

1 SARS-CoV-2 Total and Subgenomic RNA Viral Load in Hospitalized Patients

2

3 Derek E. Dimcheff<sup>1</sup>, Andrew L. Valesano<sup>2,3</sup>, Kalee E. Rumpf<sup>2,3,4</sup>, William J. Fitzsimmons<sup>2,3</sup>,  
4 Christopher Blair<sup>2,3</sup>, Carmen Mirabelli<sup>3</sup>, Joshua G. Petrie<sup>4</sup>, Emily T. Martin<sup>4</sup>, Chandan  
5 Bhambhani<sup>5</sup>, Muneesh Tewari<sup>5</sup>, Adam S. Luring<sup>2,3\*</sup>

6

7 **Author affiliations:**

8 <sup>1</sup>Division of Hospital Medicine, Department of Internal Medicine, University of Michigan, Ann  
9 Arbor, MI, USA; <sup>2</sup>Division of Infectious Disease, Department of Internal Medicine, University of  
10 Michigan, Ann Arbor, MI, USA; <sup>3</sup>Department of Microbiology and Immunology, University of  
11 Michigan, Ann Arbor, MI, USA; <sup>4</sup>Department of Epidemiology, School of Public Health University  
12 of Michigan, Ann Arbor, MI, USA; <sup>5</sup>Division of Hematology and Oncology, Department of Internal  
13 Medicine, University of Michigan, Ann Arbor, MI, USA

14

15 **\*Corresponding Author:**

16 Adam Luring  
17 1150 W. Medical Center Dr.  
18 MSRB1 5510B  
19 aluring@med.umich.edu

20 **Abstract: 264 2 words**

21 **Text: 3202 words**

22 **Tables: 0**

23 **Figures: 4**

24 **Supplementary figures:2**

25

26

## Abstract

27           Understanding viral load in patients infected with SARS-CoV-2 is critical to epidemiology  
28 and infection control. Previous studies have demonstrated that SARS-CoV-2 RNA can be  
29 detected for many weeks after symptom onset. The clinical significance of this finding is unclear  
30 and, in most patients, likely does not represent active infection. There are, however, patients  
31 who shed infectious virus for weeks. Detection of subgenomic RNA transcripts expressed by  
32 SARS-CoV-2 has been proposed to represent productive infection and may be a tractable  
33 marker for monitoring infectivity. Here, we use RT-PCR to quantify total and subgenomic  
34 nucleocapsid (N) and envelope (E) transcripts in 190 SARS-CoV-2 positive samples collected  
35 on hospital admission. We relate these findings to duration of symptoms. We find that all  
36 transcripts decline at the same rate; however, subgenomic E becomes undetectable before  
37 other transcripts. In Kaplan-Meier analysis the median duration of symptoms to a negative test  
38 is 14 days for sgE and 25 days for sgN. There is a linear decline in subgenomic RNA compared  
39 to total RNA suggesting subgenomic transcript copy number is highly dependent on copy  
40 number of total transcripts. The mean difference between total N and subgenomic N is 16-fold  
41 (4.0 cycles) and the mean difference between total E and sub-genomic E is 137-fold (7.1  
42 cycles). This relationship is constant over duration of symptoms allowing prediction of  
43 subgenomic copy number from total copy number. Although Subgenomic E is undetectable at a  
44 time that may more closely reflect the duration of infectivity, its utility in determining active  
45 infection may be no more useful than a copy number threshold determined for total transcripts.

46

## 47 Introduction

48           The emergence of the severe acute respiratory syndrome virus-2 (SARS-CoV-2) led to  
49 the rapid development of diagnostic tests for infection. Many of these tests rely on reverse

50 transcriptase polymerase chain reaction (RT-PCR) to detect total viral RNA. These assays are  
51 highly sensitive and specific, with a limit of detection of approximately 5-500 copies of viral RNA  
52 per reaction (Vogels et al. 2020). Unfortunately, the relationship between a positive RT-PCR  
53 and viral infectivity is unclear, particularly as infection progresses over time. While the median  
54 duration of detection of SARS-CoV-2 viral RNA in the upper respiratory tract is roughly 14.5  
55 days from symptom onset, many studies have shown that SARS-CoV-2 RNA can be detected  
56 for several weeks, long past when most people are infectious (Walsh et al. 2020).

57         Because of persistent test positivity, the CDC no longer recommends a test-based  
58 strategy for discontinuing isolation precautions. Current guidelines recommend that patients with  
59 mild to moderate disease remain isolated for 10 days from symptom onset, while those with  
60 either severe disease or an immunocompromising condition remain isolated for 10-20 days from  
61 symptom onset. These guidelines are largely supported by studies of viral infectivity in clinical  
62 samples over time. (Arons et al. 2020, Wölfel et al. 2020, Bullard et al. 2020) However, recent  
63 studies have demonstrated that some immunosuppressed patients may shed infectious virus for  
64 weeks, regardless of symptoms (Avanzato et al. 2020, Baang et al. 2020, Choi et al. 2020).  
65 Symptom-based strategies for isolation may also be problematic in patients incidentally found to  
66 be positive for SARS-CoV-2.

67         There is no high-throughput, rapid test that distinguishes those who are infectious from  
68 those who are not. While viral culture is perhaps the most reliable way to determine infectivity, it  
69 is neither timely nor practical in most clinical laboratories. There has been intense interest in  
70 whether other molecular markers can be used as correlates of infectivity, including the detection  
71 of subgenomic RNA. SARS-CoV-2 has a positive-sense, single-stranded RNA genome of nearly  
72 30 kb. Negative-sense RNA intermediates serve as templates for the synthesis of positive-  
73 sense genomic RNA. The viral polymerase also makes subgenomic messages that have a  
74 common 5' leader fused to downstream open reading frames, which are then translated into nine

75 proteins, including the nucleocapsid (N), spike (S), envelope (E) and membrane (M), and six  
76 accessory proteins (3a,6,7a,7b and 8) (Kim et al. 2020). Given this genomic arrangement,  
77 subgenomic RNA can be distinguished from total RNA by placement of alternate PCR primers.  
78 Because subgenomic transcripts are not packaged into virions, their presence is thought to  
79 indicate productively infected cells.

80 A few studies have evaluated the kinetics of subgenomic RNA during the course of  
81 infection. A small study of 9 patients found that no subgenomic RNA could be found in the throat  
82 up to 5 days post symptom onset. (Wölfel et al. 2020) A larger study of 35 patients showed that  
83 subgenomic RNA and culturable virus were not detectable after 8 days from symptom onset in  
84 mild COVID-19 disease, but total SARS-CoV-2 transcripts persisted for weeks (Perera et al.  
85 2020). Less is known about subgenomic RNA kinetics in hospitalized patients with severe  
86 disease and in immunosuppressed patients.

87 To determine the utility of subgenomic RNA as a correlate of SARS-CoV-2 infectivity, we  
88 used RT-PCR to amplify the total and subgenomic nucleoprotein gene (N) and envelope gene  
89 (E) from 190 individual patient samples collected on admission to the hospital. We show that  
90 subgenomic transcripts become undetectable before total transcripts when evaluated in relation  
91 to duration of symptoms. Because expression levels of total and subgenomic transcripts are  
92 highly correlated, the added benefit of measuring subgenomic RNA is limited.

93

## 94 **Materials and Methods**

### 95 **Samples and participants**

96 Nasopharyngeal (NP) samples were obtained from patients admitted to the University of  
97 Michigan between March 13, 2020 to June 10, 2020. Dacron swabs were placed in 3 ml viral  
98 transport media and transported to the clinical microbiology laboratory for testing. Residual

99 samples were stored in a central biorepository at -80°C. In total, 185 patients with RT-PCR  
100 confirmed SARS-CoV-2 and clinical information were included in our study. An additional five  
101 patients without clinical data were included for a total of 190 individual NP swabs. Clinical data,  
102 including date of symptom onset, basic demographic data, and outcome were abstracted from  
103 clinician notes. This study protocol was reviewed and approved by the University of Michigan  
104 Institutional Review Board (IRB). The IRB determined the study to be exempt from requirement  
105 for informed consent given retrospective nature of this study and this use of stored  
106 biospecimens and deidentified data.

### 107 **Quantitative RT-PCR**

108 Residual samples from NP swabs were centrifuged at 1200 x g and 200 µl of sample  
109 used for RNA extraction. RNA was extracted with the Invitrogen PureLink Pro 96 Viral  
110 RNA/DNA Purification Kit (Invitrogen, Carlsbad, CA). Samples were eluted in 100 µl and stored  
111 at -80C.

112 Amplification of total and subgenomic transcripts for nucleocapsid (N) and envelope (E)  
113 genes was performed using conditions outlined in the CDC 2019-Novel Coronavirus EUA  
114 protocol (<https://www.fda.gov/media/134922/download>). We diluted our clinical samples 1:5 in  
115 nuclease free H<sub>2</sub>O and used 5 µl of this dilution for 1 µl equivalent total RNA per reaction. This  
116 was done to preserve clinical sample, and the dilution was accounted for in copy number  
117 calculations. Reactions were preformed using Taqpath 1-step RT-qPCR master mix  
118 (Thermofisher, Waltham, MA), 500 nM of each primer, and 250 nM of each probe in a total  
119 reaction volume of 20 µl. Cycling conditions were as follows: activation of uracil-N-glycosylase  
120 for 2 min at 25° C, reverse transcription for 15 min at 50° C, denaturation for 2 min at 95° C, and  
121 45 cycles of 3 s at 95° C, 30 s at 55° C on an Applied Biosystems 7500 FAST real-time PCR  
122 system. Cycle threshold (Ct) was determined uniformly across PCR runs.

123 Subgenomic transcripts were amplified by substituting subgenomic leader sequence  
124 sgLeadSARSCoV2-F: 5'-CGATCTCTTGATAGATCTGTTCTC-3' (Wölfel et al. 2020) as the  
125 forward primer for E or N together with the reverse primers and probes for N and E genes. For  
126 total E, the primers and probes were from the Charite/Berlin protocol (Corman et al. 2020) and  
127 the sequences are as follows: E\_Sarbeco\_F: ACAGGTACGTTAATAGTTAATAGCGT;  
128 E\_Sarbeco\_R: 5-ATATTGCAGCAGTACGCACACA-3; E\_Sarbeco\_P1: FAM-  
129 AACTAGCCATCCTTACTGCGCTTCG-3IABkFQ. The N gene was amplified using the CDC  
130 N1 primer and probe set as follows: 2019-nCoV\_N1 Forward Primer GACCCCAA  
131 ATCAGCGAAAT; 2019-nCoV\_N1 Reverse Primer TCTGGTACTGCCAGTTGAATCTG; 2019-  
132 nCoV\_N1 Probe ACCCCGCATTACGTTTGGTGG ACC. Probe sequences were FAM labeled  
133 with Iowa Black quencher (Integrated DNA Technologies, Coralville, IA).

#### 134 **Droplet Digital PCR**

135 Absolute copy numbers of N, E, sgN and sgE were determined by droplet digital PCR  
136 (ddPCR) using SARS-CoV-2 RNA as template from HuH-7 infected cells. HuH-7 cells were  
137 infected with SARS-CoV-2 strain WA-1. RNA was harvested at 48 hours post infection using  
138 trizol and Zymo columns. ddPCR was performed using the QX200 Droplet Digital PCR System  
139 (Bio-Rad, Hercules, CA). Sample reaction contained HuH-7 RNA/sgRNA template (5.5 ul), the  
140 respective forward and reverse primers (final concentration 900nM each) and the respective  
141 FAM labeled probe (final concentration 250nM). The other reaction components common to all  
142 reactions were added at the final concentration as per the manufacturer's recommendation (Bio-  
143 Rad # 1864021; 1X Supermix, 20 units/ul Reverse Transcriptase and 15 mM DTT). Each 20 ul  
144 reaction mix was partitioned into droplets using the QX200 droplet generator (Bio-Rad),  
145 transferred into a 96-well plate, sealed and cycled in a C1000 Thermal Cycler (Bio-Rad). The  
146 cycling conditions used were as per the manufacturer's recommendation (Bio-Rad's expert  
147 design assay for SARS-CoV-2 # dEXD28563542; hold at 25°C for 3 min, reverse transcription

148 at 50°C for 60 min, enzyme activation for 10 minutes at 95°C for 1 cycle, denaturation at 94°C  
149 for 30 seconds and annealing/extension at 60°C for 60 seconds for 40 cycles, enzyme  
150 deactivation for 10 minutes at 98°C for 1 cycle followed by a hold at 4°C). All steps were  
151 performed with a 2°C/second ramp rate and the lid temperature set at 105°C. Droplets were  
152 read using QuantaSoft Software in the QX200 reader (Bio-Rad). Once copy number was  
153 determined, a ten-fold dilution series (from  $1 \times 10^7$  copies to 10 copies per 20ul reaction) of this  
154 sample was performed to generate standard curves and used to determine copy number of  
155 clinical samples. Copy number was corrected for dilution of sample to reflect copies per milliliter  
156 of viral transport media from original NP swab sample.

## 157 **Statistical Analyses**

158 Continuous variables were expressed as medians and interquartile ranges (IQRs) or  
159 means and standard deviations. Simple linear regression and Kaplan-Meier analysis were  
160 performed using SPSS software version 27 (IBM Corporation) and GraphPad (Prism). Statistical  
161 significance was set as  $P < 0.05$ .

162

## 163 **RESULTS**

### 164 **Patient characteristics**

165 We obtained NP swabs from 190 SARS-CoV-2 positive patients admitted to the  
166 University of Michigan Hospital from March 13, 2020 to June 10, 2020. We were able to obtain  
167 clinical information from 185 of the 190 patients. In our sample, 56.8% were men (105/185). The  
168 median age was 64 (IQR: 53-75) years old. The median days from symptom onset to SARS-  
169 CoV-2 testing was 5 days (IQR 3-8). Three patients reported no symptoms consistent with  
170 COVID-19 infection on admission but reported symptoms later in their hospital course.

## 171 **Symptom duration compared to total and subgenomic transcripts**

172 We found that there was no difference in the rate of decline of SARS-CoV-2 transcripts  
173 when compared to symptom duration with all slopes equal ( $p>0.05$ ) (Figure 1). We did however  
174 find a significant difference in the X-intercepts for the regression line of viral RNA transcript  
175 decline over time. There was no difference in intercepts between total E and N ( $p>0.72$ ). When  
176 total N was compared to sgN the X-intercepts were significantly different ( $P<0.0001$ ) as was the  
177 comparison of E to sgE ( $P<0.0001$ ). Finally, there was a significant difference between  
178 intercepts for subgenomic transcripts sgN and sgE ( $P<0.001$ ). The differences X-intercepts  
179 reflects relative differences in days until RNA is no longer detectable based on a threshold of 40  
180 cycles.

181 We used a Kaplan-Meier analysis to further investigate the relationship between  
182 symptom duration and a negative subgenomic RNA test. The median duration from symptom  
183 onset to a negative sgN RT-PCR was 25 days (95% CI: 19.9-30.1) and 14 days for sgE (95% CI  
184 9.6-18.4 days). This difference was statistically significant ( $P=0.001$ ) (Figure 2). All 185 patients  
185 were positive for total RNA regardless of symptom duration. By  $\geq 15$  days from symptom onset,  
186 only 14% of patients were positive for sgE compared to 35.7% for sgN.

187 Recent studies have demonstrated that the total copy number threshold that correlates  
188 with a loss of culture infectivity is between 5.5 and 6.5  $\log_{10}$  copies per ml (Wölfel et al. 2020,  
189 Bullard et al. 2020, van Kampen et al. 2021). Primers used in this study amplify total RNA  
190 (genomic plus subgenomic RNA), and by pairing the subgenomic leader primer with reverse E  
191 and N primers distinguish subgenomic from total RNA. We determined copy number in clinical  
192 samples using standard curves generated with known copy number from ddPCR experiments  
193 for total N, total E, sgE and sgN. At 6.5  $\log_{10}$  total E copies, 47% (56/119) of samples were  
194 negative for sgE. At a total N copy number of 6.5  $\log_{10}$ , 24% (28/119) were negative for sgN.

195 There were no samples negative for subgenomic at a copy number greater than  $6.5 \log_{10}$  total  
196 copies per ml (Supplementary Figure 1).

### 197 **Relationship of total to subgenomic transcripts**

198 To further understand the relationship between total and subgenomic transcripts for the  
199 N and E genes, we compared Ct values from 190 clinical samples. We found that 162 of 190  
200 (85%) patients had both a detectable N and sgN, and 132 (69%) patients had both a detectable  
201 E and sgE. Subgenomic transcripts declined linearly along with total RNA transcripts  
202 (Supplementary Figure 2). There was no difference in the slope of subgenomic to total RNA  
203 regression (linear regression, slope of N vs sgN -0.99 and slope of E vs sgE -1.03  $P=0.09$ ). We  
204 were unable to detect sgE RNA transcripts once total E reached 32 cycles, and were unable to  
205 detect sgN transcripts once total N reached 35 cycles.

206 We found there was very little difference between expression levels of total N and E. The  
207 mean Ct value of E was 25.76 (STD 5.69) and N was 25.6 (STD 5.6 cycles). In cell culture and  
208 in clinical samples, we found that total RNA levels were higher than subgenomic RNA levels.  
209 Total N was expressed at a 4.5-fold (2.2, STD 0.24 cycles) higher level than sgN, and E was  
210 expressed at a 13.9-fold (3.8, STD 0.2 cycles) higher level compared to sgE in HuH-7 cells at  
211 48 hours post infection (Figure 3). This difference was more dramatic in clinical samples. Total  
212 N was expressed at a 16-fold (4.0, STD 1.1 cycles) higher level than sgN transcripts and E was  
213 expressed at a 137.2-fold (7.1, STD 1.3 cycles) higher level than sgE when all clinical samples  
214 were compared together (Figure 3). This relationship remained unchanged throughout the  
215 course of infection (Figure 3). Furthermore, the large fold difference in subgenomic compared to  
216 total transcripts indicates that subgenomic RNA contributes little to the overall signal of total  
217 transcripts.

### 218 **Subgenomic RNA in a persistently infected patient**

219 Emerging data indicates that some immunosuppressed patients can be persistently  
220 positive for SARS-CoV-2 by RT-PCR and shed infectious virus for weeks. We evaluated total  
221 and subgenomic RNA from a patient persistently infected with SARS-CoV-2 for 119 days. This  
222 patient showed continued expression of total and subgenomic transcripts throughout this time.  
223 The mean fold difference of total N to subgenomic N was 11.3-fold (3.5 cycles) and the mean  
224 fold difference of total E to subgenomic E was 84-fold (6.4 cycles) (Figure 4). This ratio was  
225 maintained over the 119 day course.

## 226 **DISCUSSION**

227 In this study, we found that subgenomic transcripts become undetectable at an earlier  
228 timepoint after infection than total transcripts. The median duration of symptoms until a negative  
229 RT-PCR test was 14 days for subgenomic E and 25 days for subgenomic N. Total RNA was  
230 positive at all time points in our study up to 28 days. We found a strong correlation between total  
231 and subgenomic transcript levels. They decline at the same rate and have a fixed ratio at all  
232 time points post symptom onset. The predictability of subgenomic copy number from total copy  
233 number indicates that detection of subgenomic RNA does not add additional information about  
234 infectivity that cannot be gained from total copy number alone.

235 A recent study in hospitalized patients showed that fewer than 5% of patients are  
236 infectious 15 days after the onset of symptoms (van Kampen et al. 2021). We find that  
237 subgenomic E becomes undetectable at a median of 14 days and only 14% (2/14) of patients  
238 were positive for subgenomic E after this time point. Our results suggest that the disappearance  
239 of subgenomic E RNA may correlate with a time when patients are no longer infectious.

240 Despite our finding that a negative subgenomic E RT-PCR seems to correlate with a  
241 time when patients are not infectious, the supposition that any negative subgenomic RNA  
242 transcript can be used to predict infectivity is problematic. In our analysis, the median number of

243 days from symptom onset to a negative subgenomic N RT-PCR was 25 days. Using  
244 subgenomic N as a marker for infectivity would significantly prolong isolation and not predict  
245 infectivity. This difference is likely due to the higher levels of expression of subgenomic N  
246 relative to subgenomic E.

247 A negative subgenomic RNA test likely indicates that patients are not infectious simply  
248 because a negative subgenomic RNA test correlates with a low total RNA copy number.  
249 Previous studies have shown a copy number below  $6.5 \log_{10}$  copies per ml is a threshold for  
250 culturable virus. In fact, we found that all negative subgenomic transcripts were in samples with  
251 total copy number below this threshold. However, we also find many samples that are below  $6.5$   
252  $\log_{10}$  copies per ml, but still express subgenomic E and N. This is in line with a recent study that  
253 showed subgenomic RNA could be detected from samples where infectivity in cell culture was  
254 negative (van Kampen et al. 2021). It is likely that subgenomic RNA also has a copy number  
255 threshold below which infectious virus cannot be isolated. Because of the strong correlation in  
256 copy number of total and subgenomic transcripts throughout the course of infection and in a  
257 persistently infected patient, we do not believe that subgenomic RNA provides additional  
258 information not gained by total copy number.

259 Previous research suggested that subgenomic RNA is a suitable marker for active  
260 infection because it degrades more rapidly than total RNA. One study inoculated animals with  
261 infectious and inactivated SARS-CoV-2, which had high levels of total and subgenomic RNA.  
262 Total RNA copy number on day 1 post infection from inactivated virus ranged from roughly 4  
263  $\log_{10}$  to 6  $\log_{10}$  copies per ml but no subgenomic E RNA was detected. This was interpreted to  
264 mean that subgenomic RNA rapidly degrades without replication, but total RNA does not.  
265 (Speranza et al. 2020) In our study, at a total copy number of 4  $\log_{10}$  no subgenomic E could be  
266 detected, and at less than 6  $\log_{10}$  per ml of total RNA 47% (56/119) of samples were negative  
267 for subgenomic E. We suspect this finding is not due to differential degradation, but rather to

268 differential detection of transcripts. Furthermore, if there was a difference in rates of  
269 degradation, we would expect to see the ratio of total RNA to subgenomic RNA increase  
270 overtime, which we did not see.

271 Although there is emerging data supporting isolation guidelines in mild/moderate and  
272 severe COVID-19 infection, little is known about infectivity in immunosuppressed patients. In  
273 this study we showed persistent total and subgenomic RNA expression up to 119 days with a  
274 previous study demonstrating infectious virus could be isolated through this course (Baang et al.  
275 2020). Because low levels of subgenomic RNA can be detected in the absence of  
276 infectiousness, we do not endorse using subgenomic RNA to determine infectivity in these  
277 patients. As with non-immunosuppressed patients, a copy number threshold of total RNA would  
278 likely provide the same information which can be gained from copy number on total RNA.

279 Our study is limited by the fact that we did not attempt to isolate virus in culture from  
280 clinical samples and correlate to total or subgenomic copy number. We did however investigate  
281 correlation of transcripts to symptom duration, which has been used as a benchmark for  
282 infectivity. We are aware that recall bias may affect reported symptom duration particularly in ill  
283 hospitalized patients. This is mitigated to some degree by our large sample size. Differences in  
284 copy number observed could be the result of differences in efficiency of RT-PCR reactions  
285 across total and subgenomic transcripts. We confirmed differences in copy number using  
286 absolute copy number determined using standard curves generated with a known copy number  
287 using ddPCR for N, E, sgE and sgN. Furthermore our findings are in agreement with direct RNA  
288 sequencing studies showing the subgenomic N is expressed at higher levels than subgenomic  
289 E. (Kim et al. 2020)

290 In this study of inpatients infected with SARS-CoV-2, we find that subgenomic E  
291 transcripts become undetectable before other transcripts largely because it is transcribed at far  
292 lower levels than other transcripts studied here. There is a fixed relationship between copy

293 number of total and subgenomic transcripts over the course of infection. Because of this  
294 consistent relationship between total and subgenomic RNA, we do not believe subgenomic RNA  
295 detection adds additional information regarding infectivity that cannot be gained from copy  
296 number threshold validated to infectivity determined by cell culture.

297

## 298 **ACKNOWLEDGEMENTS**

299 We thank the University of Michigan Clinical Microbiology Laboratory and the University of  
300 Michigan Central Biorepository for their assistance in providing samples. This work was  
301 supported by a University of Michigan COVID-19 Response Innovation Grant (to ASL),  
302 K01AI141579 (to JGP) and CDC U01 IP000974 (to ETM)

## 303 **CONFLICTS OF INTEREST**

304 ASL reports receiving consulting fees from Sanofi on antiviral drugs and is a paid member of a  
305 steering committee for a clinical trial of baloxavir (Roche).

306

## 307 **REFERENCES**

308 Arons, M. M., K. M. Hatfield, S. C. Reddy, A. Kimball, A. James, J. R. Jacobs, J. Taylor, K.  
309 Spicer, A. C. Bardossy, L. P. Oakley, S. Tanwar, J. W. Dyal, J. Harney, Z. Chisty, J. M.  
310 Bell, M. Methner, P. Paul, C. M. Carlson, H. P. McLaughlin, N. Thornburg, S. Tong, A.  
311 Tamin, Y. Tao, A. Uehara, J. Harcourt, S. Clark, C. Brostrom-Smith, L. C. Page, M. Kay,  
312 J. Lewis, P. Montgomery, N. D. Stone, T. A. Clark, M. A. Honein, J. S. Duchin, J. A.  
313 Jernigan & P. H. S. a. K. C. a. C. C.-I. Team (2020) Presymptomatic SARS-CoV-2  
314 Infections and Transmission in a Skilled Nursing Facility. *N Engl J Med*, 382, 2081-2090.

315 Avanzato, V. A., M. J. Matson, S. N. Seifert, R. Pryce, B. N. Williamson, S. L. Anzick, K.  
316 Barbian, S. D. Judson, E. R. Fischer, C. Martens, T. A. Bowden, E. de Wit, F. X. Riedo &  
317 V. J. Munster (2020) Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an  
318 Asymptomatic Immunocompromised Individual with Cancer. *Cell*, 183, 1901-1912.e9.

319 Baang, J. H., C. Smith, C. Mirabelli, A. L. Valesano, D. M. Manthei, M. Bachman, C. E. Wobus,  
320 M. Adams, L. Washer, E. T. Martin & A. S. Luring (2020) Prolonged SARS-CoV-2  
321 replication in an immunocompromised patient. *J Infect Dis*.

322 Bullard, J., K. Dust, D. Funk, J. E. Strong, D. Alexander, L. Garnett, C. Boodman, A. Bello, A.  
323 Hedley, Z. Schiffman, K. Doan, N. Bastien, Y. Li, P. G. Van Caesele & G. Poliquin  
324 (2020) Predicting infectious SARS-CoV-2 from diagnostic samples. *Clin Infect Dis*.

325 Choi, B., M. C. Choudhary, J. Regan, J. A. Sparks, R. F. Padera, X. Qiu, I. H. Solomon, H. H.  
326 Kuo, J. Boucau, K. Bowman, U. D. Adhikari, M. L. Winkler, A. A. Mueller, T. Y. Hsu, M.  
327 Desjardins, L. R. Baden, B. T. Chan, B. D. Walker, M. Lichterfeld, M. Brigl, D. S. Kwon,  
328 S. Kanjilal, E. T. Richardson, A. H. Jonsson, G. Alter, A. K. Barczak, W. P. Hanage, X.  
329 G. Yu, G. D. Gaiha, M. S. Seaman, M. Cernadas & J. Z. Li (2020) Persistence and  
330 Evolution of SARS-CoV-2 in an Immunocompromised Host. *N Engl J Med*, 383, 2291-  
331 2293.

332 Corman, V. M., O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. Chu, T. Bleicker, S.  
333 Brünink, J. Schneider, M. L. Schmidt, D. G. Mulders, B. L. Haagmans, B. van der Veer,  
334 S. van den Brink, L. Wijsman, G. Goderski, J. L. Romette, J. Ellis, M. Zambon, M. Peiris,  
335 H. Goossens, C. Reusken, M. P. Koopmans & C. Drosten (2020) Detection of 2019  
336 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill*, 25.

337 Kim, D., J. Y. Lee, J. S. Yang, J. W. Kim, V. N. Kim & H. Chang (2020) The Architecture of  
338 SARS-CoV-2 Transcriptome. *Cell*, 181, 914-921.e10.

339 Perera, R. A. P. M., E. Tso, O. T. Y. Tsang, D. N. C. Tsang, K. Fung, Y. W. Y. Leung, A. W. H.  
340 Chin, D. K. W. Chu, S. M. S. Cheng, L. L. M. Poon, V. W. M. Chuang & M. Peiris (2020)

341 SARS-CoV-2 Virus Culture and Subgenomic RNA for Respiratory Specimens from  
342 Patients with Mild Coronavirus Disease. *Emerg Infect Dis*, 26, 2701-2704.

343 Speranza, E., B. N. Williamson, F. Feldmann, G. L. Sturdevant, L. Pérez-Pérez, K. Mead-White,  
344 B. J. Smith, J. Lovaglio, C. Martens, V. J. Munster, A. Okumura, C. Shaia, H. Feldmann,  
345 S. M. Best & E. de Wit (2020) SARS-CoV-2 infection dynamics in lungs of African green  
346 monkeys. *bioRxiv*.

347 van Kampen, J. J. A., D. A. M. C. van de Vijver, P. L. A. Fraaij, B. L. Haagmans, M. M. Lamers,  
348 N. Okba, J. P. C. van den Akker, H. Endeman, D. A. M. P. Gommers, J. J. Cornelissen,  
349 R. A. S. Hoek, M. M. van der Eerden, D. A. Hesselink, H. J. Metselaar, A. Verbon, J. E.  
350 M. de Steenwinkel, G. I. Aron, E. C. M. van Gorp, S. van Boheemen, J. C. Voermans, C.  
351 A. B. Boucher, R. Molenkamp, M. P. G. Koopmans, C. Geurtsvankessel & A. A. van der  
352 Eijk (2021) Duration and key determinants of infectious virus shedding in hospitalized  
353 patients with coronavirus disease-2019 (COVID-19). *Nat Commun*, 12, 267.

354 Vogels, C. B. F., A. F. Brito, A. L. Wyllie, J. R. Fauver, I. M. Ott, C. C. Kalinich, M. E. Petrone, A.  
355 Casanovas-Massana, M. Catherine Muenker, A. J. Moore, J. Klein, P. Lu, A. Lu-  
356 Culligan, X. Jiang, D. J. Kim, E. Kudo, T. Mao, M. Moriyama, J. E. Oh, A. Park, J. Silva,  
357 E. Song, T. Takahashi, M. Taura, M. Tokuyama, A. Venkataraman, O. E. Weizman, P.  
358 Wong, Y. Yang, N. R. Cheemarla, E. B. White, S. Lapidus, R. Earnest, B. Geng, P.  
359 Vijayakumar, C. Odio, J. Fournier, S. Bermejo, S. Farhadian, C. S. Dela Cruz, A.  
360 Iwasaki, A. I. Ko, M. L. Landry, E. F. Foxman & N. D. Grubaugh (2020) Analytical  
361 sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets. *Nat*  
362 *Microbiol*.

363 Walsh, K. A., K. Jordan, B. Clyne, D. Rohde, L. Drummond, P. Byrne, S. Ahern, P. G. Carty, K.  
364 K. O'Brien, E. O'Murchu, M. O'Neill, S. M. Smith, M. Ryan & P. Harrington (2020) SARS-  
365 CoV-2 detection, viral load and infectivity over the course of an infection. *J Infect*, 81,  
366 357-371.

367 Wölfel, R., V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T.  
368 C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R.  
369 Ehmann, K. Zwirgmaier, C. Drosten & C. Wendtner (2020) Virological assessment of  
370 hospitalized patients with COVID-2019. *Nature*, 581, 465-469.

371

## 372 **FIGURE LEGENDS**

373

### 374 **Figure 1**

375 Comparison of cycle threshold versus day from symptom onset for clinical samples obtained  
376 from 185 inpatients. Total N (Panel A), subgenomic N (Panel B), total E (Panel C) and  
377 subgenomic E (Panel D). Red dots in panel A and C represent subgenomic negative samples  
378 and black dots represent subgenomic positive samples. Of the 185 patients, 56 were negative  
379 for sgE and 28 were negative for sgN (Shown on y-axis). Pearson correlation coefficients: N = -  
380 0.404  $p < 0.0001$ ; SgN = -0.466,  $p < 0.0001$ ; E = -0.456  $p < 0.0001$ ; sgE = -0.427  $< 0.0001$ . Linear  
381 regression line equations are indicated in each panel.

382

### 383 **Figure 2**

384 Kaplan-Meier analysis showing proportion of patients with positive RT-PCR for subgenomic  
385 transcripts versus day from symptom onset. The median duration from symptom onset to a  
386 negative sgN RT-PCR was 25 days (95% CI: 19.9-30.1) and 14 days for sgE (95% CI 9.6-18.4).  
387 This difference between the curves was significant  $P = 0.001$ . Vertical hash marks indicate  
388 censored cases.

389

390 **Figure 3**

391 Box plots comparing difference in cycle threshold values, delta Ct (subgenomic Ct-genomic Ct)  
392 for cell culture and clinical samples. Total N is expressed at a 4.5-fold (2.2 cycles, STD 0.24)  
393 higher level than sgN and total E is expressed at a 13.9-fold (3.8 cycles STD 0.2) higher level  
394 compared to sgE in HuH-7 cells (Panel A). The difference was more dramatic in NP samples  
395 from inpatients where N was expressed at 16-fold (4.0 cycles, STD 1.1) greater than sgN and E  
396 was expressed at 137.2-fold (7.1 cycles, STD 1.3) greater than sgE expression (Panel B). This  
397 relationship was found at all times post symptom onset (n=185) for N (Panel C) and E (Panel  
398 D). Horizontal hash represents median value, if no IQM shown then N=1 for that sample.  
399 Moderate and extreme outliers noted by circles and stars respectively.

400

401 **Figure 4**

402 Relationship between total and subgenomic RNA in a persistently infectious patient. Total N  
403 was expressed at a 11-fold (3.5 cycles STD 0.75) higher level than subgenomic N, and total E  
404 was expressed at a 84-fold (6.4 cycles, STD 0.8) higher level than subgenomic E when all time  
405 points were combined (Panel A). Time course showing 45 minus cycle thresholds for total  
406 (black) and subgenomic RNA (red) over time. This patient showed persistently positive total and  
407 subgenomic RNA for 119 days. The ratio of total N and sgN of 3 to 4 cycles was maintained  
408 over the duration of infection (Panel B). This was also observed for E and sgE with the ratio of 6  
409 to 7 cycles maintained thorough the course of infection (Panel C).

410

411 **Supplementary Figure 1**

412 Comparison of viral load in  $\log_{10}$  copy number per ml from NP swabs of 185 patients versus  
413 duration of symptoms. Panel A shows total N, Panel B shows subgenomic N, Panel C shows  
414 total E, Panel D shows subgenomic E. Line at  $6.5 \log_{10}$  is the threshold for infectivity found in  
415 previous studies. In Panels B and D, samples that were undetectable after 45 cycles were  
416 arbitrarily set to  $1 \log_{10}$  per ml. Red circles in panels A and B represent the total copy number of  
417 samples that were negative for subgenomic RNA. Subgenomic E is negative at a higher total  
418 copy number than subgenomic N.

419

## 420 **Supplementary Figure 2**

421 Comparison of cycle thresholds for subgenomic RNA to total RNA for subgenomic N and total N  
422 (Panel A) and subgenomic E and total E (Panel B). Samples were obtained from patients  
423 admitted to the hospital (n=190). Subgenomic Ct values on the y-axis are represented as 40  
424 cycles minus the Ct value of subgenomic RNA meaning a lower Ct value represents a lower  
425 copy number of subgenomic RNA. Increasing Ct values on the x-axis for total RNA represent  
426 decreasing copy number for total RNA. Linear regression analysis showed no difference in the  
427 slope between the two lines. There was a significant difference between X-intercepts. The X-  
428 intercept for the comparison of Ct value of N vs Ct value of sgN was 36.01 (CI 35.67 to 36.57)  
429 and the X-intercept for the comparison of Ct value of E vs Ct value of sgE was 32.71 (CI: 32.25  
430 to 33.2) ( $P < 0.0001$ ). Pearson correlation coefficient for sgN/N = -0.976,  $P < 0.0001$ ; for sgE/E = -  
431 0.969 ( $p < 0.0001$ ).

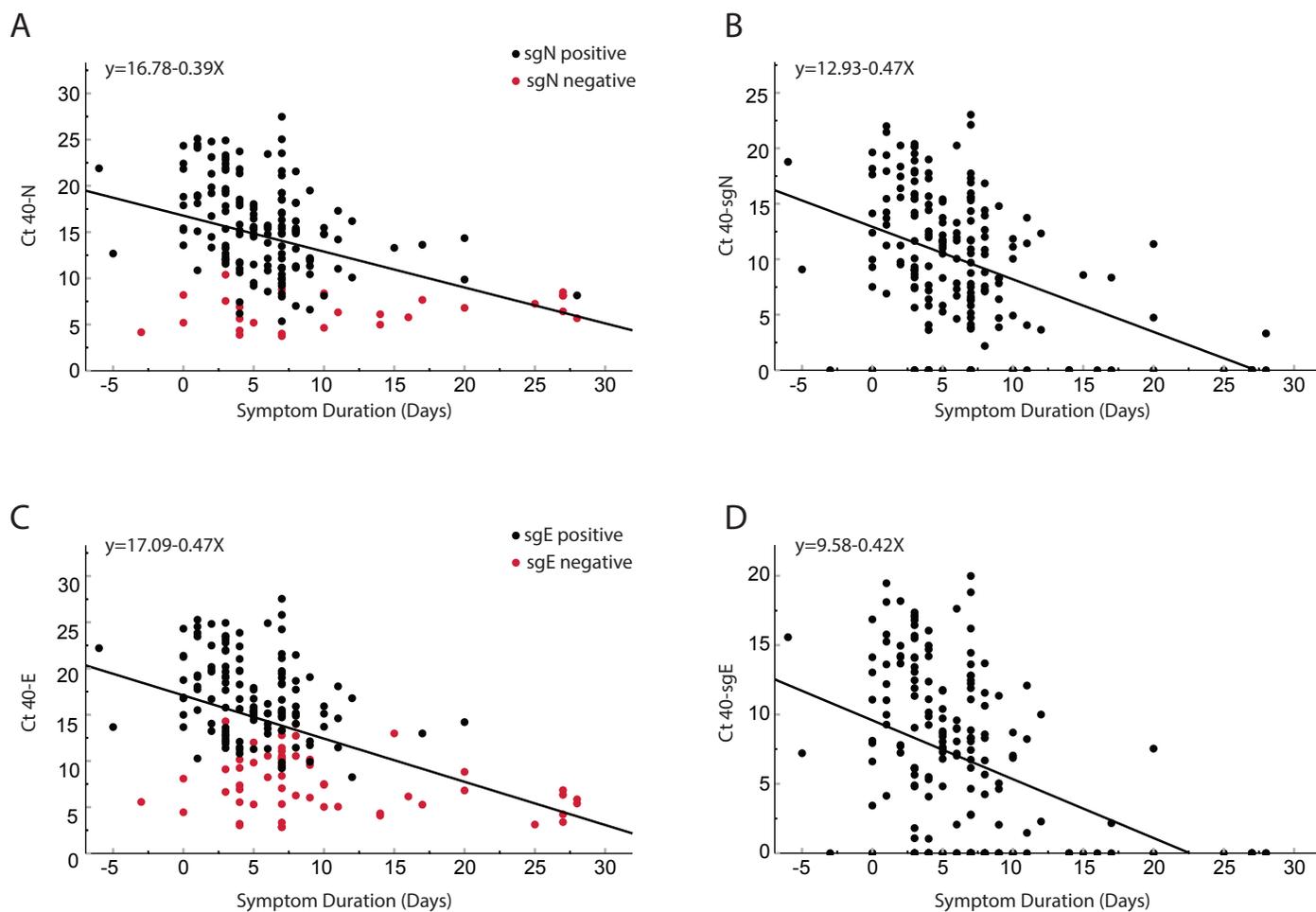


Figure 1

Comparison of cycle threshold versus day from symptom onset for clinical samples obtained from 185 inpatients. Total N (Panel A), subgenomic N (Panel B), total E (Panel C) and subgenomic E (Panel D). Red dots in panel A and C represent subgenomic negative samples and black dots represent subgenomic positive samples. Of the 185 patients, 56 were negative for sgE and 28 were negative for sgN (Shown on y-axis). Pearson correlation coefficients: N = -0.404  $p < 0.0001$ ; SgN = -0.466,  $p < 0.0001$ ; E = -0.456  $p < 0.0001$ ; sgE = -0.427  $p < 0.0001$ . Linear regression line equations are indicated in each panel.

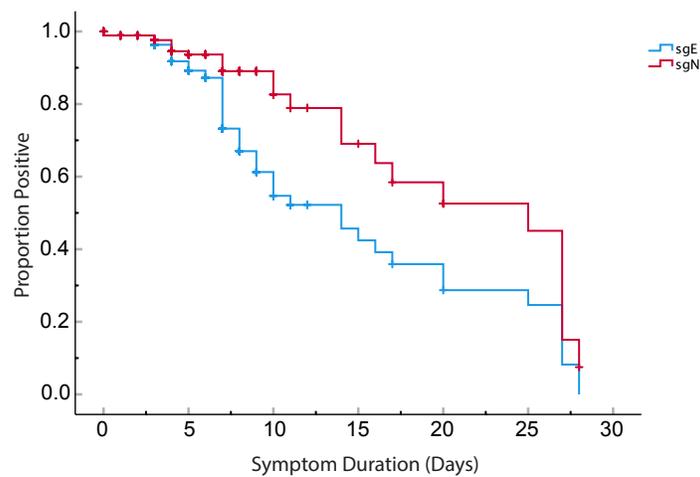


Figure 2

Kaplan-Meier analysis showing proportion of patients with positive RT-PCR for subgenomic transcripts versus day from symptom onset. The median duration from symptom onset to a negative sgN RT-PCR was 25 days (95% CI: 19.9-30.1) and 14 days for sgE (95% CI 9.6-18.4). This difference between the curves was significant  $P=0.001$ . Vertical hash marks indicate censored cases.

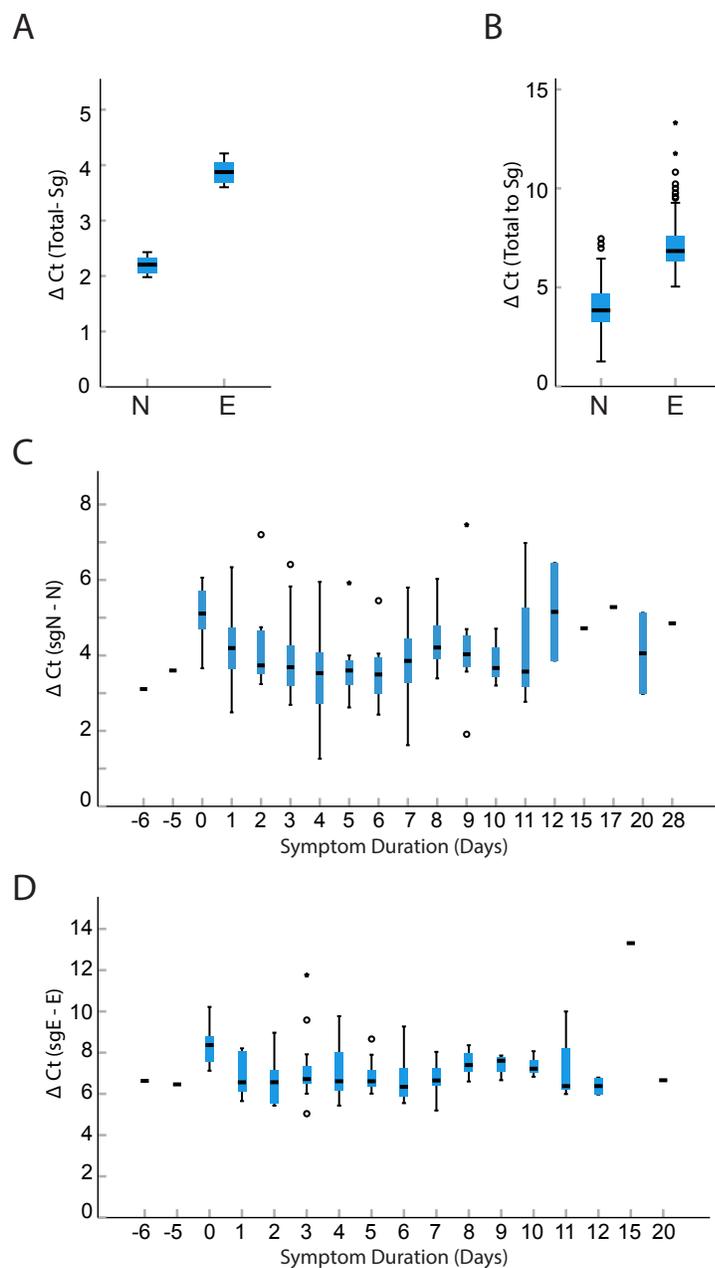


Figure 3

Box plots comparing difference in cycle threshold values, delta Ct (subgenomic Ct-genomic Ct) for cell culture and clinical samples. Total N is expressed at a 4.5-fold (2.2 cycles, STD 0.24) higher level than sgN and total E is expressed at a 13.9-fold (3.8 cycles STD 0.2) higher level compared to sgE in HuH-7 cells (Panel A). The difference was more dramatic in NP samples from inpatients where N was expressed at 16-fold (4.0 cycles, STD 1.1) greater than sgN and E was expressed at 137.2-fold (7.1 cycles, STD 1.3) greater than sgE expression (Panel B). This relationship was found at all times post symptom onset (n=185) for N (Panel C) and E (Panel D). Horizontal hash represents median value, if no IQM shown then N=1 for that sample. Moderate and extreme outliers noted by circles and stars respectively.

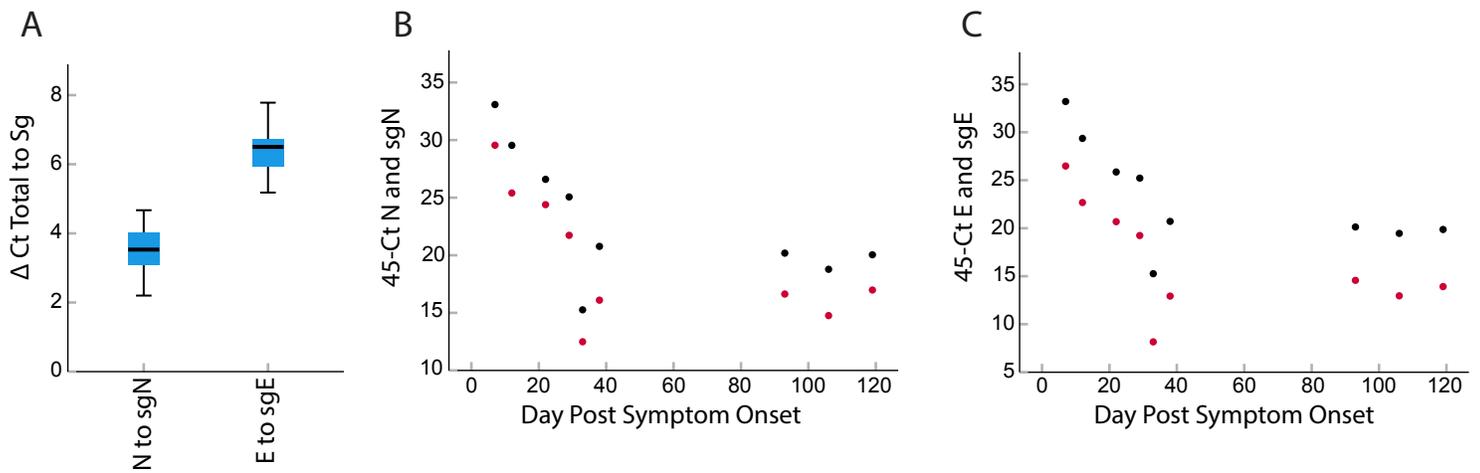
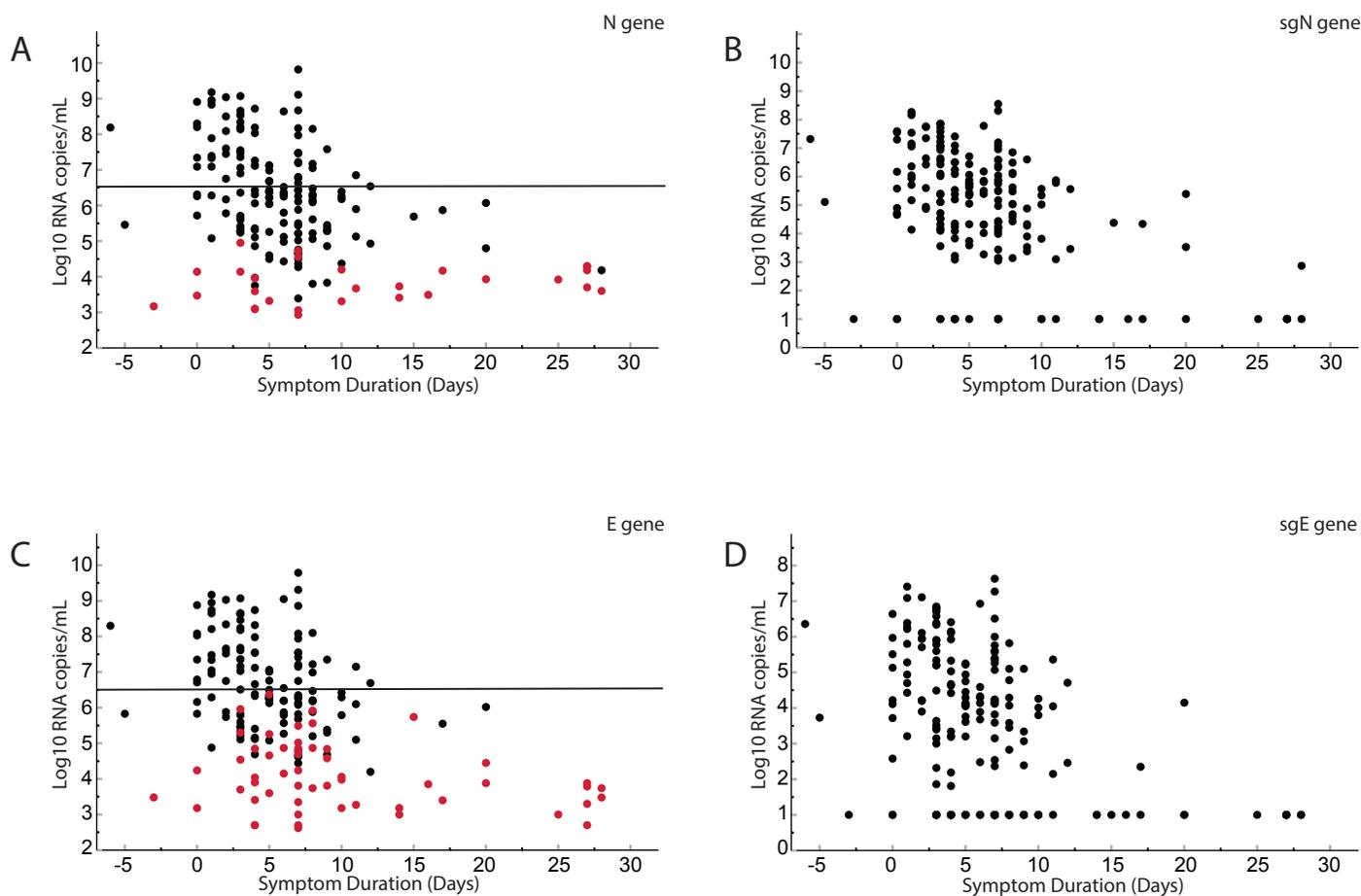


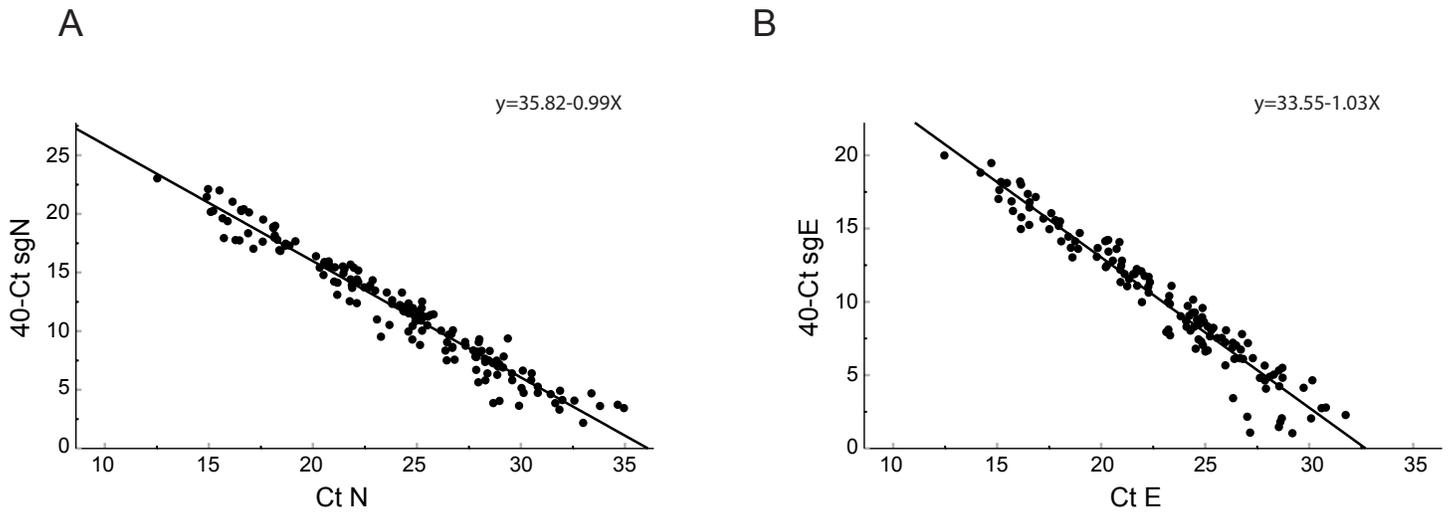
Figure 4

Relationship between total and subgenomic RNA in a persistently infectious patient. Total N was expressed at a 11-fold (3.5 cycles STD 0.75) higher level than subgenomic N, and total E was expressed at a 84-fold (6.4 cycles, STD 0.8) higher level than subgenomic E when all time points were combined (Panel A). Time course showing 45 minus cycle thresholds for total (black) and subgenomic RNA (red) over time. This patient showed persistently positive total and subgenomic RNA for 119 days. The ratio of total N and sgN of 3 to 4 cycles was maintained over the duration of infection (Panel B). This was also observed for E and sgE with the ratio of 6 to 7 cycles maintained through the course of infection (Panel C).



### Supplementary Figure 1

Comparison of viral load in log<sub>10</sub> copy number per ml from NP swabs of 185 patients versus duration of symptoms. Panel A shows total N, Panel B shows subgenomic N, Panel C shows total E, Panel D shows subgenomic E. Line at 6.5 log<sub>10</sub> is the threshold for infectivity found in previous studies. In Panels B and D, samples that were undetectable after 45 cycles were arbitrarily set to 1 log<sub>10</sub> per ml. Red circles in panels A and B represent the total copy number of samples that were negative for subgenomic RNA. Subgenomic E is negative at a higher total copy number than subgenomic N.



### Supplementary Figure 2

Comparison of cycle thresholds for subgenomic RNA to total RNA for subgenomic N and total N (Panel A) and subgenomic E and total E (Panel B). Samples were obtained from patients admitted to the hospital (n=190). Subgenomic Ct values on the y-axis are represented as 40 cycles minus the Ct value of subgenomic RNA meaning a lower Ct value represents a lower copy number of subgenomic RNA. Increasing Ct values on the x-axis for total RNA represent decreasing copy number for total RNA. Linear regression analysis showed no difference in the slope between the two lines. There was a significant difference between X-intercepts. The X-intercept for the comparison of Ct value of N vs Ct value of sgN was 36.01 (CI 35.67 to 36.57) and the X-intercept for the comparison of Ct value of E vs Ct value of sgE was 32.71 (CI: 32.25 to 33.2) ( $P<0.0001$ ). Pearson correlation coefficient for sgN/N = -0.976,  $P<0.0001$ ; for sgE/E = -0.969 ( $p<0.0001$ ).