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,

34   Abstract:

35   Current understanding of viral dynamics of SARS-CoV-2 and host responses driving the

36   pathogenic mechanisms in COVID-19 is rapidly evolving. Here, we conducted a longitudinal

37   study to investigate gene expression patterns during acute SARS-CoV-2 illness. Cases included

38   SARS-CoV-2 infected individuals with extremely high viral loads early in their illness, individuals

39   having low SARS-CoV-2 viral loads early in their infection, and individuals testing negative for

40   SARS-CoV-2. We could identify widespread transcriptional host responses to SARS-CoV-2

41   infection that were initially most strongly manifested in patients with extremely high initial viral

42   loads, then attenuating within the patient over time as viral loads decreased. Genes correlated

43   with SARS-CoV-2 viral load over time were similarly differentially expressed across independent

44   datasets of SARS-CoV-2 infected lung and upper airway cells, from both in vitro systems and

45   patient samples. We also generated expression data on the human nose organoid model during

46   SARS-CoV-2 infection. The human nose organoid-generated host transcriptional response

47   captured many aspects of responses observed in the above patient samples, while suggesting

48   the existence of distinct host responses to SARS-CoV-2 depending on the cellular context,

49   involving both epithelial and cellular immune responses. Our findings provide a catalog of

50   SARS-CoV-2 host response genes changing over time.

51

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

3

76   response is believed to be dependent, in part, on viral load and route of infection as well as the

77   host immune response [17]. The early host immune response is regulated closely by the epithelial

78   cell cytokine signaling in response to active viral replication [18]. Rapid and robust activation of the

79   antiviral innate immune response at the site of viral replication is required to control and clear

80   the virus. A delayed cytokine response can result in prolonged viral replication and worst clinical

81   outcome as seen for other respiratory viruses [19]

82   Our understanding of the viral dynamics of SARS-CoV-2 and host responses driving the

83   pathogenic mechanisms in COVID-19 is evolving rapidly. Multiple studies have reported various

84   characteristics of immune/inflammatory responses to SARS-CoV-2. Cytokine or chemokines-

85   related host inflammatory responses such as CCL2/MCP-1, CXCL10/IP-10, CCL3/MIP-1A, and

86   CCL4/MIP-1B were detected in bronchoalveolar lavage samples of SARS-CoV-2 infected adults

87   while activation of apoptosis and the P53 signaling pathway were observed in lymphocytes [20].

88   Inflammatory cytokine such as IL-1, IL-18, and IL-33 were enriched in the airways of COVID-19

89   patients [21]. In addition, a shotgun host transcriptomic analysis on nasopharyngeal samples

90   revealed a wide range of antiviral responses. These included gamma and alpha interferon

91   responses, elevated levels of ACE-2, interferon stimulated genes (ISGs), and interferon

92   inducible (IFI) genes [22]. Very few studies have demonstrated the temporal correlation between

93   viral load and host gene expression. Variation in viral load was associated with the SARS-CoV-

94   2 disease and the host response dynamics via innate and adaptive immunity (To et al., 2020).

95   Another study revealed that expression of interferon-responsive genes, including ACE-2,

96   increased as a function of viral load, while transcripts for B cell–specific proteins and neutrophil

97   chemokines were elevated in patients with lower viral load [23]. Rouchka et.al. reported that

98   cellular antiviral responses strongly correlated with viral loads. However, COVID-19 patients

99   who experienced mild symptoms had a higher viral load than those with severe complications[6].

4

100   We previously reported on a small group of adults with extremely high SARS-CoV-2 viral load,

101   who had the potential to be super spreaders and a large group of adults with low SARS-CoV-2

102   viral load, both groups had mild illness[14] . Here, we wanted to determine the host response in

103   relation to the viral load early during infection. We conducted a longitudinal study to investigate

104   gene expression patterns detected in the secretion of the nasal epithelium during the acute

105   phase of SARS-CoV-2 infection. The cases included SARS-CoV-2 infected individuals with an

106   extremely high viral load early in their illness matched to individuals who either had a low SARS-

107   CoV-2 viral load early in their infection or were otherwise stable patients who tested negative for

108   SARS-CoV-2 prior to their outpatient surgical or aerosol generating procedure. We also

109   determined the transcriptional response of a human nose organoid (HNO) line infected with

110   SARS-CoV-2 and compared it to transcriptomic profiles generated from the upper respiratory

111   tract secretion collected by nasal swabs from SARS-CoV-2 infected individuals.

112   Results

113   *Study cohort*

114   Ten SARS-CoV-2 cases were randomly selected from our population of adults with extremely

115   high viral load (Ct <16 to N1 target) at the time of their first RT-PCR positive test. For each high

116   viral load case, two additional human subjects were matched based on gender, week of first

117   SARS-CoV-2 RT-PCR test, age, and home zip code. These additional subjects consisted of

118   either 1) SARS-CoV-2 infected adults with low viral load (Ct 31-<40) (SARS-CoV-2 low viral

119   load case) or 2) stable adults who were SARS-CoV-2 RT-PCR negative (SARS-CoV-2 negative

120   control) for their out-patient surgical or aerosol generating procedure. Each high viral load case

121   had two to three subsequent SARS-CoV-2 RT-PCR positive mid-turbinate (MT) swab samples

122   collected over a 4-week period. Each SARS-CoV-2 low viral load case had similarly spaced

123   SARS-CoV-2 positive MT swab samples matched to its respective extremely high viral load

5

124   case. On the other hand, the SARS-CoV-2 negative control only had one MT-swab sample

125   collected with no longitudinal follow-up and was used to establish the transcriptomic baseline in

126   the respiratory epithelium during the time the extremely high and low viral load matched cases

127   were identified. The demographic and visit characteristics for the cohort is presented in Table 1.

128   In general, age, gender, race, ethnicity, and zip code were comparable between the extremely

129   high viral load, low viral load, and SARS-CoV-2 negative adults. The adults in the SARS-CoV-2

130   negative group were mostly asymptomatic at the time of testing, although their demographic

131   information was not significantly different compared to that of both the SARS-CoV-2 extremely

132   high and low viral load groups. The median Ct value difference between the extremely high and

133   low viral load groups were 794,672-fold (19.6 Ct difference) and 724-fold (9.5 Ct difference)

134   different at Visit 1 and Visit 2, respectively. At Visit 3, approximately 14 to 17 days after their

135   Visit 1, the Ct values were comparable between the two groups.

136

137   *RNA sequencing of serially collected specimens.*

138   Of the 73 MT swab samples from the extremely high and low viral load SARS-CoV-2 groups

139   with longitudinal follow-up and SARS-CoV-2 negative controls, only 44 (60.3%) MT swab

140   samples from 20 (66.7%) individuals were of good quality to generate RNA-sequence data to

141   study the host response to SARS-CoV-2 infection over time (Table 2). Demographic factors

142   such as age, gender, race, ethnicity, zip code, disease severity and co-morbid conditions were

143   comparable between the extremely high viral load, low viral load groups, and SARS-CoV-2

144   negative control group. Host response data were available on eight cases (extremely high viral

145   load) with 23 samples. Six of the 8 extremely high viral load cases had gene expression data for

146   Visits 1, 2, and 3, and two others for Visit 1 and 3. On the other hand, eight low viral load cases

147   had 17 samples with gene expression data. Only two of the low viral load cases had gene

6

148   expression data for Visit 1, 2, and 3. Another two low viral load cases had gene expression data

149   at Visit 1, Visit 2 or 3, and Visit 4. The remaining four low viral load cases had gene expression

150   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

151   negative control adults had gene expression data. All together 44 MT swab samples were

152   sequenced for RNA to observe gene expression changes in the host response of the cases with

153   extremely high viral load over time, as compared to the SARS-CoV-2 low viral load matched

154   cases and the negative controls.

155   *Gene expression changes by viral load*

156   From our RNA-seq dataset, we could identify widespread gene expression changes from the

157   nasal epithelium attributable to transcriptional host responses to SARS-CoV-2 infection. By

158   comparing the expression levels of each gene with the sample viral load (representing the

159   inverse correlation with Ct value) across the 44 MT swab samples, 425 genes were statistically

160   correlated at $p<0.01$ significance level and 112 genes at $p<0.001$ (Figure 1a, Pearson's

161   correlation). A stricter statistical cutoff would involve fewer expected false positive genes from

162   multiple testing. However, the above 425 genes with $p<0.01$ would still be highly enriched for

163   true positives, as revealed by integrating these genes with information from external databases,

164   as described below. We also compared the expression levels of genes at individual time points

165   during infection of both the extremely high viral load and low viral load groups with the SARS-

166   CoV-2 negative control group (Figure 1b). Comparing Visit 1 MT swab samples from the

167   extremely high viral load cases (n=8 samples from eight subjects) with the MT swab samples in

168   the SARS-CoV-2 negative control group (n=4) yielded the highest number of genes with

169   statistically significant correlated expression, as opposed to comparisons involving later times

170   for the extremely high viral load group or involving the low viral load group. The gene expression

171   from the extremely high viral load cases at Visit 1 highly overlapped with the differentially

172   expressed genes of the low viral load group at Visit 1 (Figure 1c) and remained highly correlated

7

173   throughout their last visit. Interestingly, genes from the extremely high viral load group that did

174   not overlap with the low viral load group did not show significant overlap with information from

175   external databases.

176   To further delineate the differences in host gene expression between extremely high and low

177   SARS-CoV-2 viral load groups, we performed an upset plot analysis to identify unique and

178   common intersecting genes between the samples (Figure S1). Among all the differentially

179   expressed genes (DEGs) in the samples, 614 DEGs were unique to the subjects in the

180   extremely high viral load group at visit 1 (first visit) and 226 genes were unique for the extremely

181   high viral load at the last visit. The low viral load subjects on the first and last visit showed 157

182   and 93 unique DEGs respectively. There were 31 DEGS that were common between all the

183   groups. We performed the Gene ontology (GO) analysis of the unique and overlapping DEGs

184   sets, and we found significant enrichment (FDR <0.05, count =3) of the biological processes

185   including defense response to virus, negative regulation of viral genome replication, innate

186   immune response, response to virus (Figure S2) that were uniquely expressed in the extremely

187   high viral load group at visit 1. SARS-CoV-2 infection in the low viral load group at either the

188   early or later phase of the infection and the extremely high viral load group at the last visit did

189   not show statistically significant enrichment of GO biological process. These findings indicate

190   that subjects with extremely high viral load at their initial visit were responding to the infection

191   with increased immune responses, and thus preventing prolonged viral infection with a poor

192   prognosis.

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

194   Focusing on the 112 top gene expression correlates of viral load across the 44 MT swab

195   samples (p<0.001, Pearson's), 108 of these genes were higher in the SARS-CoV-2 infected

196   adults with extremely high viral load. When visualizing the differential expression patterns of

8

197   these 108 genes by heat map (Figure 2a), the genes were highest at Visit 1 of the extremely

198   high viral load group, then decreased in expression with subsequent time points, tracking with

199   the decrease in viral load (i.e., increase in Ct value). The 108 genes showed intermediate

200   relative expression levels in the low viral load group and low expression in the SARS-CoV-2

201   negative control group. The 367 genes increased with extremely high viral load at $p<0.01$ were

202   highly enriched for functional gene categories, as defined by GO annotation terms. Enriched GO

203   terms (Figure 2b, $p<=3E-5$, one-sided Fisher's exact test) included 'immune system process',

204   'response to virus', 'type I interferon signaling pathway', 'cytokine-mediated signaling pathway',

205   'response to stress', 'regulation of viral life cycle', 'immune response', 'response to cytokine',

206   'innate immune response', 'response to interferon-gamma', 'regulation of I-kappaB kinase/NF-

207   kappaB signaling', 'JAK-STAT cascade', 'protein ubiquitination', 'regulation of cell death', 'T cell

208   activation', 'vesicle-mediated transport', and 'complement activation'. Of the 17 functional gene

209   categories, there were five gene categories - 'response to virus', 'type 1 interferon signaling',

210   'regulation of viral life cycle', 'response to interferon-gamma', and 'JAK-STAT cascade' – where

211   approximately 20% or higher of the genes were over expressed for that pathway. Overall, the

212   above gene categories were highly indicative as representing a host immune response to an

213   acute viral infection. Some of the genes that were upregulated were *EIF2AK2* (eukaryotic

214   translation initiation factor 2 alpha kinase 2), and *ZC3HAV1* (zinc finger CCCH-type containing,

215   antiviral 1), which have anti-viral activity. Other genes like *IFIT2* and *IFIT3* (interferon induced

216   protein with tetratricopeptide repeats) aid in apoptosis. Chemokine genes like *CXCL9* and

217   *CXCL10* that are involved in T-cell trafficking were also highly expressed. In contrast, the 62

218   genes decreased with high viral load at $p<0.01$ were not highly enriched for GO terms. Some of

219   the genes that were downregulated included *OR4A16* and *OR10X1*, involved in olfactory

220   responses; *SALL3* and *MAGB6*, which aid in downregulation of transcription; and *TUBA3E*,

221   *MLN*, and *ISTN1*, affecting tubulin functions.

9

222 *Comparison of the top over expressed genes in respiratory samples from the extremely high*

223 *and low viral load groups to other published data sets*

224 Genes correlated with SARS-CoV-2 viral load over time were similarly differentially expressed

225 across independent datasets of SARS-CoV-2 infected lung and upper airway cells (Figure 3).

226 We examined differential expression patterns for the top 112 genes, at p<0.001 significance

227 level, correlated with SARS-CoV-2 viral load across our serial sampling cohort (by Pearson's) in

228 two independent RNA-seq datasets of SARS-CoV-2 infection: one of lung cancer cell lines A549

229 and Calu-3 infected with SARS-CoV-2 for 24 hours from Blanco-Melo et al [24]. and one of

230 nasopharyngeal/oropharyngeal samples in 238 patients with COVID-19, other viral, or non-viral

231 acute respiratory illnesses from Mick et al [25]. As a group, the genes that positively correlated

232 with SARS-CoV-2 viral load were increased in SARS-CoV-2-infected Calu-3 cells and were high

233 in samples of human subjects infected with SARS-CoV-2 or other viruses (Figure 3a). For the

234 Mick et al. dataset, SARS-CoV-2 viral load data was available. Of the 112 genes correlated with

235 viral load in our dataset, 105 were in common with the Mick et al. dataset, and 99 (94%) of

236 these genes were positively correlated (Pearson's p<0.05) with viral load across the 94 SARS-

237 CoV-2 infected patients. In contrast to Calu-3, A549 infected cells did not show as strong a

238 correspondence to our 112-gene signature pattern. Taking the top genes that correlated

239 positively with SARS-CoV-2 viral load across the Mick et al. patient samples (p<0.01, Pearson's

240 correlation) and the top genes over-expressed in SARS-CoV-2-infected Calu-3 cells (p<0.01, t-

241 test), these significantly overlapped with the genes that positively correlated (p<0.01, Pearson's)

242 with SARS-CoV-2 viral load across our serially collected MT swab samples with a high overlap

243 among the respective dataset results (Figure 3b). The 136 genes overlapping among all three

244 datasets involved cytokines and inflammatory response pathways. In contrast, there was limited

245 overlap among the datasets involving genes under-expressed with SARS-CoV-2 infection

246 between our data set to either Mick *et al.* or Calu-3 cells (Figure 3c). This in part may reflect the

10

247   potential differences in the respective SARS-CoV-2 variants causing the infection or differences

248   in the illness severity of the host.

249   *Comparison of our respiratory sample gene sets to the transcriptional response of the human*

250   *nose organoid infected with SARS-CoV-2*

251   As another means to identify host transcriptional responses to SARS-CoV-2 infection, we

252   generated RNA-seq data on the human nose organoid model HNO [26]. We sampled HNO cells

253   infected with SARS-CoV-2 and mock control cells at 6hrs, 72hrs, and 6 days post-infection, and

254   we profiled these samples for gene expression. In the HNO204 RNA-seq dataset, 1760 genes

255   were statistically significant at $p<0.05$ significance level and 341 genes, at $p<0.01$, exceeding

256   chance expected. The top 867 genes over-expressed in HNO with SARS-CoV-2 infection

257   ($p<0.05$, t-test) showed significant overlapping patterns with the above-mentioned independent

258   RNA-seq datasets of SARS-CoV-2 infection (Figure 4a). Only a small, albeit statistically

259   significant, fraction of the HNO204 over-expressed genes overlapped with the top 367 genes

260   that correlated positively with SARS-CoV-2 viral load in our serial MT swab dataset (Figure 4b).

261   Of the 867 overexpressed genes in SARS-CoV-2 infected HNO, 35 overlapped with the 367

262   over expressed genes in the respiratory samples of extremely high and low viral load groups

263   ($p=1E-5$, one-sided Fisher's exact test). At the same time, a substantial fraction of the 867 HNO

264   genes overlapped with the genes high with SARS-CoV-2 infection in both A549 and Calu-3 lung

265   cancer cell lines (Figures 4a and 4b), with 178 Calu-3 genes overlapping ($p<1E-20$, one-sided

266   Fisher's exact test). In contrast, little overlap was observed between the genes under-expressed

267   with SARS-CoV-2 infection in HNO and genes similarly under-expressed with SARS-CoV-2 in

268   the other datasets (Figure 4c).

269   The 867 genes over-expressed in HNO at $p<0.05$ were significantly enriched for functional GO

270   gene categories. Enriched GO terms (Figure 4d, $p<=0.0001$, one-sided Fisher's exact test)

11

271   included 'vesicle', 'extracellular vesicle', 'intracellular vesicles', 'MAP kinase phosphatase

272   activity', 'regulation of locomotion', 'peptidase activator activity', 'endosome membrane',

273   'regulation of smooth muscle cell proliferation', 'regulation of cell motility', 'proteasomal protein

274   catabolic process', 'negative regulation of signaling', 'programmed cell death' , proteolysis

275   involved in cellular protein catabolic process', and 'inactivation of MAPK pathway'. Overall, the

276   over expressed genes are representative of the regulation of extracellular signaling from virus

277   infection on a wide range of cellular responses and function. The above findings of the HNO

278   transcriptional response to SARS-CoV-2 in relation to transcriptional responses observed in

279   other models and patient samples would suggest the existence of distinct host responses to

280   SARS-CoV-2 depending on cellular context, such as we previously observed between A549 and

281   Calu-3 lung cancer cell lines. The host response observed in HNO is reflective of a complex

282   epithelial cell population responding to a SARS-CoV-2 infection. On the other hand, the host

283   response genes detected in the upper respiratory tract secretion of our prospective longitudinal

284   cohort and those of Mick *et al.* patient samples are a composite of the epithelial and cellular

285   immune responses to the viral infection.

286   Discussion

287   The primary site for SARS-CoV-2 replication is thought to be the ciliated cells in the

288   nasopharynx or nasal olfactory mucosa. The viral replication initiates a signaling cascade to

289   promote the production of interferons and chemokines by epithelial cells and thereby promote

290   immune cell activation to control the virus. SARS-CoV-2 infection causes upregulation of

291   cytokines including IL-2, IL-6, IL-10, IL-12 and MCP-1 detected in tissues and serum, as well as

292   infiltration of infected tissues by inflammatory cells such as macrophages [27]. In the present

293   study, RNA seq analysis of MT swabs from SARS-CoV-2 infected individuals identified robust

294   induction of interferon inducible, cytokine, stress response, and immune-related genes. A

295   variety of genes such as *OAS2*, *PARP9*, *OASL*, *IFIT2*, *IFI3*, *CCL8*, *CXCL10*, etc., were highly

12

296   upregulated and correlated with high viral load, suggesting that innate immune response genes

297   were activated in a viral load dose response manner to control the viral infection. These results

298   are very consistent with recent studies from upper respiratory tract samples, which reported

299   upregulation of anti-viral factors and interferon response pathways [22,23,28].

300   In our study samples, the numbers of genes that were upregulated were much higher compared

301   to down regulated genes (367 vs 62). Some of genes that were downregulated included those

302   which operate olfactory functions (*OR4A16* and *OR10X1*), downregulation of transcription

303   (*SALL3* and *MAGB6*), and tubulin functions (*TUBA3E* and *MLN*, and *ISTN1*). Previous studies

304   have reported larger numbers of down regulated host response genes especially involving

305   olfactory receptor pathway, neutrophil degranulation, and vesicle formation—indicating the role

306   of these genes in loss of olfactory function in SARS-CoV-2 infections as well as the viral control

307   of host-cell machinery [20,22,23]. One other study also showed very low number of downregulated

308   genes with SARS-CoV-2 infection [29], one reason for the low number of downregulated genes

309   observed in our longitudinal study could conceivably relate to the mild illness experienced by

310   both the extremely high and low viral load groups, in addition to the timing of sample collection

311   as compared to other studies as well as the SARS-CoV-2 variants respectively involved.

312   Remarkably, the highest number of significant expressed genes were driven by the extremely

313   high viral load group at Visit 1 (first visit). Also, all the genes that were upregulated with the low

314   viral load group at Visit 1 completely overlapped with the extremely high viral load group at Visit

315   1 except for one gene, -*CNN2*, which plays a role in cell adhesion and muscle contraction. The

316   predominant sets of genes involved in defense response to virus, type I interferon signaling

317   pathway, cytokine-mediated signaling pathway—such as *CXCL10*, *TGFB*, *IFIT2*, *IFIT3*, *OAS1*,

318   and *IRF1*—were not found significantly upregulated in the low viral load group. Consonant with

319   this, Rouchka *et. al.* also observed that subjects with high viral loads had robust interferon and

320   cellular anti-viral response and even exhibited strong inverse correlation with disease severity [6].

13

321   We previously noted that some SARS-CoV-2 infected adults with low viral load experienced

322   prolonged viral shedding and low fluctuation in viral load over time [14]. Absence or low

323   expression of the anti-viral response in the low viral load group strengthens our observation of

324   prolonged shedding in adults with a low viral load early in infection.

325   In our longitudinal study, the up-regulated host response genes that correlated with SARS-CoV-

326   2 viral load over time in the respiratory secretion collected by the MT swabs were similarly

327   differentially expressed across independent data sets of SARS-CoV-2 infected lung and upper

328   airway cells [24]. About 170 of the differentially expressed genes observed in our study

329   overlapped with SARS-COV-2 infected Calu-3 lung adenocarcinoma cell line but not with A549

330   cells. The observed difference across the cell lines could possibly be attributed to A549 cells not

331   supporting robust replication of SARS-CoV-2 due to the low expression of ACE-2 [30]. Similarly,

332   207 up-regulated genes from our longitudinal study overlapped with nasopharyngeal swabs

333   from SARS-CoV-2 infected patients (3). Genes involved in cytokines and inflammatory response

334   pathways were the ones that overlapped the most, demonstrating that anti-viral innate immune

335   responses are common with SARS-CoV-2 infections. In addition, the up-regulation of

336   differentially expressed genes related to an inflammatory response in COVID-19 patients can

337   result in the induction of interleukin-6 (IL-6), CXCL10 (IP-10), and TNF-α with hyperactivation of

338   Th1/Th17 responses that results the recruitment and activation of pro-inflammatory neutrophils

339   and macrophages into the airways [31]. This has been proposed as the prime reason for failure to

340   resolve inflammation in severely symptomatic patients [31,32].

341   To better understand the contribution of epithelial cellular responses to SARS-CoV-2, we

342   compared differentially expressed genes in the respiratory secretion of adults infected with

343   SARS-CoV-2 to those that were expressed in HNO infected with SARS-CoV-2. A small, albeit

344   statistically significant, fraction (35 of 867) of the HNO up-regulated genes overlapped with the

345   367 differentially expressed up-regulated genes detected from the SARS-CoV-2 cases from our

14

346   longitudinal cohorts. These included functional genes involved with intrinsic antiviral immunity

347   and interferon signaling representing the epithelial cellular responses to SARS-CoV-2 infection.

348   A greater number of up-regulated genes overlap between our longitudinal cohorts [170 (46.3%)

349   of 367 genes] and the SARS-CoV-2 infected Calu-3 cell line [170 (9.3%) of 1836 genes]

350   compared to our SARS-CoV-2 infected HNO204 line [35 (4.0%) of 867 genes]. This could

351   reflect the difference in cellular complexity between the cell lines and greater diversity of the

352   HNO epithelium resulting in fewer overlapping up-regulated genes. HNO204 is a complex

353   pseudostratified epithelium composed of at least 9 different cell types including ciliated, goblet,

354   secretory and basal cells [33,34]. In contrast, the Calu-3 cell line, was generated from a bronchial

355   adenocarcinoma, a submucosal gland cell line of a single cell type [35]

356   Previous studies have demonstrated high expression of ACE2 in SARS-CoV-2 infected

357   nasopharyngeal samples and these were greatly elevated in high viral load subjects, suggesting

358   that higher replication occurs with increased receptor expression [22]. In our cohort we did not

359   observe a statistically significant increase in *ACE2* expression in both extremely high and low

360   viral load groups. However, the expression of *ACE2* was elevated in our HNO infected with

361   SARS-CoV-2 but not *TMPRSS2*, which has increased expression in nasal airway epithelial

362   brushings [36].

363   In summary, our longitudinal study investigated gene expression patterns in SARS-CoV-2

364   infected individuals with an extremely high viral load displayed strong immune responses that

365   decreased over time, and eventually became comparable to those with low viral loads. We

366   detected hundreds of up-regulated genes that were highly correlated to the SARS-CoV-2 viral

367   load. Enriched cellular pathways involved in the innate immune response, antiviral interferon

368   responses were observed in other cohorts of SARS-CoV-2 infected adults. A limited but highly

369   significant up-regulated gene response overlapped with our human nose organoid line, a

370   complex pseudostratified ciliated epithelium, suggesting that the gene expression profile

15

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

375   Materials and Methods

376   *Study cohort*

377   Ten extremely high, viral load SARS-CoV-2 positive cases were matched to 10 low viral load

378   SARS-CoV-2 positive adults, and 10 stable adults (SARS-CoV-2 negative controls) who were

379   cleared for having an out-patient surgical or aerosol generating procedure. The cases and

380   controls were selected from our population of 17,644 adults (24,822 samples) evaluated in the

381   outpatient clinics at Baylor College of Medicine (BCM) and their affiliate institutions from March

382   18, 2020, through January 16, 2021, as previously described[14]. Three distinct adult populations

383   were tested: 1) symptomatic employees utilizing occupational health services, 2) patients

384   evaluated at medical and surgical clinics, and 3) patients who required clearance for an out-

385   patient surgical or aerosol generating procedure. Serial samples were obtained from individuals

386   who came back to be tested for evidence that the virus was cleared or were enrolled as sub-

387   study to determine the viral shedding kinetics. Testing for SARS-CoV-2 was performed in our

388   Clinical Laboratory Improvement Amendments (CLIA) Certified Respiratory Virus Diagnostic

389   Laboratory (ID#: 45D0919666). Although RT-PCR testing was performed as a service to BCM,

390   the collection of metadata was performed under an Institutional Review Board approved

391   protocol with waiver of consent.

392   The extremely high viral load cases consisted of adults with an extremely high viral load (Ct

393   <16) for the N1 target on their first mid-turbinate (MT) sample and had at least two subsequent

394   positive MT samples [14]. Of the 104 individuals with an extremely high viral load in their first test,

395   30 individuals met the criteria for multiple positive samples over the ensuing 4 weeks. Adults

396   from two other groups were matched to each extremely high viral load case: a low viral load (Ct

397   31-<40) SARS-CoV-2 positive adult (SARS-CoV-2 low viral load) and an otherwise stable

398   control who tested negative for SARS-CoV-2 (SARS-CoV-2 negative control) and was cleared

17

399   for an out-patient surgical or aerosol generating procedure. Of the 453 individuals with a low

400   viral load in their first test, 126 individuals met the criteria for multiple positive samples over the

401   ensuing 4 weeks. The extremely high viral load cases were matched to the other two groups by

402   gender, week of first test (+ 1 week), age (+ 1 year) and zip code (5 digits). If a match could not

403   be found the range of the factors were expanded to + 3 weeks of first test, + 10 years and 3

404   digits for the zip code. The ten extremely high viral load cases were randomly selected from our

405   pool of 30 individuals with an extremely high viral load with multiple positive MT samples. The

406   best matched SARS-CoV-2 low viral load case and negative control were then selected for each

407   extremely high viral load case.

408   *SARS-CoV-2 RT- PCR*

409   Viral RNA extraction and RT-PCR testing was performed as previously described (Avadhanula

410   et al., 2021). In brief, viral RNA was extracted using the Qiagen Viral RNA Mini Kit (QIAGEN

411   Sciences, Maryland, USA) with an automated extraction platform QIAcube (QIAGEN, Hilden,

412   Germany). The extracted RNA samples were tested by CDC 2019-novel coronavirus (2019-

413   ncoV) Real-Time RT-PCR Diagnostic panel [CDC 2019-Novel Coronavirus (2019-nCoV) Real-

414   Time RT-PCR Diagnostic Panel for Emergency Use Only Instructions for Use]. RT-PCR

415   reaction was set up using TaqPath™ 1-Step RT-qPCR Master Mix, CG (Applied Biosystems,

416   CA) and run on 7500 Fast Dx Real-Time PCR Instrument with SDS 1.4 software. Respiratory

417   samples with cycle threshold (Ct) values <40 for both N1 and N2 primers were considered RT-

418   PCR positive for SARS-CoV-2.

419   *Human Nose organoid model.*

420   The differentiated human nose organoid derived air liquid interface (HNO-ALI) cells were

421   apically infected with SARS-CoV-2 [Isolate USA-WA1/2020, obtained from Biodefense and

422   Emerging Infectious resources (BEI)] at a multiplicity of infection of 0.01 or mock infected with

18

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

19

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

20

472   nasopharyngeal/oropharyngeal samples in 238 patients with COVID-19, other viral, or non-viral

473   acute respiratory illnesses [25], RPM values were quantile normalized before the analysis.

474   *Statistical analysis*

475   All p-values were two-sided unless otherwise specified. We performed all tests using log2-

476   transformed gene expression values. False Discovery Rates (FDRs) due to multiple testing of

477   genes were estimated using the method of Storey and Tibshirini [40]. Even in instances of

478   nominally significant genes only moderately exceeding chance expectations by FDR, the

479   nominally significant genes were found in downstream enrichment analyses (involving functional

480   gene sets and results of external SARS-CoV-2-related RNA-seq datasets) to contain molecular

481   information representing real biological differences. We evaluated enrichment of GO annotation

482   terms [41] within sets of differentially expressed genes using SigTerms software [42] and one-sided

483   Fisher's exact tests. Visualization using heat maps was performed using JavaTreeview (version

484   1.1.6r4) [43,44]. Gene ontology (GO) analysis of DEGs used in the upset plot (Figure S1) was

485   performed using the web-based Database for Annotation, Visualization, and Integrated

486   Discovery (DAVID; version - v2023q1) [45,46]

487   *Data Availability*

488   The RNA-seq dataset of serially collected samples and of nose organoids will be deposited at

489   Gene Expression Omnibus (GEO) (GEO accession number pending). In terms of previously

490   published data, we obtained RNA-seq expression data from experimental models of SARS-

491   CoV-2 viral infection or other treatments from GEO (GSE147507). The Mick et al. RNA-seq

492   dataset is available at GEO (GSE156063).

493

21

494   Figure Legends

495   **Figure 1. Differential gene sets associated with the transcriptional host response to**

496   **SARS-CoV-2 infection across serially collected samples. (a)** For each of 20000 genes,

497   expression (log2 FPKM) was correlated with viral load (inverse correlation with Ct value) across

498   44 samples from 20 subjects. Numbers of statistically significant genes (by Pearson's) at both

499   p<0.01 and p<0.001 significance levels are represented, as compared to the chance expected

500   by multiple testing. **(b)** Numbers of differential genes (p<0.01, t-test) when comparing: 1) Visit 1

501   samples from the extremely high viral load group (n=8 samples from eight subjects) with the

502   samples in the negative group (n=4); samples at the latest time points for each of the subjects

503   from the extremely high viral load group (n=8 samples) with the samples in the negative group;

504   samples from the low viral load group (n=8 samples from eight subjects, using earliest time

505   point) with the samples in the negative group. Chance expected genes at p<0.01 due to multiple

506   testing would be on the order of 200 [40]. **(c)** Heat map comparing differential patterns across the

507   three comparisons from part b, for the 1357 genes significant (p<0.01) for any comparison.

508   Columns off to the side indicated which genes were correlated with viral load (p<0.01) across all

509   44 samples (from part a), and which genes have Gene Ontology (GO) annotation [41] 'response

510   to virus'.

511   **Figure 2. Differential expression patterns and functional gene groups associated with**

512   **SARS-CoV-2 viral load across serially collected samples. (a)** Across 44 MT swab samples

513   representing 20 subjects, differential gene expression patterns for the set of 112 genes

514   significantly correlated with SARS-CoV-2 viral load (i.e., inversely correlated with Ct value) at

515   p<0.001 (Pearson's) are represented. Heat map contrast (bright yellow/blue) is 3-fold change

516   from the average of the samples from the low viral load group. Genes listed off to the right have

517   GO annotation 'response to virus'. Extremely high viral load, Ct<20. **(b)** Selected significantly

518   enriched GO terms [41] within the genes over-expressed with SARS-CoV-2 viral load (p<0.01,

22

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-

        23

544   expressed in HNO204 with SARS-CoV-2 infection (p<0.05, t-test) are represented here. Next to

545   the HNO204 dataset are the corresponding patterns for independent RNA-seq datasets of

546   SARS-CoV-2 infection: lung cancer cell lines (A549 and Calu-3)[37], our serially collected MT

547   swab samples from patients, and nasopharyngeal/oropharyngeal samples from Mick et al [25].

548   Gene order is the same across all datasets. Heat map contrast (bright yellow/blue) is 3-fold

549   change from the corresponding comparison group. **(b)** Venn diagram representing the gene set

550   overlaps among the genes increased with SARS-CoV-2 infection in each of the following RNA-

551   seq datasets: HNO204, serial MT swab, and Calu-3 lung cancer cell line. Gene set enrichment

552   p-values by one-sided Fisher's exact test. Genes overlapping between HNO204 and serial MT

553   swab datasets are listed. **(c)** Similar to part b, but for genes decreased with SARS-CoV-2

554   infection. **(d)** Selected significantly enriched GO terms [41] within the genes over-expressed with

555   SARS-CoV-2 infection in HNO204 (p<0.05, t-test). For each GO term, enrichment p-values and

556   numbers of genes in the SARS-CoV-2-associated gene set are indicated. Enrichment p-values

557   by one-sided Fisher's exact test.

24

558   References

559   1.   Fajnzylber J, Regan J, Coxen K, et al. SARS-CoV-2 viral load is associated with

560   increased disease severity and mortality. *Nat Commun.* 2020;11(1):5493.

561   doi:10.1038/s41467-020-19057-5

562   2.   He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of

563   COVID-19. *Nat Med.* 2020;26(5):672-675. doi:10.1038/s41591-020-0869-5

564   3.   Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection: A Narrative

565   Review. *Ann Intern Med.* 2020;173(5):362-367. doi:10.7326/M20-3012

566   4.   Pujadas E, Chaudhry F, McBride R, et al. SARS-CoV-2 viral load predicts COVID-19

567   mortality. *Lancet Respir Med.* 2020;8(9):e70. doi:10.1016/S2213-2600(20)30354-4

568   5.   Lucas C, Klein J, Sundaram ME, et al. Delayed production of neutralizing antibodies

569   correlates with fatal COVID-19. *Nat Med.* 2021;27(7):1178-1186. doi:10.1038/s41591-

570   021-01355-0

571   6.   Rouchka EC, Chariker JH, Alejandro B, et al. Induction of interferon response by high

572   viral loads at early stage infection may protect against severe outcomes in COVID-19

573   patients. *Sci Rep.* 2021;11(1):15715. doi:10.1038/s41598-021-95197-y

574   7.   Rydyznski Moderbacher C, Ramirez SI, Dan JM, et al. Antigen-Specific Adaptive

575   Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease

576   Severity. *Cell.* 2020;183(4):996-1012.e19. doi:10.1016/j.cell.2020.09.038

577   8.   Takahashi T, Ellingson MK, Wong P, et al. Sex differences in immune responses that

578   underlie COVID-19 disease outcomes. *Nature.* 2020;588(7837):315-320.

579   doi:10.1038/s41586-020-2700-3

580   9.   Puhach O, Adea K, Hulo N, et al. Infectious viral load in unvaccinated and vaccinated

581   individuals infected with ancestral, Delta or Omicron SARS-CoV-2. *Nat Med.*

582   2022;28(7):1491-1500. doi:10.1038/s41591-022-01816-0

583   10.   Liu J, Liao X, Qian S, et al. Community transmission of severe acute respiratory

584   syndrome Coronavirus 2, Shenzhen, China, 2020. *Emerg Infect Dis.* 2020;26(6):1320-

585   1323. doi:10.3201/eid2606.200239

586   11.   Jones TC, Biele G, Mühlemann B, et al. Estimating infectiousness throughout SARS-

587   CoV-2 infection course. *Science (80-).* 2021;373(6551):eabi5273.

588   doi:10.1126/science.abi5273

589   12.   van Kampen JJA, van de Vijver DAMC, Fraaij PLA, et al. Duration and key determinants

590   of infectious virus shedding in hospitalized patients with coronavirus disease-2019

591   (COVID-19). *Nat Commun.* 2021;12(1):267. doi:10.1038/s41467-020-20568-4

592   13.   Puhach O, Meyer B, Eckerle I. SARS-CoV-2 viral load and shedding kinetics. *Nat Rev*

593   *Microbiol.* Published online 2022. doi:10.1038/s41579-022-00822-w

594   14.   Avadhanula V, Nicholson EG, Ferlic-Stark L, et al. Viral Load of Severe Acute

595   Respiratory Syndrome Coronavirus 2 in Adults During the First and Second Wave of

596   Coronavirus Disease 2019 Pandemic in Houston, Texas: The Potential of the

597   Superspreader. *J Infect Dis.* Published online February 15, 2021.

598   doi:10.1093/infdis/jiab097

599   15.   Goyal A, Reeves DB, Cardozo-Ojeda EF, Schiffer JT, Mayer BT. Viral load and contact

600   heterogeneity predict SARS-CoV-2 transmission and super-spreading events. Walczak

601   AM, Childs L, Forde J, eds. *Elife.* 2021;10:e63537. doi:10.7554/eLife.63537

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

603   Infection Gradient in the Respiratory Tract. *Cell*. 2020;182(2):429-446.e14.

604   doi:10.1016/j.cell.2020.05.042

605 17. Schultze JL, Aschenbrenner AC. COVID-19 and the human innate immune system. *Cell*.

606   2021;184(7):1671-1692. doi:10.1016/j.cell.2021.02.029

607 18. Chua RL, Lukassen S, Trump S, et al. COVID-19 severity correlates with airway

608   epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol*.

609   2020;38(8):970-979. doi:10.1038/s41587-020-0602-4

610 19. Nicholson EG, Schlegel C, Garofalo RP, et al. Robust cytokine and chemokine response

611   in nasopharyngeal secretions: Association with decreased severity in children with

612   physician diagnosed bronchiolitis. *J Infect Dis*. 2016;214(4). doi:10.1093/infdis/jiw191

613 20. Xiong Y, Liu Y, Cao L, et al. Transcriptomic characteristics of bronchoalveolar lavage fluid

614   and peripheral blood mononuclear cells in COVID-19 patients. *Emerg Microbes Infect*.

615   2020;9(1):761-770. doi:10.1080/22221751.2020.1747363

616 21. Daamen AR, Bachali P, Owen KA, et al. Comprehensive transcriptomic analysis of

617   COVID-19 blood, lung, and airway. *Sci Rep*. 2021;11(1):7052. doi:10.1038/s41598-021-

618   86002-x

619 22. Butler D, Mozsary C, Meydan C, et al. Shotgun transcriptome, spatial omics, and

620   isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral

621   diversification, and drug interactions. *Nat Commun*. 2021;12(1):1660.

622   doi:10.1038/s41467-021-21361-7

623 23. Lieberman NAP, Peddu V, Xie H, et al. In vivo antiviral host transcriptional response to

624   SARS-CoV-2 by viral load, sex, and age. *PLOS Biol*. 2020;18(9):e3000849.

625   https://doi.org/10.1371/journal.pbio.3000849

27

626 24. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, et al. Imbalanced Host Response to SARS-

627   CoV-2 Drives Development of COVID-19. *Cell.* 2020;181(5):1036-1045.e9.

628   doi:10.1016/j.cell.2020.04.026

629 25. Mick E, Kamm J, Pisco AO, et al. Upper airway gene expression reveals suppressed

630   immune responses to SARS-CoV-2 compared with other respiratory viruses. *Nat*

631   *Commun.* 2020;11(1):5854. doi:10.1038/s41467-020-19587-y

632 26. Rajan A, Weaver AM, Aloisio GM, et al. The Human Nose Organoid Respiratory Virus

633   Model: an Ex Vivo Human Challenge Model To Study Respiratory Syncytial Virus (RSV)

634   and Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Pathogenesis

635   and Evaluate Therapeutics. *MBio.* 2022;13(1):e0351121. doi:10.1128/mbio.03511-21

636 27. Tjan LH, Furukawa K, Nagano T, et al. Early Differences in Cytokine Production by

637   Severity of Coronavirus Disease 2019. *J Infect Dis.* 2021;223(7):1145-1149.

638   doi:10.1093/infdis/jiab005

639 28. Zhang C, Feng Y-G, Tam C, Wang N, Feng Y. Transcriptional Profiling and Machine

640   Learning Unveil a Concordant Biosignature of Type I Interferon-Inducible Host Response

641   Across Nasal Swab and Pulmonary Tissue for COVID-19 Diagnosis. *Front Immunol.*

642   2021;12:733171. doi:10.3389/fimmu.2021.733171

643 29. Saravia-Butler AM, Schisler JC, Taylor D, et al. Host transcriptional responses in nasal

644   swabs identify potential SARS-CoV-2 infection in PCR negative patients. *iScience.*

645   2022;25(11):105310. doi:10.1016/j.isci.2022.105310

646 30. Chang C-W, Parsi KM, Somasundaran M, et al. A Newly Engineered A549 Cell Line

647   Expressing ACE2 and TMPRSS2 Is Highly Permissive to SARS-CoV-2, Including the

648   Delta and Omicron Variants. *Viruses.* 2022;14(7). doi:10.3390/v14071369

28

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.

29

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

30

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

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

706

707

31

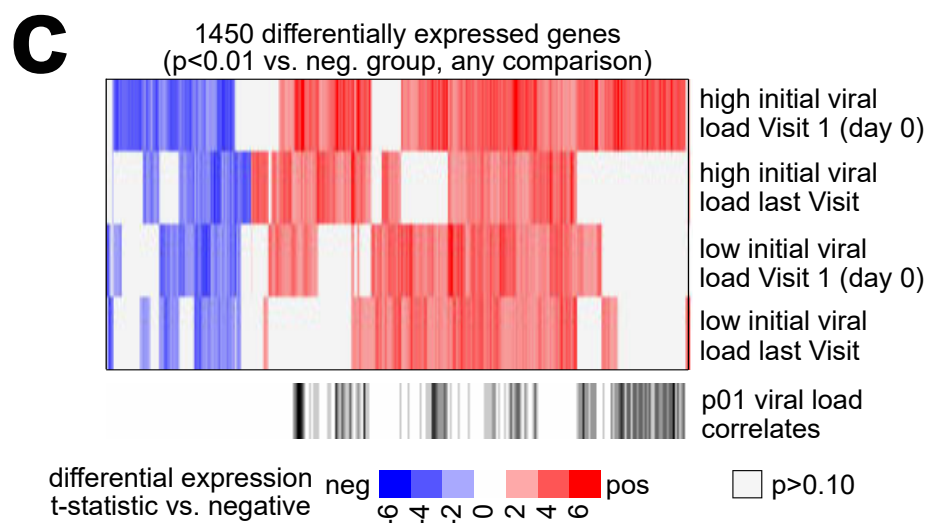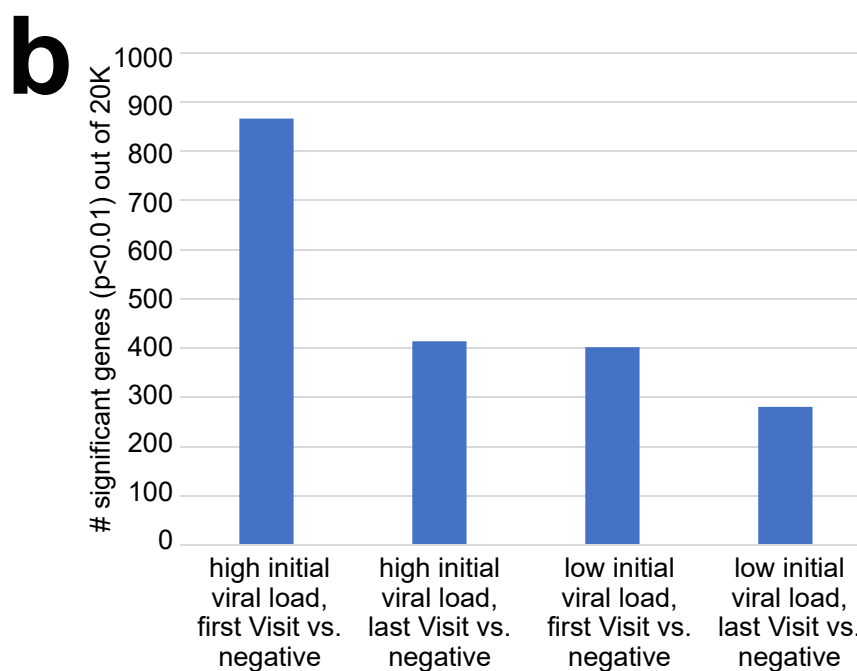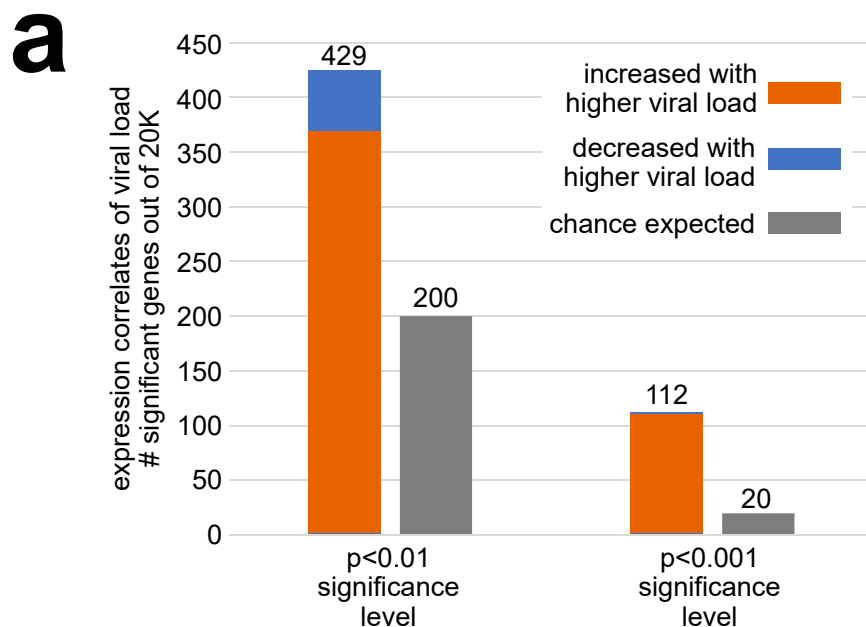708   TABLE 1. Demographic and Visit Characteristics by Matched Groups

	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	

709   Abbreviations: Ct=cycle threshold, [a]Median (Min, Max)[b]Median (IQR) **or Median (Q1-Q3) or Median**

710   **(25th percentile-75th percentile) or Median (lower quartile-upper quartile)**

711   Differences between groups were determined using the Kruskal-Wallis test for variables with non-

712   parametric distribution and by Fisher's Exact test for categorical variables. P-value <0.05 was considered

713   significantly different between groups.

32

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
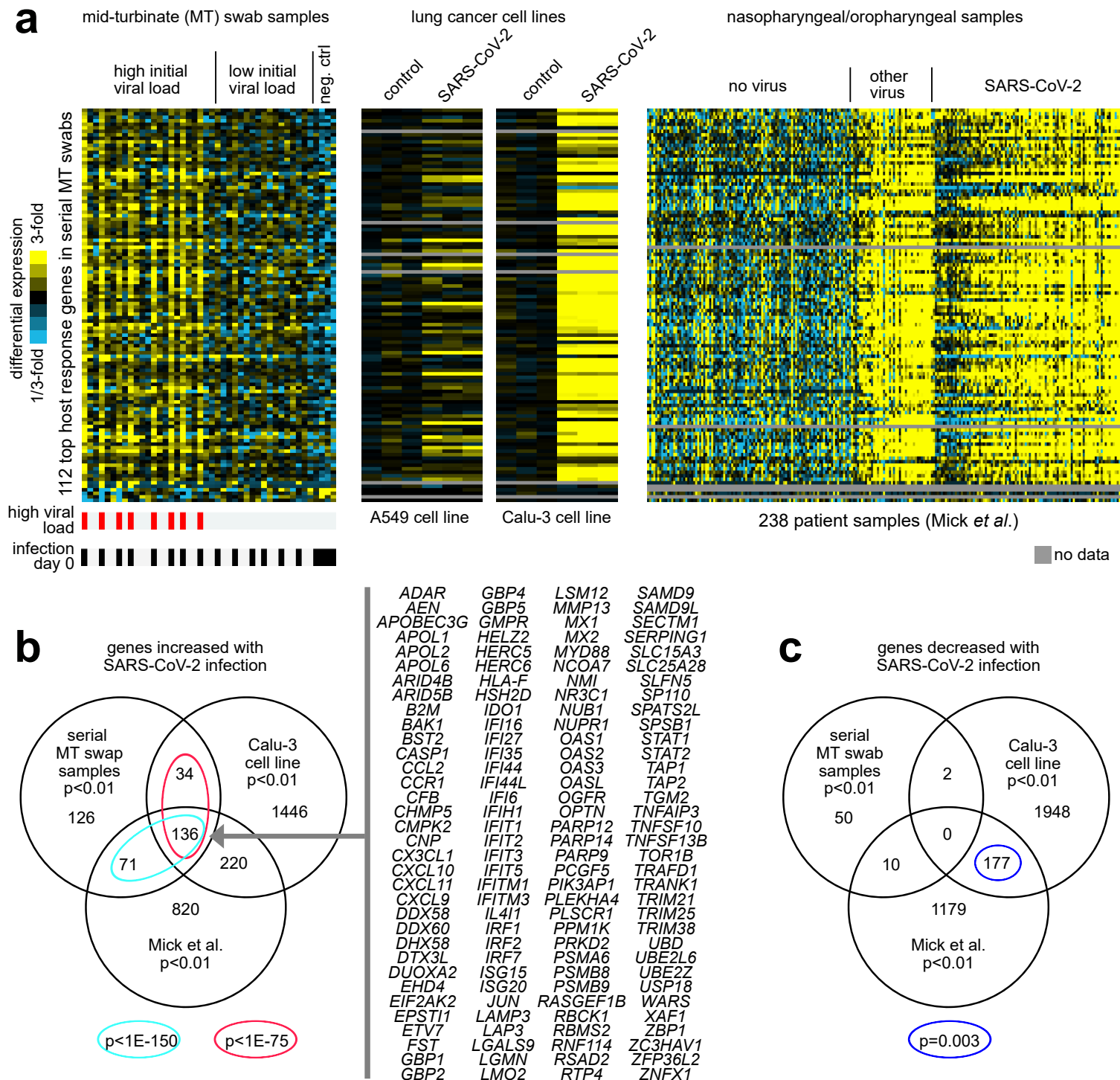717   significantly different between groups.

33

a expression correlates of viral load

significant genes out of 20K

Legend:
- increased with higher viral load (orange)
- decreased with higher viral load (blue)
- chance expected (gray)

p<0.01 significance level: 429, 200
p<0.001 significance level: 112, 20

b # significant genes (p<0.01) out of 20K

- high initial viral load, first Visit vs. negative
- high initial viral load, last Visit vs. negative
- low initial viral load, first Visit vs. negative
- low initial viral load, last Visit vs. negative

c 1450 differentially expressed genes
(p<0.01 vs. neg. group, any comparison)

- high initial viral load Visit 1 (day 0)
- high initial viral load last Visit
- low initial viral load Visit 1 (day 0)
- low initial viral load last Visit

p01 viral load correlates

differential expression t-statistic vs. negative: neg — pos, -6 -4 -2 0 2 4 6, p>0.10

a

differential expression

1/3-fold | 3-fold | no data

nose organoid | lung cancer cell lines | serial MT swab samples | 238 patient samples (Mick *et al.*)

control | SARS-CoV-2 | control | SARS-CoV-2 | control | SARS-CoV-2 | high initial viral load | low initial viral load | neg. ctrl | no virus | other virus | SARS-CoV-2

867 top host response genes in nose organoid

HNO204 | A549 | Calu-3

high SARS-CoV-2 viral load

SARS-CoV-2 viral load

b

genes increased with SARS-CoV-2 infection

serial MT swap samples p<0.01 — 178

Calu-3 cell line p<0.01 — 154

1504

16

19

162

670

nose organoid (HNO204) p<0.05

p=1E-5 | p<1E-20

AEN
APOL1
BAK1
DUOXA2
DUSP8
EHD4
GBP2
IDO1
IER5
ISG20
KYNU
MED18
MMP13
RAB8B
RASGEF1B
SQSTM1

ACAA1
APOBEC3B
AVPI1
CLIC4
DUSP7
FAAP24
FGD2
IL1RN
KCNK7
MICB
MUC15
NABP1
NOSIP
RAB29
RHOG
SCO2
SERTAD1
SLC31A2
TNFRSF12A

c

genes decreased with SARS-CoV-2 infection

serial MT swap samples p<0.01 — 60

Calu-3 cell line p<0.01 — 2031

2

0

0

94

799

nose organoid (HNO204) p<0.05

NS

d

-Log10(enrichment p-value)

# of genes (out of 867 total)

vesicle
extracellular vesicle
intracellular vesicle
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
inactivation of MAPK pathway

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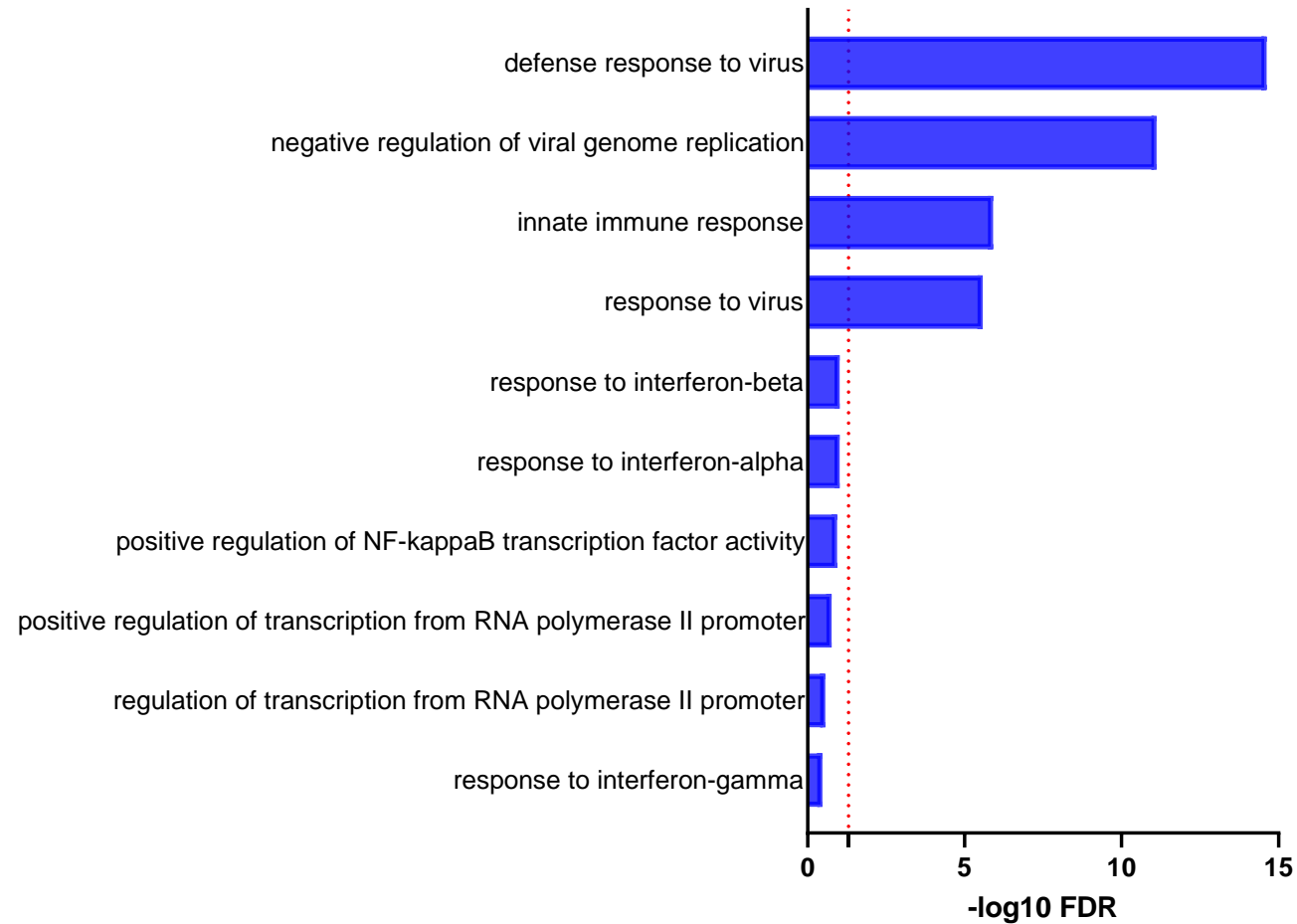
Figure S2: Gene ontology analysis of differentially expressed genes present only in high day 0 infection. The top10 enriched GO terms for biological processes altered in shown. Significantly enriched GO terms with a minimum three enriched genes were ranked by significance. The x-axis denoting the negative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