic Acids Res*.
692 2022;50(W1):W216-W221. doi:10.1093/nar/gkac194

693

694 Author Contributions

695 Conceptualization: V.A, P.A.P, C.J.C.; Methodology: C.J.C., L.F.S, R.S, D.N, E.N, A.R, V.K.M,

696 H.D, D.M.M S.J.J.C, K.L.H, Investigation: V.A, P.A.P, L.F.S, C.J.C. Formal Analysis: C.J.C. Y.Z.

697 ; Data Curation: C.J.C., R.S, V.K.M, H.D,; Visualization: V.A, C.J.C, L.F.S, P.A.P., Y.Z. D.N;

698 Writing: V.A C.J.C. P.A.P; Manuscript Review: V.A,C.J.C., L.F.S, R.S, D.N, E.N, A.R V.K.M,

699 H.D, D.M.M, G.M, S.E.B, S.J.J.C, K.L.H, , R.A.G and J.P, P.A.P ; Supervision: V.A, C.J.C.,

700 R.A.G, J.P and P.A.P

701

702 Acknowledgements

703 This work was supported by NIH grants CA125123 (CJC), U19AI144297 (VA, CJC, RG, JP,

704 PAP)

705 Conflict of Interest statement: The authors declare no competing interests.

706

707

TABLE 1. Demographic and Visit Characteristics by Matched Groups

	Extremely high viral load cases (n = 10)	Low viral load cases (n = 10)	Negative for SARS-CoV-2 (n = 10)	p-value
Age[a], years	51.5 (22.0, 69.0)	46.0 (24.0, 80.0)	51.0 (25.0, 70.0)	0.998
Gender				1.000
Male	6 (60.0%)	6 (60.0%)	6 (60.0%)	
Female	4 (40.0%)	4 (40.0%)	4 (40.0%)	
Race				0.126
Asian	1 (10.0%)	1 (10.0%)	0 (0.0%)	
Black	2 (20.0%)	0 (0.0%)	2 (20.0%)	
White	6 (60.0%)	3 (30.0%)	5 (50.0%)	
Other/Multiracial	1 (10.0%)	1 (10.0%)	0 (0.0%)	
Unknown/Declined	0 (0.0%)	5 (50.0%)	3 (30.0%)	
Ethnicity				0.322
Hispanic	3 (30.0%)	4 (40.0%)	3 (30.0%)	
Non-Hispanic	7 (70.0%)	3 (30.0%)	5 (50.0%)	
Unknown/Declined	0 (0.0%)	3 (30.0%)	2 (20.0%)	
Disease Severity				0.066
Asymptomatic/Mild	4 (40.0%)	5 (50.0%)	9 (90.0%)	
Mild/Moderate	6 (60.0%)	5 (50.0%)	1 (10.0%)	
Number of Co-morbid Conditions				0.906
None	6 (60.0%)	5 (50.0%)	8 (80.0%)	
One	2 (20.0%)	3 (30.0%)	1 (10.0%)	
Two	1 (10.0%)	1 (10.0%)	0 (0.0%)	
Three +	1 (10.0%)	1 (10.0%)	1 (10.0%)	
CDC week[b] **[end date]** at Visit 1	32 [08Aug20] (26 [27Jun20]-41 [10Oct20])	27.5 [11Jul20] (24 [13Jun20]-28 [11Jul20])	32 [08Aug20] (26 [27Jun20]-41 [10Oct20])	
Duration[a], days				
between Visit 1 - Visit 2	7.0 (5.0, 12.0)	9.5 (4.0, 13.0)	N/A	
between Visit 2 - Visit 3	7.5 (4.0, 20.0)	8.0 (4.0, 13.0)	N/A	
between Visit 3 - Visit 4	7.0	9.5 (7.0, 12.0)	N/A	
N1 Ct value[a]				
at Visit 1	14.5 (9.8, 15.8)	34.1 (31.7, 36.3)	N/A	
at Visit 2	26.6 (24.2, 33.9)	36.1 (30.8, 38.2)	N/A	
at Visit 3	35.7 (32.4, 38.2)	35.6 (32.2, 38.5)	N/A	
at Visit 4	34.3	33.6 (33.6, 33.7)	N/A	

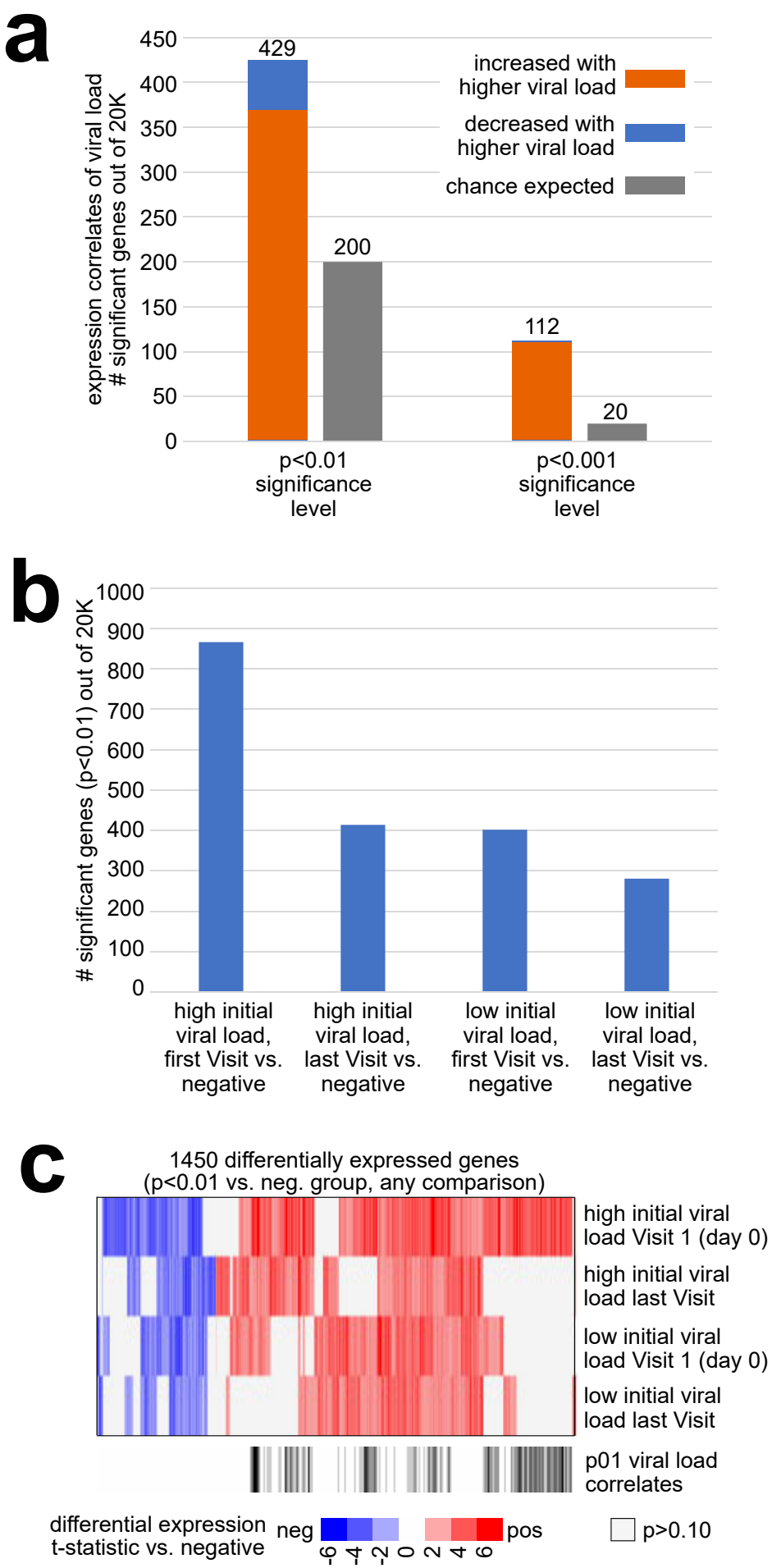
Abbreviations: Ct=cycle threshold, [a]Median (Min, Max)[b]Median (IQR) **or Median (Q1-Q3) or Median (25th percentile-75th percentile) or Median (lower quartile-upper quartile)**

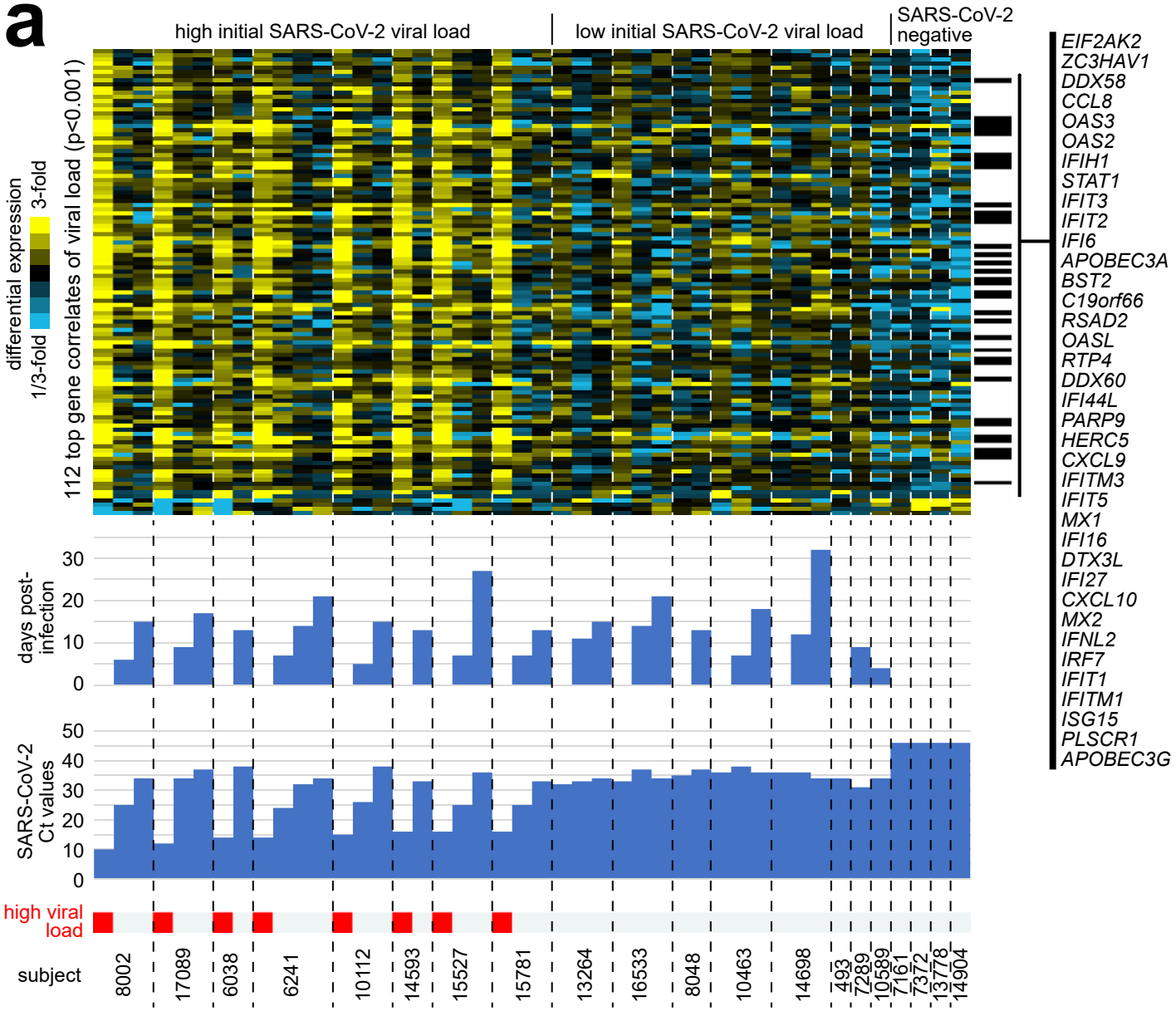
Differences between groups were determined using the Kruskal-Wallis test for variables with non-parametric distribution and by Fisher's Exact test for categorical variables. P-value <0.05 was considered significantly different between groups.

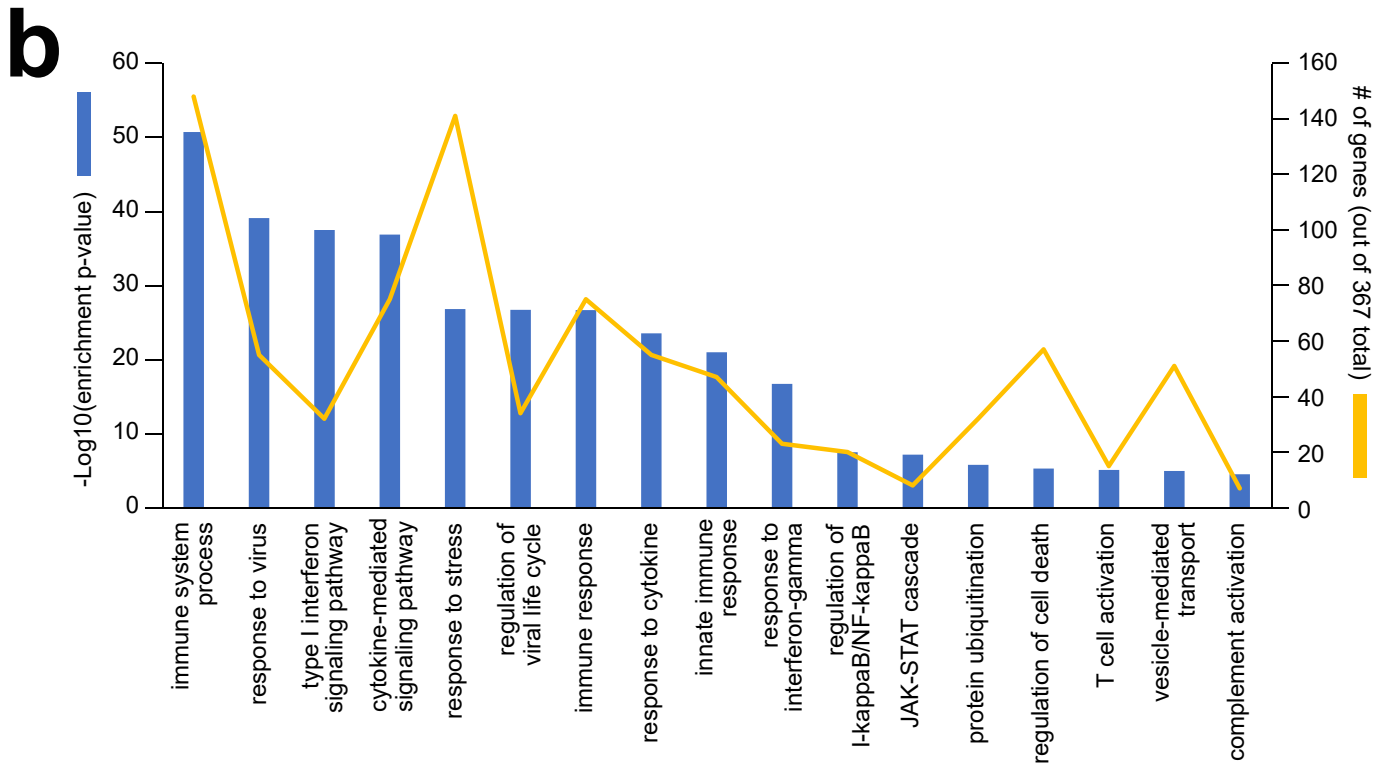
714 TABLE 2. Demographic Characteristics by Matched Groups with RNA sequencing Data

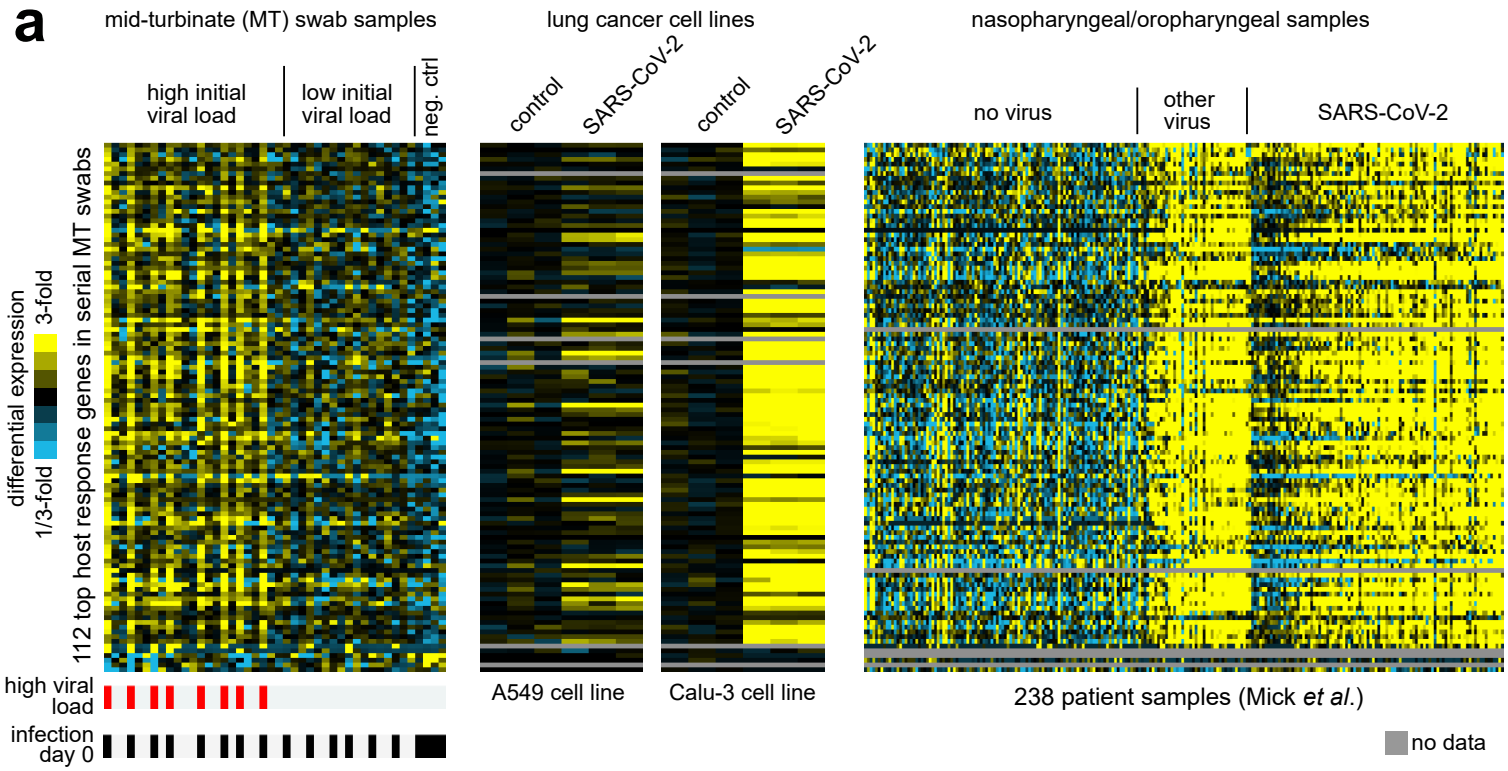
	Extremely high viral load cases (N = 8)	Low viral load cases (N = 8)	Negative for SARS-CoV-2 (N = 4)	p-value
Age				0.879[a]
Median (Q1, Q3)	39.5 (27.5, 57.5)	46.0 (33.0, 55.0)	45.5 (32.0, 57.0)	
Min, Max	22.0, 60.0	24.0, 80.0	30.0, 57.0	
Gender				0.851[b]
Male	4 (50.0%)	5 (62.5%)	3 (75.0%)	
Female	4 (50.0%)	3 (37.5%)	1 (25.0%)	
Race				0.158[b]
Asian	1 (12.5%)	1 (12.5%)	0 (0.0%)	
Black	2 (25.0%)	0 (0.0%)	0 (0.0%)	
White	4 (50.0%)	2 (25.0%)	1 (25.0%)	
Other/Multiracial	1 (12.5%)	1 (12.5%)	0 (0.0%)	
Unknown	0 (0.0%)	4 (50.0%)	3 (75.0%)	
Ethnicity				0.297[b]
Hispanic	2 (25.0%)	3 (37.5%)	1 (25.0%)	
Non-Hispanic	6 (75.0%)	3 (37.5%)	1 (25.0%)	
Unknown	0 (0.0%)	2 (25.0%)	2 (50.0%)	
Disease Severity				0.603[b]
Asymptomatic/Mild	3 (37.5%)	3 (37.5%)	3 (75.0%)	
Mild/Moderate	5 (62.5%)	5 (62.5%)	1 (25.0%)	
Number of Co-morbid Conditions				0.656[b]
None	5 (62.5%)	4 (50.0%)	4 (100.0%)	
One	1 (12.5%)	3 (37.5%)	0 (0.0%)	
Two	1 (12.5%)	0 (0.0%)	0 (0.0%)	
Three +	1 (12.5%)	1 (12.5%)	0 (0.0%)	
Sample Collected				
at Visit 1 only	0	1	4	
at Visits 1, 2, 3	5	2	0	
at Visits 1, 2, 3, 4	1	0	0	
at Visit 2 only	0	2	0	
at Visits 1, 2	0	1	0	
at Visits 1, 3	2	0	0	
at Visits 1, 2, 4	0	1	0	
at Visits 1, 3, 4	0	1	0	
Number of Samples per Subject				
One	0	3	4	
Two	2	1	0	
Three	5	4	0	
Four	1	0	0	

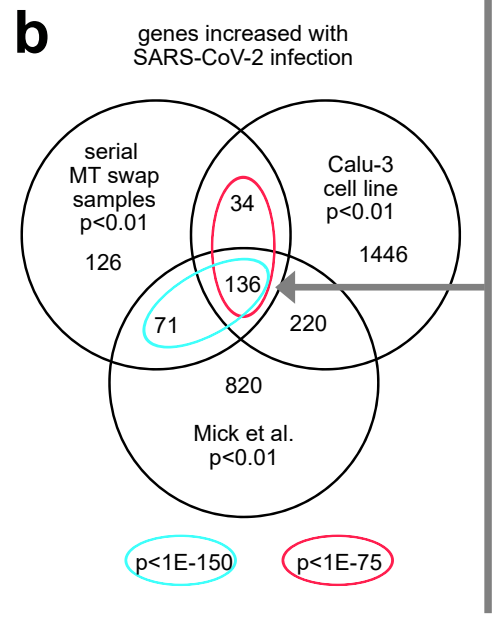
715 Differences between groups were determined using the Kruskal-Wallis test for variables with non-
716 parametric distribution and by Fisher's Exact test for categorical variables. P-value <0.05 was considered
717 significantly different between groups.



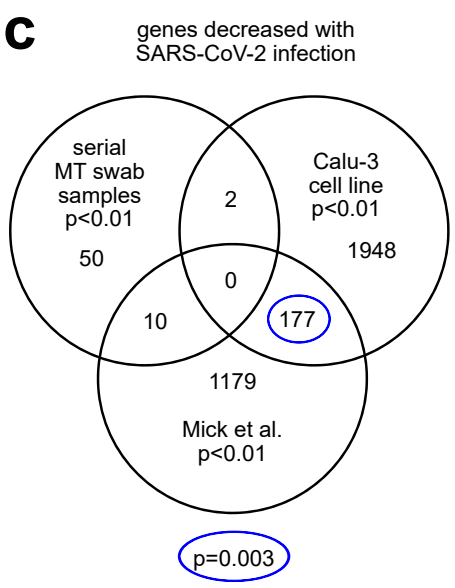


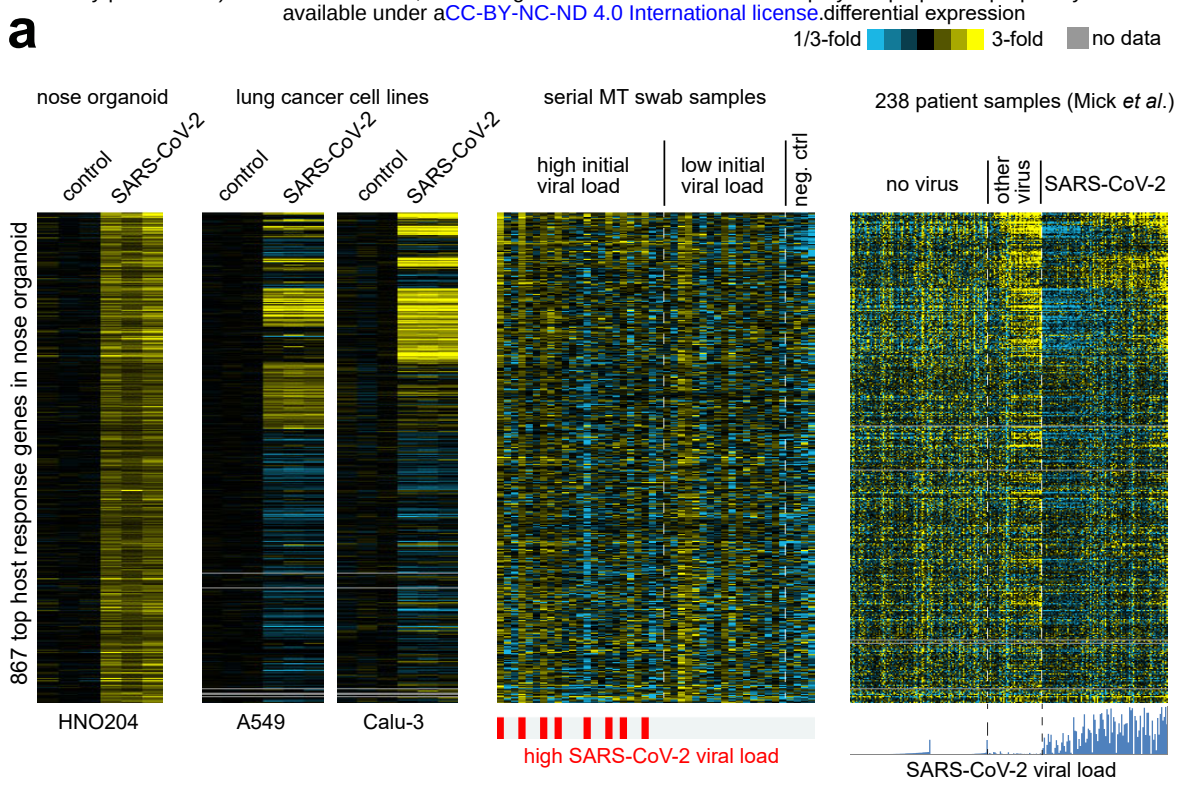


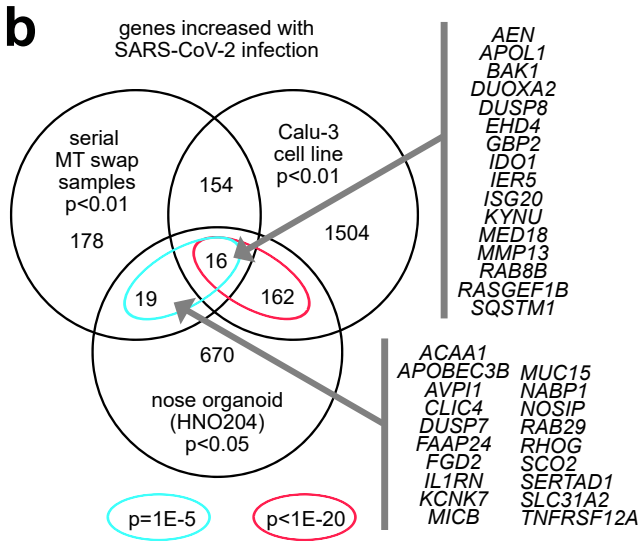


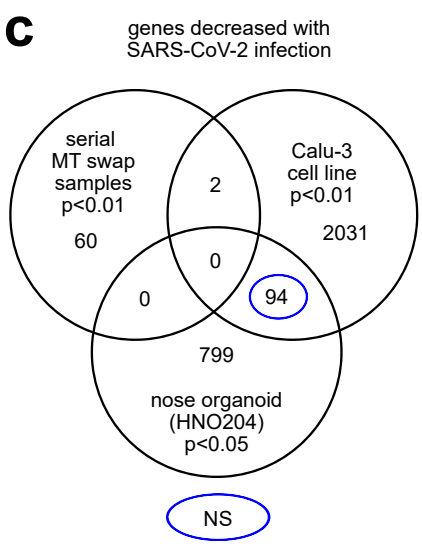


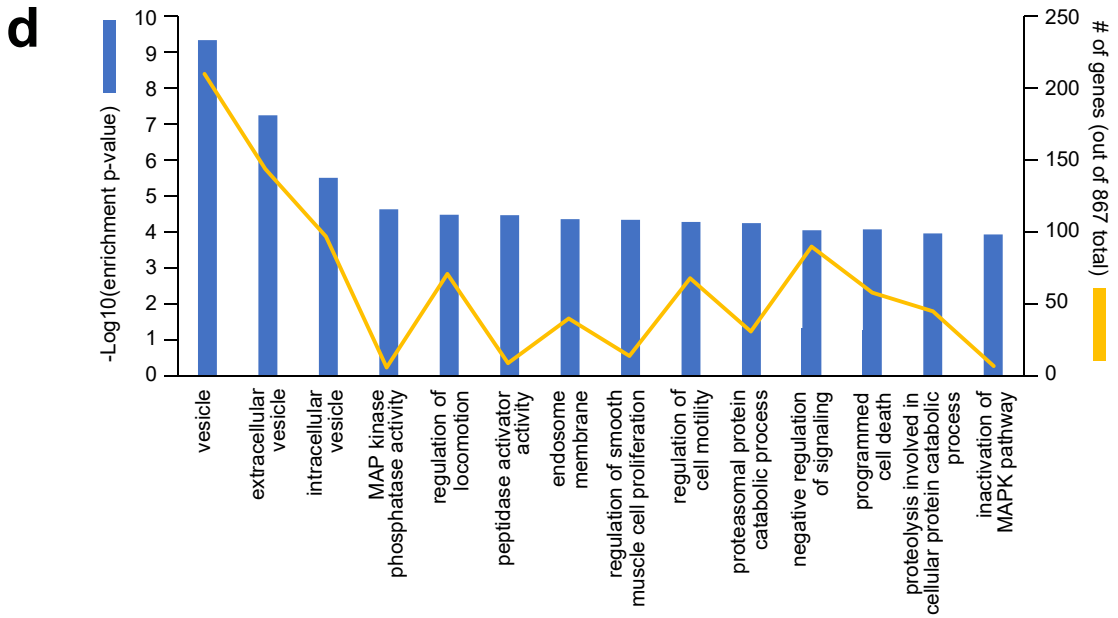
ADAR	*GBP4*	*LSM12*	*SAMD9*
AEN	*GBP5*	*MMP13*	*SAMD9L*
APOBEC3G	*GMPR*	*MX1*	*SECTM1*
APOL1	*HELZ2*	*MX2*	*SERPING1*
APOL2	*HERC5*	*MYD88*	*SLC15A3*
APOL6	*HERC6*	*NCOA7*	*SLC25A28*
ARID4B	*HLA-F*	*NMI*	*SLFN5*
ARID5B	*HSH2D*	*NR3C1*	*SP110*
B2M	*IDO1*	*NUB1*	*SPATS2L*
BAK1	*IFI16*	*NUPR1*	*SPSB1*
BST2	*IFI27*	*OAS1*	*STAT1*
CASP1	*IFI35*	*OAS2*	*STAT2*
CCL2	*IFI44*	*OAS3*	*TAP1*
CCR1	*IFI44L*	*OASL*	*TAP2*
CFB	*IFI6*	*OGFR*	*TGM2*
CHMP5	*IFIH1*	*OPTN*	*TNFAIP3*
CMPK2	*IFIT1*	*PARP12*	*TNFSF10*
CNP	*IFIT2*	*PARP14*	*TNFSF13B*
CX3CL1	*IFIT3*	*PARP9*	*TOR1B*
CXCL10	*IFIT5*	*PCGF5*	*TRAFD1*
CXCL11	*IFITM1*	*PIK3AP1*	*TRANK1*
CXCL9	*IFITM3*	*PLEKHA4*	*TRIM21*
DDX58	*IL4I1*	*PLSCR1*	*TRIM25*
DDX60	*IRF1*	*PPM1K*	*TRIM38*
DHX58	*IRF2*	*PRKD2*	*UBD*
DTX3L	*IRF7*	*PSMA6*	*UBE2L6*
DUOXA2	*ISG15*	*PSMB8*	*UBE2Z*
EHD4	*ISG20*	*PSMB9*	*USP18*
EIF2AK2	*JUN*	*RASGEF1B*	*WARS*
EPSTI1	*LAMP3*	*RBCK1*	*XAF1*
ETV7	*LAP3*	*RBMS2*	*ZBP1*
FST	*LGALS9*	*RNF114*	*ZC3HAV1*
GBP1	*LGMN*	*RSAD2*	*ZFP36L2*
GBP2	*LMO2*	*RTP4*	*ZNFX1*











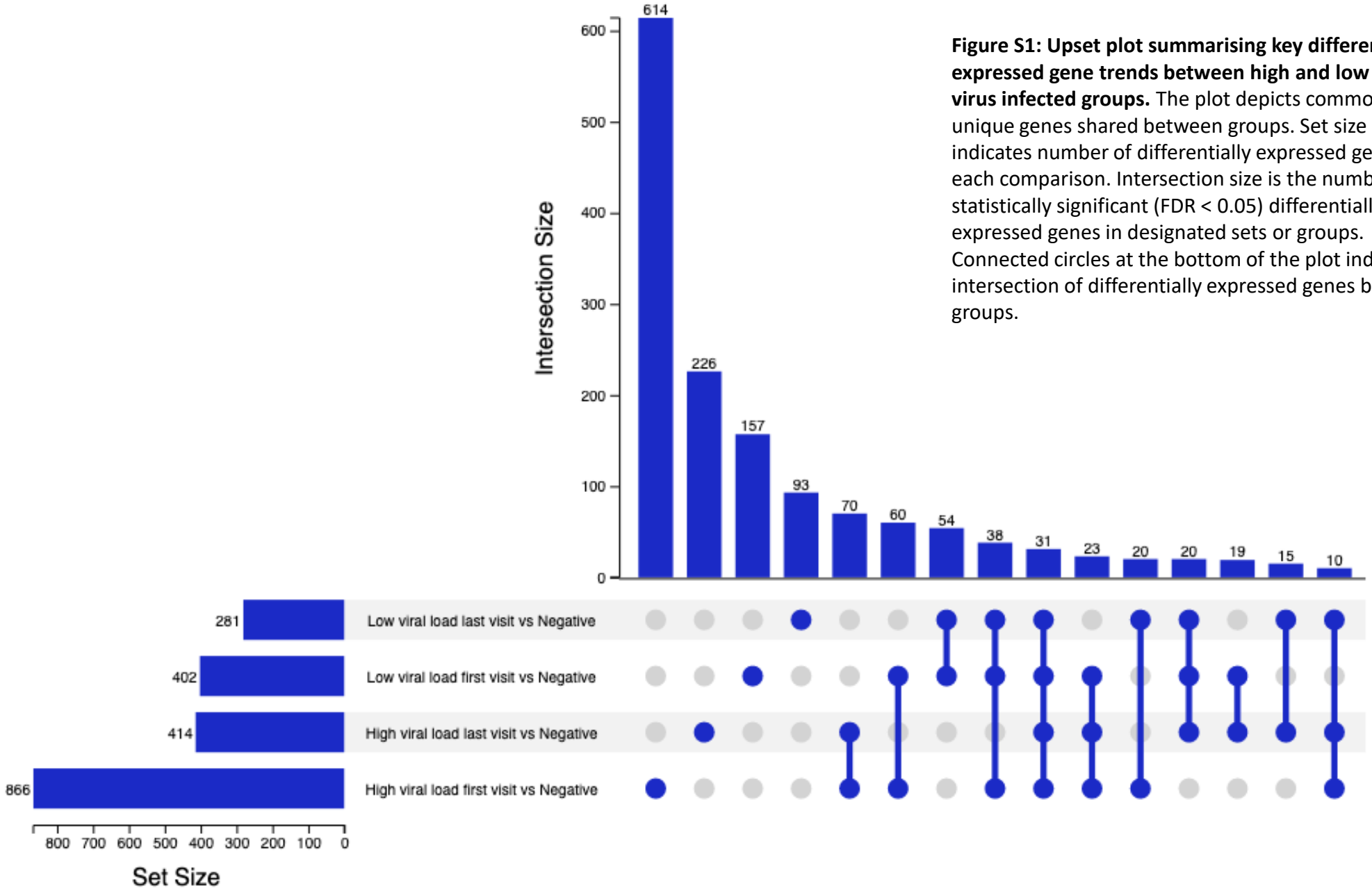


Figure S1: Upset plot summarising key differentially expressed gene trends between high and low SARS virus infected groups. The plot depicts common and unique genes shared between groups. Set size indicates number of differentially expressed genes in each comparison. Intersection size is the number of statistically significant (FDR < 0.05) differentially expressed genes in designated sets or groups. Connected circles at the bottom of the plot indicate an intersection of differentially expressed genes between groups.

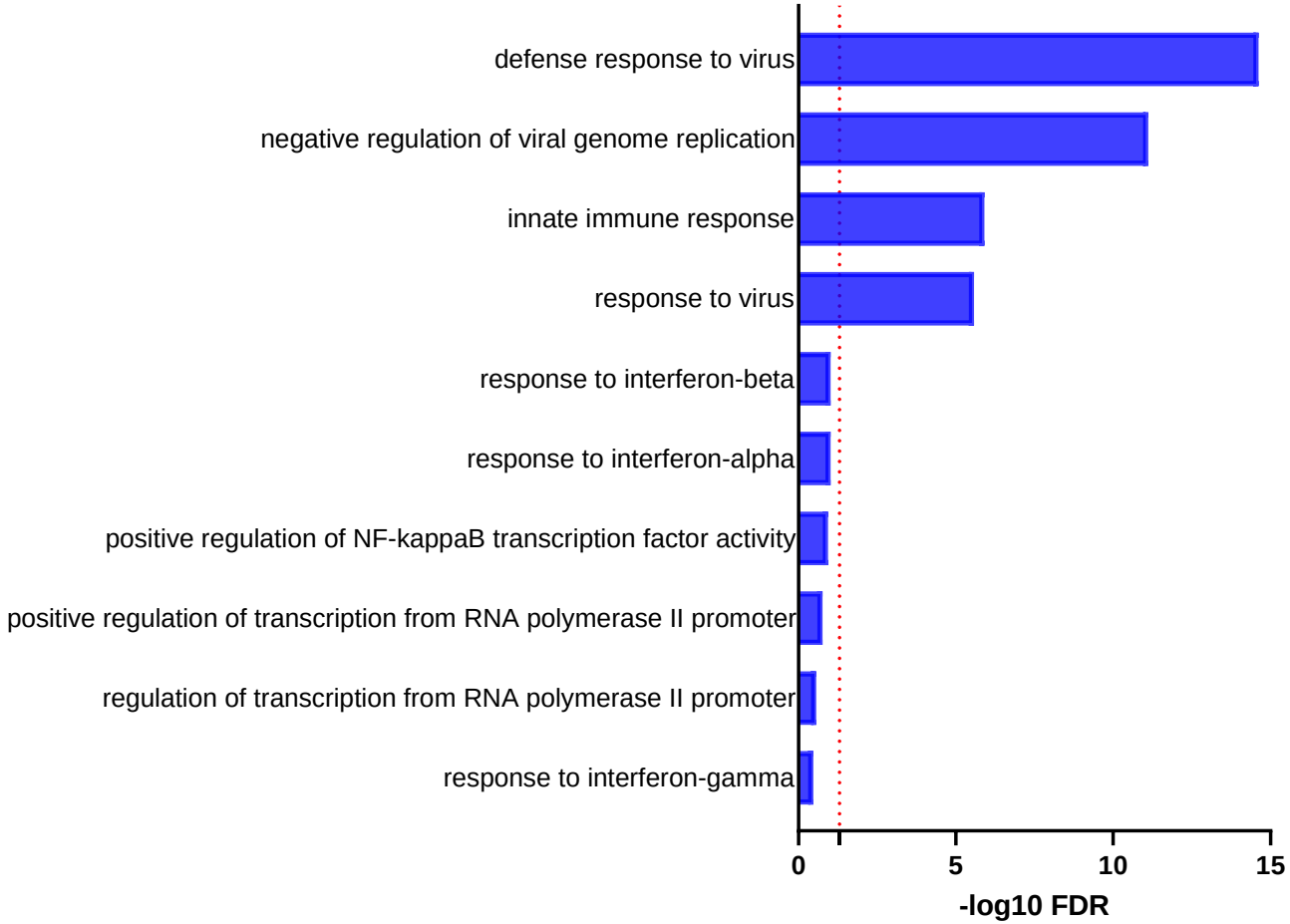


Figure S2: Gene ontology analysis of differentially expressed genes present only in high day 0 infection. The top10 enriched GO terms for biological processes altered in shown. Significantly enriched GO terms with a minimum three enriched genes were ranked by significance. The x-axis denoting the negative log fold change of significance. Dotted red line depicts the significance threshold of FDR <0.05.