

1 **Novel RT-ddPCR Assays for determining the transcriptional profile of SARS-CoV-2**

2

3 Sushama Telwatte<sup>a,b</sup>, Nitasha Kumar<sup>a,b</sup>, Albert Vallejo-Gracia<sup>c</sup>, G. Renuka Kumar<sup>c</sup>, Chuanyi

4 M. Lu<sup>a,b</sup>, Melanie Ott<sup>a,c</sup>, Joseph K. Wong<sup>a,b</sup>, Steven A. Yukl<sup>\*a,b</sup>

5

6 a. Department of Medicine, University of California, San Francisco, CA, USA

7 b. Department of Infectious Diseases, San Francisco VA Health Care System, San

8 Francisco, CA, USA

9 c. Gladstone Institute of Virology, Gladstone Institutes, San Francisco, CA, USA

10 \*Corresponding author: Steven Yukl; San Francisco VA Medical Center, 4150 Clement St,

11 111W3, San Francisco, CA 94121, USA; [steven.yukl@ucsf.edu](mailto:steven.yukl@ucsf.edu)

12

13 Declarations of interest: none

14

## 15 **Highlights**

- 16 • We developed a novel panel of 7 quantitative RT-ddPCRs assays for SARS-Cov-2
- 17 • Our panel targets nongenic and genic regions in genomic and subgenomic RNAs
- 18 • All assays detect 1-10 copies and are linear over 3-4 orders of magnitude
- 19 • All assays correlated with the clinical Abbott SARS-CoV-2 Viral Load Assay
- 20 • Clinical samples showed higher copy numbers for targets at the 3' end of the genome

21

## 22 **Abstract**

23 The exact mechanism of coronavirus replication and transcription is not fully  
24 understood; however, a hallmark of coronavirus transcription is the generation of negative-  
25 sense RNA intermediates that serve as the templates for the synthesis of positive-sense  
26 genomic RNA (gRNA) and an array of subgenomic mRNAs (sgRNAs) encompassing  
27 sequences arising from discontinuous transcription.

28 Existing PCR-based diagnostic assays for SAR-CoV-2 are qualitative or semi-  
29 quantitative and do not provide the resolution needed to assess the complex transcription  
30 dynamics of SARS-CoV-2 over the course of infection. We developed and validated a novel  
31 panel of specially designed SARS-CoV-2 ddPCR-based assays to map the viral transcription  
32 profile. Application of these assays to clinically relevant samples will enhance our  
33 understanding of SARS-CoV-2 replication and transcription and may also inform the  
34 development of improved diagnostic tools and therapeutics.

35

36

37 **Key words:** SARS-CoV-2; droplet-digital PCR; quantitative assays; coronavirus; viral  
38 transcription/replication

## 39 1. Introduction

40 The etiologic agent responsible for the ongoing COVID-19 pandemic, identified as  
41 Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2),<sup>1,2</sup> is an enveloped virus  
42 with a positive-sense, single-stranded RNA genome of ~30 kb and is a member of the  $\beta$ -  
43 coronavirus genus. SARS-CoV-2, which is the seventh coronavirus known to infect humans,  
44 shares approximately 50% sequence homology with MERS and 79% sequence homology  
45 with SARS-CoV<sup>3</sup> but appears to be more closely related to the SARS-like bat coronaviruses  
46 RmYN02 from *R. malayanus* and RaTG13 from *R. affinis* (93.3% and 96.1% sequence  
47 identity, respectively),<sup>4</sup> though its origin is, to date, unsettled<sup>5,6</sup>.

48 The exact mechanism of SARS-CoV-2 replication and transcription is not fully  
49 understood; however, a hallmark of coronavirus transcription and other viruses of the order  
50 *Nidovirales* is the generation of negative-sense RNA intermediates that serve as the templates  
51 for the synthesis of positive-sense genomic RNA (gRNA) and an array of subgenomic RNAs  
52 (sgRNAs), which arise from discontinuous transcription and encompass sequences from both  
53 ends of the genome<sup>7,8</sup> (Fig. 1). Following cell entry, SARS-CoV-2 genomic RNA is transcribed  
54 and translated to generate the nonstructural proteins (NSP) from the two open reading frames  
55 (ORF), ORF1a and ORF1b<sup>8</sup>, a process thought to involve the virus replication complex,  
56 transcription-regulating sequences (TRSs), the N protein, and double membrane vesicles in  
57 the cytoplasm of infected cells<sup>9-11</sup>. During the synthesis of the negative strand RNA, sgRNAs  
58 arise from a template switch that adds a copy of the 'leader' sequence (~70 nucleotides in the  
59 5' untranslated region [UTR] containing a short transcription-regulating sequence [TRS] at the  
60 3' end) to the 'body' sequence derived from one of various genes in the 3' third of the genome  
61 (including genes for structural proteins)<sup>12-14</sup>. Transcription of the sgRNAs is likely regulated by  
62 TRS sequences in the leader sequence and upstream of 3' genes<sup>9</sup>, and may allow for greater  
63 expression of certain viral genes.

64 A recently published study confirms that a similar mechanism exists for SARS-CoV-2  
65 to generate nine canonical sgRNAs distinct from genomic RNA<sup>8</sup> (Fig. 1). For other  
66 coronaviruses, sgRNAs encode virulence factors such as proteins that directly cause lesions<sup>15</sup>

67 or indirectly inhibit immune responses<sup>16</sup>. Incorporation of 5'UTR sequences into the capped  
68 subgenomic mRNA templates of SARS-CoV may confer resistance to cleavage by viral nsp1  
69 protein<sup>17</sup>, which typically inhibits host gene expression by degradation of host mRNA<sup>18-20</sup>. For  
70 positive-sense RNA viruses, sgRNAs act as messengers for expression of structural proteins  
71 or proteins related to pathogenesis and can regulate the transition between translation and  
72 virion production<sup>21</sup>. The various roles of sgRNAs in SARS-CoV-2 infection and pathogenesis  
73 remain to be elucidated, but the rapid accumulation and persistence of sgRNAs following  
74 infection may also contribute to disease progression.

75  
76         Understanding the viral dynamics of SARS-CoV-2 and the host response are essential  
77 in devising strategies to develop antiviral treatments or vaccines and curb new infections.  
78 Existing PCR-based diagnostic assays for SAR-CoV-2, which are interpreted in a qualitative  
79 or semi-quantitative manner (positive, negative or indeterminate) and target only 1-2 viral  
80 regions, do not distinguish between genomic and subgenomic RNA or account for possible  
81 differences between the RNA copy numbers of various viral genes, which may depend on the  
82 degree to which they are transcribed as various sgRNAs and the degree to which the sample  
83 includes virion or cell-associated RNA. Molecular assays that can quantify different viral genes  
84 found in genomic and sgRNA species will have utility in charting the extent of viral replication  
85 and changes in SARS-CoV-2 transcription over the course of infection.

86         We have devised a novel panel of seven ddPCR-based assays that target various  
87 conserved regions of SARS-CoV-2 RNA, including the 5' and 3' untranslated regions, non-  
88 structural genes that are only found in full length (genomic) RNA and structural genes that are  
89 also contained in different sgRNAs (Fig.1 and Table 1).

90         We selected genes encoding two non-structural proteins [Main Proteinase (NSP5) and  
91 RNA dependent RNA polymerase (RdRp-NSP12)] and four major structural proteins [Spike  
92 glycoprotein (S), envelope (E), membrane (M), and nucleocapsid (N)] that are known to serve  
93 critical functions in SARS-CoV-2 infection. For the spike protein, in which notable mutations  
94 have emerged<sup>22</sup>, we designed a primer/probe set to target the short, highly-conserved

95 'polybasic cleavage site' ('S-PBCS') of SARS-CoV-2 which is functionally cleaved to yield the  
96 S1 and S2 subunits<sup>23</sup>, in a similar manner to the hemagglutinin (HA) protein of avian influenza  
97 viruses (AIVs)<sup>24</sup>. In AIVs, the insertion or substitution of basic amino acids at the HA cleavage  
98 site is associated with enhanced pathogenicity<sup>25,26</sup>. The SARS-CoV-2 PBCS allows effective  
99 cleavage by host furin and other proteases<sup>5</sup>, and may potentially enhance its infectivity in  
100 humans and distinguish it from related animal coronaviruses<sup>4,5,27</sup>. Elucidating the granular  
101 detail of SARS-CoV-2 transcription could help us to understand how the virus replicates and  
102 how it may evade human immune defenses. Detailed mapping of the expressed viral  
103 transcripts across times and cell types is essential for further studies of viral gene expression,  
104 mechanisms of replication, and probing host-viral interactions involved in pathogenicity.

105

## 106 **2. Materials and methods**

### 107 **2.1 Primer design and selection**

108 Multiple primer and probe sets were designed to target various regions of SAR-CoV-  
109 2, including untranslated regions that likely play an important role in regulating transcription  
110 (5' and 3' untranslated regions [UTR]), non-structural genes found only in genomic RNA (main  
111 protease [NSP5; ORF1a], RNA-dependent RNA polymerase [RdRp; ORF1b]), and structural  
112 genes that may also be found in various sgRNAs (spike [S] protein [ORF2] polybasic cleavage  
113 site [PBCS], membrane [M] glycoprotein [ORF5], and nucleocapsid [N] protein [ORF9]).  
114 Primers/probes were designed using the Primer Quest® Tool (Integrated DNA Technologies,  
115 Coralville, IA). A multiple sequence alignment was performed using Clustal Omega<sup>28</sup>,  
116 encompassing complete sequences of 86 SARS-CoV-2 isolates from all geographical  
117 locations and all sequences available from the US on 3/14/2020. Reference sequences of  
118 other coronaviruses, including SARS-CoV (NC\_004718.3), MERS-CoV (NC\_019843.3),  
119 HCoV-229E (NC\_002645.1), HCoV-NL63 (NC\_005831.2), HCoV-OC43 (NC\_006213.1), and  
120 HCoV-HKU1 (NC\_006577.2), were included in the alignment to exclude primer sets with  
121 significant overlap with non-SARS-CoV-2 sequences. Two primer/probe sets that aligned to  
122 all SARS-CoV-2 isolates but had 1 or more mismatch with SARS-CoV and greater than 5

123 mismatches with MERS-CoV, HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1 were  
124 selected for each region (Table 1). A sequence similarity analysis using Basic Local Alignment  
125 Search Tool (BLAST)<sup>29</sup> found no significant similarity in any primer or probe to human  
126 sequences.

127

## 128 **2.2 Validations using plasmid DNA**

129 Plasmid constructs containing the regions of interest (5'UTR, 3'UTR, Main Proteinase,  
130 M gene, N gene, S protein, and a 528nt fragment of RdRp) were designed in pBluescript KS(+)  
131 (Bio Basic Inc., Ontario, Canada) to enable assay validations using DNA and for use in *in vitro*  
132 transcription reactions to generate viral RNA for standards. Plasmid concentrations were  
133 quantified using ultraviolet (UV) spectrophotometry (NanoDrop ND-1000 instrument, Thermo  
134 Fisher) and the molecular weights were used to calculate the number of molecules per  $\mu\text{L}$ .  
135 Extracted PBMC from a healthy donor (150-200 ng/well) and H<sub>2</sub>O were included as negative  
136 controls for each assay.

137 Each primer and probe set was tested using droplet digital PCR (ddPCR), as  
138 performed using the QX100 system (Bio-Rad). Droplet digital PCR was chosen because it  
139 enables “absolute” quantification, it is relatively less dependent on PCR efficiency (which may  
140 be reduced by sequence mismatches or inhibitors), and it may be more precise than  
141 quantitative PCR (qPCR) at low copy numbers<sup>30</sup>. Plasmid DNA was added to ddPCR wells at  
142 expected inputs of 1-10<sup>3</sup> copies/well in duplicate (1000 and 100 copies) or quadruplicate (10  
143 and 1 copy). Each reaction consisted of 20  $\mu\text{L}$  per well containing 10  $\mu\text{L}$  of ddPCR Probe  
144 Supermix (no deoxyuridine triphosphate), 900 nM of primers, 250 nM of probe, and 5  $\mu\text{L}$  of  
145 plasmid DNA. Droplets were amplified using a Mastercycler® nexus (Eppendorf, Hamburg,  
146 Germany) with the following cycling conditions: 10 min at 95°C, 45 cycles of 30 s at 95°C and  
147 59°C for 60 s, and a final droplet cure step of 10 min at 98°C. Droplets were read and analyzed  
148 using the QuantaSoft software in the absolute quantification mode.

149

## 150 **2.3 Validations using synthetic RNA**

151 *In vitro* transcribed (IVT) RNA standards were generated from the aforementioned  
152 plasmids using the T7 RiboMAX™ Express Large-Scale RNA Production System (Promega,  
153 Madison, WI). The concentration of each IVT RNA standard was measured by Nanodrop and  
154 the molecular weight was used to calculate the expected number of molecules per  $\mu\text{L}$ . The  
155 length, integrity, and concentration of each IVT standard were confirmed using the Agilent  
156 Bioanalyzer RNA 6000 Nano assay (Agilent, Santa Clara, CA) prior to dilution in nuclease-  
157 free water to working concentrations.

158 A reverse transcription (RT) reaction was performed in 50  $\mu\text{L}$  containing 5  $\mu\text{L}$  of 10 $\times$   
159 SuperScript III buffer (Invitrogen), 5  $\mu\text{L}$  of 50 mM  $\text{MgCl}_2$ , 2.5  $\mu\text{L}$  of random hexamers (50  
160 ng/ $\mu\text{L}$ ; Invitrogen), 2.5  $\mu\text{L}$  of 50  $\mu\text{M}$  poly-dT15, 2.5  $\mu\text{L}$  of 10 mM deoxynucleoside triphosphates  
161 (dNTPs), 1.25  $\mu\text{L}$  of RNaseOUT (40 U/ $\mu\text{L}$ ; Invitrogen), and 2.5  $\mu\text{L}$  of SuperScript III RT (200  
162 U/ $\mu\text{L}$ ; Invitrogen). Although the IVT standards were not polyadenylated, reverse transcription  
163 was performed with both random hexamers and poly-dT because we anticipated that these  
164 assays would be applied to clinical samples containing long polyadenylated SARS-CoV-2  
165 RNAs, for which the combination of poly-dT plus random hexamers may reduce bias towards  
166 reverse transcription of any one region (as can be seen with specific reverse primers), the 5'  
167 end (as would be expected with random hexamers), or the 3' end (as would be expected with  
168 poly-dT).

169 IVT RNA standards were added to RT reactions at inputs of 1, 10, 10<sup>2</sup>, 10<sup>3</sup>, and 10<sup>4</sup>  
170 copies per 5  $\mu\text{L}$  (2 replicate RT reactions for each input). RT reactions were performed in a  
171 conventional thermocycler at 25.0°C for 10 min, 50.0°C for 50 min, followed by an inactivation  
172 step at 85.0°C for 5 min. Undiluted RT product (5  $\mu\text{L}$ ) was added to ddPCR reactions (total  
173 volume of 20  $\mu\text{L}$ ) and ddPCR was performed as described for 'Validations using plasmid DNA'.  
174 Primer-probe sets for each target region were tested head-to-head using this approach. Based  
175 on performance of each primer-probe set using plasmid DNA and IVT RNA, one primer/probe  
176 set for each region was selected for further testing.

177 To determine the robustness of our approach, in addition to testing each assay with  
178 varying RNA copy inputs (each with two replicate RT reactions per input and replicate ddPCR  
179 wells for each RT), we performed repeat, independent experiments using the same  
180 parameters to confirm each assay's efficiency and sensitivity (n=4 for N-ORF9, CDC\_N1, and  
181 CDC\_N2; n=3 for 5'UTR, 3'UTR; and n= 2 for all others). No data were excluded as outliers.

182

#### 183 **2.4 Validations using SARS-CoV-2 virion RNA**

184 Vero CCL-81 kidney epithelial cells, derived from *Cercopithecus aethiops*, were  
185 infected with SARS-CoV-2 (Isolate: USA-WA1/2020) at an MOI of 0.003 (250 000 cells/well).  
186 Cells were incubated for 72 hours at 37°C/5% CO<sub>2</sub> and harvested. Viral supernatant was  
187 clarified by 2 centrifugation steps (180 xg, 5 min) and added directly to 1mL TRI reagent  
188 (Molecular Research Center Inc.). Total RNA was extracted using TRI reagent, including the  
189 addition of polyacryl carrier (2.5µL). Extracted RNA was then subjected to two rounds of  
190 DNase I treatment as follows to ensure degradation and removal of contaminating DNA. First,  
191 eluted RNA was added to a DNase Reaction Mix containing 40mM Tris-HCL (pH 7.9;  
192 Invitrogen), 6mM MgCl<sub>2</sub> (Ambion), 10mM CaCl<sub>2</sub> (Sigma) and 1 U DNase RQ1 (Promega) and  
193 incubated at 37°C for 15 minutes. Next, virion RNA was purified using the RNeasy Mini Kit  
194 with on-column DNase digestion with RNase-Free DNase I (Qiagen). The copies/µL in the  
195 virion standard were estimated by triplicate measurements using the Abbott RealTime SARS-  
196 CoV-2 assay (Abbott m2000 Molecular Platform). Dilutions of the virion standard were added  
197 to RT reactions to achieve expected inputs of 1 to 70,000 copies per 5uL RT (the input into  
198 each ddPCR well). RT reactions were performed as above, with random hexamers and poly-  
199 dT, except that the total volume of the RT was scaled up so that two replicate 5uL aliquots of  
200 cDNA could then be used to test each assay in parallel using replicate 20uL ddPCR reactions  
201 (see above) containing primers/probe specific for a given region.

202

#### 203 **2.5 Assay efficiency in presence of background RNA**



204 Further validations were performed to determine each assay's sensitivity to inhibition  
205 by "background" cellular RNA, as would be expected in clinical samples containing cells. The  
206 virion standard (1000 copies per 5 $\mu$ L RT) was added to RT reactions with or without cellular  
207 RNA from A549 cells (lung epithelial cell line) or donor PBMC (both added at 100ng/ $\mu$ l per RT,  
208 or 500ng per ddPCR well). RT reactions contained a total of 125 $\mu$ L with 12.5  $\mu$ L of 10 $\times$   
209 SuperScript III buffer (Invitrogen), 12.5  $\mu$ L of 50 mM MgCl<sub>2</sub>, 6.25  $\mu$ L of random hexamers (50  
210 ng/ $\mu$ L; Invitrogen), 6.25  $\mu$ L of 50  $\mu$ M dT15, 6.25  $\mu$ L of 10 mM deoxynucleoside triphosphates  
211 (dNTPs), 3.125  $\mu$ L of RNaseOUT (40 U/ $\mu$ L; Invitrogen), and 6.25  $\mu$ L of SuperScript III RT (200  
212 U/ $\mu$ L; Invitrogen). RT reactions were incubated at 25.0°C for 10 min, 50.0°C for 50 min,  
213 followed by an inactivation step at 85.0°C for 5 min. Undiluted cDNA (5  $\mu$ L) was added to each  
214 20  $\mu$ L ddPCR reaction and replicate ddPCR reactions were performed for each assay.

215

## 216 ***2.6 Assay validations in clinical diagnostic samples from SARS-CoV-2 infected*** 217 ***individuals***

218 To investigate the viral transcription profile in clinical samples and determine whether  
219 our RT-ddPCR assays correlate with a clinical test, we obtained unused nucleic acid (ranging  
220 from 8.25-16.8 $\mu$ L) that remained after extraction by the Abbott m2000 instrument from  
221 nasopharyngeal swabs from 3 individuals who tested positive with the Abbott Real Time  
222 SARS-CoV-2 assay. Nucleic acid from these 3 individuals, who had C<sub>t</sub> values of 11.59, 15.81,  
223 and 19.14 (respectively) on the Abbot assay, was tested using our RT-ddPCR assays for the  
224 5'UTR, Main Proteinase, RdRp, S, M, N and 3'UTR regions. The available volume of nucleic  
225 acid was added into 85 $\mu$ L RT reactions containing 1 $\times$  SuperScript III buffer, 5 mM MgCl<sub>2</sub>, 2.5  
226 ng of random hexamers, 2.5  $\mu$ M dT15, 0.5 mM deoxynucleoside triphosphates (dNTPs),  
227 1U/ $\mu$ L of RNaseOUT, and 10U/ $\mu$ L of SuperScript III RT. RT reactions were performed under  
228 the aforementioned conditions. Undiluted cDNA was divided evenly across assays (5 $\mu$ L input  
229 into each ddPCR well, tested in duplicate) and ddPCR reactions were performed under the  
230 conditions described for 'Validations using synthetic RNA'. Absolute values obtained by

231 ddPCR were adjusted to account for differing input volume of nucleic acid to yield the SARS-  
232 CoV-2 copies/ $\mu$ L extract. The log-linear relationship between viral load measured by RT-PCR  
233 (Abbott Real Time SARS-CoV-2 assay) and RT-ddPCR was determined using GraphPad  
234 Prism (version 8.4.1).

235

### 236 **3. Results**

237

#### 238 **3.1 Detection limit, linearity, and efficiency using plasmid DNA**

239 Two assays were designed for each region (indicated in Fig. 1; 'Assay locations')  
240 except the spike protein polybasic cleavage site. To evaluate the performance of each assay  
241 at the PCR stage, each pair of assays was tested on plasmid DNA. Since no commercially  
242 available plasmid contains the whole SARS-CoV-2 genome, and construction of such a  
243 plasmid is technically challenging (due to the 30kb length) and subject to higher biosafety  
244 restrictions, we constructed or purchased plasmids containing individual genes or regions.  
245 For each plasmid, the DNA concentration was measured by UV spectroscopy (NanoDrop) and  
246 the number of molecules (expected copies) was calculated using the molecular weight.  
247 Each assay was assessed for detection limit, dynamic range, linearity, and efficiency by  
248 measuring the absolute number of copies detected using droplet digital PCR (ddPCR) from  
249 expected inputs of serially diluted plasmid DNA. All assays could detect as few as 1-10 copies  
250 and were linear over at least 3 orders of magnitude ( $R^2 > 0.99$  for all; Fig. 2). Assay efficiencies  
251 (measured by the slope) varied somewhat between assays, ranging from 0.67 ("N-ORF9\_8")  
252 to 1.1 ("M-ORF5"). One assay from each pair was selected for further study (Table 2; rejected  
253 primer/probe sets are listed in Table S1) based on the overall efficiency (Fig. 2), separation  
254 between the positive and negative droplets [amplitude/signal to noise] (Fig. S2), and specificity  
255 (Table S2). For the chosen assays, no positive droplets were detected with water or DNA  
256 from peripheral mononuclear blood cells (PBMC) from uninfected blood donors.

257

258

### 259 **3.2 Detection limit, linearity, and efficiency using *in vitro* transcribed and virion RNA**

260 Selected assays for each region were tested using standards prepared from *in vitro*  
261 transcribed (IVT) RNA from the designed plasmids (5'UTR, Main Proteinase, RDRP, S, M, N  
262 and 3'UTR; Fig. 3 and Table 2). The expected copy numbers were calculated using the RNA  
263 concentration (as measured by UV spectroscopy [NanoDrop] and confirmed by the Agilent  
264 Bioanalyzer) and the molecular weight. Using RT-ddPCR, all assays could detect as few as  
265 10 copies of RNA and demonstrated linearity over 3-4 orders of magnitude ( $R^2 > 0.999$  for all;  
266 Fig. 3). The efficiencies for detecting IVT RNA standards, which ranged from 0.18 (for Main  
267 Proteinase) to 0.96 (S-PBCS), were more variable than those observed for plasmid DNA. No  
268 amplification was detected in 'No RT' control reactions containing 10,000 IVT RNA  
269 copies/well, confirming the absence of any contaminating plasmid DNA. However, it is worth  
270 noting that none of these IVT standards were polyadenylated (so they should not be reverse-  
271 transcribed by poly-dT) and some of the standards were very short (<300 base pairs), which  
272 would likely limit the efficiency with which they were reverse transcribed by random hexamers.  
273 In addition, some of the measured differences in efficiency could reflect actual differences in  
274 the copy numbers present in the various IVT standards, which are difficult to determine  
275 precisely.

276

277 To circumvent these limitations, we prepared one SARS-CoV-2 'virion' standard  
278 containing all of the target regions by extracting RNA from cell-free supernatant from a cell  
279 line (Vero CCL81) infected *in vitro* with a SARS-CoV-2 patient isolate (USA-WA1/2020). The  
280 expected copies in this virion standard were calculated using the  $C_t$  value measured by the  
281 Abbott m2000 Real Time SARS-CoV-2 viral load assay, which targets the N and RdRp genes  
282 using probes labelled with the same fluorophore. This virion standard enabled the preparation  
283 of common RT reactions containing specific inputs of SARS-CoV-2 genomic equivalents, from  
284 which aliquots of cDNA could be divided evenly across our panel of assays for simultaneous  
285 assessment of all target regions in ddPCR reactions (Fig. 4).

286 Expected inputs of 10 to 70,000 copies per well were used to measure the absolute  
287 copies of 5'UTR, Main Proteinase, RdRp, S, M, N and 3'UTR regions. All assays detected as  
288 few as 10 copies of the virion standard and were linear over four orders of magnitude  
289 ( $R^2 > 0.999$  for all). No amplification was detected in 'No RT' control reactions containing 10,000  
290 IVT RNA copies/well. Assay efficiencies were all greater than 1.0 (range: 1.05 to 2.46), likely  
291 because the estimate from the Abbott assay was lower than the true value and/or the RT-  
292 ddPCR assays are more efficient. In addition, the efficiency of the RT-ddPCR assays  
293 increased from 5' to 3' targets, which could reflect the presence of 3' subgenomic RNAs in the  
294 virion standard or greater efficiency of reverse transcription from the 3' end of the genome.

295

### 296 **3.3 Assay specificity and false positive rate**

297 To determine the non-specific reactivity of oligonucleotides (false positive rate) for  
298 each assay, we performed a median of 26 [range 18-32] 'no template' controls (NTC). These  
299 reactions were performed with both water (water NTC) and DNA or RNA isolated from SARS-  
300 CoV-2-negative donor PBMC (DNA/RNA NTC) (Table S2). Except for one experiment using  
301 IVT RNA, where a total of three droplets were detected across duplicate NTC wells containing  
302 donor PBMC tested for Main Proteinase-NSP5, no other false positives were observed.

303

### 304 **3.4 Comparison of new and existing SARS-CoV-2 assays in ddPCR platform**

305 Our assay panel included new primers/probes for the nucleocapsid (N-ORF9), which  
306 is targeted by existing diagnostic real-time PCR assays. We compared the performance of our  
307 'N-ORF9' primers/probe to the primers/probes from the U.S. Center for Disease Control  
308 assays for the nucleocapsid (CDC-N1 and CDC-N2)<sup>31</sup> using ddPCR. The N-ORF9 assay  
309 efficiency was similar to that of CDC-N1 and CDC-N2 for plasmid DNA, in between that of  
310 CDC-N1 and CDC-N2 for IVT RNA, and similar to CDC-N1 for the virion standard (Fig. 2-4).

311 In addition, we compared our primers/probes for the RdRp to published primers/probes  
312 for the "IP2" assay<sup>32</sup> (which targets ORF1a) and "E-Sarbeco"<sup>33</sup> assay (which targets the E  
313 gene) using RT-ddPCR and the virion standard (Fig. 5; Table 3). The IP2 (ORF1a) assay

314 efficiency was 1.11, compared to 1.20-1.28 for our RdRp (ORF1b) and 1.36 for our main  
315 protease (ORF1a) assays (Fig. 4-5). The E-Sarbeco [ORF4] assay efficiency (1.08) was  
316 similar to the IP2, but may have been less than our assays targeting neighboring genes (S-  
317 PBCS [ORF2]: 1.32; M-ORF5: 1.51).

318  
319

### 320 **3.5 Lower limit of detection of SARS-CoV-2 in RNA**

321 Our validation studies included SARS-CoV-2 RNA inputs down to 1 copy per ddPCR  
322 reaction (Fig. 2-3). We estimated the lower limit of detection (LLOD) for each assay in our  
323 panel based on data for all replicates tested at 10 copy and 1 copy inputs (Table S3). At 10  
324 copies, all of our assays detected SARS-CoV-2 in  $\geq 85.7\%$  of tests (range= 85.7-100%). At 1  
325 copy input, our assays detected SARS-CoV-2 in  $\geq 25\%$  of tests (range=25-88%), underscoring  
326 the high sensitivity of our assays.

327

### 328 **3.6 Effect of Background RNA on assay efficiencies**

329 Next, we assessed the efficiencies of our assays in the presence of “background” RNA  
330 from uninfected cells (Fig. 6). At a constant input of 1000 copies of the SARS-CoV-2 virion  
331 RNA, we determined the effect of adding cellular RNA (100ng per  $\mu\text{L}$  of RT) extracted from  
332 PBMC or a lung epithelial cell line (A549 cells). All assays showed slightly greater efficiency  
333 in the presence of 100ng/ $\mu\text{L}$  background RNA from either PBMC or A549 cells compared to  
334 the virion standard with no background RNA. No false positives were detected with 100ng/ $\mu\text{L}$   
335 RT RNA from PBMC, while 1-4 droplets were sometimes detected in the RNA from A549 cells  
336 using some assays (Main Proteinase, RdRp, S-PBCS). Overall, these data suggest that in  
337 samples derived from individuals infected with SARS-CoV-2, our assays are likely to be  
338 minimally inhibited by background RNA, making them ideally suited to a diverse range of  
339 clinical samples.

340

### 341 **3.7 Strong correlation between viral loads measured by RT-ddPCR and real-time PCR** 342 **in clinical diagnostic samples**

343 To compare our assays with a clinical test, we obtained unused nucleic acid that had  
344 been extracted by the Abbott m2000 molecular platform from nasopharyngeal swabs from  
345 three SARS-CoV-2-infected individuals and remained after clinical testing using the Abbott  
346 Real Time SARS-CoV-2 assay. Using this nucleic acid, we measured RNA levels of the  
347 5'UTR, Main Proteinase, RdRp, S, M, N and 3'UTR regions using our RT-ddPCR assays. (Fig.  
348 7). As observed with the virion standard, transcripts containing the most 3' regions (N-ORF9  
349 and 3'LTR) tended to be present at higher copy numbers, while those containing the 5'LTR  
350 tended to be present at lower levels. However, the order of transcript levels varied somewhat  
351 between individuals and sometimes differed from the 3' to 5' gradient observed with the virion  
352 standard. For example, levels of S-PBCS RNA tended to be lower than those of the more 5'  
353 Main Protease (NSP5) transcripts. These potential differences in SARS-CoV-2 transcription  
354 profile may reflect changes in viral dynamics over the course of infection or inter-individual  
355 variability in viral sequences or host responses, and should be confirmed in future studies  
356 using longitudinal samples from more individuals.

357  
358 Next, we determined the correlation between the  $C_t$  value as measured by the Abbott  
359 assay and SARS-CoV-2 copy numbers as determined by RT-ddPCR. For each target, this  
360 relationship was modelled using linear regression following log transformation of SARS-CoV-  
361 2 copies/ $\mu$ L extract [where  $y = \text{Log}_2(x)$ ] (Fig S1). The coefficient of determination ( $R^2$ ) for each  
362 model was  $\geq 0.93$  for all targets, underscoring the log-linear relationship between ddPCR-  
363 based SARS-CoV-2 transcript levels and  $C_t$  values in diagnostic specimens. Taken together,  
364 these data strongly underscore the sensitivity of our assays, demonstrating the ability to detect  
365 all targets using minimal RNA inputs (effectively 1.2-2.4  $\mu$ L RNA input per assay), and their  
366 strong correlation with  $C_t$  values obtained by real-time PCR using clinical assays. Furthermore,  
367 these data highlight that delineation of the SARS-CoV-2 transcription profile in samples across  
368 differing timepoints within and between participants may yield valuable insight into viral  
369 transcription dynamics across the course of SARS-CoV-2 infection.

370

#### 371 **4. Discussion**

372           The 2019 SARS-CoV-2 outbreak has heralded the development of an array of  
373 diagnostic molecular tools to study this novel coronavirus. However, currently described PCR-  
374 based diagnostic assays are qualitative or semi-quantitative, are limited to the simultaneous  
375 detection of one or two regions, and do not distinguish genomic from subgenomic RNAs.  
376 Here, we report a panel of new primer/probe sets that span the SARS-CoV-2 genome and  
377 target important nongenic regions, non-structural genes found only in genomic RNA, and  
378 structural genes that are also found in different subgenomic RNAs.

379           We used these new primers/probes for RT-ddPCR rather than qRT-PCR because  
380 ddPCR provides absolute quantification (does not require an external calibrator), tends to  
381 tolerate sequence mismatches in primer/probe sequences better than qRT-PCR, and may be  
382 more precise at low copies, while providing similar sensitivity and reproducibility<sup>30,35</sup>. During  
383 validation of these assays with multiple different standards, we sometimes found that the  
384 efficiency of the same assay varied somewhat across different standards. These differences  
385 may reflect differences in the nature of the standards (DNA, short *in vitro* transcribed non-  
386 polyadenylated RNA, or “virion RNA”) as well as the difficulty in determining the exact number  
387 of copies in an external standard; the latter issue highlights a major advantage of the absolute  
388 quantification provided by ddPCR. On all standards tested, the seven RT-ddPCR assays were  
389 extremely sensitive (down to 1-10 copies) and linear over 3-4 orders of magnitude, and all  
390 seven assays showed no inhibition by up to 500,000 cell equivalents of RNA per ddPCR well,  
391 suggesting that these assays could be extremely useful for SARS-CoV-2 research. While most  
392 existing clinical assays for SARS-CoV-2 use qPCR because it is less expensive and may have  
393 fewer false positives than ddPCR, it is likely that the primer/probe sets described here would  
394 also work well in qPCR assays for research or clinical testing.

395           The utility of assays that target multiple genomic regions is supported by studies  
396 demonstrating loss in sensitivity of published assays owing to mutations that could affect  
397 primer annealing. For instance, a recent study found that 34.38% (11,627) of SARS-CoV-2  
398 genomes featured a single mutation capable of affecting annealing of a PCR primer in tested



399 assays from the World Health Organization, Centers for Disease Control and Prevention,  
400 National MicrobiologyData Center, and Hong Kong University<sup>36</sup>. Another study found single  
401 nucleotide mismatches in 0.2% and 0.4% of the surveyed SARS-CoV-2 sequences compared  
402 to the CDC-N1 probe and reverse primer, respectively, and 0.4% of those sequences  
403 compared to Charité's E\_Sarbeco\_R primer<sup>34</sup>. Therefore, a strategy that can target multiple  
404 genomic regions may have utility in sensitive detection of SARS-CoV-2.

405 Extensive, well-designed studies have assessed the analytical sensitivity and  
406 efficiency of existing RT-qPCR primer-probes sets<sup>34,37-39</sup> and explored adaptation of such  
407 assays to the ddPCR platform<sup>40</sup>. In this study, we describe how some of the available  
408 diagnostic assays compare to our novel SARS-CoV-2 assays and report how a multi-assay  
409 approach using the ddPCR platform could significantly advance our understanding of SARS-  
410 CoV-2 transcription and replication. While highly-sensitive PCR-based assays might not be  
411 essential to identify SARS-CoV-2-infected individuals in the transmissible/contagious phase  
412 of infection, quantitative assays capable of detecting very low copies of SARS-CoV-2 will be  
413 particularly useful in understanding the course of infection and correlates of disease  
414 progression. Existing clinical assays are quite sensitive for detecting COVID-19 during the first  
415 several weeks of infection, but often become negative after 2-3 weeks of infection<sup>41-43</sup>. In  
416 some cases, individuals who test positive may have a subsequent negative test followed by  
417 another positive or alternating positive and negative tests<sup>44,45</sup>. Some individuals may also  
418 have prolonged viral shedding after symptomatic relief, with one study noting a patient with  
419 qRT-PCR positivity detected in upper respiratory tract samples 83 days post-symptom onset<sup>46</sup>.  
420 Therefore, sensitive assays such as those described in the study could be of great utility in  
421 studying the course of infection two or more weeks after the resolution of acute symptoms.  
422 Another advantage of the approach described here is that it permits a single sample to be  
423 simultaneously assayed for multiple targets, which may increase sensitivity and specificity  
424 while helping to delineate the transcriptional profile of SARS-CoV-2 in infected patient  
425 samples. As such, this panel of assays can be applied to a diverse range of clinically relevant  
426 samples in which SARS-CoV-2 RNA may be in low or high abundance.



427           Using both the virion standard and clinical samples from the nasopharynx, we tended  
428 to observe higher copy numbers for targets at the 3' end of the genome (N, 3'UTR) compared  
429 to the 5' end (5'UTR, main protease). This discrepancy is not explained by differences in PCR  
430 efficiency, since the efficiency of the N assay on plasmid DNA was actually lower than that of  
431 assays for the 5'UTR or main protease. It is possible that reverse transcription is more efficient  
432 for assays at the 3' end (perhaps due to more efficient reverse transcription from the poly-dT),  
433 although random hexamers should bias towards the 5' end and the combination has been  
434 used to prevent bias towards either the 5' or 3' end of the 9.6kb genome of HIV-1<sup>30,47</sup>. It is also  
435 possible that the 3' assays measure higher copies because they are detecting subgenomic  
436 RNAs generated by infected cells and not packaged into virions<sup>48</sup>, which may have been  
437 present in the supernatant used to prepare the virion standard if they were released from dying  
438 cells or present in low levels in exosomes. This excess of targets corresponding to sgRNAs  
439 may be much greater in samples that contain more cells or cell-associated RNA, and it has  
440 important implications for clinical testing and research. For targets in the 3' third of the genome  
441 that are transcribed as sgRNAs, regions that are further downstream (3') may be incorporated  
442 into a greater variety of sgRNAs and therefore should be present at higher copy numbers, so  
443 assays targeting these regions may be more sensitive to detect infection<sup>8</sup>. On the other hand,  
444 sgRNAs are not infectious, so assays targeting more 5' regions that are transcribed only as  
445 genomic RNA (ORF 1a and 1b) may correlate better with infectivity.

446           The clinical implications of SARS-CoV-2 subgenomic RNA transcription are currently  
447 unknown. The synthesis of subgenomic RNAs is a common strategy employed by positive-  
448 sense RNA viruses to transcribe their 3' proximal genes that encode products essential for  
449 particle formation and pathogenesis<sup>49-51</sup>. In coronaviruses such as mouse hepatitis virus  
450 (MHV), the synthesis of subgenomic RNAs may function as important mediators of positive  
451 strand synthesis<sup>52</sup>, and more broadly, members of the order *Nidovirales* (including  
452 *Coronaviridae*) feature high levels of redundancy to ensure continued protein synthesis even  
453 in the event of point mutations in regulatory sequences<sup>53</sup>. The characterization of the SARS-  
454 CoV-2 transcription profile in differing patient samples over the course of infection may provide

455 insight into the molecular mechanisms by which SARS-CoV-2 regulates gene expression  
456 through differential transcription of genomic and subgenomic RNAs, and how this differential  
457 gene expression may contribute to pathogenesis.

458 We found that our assays performed better in the presence of background RNA,  
459 irrespective of origin (blood or epithelial cells, Fig. 6). This finding accords with other studies  
460 that have extensively validated the effect of differing variables on RT efficiency and suggest  
461 that the presence of some background RNA may increase efficiency of the reverse  
462 transcription step<sup>54-56</sup>. While the efficiency of our assays tended to decrease with RNA  
463 concentrations above 100ng/ $\mu$ L RT, even at 500ng RNA/ $\mu$ L RT, these assays still performed  
464 better than in the absence of any background RNA, suggesting that they are ideally suited for  
465 testing samples from different tissues where the levels of genomic RNA may differ  
466 considerably. Furthermore, our comparison of viral loads obtained by RT-ddPCR and qRT-  
467 PCR demonstrates the strong correlation between data obtained from these two platforms  
468 and the minimal RNA input required to yield robust data using our RT-ddPCR assays.

469 Limitations of this study should be acknowledged. In order to test our assays in parallel  
470 with published assays (total of 11 assays) in background RNA experiments (Fig. 7), we  
471 increased RT reaction volumes from 50-70  $\mu$ L to 125  $\mu$ L to accommodate the additional  
472 assays. In the absence of background RNA, the efficiency appeared to be higher in the 50-  
473 70 $\mu$ L RT reactions (Fig. 4-5, >100% efficiency for all assays) than the 125 $\mu$ L reactions (Fig.  
474 7; median efficiency=88% [range: 60-133%]). If the discrepancy is not due to a difference in  
475 the actual input of the standard, it is possible that larger reaction volumes lead to less  
476 efficiency in reverse transcription. However, for application to patient samples, our core panel  
477 of 7 assays (Table 1) is sufficient to provide a detailed view of the transcription profile of SARS-  
478 CoV-2, so preparation of RT reactions >70 $\mu$ L will likely be unnecessary.

479 For our study of the viral transcription profile and correlation with the  $C_t$  value as  
480 determined by the Abbott SARS-CoV-2 Real Time Assay, a limited amount of nucleic acid  
481 was available from only a small number of de-identified individuals. Despite this small sample  
482 size, we demonstrated both the sensitivity of all assays in our panel and their strong correlation

483 with  $C_t$  values in diagnostic specimens. These data allude to potential differences in the  
484 transcription dynamics of SARS-CoV-2 during the course of infection and merit further  
485 investigation.

486

## 487 **Conclusions**

488 We developed a panel of sensitive, quantitative RT-ddPCR-based SARS-CoV-2 assays  
489 that collectively span the genome and target nongenic and genic regions, genes encoding for  
490 important enzymes and structural proteins, and genes found in different subgenomic RNAs.  
491 These assays can serve as novel molecular tools to investigate SARS-CoV-2 infection,  
492 replication dynamics, and gene expression to better understand the viral dynamics and  
493 pathogenesis of SARS-CoV-2 over the course of infection. Future studies employing these  
494 assays will enhance our understanding of SARS-CoV-2 replication and transcription and may  
495 also inform the development of improved diagnostic tools and therapeutics.

496

497 **Additional Information:**

498 **Funding:**

499 This research was supported by funds from the Emergency COVID-19 Research Seed  
500 Funding of the University of California (Grant Number R00RG3113 [ST]). The investigators  
501 received salary support from the U.S. Department of Veterans Affairs (SAY and JKW), the  
502 National Institute of Diabetes and Digestive and Kidney Diseases at the NIH (R01DK108349  
503 [SAY, JKW], R01DK120387 [SAY]), the National Institute of Allergy and Infectious Diseases  
504 at the NIH (R01AI132128 [SAY, JKW]), the UCSF/GIVI Center for AIDS Research (CFAR;  
505 Grant# P30 AI027763 [ST]; award #A120163 [PI: Paul Volberding]), and the California  
506 HIV/AIDS Research Program (Grant number BB19-SF-009 [ST]). The funders had no role in  
507 study design, data collection and analysis, decision to publish, or preparation of the  
508 manuscript.

509

510 **Ethics**

511 This study included the use of de-identified nucleic acid from three SARS-CoV-2-  
512 infected individuals. The study authors had no subject contact or access to any personally-  
513 identifiable information (Category 4, IRB exempt).

514

515 **CRedit authorship contribution statement**

516 **Sushama Telwate:** Conceptualization, Data curation, Formal analysis, Funding acquisition,  
517 Investigation, Methodology, Supervision, Validation, Visualization, Writing - original draft.

518 **Nitasha Kumar:** Investigation, Writing - review & editing. **Chuanyi M. Lu:** Resources, Writing  
519 - review & editing. **Alberto Vallejo-Gracia:** Investigation, Writing - review & editing. **G.**

520 **Renuka Kumar:** Investigation, Writing - review & editing. **Melanie Ott:** Resources,  
521 Supervision, Writing - review & editing. **Joseph K. Wong:** Resources, Supervision, Writing -

522 review & editing. **Steven A. Yuki:** Conceptualization, Funding acquisition, Methodology,  
523 Resources, Supervision, Writing - original draft.

524

525 **Conflicts of interest statement**

526 The authors declare that they have no competing interests.

527

## 528 **References**

- 529 1. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia  
530 in China, 2019. *N Engl J Med* 2020;382:727-33.
- 531 2. Kim JM, Chung YS, Jo HJ, et al. Identification of Coronavirus Isolated from a Patient in  
532 Korea with COVID-19. *Osong Public Health Res Perspect* 2020;11:3-7.
- 533 3. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel  
534 coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395:565-74.
- 535 4. Zhou H, Chen X, Hu T, et al. A Novel Bat Coronavirus Closely Related to SARS-CoV-2  
536 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Curr Biol*  
537 2020;30:2196-203.e3.
- 538 5. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of  
539 SARS-CoV-2. *Nat Med* 2020;26:450-2.
- 540 6. Malaiyan J, Arumugam S, Mohan K, Radhakrishnan GG. An update on origin of SARS-  
541 CoV-2: Despite closest identity, bat (RaTG13) and Pangolin derived Coronaviruses varied in  
542 the critical binding site and O-linked glycan residues. *J Med Virol* 2020.
- 543 7. Sawicki SG, Sawicki DL, Siddell SG. A Contemporary View of Coronavirus  
544 Transcription. *J Virol* 2007;81:20.
- 545 8. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The Architecture of SARS-CoV-2  
546 Transcriptome. *Cell* 2020;181:914-21.e10.
- 547 9. Sola I, Almazán F, Zúñiga S, Enjuanes L. Continuous and Discontinuous RNA Synthesis  
548 in Coronaviruses. *Annu Rev Virol* 2015;2:265-88.
- 549 10. Snijder EJ, Limpens R, de Wilde AH, et al. A unifying structural and functional model  
550 of the coronavirus replication organelle: Tracking down RNA synthesis. *PLoS Biol*  
551 2020;18:e3000715.
- 552 11. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory  
553 disease in China. *Nature* 2020;579:265-9.
- 554 12. Sethna PB, Hofmann MA, Brian DA. Minus-strand copies of replicating coronavirus  
555 mRNAs contain antileaders. *J Virol* 1991;65:320-5.
- 556 13. Sawicki SG, Sawicki DL. Coronaviruses use discontinuous extension for synthesis of  
557 subgenome-length negative strands. *Adv Exp Med Biol* 1995;380:499-506.
- 558 14. Woo PCY, Lau SKP, Chu C-m, et al. Characterization and complete genome sequence  
559 of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol*  
560 2005;79:884-95.
- 561 15. Cowley TJ, Long SY, Weiss SR. The murine coronavirus nucleocapsid gene is a  
562 determinant of virulence. *J Virol* 2010;84:1752-63.
- 563 16. Kopecky-Bromberg SA, Martínez-Sobrido L, Frieman M, Baric RA, Palese P. Severe  
564 acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and  
565 nucleocapsid proteins function as interferon antagonists. *J Virol* 2007;81:548-57.
- 566 17. Huang C, Lokugamage KG, Rozovics JM, Narayanan K, Semler BL, Makino S. SARS  
567 coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs:  
568 viral mRNAs are resistant to nsp1-induced RNA cleavage. *PLoS Pathog* 2011;7:e1002433.
- 569 18. Kamitani W, Narayanan K, Huang C, et al. Severe acute respiratory syndrome  
570 coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA  
571 degradation. *Proceedings of the National Academy of Sciences of the United States of*  
572 *America* 2006;103:12885-90.

- 573 19. Narayanan K, Huang C, Lokugamage K, et al. Severe acute respiratory syndrome  
574 coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in  
575 infected cells. *J Virol* 2008;82:4471-9.
- 576 20. Kamitani W, Huang C, Narayanan K, Lokugamage KG, Makino S. A two-pronged  
577 strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nat Struct*  
578 *Mol Biol* 2009;16:1134-40.
- 579 21. Sztuba-Solińska J, Stollar V, Bujarski JJ. Subgenomic messenger RNAs: mastering  
580 regulation of (+)-strand RNA virus life cycle. *Virology* 2011;412:245-55.
- 581 22. Korber B, Fischer WM, Gnanakaran S, et al. Tracking Changes in SARS-CoV-2 Spike:  
582 Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 2020;182:812-27.e19.
- 583 23. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Velesler D. Structure, Function,  
584 and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 2020;181:281-92.e6.
- 585 24. Böttcher-Friebertshäuser E, Garten W, Matrosovich M, Klenk HD. The hemagglutinin:  
586 a determinant of pathogenicity. *Curr Top Microbiol Immunol* 2014;385:3-34.
- 587 25. Monne I, Fusaro A, Nelson MI, et al. Emergence of a highly pathogenic avian  
588 influenza virus from a low-pathogenic progenitor. *J Virol* 2014;88:4375-88.
- 589 26. Horimoto T, Rivera E, Pearson J, et al. Origin and molecular changes associated with  
590 emergence of a highly pathogenic H5N2 influenza virus in Mexico. *Virology* 1995;213:223-  
591 30.
- 592 27. Nao N, Yamagishi J, Miyamoto H, et al. Genetic Predisposition To Acquire a Polybasic  
593 Cleavage Site for Highly Pathogenic Avian Influenza Virus Hemagglutinin. *mBio* 2017;8.
- 594 28. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools  
595 APIs in 2019. *Nucleic Acids Res* 2019;47:W636-W41.
- 596 29. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new  
597 generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389-402.
- 598 30. Yukl SA, Kaiser P, Kim P, et al. HIV latency in isolated patient CD4(+) T cells may be  
599 due to blocks in HIV transcriptional elongation, completion, and splicing. *Sci Transl Med*  
600 2018;10.
- 601 31. 2019-Novel Coronavirus (2019-nCoV) Real-time RT-PCR Primers and Probes. National  
602 Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases, 2020.  
603 (Accessed March 14, 2020, at [https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-](https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html)  
604 [panel-primer-probes.html](https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html).)
- 605 32. Protocol: Real-time RT-PCR assays for the detection of SARS-CoV-2. Institut Pasteur,  
606 2020. (Accessed July 6, 2020, at [https://www.who.int/publications/m/item/molecular-](https://www.who.int/publications/m/item/molecular-assays-to-diagnose-covid-19-summary-table-of-available-protocols)  
607 [assays-to-diagnose-covid-19-summary-table-of-available-protocols](https://www.who.int/publications/m/item/molecular-assays-to-diagnose-covid-19-summary-table-of-available-protocols).)
- 608 33. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-  
609 nCoV) by real-time RT-PCR. *Euro Surveill* 2020;25:2000045.
- 610 34. Vogels CBF, Brito AF, Wyllie AL, et al. Analytical sensitivity and efficiency  
611 comparisons of SARS-CoV-2 RT-qPCR primer-probe sets. *Nat Microbiol* 2020.
- 612 35. Arvia R, Sollai M, Pierucci F, Urso C, Massi D, Zakrzewska K. Droplet digital PCR  
613 (ddPCR) vs quantitative real-time PCR (qPCR) approach for detection and quantification of  
614 Merkel cell polyomavirus (MCPyV) DNA in formalin fixed paraffin embedded (FFPE)  
615 cutaneous biopsies.
- 616 36. Penarrubia AL, Ruiz M, Porco R, et al. Multiple assays in a real-time RT-PCR SARS-  
617 CoV-2 panel can mitigate the risk of loss of sensitivity by new genomic variants during the  
618 COVID-19 outbreak. *Int J Infect Dis* 2020.



- 619 37. Uhteg K, Jarrett J, Richards M, et al. Comparing the analytical performance of three  
620 SARS-CoV-2 molecular diagnostic assays. *J Clin Virol* 2020;127:104384.
- 621 38. Etievant S, Bal A, Escuret V, et al. Performance Assessment of SARS-CoV-2 PCR  
622 Assays Developed by WHO Referral Laboratories. *J Clin Med* 2020;9.
- 623 39. Kudo E, Israelow B, Vogels CBF, et al. Detection of SARS-CoV-2 RNA by multiplex RT-  
624 qPCR. *bioRxiv* 2020:2020.06.16.155887.
- 625 40. Suo T, Liu X, Feng J, et al. ddPCR: a more accurate tool for SARS-CoV-2 detection in  
626 low viral load specimens. *Emerg Microbes Infect* 2020;9:1259-68.
- 627 41. Xiao AT, Tong YX, Zhang S. Profile of RT-PCR for SARS-CoV-2: a preliminary study  
628 from 56 COVID-19 patients. *Clin Infect Dis* 2020.
- 629 42. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult  
630 inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* (London,  
631 England) 2020;395:1054-62.
- 632 43. Zheng S, Fan J, Yu F, et al. Viral load dynamics and disease severity in patients  
633 infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective  
634 cohort study. *Bmj* 2020;369:m1443.
- 635 44. Lan L, Xu D, Ye G, et al. Positive RT-PCR Test Results in Patients Recovered From  
636 COVID-19. *JAMA* 2020;323:1502-3.
- 637 45. Alvarez-Moreno CA, Rodríguez-Morales AJ. Testing Dilemmas: Post negative, positive  
638 SARS-CoV-2 RT-PCR - is it a reinfection? *Travel Med Infect Dis* 2020;35:101743-.
- 639 46. Li N, Wang X, Lv T. Prolonged SARS-CoV-2 RNA shedding: Not a rare phenomenon. *J*  
640 *Med Virol* 2020.
- 641 47. Telwatte S, Lee S, Somsouk M, et al. Gut and blood differ in constitutive blocks to  
642 HIV transcription, suggesting tissue-specific differences in the mechanisms that govern HIV  
643 latency. *PLoS Pathog* 2018;14:e1007357.
- 644 48. Wölfel R, Corman VM, Guggemos W, et al. Virological assessment of hospitalized  
645 patients with COVID-2019. *Nature* 2020;581:465-9.
- 646 49. Koev G, Miller WA. A positive-strand RNA virus with three very different subgenomic  
647 RNA promoters. *J Virol* 2000;74:5988-96.
- 648 50. Pasternak AO, Spaan WJ, Snijder EJ. Nidovirus transcription: how to make sense...? *J*  
649 *Gen Virol* 2006;87:1403-21.
- 650 51. Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. Nidovirales: evolving the largest  
651 RNA virus genome. *Virus Res* 2006;117:17-37.
- 652 52. Baric RS, Yount B. Subgenomic negative-strand RNA function during mouse hepatitis  
653 virus infection. *J Virol* 2000;74:4039-46.
- 654 53. Di H, Madden JC, Jr., Morantz EK, et al. Expanded subgenomic mRNA transcriptome  
655 and coding capacity of a nidovirus. *Proc Natl Acad Sci U S A* 2017;114:E8895-e904.
- 656 54. Kuang J, Yan X, Genders AJ, Granata C, Bishop DJ. An overview of technical  
657 considerations when using quantitative real-time PCR analysis of gene expression in human  
658 exercise research. *PLoS One* 2018;13:e0196438-e.
- 659 55. Miranda JA, Steward GF. Variables influencing the efficiency and interpretation of  
660 reverse transcription quantitative PCR (RT-qPCR): An empirical study using Bacteriophage  
661 MS2. *J Virol Methods* 2017;241:1-10.
- 662 56. Levesque-Sergerie J-P, Duquette M, Thibault C, Delbecchi L, Bissonnette N. Detection  
663 limits of several commercial reverse transcriptase enzymes: impact on the low- and high-  
664 abundance transcript levels assessed by quantitative RT-PCR. *BMC Mol Biol* 2007;8:93-  
665



666 **Tables and Figures**

667 **Table 1. SARS-Cov2 ddPCR assay panel for assessing patient samples**

668

Assay Name	RNA Target	Detects
<b>5'UTR</b>	5' untranslated region	Genomic RNA
<b>Main Proteinase-NSP5</b>	Main Proteinase	Genomic RNA
<b>RdRp</b>	RNA-dependent RNA Polymerase	Genomic RNA
<b>S-PBCS</b>	<u>P</u> oly <u>b</u> asic <u>c</u> leavage <u>s</u> ite of the surface (S) glycoprotein	Genomic/subgenomic
<b>M-ORF5</b>	Membrane glycoprotein	Genomic/subgenomic
<b>N-ORF9</b>	Nucleocapsid	Genomic/subgenomic
<b>3'UTR</b>	3' untranslated region	Genomic/subgenomic

669

670

671 **Table 2. SARS-CoV-2 primer/probe sets selected for validation using IVT and virion**  
672 **RNA**

Target Region	Primer Name <sup>a</sup>	SARS-CoV-2 coordinates <sup>b</sup>	Sequence (5'-3')
<b>5'UTR</b>			
	5'UTR_F	152-171	GTTGACAGGACACGAGTAAC
	5'UTR_P	175-197	TCTATCTTCTGCAGGCTGCTTAC
	5'UTR_R	220-241	GAAACCTAGATGTGCTGATGAT
<b>Main proteinase/NSP5 (ORF1a)</b>			
	NSP5_F	10366-10387	TCGCATTCAACCAGGACAGACT
	NSP5_P	10399-10425	AGCTTGTTACAATGGTTCACCATCTGG
	NSP5_R	10426-10450	GGGCCTCATAGCACATTGGTAAACA
<b>RNA-dependent RNA polymerase / NSP12 (ORF1b)</b>			
	RDRP_F	15341-15364	CCTCACTTGTTCTTGCTCGCAAAC
	RDRP_P	15370-15393	ACGTGTTGTAGCTTGTACACCCGT
	RDRP_R	15437-15456	TGAACCGCCACACATGACCA
<b>S protein/ polybasic cleavage site (ORF 2)</b>			
	S_PBCS_F	23554-23576	ACCCATTGGTGCAGGTATATGCG
	S_PBCS_P	23603-23622	ACACTACGTGCCCCGCCGAGG*
	S_PBCS_R	23641-23664	GCACCAAGTGACATAGTGTAGGCA
<b>M protein (ORF 5)</b>			
	M-ORF5_F	26768- 26789	CGCAATGGCTTGTCTTGTAGGC
	M-ORF5_P	26794-26816	TGTGGCTCAGCTACTTCATTGCT
	M-ORF5_R	26821-26840	CGTACGCGCAAACAGTCTGA
<b>N protein (ORF 9)</b>			
	N-ORF9_F	28833-28851	CATCACGTAGTCGCAACAG
	N-ORF9_P	28885-28907	AACTTCTCCTGCTAGAATGGCTG
	N-ORF9_R	28917-28934	AAGCAAGAGCAGCATCAC
<b>3'UTR</b>			
	3'UTR_F	29702-29723	GGAGGACTTGAAAGAGCCACCA
	3'UTR_P	29727-29746	TTTCACCGAKGCCACRCGGA
	3'UTR_R	29768-29788	GGCAGCTCTCCCTAGCATTGT

673 \*Reverse complement

674 <sup>a</sup> 'F' = forward primer, 'R' = reverse primer, 'P' = probe (fluorophore/quencher: FAM, MGB)

675 <sup>b</sup> SARS-CoV2 coordinates indicated are based on the SARS-CoV2 reference sequence  
676 (NC\_045512.2)

677

678 **Table 3. SARS-CoV-2 assays from other sources**

Target Region	Primer Name <sup>a</sup>	SARS-CoV-2 coordinates <sup>b</sup>	Sequence (5-3')	Reference
N protein/ ORF 9	CDC_N1_F	28287-28306	GAC CCC AAA ATC AGC <u>G</u> AA AT	31
	CDC_N1_P	28309-28330	AC <u>C</u> CCG CAT TAC GTT TGG TGG ACC	
	CDC_N1_R	28335- 28358	TCT GGT TAC TGC CAG <u>G</u> TTG AAT CTG	
	CDC_N2_F	29164-29183	TTACAAACATTGGCCGCAAA	
	CDC_N2_P	29188- 29210	ACAATTTGCCCCAGCGCTTCAG	
	CDC_N2_R	29213-29230	GCG CGA CAT TCC GAA GAA	
ORF1a	nCoV_IP2-12669Fw	12690-12707	ATGAGCTTAGTCCTGTTG	32
	nCoV_IP2-12696bProbe(+)	12717- 12737	AGATGTCTTGTGCTGCCGGTA	
	nCoV_IP2-12759Rv	12780- 12797	CTCCCTTTGTTGTGTTGT	
E gene	E_Sarbeco_F1	26269-26394	ACAGGTACGTTAATAGTTAATAGCGT	33
	E_Sarbeco_P1	26332- 26357	ACACTAGCCA <u>I</u> CCTTACTGCGCT <u>I</u> CG	
	E_Sarbeco_R2	26360-26381	ATATTGCAGCA <u>G</u> TACGCACACA	

679 <sup>a</sup> 'F' = forward primer, 'R' = reverse primer, 'P'= probe (fluorophore/quencher: FAM, MGB)

680 <sup>b</sup> SARS-CoV-2 coordinates indicated are based on the SARS-CoV-2 reference sequence  
681 (NC\_045512.2)

682 **Bold and underlined** = known mismatches as reported in <sup>34</sup> and in-house SARS-CoV2

683 multiple sequence alignment (mismatches identified were relative to sequence MT825091.1  
684 from Iran)

685

686 **Figure 1. Schematic presentation of SARS-CoV2 genome organization, virion**  
687 **structure and canonical sgRNAs.** SARS-CoV-2 encodes two large genes, ORF1a (yellow)  
688 and ORF1b (blue), which encode 16 non-structural proteins (NSP1–NSP16). The structural  
689 genes encode the structural proteins, spike (S; green), envelope (E; blue), membrane (M;  
690 purple), and nucleocapsid (N; gold). Assay locations of each assay designed for this study  
691 are indicated. Virion structure and canonical subgenomic RNAs produced by SARS-CoV-2  
692 are shown in the lower panel (S, 3a, E, M, 6, 7a, 7b, 8 and N).

693

694 **Figure 2. Efficiency and linearity of SARS-CoV-2 panel of ddPCR assays determined**  
695 **using plasmid DNA.** Plasmids containing individual SARS-CoV-2 genes or regions were  
696 quantified by UV spectroscopy and diluted (expected copies) to test the absolute number of  
697 copies detected by each primer/probe set using duplicate ddPCR reactions (measured  
698 copies). Two primer/probe sets were tested for each region except the S-PBCS. One  
699 primer/probe set from each region (indicated by coloured symbol) was selected for  
700 subsequent experiments.

701

702 **Figure 3. Efficiency and linearity of SARS-CoV-2 panel of ddPCR assays determined**  
703 **using in vitro transcribed (IVT) RNA.** RNA standards containing a given region or gene of  
704 SARS-CoV-2 were prepared by in vitro transcription from plasmids and quantified by  
705 independent means (UV spectroscopy and the Agilent Bioanalyzer). Various inputs of each  
706 IVT RNA standard (which were used to calculate 'Expected Copies' per ddPCR well) were  
707 reverse transcribed and replicate aliquots of cDNA were used to measure the absolute number  
708 of copies detected by each ddPCR assay ('Measured Copies'). Each assay was tested using  
709 expected inputs of  $1-10^4$  copies per ddPCR well (except S-PBCS, which was tested at inputs  
710 of 2-2100 copies). Data represent average of duplicate wells from a representative  
711 experiment. S=slope, indicating assay efficiency. Each assay was tested in at least two  
712 independent experiments.

713

714 **Figure 4. Efficiency and linearity of SARS-CoV2 panel of ddPCR assays determined**  
715 **using SARS-CoV-2 virion RNA.** A SARS-CoV-2 “virion” standard was prepared by  
716 extracting the RNA from the supernatant of an in vitro infection and quantified using the Abbott  
717 Real Time SARS-CoV-2 assay. Various inputs of the virion standard (which were used to  
718 calculate ‘Expected Copies’ per ddPCR well) were applied to a common reverse transcription  
719 reaction, from which aliquots of cDNA were used to measure the absolute number of copies  
720 detected by each ddPCR assay (measured copies). Each assay was tested with expected  
721 inputs of  $10$ - $10^4$  copies/ddPCR well in duplicate. S (slope) and  $R^2$  are indicated for each assay.  
722 Representative data for  $n=2$  independent experiments are shown.

723

724 **Figure 5. Comparison of assay efficiency and linearity of published assays, ORF1a**  
725 **“nCoV\_IP2” and E gene and novel RDRP-NSP12 assay.** The performance of our RDRP-  
726 NSP12 assay was compared to published primers/probes for ORF1a and the E gene in the  
727 ddPCR platform using common RT reactions containing virion standard RNA inputs of  $2 \times 10^4$   
728 copies/ddPCR well. S (slope) and  $R^2$  are indicated for each assay.

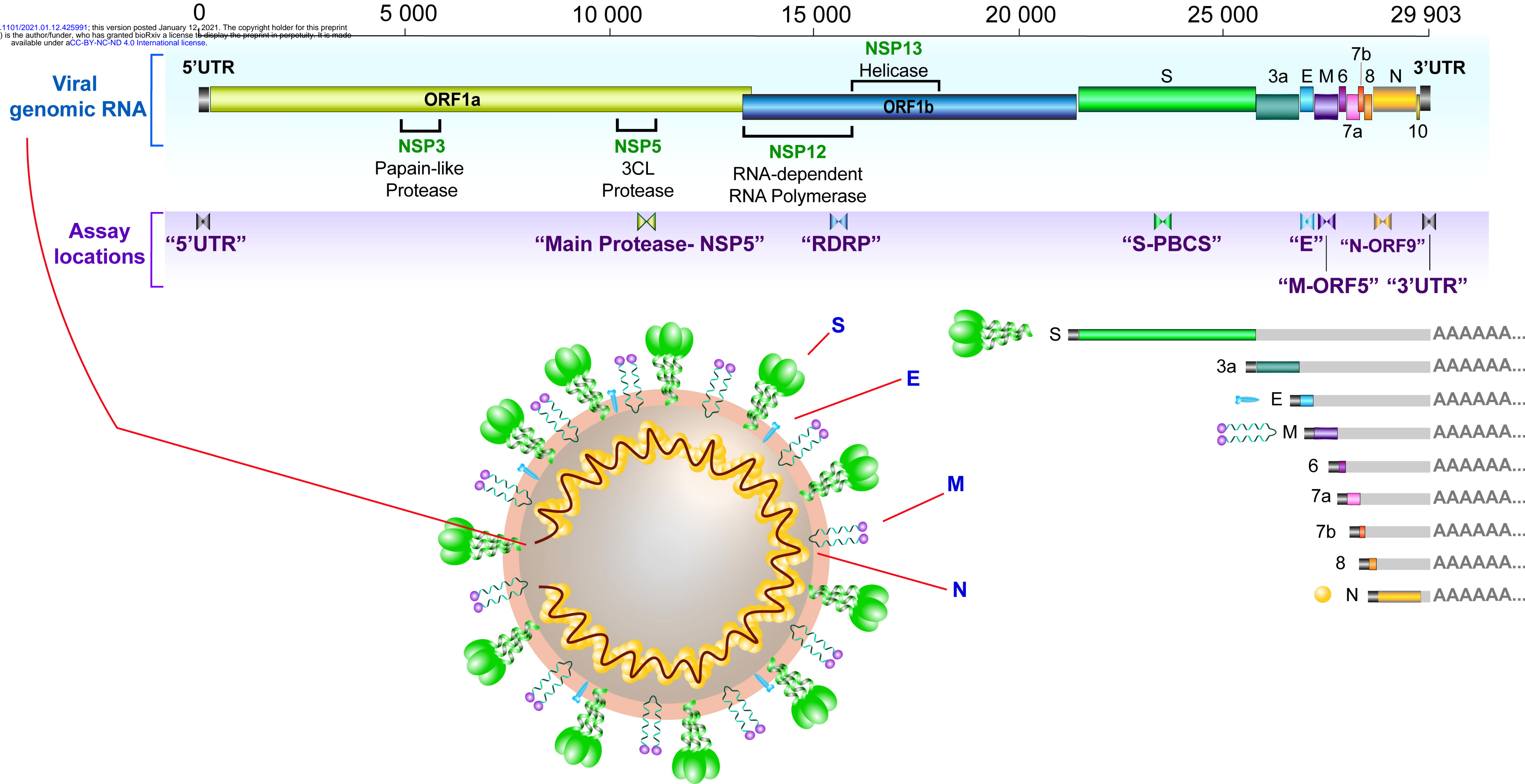
729

730 **Figure 6. Effect of background RNA on ddPCR assay performance.** We simultaneously  
731 tested all assays in our panel against reported assays, CDC\_N1, CDC\_N2, E\_Sarbeco, and  
732 IP2\_ORF1a, in the presence and absence of background RNA. Each assay was tested with  
733 a constant input of SARS-CoV-2 virion standard (predicted to yield 1000 copies/ddPCR well)  
734 in the presence or absence of background RNA from PBMC or a lung epithelial cell line (A549)  
735 added at a concentration of  $100 \text{ ng}/\mu\text{L}$  of RT reaction (500 ng/ddPCR well, or  $1 \mu\text{g}$  for the 2  
736 replicate wells used to test each assay). Negative controls included water,  $1 \mu\text{g}/\text{assay}$  PBMC  
737 RNA, and  $1 \mu\text{g}/\text{assay}$  A549 RNA. Assays are indicated on x-axis in order from 5' to 3' and  
738 dotted line indicates 1000 SARS-CoV-2 RNA copy input. Error bars represent standard  
739 deviation from duplicate wells.

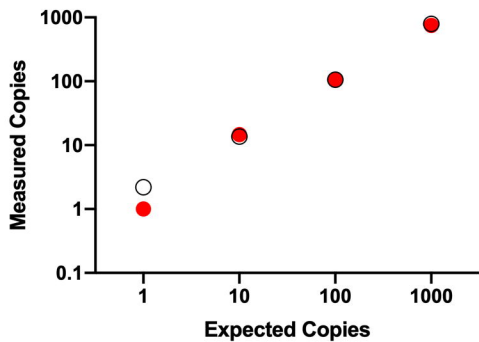
740

741 **Figure 7. Transcription profile of three SARS-CoV-2 infected individuals determined**  
742 **using RT-ddPCR.** Unused nucleic acid (ranging from 8.6-16.8  $\mu$ L) extracted by the Abbott  
743 m2000 platform from nasopharyngeal swabs from SARS-CoV-2 infected individuals (n=3)  
744 was used in a common RT reaction for each individual. Resulting cDNA was divided evenly  
745 across reactions for the seven assays in our panel and targets were measured using  
746 ddPCR. Colored symbols indicate SARS-CoV-2 target region. Copy numbers from each  
747 assay are expressed as SARS-CoV-2 copies per  $\mu$ L of nucleic acid and grouped for each  
748 individual (x-axis). Threshold cycle ( $C_t$ ) values, as determined by Abbott Real Time SARS-  
749 CoV-2 viral load assay, are indicated above each individual's dataset.



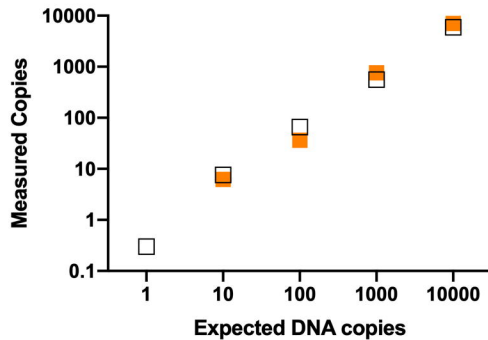


### 5'UTR



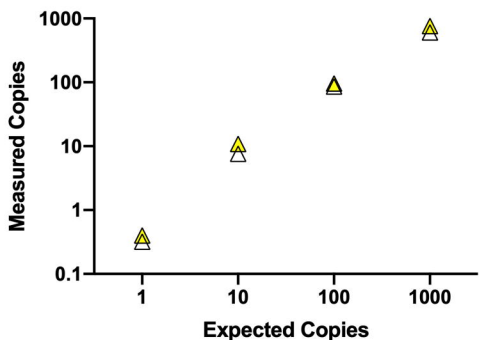
- 5'UTR\_A  
 $R^2=0.9986$   
Slope=0.74
- 5'UTR\_B  
 $R^2=0.9990$   
Slope=0.78

### Main Proteinase-NSP5



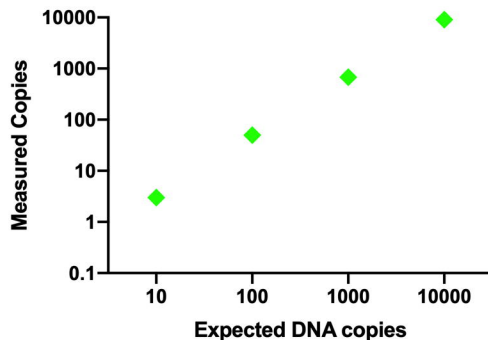
- Pro-Nsp5\_A  
 $R^2=0.9999$   
Slope=0.70
- Pro-Nsp5\_B  
 $R^2=1.0$   
Slope=0.59

### RNA-dependent RNA Polymerase



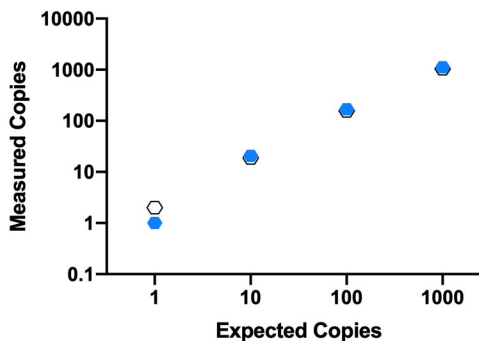
- ▲ RDRP-NSP12\_A  
 $R^2=0.9994$   
Slope=0.76
- △ RDRP-NSP12\_B  
 $R^2=0.9983$   
Slope=0.60

### S-PBCS



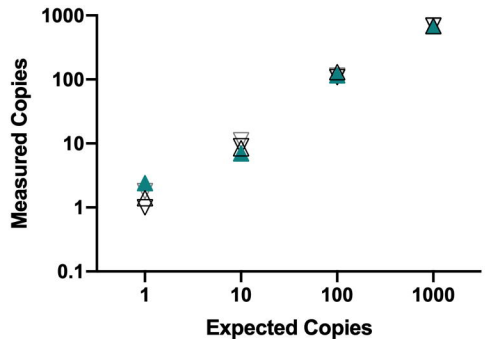
- ◆ S-PBCS\_A  
 $R^2=0.9995$   
Slope=0.91

### M-ORF5



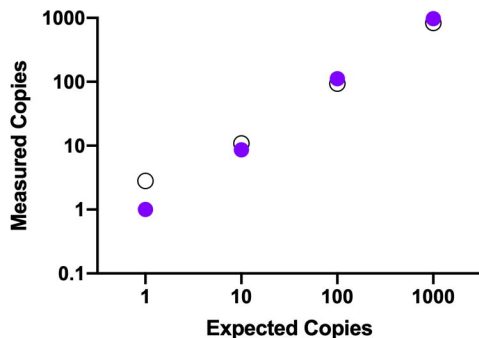
- M\_ORF5\_A  
 $R^2=0.9978$   
Slope=1.1
- M\_ORF5\_B  
 $R^2=0.9978$   
Slope=1.1

### N-ORF9



- ▲ N-ORF9\_A  
 $R^2=0.9951$   
Slope=0.67
- △ N-ORF9\_B  
 $R^2=0.9918$   
Slope=0.67
- ▽ CDC\_N1  
 $R^2=0.9940$   
Slope=0.68
- ▽ CDC\_N2  
 $R^2=0.9957$   
Slope=0.67

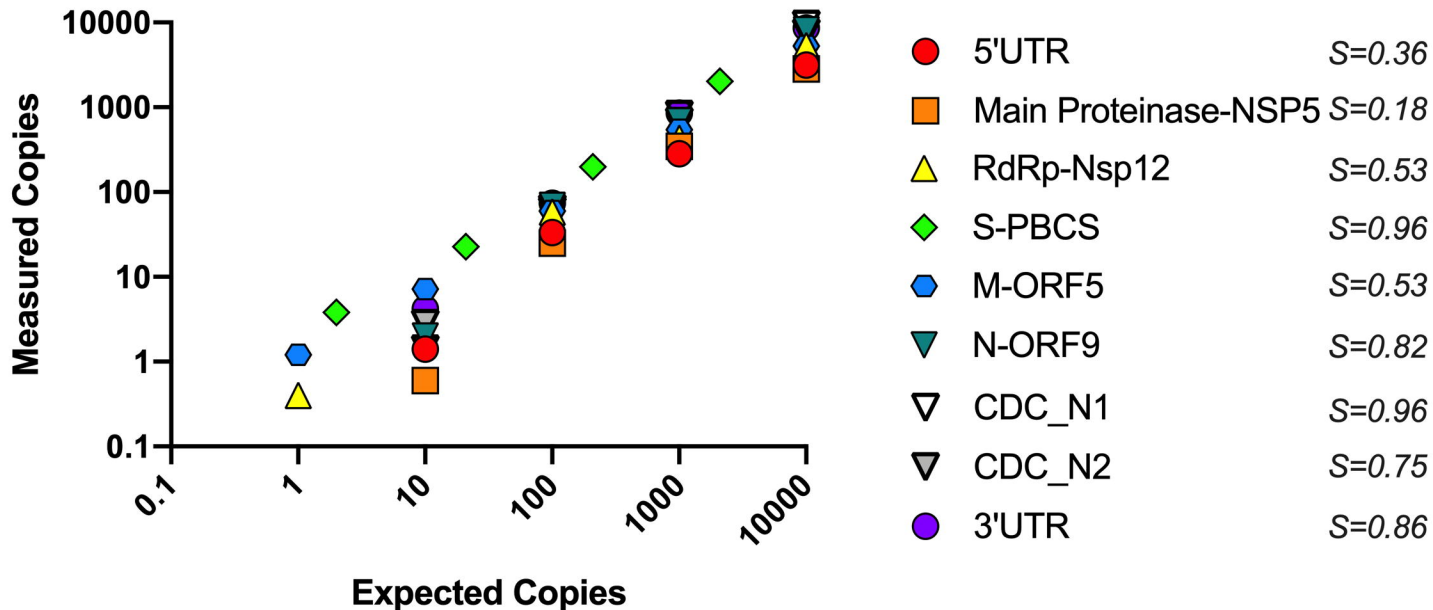
### 3'UTR

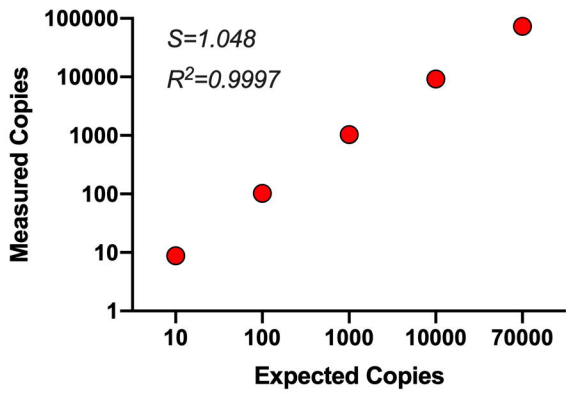
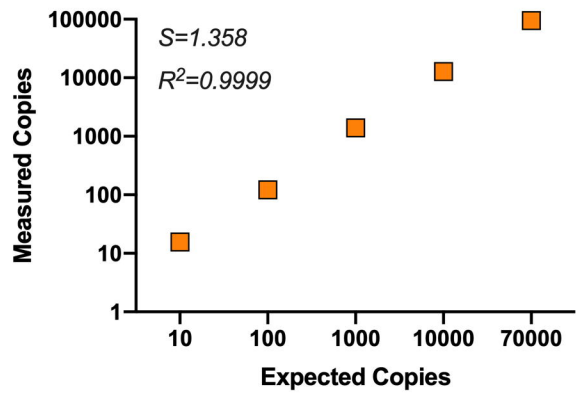
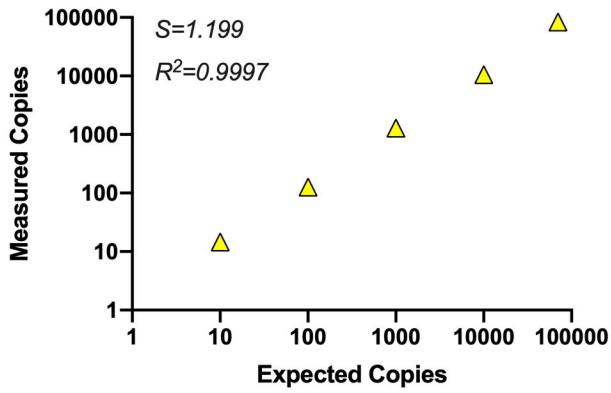
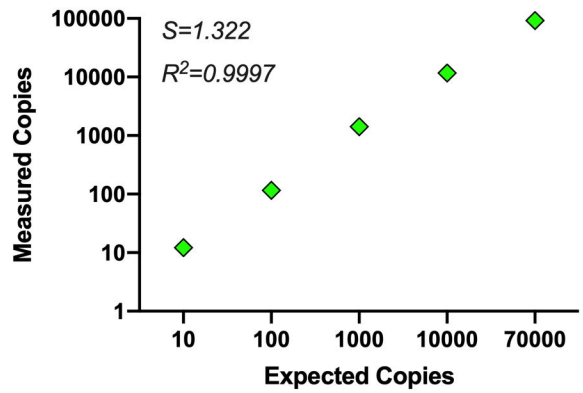
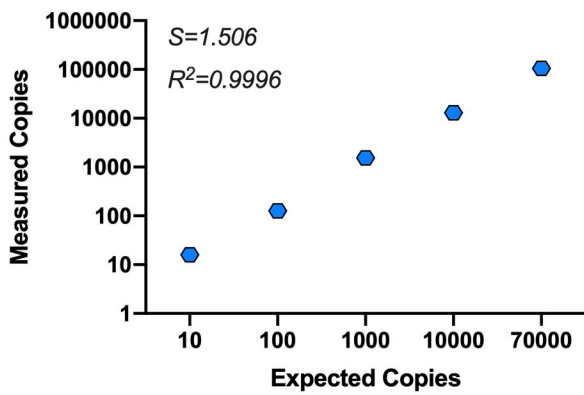
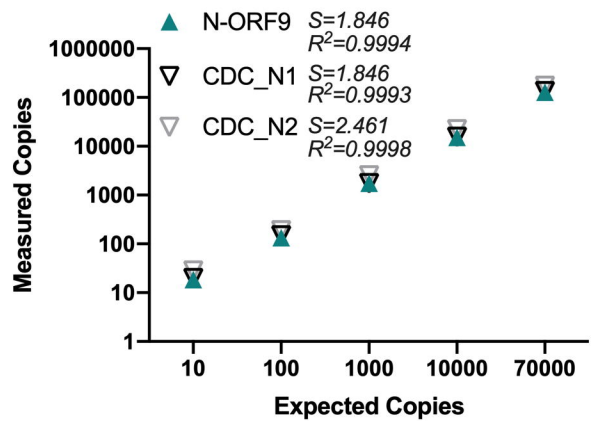
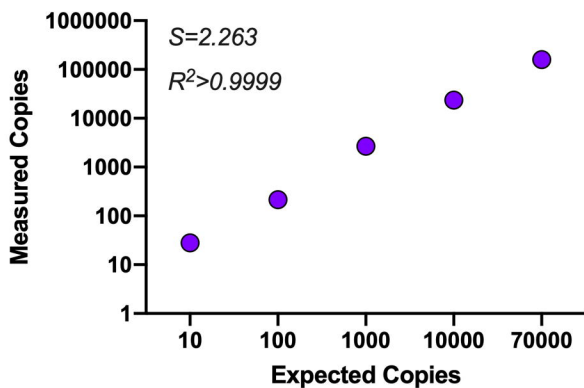


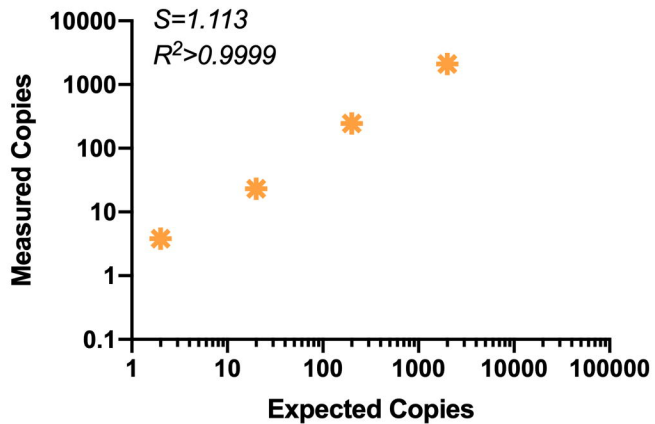
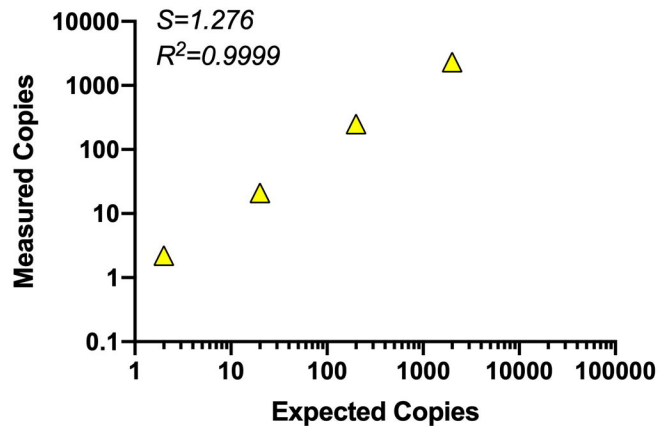
- 3'UTR\_A  
 $R^2=0.9997$   
Slope=0.97
- 3'UTR\_B  
 $R^2=0.9999$   
Slope=0.82



# Panel of SARS-CoV2 ddPCR assays



**5'UTR****Main Protease- NSP5****RdRP****S-PBCS****M-ORF5****N-ORF9****3'UTR**

**IP2\_ORF1a****RDRP\_NSP12\_4****E\_Sarbeco**