

UCSF

UC San Francisco Previously Published Works

Title

Novel RT-ddPCR assays for measuring the levels of subgenomic and genomic SARS-CoV-2 transcripts

Permalink

<https://escholarship.org/uc/item/6qv6962s>

Authors

Telwatte, Sushama
Martin, Holly Anne
Marczak, Ryan
[et al.](#)

Publication Date

2022-05-01

DOI

10.1016/j.ymeth.2021.04.011

Peer reviewed



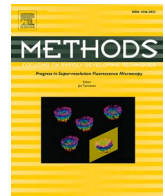
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Novel RT-ddPCR assays for measuring the levels of subgenomic and genomic SARS-CoV-2 transcripts

Sushama Telwatte^{a,b}, Holly Anne Martin^{a,b}, Ryan Marczak^c, Parinaz Fozouni^d,
Albert Vallejo-Gracia^d, G. Renuka Kumar^d, Victoria Murray^a, Sulggi Lee^a, Melanie Ott^{a,d},
Joseph K. Wong^{a,b}, Steven A. Yukl^{a,b,*}

^a Department of Medicine, University of California, San Francisco (UCSF), San Francisco, CA, United States

^b Department of Medicine, San Francisco VA Health Care System, San Francisco, CA, United States

^c University of California, Santa Barbara, CA, United States

^d Gladstone Institute of Virology, Gladstone Institutes, San Francisco, CA, United States

ARTICLE INFO

Keywords:

SARS-CoV-2

COVID-19

Coronavirus

Viral transcription/replication

Digital PCR

Droplet digital PCR

Subgenomic RNA

Quantitative assays

ABSTRACT

The replication of SARS-CoV-2 and other coronaviruses depends on transcription of negative-sense RNA intermediates that serve as the templates for the synthesis of positive-sense genomic RNA (gRNA) and multiple different subgenomic mRNAs (sgRNAs) encompassing fragments arising from discontinuous transcription. Recent studies have aimed to characterize the expression of subgenomic SARS-CoV-2 transcripts in order to investigate their clinical significance. Here, we describe a novel panel of reverse transcription droplet digital PCR (RT-ddPCR) assays designed to specifically quantify multiple different subgenomic SARS-CoV-2 transcripts and distinguish them from transcripts that do not arise from discontinuous transcription at each locus. These assays can be applied to samples from SARS-CoV-2 infected patients to better understand the regulation of SARS-CoV-2 transcription and how different sgRNAs may contribute to viral pathogenesis and clinical disease severity.

1. Introduction

1.1. Characteristics of SARS-CoV-2 transcription

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [1,2], the etiologic agent of COVID-19, is an enveloped virus with a positive-sense, single-stranded RNA genome of about 30,000 nucleotides. Like other viruses of the order *Nidovirales*, SARS-CoV-2 replicates through the transcription of negative-sense RNA intermediates that serve as the templates for the transcription of positive-sense genomic RNA (gRNA) and multiple different subgenomic RNAs (sgRNAs). These sgRNAs, which are analogous to spliced RNAs, are generated from discontinuous transcription during the synthesis of the negative strand RNA [3]. sgRNAs arise from a template switch at transcription-regulating sequences (TRS) located at the end of the 'leader' sequence in the 5' untranslated region [UTR] and 'body' TRS sequences located upstream of various genes in the distal third of the genome (open reading frames [ORF] 2 to 9) [4]. The resulting sgRNAs contain the 5'

UTR 'leader' sequence followed by the 'body' sequence derived from one of the 3' genes [5] (see Fig. 1). Although further studies are needed to elucidate the contribution of different sgRNAs to SARS-CoV-2 replication and disease, subgenomic transcription may allow for variation in expression of viral structural proteins and proteins involved in pathogenesis. The transcription of gRNA and sgRNA occurs at double-membraned vesicles that contain cellular and viral materials in the cytoplasm of infected cells [6]. While the gRNAs are packaged into virions, it seems unlikely that the sgRNAs are packaged [7,8]. However, sgRNAs are detectable during early symptomatic infection and sometimes after symptoms have subsided [8,9]. It is unclear which mechanisms allow for prolonged persistence of sgRNAs, but they might be protected from degradation by encapsulation in double-membrane vesicles [6,9].

* Corresponding author at: Department of Medicine, University of California, San Francisco (UCSF) and San Francisco VA Health Care System, 4150 Clement St, 111W3, San Francisco, CA, 94121, United States.

E-mail address: Steven.yukl@ucsf.edu (S.A. Yukl).

<https://doi.org/10.1016/j.ymeth.2021.04.011>

Received 18 January 2021; Received in revised form 12 April 2021; Accepted 14 April 2021

Available online 18 April 2021

1046-2023/Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1.2. Advantages of droplet digital PCR for assessing SARS-CoV-2 transcription

By partitioning PCR reactions into thousands of individual droplets prior to amplification and acquiring data at the reaction end point, ddPCR permits absolute quantification of targets independent of standard curves, which leads to higher precision and reproducibility compared to quantitative PCR [10]. Recent studies have attempted to measure SARS-CoV-2 sgRNAs using various strategies in cell lines [11] or cells from infected individuals [9,12]. These methods consistently indicate a low abundance of subgenomic transcripts [9,11,12]. The combination of low abundance targets (as is the case with sgRNAs), limited yields of cells or fluid from sampling procedures (e.g. nasopharyngeal swabs), and the potential presence of inhibitors such as chemical or protein contaminants [10] requires amplification methods that are sensitive, reliable, accurate, and precise. In this regard, ddPCR offers absolute quantification, improved sensitivity [10,13], improved tolerance to PCR inhibitors [14,15], increased precision (particularly for low concentration samples [16–18]), inter-run reproducibility, and enhanced discrimination between small differences in concentration [19]. Therefore, the use of RT-ddPCR offers advantages to study SARS-CoV-2 sgRNAs compared to other methods that have been recently described, such as RT-PCR [20] and ThermoFisher's Ampliseq panel. The latter contains 237 primer pairs targeting SARS-CoV-2 and 5 primer pairs for cellular targets; however, only two of the forward primers are located within the leader sequence and are likely to amplify subgenomic RNAs [9]. Furthermore, the number of cycles must be carefully calibrated because biases due to differential amplification efficiency of primer pairs can skew amplification [21]. This presents a technical challenge for samples with low viral loads that require >20 amplification cycles [9,22]. Here, we detail a protocol designed to measure SARS-CoV-2 sgRNAs using RT-ddPCR and validate the protocol by applying it

to Vero CCL-81 kidney epithelial cells infected with SARS-CoV-2.

2. Material and methods

2.1. Assay approach

We designed ddPCR assays targeting seven of the reportedly most abundant canonical sgRNAs expressed by SARS-CoV-2 [11]: spike (S) protein (ORF2), open reading frame 3a (ORF3a), envelope (E; ORF 4), membrane (M) glycoprotein (ORF5), ORF7a, ORF8, and nucleocapsid (N) protein (ORF9). For each region, we designed pairs of primer/probe sets that shared the same reverse primer and probe (located in a 3' body gene) but differed in the forward primer, either directed to the leader sequence in the 5' UTR (which is far from the probe in gRNA and therefore should only amplify sgRNA) or upstream of the body TRS (in a region that would be removed during the discontinuous transcription of that particular sgRNA) [Table 1, Fig. 1]. It should be noted that the genomic forward primer for the spike (S) gene (the most upstream of the body genes) should only detect gRNA, whereas "genomic" primers for more downstream genes will detect gRNA as well as sgRNAs arising from discontinuous transcription at other body TRS sequences located further upstream. However, the levels of each sgRNA and its corresponding "genomic" target can still be used to measure the fraction of transcripts that result from discontinuous transcription at the given locus. Of note, the "genomic" and subgenomic forward primers were chosen to be a similar distance from the probe in their respective genomic and subgenomic RNAs in order to minimize any potential difference in amplification efficiency prior to subsequent validations.

2.2. Primer design and selection

A multiple sequence alignment was performed using Clustal Omega

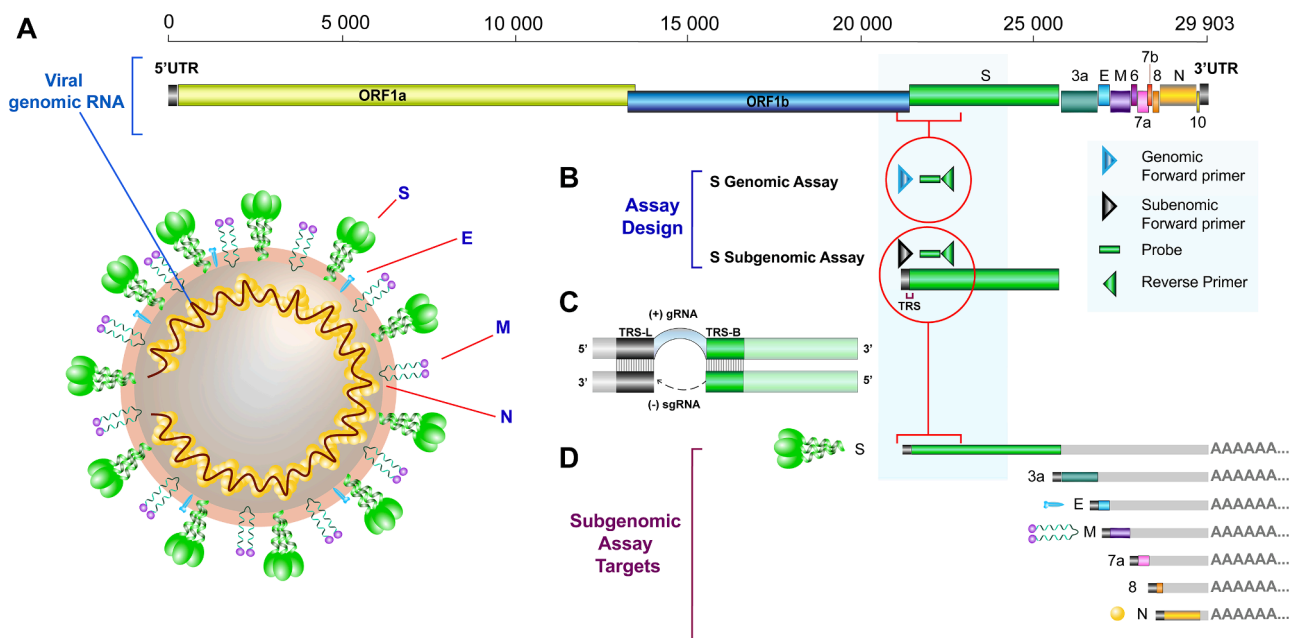


Fig. 1. Schematic representation of SARS-CoV-2 genome organization, virion structure, assay design and sgRNA targets. SARS-CoV-2 employs discontinuous transcription to generate subgenomic RNAs. (A) The genome organization of SARS-CoV-2. The genome features two large genes, ORF1a (yellow) and ORF1b (blue), which encode 16 non-structural proteins (NSP1–NSP16). The structural genes encode the structural proteins, spike (S; green), envelope (E; blue), membrane (M; purple), and nucleocapsid (N; gold). Assay locations of each assay designed for this study are indicated. The SARS-CoV-2 virion structure is shown in the lower panel. (B) Assay design. Assays for a given subgenomic RNA and the corresponding "genomic" RNA region share the same probe and reverse primer (located in a body gene) but differ in the forward primers. The sgRNA-targeting forward primer is located in the 5' UTR (upstream of the leader-body TRS junction), whereas the cognate gRNA-targeting forward primer is located upstream of the body TRS and coding region. (C) Schematic representation of negative strand RNA synthesis from the full-length positive strand gRNA. A template-switch occurs at the body TRS [TRS-B] to the 5' leader TRS [TRS-L] to give rise to the negative-strand sgRNA. (D) The panel of validated canonical subgenomic RNA targets is shown, including S, 3a, E, M, 7a, 8 and N. Figure adapted from Telwatte et al., 2021 and Kim et al., 2020. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
SARS-CoV-2 genomic and subgenomic primer/probe sets selected.

Target Region	Primer Name ^a	SARS-CoV-2 coordinates ^b	Sequence (5'-3')
Genomic S	gS_F1	21516–21541	<u>CAGAGTTGTTATTTCTAGTGTGTTT</u>
	S_P1	21585–21607	<u>TGCCACTAGTCTCTAGTCAGTGT</u>
	S_R1	21616–21637	<u>GGGTAATTGAGTCTCTGGTTGTA</u>
Subgenomic S	sgS-ORF3a_F1	35–56	<u>ACCAACTTTCGATCTCTGTAG</u>
Genomic ORF3a	gORF3a_F1	25350–25369	<u>CAGTGCTCAAAGGAGTCAAA</u>
	ORF3a_P1	25421–25445	<u>TFGGAACGTAACTTTGAAGCAAGG</u>
	ORF3a_R1	25448–25469	<u>GAAGGAGTAGCATCCTTGATT</u>
Subgenomic ORF3a	sgS-ORF3a_F1	35–56	<u>ACCAACTTTCGATCTCTGTAG</u>
Genomic E (ORF 4)	gE_F3	26217–26236	<u>GTAAGCACAAGCTGATGAGT</u>
	E_P1	26252–26275	<u>CATTGTTTCGGAAGAGACAGGTA</u>
	E_R3	26303–26323	<u>AGAATACCACGAAGCAAGAA</u>
Subgenomic E	sgE_F3	20–38	<u>CCAGGTAACAAACCAACCA</u>
Genomic M (ORF 5)	gM_F1	26452–26473	<u>GTTCTGATCTTCTGGTCTAAA</u>
	M_P1	26516–26537	<u>TTTAGCCATGGCAGATTCCAAC</u>
	M_R1	26541–26561	<u>AAGCTCTCAACGGTAATAGT</u>
Subgenomic M	sgM_F1	33–53	<u>CAACCAACTTTCGATCTCTTG</u>
Genomic ORF7a	gORF7a_F1	27361–27380	<u>GAAGAGCAACCAATGGAGAT</u>
	ORF7a_P2	27410–27431	<u>TCTTGGCACTGATAACTCTCGC</u>
	ORF7a_R2	27432–27454	<u>GGTAGTGATAAAGCTCACAAAGTA</u>
Subgenomic ORF7a	sgORF7a_F2	20–38	<u>CCAGGTAACAAACCAACCA</u>
Genomic ORF8	gORF8_F2	27877–27894	<u>GTCAGCCTAAACGAACA</u>
	ORF8_P1	27930–27953	<u>ACATTCTTGGTGAATGCAGCTAC*</u>
	ORF8_R1	27956–27978	<u>GATGTTGAGTACATGACTGTAAA</u>
Subgenomic ORF8	sgORF8-N_F1	32–51	<u>CCAACCAACTTTCGATCTCT</u>
Genomic N (ORF 9)	gN_F1	28227–28245	<u>CATGACGTTCTGTTGTTT</u>
	N_P2	28275–28299	<u>TGTCTGATAATGGACCCCAAATCA</u>
	N_R2	28316–28333	<u>GGGTCCACCAACGTAAT</u>
Subgenomic N	sgORF8-N_F1	32–51	<u>CCAACCAACTTTCGATCTCT</u>

*Reverse complement.

^a 'F' = forward primer, 'R' = reverse primer, 'P' = probe (fluorophore/quencher: FAM, MGB).

^b SARS-CoV2 coordinates indicated are based on the SARS-CoV2 reference sequence (NC_045512.2).

[23], encompassing complete sequences of 86 SARS-CoV-2 isolates from all geographical locations and all sequences available from the US on 3/14/2020 as previously described [24]. Using this alignment, we assessed sequence conservation across target regions. Multiple primer/probe sets were designed to target seven regions of SAR-CoV-2 reported to generate subgenomic in addition to genomic RNA. Where feasible, two alternative primer/probe sets were designed for a given subgenomic or "genomic" target in order to provide flexibility in choosing the better assay. Primers/probes were designed using the Primer Quest® Tool (Integrated DNA Technologies, Coralville, IA). A sequence similarity analysis using Basic Local Alignment Search Tool (BLAST) [25] found no significant similarity in any primer or probe to human sequences.

2.3. Preparation of SARS-COV-2 supernatant viral RNA

Vero CCL-81 kidney epithelial cells, derived from *Cercopithecus aethiops*, were infected with SARS-CoV-2 (Isolate: USA-WA1/2020) at an MOI of 0.003 (250 000 cells/well) [24]. After 72 h, supernatant was harvested and centrifuged twice to remove cells, as previously described [24]. The clarified supernatant from one infection was mixed with TriReagent (Molecular Research Center; 1 mL per 100 µL supernatant), and RNA was extracted using TriReagent to generate a viral RNA standard [24]. The copies/µL in the supernatant virus standard were

estimated by triplicate measurements using the Abbott RealTime SARS-CoV-2 assay (Abbott m2000 Molecular Platform), which targets the RdRp and N genes. Dilutions of the supernatant virus standard were added to reverse transcription reactions to achieve expected inputs of 1 to 10,000 copies per 5 µL RT, and 5 µL aliquots of cDNA were used to test each assay in parallel using replicate 20 µL ddPCR reactions (see Section 2.5) containing primers/probe specific for a given region.

2.4. Reverse transcription

Reverse transcription (RT) reactions were performed in 50 µL containing 5 µL of 10× SuperScript III buffer (Invitrogen), 5 µL of 50 mM MgCl₂, 2.5 µL of random hexamers (50 ng/µL; Invitrogen), 2.5 µL of 50 µM poly-dT15, 2.5 µL of 10 mM deoxynucleoside triphosphates (dNTPs), 1.25 µL of RNaseOUT (40 U/µL; Invitrogen), and 2.5 µL of SuperScript III RT (200 U/µL; Invitrogen) [26]. Key reagents and resources are listed in Table S1. Reverse transcription was performed with both random hexamers and poly-dT in anticipation of this approach being applied to clinical samples containing long polyadenylated SARS-CoV-2 RNAs, for which the combination of poly-dT plus random hexamers may reduce bias towards reverse transcription of any one region (as can be seen with specific reverse primers), the 5' end (as would be expected with random hexamers), or the 3' end (as would be expected with poly-dT).

2.5. Droplet digital PCR

Droplet digital PCR reactions consisted of 20 μ L per well containing 10 μ L of ddPCR Probe Supermix (no deoxyuridine triphosphate, Bio-Rad, Hercules, CA), 900 nM of primers, 250 nM of probe, and 5 μ L of plasmid DNA or cDNA generated from SARS-CoV-2 virion RNA [24]. The ddPCR reactions were incorporated into droplets using the QX100 Droplet Generator (Bio-Rad). Nucleic acids were amplified with the following cycling conditions: 10 min at 95 °C, 45 cycles of 30 s at 95 °C and 59 °C for 60 s, and a final droplet cure step of 10 min at 98 °C using a Mastercycler® nexus (Eppendorf, Hamburg, Germany). Droplets were read and analyzed using Bio-Rad QX100 system and QuantaSoft software in the absolute quantification mode. Only wells containing \geq 11,000 droplets were accepted for further analysis.

2.6. sgRNA expression levels in a clinical sample

2.6.1. Ethics statement

All participants provided written informed consent, and the research was approved by the Committee on Human Research (CHR), the Institutional Review Board for the University of California, San Francisco (IRB# 20–30588).

2.6.2. Participant recruitment, sample collection, and processing

Nasopharyngeal (NP) swab samples were obtained from an acutely-infected, symptomatic male with mild disease (outpatient). SARS-CoV-2 infection was confirmed by diagnostic PCR 4 days prior to collection of the study sample, while symptom onset was reported 5 days prior to the study sample. The collected NP swab was transported in 850 μ L Viral Transport Medium (VTM; Hanks Balanced Salt Solution, 2% fetal bovine serum, 100 μ g/mL Gentamycin, and 0.5 μ g/mL Amphotericin B). The nasopharyngeal sample in VTM was subjected to two cycles of centrifugation at 500 \times g for 5 min to pellet cells. Cell-associated RNA was extracted using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen) as per manufacturer's instructions. The levels of each subgenomic target and three genomic targets were measured using RT-ddPCR as described in Section 2.4.2.5.

3. Results

3.1. Validations using plasmid DNA

Plasmid constructs containing genomic regions upstream of one or more SARS-CoV-2 body TRS and body gene ("S2-ORF3a-E", "ORF6-ORF7ab-ORF8-NC-3'UTR", "Nsp16-S1", and "E-M") and subgenomic constructs containing the 5' UTR leader sequence upstream of different body genes (S, ORF3a, E, M, ORF7a, ORF8, and N) were designed in pBluescript KS(+) (Bio Basic Inc., Ontario, Canada) to enable assay validations (Table 2). Plasmid concentrations were quantified by ultraviolet (UV) spectrophotometry (NanoDrop ND-1000 instrument, Thermo Fisher) and the molecular weights of respective plasmids were used to calculate the number of molecules per μ L. Extracted PBMC from a healthy donor (150–200 ng/well) and H₂O were included as negative controls for each primer/probe set.

3.1.1. Comparison of sgRNA primer/probe sets to previously validated primer/probe sets targeting SARS-CoV-2

Initially, to assess the performance of the designed sgRNA assays, we simultaneously measured the copies detected for selected subgenomic primer/probe sets (sgE, sgM, and sgN₂) and previously-validated SARS-CoV-2 primer/probe sets that target coding regions of the E, M, or N genes [24] using the subgenomic plasmids with the leader sequence linked to the respective coding region (Table 2, Fig. 2). Each of the previously-validated primer/probe sets target coding regions downstream of the body TRS, so they detect both genomic and subgenomic RNAs containing that gene ("total" SARS-Cov-2 RNA) [24].

Table 2

Plasmid constructs for SARS-CoV-2 sgRNA and gRNA assay validations.

Plasmid Name	Coding Regions ^a	Features	Contains regions targeted by
Subgenomic S	Spike subunit 1	TRS at 5' end	sgS
Subgenomic ORF3a	ORF3a	TRS at 5' end	sgORF3a
Subgenomic E	Envelope	TRS at 5' end	sgE, E_Sarbeco
Subgenomic M	Membrane	TRS at 5' end	sgM, M-ORF5
Subgenomic ORF7a	ORF7a	TRS at 5' end	sgORF7a
Subgenomic ORF8	ORF8	TRS at 5' end	sgORF8
Subgenomic N	Nucleocapsid	TRS at 5' end	sgN, N-ORF9
Nsp16-SpikeS1	Nsp16, Spike subunit 1	genomic sequence upstream of S	gS
S2-ORF3a-E	Spike subunit 2, ORF3a, and envelope	genomic sequence upstream of ORF3a and E	gORF3a, gE
E-M	Envelope, membrane	genomic sequence upstream of M	gM
ORF6-ORF7ab-ORF8-NC-3UTR	ORF6, ORF7a, ORF7b, ORF8, nucleocapsid, and 3'UTR	genomic sequence upstream of ORF7a, ORF8, and N	gORF7a, gORF8, gN

^a All plasmids feature a backbone of pBluescript KS(+).

Two plasmid inputs were used to compare each assay for subgenomic or total RNA and the efficiencies at each input were determined using the ratio of measured copies to expected copies (as calculated from plasmid concentration and molecular weights) to ascertain whether there were marked differences between primer/probes sets using the same plasmid input. The average efficiency was calculated by the slope (S) of the measured copies vs. expected copies for both inputs (Fig. 2). The average efficiency of each subgenomic primer/probe set was similar to the assay for total SARS-CoV-2 RNA (Slope for E: subgenomic E [sgE] = 1.20, total E ["E_Sarbeco"] = 1.16; Slope for M: sgM = 0.66, total M ["M-ORF5"] = 0.71; and slope for N: sgN₂ = 0.92, total N ["N-ORF9"] = 0.95; Fig. 2). The sgE and sgN assays performed with efficiency close to 100%, whereas sgM was lower (at 66%) but similar to the assay for total M (ORF5; efficiency = 71%) in this study and previous work [24]. For these initial tests, as we measured only two plasmid inputs, the efficiency estimation served only as a preliminary guide prior to commencing subsequent validation experiments.

3.1.2. Efficiency and linearity of SARS-CoV-2 panel of ddPCR assays determined using plasmid DNA

Next, we determined the efficiency, linearity, and sensitivity for all of the subgenomic primer/probe sets using terminal dilutions of the respective subgenomic plasmids (Fig. 3). Briefly, plasmid DNA (Table 2) was added to ddPCR wells at expected inputs of 1–10,000 copies/well in duplicate (10,000, 1000 and 100 copies) or quadruplicate (10 and 1 copy).

Assay efficiencies (slope of measured vs. expected copies) were $>$ 63% for all assays (Range: 63%–138%; Median: 81%; Fig. 3). For subgenomic targets for which we had designed alternative primer/probe sets (sgORF3a, sgORF7a, and sgORF8), the better performing assay (Fig. 3, colored symbols) was chosen based on efficiency, sensitivity, and separation of the positive and negative droplets (Figs. S1–S2). All assays showed linear quantification ($R^2 \geq 0.999$) with a dynamic range of ≥ 4 log₁₀ (Fig. 3). Five of the seven assays were able to detect as few as one copy per well (two independent experiments, each with quadruplicate wells for each primer/probe set). Negative controls of human donor PBMC and H₂O included in every experiment were consistently negative for all selected sgRNA primer/probe sets listed in Table 1 (Figs. S1–S2). Alternate primer/probe sets for which false positives were detected,

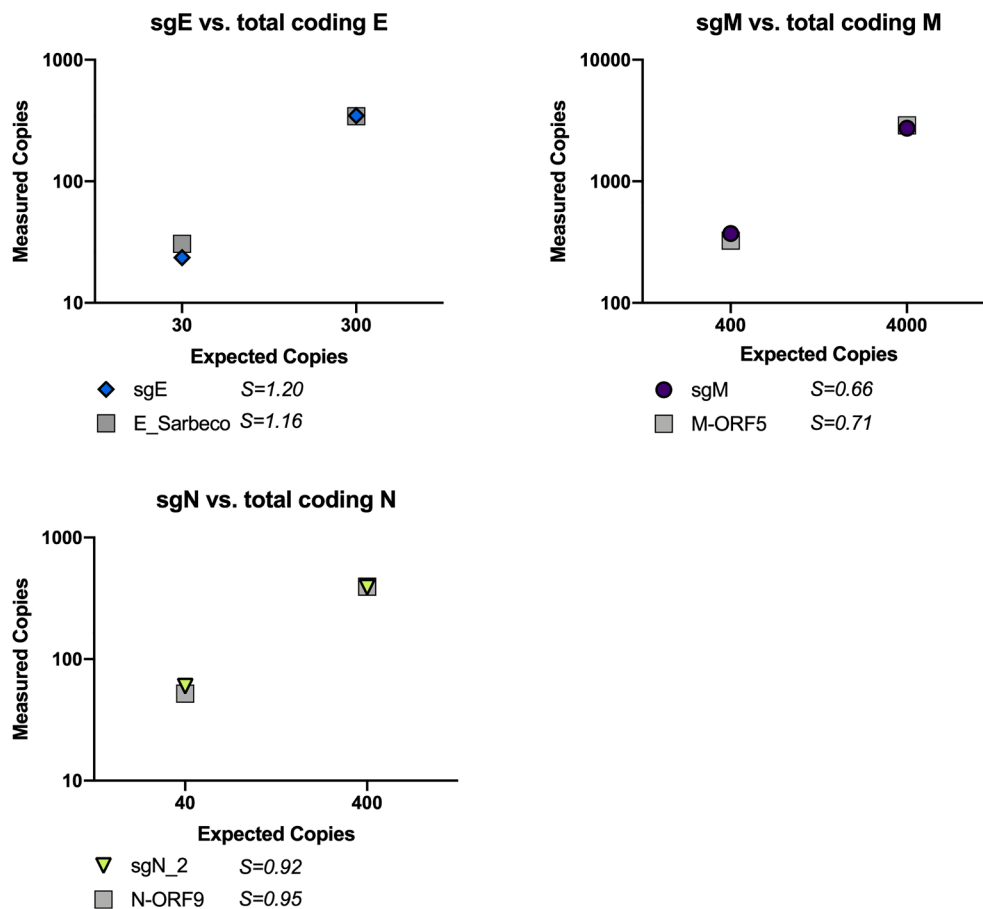


Fig. 2. Comparing efficiency of subgenomic assays and total SARS-CoV-2 assays. The efficiency of selected subgenomic assays (sgE, sgM, and sgN₂) relative to previously validated SARS-CoV-2 assays (targeting total E, M or N coding region; grey symbols) were measured using the same plasmids. *S* indicates slope (efficiency).

such as for ‘sgORF8_2’ and ‘sg_N1’, were not selected for the final panel.

3.1.3. Specificity of subgenomic primer/probe sets

To ascertain the specificity of the subgenomic RNA primer/probe sets, we tested each subgenomic primer/probe set using plasmids designed for their respective genomic sequence (Table 2). Each genomic plasmid was added to ddPCR reactions at an input of at least 10 000 copies along with either the subgenomic or respective genomic sequence-targeting primer/probe set. Five of the seven subgenomic sequence targeting primer/probe sets (sgORF3a_1, sgE, sgM, sgORF8_1, sgN₂) did not detect any copies of plasmid containing the respective genomic sequence even at inputs above 17 000 copies [near saturation] (Fig. S3). At inputs >17 000 copies of genomic plasmid, only one droplet was detected across 4 replicate wells for sgS, while for sgORF7a_2, one droplet was detected in each well of two replicates. In contrast, the matched genomic sequence-targeting primer/probe sets detected copies in the expected input range, close to saturation. Taken together, these data demonstrate that the detection of subgenomic sequences by all seven designed primer/probe sets is highly specific.

3.2. The expression of sgRNA vs. “genomic” RNA in supernatant from SARS-CoV-2 *in vitro* infections

Although supernatant from *in vitro* infections with SARS-CoV-2 might be expected to contain mostly gRNA in virions, we previously reported that supernatant RNA levels of the total M, N, and 3’ UTR regions (found in both gRNA and sgRNA) were higher than that of RNA coding for 5’ targets found only in gRNA (main protease, RDRP) [24], suggesting the supernatant contains sgRNA. In order to test this

hypothesis and determine if we could quantify different sgRNAs, all of the assays for subgenomic and “genomic” ORF2-9 transcripts were applied to RNA extracted from the supernatant. Dilutions of the supernatant RNA were added to RT reactions to achieve expected inputs (based on testing the supernatant with the Abbott RealTime SARS-CoV-2 clinical assay) of 1 to 10,000 copies per 5 μ L RT (the input into each ddPCR well).

All assays for “genomic” ORF 2–9 transcripts showed relatively similar efficiencies, ranging from 118 to 169% (Fig. 4). In these experiments, for 6 out of 7 regions, at least one “genomic” RNA primer/probe set was able to detect as few as 1 copy/well. Negative controls, including ‘no template’ and ‘no RT’ controls, were routinely negative (Fig. S4). Contrastingly, each sgRNA assay detected a smaller fraction of the total, ranging from 1 to 14% of expected total copies, based on average efficiency inferred from the slope [*S*] (Fig. 4). Since each subgenomic assay should only detect one particular form of sgRNA arising from a template switch at a given body TRS (Fig. 1) and is highly specific for sgRNA and not genomic RNA (Fig. S3), we reasoned that the sum of all of these sgRNA species should approximate the overall levels of subgenomic RNA present in the supernatant. The cumulative sum of all measured sgRNAs was approximately 42% of the total expected SARS-CoV-2 RNA copies in the supernatant, which accords with the previously measured excess of 3’ over 5’ SARS-CoV-2 RNA levels [24].

Next, we calculated the ratio of copies of each sgRNA relative to its cognate “genomic” RNA, which represents the ratio of discontinuous transcription to continuous transcription at a given body TRS. This comparison should minimize differences due to assay efficiency because the corresponding assays share the same probe and reverse primer, while the forward primers are located similar distances from the probe.

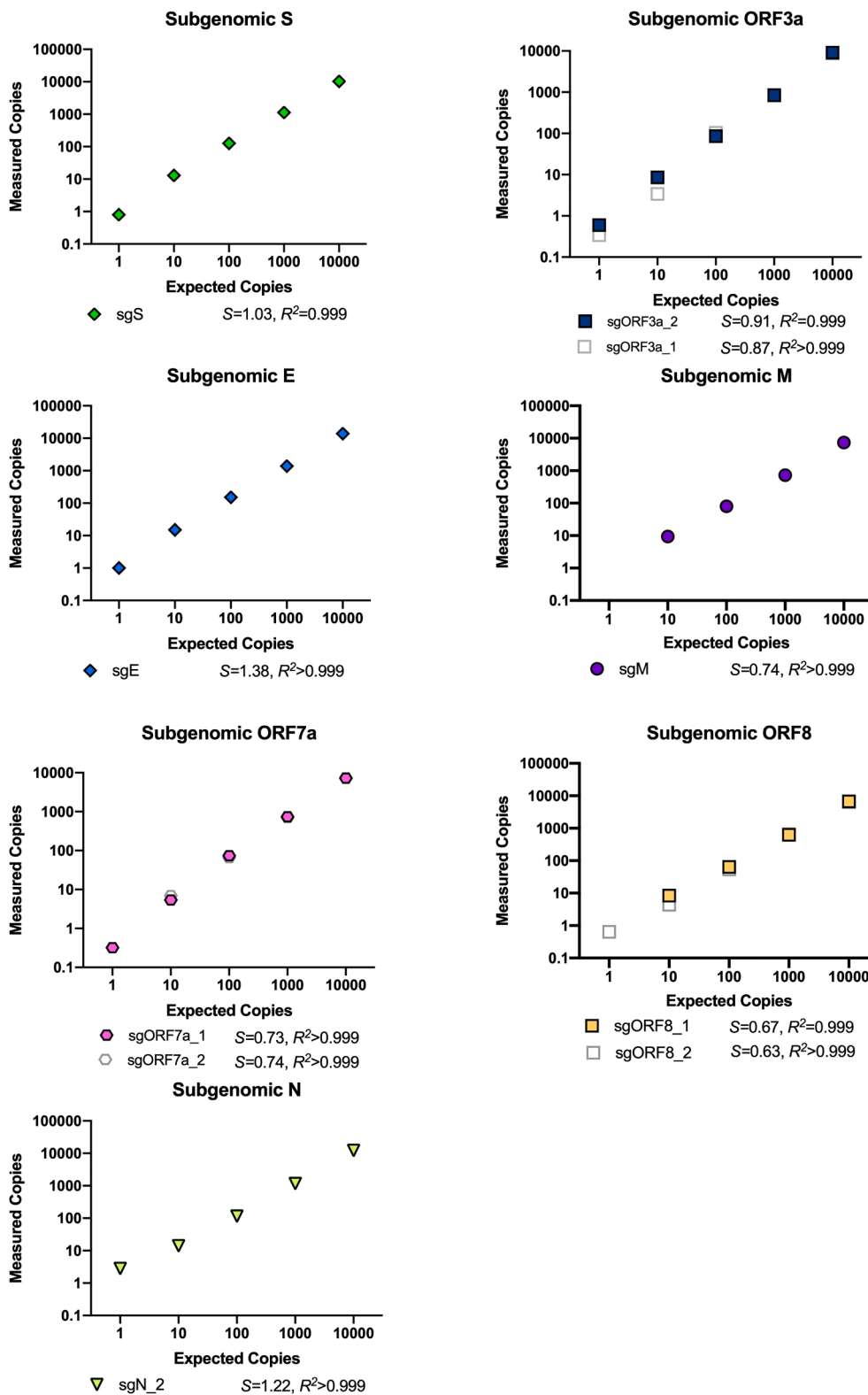


Fig. 3. Efficiency and linearity of ddPCR assays for SARS-CoV-2 subgenomic transcripts determined using plasmid DNA. Specially designed subgenomic plasmids containing the 5' UTR sequence upstream of the coding regions of individual body genes were quantified by UV spectroscopy and diluted (expected copies) to test the absolute number of copies detected by each primer/probe set using duplicate ddPCR reactions (measured copies). Coloured symbols denote the best-performing of alternate primer/probe sets for a given sgRNA.

The ratios of subgenomic to cognate “genomic” RNA varied between 1.5–5.2%. The sum of these ratios, which provides a minimum estimate of all subgenomic RNA as a fraction of genomic RNA, was 27.9%. However, as mentioned previously, some of our “genomic” assays may also detect subgenomic transcripts that result from upstream “splicing.” Therefore, we also calculated the ratio of copies of each sgRNA relative to a single true genomic RNA, as measured by the copies

from the genomic S assay. The ratios of subgenomic RNA to genomic S ranged from 1.2 to 5.2%. The sum of these ratios, which provides another minimum estimate of the total ratio of subgenomic to genomic RNA, was 24.9% (Fig. 5C). These two analyses suggest a subgenomic to genomic ratio of at least 25–28%, indicating that subgenomic transcripts may account for at least 20–22% of all viral RNA. Taken together, these results suggest that our novel panel of RT-ddPCR assays can be used to

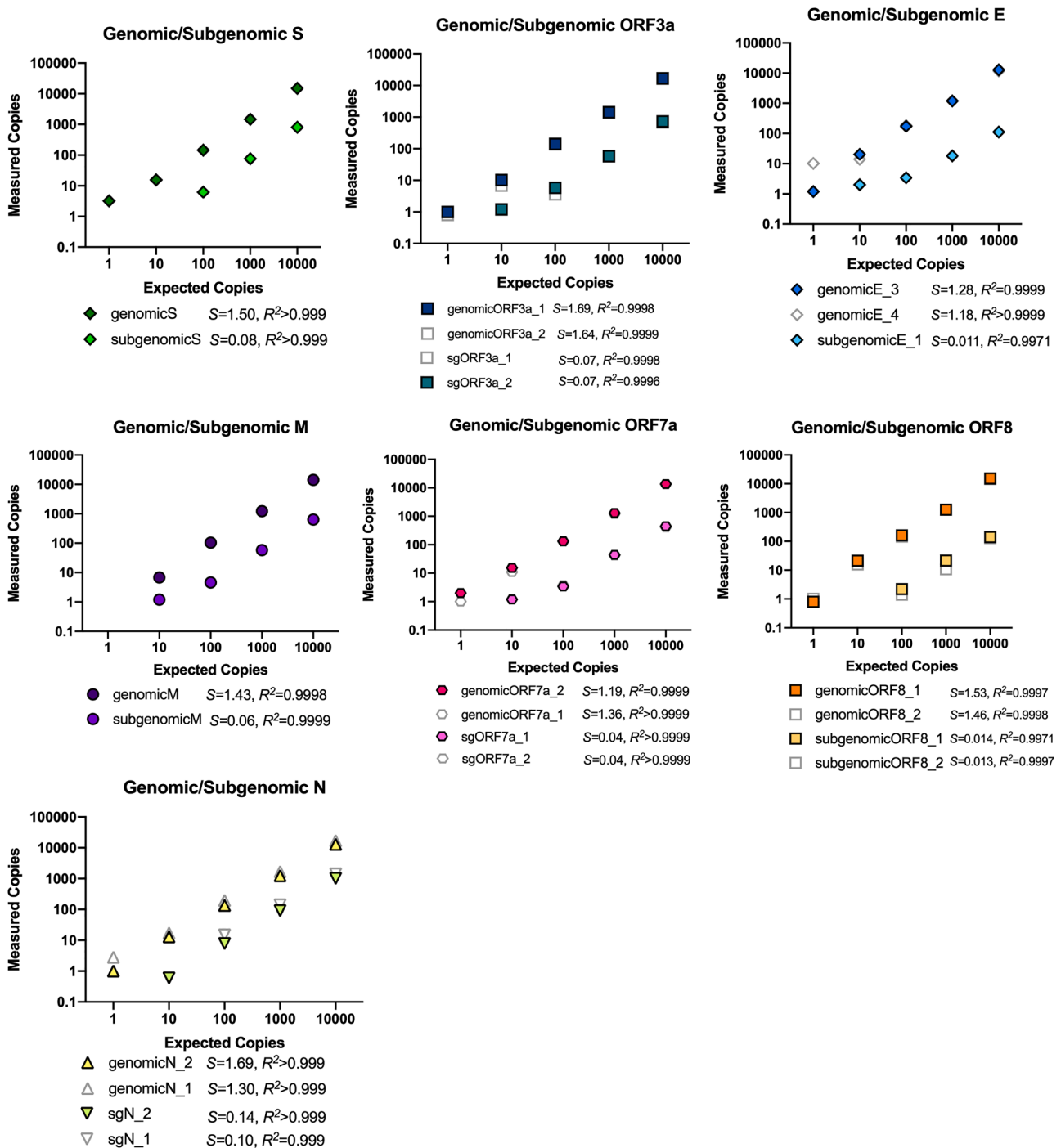


Fig. 4. Efficiency and linearity of ddPCR assays for SARS-CoV-2 “genomic” and subgenomic transcripts determined using supernatant from *in vitro* infection. A SARS-CoV-2 ‘supernatant’ standard was prepared by extracting the RNA from the supernatant of an *in vitro* infection and quantified using the Abbott Real Time SARS-CoV-2 assay. Various inputs of the supernatant standard (which were used to calculate ‘Expected Copies’ per ddPCR well) were applied to a common reverse transcription reaction, from which aliquots of cDNA were used to measure the absolute number of copies detected by each ddPCR assay (measured copies) for ‘genomic’ or subgenomic transcripts. Each assay was tested with expected inputs of 10–10,000 copies/ddPCR well in duplicate. *S* (slope) and R^2 are indicated for each assay. Coloured symbols denote the best-performing of alternate primer/probe sets for a given gRNA or sgRNA.

measure the absolute levels of 7 different SARS-CoV-2 sgRNAs and provide a minimum estimate of all subgenomic transcripts as a fraction of genomic or total viral RNA.

3.3. Subgenomic RNA expression levels in pharynx from one acutely infected individual

Next, we applied our newly-validated assays to a nasopharyngeal

sample from a laboratory-confirmed SARS-CoV-2 acutely-infected individual (5 days post-symptom onset and 4 days after positive clinical PCR test). Levels of each distinct sgRNA (sgS, sgORF3a, sgE, sgM, sgORF7a, sgORF8, and sgN) and three targets found only in genomic RNA (gS, RDRP and Main Proteinase) were measured in cells from the nasopharynx (Fig. 6). All 7 subgenomic RNAs were detected in the nasopharyngeal cells. With the exception of subgenomic S, the hierarchy of subgenomic transcript levels ($N > 3a, M, 7a > S > 8, E$) was relatively

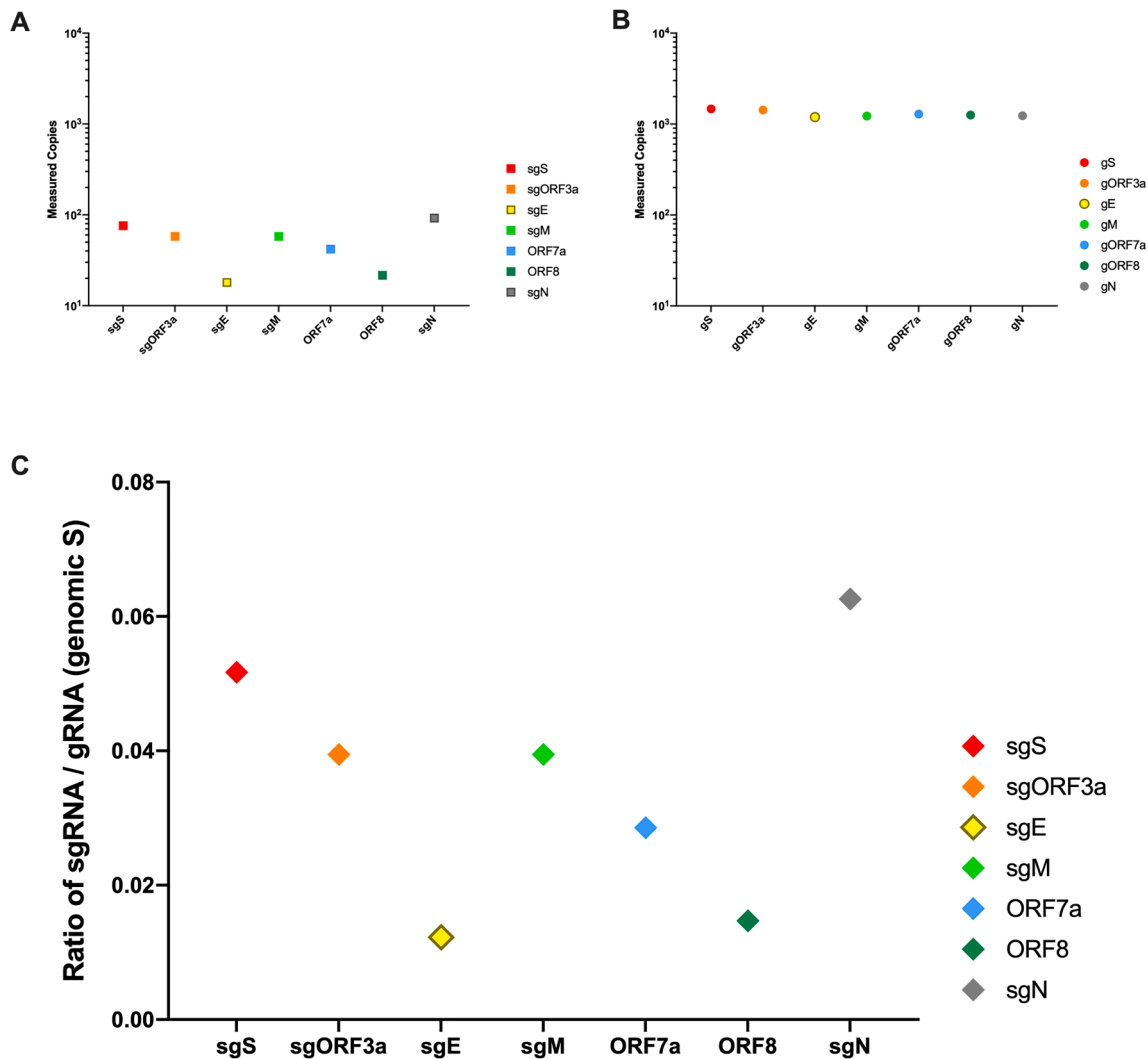


Fig. 5. Expression of sgRNA and gRNA in supernatant from a SARS-CoV-2 *in vitro* infection. Aliquots of cDNA from a common RT reaction containing SARS-CoV-2 RNA were added to ddPCR reactions (predicted to contain 1000 copies of SARS-CoV-2/well) and the levels of each (A) sgRNA and (B) gRNA target were measured. (C) The copies of each sgRNA were divided by the copies of genomic S RNA to express the ratio of each subgenomic RNA to genomic RNA. Each target region is depicted by a different colour (red: S; orange: ORF3a; yellow: E; green: M; blue: ORF7a; teal: ORF8; and grey: N). Symbols represent the following: squares: subgenomic RNA; circles: genomic RNA; and diamonds: ratio of sgRNA/gRNA. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

similar to that seen in the supernatant from SARS-CoV-2 *in vitro* infection ($N > S > 3a, M, 7a > 8, E$). Using the sum of the ratios of each sgRNA to genomic S RNA, we determined that the total ratio of subgenomic to genomic RNA was 55.4%, indicating that subgenomic transcripts may account for at least 35.6% of the total viral RNA present in nasopharyngeal cells from this acutely infected individual.

4. Discussion

The role of sgRNAs in infection with SARS-CoV-2 and other coronaviruses remains unclear. While important in the generation of viral proteins that are encoded in the 3' region of the viral genome, their role in the life cycle and pathogenicity of these viruses still remains to be studied in greater detail. It is important to determine if sgRNAs are a suitable marker for active replication, as suggested by some studies [12,20,27,28], or if their enclosure in double membrane vesicles [6,29] and/or extracellular vesicles allows persistence after replication has ceased. Furthermore, the discovery of several non-canonical sgRNAs [2,30] suggests that more diverse sgRNA species may play a role in SARS-CoV-2 life cycle. However, the unique recombination sequences

seen in these non-canonical sgRNAs may indicate that they function as defective interfering RNAs, as has been previously suggested [31]. Since sgRNAs in other viruses play roles in viral replication and recombination, these and other roles of sgRNA in SARS-CoV2 merit further investigation. Given these gaps in knowledge, it is critical to continue to study sgRNAs to better understand coronaviruses, their role in pathogenicity, and potential targets of future therapies.

In this study, we developed and validated a novel panel of RT-ddPCR-based SARS-CoV-2 assays that target 7 distinct canonical sgRNAs. All assays could detect as few as 1–10 copies using plasmid DNA carrying the cognate sequence and demonstrated linearity over 4–5 orders of magnitude ($R^2 > 0.999$ for all; Fig. 3). Given the relatively similar PCR efficiencies of the subgenomic assays on plasmid DNA (Fig. 3), it is unlikely that variable assay efficiencies alone account for the differing abundances of the sgRNA species observed in the supernatant from the *in vitro* infection (Fig. 4). Though we did not formally measure the limit of detection (LOD) of each primer probe set, the lowest inputs of 1 or 10 copies were typically measured in quadruplicate wells for each independent validation experiment using either plasmid DNA or viral RNA. We found the sensitivity of each primer/probe set to be

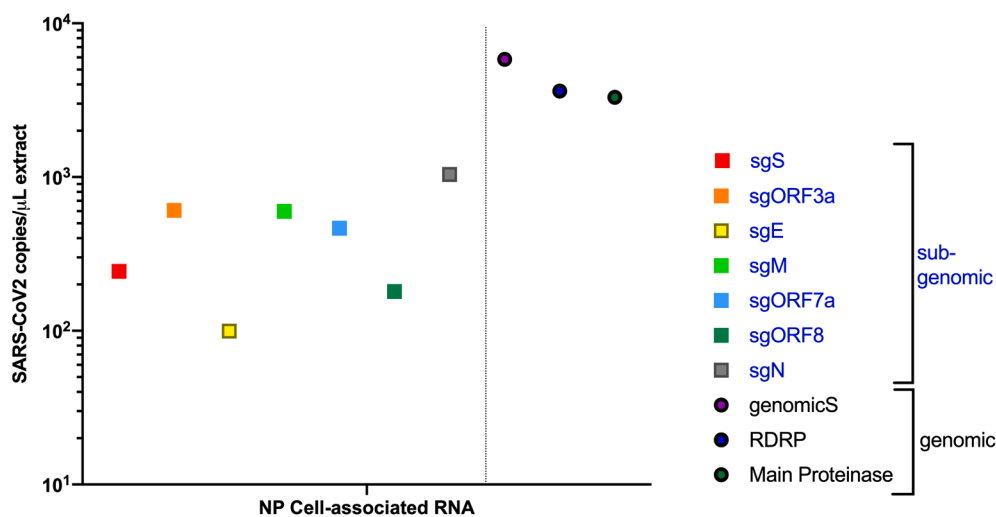


Fig. 6. Expression of sgRNAs in pharynx of one acutely SARS-CoV-2- infected individual. Total cell-associated RNA was isolated from a nasopharyngeal swab. A common RT reaction was divided across ddPCR reactions to measure seven sgRNA targets and one genomic target (S). Independent measurements of two other genomic targets, RDRP and Main Proteinase, from the same sample are included as a reference. Levels of each target are expressed as copies per μL extract.

reproducible in these experiments.

It is worth noting that the expected copies of SARS-CoV-2 RNA in the supernatant were based on the viral load as determined by the Abbot RealTime SARS-CoV-2 assay, which measures the combined fluorescence from a reaction containing two qPCR assays (targeting RDRP and N) employing probes with the same fluorophore. The Abbott RealTime SARS-CoV-2 assay does not exclusively detect genomic RNA, since the N primer/probe set will detect both genomic and subgenomic SARS-CoV-2. Moreover, the Abbott provides relative quantification by extrapolation from an external standard. The observation that our “genomic” RNA assays consistently detected more than the expected copies from the Abbott (Fig. 4) could reflect the possibility that the relative quantification from the Abbott underestimated the true number of copies, that our assays were more efficient, or that there were other sources of error (dilution, pipetting, assay performance, etc.). Nonetheless, the fraction of sgRNA detected for each target is indicative of the level of sgRNA produced during *in vitro* infection of Vero-CCL81 cells with SARS-CoV-2. Our data suggest that the ratio of subgenomic to genomic RNA is at least 25–28% in the supernatant from this *in vitro* infection.

Like the approach used by Kim et al., 2020 [11], we employed Vero cells infected with virus isolated from a patient [24]. However, our approach differed from other reported methods [9,11,12] to quantify sgRNAs. We employed an RT strategy that seeks to minimize bias towards any one region or either end of the RNA by using both random hexamers and poly-dT. Furthermore, our use of a common RT reaction for multiple ddPCR reactions should reduce differences due to reverse transcription and facilitate more accurate comparison of different RNA target levels (for instance, matched sgRNA and gRNA targets for multiple regions). Our approach is also faster than deep and direct RNA sequencing methods [11], although it does not allow for the examination of any RNA modifications, as does direct RNA sequencing. While a prior publication described a PCR-based assay for one subgenomic RNA (sgE) [12], our approach is more comprehensive in that it allows us to quantify seven canonical sgRNAs, measure the rates of discontinuous transcription at each region, and estimate the total abundance of sgRNA species.

Recent studies estimate the abundance of sgRNAs in clinical samples or *in vitro* infections to be 0.4% [12] – 5.6% [11]. Our data suggest that this may be an underestimate of total sgRNA species, which may be present at levels exceeding 25–55% of the genomic SARS-CoV-2 RNA copies or 20–36% of the total viral RNA. Further studies with larger numbers of clinical samples are required to confirm this finding. Taken together, these data suggest that sgRNA species are likely more

abundant than previously thought and highlight the difficulty in accurately quantifying one or all sgRNAs using other published approaches.

4.1. Limitations

Given the difficulty in preparing and independently quantifying full-length RNA standards for genomic and seven different subgenomic SARS-CoV-2 transcripts, we were unable to determine whether the efficiency of reverse transcription may differ between the various assays for sgRNA or “genomic” RNA. Consequently, we cannot exclude the possibility that differences in the efficiencies of reverse transcription could contribute to differences between the measured levels of various sgRNAs or the ratios of each sgRNA to its cognate “genomic” RNA. However, the use of a “common” RT reaction (from which aliquots are taken for all assays) and the reverse transcription with random hexamers and poly-dT should minimize differences at the reverse transcription step. Moreover, use of a common probe and reverse primer should minimize differences between the sgRNA and cognate “genomic” RNA assays.

We also considered testing one step, dd-RT-PCR assays where reverse transcription would take place in each droplet using specific reverse primers. However, prior experiments using dd-RT-PCR suggested that RT efficiency did not contribute significantly to differences between supernatant levels of different SARS-CoV-2 RNA regions as measured by our previously validated panel of SARS-CoV-2 RNA assays [24]. Moreover, we found that the one step dd-RT-PCR approach resulted in a poorer signal to noise ratio and more false positive droplets using the same primers/probes (all TaqMan dual-labelled FAM-MGB probes from Applied Biosystems) used in parallel with the 2-step approach.

While genetic diversification of coronaviruses occurs at lower rates than that of influenza virus, viral variants have been described that, owing to enhanced transmissibility, are predicted to replace original founder strains and become predominant in the near future. These mutations could affect performance of PCR-based assays if they occur in regions targeted by the PCR primers or probes. Table S2 shows the mutations described in the recent COVID-19 Genomics UK Consortium report [32]. Based on this report, one mutation falls in the reverse primer for the subgenomic and genomic ORF 8 assay (ORF8_R1), while another mutation corresponds to the very 5' end of the forward primer for “genomic” N (gN_F1; Table S1). However, the remaining assays (including those for the spike gene) should not be affected. Moreover, it should be noted that not all mismatches in primer/probe sequences will have an appreciable impact on amplification efficiency, and ddPCR is

reported to be less susceptible to the inhibitory effects of sequence mismatches. When mismatches do cause significant effects on amplification efficiency, these effects are often visible in the raw ddPCR plots, and primer/probe sequences can be corrected to account for known mutations.

4.2. Potential applications

The sensitive, quantitative performance of our RT-ddPCR assays and the breadth of detection for different sgRNA species enables fine discrimination of differences between the rates of discontinuous transcription at various loci and between the abundance of different sgRNAs within the same sample. As a result, this approach can be utilized to enhance our understanding of the mechanisms that regulate coronavirus transcription. Our panel of sgRNAs and corresponding “genomic” RNAs (or previously developed RT-ddPCR assays for “genomic only” targets such as RdRP and main protease [24]) should be applied to other samples from *in vitro* infections and SARS-CoV-2 infected individuals. These assays can be used to study the kinetics of SARS-CoV-2 genomic and subgenomic transcription, the function of sgRNAs, and the contribution of sgRNA to viral pathogenesis. It remains to be determined whether sgRNA levels correlate with infectivity/transmission, disease severity, or sequelae of COVID-19, but our assays provide powerful tools to address these questions.

5. Conclusions

Here, we describe the design and validation of seven novel ddPCR assays specific for the most abundantly reported sgRNAs of SARS-CoV-2, as well as matched primer/probe sets that target “genomic” RNA. Our approach, which may be faster, cheaper, and more quantitative than sequencing, could be utilized to study less abundant sgRNAs of SARS-CoV-2 or adapted to measure sgRNAs from other viruses. In this way, the methods described here can help researchers understand the role of subgenomic transcription in SARS-CoV-2 and other viruses.

CRediT authorship contribution statement

Sushama Telwatte: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Validation, Visualization, Writing - original draft. **Holly Anne Martin:** Investigation, Formal analysis, Writing - original draft. **Ryan Marczak:** Investigation, Writing - review & editing. **Parinaz Fozouni:** Resources, Investigation, Writing - review & editing. **Albert Vallejo-Gracia:** Resources, Investigation, Writing - review & editing. **G. Renuka Kumar:** Resources, Investigation, Writing - review & editing. **Victoria Murray:** Resources, Writing - review & editing. **Sulggi Lee:** Resources, Supervision, Writing - review & editing. **Melanie Ott:** Resources, Supervision, Writing - review & editing. **Joseph K. Wong:** Resources, Supervision, Writing - review & editing. **Steven A. Yukt:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Visualization, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The revised sentence would read: "This research was supported by funds from the Emergency COVID-19 Research Seed Funding of the University of California (Grant Number R00RG3113 [ST]), the U.S. Department of Veterans Affairs (COVID19-8100-05 [JW]), and a COVID-19 Fast Grant Award (Emergent Ventures [SAL]). The investigators

received salary support from the U.S. Department of Veterans Affairs (SAY and JKW), the National Institute of Diabetes and Digestive and Kidney Diseases at the NIH (R01DK108349 [SAY, JKW], R01DK120387 [SAY]), the National Institute of Allergy and Infectious Diseases at the NIH (R01AI132128 [SAY, JKW], 1R61AI140465-01 [MO]), the UCSF/GIVI Center for AIDS Research (Mentored Scientist in HIV Award, CFAR; Grant# P30 AI027763 [ST]; award #A120163 [PI: Paul Volberding]), and the California HIV/AIDS Research Program (Grant number BB19-SF-009 [ST]). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ymeth.2021.04.011>.

References

- [1] N.a. Zhu, D. Zhang, W. Wang, X. Li, B.o. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* 382 (8) (2020) 727–733.
- [2] J.-M. Kim, Y.-S. Chung, H.J. Jo, N.-J. Lee, M.S. Kim, S.H. Woo, S. Park, J.W. Kim, H.M. Kim, M.-G. Han, Identification of coronavirus isolated from a patient in Korea with COVID-19, *Osong. Public Health Res. Perspect.* 11 (1) (2020) 3–7.
- [3] S.G. Sawicki, D.L. Sawicki, Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands, *Adv Exp Med Biol* 380 (1995) 499–506.
- [4] M.M. Lai, S.A. Stohlman, Comparative analysis of RNA genomes of mouse hepatitis viruses, *J. Virol.* 38 (1981) 661–670.
- [5] I. Sola, F. Almazán, S. Zúñiga, L. Enjuanes, Continuous and discontinuous RNA synthesis in coronaviruses, *Annu. Rev. Virol.* 2 (1) (2015) 265–288.
- [6] M.J. van Hemert, S.H.E. van den Worm, K. Knoops, A.M. Mommaas, A. E. Gorbalenya, E.J. Snijder, R.S. Baric, SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity *in vitro*, *PLoS Pathog.* 4 (5) (2008) e1000054, <https://doi.org/10.1371/journal.ppat.1000054>.
- [7] D. Escors, A. Izeta, C. Capiscol, L. Enjuanes, Transmissible gastroenteritis coronavirus packaging signal is located at the 5' end of the virus genome, *J. Virol.* 77 (14) (2003) 7890–7902.
- [8] J.J.A. van Kampen, D.A.M.C. van de Vijver, P.L.A. Fraaij, et al., Shedding of infectious virus in hospitalized patients with coronavirus disease-2019 (COVID-19): duration and key determinants, *medRxiv* 2020 (06) (2020).
- [9] S. Alexandersen, A. Chamings, T.R. Bhatta, SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication, *Nature Commun* 11 (2020) 6059.
- [10] S.C. Taylor, G. Laperriere, H. Germain, Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data, *Sci. Rep.* 7 (2017) 2409.
- [11] D. Kim, J.-Y. Lee, J.-S. Yang, J.W. Kim, V.N. Kim, H. Chang, The Architecture of SARS-CoV-2 Transcriptome, *Cell* 181 (4) (2020) 914–921.e10.
- [12] R. Wölfel, V.M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M.A. Müller, D. Niemeyer, T.C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R. Ehmam, K. Zwirgmaier, C. Drosten, C. Wendtner, Virological assessment of hospitalized patients with COVID-2019, *Nature* 581 (7809) (2020) 465–469.
- [13] J. Ma, N. Li, M. Guarnera, F. Jiang, Quantification of plasma miRNAs by digital PCR for cancer diagnosis, *Biomarker Insights* 8 (2013) 127–136.
- [14] T.C. Dingle R.H. Sedlak L. Cook K.R. Jerome Tolerance of droplet-digital PCR vs real-time quantitative PCR to inhibitory substances 59 11 2013 2013 1670 1672.
- [15] R.H. Sedlak, J. Kuypers, K.R. Jerome, A multiplexed droplet digital PCR assay performs better than qPCR on inhibition prone samples, *Diagn. Microbiol. Infect. Dis.* 80 (4) (2014) 285–286.
- [16] M.C. Strain, S.M. Lada, T. Luong, S.E. Rought, S. Gianella, V.H. Terry, C.A. Spina, C.H. Woelk, D.D. Richman, Highly precise measurement of HIV DNA by droplet digital PCR, *PLoS One* 8 (4) (2013), e55943.
- [17] C.M. Hindson, J.R. Chevillet, H.A. Briggs, E.N. Gallichotte, L.K. Ruf, B.J. Hindson, R.L. Vessella, M. Tewari, Absolute quantification by droplet digital PCR versus analog real-time PCR, *Nat. Methods* 10 (10) (2013) 1003–1005.
- [18] Y. Zhao Q. Xia Y. Yin Z. Wang Y.-W. He Comparison of droplet digital PCR and quantitative PCR assays for quantitative detection of *Xanthomonas citri* Subsp. citri *PLoS One* 11 7 2016;11:e0159004. e0159004 10.1371/journal.pone.0159004.
- [19] A. Bharuthram, M. Paximadis, A.C.P. Picton, C.T. Tiemessen, Comparison of a quantitative Real-Time PCR assay and droplet digital PCR for copy number analysis of the CCL4L genes, *Infect. Genet. Evol.* 25 (2014) 28–35.
- [20] R.A.P.M. Perera, E. Tso, O.T.Y. Tsang, D.N.C. Tsang, K. Fung, Y.W.Y. Leung, A.W. H. Chin, D.K.W. Chu, S.M.S. Cheng, L.L.M. Poon, V.W.M. Chuang, M. Peiris, SARS-CoV-2 virus culture and Subgenomic RNA for respiratory specimens from patients with mild coronavirus disease, *Emerg. Infect. Dis.* 26 (11) (2020) 2701–2704.
- [21] F. Alessandrini, S. Caucci, V. Onofri, F. Melchionda, A. Tagliabracchi, P. Bagnarelli, L. Di Sante, C. Turchi, S. Menzo, Evaluation of the ion AmpliSeq SARS-CoV-2

- research panel by massive parallel sequencing, *Genes (Basel)* 11 (8) (2020) 929, <https://doi.org/10.3390/genes11080929>.
- [22] T.R. Bhatta, A. Chamings, K.-C. Liew, et al., Sequence analysis of travel-related SARS-CoV-2 cases in the Greater Geelong region, Australia. *medRxiv* 2020 (06) (2020).
- [23] F. Madeira Y.m. Park J. Lee N. Buso T. Gur N. Madhusoodanan P. Basutkar A.R.N. Tivey S.C. Potter R.D. Finn R. Lopez The EMBL-EBI search and sequence analysis tools APIs in 2019 47 W1 2019 2019 W636 W641.
- [24] S. Telwatte, N. Kumar, A. Vallejo-Gracia, G.R. Kumar, C.M. Lu, M. Ott, J.K. Wong, S.A. Yuki, Novel RT-ddPCR assays for simultaneous quantification of multiple noncoding and coding regions of SARS-CoV-2 RNA, *J. Virol. Methods* 292 (2021) 114115, <https://doi.org/10.1016/j.jviromet.2021.114115>.
- [25] S.F. Altschul, T.L. Madden, A.A. Schäffer, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (1997) 3389–3402.
- [26] S. Telwatte P. Kim T.-H. Chen J.M. Milush M.a. Somsouk S.G. Deeks P.W. Hunt J.K. Wong S.A. Yuki Mechanistic differences underlying HIV latency in the gut and blood contribute to differential responses to latency-reversing agents 34 14 2020 2013 2024.
- [27] K.S. Corbett, B. Flynn, K.E. Foulds, J.R. Francica, S. Boyoglu-Barnum, A.P. Werner, B. Flach, S. O'Connell, K.W. Bock, M. Minai, B.M. Nagata, H. Andersen, D. R. Martinez, A.T. Noe, N. Douek, M.M. Donaldson, N.N. Nji, G.S. Alvarado, D. K. Edwards, D.R. Flebbe, E. Lamb, N.A. Doria-Rose, B.C. Lin, M.K. Louder, S. O'Dell, S.D. Schmidt, E. Phung, L.A. Chang, C. Yap, J.-P. Todd, L. Pessaint, A. Van Ry, S. Browne, J. Greenhouse, T. Putman-Taylor, A. Strasbaugh, T.-A. Campbell, A. Cook, A. Dodson, K. Steingrebe, W. Shi, Y.i. Zhang, O.M. Abiona, L. Wang, A. Pegu, E.S. Yang, K. Leung, T. Zhou, I.-T. Teng, A. Widge, I. Gordon, L. Novik, R.A. Gillespie, R.J. Loomis, J.I. Moliva, G. Stewart-Jones, S. Himansu, W.-P. Kong, M.C. Nason, K.M. Morabito, T.J. Ruckwardt, J.E. Ledgerwood, M. R. Gaudinski, P.D. Kwong, J.R. Mascola, A. Carfi, M.G. Lewis, R.S. Baric, A. McDermott, I.N. Moore, N.J. Sullivan, M. Roederer, R.A. Seder, B.S. Graham, Evaluation of the mRNA-1273 vaccine against SARS-CoV-2 in nonhuman primates, *N. Engl. J. Med.* 383 (16) (2020) 1544–1555.
- [28] N. van Doremalen, T. Lambe, A. Spencer, S. Belij-Rammerstorfer, J. N. Purushotham, J.R. Port, V.A. Avanzato, T. Bushmaker, A. Flaxman, M. Ulaszewska, F. Feldmann, E.R. Allen, H. Sharpe, J. Schulz, M. Holbrook, A. Okumura, K. Meade-White, L. Pérez-Pérez, N.J. Edwards, D. Wright, C. Bissett, C. Gilbride, B.N. Williamson, R. Rosenke, D. Long, A. Ishwarbhai, R. Kailath, L. Rose, S. Morris, C. Powers, J. Lovaglio, P.W. Hanley, D. Scott, G. Saturday, E. de Wit, S.C. Gilbert, V.J. Munster, ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques, *Nature* 586 (7830) (2020) 578–582.
- [29] X. Zhao, K. Shaw, D. Cavanagh, Presence of subgenomic mRNAs in virions of coronavirus IBV, *Virology* 196 (1) (1993) 172–178.
- [30] A. Viehweger, S. Krautwurst, K. Lamkiewicz, R. Madhugiri, J. Ziebuhr, M. Hölzer, M. Marz, Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis, *Genome Res.* 29 (9) (2019) 1545–1554.
- [31] K. Pathak, P. Nagy, Defective interfering RNAs: foes of viruses and friends of virologists, *Viruses* 1 (3) (2009) 895–919.
- [32] COVID-19 Genomics (COG) UK Consortium. COG-UK update on SARS-CoV-2 Spike mutations of special interest 2020 Dec 20, 2020. Report No.: 15/01/21.