An increased ratio of SARS-CoV-2 positive to negative sense genomic and subgenomic RNAs within routine diagnostic upper respiratory tract swabs may be a marker of virion shedding

Anthony Chamings [1,2], Tarka Raj Bhatta [1,2,] and Soren Alexandersen[1,2,3,*]

[1]Geelong Centre for Emerging Infectious Diseases, Geelong, VIC 3220, Australia; [2]Deakin University, Geelong, VIC 3220, Australia; [3] Barwon Health, University Hospital Geelong, Geelong, VIC 3220 Australia

* Correspondence: soren.alexandersen@deakin.edu.au; Tel.: +61-0-342159635

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread rapidly in the global population since its emergence in humans in late 2019. Replication of SARS-CoV-2 is characterised by transcription and replication of genomic length RNA and shorter subgenomic RNAs to produce virus proteins and ultimately progeny virions. Here we explore the pattern of both genome-length and subgenomic RNAs and positive and negative strand SARS-CoV-2 RNAs in diagnostic nasopharyngeal swabs using sensitive probe based PCR assays as well as Ampliseq panels designed to target subgenomic RNAs. Using these assays, we measured the ratios of genomic to subgenomic RNAs as well as the ratios of positive to negative strand RNAs in SARS-CoV-2 positive nasopharyngeal swab samples. We found that while subgenomic RNAs and negative strand RNA can be readily detected in swab samples taken up to 19 and 17 days post symptom onset respectively, and therefore their detection alone is not likely an indicator of active SARS-CoV-2 replication. However, the ratios of genomic-length to subgenomic RNA and also of positive to negative strand RNA were elevated in some swabs, particularly those collected around the onset of clinical symptoms or in an individual with decreasing PCR Ct's in successive swab samples. We tentatively conclude that it may be possible to refine such molecular assays to help determine if active replication of virus is occurring and progeny virions likely present in a SARS-CoV-2 positive individual. Assays targeting subgenomic N or ORF7a RNAs as well as strand specific ORF7a total genome-length and subgenomic RNAs may be the most sensitive for this purpose as these targets were consistently the most abundant in the swab samples.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of human coronavirus disease 2019 (COVID-19), is a novel betacoronavirus which was first detected in humans in late 2019[1]. The virus was readily transmittable from person to person and rapidly spread worldwide causing unprecedented economic and social disruption in many countries[2]. While infection can result in severe life-threatening respiratory disease and death in some individuals, a large proportion of people will be asymptomatic or only exhibit mild respiratory signs[1,3]. Therefore, knowing who is infected and when they may possibly be infectious cannot be done based on clinical symptoms alone, and appropriate public health interventions rely on knowing who is infected and therefore possibly infectious. The most commonly used method to determine the infection status of an individual is the real-time reverse transcription PCR (rRT-PCR)[4]. The rRT-PCR detects the presence of SARS-CoV-2 RNA and will not indicate the presence of infectious virus particles as some individuals can test positive for virus RNA for 3 weeks or more[4] after they were thought to be initially infected, and long after the time when infectious virus shedding is thought to cease which is around 7-8 days post detection of symptoms in most people[5-7]. The gold standard method for detecting whether infectious virus is being excreted is inoculation of susceptible cell cultures, which is not practical in most diagnostic laboratories, and importantly, lacks sensitivity and consequently will not detect low levels of potential infectivity[5,8]. This has led to a number of other molecular indicators being proposed as possible markers of whether an individual is likely to be infectious or not to aid in the control of SARS-CoV-2[5].

SARS-CoV-2 is an enveloped virus with a single stranded positive sense RNA genome of nearly 30,000 nucleotides. The genomic RNA consists of a 5' UTR, two large open reading frames (ORFs) - ORF1a and ORF1b, which occupy two thirds of the 5' end of the genome and which express two large polyproteins which are proteolytically cleaved into 16 known non-structural proteins, and several ORFs which encode the structural and accessory proteins needed for the virus to replicate/transcribe and produce progeny virions, and a polyadenylated 3' UTR[9]. After initial infection, virus replication inside a cell is believed to involve initial translation of the ORF1ab proteins from the genomic RNA leading to formation of replicative organelles, so-called double membrane vesicles, followed by the production of negative stranded RNA copies of the SARS-CoV-2 genome which serve as templates for new positive strand copies of the genomic RNA (gRNA) to be used as either mRNA for additional production of ORF1ab proteins or later in the infectious cycle of the individual infected cell, to be packaged inside progeny virus particles. A number of shorter subgenomic RNAs (sgRNA) are also produced via a complex method of discontinuous negative strand RNA synthesis serving as templates for generation of positive strand RNA copies which serve as messenger RNAs used to express each of the structural and accessory proteins of SARS-CoV-2[9,10].

Each subgenomic RNA molecule shares a common leader sequence of approximately 65-69 nucleotides within the 5'UTR of the SARS-CoV-2 genome[11]. During negative strand RNA synthesis, the viral RNA polymerase pauses at a transcription regulatory sequence (TRS) upstream from the ORFs responsible for encoding the structural and accessory proteins within the 3' third of the genome. The nascent RNA is then joined to the TRS within the leader sequence creating a negative sense sgRNA which is used as a template to make positive sense sgRNA's as mentioned above. The site of coronavirus RNA replication is not completely defined, but is thought to be closely associated with double membrane vesicles and channel like structures produced from intracellular membranes which are thought to spatially bring together the necessary components required for RNA synthesis and to protect the single and double stranded RNA as it is produced[10,12].

With the replication biology of coronaviruses in mind, several studies have proposed using sgRNA or negative sense SARS-CoV-2 RNA as a marker of active replication of the virus within an individual[5,13]. However, we recently reported that we were able to detect sgRNA and negative sense RNA in samples up to 17 days post infection and that sgRNAs are relatively stable and likely protected by the double membrane structures created during the replication of the virus in the cytoplasm of infected cells[11], and therefore sgRNAs, or the presence of any negative strand RNA, do not necessarily indicate that virus replication is currently occurring, only that replication of SARS-CoV-2 has occurred at some point during infection[11]. Other authors have similarly suggested that the amount of sgRNA is simply related to the amount of total SARS-CoV-2 RNA present within a sample and correlates poorly with the shedding of infectious virus[7,14].

Given that there is ongoing interest in sgRNAs as a potential marker of active replication and a dichotomy of conclusions being drawn within the scientific community as to the utility of sgRNA as a marker of infectivity, we decided to further examine the presence of sgRNA's in a larger number of routine diagnostic SARS-CoV-2 positive nasopharyngeal swab samples collected subsequent to our last study during the second SARS-CoV-2 epidemic wave in Victoria, Australia in mid-to late 2020[11,15]. In addition to including more routine diagnostic samples, we also developed and used more sensitive molecular assays to study subgenomic RNAs as well as negative and positive strand SARS-CoV-2 RNA in such samples and investigated if there was any relationship between sgRNA, genomic RNA, positive and negative strand RNA, sample characteristics including the time of onset of clinical symptoms and the local immune response occurring in the upper respiratory epithelium that could be useful in informing management of SARS-CoV-2 positive individuals.

We found that in some SARS-CoV-2 positive samples collected at the very onset of clinical symptoms, that both the ratios of genomic to subgenomic RNA and importantly the ratio of positive to negative strand RNA were slightly higher compared to samples taken later in the clinical course of infection where these ratios fell to lower and more stable levels. We were able to detect subgenomic

104 RNAs and negative strand RNAs reliably in samples with non-strand specific PCR Ct's below 30
105 irrespective of the time between swab collection and symptom onset. Detection of subgenomic RNA
106 or negative strand RNA alone may therefore not be a reliable indicator of current SARS-CoV-2
107 replication, but the presence of relatively more genomic and positive strand RNAs than subgenomic
108 and negative strand RNAs may be a better marker of current/recent virus replication.
109

110 **Results**

111 Quantitation of SARS-CoV-2 genomic and 7a and N subgenomic RNAs using probe
112 based real-time PCR assays

113 We had previously identified by amplicon-based sequencing and partly supported by non-probe based
114 PCR, that subgenomic N and ORF7a RNA molecules were frequently the most abundant subgenomic
115 RNAs present in naso-oropharyngeal swabs from infected individuals[11], and therefore here designed
116 probe based PCR assays to quantitate the levels of these two RNAs in samples. We also developed a
117 probe based 5' UTR PCR and a total ORF7a PCR to quantitate SARS-CoV-2 genomic RNA and
118 collective genomic RNA and subgenomic RNAs for S, ORF3a, E, M, ORF6 and ORF7a respectively,
119 to allow for estimation of the ratio of genomic to subgenomic RNAs within samples. Using the
120 developed PCRs as single target, duplex (5'UTR and subgenomic ORF7a) and triplex (5'UTR and
121 subgenomic ORF7a and N gene) probe based PCR assays, we could detect SARS-CoV-2 genomic
122 RNA molecules with the 5'UTR assays in a positive SARS-CoV-2 cell culture preparation used as
123 positive control/comparison, and notably in twenty four out of twenty five PCR positive diagnostic
124 naso-oropharyngeal samples included in this study (Table 1, Figure 1, Figure 2, and Supplementary
125 Figure 1, Tables 2, 3 and 4). The only exception was the swab sample from individual 7 collected on
126 day 47 after onset of symptoms (GC-58) which had a borderline Ct of 38 in the original diagnostic
127 PCRs. The same individual was only weak positive when tested 1 month earlier (GC-21) with a Ct
128 of 33 (Table 1). Subgenomic ORF7a and N gene RNA could be detected in samples up to 19 days
129 after the onset of symptoms and in all samples with a 5' UTR PCR below 30. Above a 5'UTR PCR
130 Ct of 30, detection of either subgenomic RNA targets became less consistent as the subgenomic RNA
131 assays likely approached their limits of detection. Despite this limitation, we detected both targets in
132 16 swab samples with both the single target and triplex assays and only swab sample GC-365 from
133 individual 16, collected 17 days post the onset of symptoms was negative for both ORF7a and N
134 subgenomic RNA in both assays (Table 2). This sample was already near the limit of detection for
135 the genomic length SARS-CoV-2 RNA with a 5' UTR PCR Ct of 38.1. The negative swab and pool
136 were consistently negative in all PCR assays.
137 In the SARS-CoV-2 positive samples, the single target, duplex and triplex PCR assays agreed closely
138 with each other, with the targets shared between these assays having similar Ct's for each of the

139 samples and indicating that it is reasonable to multiplex these assays to use less sample in future
140 testing. The duplex 5'UTR PCR was on average 2.1 Ct's (Range: -4.4 to -1.2) lower than the single
141 target 5'UTR PCR for the same sample, but on average only 0.5 Ct's (Range: -4.2 to 0.1 Ct's) lower
142 than the triplex assay. A very similar pattern was seen in the ORF 7a subgenomic assays, with the
143 duplex PCR having 2.1 Cts (Range: 0.4 to 3.2 Ct's) lower than the single target ORF 7a subgenomic
144 PCR on average for each sample, and only 0.5 Ct's lower (Range: -0.4 to 1.1 Ct's) than the triplex
145 assay. The single target subgenomic N gene PCR was repeatable, with 0.3 Ct's difference on average
146 (range: -0.6 to 0.9 Ct's) between the repeated testing of each sample. It also agreed strongly with the
147 triplex N gene subgenomic assay results and was only 0.2 Ct's lower on average compared to the
148 triplex assay for each sample (range: -0.85 to 0.9 Ct's)). Each of the PCRs within the duplex and
149 triplex assays shared the same common forward primer, however this did not appear to significantly
150 reduce the sensitivity of these assays, and conducting multiple PCRs in one reaction may have slightly
151 improved the sensitivity of the 5' UTR genomic and ORF 7a subgenomic PCRs. These multiplex
152 assays had the benefit that less cDNA was required for testing from limited diagnostic samples, and
153 the triplex PCR had the advantage of detecting an extra subgenomic target for essentially the same
154 level of sensitivity as the duplex PCR. The triplex assay was repeated twice, with an average
155 difference of 0.4 (range: 0 to 1.9), 0.2 (range: 0 to 1.6) and 0.2 Ct's (range: 0 to 0.8) observed for the
156 same sample between the 5'UTR, subgenomic ORF7a and subgenomic N gene PCRs respectively.
157 The ORF7a total single PCR, detecting genomic and subgenomic spike, ORF3a, E, M, ORF6 and
158 ORF7a subgenomic RNAs, also closely followed the single target 5'UTR PCR Ct value in samples
159 with a Ct less than 30, typically producing a Ct 1.4 cycles lower (IQR: 0.7 to 1.6) than the 5'UTR
160 assay (Figure 1). We did expect that the ORF7a total assay should have a Ct slightly lower than the
161 5' UTR assay given that it can amplify both SARS-CoV-2 genomic RNA and several subgenomic
162 RNA targets, but were not expecting a full Ct difference or more between these assays. Given recent
163 reports on the structure of the SARS-CoV-2 5'UTR[16] and the challenges we faced using the 5'UTR
164 assay as a strand-specific PCR (see below), we believe the 5' UTR PCR may actually not be as
165 sensitive as we would have liked, and may slightly under report the amount of genomic SARS-CoV-
166 2 RNA within the samples accounting for the majority of the difference between the 5'UTR and
167 ORF7a total assays.
168 The PCR Ct's of the subgenomic ORF7a and N gene targets generally followed the Ct of the genomic
169 5' UTR PCR by 1-5 Ct's, and steadily increased as the 5'UTR PCR Ct increased (Figures 1 and 2
170 and Supplementary Figure 1) suggesting that the amount of SARS-CoV-2 sgRNA within a sample
171 was directly related to the amount of gRNA. Above a Ct of 30 for the genomic 5'UTR PCR, the
172 detection of subgenomic N gene and ORF7a RNAs became less consistent in each of the single target,
173 duplex and triplex assays. These samples were all repeat samples taken from individuals at 10 or more

174 days after the initial onset of symptoms and the quantity of PCR targets may likely have been
175 approaching the limit of detection for the single PCR assays making quantitation unreliable. We
176 noticed that the swab sample from individual 2 (GC-26) was reporting a higher Ct in this study
177 compared to our previous SARS-CoV-2 work but the bioanalyzer analysis (see below) indicated that
178 there may have been some degradation of the RNA during storage between the two studies[11].
179 Nevertheless, we could still detect the genomic and subgenomic SARS-CoV-2 targets in this sample
180 with the probe-based assays.
181

182 ## Ratio of SARS-CoV-2 genomic RNA to subgenomic ORF7a and N gene RNAs

183 To calculate the ratio of the genomic to subgenomic SARS-CoV-2 RNA in the naso-oropharyngeal
184 swab samples and in the included cell cultured sample, we plotted the difference in Ct between the
185 subgenomic ORF7a or N gene and the 5' UTR PCR for the single target, duplex and triplex PCRs
186 assays for the 16 naso-oropharyngeal swabs in which we detected all subgenomic RNA targets as
187 well as the cell culture sample (Figure 3, Figure 4 and Supplementary Figure 2). The delta Ct's and
188 ratios of genomic to subgenomic RNA for ORF7a or N gene tended to more or less fluctuate around
189 their median values, particularly in swab samples collected >24 hours or later after the onset of
190 symptoms. (Figures 3 and 4 and Supplementary Figure 2). The differences or ratios were sometimes
191 higher in some of the swab samples collected within the first 24 hours after the onset of symptoms
192 (for instance GC-238 (individual 14 day 0), GC-199 (individual 19 day 1), GC-277 (individual 13
193 day 0)), but this was not a consistent observation for all of the swabs collected within this time frame.
194 The median single target ORF7a subgenomic PCR Ct was 3.7 (IQR: 3.0 to 4.4 Ct's) higher than the
195 5' UTR assay for the single target PCRs, and 3.4 Ct's (IQR: 3.1 to 3.7) and 3.2 Ct's (IQR: 3.0 to 3.7)
196 for the duplex and triplex assays respectively. With a calculated PCR efficiency of 95%, this was
197 equivalent to a genomic RNA to ORF7a subgenomic RNA ratio of 10.8 (IQR: 6.9 to 16.9), 8.9 (IQR:
198 7.3 to 10.7) and 7.6 (IQR: 6.9 to 10.8) for the single, duplex and triplex PCRs respectively. Overall
199 the general pattern in the PCRs agreed, and indicated that the abundance of SARS-CoV-2 genomic
200 molecules in naso-oropharyngeal was 7 to 11 times higher than the ORF7a subgenomic RNA. For
201 the 48 hr cell culture sample, the ratio estimated by the PCRs was similar, with 6-12 times more
202 genomic RNA to 7a subgenomic RNA.
203 The median subgenomic N gene single target PCR was 0.75 Ct's higher than the single target 5'UTR
204 PCR (IQR: 0 to 1.8) and the triplex N gene subgenomic PCR 2.4 Ct's higher than the triplex 5'UTR
205 PCR (IQR: 1.8 to 2.7). This corresponded to a median of 1.6 (IQR: 1 to 3.2) times more genomic
206 RNA than subgenomic N gene RNA for the single target PCRS and a median of 5 (IQR: 3.2 to 5.7)
207 times more genomic RNA for the triplex assay. The ratio of genomic SARS-CoV-2 RNA in the naso-
208 oropharyngeal swabs was therefore between 1-5.7 times higher than subgenomic N gene RNA, and

209 5.3 to 5.6 times higher in the 48 hr cell culture. The ratio of the genomic RNA to N gene subgenomic
210 was lower than the corresponding ORF7a subgenomic ratio in almost all swab samples and the cell
211 culture indicating that the subgenomic N gene is usually the more abundant of these two subgenomic
212 RNAs in naso-oropharyngeal swabs.
213 Overall, within the 16 naso-oropharyngeal swab samples, the subgenomic ORF7a and N gene RNA's
214 were always detectable, and based on the above ratios, the N and 7a subgenomic RNAs together could
215 make up on average 25% (IQR: 20.5 to 31.5%) of the total SARS-CoV-2 RNA within the samples.
216 Except for the three swab samples collected at symptom onset mentioned above, the ratios of SARS-
217 CoV-2 genomic to subgenomic RNA appeared to fluctuate within the range of 7-10 fold less ORF7a
218 subgenomic RNA and 2-5 fold less subgenomic N gene RNA relative to the amount of genomic
219 SARS-CoV-2 RNA present within a sample (Figure 3 and Figure 4). However, given that we suspect
220 that the 5' UTR assay may slightly underestimate the total amount of genomic RNA within a sample
221 by about 1 Ct based on the fact that the 7a total PCR was consistently around 1.4 Ct below the 5'
222 UTR PCR, then these ratios may be out by a factor of 2 and therefore the true ratio may be closer to
223 14-20 fold less ORF7a and 4-10 times less subgenomic N RNA compared to genomic RNA.
224

225 Detection of positive and negative strand SARS-CoV-2 RNA in diagnostic samples

226 We next wanted to study whether we could accurately quantify the ratio of positive to negative strand
227 SARS-CoV-2 RNA molecules in clinical samples, as detection of negative strand RNA has been
228 proposed as a surrogate measure of replicating virus[13]. To do this, we used the specific probe based
229 PCR assays mentioned above, but instead of making cDNA with random hexamers, we generated
230 strand specific cDNA with the specific forward or reverse sense primers for the negative and positive
231 strand assays respectively. During our assay development, we found that the reverse 5'UTR PCR
232 primers designed to specifically detect SARS-CoV-2 genomic RNA, while relative efficient PCR
233 primers (see above and Materials and Methods), performed very poorly as cDNA synthesis primers
234 from positive strand templates. This could be explained by the recently reported secondary structure
235 of the 5'UTR of the SARS-CoV-2 genomic RNA which revealed that the location where the reverse
236 5'UTR primer annealed lay within a highly stable RNA hairpin structure (stem loop 5a), which was
237 resistant to RNase I treatment even under denaturing conditions[16]. As a result, this assay was not
238 appropriate to study the ratios of positive to negative sense RNA molecules in the clinical samples
239 most likely due to poor reverse primer annealing and subsequent cDNA synthesis. The ORF7a total
240 RNA, ORF7a subgenomic and N subgenomic RNA PCRs however, all performed well and after the
241 initial optimisation performed on dilutions of the RNA from the SARS-CoV-2 positive cell culture,
242 all had efficiencies of close to 95%. From the non-strand specific PCR results above, we knew that
243 the ORF7a total RNA assay Ct's were close to those of the 5'UTR PCR, and therefore the majority

244 of targets detected by this assay were likely SARS-CoV-2 genomic RNA molecules as opposed to
245 the subgenomic RNAs. Therefore, this PCR could be used to approximate the ratio of positive to
246 negative sense copies of the SARS-CoV-2 genomic RNA. From our previous work, we knew that the
247 strand specific PCRs were less sensitive than their non-strand specific counterparts[11], therefore for
248 this experiment, we selected fourteen of the swab samples with the lowest 5' UTR Ct's and in which
249 both the subgenomic ORF7a and N gene non-strand specific PCRs were positive, in addition to the
250 cell culture sample and used the RNA of each in the strand specific PCRs.
251 The strand-specific ORF7a total PCR assay was able to detect positive sense SARS-CoV-2 genomic
252 RNA (ORF7a total) in all the fourteen naso-oropharyngeal swabs tested as well as the cell culture
253 sample (Table 5 and Figure 5). Similarly, positive sense N gene subgenomic RNA was detected in
254 all samples, but positive sense ORF7a subgenomic RNA detected in only eight swab samples and the
255 cell culture sample. Negative sense genomic 7a total and N gene subgenomic RNA was detected in
256 eleven swab samples, while negative sense 7a subgenomic RNA was detected in 9 swabs (Figure 5).
257 When sufficient RNA remained, the strand specific PCRs were repeatable, although some samples
258 with very high Ct's >35 for some of the targets did test negative in one of the repeats suggesting that
259 some of these targets were near the limit of detection of the assays.
260 While most of the fourteen samples were taken early in the clinical course of the SARS-CoV-2
261 infection, four samples, GC-26, GC-292, GC-291 and GC-310 taken on days 7,10,13 and 17 post
262 symptom onset were also within this set, and we were able to detect two or more of the negative sense
263 targets in each of these samples. In the samples with a non-strand specific ORF7a total PCR Ct of 20
264 or less, all negative and positive sense targets could be detected, but in samples with a CT > 20,
265 detection of one or more strand specific targets became inconsistent, and only one of those samples
266 (GC-26 taken on day 7) was PCR positive for all of the positive and negative strand targets. Despite
267 this, generally there was a trend in the diagnostic swab samples that as the Ct for the total amount of
268 virus RNA in the sample increased (as indicated by the Ct of the non-strand specific total ORF7a Ct
269 in Figure 5), the strand specific PCR Ct's, if positive, also increased.
270 In only six of the fourteen swab samples were we able to detect all positive and negative strand RNA
271 targets below a Ct of 35. These six samples included three taken on the day of symptom onset, and
272 three taken on 1, 2 or 10 days after the onset of symptoms. Using the strand specific PCR Ct's, we
273 calculated the relative proportion of each RNA molecule relative to the positive sense ORF7a total
274 positive assay in these six samples (Supplementary Figures 3 and 4 and Source Data File). The
275 overwhelming majority of SARS-CoV-2 RNA in the swab samples was positive sense. In the three
276 swabs taken on the day of symptom onset the amount of positive sense SARS-CoV-2 RNA was 63
277 to 264 fold higher than the negative sense RNA, while in the three swabs collected later in the clinical

278 course, the relative amount of positive sense RNA decreased to only 32 to 41 fold higher than the
279 negative sense RNA (Supplementary Figure 3).
280 The most abundant positive sense RNA was detected by the ORF7a total PCR (Figure 5), which was
281 likely mostly positive sense genomic length SARS-CoV-2 RNA. The range of relative ratios of
282 positive to negative sense RNA for each of the subgenomic targets and ORF7a total (see Source Data
283 File) were higher in the three swabs taken at the onset of symptoms (ORF7a total: 119.3-283.8 fold,
284 subgenomic ORF7a: 53.5-719.5 fold, subgenomic N gene: 10.8-164.5 fold) compared to the swab
285 samples collected later (ORF7a total: 14-44 fold, subgenomic ORF7a: 34.1-89.4, subgenomic N: 17-
286 38.8 fold).
287 While we only detected all strand specific targets in six samples, we did however detect positive and
288 negative strand ORF 7a total in eleven swab samples and were therefore able to calculate the ORF7a
289 total positive to negative strand ratios in these samples. We again saw the pattern of swabs from the
290 day of symptom onset having 100 fold or more positive sense gRNA, while the majority of the swabs
291 collected later had a slightly lower ratio of positive to negative sense gRNA (ratio of 30-40)
292 (Supplementary Figure 4). This however was not a hard and fast rule with two later samples (GC-26
293 and GC-291) also having relatively high proportions of positive to negative strand gRNA (ratio above
294 100). GC-291 was from individual 14 taken 13 days post symptom onset, and the diagnostic SARS-
295 CoV-2 PCR Ct of this sample was lower than their first swab collected at day of symptom onset (GC-
296 238). It was therefore likely that some virus replication was still occurring in this individual to account
297 for this increase in SARS-CoV-2 RNA. Presumably during replication more virions and cytoplasmic
298 positive sense RNA, including both full length genomic and subgenomic RNAs, used for translation
299 of SARS-CoV-2 proteins would be present, and this would increase the positive to negative strand
300 RNA ratio. After replication had finished, these virions would be excreted and cytoplasmic RNAs
301 degraded and the relative proportion of positive strand RNA would fall until only the positive and
302 negative strand SARS-CoV-2 RNA in the replication complexes remained[12].
303 All positive and negative sense RNAs were detected in the cell culture sample, but the ratios of
304 positive to negative strand RNA were even higher in the cell culture sample compared to any of the
305 swab samples for all three targets. This is likely due to the fact that although the cell culture at 48 hrs
306 represents a late stage of infection with full lysis of cells and release of virions into the medium, the
307 material had been clarified by centrifugation possibly pelleting and thus removing most cellular
308 membrane material including double membrane vesicles, which we believe contain most of the
309 negative sense RNA. In addition, the lysed cells may still have contained fragments of positive sense
310 virus RNAs used for translation. In contrast, for the swab samples at the time of onset of symptoms,
311 that is up to 14 days post initial exposure[17], any virions produced in vivo would be continually cleared
312 by normal muco-ciliary clearance mechanisms and later on by the immune response. The cell culture

313 material was also gamma irradiated prior to transportation to our laboratory which would be more
314 likely to damage double stranded RNA compared to single stranded RNA[18]. Given that any negative
315 strand would likely be complexed with the more highly abundant complementary forward strand, this
316 damage could disproportionately affect the negative sense RNA possibly also contributing to the
317 observed higher ratio of positive to negative sense RNA.
318

319 ## Detection and abundance of NGS reads mapped to subgenomic RNAs

320 To further explore the pattern of subgenomic RNAs in the naso-oropharyngeal swab samples, we
321 created two Ampliseq primer mini-panels using 11 reverse primers within each of the ten potential
322 SARS-CoV-2 subgenomic RNAs and in the 5' UTR region together with two different forward
323 primers within the 5' leader sequence of all the SARS-CoV-2 RNAs. These primers were a subset of
324 the Ampliseq primers from the ThermoFisher Ion AmpliSeq SARS-CoV-2 Research Panel which we
325 had previously used to detect subgenomic RNAs[11]. The rationale for doing so was to essentially create
326 a multiplex PCR to quantitate the relative abundance of genomic and subgenomic RNAs within each
327 sample. We found that despite these two mini-panels differing only in their forward primer, the
328 amount of amplification detectable in the same SARS-CoV-2 positive swab samples differed greatly
329 with an average of 901,000 reads generated with mini-panel 2 after 31 cycles and an average of 21,000
330 reads generated by mini-panel 1 after 31 cycles of PCR amplification (see Source Data). Given that
331 the forward primer was the only difference between the two mini-panels, this primer was the likely
332 reason for the lower amplification in mini-panel 1. We subsequently identified during the course of
333 full genome sequencing (see below), that there was a mismatch between this primer and the SARS-
334 CoV-2 genome sequence of several, perhaps all, of the samples we had used in this study, and this
335 likely resulted in this primer annealing poorly during the Ampliseq PCR across the 5'UTR and
336 subgenomic amplicons. As a result, we found that we needed to use 31 cycles of PCR amplification
337 with this panel to obtain an average of 459 reads mapped to all of the amplicon targets (see Source
338 Data).
339 With 31 PCR cycles, mini-panel 1 estimated that approximately 38.5% of the SARS-CoV-2 RNA
340 within the samples was subgenomic RNA (range: 5-52%). All canonical subgenomic RNAs except
341 for ORF7b and ORF10 were detected by the panel. Subgenomic N gene was the most abundant
342 subgenomic molecule, followed by ORF7a (Figure 6). These two molecules were also the most
343 abundant in the cell culture sample (See Source Data). Together, these two molecules made up
344 between 50-95% of all the subgenomic RNA within the samples. The other subgenomic RNAs were
345 present, but at lower abundances. A similar pattern was also seen at 21 PCR cycles although with
346 relative lesser amplification of the lower abundant subgenomic targets by that stage of the PCR
347 reaction (Supplementary Figure 5). Making cDNA with the panel primers resulted on average in a

348 similar number (316 vs 339) of reads as the random hexamer cDNA (See Source Data) and again,
349 subgenomic ORF7a and N were the most abundant amplicons.
350 In mini-panel 2, the forward primer was much more efficient than the mini-panel 1 forward primer,
351 and we saw much more amplification across all targets. The 5' UTR amplicon was particularly
352 efficient, and even by 21 cycles of amplification, the number of reads mapped to this amplicon was
353 significantly higher than all other amplicons. By 31 cycles we could see that the relative number of
354 reads mapped to the subgenomic RNAs was reduced due to this single amplicon crowding them out
355 on the sequencer chip. Despite this, mini-panel 2 identified a similar pattern in the relative abundance
356 of the subgenomic RNA's in both the 21 and 31 cycles of amplification, with the N gene and ORF7a
357 subgenomic RNAs being the two most abundant subgenomic RNAs across the swab samples (making
358 up between 82.3-97.5% of all subgenomic reads at 21 cycles, and 88.9-100% at 31 cycles), and the
359 other canonical subgenomic RNAs, except ORF7b and ORF 10 that were not detected at all, being
360 expressed at much lower levels (Supplementary Figure 6 and Supplementary Figure 7). These two
361 subgenomic RNAs for N and ORF7a were also the most abundant in the cell culture sample (See
362 Source Data).
363 Mini-panel 2 was also able to detect a low number of reads belonging to potential non-canonical
364 subgenomic RNAs which could initiate expression of the proteins of ORF 9b[19] and also ORF 7b. The
365 ORF9b subgenomic RNA contained the 69 nucleotides of the leader joined to nucleotide 28284,
366 which is 24 nucleotides downstream of the regular N gene TRS (based on the nucleotide numbering
367 of SARS-CoV-2 Wuhan-Hu-1 NC_045512.2) similar to what has been described[19]. The non-
368 canonical ORF 7b subgenomic RNA, here designated subgenomic RNA 7b*, contained the 69
369 nucleotides of the leader sequenced joined to nucleotide position 27674, 26 nucleotides downstream
370 of the canonical 7b TRS site. The ability to detect such reads appeared to be related to the number of
371 reads detected mapping to the canonical subgenomic RNA upstream to these non-canonical sites (ie
372 subgenomic N gene for ORF 9b and subgenomic 7a for ORF 7b), and therefore these reads were seen
373 more frequently when we used 31 cycles of PCR amplification. The number of subgenomic ORF 9b
374 reads made up only approximately 0.03-0.11% of the total subgenomic N gene reads and the number
375 of subgenomic ORF 7b was equal to only 0.13-0.25% of the number of subgenomic ORF 7a reads
376 (See Source Data). Interestingly, there was some minor heterogeneity to these non-canonical
377 subgenomic RNAs in the occasional read in some samples such as GC-14 (using 21 PCR cycles, see
378 Source Data) which showed the leader joined to sequence upstream of the 7b AUG at position 27602
379 and also the 9b subgenomic RNA had some minor heterogeneity, for example for sample GC-251
380 (using the full panel (see below) and 21 cycles, see source data). Close inspection of reads generated
381 in our previous study[11] also supported this minor heterogeneity, for example GC-277and GC-26
382 which showed the leader joined to sequences at position 27678 for the 7b* subgenomic RNA and

383 some heterogeneity around the 9b site, respectively (data not shown, raw data available in SRA
384 Accession PRJNA636225).
385

386 ## SARS-CoV-2 Sequencing and subgenomic analysis using the full Ampliseq panel

387 We had previously analysed the genome sequence of the SARS-CoV-2 viruses in samples from
388 individuals 2 to 8 which were from the first wave of COVID-19 infections in Victoria in March-May
389 2020[11], but had not sequenced the later samples from the second wave between July and September
390 2020 included here. The samples from the ten individuals from this period (individuals 11 to 21) were
391 all community acquired infections and likely picking up the infection from multiple sources according
392 to the information provided by the local COVID-19 outbreak contract tracing team. Using the SARS-
393 CoV-2 Ampliseq Panel, we obtained the near complete SARS-CoV-2 genomic sequence (nucleotide
394 positions 42 to 29,842) from the swab samples collected from individual 17 day 1 and from four other
395 individuals who were linked to the same case cluster by the contact tracing team - individuals 12 day
396 0, individual 14 day 13, individual 20 day 10 and individual 21 day 0 (GenBank accessions:
397 MZ410617-MZ410621). We also obtained partial SARS-CoV-2 sequence from two cases linked to
398 this cluster - individual 13 day 0 (GenBank accession: MZ410622) and individual 14 day 0 (NCBI
399 SRA: SRR14836407). The sequences from individual 17, 12, 14 and 21 were identical to each other,
400 although individual 17 was not linked to the same case cluster as individuals 12, 14 and 21. These
401 four sequences were identical to a sequence reported from Victoria from early August 2020 (SARS-
402 CoV-2/human/AUS/VIC17057/2020 GenBank accession: MW321043). The partial sequences from
403 individual 13 was also identical to MW321043 over the genome regions we obtained. All the
404 individuals from the case cluster were tested for SARS-CoV-2 in the first part of August whereas
405 individual 17 may have picked up this virus 2-3 weeks later further along the community transmission
406 chain.
407 The sequence from individual 20, who also belonged to the case cluster, differed in two nucleotides
408 from MW321043 and the SARS-CoV-2 sequences obtained from the other individuals in the cluster.
409 One change at nucleotide 22993 (based on the nucleotide numbering of SARS-CoV-2 Wuhan-Hu-1
410 NC_045512.2) was a change from a C to T without altering the amino acid sequence of the spike
411 protein. The second change, G to A, was at nucleotide 28884 located within the N protein (amino
412 acid 203) and changed an arginine to a glutamine in the nucleoprotein sequence. This would have
413 also resulted in a lysine to glutamic acid in the ORF9c protein (amino acid 50)[20], and altered the non-
414 canonical TRS of the proposed ORF N*[19] which was present in the other SARS-CoV-2 sequences
415 presented here. Searching for sequences from the second wave in Victoria, Australia with BLASTN
416 on GISAID and GenBank's nucleotide database, we found that there was only one virus sequenced
417 in the local epidemic with this change in the nucleoprotein during the initial epidemic molecular

418 tracing (SARS-CoV-2/human/AUS/VIC8533/2020, GenBank accession: MW153442) collected in
419 early August 2020. The characteristics for MW153442 matched those of our individual 20, and
420 subsequent follow up of laboratory records revealed that individual 20 was in fact tested for SARS-
421 CoV-2 in early August, and the swab sample sent to the state reference laboratory for sequencing.
422 Given that no other virus sequences with this change were identified in the local epidemic, this variant
423 may have only arisen in a single individual during the outbreak in this cluster, but was not able or did
424 not have the opportunity to transmit and propagate significantly in the general population afterwards.
425 A MegaBLAST search in NCBI's Nucleotide database revealed that the same amino acid change in
426 the nucleoprotein has been observed in other countries indicating that the SARS-CoV-2 virus has
427 spontaneously made this change multiple times during the global epidemic.
428 Interestingly, in the SARS-CoV-2 sequencing from the swab samples GC-251 and GC-243 from
429 individuals 12 day 0 and 21 day 0, the forward primer of the first 5'UTR amplicon of pool 1 of the
430 Ampliseq panel did not work well, and this resulted in the sequencing of reads created from linear
431 amplification from the reverse primers, and thus from positive sense cDNA produced from randomly
432 primed negative sense SARS-CoV-2 RNAs, which extended beyond where the forward primer would
433 sit. This revealed a change of a C to a T at nucleotide position 40 within the leader sequence, which
434 is within the 3' end of where the forward primer would anneal. This change is most likely the reason
435 the forward primer was unable to anneal and amplify in these two samples. Given that all the other
436 SARS-CoV-2 sequences we generated in this study, except for that of individual 20, were identical
437 to the SARS-CoV-2 genomes from individuals 12 and 21, it is entirely possible that those SARS-
438 CoV-2 genomes also had the same C to T change at position 40, but a forward primer did manage to
439 eventually anneal during the amplification PCR of these samples resulting in the expected amplicon.
440 If this nucleotide change was present in the samples, particularly if strong RNA secondary structures
441 are also present in this region[16], then that could explain why the mini-panel 1, which used this same
442 forward primer, amplified the 5'UTR and subgenomic RNAs inefficiently compared to mini-panel 2.
443 For mini-panel 2, the end of the primer was at nucleotide position 52, and therefore this particular
444 nucleotide substitution would have sat within the middle of the primer which is less critical to primer
445 annealing and extension when compared to the 3' end. Similarly, there was also a nucleotide change
446 of a C to a T at position 241 in all seven SARS-CoV-2 genomes sequenced in this study, and this also
447 sat within the middle of the annealing site of the reverse 5'UTR primer of the mini-panels and did
448 not affect the primer's function as the 5' UTR was the most efficient amplicon of mini-panel 2.
449 By looking for reads with the 5' UTR leader sequence joined to 3' ORFs at known TRS sequences,
450 we could identify between 161 to 142999 reads coming from subgenomic RNA in the full Ampliseq
451 panels (See Source Data). Consistent with the mini-panels, reads mapping to subgenomic ORF7a and
452 N gene RNAs were the most abundant, accounting for 37.3 to 61.2% of all subgenomic reads

453 (Supplementary Figure 8). The full panel was also reasonably good at amplifying ORF3a and ORF6
454 subgenomic RNAs, and therefore it estimated the relative abundance of these two RNAs higher than
455 either of the two mini-panels. The remaining subgenomic RNAs were also detected (except for
456 ORF7b and ORF10) but made up a much smaller fraction of the total subgenomic reads (11-23% of
457 subgenomic reads). Like mini-panel 2, the full panel was also able to detect a low number of some
458 non-canonical ORF 9b subgenomic RNAs (0-0.65% of mapped subgenomic reads). Sample GC-251
459 contained a TRS joining variant where the leader, nucleotides 1-69, was joined to nucleotide 28278,
460 6nt earlier than ORF 9b subgenomic reads seen with the mini-panel 2 in the samples GC-277, GC-
461 316, GC-291 and GC-292. No non-canonical subgenomic ORF 7b* were detected by the full panel,
462 however subgenomic RNAs which would initiate ORF N*[19] were detected by the full panel (0.27-
463 1.83% of mapped subgenomic reads) (See Source Data). These reads joined the leader nucleotides 1-
464 69 to the nucleotide 28882 at a TRS sequence which was created by a triple nucleotide polymorphism
465 GGG→AAC at nucleotides 28881-2883 which arose in the B.1.1 lineage early in 2020 [21,22].
466 Both samples GC-243 and GC-251 reported much fewer subgenomic reads compared to the other
467 swab samples consistent with the nucleotide change in the annealing site of the 5'UTR forward primer
468 sequence resulting in poor primer binding and amplification. In contrast, Sample GC-277 also had
469 reduced 5'UTR amplification using the full panel compared to other samples but good amplification
470 of many of the subgenomic targets. Why this occurred is unclear, but perhaps the 5'UTR of the
471 genomic RNA renatured due to its secondary structure during cDNA synthesis in this particular
472 sample.
473

474 Exploring the quality of cellular RNA and mRNA in diagnostic naso-oropharyngeal 475 swabs

476 As we were looking specifically at RNA intermediates generated during SARS-CoV-2 replication in
477 the oro-nasal mucosa, we wanted to understand whether we could measure how well a swab sample
478 had been collected and transported by looking at host cellular RNA. We decided to quantify the total
479 RNA and the RNA Integrity Number (RIN) of the extracted nucleic acids as well as the relative levels
480 of expression of cellular mRNAs for selected inflammatory genes as surrogate measures of how well
481 a swab had been collected (a well collected swab should have collected pharyngeal and nasal
482 epithelial cells and hence more cellular RNA in addition to any SARS-CoV-2 RNA present) and
483 stored prior to arriving at the laboratory. Total RNA could potentially indicate how many host
484 epithelial cells had been sampled during the swabbing process, and the RIN could similarly indicate
485 how well cellular RNA had been preserved and how metabolically active the epithelial cells were at
486 the time of sampling. The virus infection itself could also have an impact on these measures, as
487 infected cells could have reduced cellular mRNA activity as SARS-CoV-2 infection has been shown

488 to suppress cellular mRNA splicing and translation[23] and result in apoptosis of human epithelial
489 cells[24], and therefore we also looked for correlations between the amount of virus detected in the
490 samples and the level of cellular RNA as we had seen a negative relationship in our previous study[11].
491 We first measured total RNA in the swab samples using RNA 6000 Pico chips on a bioanalyzer and
492 assessed the appearance of ribosomal peaks in the extracted RNA. In the 30 swab samples for which
493 we had sufficient RNA to perform the bioanalyzer analysis, we found that quantity of extracted
494 nucleic acid was highly variable (40-4816pg/µl see Source Data File), and clear peaks corresponding
495 to ribosomal RNA were only visible in 16 of the 30 swab samples. As a result, the reported RNA
496 integrity (RIN) scores varied between 1 and 8.4, with nine samples not reporting a RIN and only six
497 with a RIN of 3 or greater (Source Data File). This indicated that there was some variability between
498 swab samples in the amount of epithelial cells and cellular material collected. The quality and quantity
499 of ribosomal RNA was also highly variable and may have either been poorly preserved prior to
500 arriving at our laboratory or reduced as a result of the SARS-CoV-2 infection. The bioanalyzer also
501 identified that there had been some degradation of the extracted RNA from the swab sample from
502 individual 2 which we had extracted as part of our previous study[11] and used again here. This may
503 have explained the increase in this sample's SARS-CoV-2 PCR Ct's seen in this investigation as
504 compared to our previous study[11].
505 We then examined whether there was a relationship between the total RNA in a swab and the PCR
506 Ct's across all of the SARS-CoV-2 positive samples. Interestingly, for the twenty two samples where
507 we had both SARS-CoV-2 PCR and bioanalyzer data, there was a consistent weak trend of increasing
508 Ct across all non-strand specific SARS-CoV-2 PCR as the overall quantity of RNA increased
509 (Kendall Rank Correlations: 0.25 to 0.4) (See Source Data and Supplementary Figure 9), indicating
510 that more virus might be correlated with less cellular RNA in a swab sample presumably due to the
511 negative effect of infection on regular cellular function[23] and consistent with our previous smaller
512 study[11]. The strand specific PCR Ct's showed a similar but weaker trend, with higher Ct's seen in
513 samples with more total RNA (Kendall Rank correlation: 0.16 to 0.43) (See Source Data and
514 Supplementary Figure 10).
515 We grouped samples by RIN score (RIN<3 (or no RIN) = low, 3≤ RIN ≤6 = medium, RIN > 6 =
516 high) and looked for any relationships between the PCR Ct's and RIN score. Eighteen of the twenty
517 two SARS-CoV-2 positive swabs with bioanalyzer and PCR data belonged to the low group, three to
518 the medium and only one to the high RIN group (See Source Data). For the non-strand specific PCRs,
519 the ranges of Ct's within the three groups overlapped, with the largest range of Ct's seen in the low
520 RIN group (Supplementary Figure 11). Similarly for the strand specific PCRs, the ranges of Ct's of
521 the three groups mostly overlapped, albeit with fewer data points in some assays as not all targets
522 were detected in all samples (Supplementary Figure 12). No consistent relationship could be observed

523 between the RIN and the SARS-CoV-2 PCR Ct's suggesting that the amount of SARS-CoV-2 PCR
524 RNA (Ct) was independent of how intact the ribosomal RNA (RIN) was at the time of testing.
525

526 ## Analysis of the host response in swab samples

527 To look more closely at cellular mRNA in the swab samples, we used an Ampliseq panel designed to
528 specifically amplify 683 targets from the mRNA of known human inflammatory genes and compared
529 the relative expression of these genes across the samples. SARS-CoV-2 is known to provoke innate
530 immune responses in epithelial cells[25] and is known to have general immune suppressive effects in
531 patients[26], and therefore studying selected inflammatory genes could indicate the transcriptional
532 activity of the upper respiratory epithelium and whether the cells present were responding to the virus.
533 From previous work, we knew that the RNA extraction method we had employed here was able to
534 extract both DNA and RNA from samples, and therefore the first step of this analysis involved
535 running three selected no cDNA controls (two SARS-CoV-2 positive swabs (GC-11 and GC-21) and
536 the negative pool 1) to identify potential gene targets that could amplify from genomic DNA as well
537 as from host mRNA. As most amplicons in the Ampliseq panel were designed to sit across an intron,
538 theoretically only reads amplifying the cDNA of mRNA would amplify from both forward and
539 reverse primers, and any reads generated from a single primer binding to genomic DNA, would
540 include exon or intragenic sequence at one of the ends of the read sequence. As the read mapping
541 TMAP software would penalize any read with a large unmapped end when we mapped to the human
542 transcriptome, we were able to establish a stringent read mapping requirement of a minimum
543 alignment length of 74 nucleotides (the length of the shortest amplicon), and a MAPQ of 90, to filter
544 out the reads generated from DNA targets (these were seen in IGV as mapped reads with large soft
545 clipped ends and a MAPQ score of ~75 or lower depending on the amplicon). However, these settings
546 alone could not filter all reads coming from DNA as the panel included 22 amplicons which did not
547 span an intron, and therefore reads mapped to these amplicons could come from genomic DNA or
548 mRNA. These 22 amplicons were therefore excluded from this analysis (see source data file).
549 We identified an additional 80 amplicons in which 10 or more reads were mapped in one or more of
550 the no cDNA control samples, even at the stringent mapping requirements described above. We
551 inspected these amplicons in IGV and found that for many of these 80 targets, the majority of the
552 amplicon sat within a single exon of a gene, with essentially only one primer sitting in another exon.
553 Consequently, reads coming from the genomic DNA and which had their intron sequences trimmed
554 by the IonS5XL server for base calling quality reasons, were completely mapped within the amplicon
555 region and therefore received a high MAPQ score from the TMAP software. We therefore removed
556 these 80 amplicons from the analysis (see source data file). As a result, the inflammatory gene analysis
557 was restricted to 581 inflammatory gene targets.

558 To avoid identification of spurious differences in genes between the groups arising simply from very
559 low numbers of reads being mapped, we excluded any amplicons with less than an average of 10
560 reads across all the samples. With the remaining 469 inflammatory gene targets, the number of reads
561 mapped to each gene was ranked highest to lowest and we then performed hierarchical cluster analysis
562 to see if we could observe any difference in the pattern of inflammatory gene expression in the SARS-
563 CoV-2 positive and negative swab samples. We then calculated the average number of reads mapped
564 to each inflammatory target across all the swab samples, and then for each sample classified how
565 many gene targets were above or below the average number of reads mapped of all samples to
566 quantify overall cellular expression. The hierarchical cluster analysis identified one broad cluster of
567 samples (Cluster 1), which included the SARS-CoV-2 negative samples, except for the swab from
568 negative individual 1 (Figure 7, Source Data File and Supplementary Data 1). Included in this cluster
569 were also many of the samples collected later in the clinical course of infection (4+ days post
570 symptom onset) with low levels of SARS-CoV-2 RNA (Ct's>29), and only one swab in one
571 individual (individual 11 day 0) from the day of the onset of symptoms. Individual 11 was the
572 individual with the highest SARS-CoV-2 PCR Ct (Ct 28) from all swabs collected on the day of
573 symptom onset, was never hospitalized during their infection, and was therefore a mild infection
574 possibly with only a low level of virus replication occurring in their nasal mucosa. Samples within
575 cluster 1 had an average of 231 genes (Range:175-236) targets with above average numbers of reads
576 mapped and an average of 237 genes (Range: 203-294) with below average number of reads mapped,
577 while outside of cluster 1, the samples had on average of only 130 genes (Range: 54-218) with an
578 above average mapped reads and 338 (Range: 251-415) with below average mapped reads. This
579 would indicate that a broader profile of inflammatory genes transcription in the cells of the mucosa
580 within the samples of cluster 1 compared to the other swab samples. Given that the majority of the
581 SARS-CoV-2 negative samples sat within cluster 1, the SARS-CoV-2 positive samples within this
582 cluster were only weakly positive (Ct>28), we believe this cluster most likely represented samples
583 from individuals with a more diverse mucosa immune profile, whereas the samples outside cluster 1
584 were exhibiting some suppressed immune gene transcription possibly as a result of SARS-CoV-2
585 infection[23].
586 The hierarchical clustering also identified a second cluster of samples, cluster 2, with the lowest
587 number of inflammatory genes with above the average number of reads mapped out of all the samples
588 in this study (Cluster 2, Figure 7). The swab samples in cluster 2 were all from elderly individuals,
589 several of which were reported to have other significant health problems at the time of infection. The
590 swab samples GC-366 from individual 16, taken 14 days post onset of symptoms fell within this
591 cluster 2, and only had 54 genes expressed above the average of all the swab samples, the lowest of
592 all the samples in this study. This sample however also had a high SARS-CoV-2 Ct, low RIN value

593 and multiple bands on the bioanalyzer suggesting that the RNA in this sample was degraded. It was
594 one of the samples that travelled furthest to reach our laboratory and the additional time to reach our
595 laboratory may have resulted in degradation of the RNA. The follow up swab taken from individual
596 16 three days later had much higher levels of gene expression and was grouped within cluster 1 further
597 suggesting that the day 14 sample may have been a poor swab sample. The swabs from individual 18
598 on day 17 and individual 12 day 10 appeared to have poor/degraded nucleic acids based on their
599 bioanalyzer results, however the numbers of genes with above expression was almost twice that of
600 individual 16, day 14 and more akin to many of the other swab samples outside of cluster 1. These
601 two swabs also had low SARS-CoV-2 PCR Ct's (23 and 17 respectively) indicating that the swab
602 collector did take a reasonable sample of the surface of the mucosa. The remaining samples in this
603 cluster 2 either had a low SARS-CoV-2 Ct, good ribosomal bands and a high RIN, and therefore we
604 believe these truly are samples with less inflammatory gene expression within the mucosal cells
605 possibly due to the individuals' age and the immune suppressive ability of SARS-CoV-2[26] and its
606 ability suppress cellular mRNA splicing and translation[23].
607 By plotting the number of genes with above average expression in the SARS-CoV-2 positive samples
608 versus the days post symptom onset (Figure 8), we could see that the expression pattern in the samples
609 collected on the day of symptom onset was highly variable, but the majority of swab samples had
610 fewer genes with above average number of reads mapped compared to the SARS-CoV-2 negative
611 swabs (130 genes vs the negative's 239 genes). Due to diagnostic testing regimes, we had very few
612 samples available between 1 and 10 days post symptom onset available, but the one we did have had
613 a high number of genes with above average numbers of reads mapped. Samples collected around day
614 10 post symptom onset often had relatively fewer genes with above average number of reads mapped,
615 but samples collected after this time and before day 17 post symptom onset had a steadily increasing
616 number of inflammatory genes with above average number of reads mapped, suggesting that as time
617 passed, the overall gene expression in the epithelium increased presumably as new cells were
618 produced and the mucosa returned to a state of homeostasis.
619 Due to the differences we observed in the ratios of strand specific and subgenomic RNAs in some
620 swab samples collected close to the onset of clinical symptoms, we wanted to see if the host response
621 could be used as an indicator of the stage of the SARS-CoV-2 infection in the upper respiratory
622 epithelium in the individuals. We classified the swab samples as SARS-CoV-2 negative, early clinical
623 (≤1 days post symptom onset) or later clinical (>1 days post symptom onset) and looked for
624 differential counts in the inflammatory gene targets. We first plotted the differential read counts of
625 the early and late SARS-CoV-2 positive groups against the median of the negative samples to look
626 for genes with increased or decreased relative read counts (Supplementary Data 2 and 3). The early
627 samples contained 247 genes with interquartile ranges of read counts not including 0 (equal to the

628 median of the negative samples) whereas the later samples only had 148 such genes suggesting that
629 the pattern of expression was different in the three groups and the pattern in the early samples was
630 more different when compared to the negative samples. We performed a Kruskal-Wallis analysis to
631 look for genes with significant differences in reads counts between the two SARS-CoV-2 positive
632 groups and the negative samples (see Source Data) and a Wilcoxon Rank-Sum test to compare the
633 read counts of the samples in each SARS-CoV-2 positive group to the negative samples. The
634 Benjamini-Hochberg method using a false discovery rate (FDR) of 25% was applied to examine only
635 those inflammatory genes with the largest differential expression between the three groups. We found
636 no genes with significant differences in the pattern of relative read counts between all three groups,
637 and no genes were identified as having significantly different distributions of read counts between
638 the groups of samples collected later in SARS-CoV-2 positive individuals and the negative individual
639 group (See Source Data). However, 77 genes were identified as having significantly different patterns
640 of reads per million reads mapped between the early SARS-CoV-2 positive and the negative samples
641 (See Supplementary Figure 13 and see Source Data). Sixty-six of these genes were relatively lower
642 in read counts in the early SARS-CoV-2 positive samples and only eleven were relatively higher
643 compared to the negative samples.
644 Of the eleven genes up regulated in the early group, six were associated with pathways involved in
645 the detection of virus RNA and replication and the interferon response. These were DExD/H-Box
646 helicase (DDX58)[27,28], interferon induced with helicase C Domain containing protein 1 (IFIH1)[29],
647 N-Myc interactor (NMI)[30], interferon regulatory factor 7 (IRF7)[30], interleukin 1 receptor agonist
648 (IL1RN) and myeloid differentiation primary response 88 (MYD88)[31]; four were associated with
649 attraction and regulation of phagocyte migration (CX3CL1)[32], formyl peptide receptor 1 (FPR1)[33],
650 complement C3a receptor (C3AR1)[34], secreted and transmembrane protein 1 (SECTM1)[35], and one
651 were associated with apoptosis, tumour necrosis factor ligand superfamily member 10 (TNFSF10)[36].
652 Seven of these genes (DDX58, IFIH1, TNFSF10, NMI, IRF7, C3AR1 and FPR1) showed median
653 elevations of 2.3-6.4-fold relative to the negative samples (Figure 9 and Supplementary Data 1).
654 The normalized number of reads to both DDX58 and IFIH1 were only elevated in swabs collected on
655 or shortly after the onset of symptoms, and both fell to levels below those seen in the negative samples
656 in the swabs collected between 10 to 14 days post symptom onset (Figure 9). The elevation of these
657 innate immune genes indicated that by the time the individuals in this study were showing symptoms,
658 the innate immune defences in the upper respiratory epithelium was already responding to the virus
659 infection. Interestingly, individuals 12, 13 and 6 on the day of symptom onset which all showed
660 higher ratios of positive to negative sense SARS-CoV-2 RNA, were some of the samples with the
661 highest relative number of reads mapped to IFIH1, and elevated numbers of reads mapped to DDX58.
662 IFIH1 has been identified as the primary method of sensing SARS-CoV-2 replication within

663 respiratory epithelial cells[29], while DDX58 is understood to be able to bind to single stranded positive
664 sense viral RNA and block the SARS-CoV-2 RNA dependent RNA polymerase from copying its
665 genome[27] and thus significantly reducing viral replication. The elevation of these two genes along
666 with more positive sense SARS-CoV-2 RNA would be consistent with current/recent virus replication
667 in the upper respiratory epithelium at the time of sampling in these individuals.
668 Reads mapped to NMI were similarly elevated in the early group further indicating that the epithelial
669 cells were responding and regulating the innate immune response at the time these swabs were taken.
670 NMI is a regulator of the innate immune response and has been demonstrated to be part of the system
671 regulating the expression of DDX58[37]. NMI Read counts fell below levels seen in the negative
672 samples in about half of the swab samples collected between days 10 to 14 post symptom onset. One
673 late sample, GC-291 (individual 14, day 13) which had a higher SARS-CoV-2 RNA load in this late
674 swab compared to the swab collected on the day of symptom onset from the same individual, had
675 elevated levels of NMI but this individual may have had a delayed innate immune response as the
676 swab taken on the day of symptom onset also had the lowest levels of IFIH1 of all the early swabs.
677 The included SARS-CoV-2 negative swab from negative individual 1, who was possibly fighting
678 another infection at the time of sampling given their reported fever and upper respiratory symptoms,
679 also had elevated numbers of reads mapped to the NMI amplicon above that seen in the other negative
680 samples.
681 The number of reads mapped to IRF7 were higher in most of the early clinical swab samples, but the
682 expression levels did not drop off in the swabs collected later during the course of infection as sharply
683 as DDX58 or IFIH1 mentioned above. IRF7 is an important transcription activator for the production
684 of interferon in the innate immune response to some RNA viruses but interestingly, IRF3 and IRF5
685 have been shown to be more important that IRF7 in regulating the IFIH1 induced interferon response
686 in SARS-CoV-2[29] yet IRF7 was on average more elevated relative to IRF3 in the early clinical swabs
687 compared to the late clinical/negative swabs in this study (IRF5 was not part of the Ampliseq
688 inflammatory panel used here).
689 TNFSF10 expression was higher in most of the early clinical swabs and one of the later swabs
690 collected on day 10 post symptom onset (GC-292, Individual 20). As SARS-CoV-2 has been shown
691 to induce apoptosis in cell culture after the SARS-CoV-2 replication[24], we believe this marker could
692 be indicative of the infected cells in the epithelium undergoing programmed cell death post virus
693 replication. Relative to the SARS-CoV-2 negative samples, reads mapped to both C3AR1 and FPR1
694 were elevated in most early samples and about 50% of the later SARS-CoV-2 positive samples,
695 including some taken 16 to 17 days post symptom onset (Figure 9). These two genes have been shown
696 to attract and regulate neutrophil migration from the bone marrow into blood and into tissue such as
697 the respiratory epithelium[33,34]. Given that C3AR1, FPR1 and TNFSF10 were elevated in the swab

698 samples taken up to 17 days after initial clinical signs, it could indicate that the process of apoptosis,
699 phagocyte invasion and epithelial repair resulting in the removal of previously infected cells and their
700 remnants may take at least 2 to 3 weeks to complete after the initial onset of clinical symptoms. If
701 this is the case, it might explain why membrane protected SARS-CoV-2 RNA could persist in some
702 people resulting in prolonged PCR positivity.
703

704 **Discussion**

705 In this study we have developed and used new sensitive probe based PCR assays and Ampliseq panels
706 to further explore the pattern of SARS-CoV-2 subgenomic RNA and strand specific RNA in
707 diagnostic swab samples, and coupled this with measures of the total cellular RNA and mRNA to
708 understand the activity of the host cells during SARS-CoV-2 infection. We have shown that SARS-
709 CoV-2 subgenomic RNA can be detected up to 19 days post initial symptom onset, and negative
710 sense RNA can be detected as late as 17 days after initial symptoms were observed. The period of
711 detection of subgenomic RNA is consistent with what we and other investigators have seen previously
712 (14 to 25 days)[8,11,14]. Our detection of negative strand RNA out to 17 days post symptom onset is also
713 consistent with the length of time other studies have detected negative strand RNA in non-
714 immunocompromised people[13].
715 The quantity of SARS-CoV-2 subgenomic RNA detected by PCR was related to the amount of total
716 SARS-CoV-2 RNA present in the samples detected by the genomic RNA and diagnostic PCRs. We
717 could consistently detect subgenomic RNA in samples with a genomic PCR Ct of 30 or less, but
718 above this level the detection of these RNA molecules became inconsistent. The majority of SARS-
719 CoV-2 RNA within a swab sample was genomic-length RNA and this was on average between 7 to
720 22 times more abundant than subgenomic ORF 7a RNA, and about 2 to 10 times more abundant than
721 subgenomic N gene RNA in swab samples. Based on the probe-based PCRs, we estimate that on
722 average, the ORF7a and N gene subgenomic RNA molecules could make up to around 12.5-25% of
723 the total SARS-CoV-2 RNA collected by a nasopharyngeal swab sample. These two subgenomic
724 RNA molecules likely represent a large fraction of the subgenomic RNA present within a typical
725 naso-pharyngeal sample, as both of the Ampliseq subgenomic mini-panels and the genomic
726 sequencing SARS-CoV-2 Ampliseq panel all consistently found that these two subgenomic
727 molecules made up the largest percentages of all the subgenomic reads. Based on these findings, we
728 believe that assays detecting either the ORF7a or N gene subgenomic RNAs would be the most
729 sensitive way of detecting SARS-CoV-2 subgenomic RNA, and that these two targets, simply because
730 of higher abundance and not as a measure of current/active transcription, should be at detectable
731 levels for longer than perhaps some of the other subgenomic RNA molecules in diagnostic samples.

732 We observed a higher relative proportion of both SARS-CoV-2 genomic RNA and positive sense
733 RNAs in some of the swabs collected at or very close to the onset of symptoms (≤24 hours) compared
734 to the relative proportions seen in swabs collected later in the clinical course of SARS-CoV-2
735 infection. These higher ratios were not seen in all swabs collected at symptom onset, but in the few
736 samples where it was observed, we believe these samples may have been taken during late SARS-
737 CoV-2 replication in the upper respiratory epithelium, or very shortly after replication had ceased.
738 During virus replication, much more positive sense RNA would be present within infected cells as
739 the replication transcription complexes (RTC) associated with convoluted membrane structures and
740 double membrane vesicles[12,38] export positive sense SARS-CoV-2 genomic RNAs into the cytoplasm
741 of an infected cell to be packaged into virions for export out of the infected cells. Similarly the RTC's
742 would export positive sense genomic-length and subgenomic RNAs into the cytoplasm to act as
743 mRNA to be translated into virus proteins at cellular ribosomes[38]. However, both these fractions of
744 SARS-CoV-2 RNA would likely be transient or labile *in vivo,* as packaged virions are exported
745 extracellularly and removed by the muco-ciliary clearance system of the nasopharyngeal epithelium,
746 while cytoplasmic mRNA are broken down by cellular RNases. Therefore, these RNAs would only
747 be present in high relative quantities during replication, or very shortly after replication has ceased,
748 and this would be observed as an increased proportion of genomic (positive sense) RNA and positive
749 strand mRNAs within a sample if taken within this window. After replication had finished, eventually
750 the virions would be cleared and cytoplasmic mRNA broken down, and only the SARS-CoV-2 RNA
751 protected within the membrane structures, which we have previously shown to be protected[11], would
752 remain in those cells/cell remnants once infected by the SARS-CoV-2 virus. RNA within these
753 structures is likely both subgenomic and genomic length RNA with both types of molecules existing
754 as a mixture of positive sense and negative sense RNA. Positive sense RNA is still the predominant
755 form of SARS-CoV-2 RNA in these membrane protected structures, with about 30-40 fold more
756 positive sense RNA to negative sense RNA based on the PCR ratios in the later samples presented
757 here. Individuals could therefore remain PCR positive until cells containing these membrane
758 structures eventually fall-off the epithelium and are replaced by new non-infected cells or are cleared
759 by immune cells. However, if the post replicative SARS-CoV-2 RNA is mostly within these
760 membrane structures, then its recognition by cytosolic pattern-recognition receptors could be
761 impaired, slowing the immune clean-up of infected cells[12,39] leaving the cells containing these
762 structures to be cleared by the natural turnover of the respiratory epithelium. The human respiratory
763 epithelium is estimated to turn over once every 30-50 days under normal conditions[40], but in the face
764 of epithelial cell loss, re-epithelialization is initiated quickly and can be completed in a matter of
765 days[41]. Given that we observed that the relative inflammatory gene expression levels only started to
766 approach the levels of the SARS-CoV-2 negative samples in swabs collected 14 days or later after

767 the onset of clinical symptoms, and there was an increase in reads mapped to genes such as TNFSF10,
768 which induces apoptosis and C3AR1 and FPR1 which signal and attract phagocytic cells as late as
769 17 days post symptom onset, this could suggest that the epithelium is still under repair two or more
770 weeks post infection, which may result in the delayed clearance and replacement of apoptotic or
771 senescent cells with these membrane protected SARS-CoV-2 RNAs still present within them. These
772 cells may therefore be present for two to three weeks or even longer and this could explain why some
773 individuals even in the absence of obvious immunosuppression can be PCR positive for a protracted
774 period[42,43]. Prolonged clearance of epithelial cells containing SARS-CoV-2 RNA has been observed
775 previously in intestinal biopsies taken as long as 92 to 173 days post symptom onset[44]. Given the
776 intestinal epithelium normally turns over relatively rapidly, every 2 to 6 days[45], the persistence of
777 virus replication complexes may not be unexpected in the nasal epithelium with a much lower rate of
778 turn over.
779 Our exploration of the inflammatory genes indicated that by the time the majority of individuals were
780 exhibiting COVID-19 clinical signs and presented for testing, innate antiviral immune genes of cells
781 within the epithelium were already upregulated and responding to the virus infection. IFIH1 and
782 DDX58 were only upregulated in swabs collected at or close to symptom onset, and these genes have
783 previously been shown to specifically respond to presence of virus RNA in the cellular cytoplasm
784 and produce proteins to reduce the amount of SARS-CoV-2 replication and trigger a type I interferon
785 response[27,29]. The fact that these genes were not upregulated in the swab samples collected later in
786 the clinical course of COVID-19 or the SARS-CoV-2 negative swabs further suggests that active
787 replication had mostly ceased in these later swabs. These findings are in line with the PCR findings
788 above, where we believe SARS-CoV-2 replication is being shut down before, or around the time of,
789 initial symptom onset in most people resulting in the observed decrease of the proportion of genomic
790 and positive to negative sense RNA in the samples.
791 Based on these observations, we again believe that the detection of the presence of SARS-CoV-2
792 subgenomic RNA or negative sense RNA should not be considered in themselves as measures of
793 current active virus replication in the upper respiratory epithelium due to their persistence in
794 intracellular membrane structures which we were able to detect for almost three weeks after clinical
795 symptom onset. Instead, the reality is far more nuanced, with the relative proportion of these
796 molecules in upper respiratory swab samples changing from relatively high proportions of genomic
797 and positive sense RNA in swab samples where we think replication was occurring or had just
798 finished, tapering towards the proportions seen in later samples in which mainly the SARS-CoV-2
799 RNA protected within the membrane structures remain. It may therefore be possible to use the assays
800 developed here or other subgenomic or strand specific RNA assays to measure the relative ratios of
801 these molecules in swab samples, and conclude where there is a higher proportion of positive sense

genomic RNA, that there is an increased risk that the individual may be shedding virions and therefore is more likely to be infectious at the time of sampling.

Methods

Samples

Samples included in this study were selected from combined nasopharyngeal and oropharyngeal swabs collected as part of ongoing public health surveillance for SARS-CoV-2 in Victoria, Australia from January to August 2020. SARS-CoV-2 positive and negative swabs were identified by PCR at two diagnostic laboratories, and the remaining swab and media were transported to the Geelong Centre for Emerging Infectious Diseases (GCEID) laboratory. A subset of 27 positive swabs from 17 individuals, including 7 from our previous study[11], were selected for this study (Table 1). Individuals were selected on the basis of 1) they had repeated swabs taken on multiple days, and/or 2) their initial swab had a Ct below 30. One sample from individual 4 (GC-11) used in our previous study was exhausted part way through this study, and so was only used for the inflammatory panel analysis described below. The swabs came from individuals in the community, and most were not known to be linked to each other except for individuals 11, 12, 13, 14, 20 and 21 which were identified by the local contact tracing team as belonging to a known larger cluster of cases. Forty-one SARS-CoV-2 negative combined nasopharyngeal and oropharyngeal swabs were selected from samples identified as negative by PCR during the same period. Forty of these samples were pooled into four pools of 10 swabs, while one (individual 1 from our previous study[11]) was used as an individual negative PCR control, and a second negative individual (22) was used as a negative control in the inflammatory gene study only.

Basic patient details including age group, gender, epidemic wave during which swab was collected, swab collection in days post symptom onset, summary clinical signs and hospitalization or not were collected from laboratory submission forms or medical records. Summary details of the samples included are shown in Table 1. The study complied with all relevant ethical regulations and has been approved by the Barwon Health Human Research Ethics Committee (Ref HREC 20/56) and all participants gave their informed consent.

A SARS-CoV-2 positive cell culture supernatant was included as a positive control and for comparison to our diagnostic swab samples. This sample consisted of the supernatant from a 48 hour third passage of SARS-CoV-2 virus in Vero E6 cells and kindly supplied by the Australian Centre for Disease Preparedness. The cultured virus isolate is designated VIC-01-059 and isolated from a SARS-CoV-2 positive individual in Victoria, Australia in 2020. The culture supernatant was clarified at 4000g for 10 min before being gamma irradiated and transferred to the GCEID laboratory.

837 ## Nucleic acid extraction and cDNA synthesis

838 Nucleic acid extraction was performed on 50 µl of the collection media from the combined
839 nasopharyngeal/oropharyngeal swab samples using the MagMax 96 Viral RNA Isolation kit
840 (Thermofisher) on a Kingfisher Flex extraction robot (Thermofisher). cDNA synthesis was performed
841 by heating extracted nucleic acids at 65°C for 5 minutes and rapid cooling on ice before cDNA
842 synthesis using SuperScript™ VILO™ Master Mix (Thermofisher Scientific, Victoria, Australia) as
843 per manufacturers' instructions [46,47] and described previously[11]. The nucleic acids from the cell
844 culture positive control sample was extracted using the Qiagen Viral RNA mini kit (Qiagen, Victoria,
845 Australia) as per the manufacturer's instructions.
846

847 ## Development of SARS-CoV-2 genomic and subgenomic probe based real-time PCR
848 assays

849 We had previously developed and reported SYBR/Syto 9 based PCR assays to detect genomic and
850 subgenomic SARS-CoV-2 RNA in clinical samples[11]. These assays, however, had limitations in their
851 sensitivity and needed to be run individually as they all used a single fluorescent reporter SYBR/Syto
852 9 in the real-time PCR reaction. To improve the sensitivity of these assays, we designed four probe-
853 based PCR assays, three of which were similar to our previous SYBR/Syto 9 based assays, and each
854 amplified either the 5' UTR genomic RNA, the ORF7a subgenomic RNA only, or the total (genomic
855 and subgenomic) ORF7a RNA, respectively, and a fourth PCR targeting the N-subgenomic RNA
856 only. Primer and probe sequences were designed based on the sequence of SARS-CoV-2 Wuhan-Hu-
857 1 (NC_045512.2) and are provided in Supplementary Table 1. We tested several forward primers
858 within the leader sequence and found the optimum primer to be one which annealed in the leader
859 sequence of the SARS-CoV-2 genome between nucleotide positions 45 and 66, while the probe and
860 reverse primer annealed in the 5' UTR downstream of the leader sequence which is only present in
861 the genomic SARS-CoV-2 RNA. The subgenomic ORF7a and subgenomic N RNA specific assays
862 utilized the same forward SARS-CoV-2 leader primer as the 5' UTR assay mentioned above, while
863 their probes and reverse primers annealed within their respective genes. They would therefore only
864 detect the subgenomic RNAs where the virus had joined the leader immediately upstream of these
865 genes. The ORF7a total PCR utilized the same probe and reverse primer as the ORF7a subgenomic
866 PCR but used a different forward primer annealing within the ORF7a gene. This PCR would therefore
867 detect genomic RNA as well as several subgenomic RNA molecules, specifically those used to
868 transcribe the Spike, ORF3a, Envelope, Membrane, ORF6 and ORF7a subgenomic RNAs. The three
869 probes each had a different fluorophore attached, with the aim to be able to subsequently combine
870 assays.

871 Each of these assays were initially established as single/individual PCRs. For these single assays, 2ul
872 of prepared cDNA from the samples were added to a PCR mastermix comprising 1x Brilliant
873 Multiplex qPCR mastermix (Agilent), 1µM of both the forward and reverse primer, 0.2 µM probe
874 and nuclease free water to a total reaction volume of 10 µl. The PCR reaction was then performed in
875 a Quantstudio 6 real-time PCR machine (Thermofisher). Known SARS-CoV-2 positive and negative
876 swab samples and the SARS-CoV-2 cell culture sample were used to initially optimize the PCR
877 temperature cycling conditions, with the optimum cycling temperatures determined to be 95°C for 10
878 min, then 40 cycles of 95°C for 3 sec, 58°C for 30 sec and 64°C for 30 sec.
879 After the successful amplification from known positive clinical samples, the resulting PCR product
880 was visualized on a 2% agarose gel (2% Size Select E-gel, Thermofisher) to confirm a product of the
881 expected size was produced. The amplicon was then extracted from the gel and sequenced using the
882 3.1 Big Dye Terminator PCR sequencing reaction and sequenced on a Hitachi 3500 genetic sequencer
883 (Thermofisher) to confirm that the amplicons were indeed the expected genomic and subgenomic
884 targets. The gel purified amplicons were quantitated on a QiaXpert spectrophotometer (Qiagen,
885 Victoria, Australia) and used as standards in serial dilution in the subsequent qPCR reactions, and in
886 determining the efficiency of the PCR reactions. The PCR efficiency, estimated to be 95% for these
887 assays, along with the difference in Ct values between the genomic 5'UTR and the subgenomic
888 ORF7a and N assays was used to calculate the ratio of genomic to subgenomic RNA within the
889 samples.
890 To determine if we could multiplex the individual assays and develop a single test which could both
891 quantitate genomic and subgenomic RNA in a SARS-CoV-2 positive sample, we initially combined
892 the 5' UTR and the ORF7a subgenomic assays into a duplex real-time PCR. This was performed by
893 combining 1x Brilliant Multiplex qPCR mastermix (Agilent), 0.9µM of the shared SARS-CoV-2
894 leader forward primer, 0.9 µM of each of the 5'UTR and ORF7a reverse primers, 0.18 µM of each of
895 the 5'UTR and ORF7a probes, 2ul of sample/control cDNA and nuclease free water to a final volume
896 of 11ul. This PCR reaction was run under the same cycling temperatures described above.
897 Once we determined that the 5'UTR genomic and ORF7a subgenomic assays could be successfully
898 duplexed, we then attempted to create a triplex assay by incorporating the N gene subgenomic assay
899 into the duplex. This would allow us to efficiently quantitate two subgenomic RNA molecules
900 alongside the genomic RNA. Different primer concentrations of the common leader forward primer
901 from 0.71, 1.43 and 2.86 µM were evaluated as part of the optimization of this assay. The final PCR
902 reaction for the triplex assay was setup by adding 1x Brilliant Multiplex qPCR mastermix (Agilent),
903 2.86µM of the shared SARS-CoV2 leader forward primer, 0.71 µM of each of the 5'UTR and ORF7a
904 and N-gene reverse primers, 0.14 µM of each of the 5'UTR, ORF7a and N-gene probes, 2ul of

905 sample/control cDNA and nuclease free water to a final volume of 14ul. The triplex assay was run in
906 duplicate under the same temperature conditions as described above.
907

908 ## Real time assays to quantitate positive and negative sense SARS-CoV-2 RNA.

909 During replication of SARS-CoV-2 in cells, the virus produces negative sense RNA to act as a
910 template for the production of full length copies of the virus genome and the various subgenomic
911 RNAs. We had previously detected negative strand RNA in three of eight diagnostic swab samples[11],
912 however the quantity of negative strand SARS-CoV-2 RNA was 150 and 20 fold lower than positive
913 strand genomic and subgenomic RNA respectively[11], and therefore we could only detect it in samples
914 with a very high virus load (low Ct). To create potentially more sensitive assays to detect strand
915 specific RNA in more SARS-CoV-2 samples, we attempted to adapt the probe-based PCRs described
916 above to single step, strand-specific PCR reactions combining sense specific primer reverse
917 transcription followed by PCR.
918 We initially used the SARS-CoV-2 48 hr cell culture, for which we had ample RNA available, for
919 the development and optimization of the strand based assays. For each strand specific PCR, 2µl of
920 the RNA from a sample was denatured at 95°C for 3 mins and then rapidly cooled on ice. The
921 denatured RNA was then mixed with 1x Brilliant II qRT-PCR 1-Step QRT-Master Mix (Agilent),
922 1µl of 10µM primer (the PCR forward primer for negative sense cDNA synthesis and the PCR reverse
923 primer for positive sense cDNA synthesis), 0.5ul of RT/RNase block enzyme mixture (Agilent) and
924 nuclease free water to a final volume of 8.8ul. This was then incubated at 50°C for 30 min followed
925 by 95°C for 8 min. At this point, 1µl of 10uM complementary primer and 0.2µl of the 10µM stock of
926 the corresponding probe were added to the reaction to bring the reaction volume to 10µl. This was
927 then incubated at 95°C for 2 min, then 40 cycles of 95°C for 3 sec, 58°C for 30 sec and 64°C for 30
928 sec in a Quantstudio 6 real-time PCR thermocycler. Serial dilutions of the gel purified amplicons
929 were used to determine the efficiency slope of each PCR reaction. During testing with the positive
930 and negative controls, it was determined that the reverse primer of the 5'UTR PCR, while a relatively
931 efficient PCR primer, was inefficient at synthesising first strand cDNA. Attempts to improve the
932 cDNA synthesis, including trialling different reaction temperatures, adding primer at the denaturation
933 step, using different RT enzymes or slightly different reverse primers for the cDNA step, were not
934 successful in improving the performance of the positive sense-specific 5'-UTR PCR. The positive
935 sense 5'UTR assay was therefore unreliable/not sensitive enough to quantitate the amount of positive
936 sense genomic RNA present within the samples. As the 5'UTR primers were part of the triplex and
937 duplex assays, we could not adapt these assays to detect positive/negative strand RNA. Instead, we
938 selected and used the sense-specific PCRs for the 7a total instead of the 5'UTR together with the 7a
939 subgenomic and N subgenomic as three separate strand specific assays. These three assays performed

940 as expected with efficiencies of around 95% similar to the non-strand specific versions of these
941 assays.
942 Once we were confident that we had developed an efficient set of strand specific assays, we focussed
943 on the samples which likely had enough virus RNA load to allow detection of strand specific RNA
944 targets. Fourteen SARS-CoV-2 positive swabs, the 48 hour SARS-CoV-2 cell culture supernatant,
945 and the negative individual and pooled negative swab sample 1 were tested in the strand specific
946 assays described above. Some of the clinical samples had little RNA remaining at this point in the
947 investigation, so all RNAs were diluted 1:4 in low TE prior to cDNA synthesis except for the cell
948 culture, which was not diluted and the sample from individual 3 (GC-13), which was diluted 1:6.
949 Due to the poor performance of the 5' UTR positive sense PCR, ratios of the amount of strand specific
950 subgenomic RNAs were estimated by the delta-Ct between the 7a/N subgenomic assays and the 7a-
951 total assay, which was used as a proxy for the amount of genomic RNA present in the sample.
952

953 Development of Ampliseq Mini-Panels to specifically detect SARS-CoV-2
954 subgenomic RNAs

955 We previously reported on the ability of the commercially available SARS-CoV-2 Ampliseq Panel
956 from Thermofisher Scientific [https://www.thermofisher.com/au/en/home/life-
957 science/sequencing/dna-sequencing/microbial-sequencing/microbial-identification-ion-torrent-next-
958 generation-sequencing/viral-typing/coronavirus-research.html] to detect subgenomic RNAs [11] as part
959 of genomic sequencing. This occurred as a result of the two forward primers within the leader
960 sequence (the first two primers of each of the two commercial Ampliseq primer pools) of the SARS-
961 CoV-2 genome in addition to producing genomic amplicons in the 5'-UTR from reverse primers
962 included for that, also are producing subgenomic RNA specific amplicons with the reverse primers
963 downstream of the Transcription Regulatory Sites (TRS) sequence in each of the different SARS-
964 CoV-2 genes[11]. The method was sensitive and detected subgenomic RNAs in all samples tested
965 despite these amplicons competing with the 242 SARS-CoV-2 and host gene specific amplicons
966 within the Ampliseq reaction. To determine if the sensitivity, and possibly also the ability to compare
967 levels of individual amplicons/subgenomic RNAs, of this method could be improved, we
968 designed/selected two new Ampliseq mini-panels containing only the primers required to amplify the
969 subgenomic RNA molecules, i.e. a forward sense primer sitting within the leader sequence and a
970 reverse primer from each of the potential 10 SARS-CoV-2 subgenomic RNAs downstream of the
971 predicted TRS sequences. A single reverse primer within the 5'UTR and a primer pair targeting the
972 host cellular TATA box binding protein (TBP) mRNA were also included in each of the two Ampliseq
973 panels to hopefully provide an estimate of the amount of SARS-CoV-2 genomic RNA and host
974 cellular mRNA within the sample, respectively. The full list of primers in each Ampliseq mini-panel

975 can be found in Supplementary Table 2 and were obtained from Thermofisher. In all, one SARS-
976 CoV-2 leader forward primer, 1 genomic and 10 SARS-CoV-2 subgenomic specific reverse primers
977 and the forward and reverse TBP primers were included in each mini-panel. The two mini-panels we
978 tested differed only in their forward leader primers. Mini-panel 1 included the first primer from the
979 commercial SARS-CoV-2 Ampliseq primer pool 1, which terminated at nucleotide 42 of SARS-CoV-
980 2 Wuhan-Hu-1 (NC_045512.2). Mini-panel 2 included the first primer from the second SARS-CoV-
981 2 Ampliseq primer pool which ended at nucleotide position 52 of NC_045512.2. The Ampliseq
982 primers for these mini-panels were selected by us from the full SARS-CoV-2 Ampliseq Panel
983 (Thermofisher Scientific) and obtained from the Ampliseq custom service team at Thermofisher
984 Scientific.
985 Ampliseq reactions were performed using the random hexamer VILO cDNA prepared above, and the
986 Ion Ampliseq Plus Library Kit (Thermofisher Scientific, Victoria Australia) as per manufacturer's
987 instructions initially with three modifications. The first was the addition of 2µM of SYTO 9
988 (Thermofisher Scientific, Victoria, Australia) to each reaction to enable the PCR amplification to be
989 monitored in real time. The second was to initially increase the PCR cycles from 21 to 31 cycles
990 which corresponded with most samples entering the exponential phase of the PCR amplification. The
991 third was to decrease the individual primer concentration to 71nM (Total primer concentration 1µM).
992 In a separate set of reactions, we used selected samples made into cDNA by either random hexamers
993 (SuperScript™ VILO™ Master Mix, Thermofisher Scientific) or cDNA made with the Ampliseq
994 mini-panel 1 primers using superscript IV Reverse transcriptase (Thermofisher Scientific) and 21
995 Ampliseq PCR reaction cycles. Sample libraries were quantitated using the Ion Library Taqman
996 quantitation kit, pooled and loaded onto Ion Torrent 530 chips using an Ion Chef templating robot
997 and sequenced using an Ion Torrent S5XL. Seven samples were also sequenced with the full SARS-
998 CoV-2 Ampliseq Panel as per the manufacturer's instructions using 21 cycles of PCR amplification
999 and the reads assembled into SARS-CoV-2 consensus sequences using Ion Torrent plugins as
1000 described previously[11] and analysed for the presence of subgenomic reads described briefly below.
1001 Generated sequences were then mapped to a custom SARS-CoV-2 subgenomic RNA reference which
1002 contained the sequence of the ORF1ab and the start of each subgenomic RNA molecule including the
1003 65/69 nucleotides of the leader sequence, the TRS and the 5' end of each of the genes of the SARS-
1004 CoV2 virus including up to where each reverse primer was predicted to anneal within that gene's
1005 sequence[11]. Mapping of the reads was performed by TMAP software included in the Ion torrent
1006 Server software suite 5.10.1[48]. Counts of reads mapping to each subgenomic, genomic and host target
1007 were obtained by using the Coverage Analysis plugin (v5.12.0.0) on the Ion Torrent Server, using a
1008 minimum mapping quality score of 20 and minimum alignment length of 20. Read mapping counts
1009 were checked by visualizing the reads and coverage using Integrative Genome Viewer (IGV) 2.6.3[49].

1010

1011 ## Measuring the nucleic acid quantity and quality in the nasopharyngeal swab samples

1012 To investigate the host RNA and host inflammatory response detectable in the nasopharyngeal
1013 mucosa to SARS-CoV-2 infection, as represented by the swab samples included in our study, we first
1014 measured the quantity and quality of the extracted nucleic acid using Agilent RNA 6000 Pico Chips
1015 on an Agilent 2100 bioanalyzer (Agilent Technologies, California, USA). We then performed a
1016 Kendall Rank correlation analysis on the total nucleic acid quantity (pg/µl) and the RNA Integrity
1017 Number (RIN) and the Ct's of the non-strand specific single target and triplex PCR assays, and the
1018 strand specific assays.

1019

1020 ## Exploring the cellular gene expression in the naso-oropharyngeal swab samples

1021 We then used the Ion Ampliseq RNA Inflammation Response research panel (Thermofisher
1022 Scientific, Victoria, Australia) to look at the host inflammatory response in more detail. This designed
1023 primer panel amplified targets within 683 human genes associated with immune/inflammatory
1024 responses [https://www.thermofisher.com/au/en/home/life-science/sequencing/rna-
1025 sequencing/targeted-rna-sequencing-ion-torrent-next-generation-sequencing.html]. All swabs from
1026 individuals from which two or more swabs were collected on different days, were included in this
1027 analysis. Two SARS-CoV-2 negative individuals and four pools of SARS-CoV-2 negative
1028 individuals and a no template control (MagMax 96 Viral RNA elution buffer) were also included in
1029 this analysis. Ampliseq reactions were performed using the same random hexamer VILO cDNA
1030 preparations as above, and the Ion Ampliseq Library Kit 2.0 (Thermofisher Scientific, Victoria
1031 Australia) as per manufacturer's instructions with two modifications, specifically 2µM of SYTO9
1032 was added to each reaction to monitor PCR amplification and the Ampliseq PCR reaction was run
1033 for 22 cycles when all samples had entered the exponential amplification phase of the reaction. In
1034 addition to the SARS-CoV-2 positive and negative pool and individual samples, three no cDNA
1035 controls, i.e. nucleic acids samples not going through cDNA preparation but directly used for panel
1036 amplification, from samples GC-11, GC-21 and negative pool 4 were included as DNA controls.
1037 Sample libraries were quantitated and sequenced as for the subgenomic RNA Ampliseq panels
1038 described above.
1039 Amplicons were mapped to the latest RefSeq human transcript of human genome hg19 (GRCh38)
1040 obtained from the Ion Ampliseq Designer [https://ampliseq.com] (Thermofisher Scientific) using the
1041 TMAP software as part of the Ion Torrent Server Suite 5.10.1. Read counts mapping to each amplicon
1042 target were calculated using the Coverage Analysis plugin (v5.12.0.0) within the Torrent server suite,
1043 and the read mapping visualized in IGV.

1044 The large majority of the designed amplicons within this Ampliseq panel were designed to span
1045 introns within inflammatory genes, and therefore theoretically only amplify cDNA generated from
1046 cellular mRNA transcripts. However, it was apparent from the no cDNA controls, that a number of
1047 amplicons could be readily amplified from cellular DNA in the samples. Therefore, these amplicons
1048 were excluded from subsequent analysis in all libraries. The excluded amplicons are shown in the
1049 source data file. The mapping of reads to the remaining 592 inflammatory gene targets was visualized
1050 in IGV to find the most appropriate settings for the Coverage Analysis plugin to count the number of
1051 reads mapped to each amplicon's target. At a low mapping quality scores, reads spanning a target but
1052 with additional sequence were visible. These could have been products from single primer
1053 amplification of cellular DNA. Therefore, a minimum alignment score of 74 (the length of the shortest
1054 amplicon excluding primers) and a mapping quality score of 90 were settled upon for quantitation of
1055 amplicon reads. At these settings, any reads not spanning the full amplicon, or reads with additional
1056 sequence beyond the target (i.e. exon sequence from a single primer amplification) were filtered out
1057 from the target read counts.
1058 The read count data from the samples was normalized using python to calculate the reads per million
1059 mapped reads (RPM) to the inflammatory targets, and the Matplotlib version 3.4.2[50], Pandas version
1060 1.2.4[51] and Seaborn 0.11.1[52] python packages were then used to create cluster maps of the RPM per
1061 sample to look for patterns of gene expression in the data, and to create bar plots of the relative levels
1062 of expression in the samples. To focus in on the targets with the greatest differential expression in
1063 SARS-CoV-2 positive and negative samples, samples were grouped based on their time since the
1064 onset of clinical signs. SARS-CoV-2 positive samples taken on the day of onset of symptoms were
1065 classified as 'early positive' swabs, while those taken afterwards were grouped as 'late positive'
1066 swabs. The four SARS-CoV-2 negative pools and two negative individuals made up the negative
1067 group. To highlight genes with potential differential expression between the groups, the Kruskal-
1068 Wallis test and Wilcoxon Rank-Sum test was applied via the statistical package within the SciPy
1069 python library (1.6.1)[53]. The Benjamini-Hochberg method[54] with a false discovery rate of 25% was
1070 applied to restrict the analysis to only those genes with the largest differences between the groups.
1071 Read counts per million mapped reads from genes with statistically different expression were then
1072 plotted as violin plots and by the days since clinical onset to look for patterns in differential gene
1073 expression. Gene relationships were analysed using String v11[55] and the statistical over-
1074 representation test in Panther Gene Ontology version 16[56].
1075
1076
1077 **Reporting Summary**

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data Availability

The sequence reads of our SARS-CoV-2 positive and negative samples reported here have been deposited along with their corresponding fasta reference files in the NCBI Sequence Read Archive (SRA) under BioProjectID: PRJNA738539. Assembled near full length SARS-CoV-2 genomic sequences are deposited in NCBI GenBank under accessions MZ410617-MZ410622.
Other assembled nucleotide sequences or nucleotide sequence read archives mentioned are publicly available at NCBI ([https://www.ncbi.nlm.nih.gov/nuccore/ and https://www.ncbi.nlm.nih.gov/sra/]. All other data supporting the findings of this manuscript are available in the Supplementary Information files or from the corresponding author upon reasonable request. Source data are provided with this paper. A reporting summary for this Article is available as a Supplementary Information file.

Code Availability

Not applicable.

References

1 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).

2 WHO. *WHO characterizes COVID-19 as a pandemic*, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen> (2020).

3 Zou, L. *et al.* SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N. Engl. J. Med.* **382**, 1177-1179, doi:10.1056/NEJMc2001737 (2020).

4 Sethuraman, N., Jeremiah, S. S. & Ryo, A. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA* **323**, 2249-2251, doi:10.1001/jama.2020.8259 (2020).

5 Wölfel, R. *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465-469, doi:10.1038/s41586-020-2196-x (2020).

6 Bullard, J. *et al.* Predicting Infectious Severe Acute Respiratory Syndrome Coronavirus 2 From Diagnostic Samples. *Clin. Infect. Dis.* **71**, 2663-2666, doi:10.1093/cid/ciaa638 (2020).

7 van Kampen, J. J. A. *et al.* Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). *Nature Communications* **12**, 267, doi:10.1038/s41467-020-20568-4 (2021).

8 Perera, R. A. P. M. *et al.* SARS-CoV-2 Virus Culture and Subgenomic RNA for Respiratory Specimens from Patients with Mild Coronavirus Disease. *Emerg. Infect. Dis.* **26**, 2701-2704, doi:10.3201/eid2611.203219 (2020).

9 V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology* **19**, 155-170, doi:10.1038/s41579-020-00468-6 (2021).

1121 10 Sola, I., Almazán, F., Zúñiga, S. & Enjuanes, L. Continuous and Discontinuous RNA Synthesis in
1122 Coronaviruses. *Annual Review of Virology* **2**, 265-288, doi:10.1146/annurev-virology-100114-
1123 055218 (2015).
1124 11 Alexandersen, S., Chamings, A. & Bhatta, T. R. SARS-CoV-2 genomic and subgenomic RNAs in
1125 diagnostic samples are not an indicator of active replication. *Nature Communications* **11**, 6059,
1126 doi:10.1038/s41467-020-19883-7 (2020).
1127 12 Snijder, E. J. *et al.* A unifying structural and functional model of the coronavirus replication
1128 organelle: Tracking down RNA synthesis. *PLoS Biol.* **18**, e3000715,
1129 doi:10.1371/journal.pbio.3000715 (2020).
1130 13 Hogan, C. A. *et al.* Strand-Specific Reverse Transcription PCR for Detection of Replicating SARS-CoV-
1131 2. *Emerg. Infect. Dis.* **27**, 632-635, doi:10.3201/eid2702.204168 (2021).
1132 14 Dimcheff, D. E. *et al.* SARS-CoV-2 Total and Subgenomic RNA Viral Load in Hospitalized Patients.
1133 *medRxiv*, 2021.2002.2025.21252493, doi:10.1101/2021.02.25.21252493 (2021).
1134 15 Costantino, V. & Raina MacIntyre, C. The Impact of Universal Mask Use on SARS-COV-2 in Victoria,
1135 Australia on the Epidemic Trajectory of COVID-19. *Frontiers in Public Health* **9**,
1136 doi:10.3389/fpubh.2021.625499 (2021).
1137 16 Miao, Z., Tidu, A., Eriani, G. & Martin, F. Secondary structure of the SARS-CoV-2 5'-UTR. *RNA Biol.*
1138 **18**, 447-456, doi:10.1080/15476286.2020.1814556 (2021).
1139 17 McAloon, C. *et al.* Incubation period of COVID-19: a rapid systematic review and meta-analysis of
1140 observational research. *BMJ Open* **10**, e039652, doi:10.1136/bmjopen-2020-039652 (2020).
1141 18 Hume, A. J. *et al.* Inactivation of RNA Viruses by Gamma Irradiation: A Study on Mitigating Factors.
1142 *Viruses* **8**, 204, doi:10.3390/v8070204 (2016).
1143 19 Parker, M. D. *et al.* Altered Subgenomic RNA Expression in SARS-CoV-2 B.1.1.7 Infections. *bioRxiv*,
1144 2021.2003.2002.433156, doi:10.1101/2021.03.02.433156 (2021).
1145 20 Dominguez Andres, A. *et al.* SARS-CoV-2 ORF9c Is a Membrane-Associated Protein that Suppresses
1146 Antiviral Responses in Cells. *bioRxiv*, 2020.2008.2018.256776, doi:10.1101/2020.08.18.256776
1147 (2020).
1148 21 Leary, S. *et al.* Generation of a novel SARS-CoV-2 sub-genomic RNA due to the R203K/G204R
1149 variant in nucleocapsid. *bioRxiv*, 2020.2004.2010.029454, doi:10.1101/2020.04.10.029454 (2021).
1150 22 Weber, S., Ramirez, C. & Doerfler, W. Signal hotspot mutations in SARS-CoV-2 genomes evolve as
1151 the virus spreads and actively replicates in different parts of the world. *Virus Res.* **289**, 198170-
1152 198170, doi:10.1016/j.virusres.2020.198170 (2020).
1153 23 Banerjee, A. K. *et al.* SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress
1154 Host Defenses. *Cell* **183**, 1325-1339.e1321, doi:https://doi.org/10.1016/j.cell.2020.10.004 (2020).
1155 24 Zhu, N. *et al.* Morphogenesis and cytopathic effect of SARS-CoV-2 infection in human airway
1156 epithelial cells. *Nature Communications* **11**, 3910, doi:10.1038/s41467-020-17796-z (2020).
1157 25 Li, Y. *et al.* SARS-CoV-2 induces double-stranded RNA-mediated innate immune responses in
1158 respiratory epithelial-derived cells and cardiomyocytes. *Proceedings of the National Academy of*
1159 *Sciences* **118**, e2022643118, doi:10.1073/pnas.2022643118 (2021).
1160 26 Tian, W. *et al.* Immune suppression in the early stage of COVID-19 disease. *Nat Commun* **11**, 5859,
1161 doi:10.1038/s41467-020-19706-9 (2020).
1162 27 Yamada, T. *et al.* RIG-I triggers a signaling-abortive anti-SARS-CoV-2 defense in human lung cells.
1163 *Nat. Immunol.*, doi:10.1038/s41590-021-00942-0 (2021).
1164 28 Wu, J. *et al.* SARS-CoV-2 ORF9b inhibits RIG-I-MAVS antiviral signaling by interrupting K63-linked
1165 ubiquitination of NEMO. *Cell Reports* **34**, 108761, doi:https://doi.org/10.1016/j.celrep.2021.108761
1166 (2021).
1167 29 Yin, X. *et al.* MDA5 Governs the Innate Immune Response to SARS-CoV-2 in Lung Epithelial Cells.
1168 *Cell reports* **34**, 108628-108628, doi:10.1016/j.celrep.2020.108628 (2021).
1169 30 Wang, J. *et al.* Negative Regulation of Nmi on Virus-Triggered Type I IFN Production by Targeting
1170 IRF7. *The Journal of Immunology* **191**, 3393, doi:10.4049/jimmunol.1300740 (2013).
1171 31 Jensen, S. & Thomsen, A. R. Sensing of RNA Viruses: a Review of Innate Immune Receptors Involved
1172 in Recognizing RNA Virus Invasion. *J. Virol.* **86**, 2900, doi:10.1128/JVI.05738-11 (2012).

32 McComb, J. G. *et al.* CX3CL1 up-regulation is associated with recruitment of CX3CR1+ mononuclear phagocytes and T lymphocytes in the lungs during cigarette smoke-induced emphysema. *Am. J. Pathol.* **173**, 949-961, doi:10.2353/ajpath.2008.071034 (2008).
33 Leslie, J. *et al.* FPR-1 is an important regulator of neutrophil recruitment and a tissue-specific driver of pulmonary fibrosis. *JCI insight* **5**, doi:10.1172/jci.insight.125937 (2020).
34 Brennan, F. H. *et al.* Complement receptor C3aR1 controls neutrophil mobilization following spinal cord injury through physiological antagonism of CXCR2. *JCI insight* **4**, doi:10.1172/jci.insight.98254 (2019).
35 Kamata, H. *et al.* Epithelial Cell-Derived Secreted and Transmembrane 1a Signals to Activated Neutrophils during Pneumococcal Pneumonia. *Am. J. Respir. Cell Mol. Biol.* **55**, 407-418, doi:10.1165/rcmb.2015-0261OC (2016).
36 Wiley, S. R. *et al.* Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* **3**, 673-682, doi:10.1016/1074-7613(95)90057-8 (1995).
37 Das, A., Dinh, P. X. & Pattnaik, A. K. Trim21 regulates Nmi-IFI35 complex-mediated inhibition of innate antiviral response. *Virology* **485**, 383-392, doi:https://doi.org/10.1016/j.virol.2015.08.013 (2015).
38 Wolff, G. *et al.* A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* **369**, 1395, doi:10.1126/science.abd3629 (2020).
39 Gamage, A. M. *et al.* Infection of human Nasal Epithelial Cells with SARS-CoV-2 and a 382-nt deletion isolate lacking ORF8 reveals similar viral kinetics and host transcriptional profiles. *PLoS Pathog.* **16**, e1009130, doi:10.1371/journal.ppat.1009130 (2020).
40 Crystal, R. G., Randell, S. H., Engelhardt, J. F., Voynow, J. & Sunday, M. E. Airway Epithelial Cells. *Proc. Am. Thorac. Soc.* **5**, 772-777, doi:10.1513/pats.200805-041HR (2008).
41 Erjefält, J. S., Erjefält, I., Sundler, F. & Persson, C. G. In vivo restitution of airway epithelium. *Cell Tissue Res.* **281**, 305-316, doi:10.1007/bf00583399 (1995).
42 Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* **395**, 1054-1062, doi:10.1016/S0140-6736(20)30566-3 (2020).
43 Hoang, V. T., Dao, T. L. & Gautret, P. Recurrence of positive SARS-CoV-2 in patients recovered from COVID-19. *J. Med. Virol.* **92**, 2366-2367, doi:https://doi.org/10.1002/jmv.26056 (2020).
44 Gaebler, C. *et al.* Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**, 639-644, doi:10.1038/s41586-021-03207-w (2021).
45 Mayhew, T. M., Myklebust, R., Whybrow, A. & Jenkins, R. Epithelial integrity, cell death and cell loss in mammalian small intestine. *Histol. Histopathol.* **14**, 257-267, doi:10.14670/hh-14.257 (1999).
46 Bhatta, T. R., Chamings, A., Vibin, J. & Alexandersen, S. Detection and characterisation of canine astrovirus, canine parvovirus and canine papillomavirus in puppies using next generation sequencing. *Sci Rep* **9**, 4602, doi:10.1038/s41598-019-41045-z (2019).
47 Bhatta, T. R. *et al.* Sequence analysis of travel-related SARS-CoV-2 cases in the Greater Geelong region, Australia. *medRxiv* (2020).
48 Caboche, S., Audebert, C., Lemoine, Y. & Hot, D. Comparison of mapping algorithms used in high-throughput sequencing: application to Ion Torrent data. *BMC Genomics* **15**, 264, doi:10.1186/1471-2164-15-264 (2014).
49 Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178-192, doi:10.1093/bib/bbs017 (2013).
50 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95, doi:10.1109/MCSE.2007.55 (2007).
51 Pandas development team. Pandas. doi:10.5281/zenodo.3509134 (2020).
52 Waskom, M. & Seaborn Development Team. Seaborn. doi:https://doi.org/10.5281/zenodo.592845 (2020).
53 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261-272, doi:10.1038/s41592-019-0686-2 (2020).

54 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
55 Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607-D613, doi:10.1093/nar/gky1131 (2018).
56 Mi, H. *et al.* PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, D394-D403, doi:10.1093/nar/gkaa1106 (2020).

Acknowledgements

We gratefully acknowledge clinical staff for sample collection and the diagnostic staff at the Australian Rickettsial Reference Laboratory (ARRL) and Australian Clinical Labs (ACL) for performing the initial diagnostic SARS-CoV-2 testing, in particular Professor John Stenos and Dr Mythili Tadepalli from ARRL and Professor Owen Harris, Dr Kwee Chin Liew and Dr Richard McCoy from ACL for their assistance in testing, aliquoting, storing and providing samples. We acknowledge Dr Darcie Paige Cooper for retrieving summary clinical information and Professors Peter Vuillermin and Eugene Athan for the coordination of the Barwon Health and Deakin University COVID-19 Research Task Force and Cohort Study and the human research ethics approval. We acknowledge Thermofisher Scientific, Victoria, Australia, for supplying the SARS-CoV-2 sequencing panel and for help with the custom Ampliseq mini-panels used. We acknowledge colleagues at the CSIRO Australian Centre for Disease Preparedness, especially Dr Alexander J. McAuley and Professor Seshadri S. Vasan, for providing the SARS-CoV-2 positive cell culture. We acknowledge Professor Alister Ward and Dr Poshmaal Dhar for their helpful comments on the draft manuscript. We also acknowledge Dr Jason Hodge, laboratory manager of the GCEID laboratory for his technical input. This research was funded by Deakin University, Barwon Health and CSIRO and from the National Health and Medical Research Council (NHMRC) equipment grant number GNT9000413 to S.A.

Author Contributions

S.A. initiated the work and coordinated work carried out at the GCEID laboratory. T.R.B. and S.A. designed the PCRs and primers/probes and S.A. selected and designed the mini-panels for NGS. T.R.B. managed the collection of samples and performed all strand and non-strand specific PCR assays. A.C. performed all mini-panel, SARS-CoV-2 and inflammatory Ampliseq panels. S.A., A.C. and T.R.B. all analysed the PCR and sequence data. A.C. prepared the initial draft of the paper together with S.A. and all authors contributed to the editing and drafting process.

Competing Interests

The authors declare no competing or conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Table and Figure Legends

Table 1. Table showing summary information about the individuals and samples included in this study

Samples from a total of 19 individuals, including two SARS-CoV-2 negative control individuals (Individual 1 and 22), are detailed below along with four pools of ten SARS-CoV-2 negative samples used as additional negative controls. Swab sample identification (ID), clinical symptoms onset and description where available, sampling date and the results of the initial diagnostic SARS-COV-2 RT-PCR test (Ct value) are provided.

Individual ID	Gender	Age group (Years)	Swab Sample ID	Epidemic wave sample taken†	Days since onset of symptoms/Initial swab	Clinical Symptoms at swab collection	Hospitalized	Initial SARS-CoV2 RT-PCR result (Ct Value)	Comments
1	F	20-40	GC-28	1	0*	Fever, cough sore throat	N	Not detected	Negative swab sample used in PCR assays
2	F	20-40	GC-26	1	7	Sore throat, dry cough	N	Detected (21)	
3	F	20-40	GC-13	1	4	Body aches, headache, dry cough, shortness of breath	N	Detected (29)	
4	F	40-60	GC-11	1	0*	Cold, sinusitis	N	Detected (19)	Very little remaining sample – only used for inflammatory gene analysis
			GC-20		9*	Cold, sinusitis		Detected (31)	Weak positive sample. Not used in this study
			GC-24		14*	Cold, sinusitis		Detected (31)	
5	F	20-40	GC-12	1	1	Sore throat, rigor, fever	Y	Detected (31)	Weak positive sample, only tested in Duplex PCR assay
6	F	40-60	GC-14	1`	0	Unspecified	N	Detected (18)	
			GC-23		11	Asymptomatic		Detected (31)	
			GC-51		17	Asymptomatic		Detected (31)	
7	F	40-60	GC-21	1	16	Sore throat, shortness of breath, runny nose	Y	Detected (31)	

			GC-58		47	Not specified		Detected (38)	
8	M	40-60	GC-25	1	2	Sore throat, hoarse voice	N	Detected (19)	
11	F	80-	GC-242	2	0	Not specified	N	Detected (28)	
			GC-290		11	Not specified		Detected (34)	
12	F	60-80	GC-251	2	0	Not specified	N	Detected (17)	
			GC-288		10			Detected (34)	
13	F	40-60	GC-277	2	0	Asymptomatic	N	Detected (21)	
			GC-329		19			Detected (34)	
14	F	80-	GC-238	2	0	Not specified		Detected (25)	
			GC-291		13			Detected (22)	
16	F	80-	GC-366	2	14	Not specified	Y	Detected (30)	
			GC-365		17			Detected (33)	
17	F	40-60	GC-316	2	1	Fever, runny nose, headache	N	Detected (20)	
18	M	60-80	GC-310	2	17	Cough, shortness of breath	Y	Detected (23)	
19	M	40-60	GC-199	2	1	Fever, cough, shortness of breath, headache	Y	Detected (24)	
20	F	80-	GC-292	2	10	Not specified	Y	Detected (18)	
21	F	40-60	GC-243	2	0	Not specified	N	Detected (24)	
22	M	20-40	GC-5	1	0*	Not specified	N	Not detected	Negative control in inflammatory gene analysis
Negative Pool 1				1		Pool of 10 individuals		Not detected	This pool was used as a negative control for the PCR assays
Negative Pool 2				1		Pool of 10 individuals		Not detected	Negative control in inflammatory gene analysis
Negative Pool 3				2		Pool of 10 individuals		Not detected	Negative control in inflammatory gene analysis
Negative Pool 4				2		Pool of 10 individuals		Not detected	Negative control in inflammatory gene analysis

*Actual onset of symptoms was not recorded for theses samples, so we counted the days from the time of initial swab collection. † 1 - first epidemic wave in Victoria Australia during March-May 2020 and 2 - second epidemic wave in Victoria Australia July-September 2020. Negative pools 1 and 2 were made from SARS-CoV2 negative individuals tested during the first wave of SARS-CoV-2 infections. Negative pools 3 and 4 were made from negative individuals tested during the second wave of SARS-CoV-2 infections.

Table 2. Ct values obtained from the single target SARS-CoV-2 genomic and subgenomic PCR assays for the swab samples from the 17 SARS-CoV-2 positive individuals, negative individual and pool and from 48hr virus cell culture supernatant.

Individual	Swab Sample ID	Days since onset of symptoms/Initial swab	5'UTR Genomic PCR	ORF7a subgenomic PCR	N gene subgenomic PCR		7a Genomic and Subgenomic PCR
1	GC-28	0*	ND	ND	NEG	NEG	ND
2	GC-26	7	25.6	30	27.6	27.4	25.3
3	GC-13	4	29.6	33.3	29.8	30.1	28.6
4	GC-11	0*	20.4	28.5	ND	ND	20.1
	GC-20	9*	NEG	NEG	ND	ND	NEG
	GC-24	14*	35.4	NEG	36.2	NEG	33.6
5	GC-12	1	NEG	NEG	NEG	NEG	NEG
6	GC-14	0	18.7	23.7	20.2	19.8	17.9
	GC-23	11	33.7	38.6	34.7	34.6	32.6
	GC-51	17	39.8	NEG	34.1	35.5	36
7	GC-21	16	35.7	NEG	36.7	NEG	33.7
	GC-58	47	NEG	NEG	NEG	NEG	NEG
8	GC-25	2	21.5	25.4	21.1	20.9	19.9
11	GC-242	0	27.3	30.3	26.5	26.7	26.1
	GC-290	11	33.5	36.7	33.5	34.2	31.9
12	GC-251	0	17	21.2	17.6	17.6	16.5
	GC-288	10	36.7	37.1	34.8	NEG	34.7
13	GC-277	0	20.2	25.1	21.4	21.2	18.9
	GC-329	19	37	NEG	35.7	NEG	34.6
14	GC-238	0	24.3	29.5	26.7	26.7	22.7
	GC-291	13	22	24.8	22.8	23	20.4
16	GC-366	14	39	NEG	31.9	38.5	33.7
	GC-365	17	38.1	NEG	NEG	NEG	34.1
17	GC-316	1	19.5	22.6	19.5	19.6	18.1
18	GC-310	17	21.5	25.9	23.7	23.8	21.7
19	GC-199	1	23	28.1	24.8	24.8	23.1
20	GC-292	10	17.6	20.3	17.5	17.3	17.2
21	GC-243	0	24.7	27.6	23.4	23.7	23.4
SARS-CoV-2 Negative swab pool	Pool of 10 Neg. swabs	Not specified	ND	ND	NEG	NEG	ND
Cell culture 48Hr		48hrs post inoculation	16.8	20.4	19.5	19.3	ND

ND: Test not performed. NEG: SARS-CoV-2 not detected.

*Date of onset of clinical signs not provided. Number of days since the collection of the first nasopharyngeal swabs is shown.

Table 3. Ct values obtained from the duplex SARS-CoV-2 genomic and subgenomic PCR assays for the swab samples from the 17 SARS-CoV-2 positive individuals, negative individual and pool and from 48hr virus cell culture supernatant.

Individual	Swab Sample ID	Days since onset of symptoms/Initial swab	5'UTR Genomic Duplex PCR	ORF7a subgenomic Duplex PCR
1	GC-28	0*	NEG	NEG
2	GC-26	7	23.9	27.5
3	GC-13	4	27.3	30.7
4	GC-11	0*	36.9	NEG
	GC-20	9*	34.3	36.2
	GC-24	14*	33.8	NEG
5	GC-12	1	29.2	NEG
6	GC-14	0	16.8	20.5
	GC-23	11	31.6	NEG
	GC-51	17	35.4	NEG
7	GC-21	16	34.5	NEG
	GC-58	47	NEG	NEG
8	GC-25	2	19	22.4
11	GC-242	0	25.8	28.6
	GC-290	11	31.9	33.9
12	GC-251	0	15.7	18.9
	GC-288	10	33.7	NEG
13	GC-277	0	18.3	22.4
	GC-329	19	34.8	36.6
14	GC-238	0	21.8	27.6
	GC-291	13	19.9	23.3
16	GC-366	14	34.7	NEG
	GC-365	17	34.7	NEG
17	GC-316	1	17	20.3
18	GC-310	17	20.5	24.4
19	GC-199	1	20.6	25.6
20	GC-292	10	15.4	18.9
21	GC-243	0	22.7	25.6
SARS-CoV-2 Negative swab pool	Pool of 10 Neg. swabs	Not specified	ND	ND
Cell culture 48Hr		48 hrs post inoculation	17.3	20

ND: Test not performed. NEG: SARS-CoV-2 not detected.

*Date of onset of clinical signs not provided. Number of days since the collection of the first nasopharyngeal swabs is shown.

Table 4. Ct values obtained from the triplex SARS-CoV-2 genomic and subgenomic PCR assays for the swab samples from the 17 SARS-CoV-2 positive individuals, negative individual and pool and from 48hr virus cell culture supernatant.

Individual	Swab Sample ID	Days since onset of symptoms/Initial swab	5'UTR Genomic PCR Repeat 1	5'UTR Genomic PCR Repeat 2	ORF7a subgenomic PCR Repeat 1	ORF7a subgenomic PCR Repeat 2	N gene subgenomic PCR Repeat 1	N gene subgenomic PCR Repeat 2
1	GC-28	0*	NEG	NEG	NEG	NEG	NEG	NEG
2	GC-26	7	24.7	24.5	27.8	27.7	27.8	27.9
3	GC-13	4	27.7	28.2	31.5	31.4	30.3	30.8
4	GC-11	0*	ND	ND	ND	ND	ND	ND
	GC-20	9*	ND	ND	ND	ND	ND	ND
	GC-24	14*	34.2	34.2	NEG	NEG	NEG	NEG
5	GC-12	1	NEG	NEG	NEG	NEG	NEG	NEG
6	GC-14	0	17.8	18	21.3	21.5	20.4	20.5
	GC-23	11	32.2	31.5	36.4	NEG	33.9	34.4
	GC-51	17	36.3	36.9	NEG	NEG	NEG	NEG
7	GC-21	16	35.3	33.9	NEG	NEG	36.3	NEG
	GC-58	47	NEG	NEG	NEG	NEG	NEG	NEG
8	GC-25	2	19.2	19.5	22.4	22.5	21.3	21.3
11	GC-242	0	26.3	26.2	28.9	28.7	27.1	27.3
	GC-290	11	31.8	32.2	35	35	34.2	33.4
12	GC-251	0	16.5	16.5	19.3	19.4	17.9	18
	GC-288	10	33.5	34.2	34.9	35.3	35.3	35.2
13	GC-277	0	18.7	18.8	22.6	22.5	21.6	21.6
	GC-329	19	35.5	35	37.4	35.8	NEG	34.8
14	GC-238	0	22	21.9	27.8	27.6	26.9	26.8
	GC-291	13	20.4	20.7	23.6	23.8	23	23.2
16	GC-366	14	34.8	34.8	NEG	NEG	35.9	35.9
	GC-365	17	34.5	36.4	NEG	36.1	35.4	NEG
17	GC-316	1	17.5	17.8	20.7	20.9	19.9	19.8
18	GC-310	17	21.3	21.3	25	25	24.2	24.1
19	GC-199	1	22.5	22.6	26.5	26.4	25.1	25
20	GC-292	10	15.9	16	19.3	19.4	18.2	18.3
21	GC-243	0	22.4	22.8	25.2	25.5	23.8	24.2
SARS-CoV-2 Negative swab pool	Pool of 10 Neg. swabs	Not specified	NEG	NEG	NEG	NEG	NEG	NEG
Cell culture 48Hr		48 hr post inoculation	16.9	17.1	20.8	21	19.6	19.8

ND: Test not performed. NEG: SARS-CoV-2 not detected.

*Date of onset of clinical signs not provided. Number of days since the collection of the first nasopharyngeal swabs is shown.

Table 5. Ct values obtained by the strand specific RT-PCRs of 14 SARS-CoV-2 positive nasopharyngeal swabs and the 48hr cell culture supernatant. Individual 1 was used as a negative control

Individual	Swab Sample ID	Days since onset of symptoms/Initial swab	ORF7a total +ve strand		ORF7a total -ve strand		ORF7a subgenomic +ve strand		ORF7a subgenomic -ve strand		N gene subgenomic +ve strand		N gene subgenomic -ve strand	
1	GC-28	0*	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
2	GC-26	7	24.3	24.6	33.2	34.7	33.2	34.1	NEG	38.3	29.6	29.8	39.2	35.4
3	GC-13	4	35.8	35.3	NEG	NEG	NEG	NEG	NEG	NEG	34.6	33.3	NEG	NEG
6	GC-14	0	20.6	20.4	27.7	28.2	25.6	24.8	31.2	31.6	24.1	24.1	27.9	27.7
8	GC-25	2	21.3	21	26.9	27.2	25.6	25.4	31	31	24.2	24.1	29	29.2
11	GC-242	0	27.8	28.1	NEG	NEG	NEG	NEG	37.5	NEG	30.7	30.5	NEG	NEG
12	GC-251	0	18	17.8	26.1	27.3	22.5	22.3	33.1	32.2	20.6	20.6	28.6	28.5
13	GC-277	0	20.4	20.5	29	28.6	24	25.5	31.9	33.5	22.7	23.8	29	28.4
14	GC-238	0	24	24.4	33.8	33	NEG	NEG	NEG	NEG	29.4	28.8	34.7	NEG
	GC-291	13	22.8	22.6	30.5	30	NEG	ND	NEG	ND	27.8	ND	36.1	ND
17	GC-316	1	20.4	20.5	26.2	25.3	27.3	ND	34.3	ND	23.8	ND	29.5	ND
18	GC-310	17	23.6	23.1	29.9	29.8	NEG	ND	NEG	ND	27.9	ND	32.7	ND
19	GC-199	1	NEG	30.9	NEG	NEG	NEG	38.1	NEG	NEG	27	26.9	NEG	NEG
20	GC-292	10	18.8	19	24.3	24.6	25.7	ND	31.4	ND	22.1	ND	26.9	ND
21	GC-243	0	24.1	24.6	32.6	30.4	NEG	NEG	34.2	34.3	27.2	26.8	31.4	31.3
SARS-CoV-2 Negative swab pool	Pool of 10 Neg. swabs	NA	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG	NEG
Cell culture 48Hr		NA	14.6	15.1	25.5	24.7	20.9	21.2	33.2	33.5	19.5	19.4	31.6	30.4

ND: insufficient RNA remaining to perform repeat PCR

*Date of onset of clinical signs not provided. Number of days since the collection of the first nasopharyngeal swabs is shown.

Figure 1. The Ct values of the four individual PCR assays targeting the genomic 5' UTR, subgenomic ORF7a and N gene, and total ORF7a (detecting genomic and S, ORF3a, E, M, ORF6 and ORF7a subgenomic RNAs) for the naso-oropharyngeal swabs from SARS-CoV-2 positive individuals, the 48 hr cell culture and the SARS-CoV-2 negative pool of swabs. The results are ordered by ascending values of the 5'UTR Ct and show that both the ORF7a and N gene subgenomic, and the ORF7a total PCR Ct's also increase as the UTR Ct increases. Above a Ct of around 30 for the 5'UTR, the detection of subgenomic RNAs becomes less consistent/below levels of detection/quantitation, and the general pattern of Ct values following ORF 7a total < 5' UTR < N gene subgenomic RNA < ORF7a subgenomic RNA becomes unreliable. (n=25 data points from 25 biological samples run once (Subgenomic N run twice)).

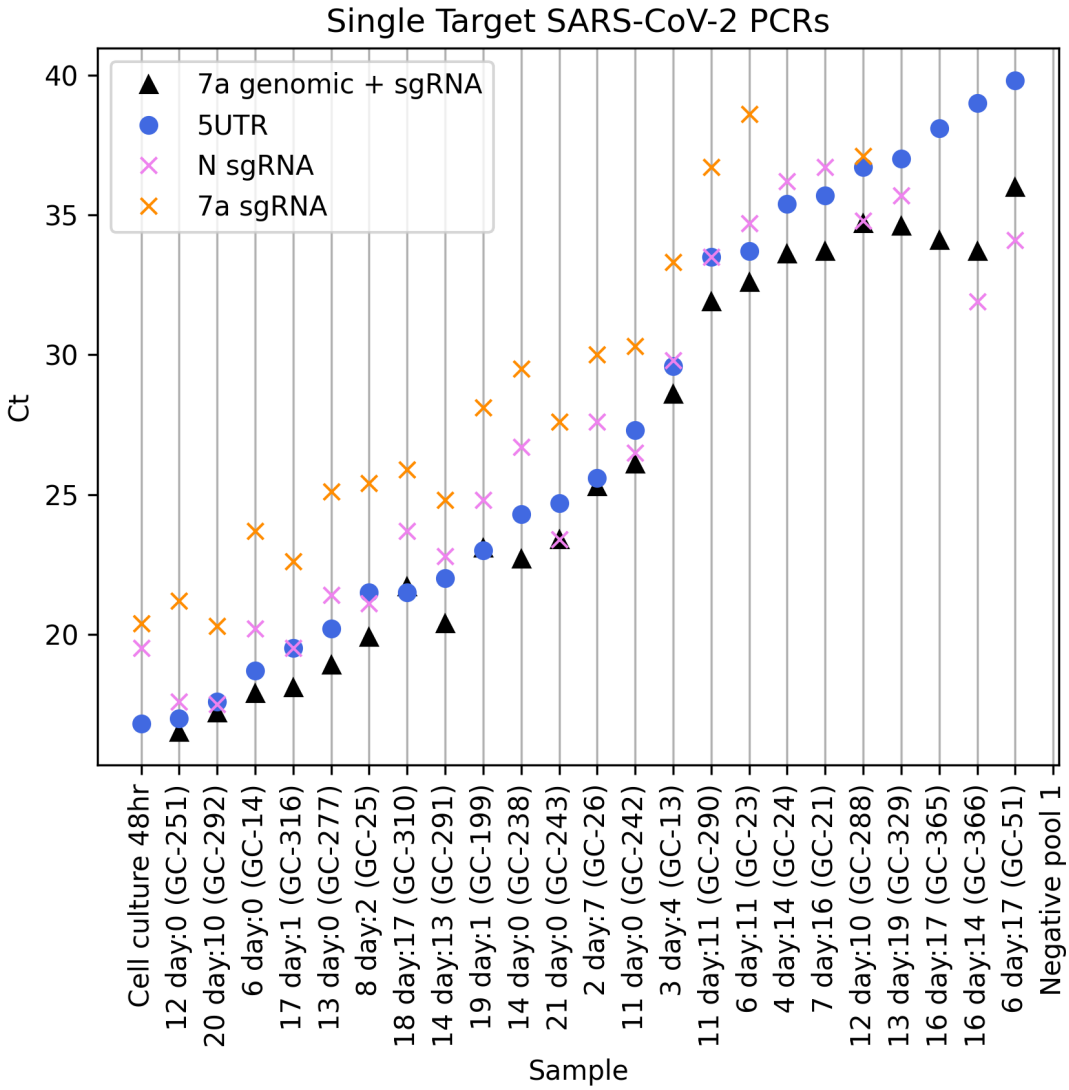


Figure 2. The average Ct's of the replicates of the 5'UTR genomic and ORF7a and N gene subgenomic triplex PCRs for the SARS-CoV-2 positive naso-oropharyngeal swabs, negative swab pool and the SARS-CoV-2 48 hr cell culture.

The sample results have been ordered by ascending values of the 5'UTR PCR Ct, and show that both of the subgenomic assays have a higher Ct, with ORF7a subgenomic PCR being consistently higher than the N gene subgenomic assay, and both increase in Ct as the 5' UTR Ct increases. Above a 5'UTR Ct of 30, this becomes inconsistent due to high or non-detectable Ct values for the subgenomic targets. (n=24 data points from 24 samples run twice)

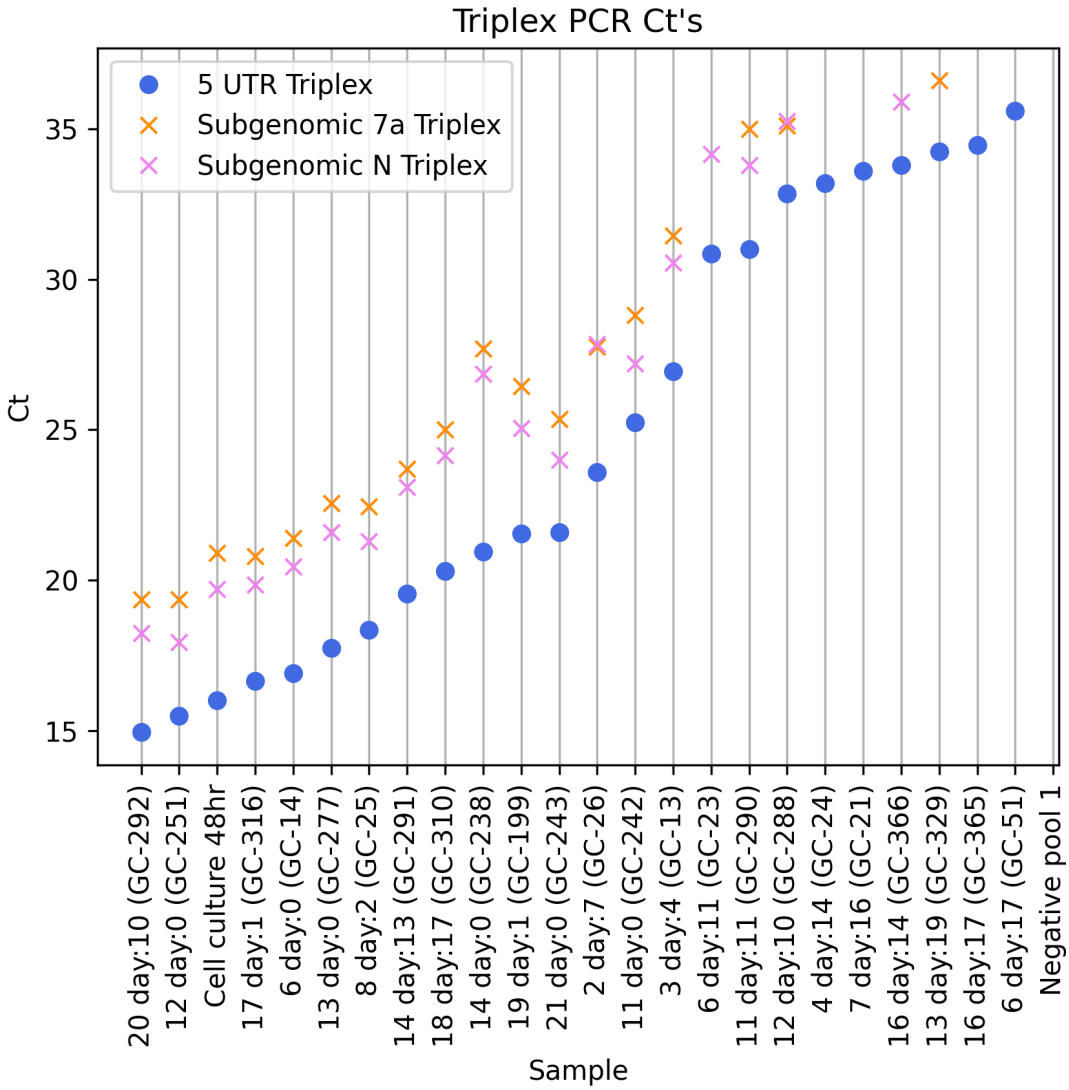


Figure 3. A: Violin plots of the difference in Ct's (delta Ct) between both the single target ORF 7a and N subgenomic (average Ct of replicates) assays and the 5'UTR genomic PCR. B: Violin plots of the ratio of genomic to subgenomic RNA for both the ORF7a and N gene subgenomic targets.

The accompanying scatter plots of delta Ct's per sample are shown on the right. The individual, and collection day post onset of initial clinical signs of each swab sample are shown in the X axis of the scatter plot. (n=16 data points from 16 biological samples run once, subgenomic N run twice)

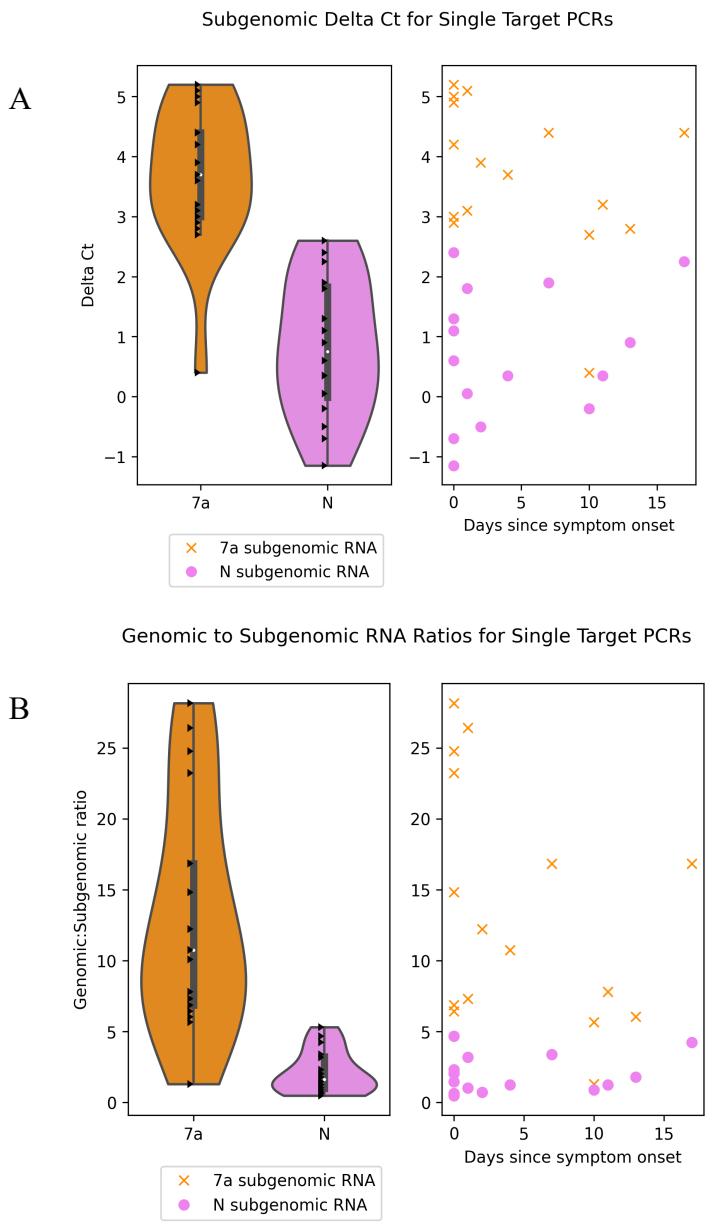
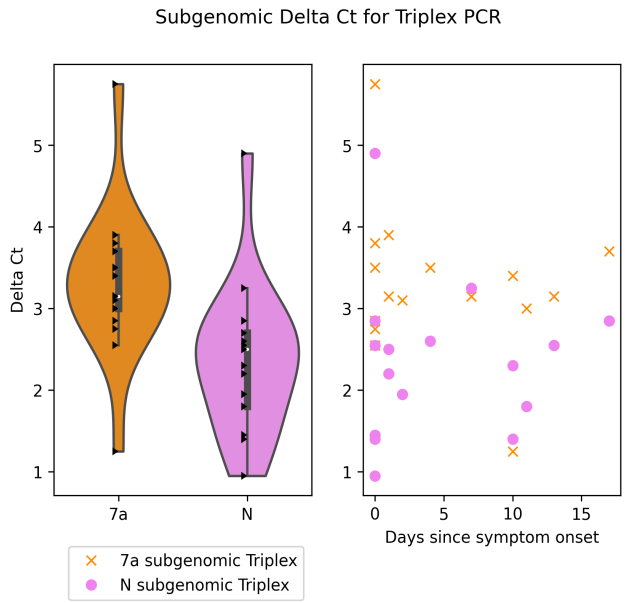


Figure 4. A: Violin plots of the difference in Ct's (delta Ct) between the average of the replicates of the triplex PCR ORF 7a and N subgenomic targets and the 5'UTR genomic target. B: Violin plots of the ratio of genomic to subgenomic RNA for the ORF7a and N gene subgenomic targets.

The accompanying scatter plots of delta Ct's per sample are shown on the right. The individual, and collection day post onset of initial clinical signs of each swab sample are shown in the X axis of the scatter plot. (n=16 from 16 biological samples run twice)

A



B

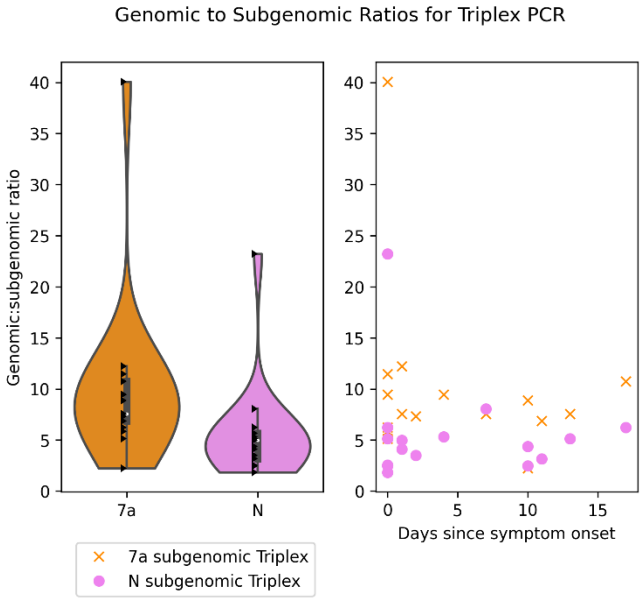


Figure 5. The average Ct values of the negative and positive strand specific PCRs replicates for the ORF7a total RNA, 7a subgenomic and N subgenomic RNA targets. Positive strand RNA was detected in all 14 naso-oropharyngeal swab samples tested, and one or more negative strand RNA molecules in 12 of the swab samples. Swab samples have been arranged by their ORF7a total PCR Ct. (n=15 from 15 biological samples run twice)

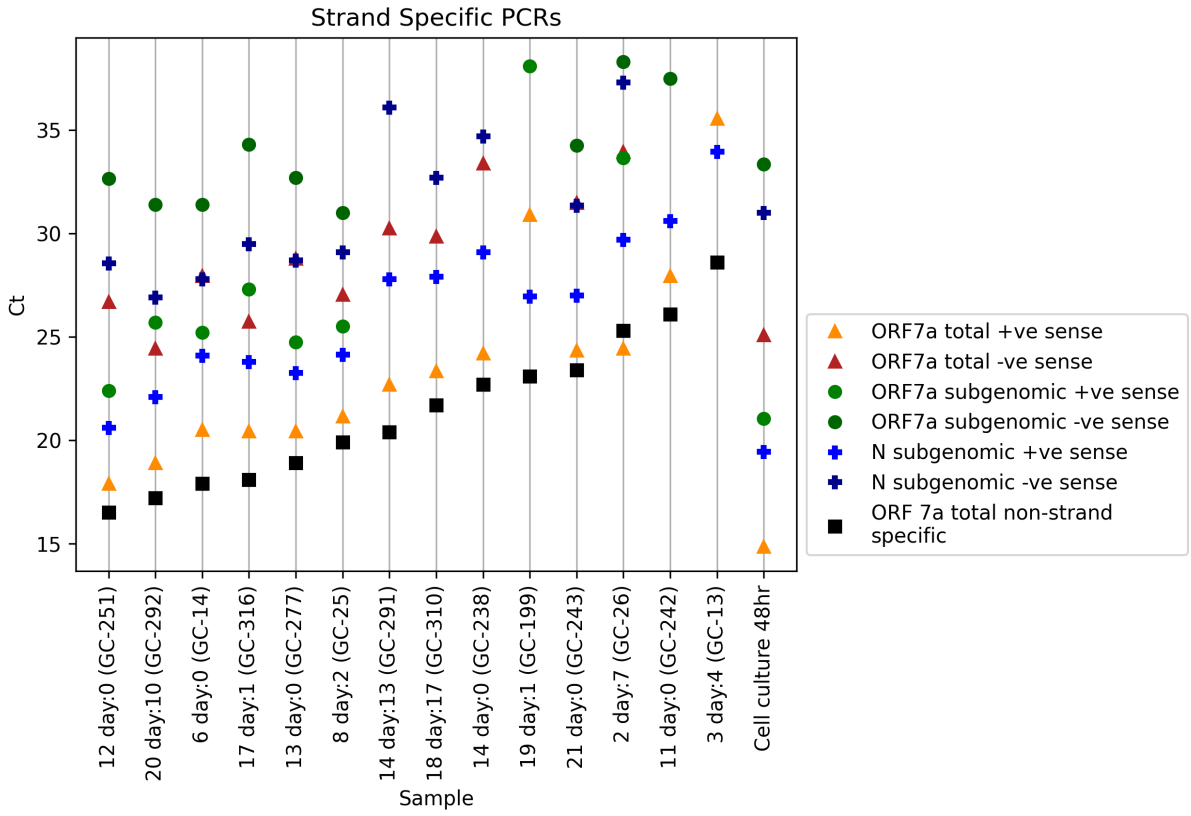


Figure 6. Relative percentage of each subgenomic RNA of all reads mapped to the SARS-CoV-2 subgenomic RNA for the Ampliseq mini-panel 1 in naso-oropharyngeal swab samples with 31 PCR amplification cycles. Subgenomic ORF7a and N gene RNAs were consistently the two most abundant subgenomic RNAs present in the samples. Subgenomic ORF7b and ORF10 were not detected in any sample. (n=12 from 12 biological samples run once)

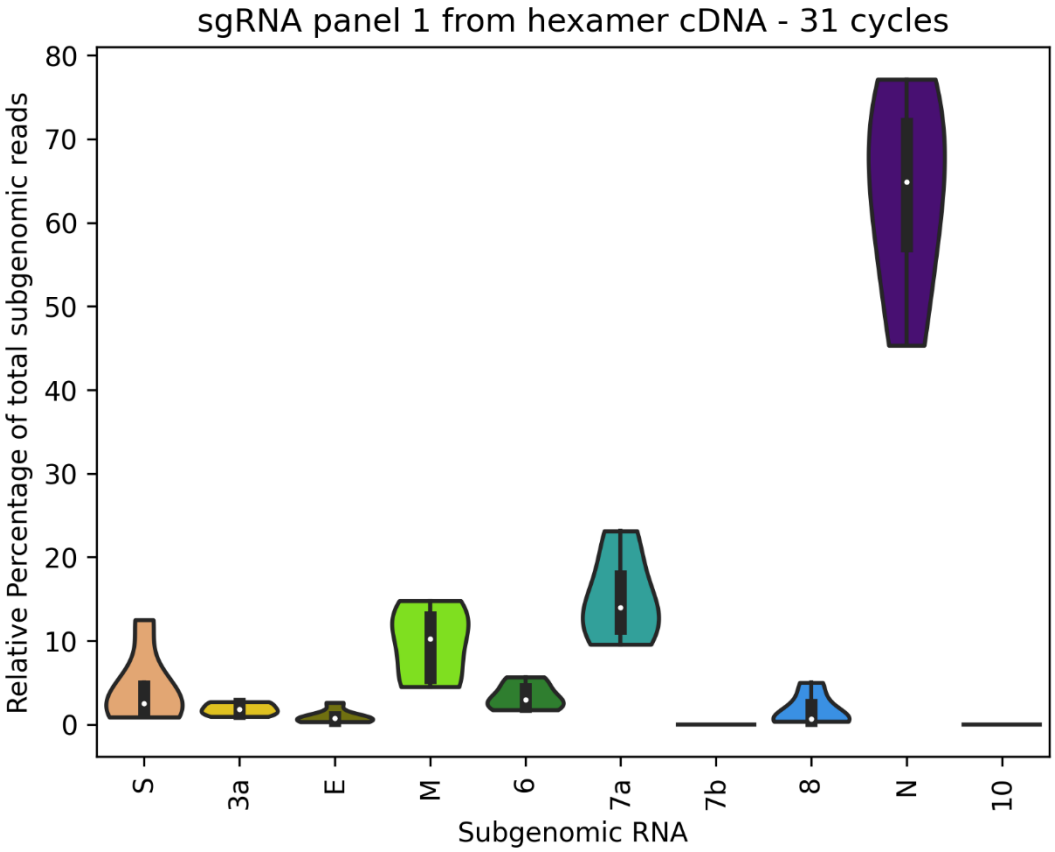


Figure 7: Summary heat map and hierarchical clustering of the ranked relative expression the 469 inflammatory genes in the SARS-CoV-2 and negative control naso-oropharyngeal swab samples. Full data is presented in Supplementary Data 2. (n=28 from 28 biological samples run once)

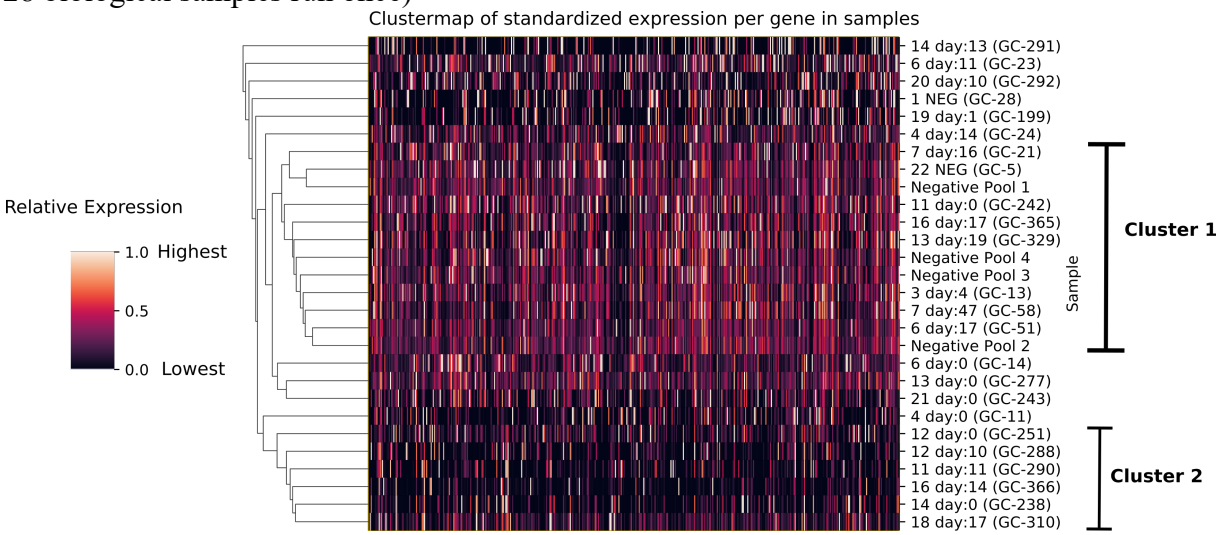


Figure 8: The number of inflammatory amplicons per sample with an above average number of reads mapped plotted against the day the swab was collected post symptom onset. The average number of reads for each amplicon was calculated from all samples. The average number of amplicons with above average numbers of reads mapped in the negative samples (239 - excluding individual 1 (GC-28)) is shown as a red line. (n=22 from 22 biological samples run once)

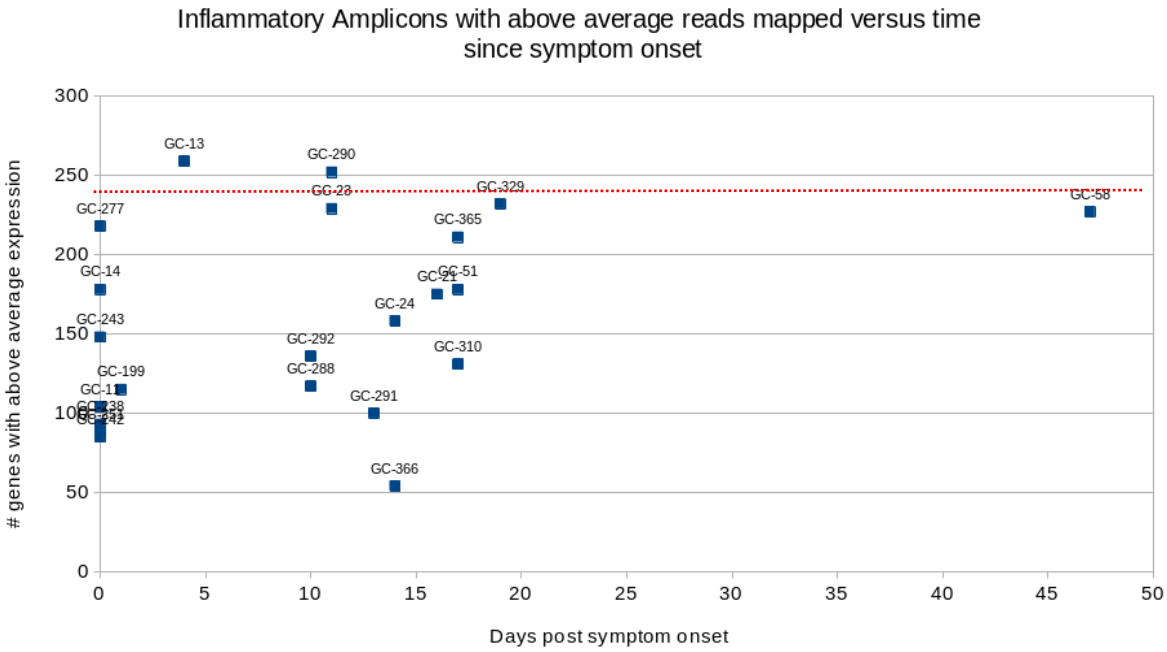
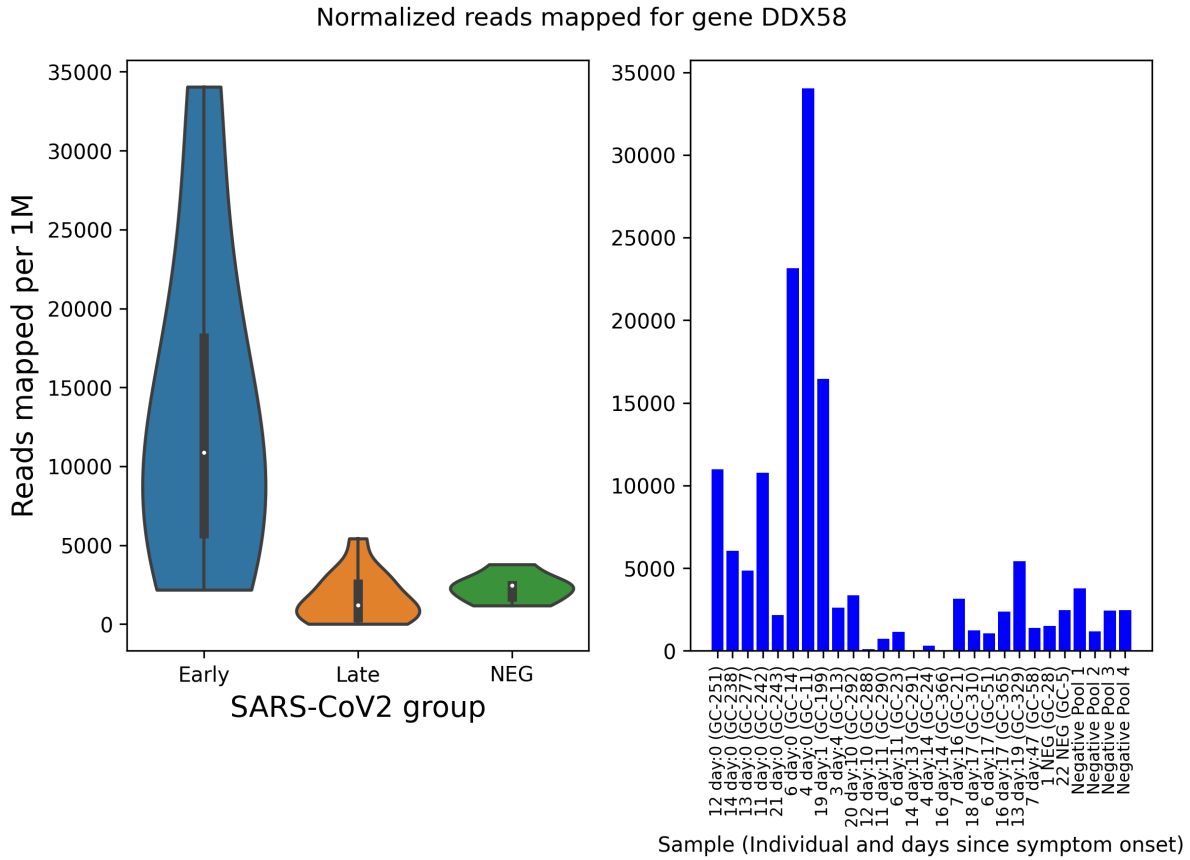
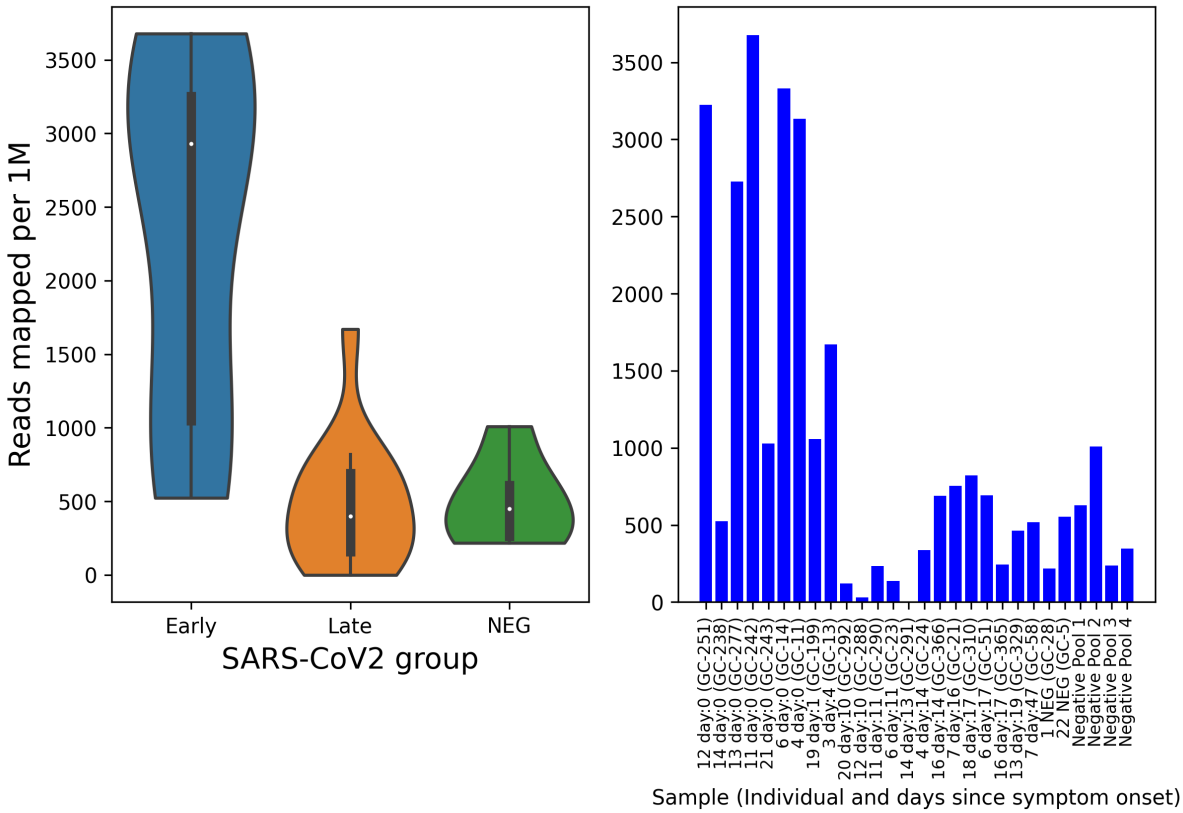


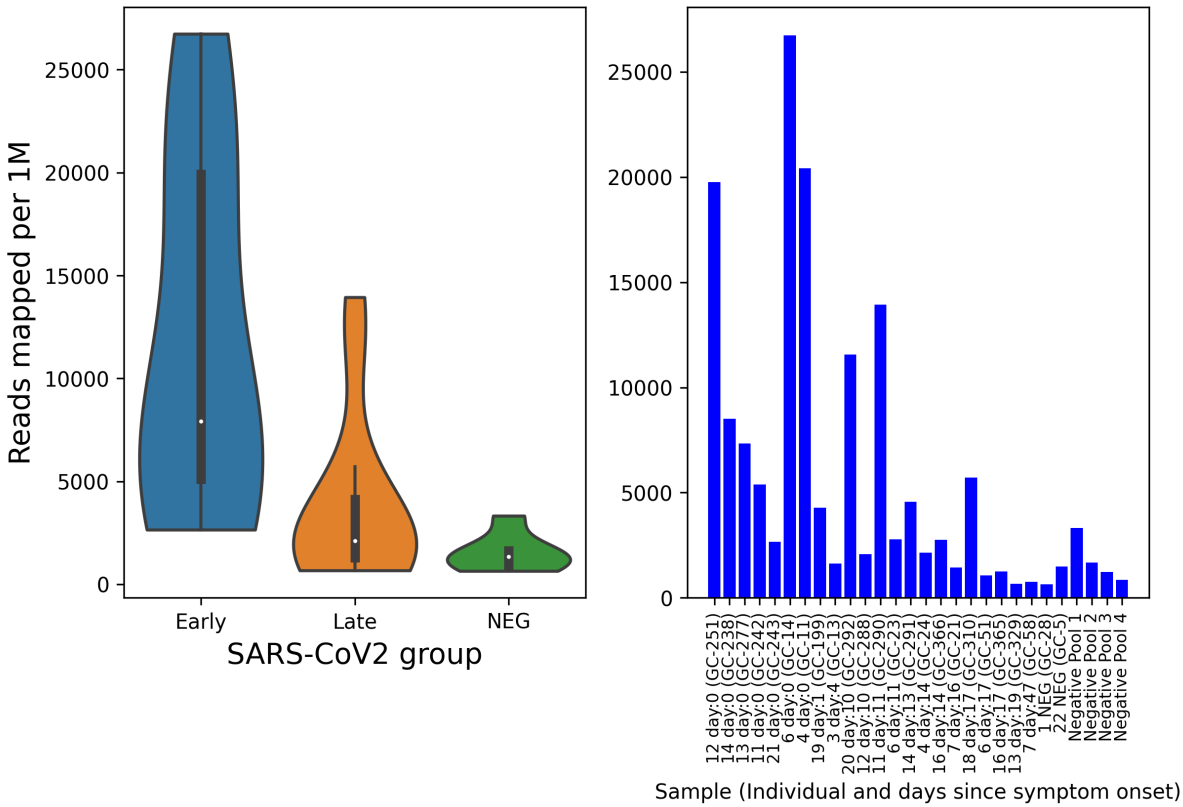
Figure 9: Patterns of expression in the seven statistically significant gene targets identified in the Wilcoxon rank-sum comparison of SARS-CoV-2 negative and early clinical (0-1 days). A: DExD/H-Box helicase (DDX58), B: interferon induced with helicase C Domain containing protein 1 (IFIH1), C: interferon regulatory factor 7 (IRF7), D: N-Myc interactor (NMI), E: tumour necrosis factor ligand superfamily member 10 (TNFSF10). F: Complement C3a Receptor (C3AR1). G: formyl peptide receptor 1 (FPR-1). (n=28 from 28 biological samples run once)



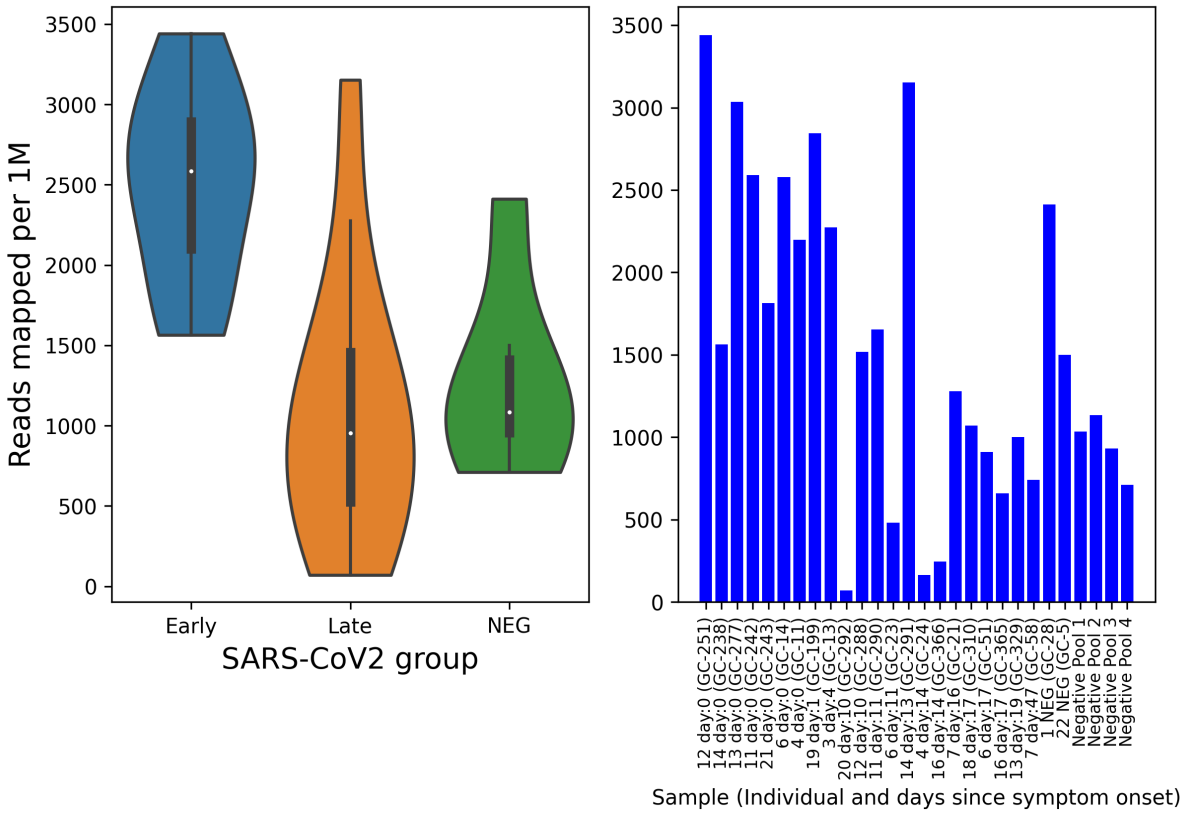
Normalized reads mapped for gene IFIH1



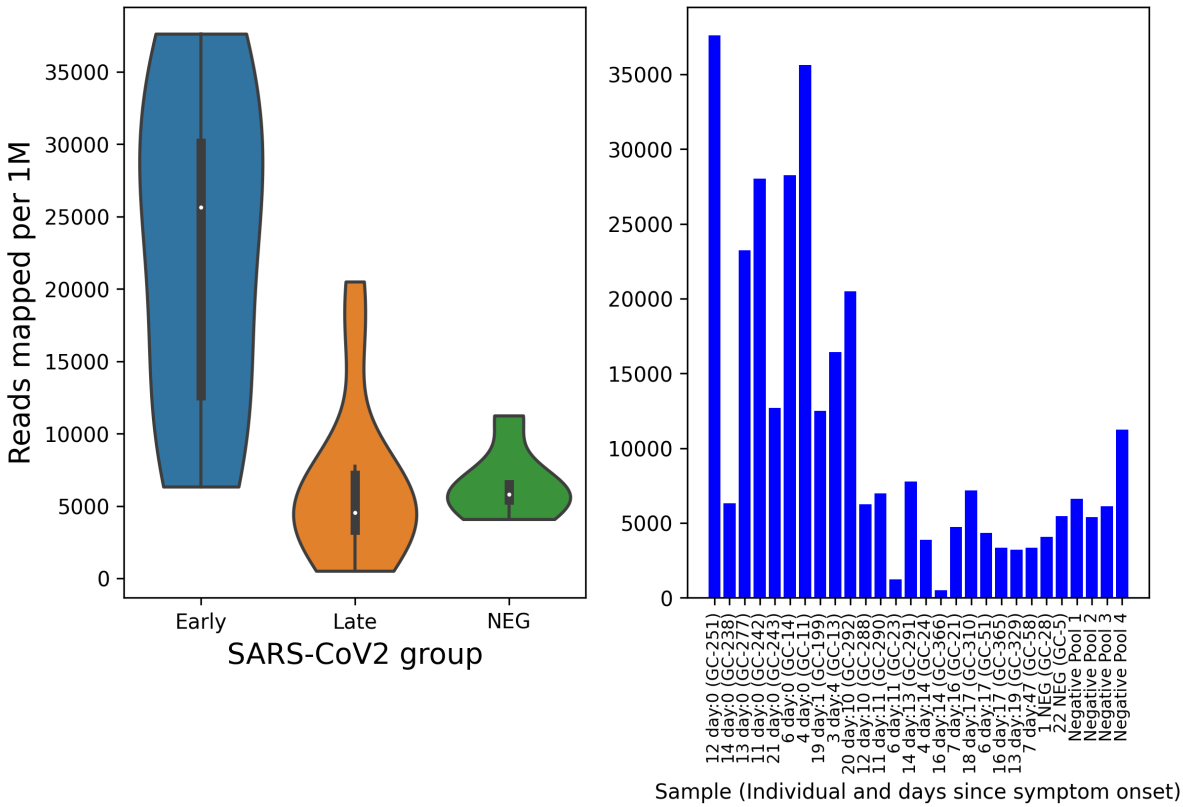
Normalized reads mapped for gene IRF7



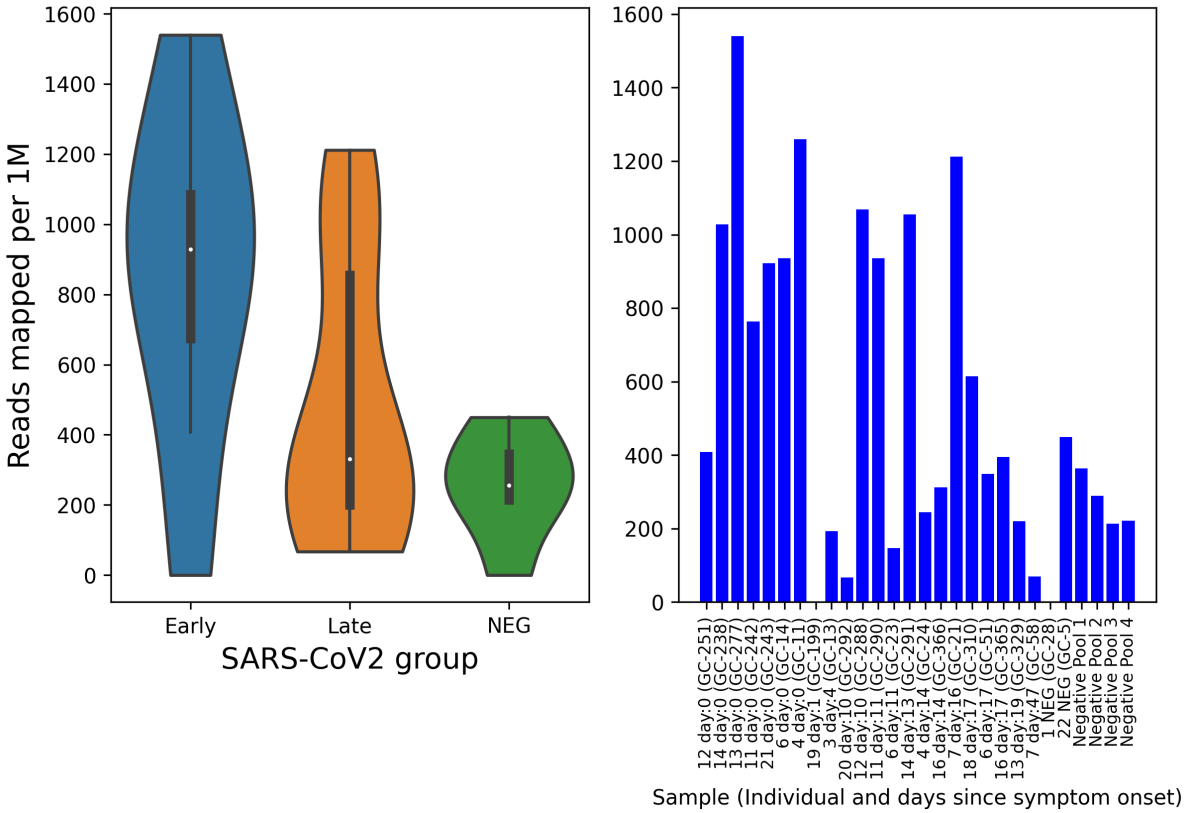
Normalized reads mapped for gene NMI



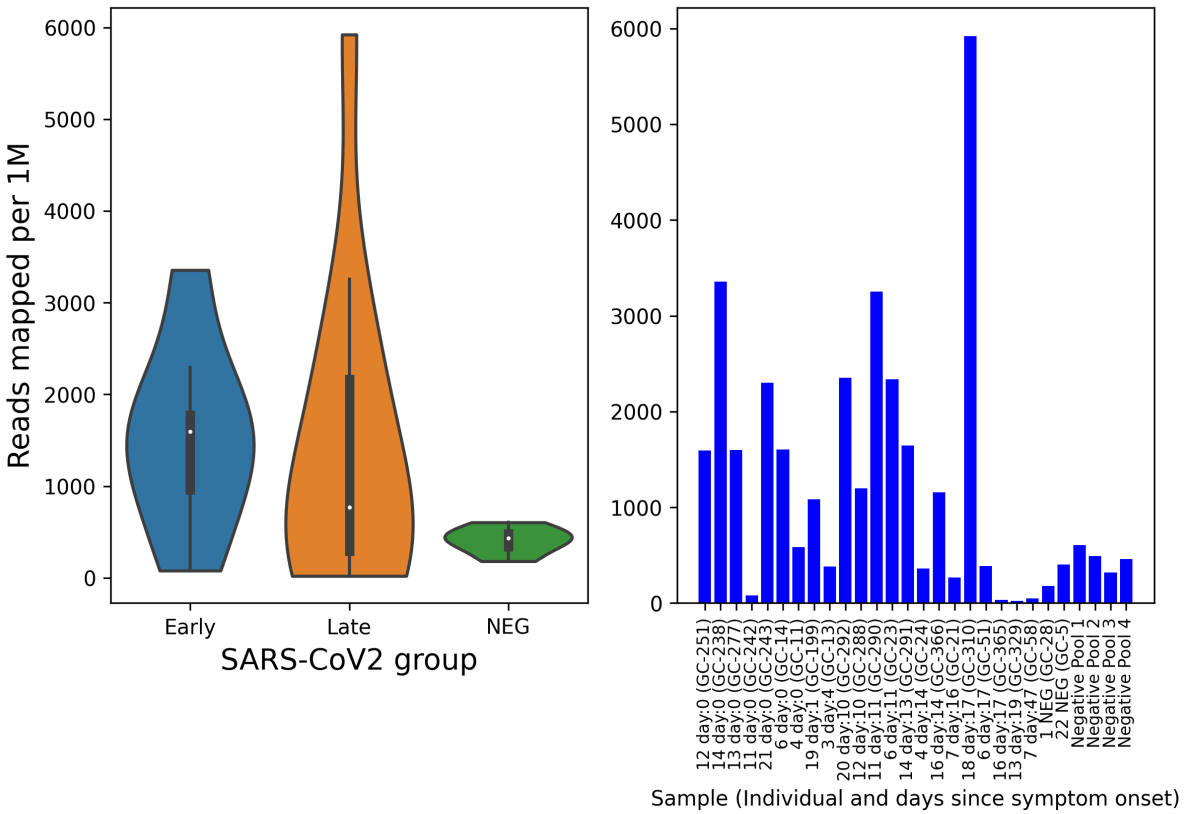
Normalized reads mapped for gene TNFSF10



Normalized reads mapped for gene C3AR1



Normalized reads mapped for gene FPR1



Supplementary Materials: The following is available online; Supplementary Information file 1 including Supplementary Tables 1-2 and Supplementary Figures 1-13. Supplementary Data 1 (1 Figure), Supplementary Data 2 (1 Figure) and Supplementary Data 3 (1 Figure).