Analytical sensitivity and efficiency comparisons of SARS-COV-2 qRT-PCR assays

Chantal B.F. Vogels[1]*, Anderson F. Brito[1], Anne L. Wyllie[1], Joseph R. Fauver[1], Isabel M. Ott[2], Chaney C. Kalinich[1], Mary E. Petrone[1], Marie L. Landry[4], Ellen F. Foxman[3], Nathan D. Grubaugh[1]*

[1] Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT 06510, USA.
[2] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA
[3] Department of Laboratory Medicine, Yale University School of Medicine, New Haven, CT, 06510, USA.
[4] Departments of Laboratory Medicine and Medicine, Yale University School of Medicine, New Haven, Connecticut
* Correspondence: chantal.vogels@yale.edu (CBFV); nathan.grubaugh@yale.edu (NDG)

Abstract

The recent spread of severe acute respiratory syndrome coronavirus (SARS-CoV-2) exemplifies the critical need for accurate and rapid diagnostic assays to prompt public health actions. Currently, several quantitative reverse-transcription polymerase chain reaction (qRT-PCR) assays are being used by clinical, research, and public health laboratories for rapid detection of the virus. However, it is currently unclear if results from different tests are comparable. Our goal was to evaluate the primer-probe sets used in four common diagnostic assays available on the World Health Organization (WHO) website. To facilitate this effort, we generated RNA transcripts to create standards and distributed them to other laboratories for internal validation. We then used these RNA transcript standards, full-length SARS-CoV-2 RNA, and RNA-spiked mock samples to determine analytical efficiency and sensitivity of nine primer-probe sets. We show that all primer-probe sets can be used to detect SARS-CoV-2, but there are clear differences in the ability to differentiate between true negatives and positives with low amounts of virus. Adding to this, many primer-probe sets, including the “N2” and “N3” sets issued by the US Centers for Disease Control and Prevention, have background amplification with SARS-CoV-2-negative nasopharyngeal swabs, which may lead to inconclusive results. Our findings characterize the limitations of commonly used primer-probe sets and can assist other laboratories in selecting appropriate assays for the detection of SARS-CoV-2.

Introduction

Accurate diagnostic assays and large-scale testing are critical for mitigating outbreaks of infectious diseases. Early detection prompts public health actions to prevent and control the spread of pathogens. This has been exemplified by the novel coronavirus, known as SARS-CoV-2, which was first identified as the cause of an outbreak of pneumonia in Wuhan, China, in December 2019, and rapidly spread around the world[1–3]. The first SARS-CoV-2 genome sequence was critical for the development of diagnostics[2], which led to several molecular assays being developed to detect COVID-19 cases[4–7]. The World Health Organization (WHO) currently lists seven molecular assays (i.e. qRT-PCR) to diagnose COVID-19[8]; however, it is not clear to many laboratories or public health agencies which assay they should adopt.

Our goal was to critically compare the analytical efficiencies and sensitivities of the four most common SARS-CoV-2 qRT-PCR assays developed by the China Center for Disease Control (China CDC)[7], United States CDC (US CDC)[6], Charité (Universitätsmedizin Berlin Institute of Virology, Germany)[5], and Hong Kong University (HKU)[4]. To this end, we first generated RNA transcripts from a SARS-CoV-2 isolate from an early COVID-19 case from the state of Washington (United States)[9]. Using RNA transcripts, isolated virus RNA, and mock clinical samples, our analyses show that all of the primer-probe sets used in the qRT-PCR assays can detect SARS-COV-2, but we find important differences between the analytical sensitivities to detect low amounts of virus and the detection of false positives. Thus, we provide evidence that all of the assays are appropriate for virus detection as long as the limitations of each are recognized.

Results and Discussion

Generation of RNA transcript standards for qRT-PCR validation

A barrier to implementing and validating qRT-PCR molecular assays for SARS-CoV-2 detection was the availability of virus RNA standards. As the full length SARS-CoV-2 RNA is considered as a biological safety level 2 hazard in the US, we generated small RNA transcripts (704-1363 nt) from the non-structural protein 10 (nsp10), RNA-dependent RNA polymerase (RdRp), non-structural protein 14 (nsp14), envelope (E), and nucleocapsid (N) genes spanning each of the primer and probe sets in the China CDC[7], US CDC[6], Charité[5], and HKU[4] assays (**Fig. 1A**; **Table 1; Supplemental Tables 1-2**)[10]. By measuring PCR amplification using 10-fold serial dilutions of our RNA transcript standards, we found the efficiencies of each of the nine primer-probe sets to be above 90% (**Fig. 1B**), which match the criteria for an efficient qRT-PCR assay[11]. Our RNA transcripts can thus be used for assay validation, positive controls, and standards to quantify viral loads: critical steps for a diagnostic assay. Our protocol to generate the RNA transcripts is openly available[10], and any clinical or research diagnostic lab can directly request them for free through our lab website (www.grubaughlab.com).

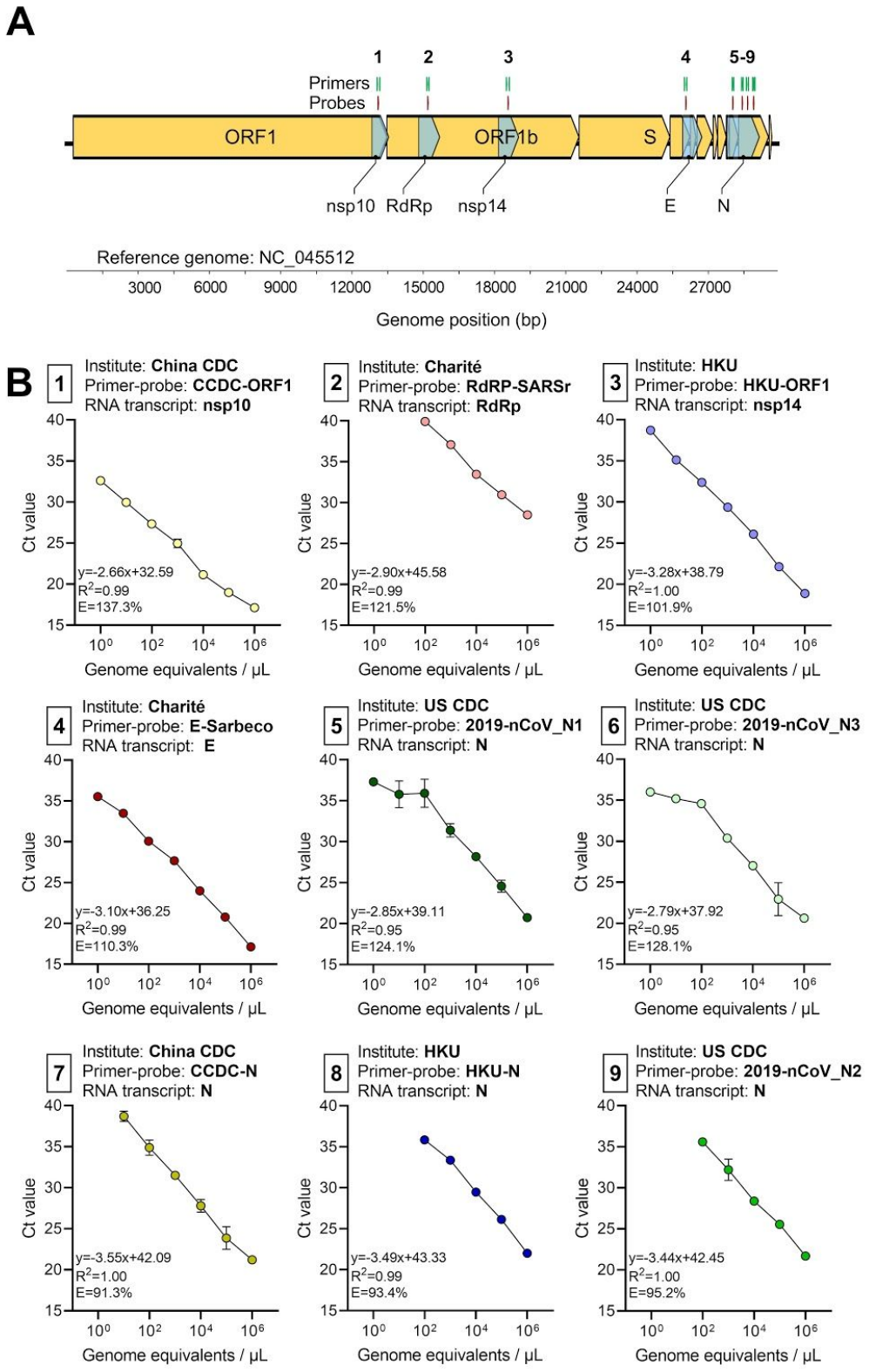


Fig. 1: Generation of RNA transcript standards for validation of SARS-CoV-2 qRT-PCR assays. (A) We reverse-transcribed RNA transcript standards for the non-structural protein 10 (nsp10), RNA-dependent RNA polymerase (RdRp), non-structural protein 14 (nsp14), envelope (E), and nucleocapsid (N) genes to be used for validation of nine primer-probe sets used in SARS-CoV-2 qRT-PCR assays. (B) We generated standard curves for nine primer-probe sets with 10-fold dilutions

(10^0-10^6 genome equivalents/μL) of the corresponding RNA transcript standards. For each combination of primer-probe set and RNA transcript standard, we provide the slope, intercept, R^2, and efficiency.

Table 1: Common qRT-PCR assays for SARS-CoV-2 diagnostics.

Institute	Target	Primer/Probe	Sequence	Reference
Charité	E	E_Sarbeco_F	ACAGGTACGTTAATAGTTAATAGCGT	[5]
		E_Sarbeco_R	ATATTGCAGCAGTACGCACACA	
		E_Sarbeco_P1	ACACTAGCCATCCTTACTGCGCTTCG	
	RdRp	RdRp_SARSr-F	GTGA**R**ATGGTCATGTGTGGCGG	
		RdRp_SARSr-R	CARATGTTAAA**S**ACACTATTAGCATA	
		RdRp_SARSr-P1	CCAGGTGG**W**AC**R**TCATC**M**GGTGATGC	
		RdRp_SARSr-P2	CAGGTGGAACCTCATCAGGAGATGC	
HKU	N	HKU-N-F	TAATCAGACAAGGAACTGATTA	[4]
		HKU-N-R	CGAAGGTGTGACTTCCATG	
		HKU-N-P	GCAAATTGTGCAATTTGCGG	
	nsp14	HKU-ORF1-F	TGGGG**Y**TTTACRGGTAACCT	
		HKU-ORF1-R	AAC**R**CGCTTAACAAAGCACTC	
		HKU-ORF1-P	TAGTTGTGATGC**W**ATCATGACTAG	
China CDC	N	CCDC-N-F	GGGGAACTTCTCCTGCTAGAAT	[7]
		CCDC-N-R	CAGACATTTTGCTCTCAAGCTG	
		CCDC-N-P	TTGCTGCTGCTTGACAGATT	
	nsp10	CCDC-ORF1-F	CCCTGTGGGTTTTACACTTAA	
		CCDC-ORF1-R	ACGATTGTGCATCAGCTGA	
		CCDC-ORF1-P	CCGTCTGCGGTATGTGGAAAGGTTATG G	
US CDC	N	2019-nCoV_N1-F	GACCCCAAAATCAGCGAAT	[6]
		2019-nCoV_N1-R	TCTGGTACTGCAGTTGAATCTG	
		2019-nCoV_N1-P	ACCCCGCATTACGTTTGGTGGACC	
	N	2019-nCoV_N2-F	TTACAAACATTGGCCGCAAA	
		2019-nCoV_N2-R	GCGCGACATTCCGAAGAA	
		2019-nCoV_N2-P	ACAATTTGCCCCCAGCGCTTCAG	
	N	2019-nCoV_N3-F	GGGAGCCTTGAATACACCAAAA	
		2019-nCoV_N3-R	TGTAGCACGATTGCAGCATTG	
		2019-nCoV_N3-P	A**Y**CACATTGGCACCCGCAATCCTG	
	Human RNase P	RP-F	AGATTTGGACCTGCGAGCG	
		RP-R	GAGCGGCTGTCTCCACAAGT	
		RP-P	TTCTGACCTGAAGGCTCTGCGCG	

Analytical comparisons of qRT-PCR primer and probe sets

Critical evaluations of the designed primer-probe sets used in the primary SARS-CoV-2 qRT-PCR detection assays are necessary to compare findings across studies, and select appropriate assays for in-house testing. Our goal in this study was to directly compare the

designed primer-probe sets, not the assays *per se*, as that would involve many different variables. To do so we used the same (i) thermocycler conditions (40 cycles of 10 seconds at 95°C and 20 seconds at 55°C); (ii) primer-probe concentrations (500 nM of forward and reverse primer, and 250 nM of probe); and (iii) PCR reagents (New England Biolabs Luna Universal One-step RT-qPCR kit) in all reactions. From our measured PCR amplification efficiencies and analytical sensitivities of detection, most primer-probes sets were comparable, except for the RdRp-SARSr (Charité) set, which had low sensitivity (**Fig. 2**).

By testing each of the nine primer-probe sets using 10-fold dilutions of SARS-CoV-2 RNA derived from cell culture (**Fig. 2A**) or 10-fold dilutions of SARS-CoV-2 RNA spiked into RNA extracted from pooled nasopharyngeal swabs from pre-COVID-19 respiratory disease patients (virus RNA-spiked mocks; **Fig. 2B**), we again found that the PCR amplification efficiencies were near or above 90% (**Fig. 2C**). To measure the analytical sensitivity of virus detection, we used the cycle threshold (CT) value in which the expected linear dilution series would cross the y-intercept when tested with 1 genome equivalent per µL of RNA. Our measured sensitivities (y-intercept CT values) were similar among most of the primer-probe sets, except for the RdRp-SARSr (Charité) set (**Fig. 2D**). We found that the CT values from the RdRp-SARSr set were usually 6-10 CTs higher (lower virus detection) than the other primer-probe sets.

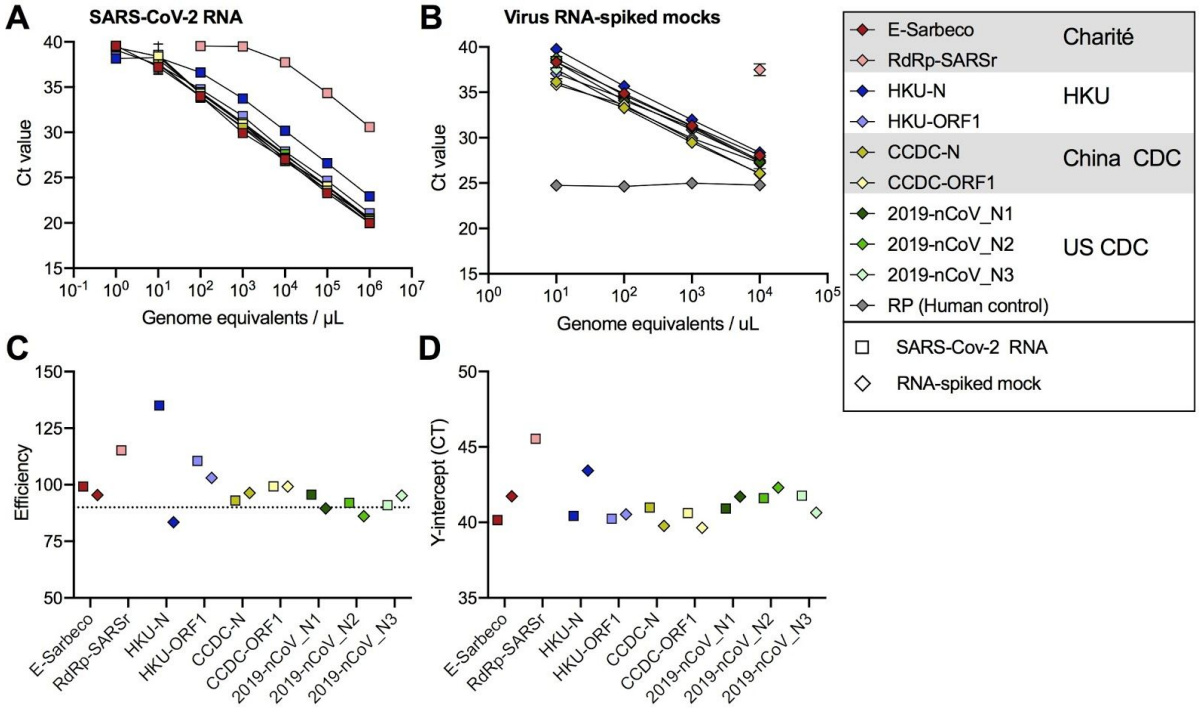


Fig. 2: Analytical efficiency and sensitivity of the nine primer-probe sets used in SARS-CoV-2 assays. We compared nine primer-probe sets and a human control primer-probe set targeting the human RNase P gene with 10-fold dilutions of (A) full-length SARS-CoV-2 RNA and (B) mock samples spiked with known concentrations of SARS-CoV-2 RNA. We extracted nucleic acid from SARS-CoV-2-negative nasopharyngeal swabs and spiked these with known concentrations of SARS-CoV-2 RNA. Symbols depict sample types: squares represent tests with SARS-CoV-2 RNA

and diamonds represent RNA-spiked mock samples. Colors depict the nine tested primer-probe sets. The CDC human RNase P (RP) assay was included as an extraction control.

Detection of virus at low concentrations and false positives

To determine the lower limit of detection, and the occurrence of false positive or inconclusive detections, we tested primer-probe sets using SARS-CoV-2 RNA spiked into RNA extracted from pooled nasopharyngeal swabs from pre-COVID-19 respiratory disease patients. Our mock samples demonstrated that many of the primer-probe sets cross-reacted with non-SARS-CoV-2 nucleic acid, which may lead to false positive results (**Fig. 3**).

When using nasopharyngeal swabs without spiked in SARS-CoV-2 RNA, we detected CT values <40 for the CCDC-N (5/8, 62.5%), CCDC-ORF1 (2/8, 25%), 2019-nCoV_N2 (2/8, 25%), and 2019-nCoV_N3 (6/8, 75%) sets, which suggests amplification of nonspecific products (**Fig. 3**). Moreover, the CT value ranges for mock samples overlapped with the CT value ranges (~36-40) for the swabs spiked with 10^0 and 10^1 virus genome equivalents/µL (**Fig. 3**), indicating that this "background noise" will limit the ability to differentiate between true positives and negatives at low virus concentrations using the CCDC-N, CCDC-ORF1, 2019-nCoV_N2, and 2019-nCoV_N3. In fact, the 2019-nCoV_N3 primer-probe set has been excluded from the US CDC assay due to these issues[12].

Of the primer-probe sets without background CT values in the SARS-CoV-2-negative mock samples (E-Sarbeco, RdRp-SARSr, HKU-N, HKU-ORF1, and 2019-nCoV_N1), our results show that none were able to detect SARS-CoV-2 RNA at 1 (10^0) virus genome equivalents/µL and mixed detection at 10 (10^1) virus genome equivalents/µL (**Fig. 3**). We found that the two most sensitive primer-probe sets are E-Sarbeco (Charité) and HKU-ORF1, which each detected 6/8 (75%) of the nasopharyngeal swabs spiked with 10 virus genome equivalents/µL (**Fig. 3**). At 100 (10^2) virus genome equivalents/µL, we could detect virus (CT <40) and differentiate between the negative mocks for all replicates and primers sets, except for the RdRp-SARSr (Charité) set, which was negative (CT >40) for all 10^0-10^2 genome equivalents/µL concentrations. Thus, our results show that there are differences in each of the primer-probe sets to differentiate between true negatives and true positives at virus concentrations at or below 10 virus genome equivalents/µL.

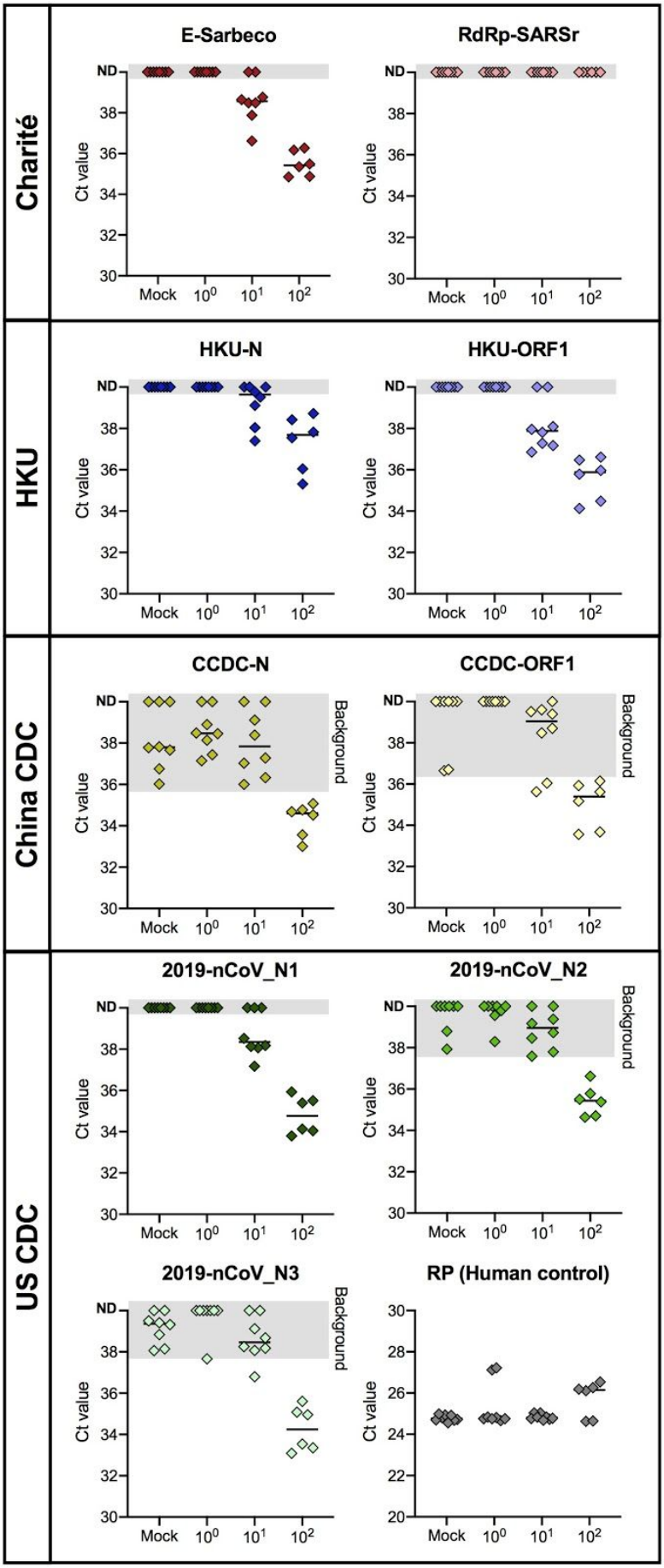


Fig. 3: All nine primer-probe sets have a similar lower detection limit of 10^2 SARS-CoV-2 genome equivalents/µL. We determined the lower detection limit of nine primer-probe sets as well as the human RNase P control for mock samples (RNA extracted from nasopharyngeal swabs collected in 2017) spiked with known concentrations of SARS-CoV-2 RNA. We performed 6-8 technical replicates with mock samples without spiking RNA and mock samples spiked with 10^0-10^2 genome equivalent/µL of SARS-CoV-2 RNA. For each primer-probe set, we show the range of cycle threshold values obtained with mock samples extracted from SARS-CoV-2-negative nasopharyngeal swabs, which indicates variation in the lower detection limit of each primer-probe set. ND = not detected. Gray-shaded areas = non-specific amplification.

Lower performance of RdRp-SARSr (Charité) set

To further investigate the relatively low performance of the RdRp-SARSr (Charité) primer-probe set, we compared our standardized primer-probe concentrations with the recommended concentrations in the confirmatory (Probe 1 and Probe 2) and discriminatory (Probe 2 only) RdRp-SARSr (Charité) assays. We deviated from the recommended concentrations in the original assays to make a fair comparison across primer-probe sets, using 500 nM of each primer and 250 nM of probe 2. To investigate the effect of primer-probe concentration on the ability to detect SARS-CoV-2, we made a direct comparison between (i) our standardized primer (500 nM) and probe (250 nM) concentrations, (ii) the recommended concentrations of 600 nM of forward primer, 800 nM of reverse primer, and 100 nM of probe 1 and 2 (confirmatory assay), and (iii) the recommended concentrations of 600 nM of forward primer, 800 nM of reverse primer, and 200 nM of probe 2 (discriminatory assay) per reaction[5]. We found that adjusting the primer-probe concentrations or using the combination of probes 1 and 2 did not increase SARS-CoV-2 RNA detection when using 10-fold serial dilutions of our RdRp RNA transcripts, or full-length SARS-CoV-2 RNA from cell culture (**Fig. 4**). The Charité Universitätsmedizin Berlin Institute of Virology assay is designed to use the E-Sarbeco primer-probes as an initial screening assay, and the RdRp-SARSr primer-probes as a confirmatory test[5]. Our data suggest that the RdRp-SARSr assay is not a reliable confirmatory assay at low virus amounts.

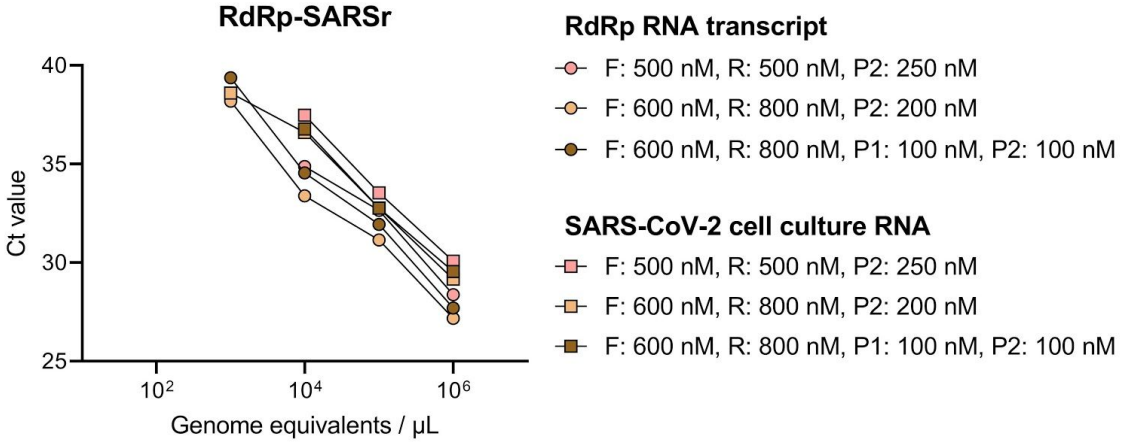


Fig 4: No effect of different concentrations of RdRp-SARSr primers and probes on analytical efficiency and sensitivity. Low performance of the standardized RdRp-SARSr primer-probe set triggered us to further investigate the effect of primer concentrations. We compared our standardized

primer-probe concentrations (500 nM of forward and reverse primers, and 250 nM of probe) with the recommended concentrations in the confirmatory assay (600 nM of forward primer, 800 nM of reverse primer, 100 nM of probe 1, and 100 nM of probe 2), and the discriminatory assay (600 nM of forward primer, 800 nM of reverse primer, and 200 nM of probe 2) as developed by the Charité Universitätsmedizin Berlin Institute of Virology. Standard curves for both RdRp-transcript standard and full-lenght SARS-CoV-2 RNA are similar, which indicates that higher primer concentrations did not improve the performance of the RdRp-SARSr set. Symbol indicates tested sample type (circles = RdRp transcript standard, and squares = full-length SARS-CoV-2 RNA) and colors indicate the different primer and probe concentrations.

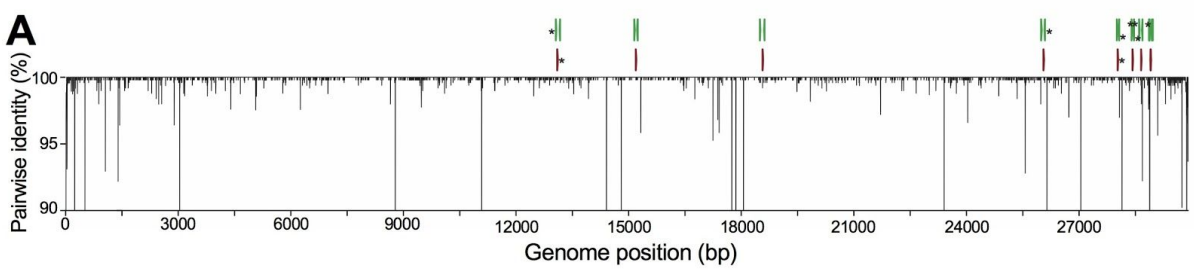
Mismatches in primer binding regions

As viruses evolve during outbreaks, nucleotide substitutions can emerge in primer or probe binding regions that can alter the sensitivity of PCR assays. To investigate whether this had already occurred during the early COVID-19 pandemic, we calculated the accumulated genetic diversity from 992 available SARS-CoV-2 genomes (**Fig. 5A**) and compared that to the primer and probe binding regions (**Fig. 5B**). Thus far we detected 12 primer-probe nucleotide mismatches that have occurred in at least two of the 992 SARS-CoV-2 genomes.

The most potentially problematic mismatch is in the RdRp-SARSr reverse primer (**Fig. 5B**), which likely explains our sensitivity issues with this set (**Figs. 2-4**). Oddly, the mismatch is not derived from a new variant that has arisen, but rather that the primer contains a degenerate nucleotide (S, binds with G or C) at position 12, and 990 of the 992 SARS-CoV-2 genomes encode for a T at this genome position (**Fig. 5B**). This degenerate nucleotide appears to have been added to help the primer anneal to SARS-CoV and bat-SARS-related CoV genomes[5], seemingly to the detriment of consistent SARS-CoV-2 detection. Earlier in the outbreak, before hundreds of SARS-CoV-2 genomes became available, non-SARS-CoV-2 data were used to infer genetic diversity that could be anticipated during the outbreak. As a result, several of the primers contain degenerate nucleotides (**Supplemental Table 3**). For RdRp-SARSr, adjusting the primer (S→A) may resolve its low sensitivity.

Of the variants that we detected in the primer-probe regions, we only found four in more than 30 of the 992 SARS-CoV-2 genomes (>3%, **Fig. 5B**). Most notable was a stretch of three nucleotide substitutions (GGG→AAC) at genome positions 28,881-28,883, which occur in the three first positions of the CCDC-N forward primer binding site. While these substitutions define a large clade that includes ~13% of the available SARS-CoV-2 genomes and has been detected in numerous countries[13], their position on the 5' location of the primer may not be detrimental to sequence annealing and amplification. The other high frequency variant that we detected was T→C substitution at the 8th position of the binding region of the 2019-nCoV_N3 forward primer, a substitution found in 39 genomes (position 28,688). While this primer could be problematic for detecting viruses with this variant, the 2019-nCoV_N3 set has already been removed from the US CDC assay. We found another seven variants in only five or fewer genomes (<0.5%, **Fig. 5B**), and their minor frequency at present does not pose a major concern for viral detection. This scenario may change if those variants increase in frequency: most of them lie in the second half of the primer binding region, and

may decrease primer sensitivity[14]. The WA1_USA strain (GenBank: MN985325) that we used for our comparisons did not contain any of these variants.



B

Institute	Primer/probe	Primer/probe position 5'-3'	Genome position 5'-3'	Primer/probe nucletotide (*RC)	Mismatch nucleotide (frequency)
China CDC	CCDC-N-Fwd	1	28,881	G	A (126/992; 12.7%)
	CCDC-N-Fwd	2	28,882	G	A (126/992; 12.7%)
	CCDC-N-Fwd	3	28,883	G	C (126/992; 12.7%)
	CCDC-ORF1-Fwd	17	13,358	C	T (2/992; 0.2%)
	CCDC-ORF1-Probe	26	13,402	T	G (4/992; 0.4%)
Charité	E_Sarbeco_R	12	26,370	C (G)	A (4/992; 0.4%)
	RdRp-SARSr_R	12	15,519	S (G or C)	**T (990/992; 99.8%)**
HKU	HKU-N-Fwd	4	29,148	T	C (5/992; 0.5%)
US CDC	2019-nCoV_N1-P	3	28,311	C	T (2/992; 0.2%)
	2019-nCoV_N1-R	15	28,344	G (C)	T (4/992; 0.4%)
	2019-nCoV_N3-F	8	28,688	T	C (39/992; 3.9%)
	2019-nCoV_N3-R	14	28,739	C (G)	A (4/992; 0.4%)

*RC, reverse complement

Fig. 5: High frequency primer and probe mismatches may result in decreased sensitivity for SARS-CoV-2 detection. (A) We aligned nucleotide diversity across 992 SARS-CoV-2 genomes sequenced up to 22 March 2020 and determined mismatches with the nine primer-probe sets. We measured diversity using pairwise identity (%) at each position, disregarding gaps and ambiguous nucleotides. Asterisks (*) at the top indicate primers and probes targeting regions with one or more mismatches. Genomic plots were designed using DNA Features Viewer in Python[15]. (B) We only listed mismatch nucleotides with frequencies above 0.1%. These mismatches may result in decreased sensitivity of primer-probe sets.

Conclusions

Our comparative results of primer-probe sets used in qRT-PCR assays indicate that overall, all assays are able to detect SARS-COV-2; however, detection limits and ability to differentiate between true negatives and positives at low RNA concentrations are variable between sets. This should be carefully evaluated to determine CT value cut-offs to differentiate between positives and negatives. The US CDC assay, for example, uses a cut-off value of CT 40, but we generated CT values in the range of 37-40 when the 2019-nCoV_N2 set was tested on RNA from nasopharyngeal swabs void of SARS-CoV-2 RNA. Considering that both the US CDC 2019-nCoV_N1 and N2 sets need to be >40 CTs to be considered as negative, background amplification in one of the sets would result in inconclusive results.

Overall, we found that the most sensitive primer-probe sets are E-Sarbeco (Charité), HKU-ORF1 (HKU), and 2019-nCoV_N1 (US CDC). In contrast, the RdRp-SARSr (Charité) primer-probe set had the lowest sensitivity, likely stemming from a mismatch in the reverse primer. Importantly, sensitivity as reported in our study may not be applicable to other PCR

kits or thermocyclers; analytical sensitivities and positive-negative cut-off values should be locally validated when establishing these assays.

Methods

Ethics

Residual de-identified nasopharyngeal samples from patients with suspected respiratory infections were obtained from the Yale-New Haven Hospital Clinical Virology Laboratory in accordance with human subjects protections using a protocol approved by the Yale Human Investigations committee.

Generation of RNA transcript standards

We generated RNA transcript standards for each of the five genes targeted by the diagnostic qRT-PCR assays using T7 transcription. A detailed protocol can be found here[10]. Briefly, cDNA was synthesized from full-length SARS-CoV-2 RNA (WA1_USA strain from UTMB; GenBank: MN985325). Using PCR, we amplified the nsp10, RdRp, nsp14, E, and N genes with specifically designed primers (**Supplemental Table 1**). We purified PCR products using the Mag-Bind TotalPure NGS kit (Omega Bio-tek, Norcross, GA, USA) and quantified products using the Qubit High Sensitivity DNA kit (ThermoFisher Scientific, Waltham, MA, USA). We determined fragment sizes using the DNA 1000 kit on the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). After quantification, we transcribed 100-200 ng of each purified PCR product into RNA using the Megascript T7 kit (ThermoFisher Scientific). We quantified RNA transcripts using the Qubit High sensitivity RNA kit (ThermoFisher Scientific) and checked quality using the Bioanalyzer RNA pico 6000 kit. For each of the RNA transcript standards (**Supplemental Table 2**), we calculated the number of genome copies per µL using Avogadro's number. We generated a genomic annotation plot with all newly generated RNA transcript standards and the nine tested primer-probe sets based on the NC_045512 reference genome using the DNA Features Viewer Python package (**Fig. 1A**)[15]. We generated standard curves for each combination of primer-probe set with its corresponding RNA transcript standard (**Fig. 1B**), using standardized qRT-PCR conditions as described below.

qRT-PCR conditions

To make a fair comparison between nine primer-probe sets (**Table 1)**, we used the same qRT-PCR reagents and conditions for all comparisons. We used the Luna Universal One-step RT-qPCR kit (New England Biolabs, Ipswich, MA, USA) with standardized primer and probe concentrations of 500 nM of forward and reverse primer, and 250 nM of probe for all comparisons. PCR cycler conditions were reverse transcription for 10 minutes at 55°C, initial denaturation for 1 min at 95°C, followed by 40 cycles of 10 seconds at 95°C and 20 seconds at 55°C on the Biorad CFX96 qPCR machine (Biorad, Hercules, CA, USA). We calculated analytical efficiency of qRT-PCR assays tested with corresponding RNA transcript standards using the following formula[16,17]:

$$E = 100 \times (10^{-1/slope} - 1)$$

Validation with SARS-CoV-2 RNA and mock samples

We prepared mock samples by extracting RNA from 12 de-identified nasopharyngeal swabs collected in 2017 (pre-SARS-CoV-2) from hospital patients with respiratory disease using the MagMAX Viral/Pathogen Nucleic Acid Isolation kit (ThermoFisher Scientific) following manufacturer's protocol. After nucleic acid extraction, we spiked mock samples with 10-fold dilutions of SARS-CoV-2 RNA. We compared analytical efficiency and sensitivity of qRT-PCR assays by testing 10-fold dilutions (10^6-10^0 genome equivalents/µL) of SARS-CoV-2 RNA as well as the RNA-spiked mock samples, in duplicate. In addition, we determined analytical sensitivity of the nine primer-probe sets by testing 6-8 replicates of high dilutions of RNA-spiked mock samples (10^2-10^0 genome equivalents/µL) and mock samples without addition of RNA.

Mismatches in primer binding regions

We investigated mismatches in primer binding regions by calculating pairwise identities (%) for each nucleotide position in binding sites of assay primers and probes. Ignoring gaps and ambiguous bases, we compared all possible pairs of nucleotides in all columns of a multiple sequence alignment including all available SARS-CoV-2 genomes (as of 22 March 2020). We assigned a score of 1 for each identical pair of bases, and divided the final score by the total number of valid nucleotide pairs, to finally express pairwise identities as percentages. Pairwise identity of less than 100% indicates mismatches between primers or probes and some SARS-CoV-2 genomes. We calculated mismatch frequencies and reported absolute and relative frequencies for mismatches with frequency higher than 0.1%. The DNA Features Viewer package in Python was used to generate the diversity plot (**Fig. 5**)[15].

Acknowledgements

We thank K. Plante and the University of Texas Medical Branch World Reference Center for Emerging Viruses for providing SARS-CoV-2 RNA, the Yale COVID-19 Laboratory Working Group for technical support, and P. Jack and S. Taylor for discussions. This research was funded by the generous support from the Yale Institute for Global Health and the Yale School of Public Health start-up package provided to NDG. CBFV is supported by NWO Rubicon 019.181EN.004.

References

1. Gorbalenya, A. E. *et al.* Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. *bioRxiv* 2020.02.07.937862 (2020) doi:10.1101/2020.02.07.937862.
2. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
3. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
4. Chu, D. K. W. *et al.* Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. *Clin. Chem.* (2020) doi:10.1093/clinchem/hvaa029.
5. Corman, V. M. *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* **25**, (2020).
6. CDC. Coronavirus Disease 2019 (COVID-19). *Centers for Disease Control and Prevention* https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html (2020).
7. Institute of Viral Diseases. China CDC. *National Institute for Viral Disease Control and Prevention* http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html (2020).
8. WHO - Coronavirus disease (COVID-19) technical guidance: Laboratory testing for 2019-nCoV in humans. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance.
9. Harcourt, J. *et al.* Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with 2019 Novel Coronavirus Disease, United States. *Emerg. Infect. Dis.* **26**, (2020).
10. Vogels, C. B. F., Fauver, J. R., Ott, I. & Grubaugh, N. D. Generation of SARS-COV-2 RNA transcript standards for qRT-PCR detection assays v1 (protocols.io.bdv6i69e). (2020) doi:10.17504/protocols.io.bdv6i69e.

11. Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M. W. & Kubista, M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomol Detect Quantif* **3**, 9–16 (2015).
12. CDC Revises SARS-CoV-2 Assay Protocol; Surveillance Testing on Track to Start Next Week. *360Dx* https://www.360dx.com/pcr/cdc-revises-sars-cov-2-assay-protocol-surveillance-testing-track-start-next-week.
13. Genomic epidemiology of novel coronavirus. *Nextstrain* https://nextstrain.org/ncov?c=gt-ORF14_50.
14. Bru, D., Martin-Laurent, F. & Philippot, L. Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. *Appl. Environ. Microbiol.* **74**, 1660–1663 (2008).
15. Zulkower, V. & Rosser, S. DNA Features Viewer, a sequence annotations formatting and plotting library for Python. *bioRxiv* 2020.01.09.900589 (2020) doi:10.1101/2020.01.09.900589.
16. Broeders, S. *et al.* Guidelines for validation of qualitative real-time PCR methods. *Trends Food Sci. Technol.* **37**, 115–126 (2014).
17. Ginzinger, D. G. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp. Hematol.* **30**, 503–512 (2002).

Supplement

Supplemental Table 1: Primers for generation of T7 RNA transcript standards for SARS-CoV-2.

Target	Primer	Sequence
nsp10	nsp10-Std-T7-Fwd	TAATACGACTCACTATAGGGGTGGGGGACAACCAATCACT
	nsp10-Std-Rev	AGACGAGGTCTGCCATTGTG
RdRp	RdRp-Std-T7-Fwd	TAATACGACTCACTATAGGGAATAGAGCTCGCACCGTAGC
	RdRp-Std-Rev	CATCTACAAAACAGCCGGCC
nsp14	nsp14-Std-T7-Fwd	TAATACGACTCACTATAGGGTAGTGCTAAACCACCGCCTG
	nsp14-Std-Rev	AACTGCCACCATCACAACCA
E	E-Std-T7-Fwd	TAATACGACTCACTATAGGGGCGTGCCTTTGTAAGCACAA
	E-Std-Rev	GGCAGGTCCTTGATGTCACA
N	N-Std-T7-Fwd	TAATACGACTCACTATAGGGGAATTGTGCGTGGATGAGGC
	N-Std-Rev	TGTCTCTGCGGTAAGGCTTG

Supplemental Table 2: RNA transcript standards for common SARS-CoV-2 diagnostic assays (see genomic context on Figure 1A).

Gene	Length	Sequence
nsp10	704nt (13,122 - 13,825)	GUGGGGGACAACCAAUCACUAAUUGUGUUAAGAUGUUGUGUACA CACACUGGUACUGGUCAGGCAAUAACAGUUACACCGGAAGCCAAU AUGGAUCAAGAAUCCUUUGGUGGUGCAUCGUGUUGUCUGUACUG CCGUUGCCACAUAGAUCAUCCAAAUCCUAAAGGAUUUUGUGACUU AAAAGGUAAGUAUGUACAAAUACCUACAACUUGUGCUAAUGACCC UGUGGGUUUUACACUUAAAAACACAGUCUGUACCGUCUGCGGUA UGUGGAAAGGUUAUGGCUGUAGUUGUGAUCAACUCCGCGAACCC AUGCUUCAGUCAGCUGAUGCACAAUCGUUUUUAAACGGGUUUGC GGUGUAAGUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAG UACUGAUGUCGUAUACAGGGCUUUUGACAUCUACAAUGAUAAAGU AGCUGGUUUUGCUAAAUUCCUAAAAACUAAUUGUUGUCGCUUCCA AGAAAAGGACGAAGAUGACAAUUUAAUUGAUUCUUACUUUGUAGU UAAGAGACACACUUUCUCUAACUACCAACAUGAAGAAACAAUUUAU AAUUUACUUAAGGAUUGUCCAGCUGUUGCUAAACAUGACUUCUUU AAGUUUAGAAUAGACGGUGACAUGGUACCACAUAUAUCACGUCAA CGUCUUACUAAAUACACAAUGGCAGACCUCGUCU
RdRp	883nt (15,094 - 15,976)	AAUAGAGCUCGCACCGUAGCUGGUGUCUCUAUCUGUAGUACUAU GACCAAUAGACAGUUUCAUCAAAAAUUAUUGAAAUCAAUAGCCGC CACUAGAGGAGCUACUGUAGUAAUUGGAACAAGCAAAUUCUAUGG UGGUUGGCACAACAUGUUAAAAACUGUUUAUAGUGAUGUAGAAAA CCCUCACCUUAUGGGUUGGGAUUAUCCUAAAUGUGAUAGAGCCA UGCCUAACAUGCUUAGAAUUAUGGCCUCACUUGUUCUUGCUCGC AAACAUACAACGUGUUGUAGCUUGUCACACCGUUUCUAUAGAUUA GCUAAUGAGUGUGCUCAAGUAUUGAGUGAAAUGGUCAUGUGUGG CGGUUCACUAUAUGUUAAACCAGGUGGAACCUCAUCAGGAGAUGC CACAACUGCUUAUGCUAAUAGUGUUUUUAACAUUUGUCAAGCUGU CACGGCCAAUGUUAAUGCACUUUUAUCUACUGAUGGUAACAAAAU UGCCGAUAAGUAUGUCCGCAAUUUACAACACAGACUUUAUGAGUG UCUCUAUAGAAAUAGAGAUGUUGACACAGACUUUGUGAAUGAGUU UUACGCAUAUUUGCGUAAACAUUUCUCAAUGAUGAUACUCUCUGA CGAUGCUGUUGUGUGUUUCAAUAGCACUUAUGCAUCUCAAGGUC UAGUGGCUAGCAUAAAGAACUUUAAGUCAGUUCUUUAUUAUCAAA ACAAUGUUUUUAUGUCUGAAGCAAAAUGUUGGACUGAGACUGACC UUACUAAAGGACCUCAUGAAUUUUGCUCUCAACAUACAAUGCUAG UUAAACAGGGUGAUGAUUAUGUGUACCUUCCUUACCCAGAUCCAU CAAGAAUCCUAGGGGCCGGCUGUUUUGUAGAUG
nsp14	848nt (18,447-19,294)	UAGUGCUAAACCACCGCCUGGAGAUCAAUUUAAACACCUCAUACC ACUUAUGUACAAAGGACUUCCUUGGAAUGUAGUGCGUAUAAAGAU UGUACAAAUGUUAAGUGACACACUUAAAAAUCUCUCUGACAGAGU CGUAUUUGUCUUAUGGGCACAUGGCUUUGAGUUGACAUCUAUGA AGUAUUUUGUGAAAAUAGGACCUGAGCGCACCUGUUGUCUAUGU GAUAGACGUGCCACAUGCUUUUCCACUGCUUCAGACACUUAUGCC UGUUGGCAUCAUUCUAUUGGAUUUGAUUACGUCUAUAAUCCGUU UAUGAUUGAUGUUCAACAAUGGGGUUUUACAGGUAACCUACAAAG CAACCAUGAUCUGUAUUGUCAAGUCCAUGGUAAUGCACAUGUAGC UAGUUGUGAUGCAAUCAUGACUAGGUGUCUAGCUGUCCACGAGU GCUUUGUUAAGCGUGUUGACUGGACUAUUGAAUAUCCUAUAAUU GGUGAUGAACUGAAGAUUAAUGCGGCUUGUAGAAAGGUUCAACAC AUGGUUGUUAAAGCUGCAUUAUUAGCAGACAAAUUCCCAGUUCUU CACGACAUUGGUAACCCUAAAGCUAUUAAGUGUGUACCUCAAGCU GAUGUAGAAUGGAAGUUCUAUGAUGCACAGCCUUGUAGUGACAAA GCUUAUAAAAUAGAAGAAUUAUUCUAUUCUUAUGCCACACAUUCU GACAAAUUCACAGAUGGUGUAUGCCUAUUUUGGAAUUGCAAUGUC GAUAGAUAUCCUGCUAAUUCCAUUGUUUGUAGAUUUGACACUAGA GUGCUAUCUAACCUUAACUUGCCUGGUUGUGAUGGUGGCAGUU
Envelope (E)	808nt (26,207 -	GCGUGCCUUUGUAAGCACAAGCUGAUGAGUACGAACUUAUGUAC UCAUUCGUUUCGGAAGAGACAGGUACGUUAAUAGUUAAUAGCGUA

	27,116)	CUUCUUUUUCUUGCUUUCGUGGUAUUCUUGCUAGUUACACUAGC CAUCCUUACUGCGCUUCGAUUGUGUGCGUACUGCUGCAAUAUUG UUAACGUGAGUCUUGUAAAACCUUCUUUUUACGUUUACUCUCGUG UUAAAAAUCUGAAUUCUUCUAGAGUUCCUGAUCUUCUGGUCUAAA CGAACUAAAUAUUAUAUUAGUUUUUCUGUUUGGAACUUUAAUUUU AGCCAUGGCAGAUUCCAACGGUACUAUUACCGUUGAAGAGCUUAA AAAGCUCCUUGAACAAUGGAACCUAGUAAUAGGUUUCCUAUUCCU UACAUGGAUUUGUCUUCUACAAUUUGCCUAUGCCAACAGGAAUAG GUUUUUGUAUAUAAUUAAGUUAAUUUUCCUCUGGCUGUUAUGGC CAGUAACUUUAGCUUGUUUUGUGCUUGCUGCUGUUUACAGAAUA AAUUGGAUCACCGGUGGAAUUGCUAUCGCAAUGGCUUGUCUUGU AGGCUUGAUGUGGCUCAGCUACUUCAUUGCUUCUUUCAGACUGU UUGCGCGUACGCGUUCCAUGUGGUCAUUCAAUCCAGAAACUAACA UUCUUCUCAACGUGCCACUCCAUGGCACUAUUCUGACCAGACCGC UUCUAGAAAGUGAACUCGUAAUCGGAGCUGUGAUCCUUCGUGGA CAUCUUCGUAUUGCUGGACACCAUCUAGGACGCUGUGACAUCAA GGACCUGCC
Nucleocapsid (N)	1363nt (28,068 - 29,430)	GAAUUGUGCGUGGAUGAGGCUGGUUCUAAAUCACCCAUUCAGUA CAUCGAUAUCGGUAAUUAUACAGUUUCCUGUUUACCUUUUACAAU UAAUUGCCAGGAACCUAAAUUGGGUAGUCUUGUAGUGCGUUGUU CGUUCUAUGAAGACUUUUUAGAGUAUCAUGACGUUCGUGUUGUU UUAGAUUUCAUCUAAACGAACAAACUAAAAUGUCUGAUAAUGGAC CCCAAAAUCAGCGAAAUGCACCCCGCAUUACGUUUGGUGGACCCU CAGAUUCAACUGGCAGUAACCAGAAUGGAGAACGCAGUGGGGCG CGAUCAAAACAACGUCGGCCCCAAGGUUUACCCAAUAAUACUGCG UCUUGGUUCACCGCUCUCACUCAACAUGGCAAGGAAGACCUUAAA UUCCCUCGAGGACAAGGCGUUCCAAUUAACACCAAUAGCAGUCCA GAUGACCAAAUUGGCUACUACCGAAGAGCUACCAGACGAAUUCGU GGUGGUGACGGUAAAAUGAAAGAUCUCAGUCCAAGAUGGUAUUU CUACUACCUAGGAACUGGGCCAGAAGCUGGACUUCCCUAUGGUG CUAACAAAGACGGCAUCAUAUGGGUUGCAACUGAGGGAGCCUUG AAUACACCAAAAGAUCACAUUGGCACCCGCAAUCCUGCUAACAAU GCUGCAAUCGUGCUACAACUUCCUCAAGGAACAACAUUGCCAAAA GGCUUCUACGCAGAAGGGAGCAGAGGCGGCAGUCAAGCCUCUUC UCGUUCCUCAUCACGUAGUCGCAACAGUUCAAGAAAUUCAACUCC AGGCAGCAGUAGGGGAACUUCUCCUGCUAGAAUGGCUGGCAAUG GCGGUGAUGCUGCUCUUGCUUUGCUGCUGCUUGACAGAUUGAAC CAGCUUGAGAGCAAAAUGUCUGGUAAAGGCCAACAACAACAAGGC CAAACUGUCACUAAGAAAUCUGCUGCUGAGGCUUCUAAGAAGCCU CGGCAAAAACGUACUGCCACUAAAGCAUACAAUGUAACACAAGCU UUCGGCAGACGUGGUCCAGAACAAACCCAAGGAAAUUUUGGGGA CCAGGAACUAAUCAGACAAGGAACUGAUUACAAACAUUGGCCGCA AAUUGCACAAUUUGCCCCCAGCGCUUCAGCGUUCUUCGGAAUGU CGCGCAUUGGCAUGGAAGUCACACCUUCGGGAACGUGGUUGACC UACACAGGUGCCAUCAAAUUGGAUGACAAAGAUCCAAAUUUCAAA GAUCAAGUCAUUUUGCUGAAUAAGCAUAUUGACGCAUACAAAACA UUCCCACCAACAGAGCCUAAAAAGGACAAAAAGAAGAAGGCUGAU GAAACUCAAGCCUUACCGCAGAGACA

Supplemental Table 3: Degenerate bases in common SARS-CoV-2 qRT-PCR assay primers and probes.

Primer	Degenerate base, and its purpose	Position in primer (5'-3')	Genomic position (5'-3')	Pairing base in genomes (frequency)
RdRp_SARSr-F	R, to pair with T or C	5	15,435	T (992/992; 100.0%)
RdRp_SARSr-R	**S, to pair with C or G**	12	15,519	T (990/992; 99.8%)
RdRp_SARSr-R	R, to pair with T or C	3	15,528	T (992/992; 100.0%)
HKU-ORF1-Fwd	Y, to pair with A or G	6	18,783	A (992/992; 100.0%)
HKU-ORF1-Fwd	R, to pair with T or C	12	18,789	T (989/992; 99.7%)
HKU-ORF1-Probe	W, to pair with T or A	13	18,861	T (992/992; 100.0%)
HKU-ORF1-Rev	R, to pair with T or C	4	18,906	T (992/992; 100.0%)
2019-nCoV_N3-P	Y, to pair with A or G	2	28,705	A (992/992; 100.0%)