medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

PCR assay to enhance global surveillance for SARS-CoV-2 variants of concern

Chantal B.F. Vogels[1†*], Mallery Breban[1†], Tara Alpert[1], Mary E. Petrone[1], Anne E. Watkins[1], Emma B. Hodcroft[2], Christopher E. Mason[3], Gaurav Khullar[3], Jessica Metti[3], Joel T. Dudley[3], Matthew J. MacKay[3], Megan Nash[3], Jianhui Wang[4], Chen Liu[4], Pei Hui[4], Steven Murphy[5], Caleb Neal[5], Eva Laszlo[5], Marie L. Landry[6], Anthony Muyombwe[7], Randy Downing[7], Jafar Razeq[7], Richard A. Neher[8,9], Joseph R. Fauver[1#*], Nathan D. Grubaugh[1,10#*]

[1] Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT 06510, USA
[2] Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
[3] Tempus Labs, Chicago, IL 60654, USA
[4] Department of Pathology, Yale University School of Medicine, New Haven, CT 06510, USA
[5] Murphy Medical Associates, Greenwich, CT 06614, USA
[6] Departments of Laboratory Medicine and Medicine, Yale School of Medicine, New Haven, CT 06510, USA
[7] Connecticut State Department of Public Health, Rocky Hill, CT 06067, USA
[8] Biozentrum, University of Basel, 4056 Basel, Switzerland
[9] Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
[10] Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06510, USA
[†] These authors contributed equally
[#] Senior authors
[*] Correspondence: Chantal B.F. Vogels (chantal.vogels@yale.edu), Joseph R. Fauver (joseph.fauver@yale.edu), Nathan D. Grubaugh (nathan.grubaugh@yale.edu)

Abstract

With the emergence of SARS-CoV-2 variants that may increase transmissibility and/or cause escape from immune responses[1–3], there is an urgent need for the targeted surveillance of circulating lineages. It was found that the B.1.1.7 (also 501Y.V1) variant first detected in the UK[4,5] could be serendipitously detected by the ThermoFisher TaqPath COVID-19 PCR assay because a key deletion in these viruses, spike Δ69-70, would cause a “spike gene target failure” (SGTF) result. However, a SGTF result is not definitive for B.1.1.7, and this assay cannot detect other variants of concern that lack spike Δ69-70, such as B.1.351 (also 501Y.V2) detected in South Africa[6] and P.1 (also 501Y.V3) recently detected in Brazil[7]. We identified a deletion in the ORF1a gene (ORF1a Δ3675-3677) in all three variants, which has not yet been widely detected in other SARS-CoV-2 lineages. Using ORF1a Δ3675-3677 as the primary target and spike Δ69-70 to differentiate, we designed and validated an open source PCR assay to detect SARS-CoV-2 variants of concern[8]. Our assay can be rapidly deployed in laboratories around the world to enhance surveillance for the local emergence spread of B.1.1.7, B.1.351, and P.1.

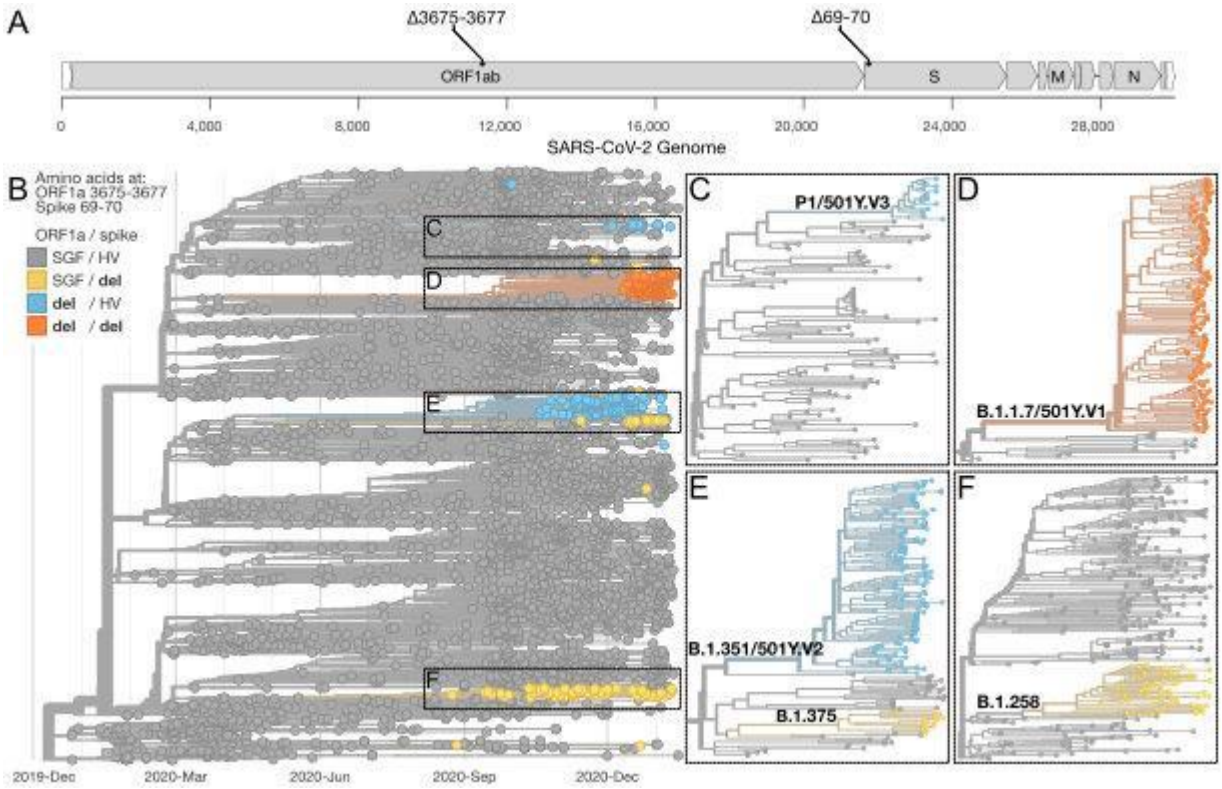
NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Main

Broadly accessible and inexpensive surveillance methods are needed to track SARS-CoV-2 variants of concern around the world. While sequencing is the gold standard to identify circulating SARS-CoV-2 variants, routine genomic surveillance is not available in most countries primarily due to a lack of resources and expertise. In the current situation with the identification of the variants of concern B.1.1.7, B.1.351, and P.1, and with the likelihood that more will emerge, a lack of genomic surveillance leaves public health authorities with a patchy and skewed picture to inform decision making. The discovery of B.1.1.7 variants causing SGTF results when tested using the TaqPath PCR assay provided labs in the UK and throughout Europe with a ready-made, simple tool for tracking the frequencing of this variant[9,10]. As B.1.1.7 spread to other countries, TaqPath SGTF results were used as a front line screening tool for sequencing and an approximation for B.1.1.7 population frequency[11]. These findings highlight the usefulness of a PCR assay that produces distinctive results when targeting variants in virus genomes for both tracking and sequencing prioritization.

The TaqPath assay was not specifically designed for SARS-CoV-2 variant surveillance, and it has several limitations. The 6 nucleotide deletion in the spike gene at amino acid positions 69 and 70 (spike Δ69-70) that causes the TaqPath SGTF is also present in other SARS-CoV-2 lineages (**Fig. 1**, **Supplementary Table 1**), most notably Pango lineages B.1.258 detected throughout Europe and B.1.375 detected primarily in the US[12,13], meaning that SGTF results are not definitive for B.1.1.7. Furthermore, too much focus on TaqPath SGTF results will leave blindspots for other emerging SARS-CoV-2 variants of concern that do not have spike Δ69-70. In particular, B.1.351 and P.1, which were recently discovered in South Africa and Brazil, respectively, may also be more transmissible and contain mutations that could help to evade immune responses[6,7,14]. For all of these reasons, a PCR assay specifically designed for variant surveillance would help to fill in many of the gaps about their distribution and frequency.



medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1. Identification of genome targets to differentiate between B.1.1.7, B.1.351, P.1, and other SARS-CoV-2 lineages. (**A**) Location on the SARS-CoV-2 genome where the targeted deletions in the ORF1a gene at amino acid positions 3675-3677 (Δ3675-3677) and the spike gene at amino acid positions 69-70 (Δ69-70) occur. (**B**) The Nextstrain 'global build' (nextstrain.org/ncov/global) accessed on 2021-01-22 showing the phylogenetic representation of 4,046 SARS-CoV-2 genomes colored by the presence of deletions at amino acid positions ORF1a 3575-3677 and spike 69-70. (**C-F**) Zooms of large SARS-CoV-2 clades, which include the variants of concern B.1.1.7, B.1.351, and P.1, containing one or both deletions. A list of SARS-CoV-2 genomes used in the analysis is available in **Source Data Fig. 1.**

We analyzed over 400,000 SARS-CoV-2 genomes on GISAID and used custom Nextstrain builds[15] to identify that a 9 nucleotide deletion in the ORF1a gene at amino acid positions 3675-3677 (ORF1a Δ3675-3677) occurs in the B.1.1.7, B.1.351, and P.1 variants, but is only found in 0.03% (103/377,011) of all other genomes (**Fig. 1**, **Supplementary Table 1**). Within the B.1.351 lineage, however, 18.4% of the sequences do not have ORF1a Δ3675-3677 (**Supplementary Table 1**, not shown in Fig. 1E). Therefore, by designing a PCR assay that targets both ORF1a Δ3675-3677 and spike Δ69-70 (**Fig. 1A**), we can detect most viruses from all three current variants of concern (ORF1a results, **Fig. 1B-E**), differentiate B.1.1.7 (ORF1a and spike results, **Fig. 1D-F**), and provide results similar to TaqPath SGTF to compare dataset (spike results).

To create a multiplexed RT-qPCR screening assay for the B.1.1.7, B.1.351, and P.1 variants, we designed two sets of primers that flank each of ORF1a Δ3675-3677 and spike Δ69-70 and probes specific to the undeleted "wildtype" sequences. As a control, we included the CDC N1 primer and probe set that will detect both the wildtype and variant viruses. As designed, testing SARS-CoV-2 RNA that contains ORF1a Δ3675-3677 and/or spike Δ69-70 will generate undetected cycle threshold (Ct) values with the specific PCR target sets as the probes cannot anneal to the deleted sequences, but will have "positive" N1 Ct values. This configuration ensures that target failures are likely due to the presence deletions and that there is sufficient virus RNA for sequencing confirmation. Our RT-qPCR conditions are highly similar to our previously published SARS-CoV-2 multiplex assay[16], and a detailed protocol is openly available[8].

We evaluated the analytical sensitivity of our multiplexed RT-qPCR assay using synthetic RNA designed based on the original Wuhan-Hu-1 sequence and a B.1.1.7 sequence (England/205041766/2020). As the B.1.1.7 sequence contains both ORF1a Δ3675-3677 and spike Δ69-70 and the Wuhan-Hu-1 sequence contains neither deletion, using these RNAs allows us to fully evaluate the designed primer and probe sets. We tested a two-fold dilution series from 100 copies/µL to 1 copy/µL for both RNA controls in triplicate (**Table 1**). Using the Wuhan-Hu-1 RNA, we found similar detection (within 1 Ct) across all three N1, ORF1a, and spike targets, and all three could detect virus RNA at our lowest concentration of 1 copy/µL, indicating that our primer and probes sets were efficiently designed. Using the B.1.1.7 RNA, we again could detect the RNA down to 1 copy/µL with the N1 set, but did not detect any concentration of the virus RNA with the ORF1a and spike sets, confirming the expected "target failure" signature when testing viruses containing both ORF1a Δ3675-3677 and spike Δ69-70. Overall, our PCR screening assay could easily differentiate between SARS-CoV-2 RNA with and without the ORF1a and spike deletions by comparing the Ct values to the N1 control.

Table 1: Analytical sensitivity of the multiplexed RT-qPCR assay to screen for variants of concern using primer/probe sets targeting key deletions. Listed are cycle threshold (Ct) values for the three primer-probe

medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

sets targeting the SARS-CoV-2 nucleocapsid (N1-FAM), ORF1a 3675-3766 deletion (ORF1a-Cy5), and spike 69-70 deletion (Spike-HEX). Two-fold dilutions of synthetic control RNA (Wuhan-Hu-1 and B.1.1.7) were tested in triplicate.

RNA	Concentration	N1-FAM			ORF1a-Cy5			Spike-HEX		
Wuhan-Hu-1	100 copies/µL	30.4	30.5	30.4	31.1	31.2	31.1	30.8	30.8	30.9
	50 copies/µL	31.5	31.5	31.4	32.1	32.1	32.0	31.9	31.6	31.7
	25 copies/µL	32.3	32.7	32.7	33.0	33.1	33.2	32.8	32.8	33.0
	12 copies/µL	33.3	33.7	33.3	34.1	34.2	34.2	34.0	33.6	34.1
	6 copies/µL	34.5	34.3	34.9	35.1	35.2	35.4	35.1	34.3	35.6
	3 copies/µL	35.0	37.0	36.1	35.6	36.6	36.8	35.4	36.1	35.8
	1 copy/µL	37.0	37.0	35.7	37.2	36.7	37.0	36.4	36.4	36.3
B.1.1.7	100 copies/µL	29.1	29.2	29.2	ND	ND	ND	ND	ND	ND
	50 copies/µL	30.1	30.2	30.2	ND	ND	ND	ND	ND	ND
	25 copies/µL	31.1	31.2	31.2	ND	ND	ND	ND	ND	ND
	12 copies/µL	32.5	32.2	32.1	ND	ND	ND	ND	ND	ND
	6 copies/µL	33.0	33.2	33.1	ND	ND	ND	ND	ND	ND
	3 copies/µL	33.7	34.6	34.0	ND	ND	ND	ND	ND	ND
	1 copy/µL	35.0	35.2	35.3	ND	ND	ND	ND	ND	ND

Wuhan-Hu-1 = Twist synthetic RNA control 2; B.1.1.7 = Twist synthetic RNA control 14; N1-FAM = CDC N1 primer-probe set targeting the nucleocapsid with FAM fluorophore[17]; ORF1a-Cy5 = Yale primer-probe set targeting the ORF1a gene 3675-3677 deletion with Cy5 fluorophore[8]; Spike-HEX = Yale primer-probe set targeting the spike gene 69-70 deletion with HEX fluorophore[8].

Next, we validated our multiplex RT-qPCR variant screening assay using known COVID-19 clinical samples that we have previously sequenced (**Table 2**). We tested 19 samples from SARS-CoV-2 lineages without either ORF1a Δ3675-3677 and spike Δ69-70 (classified as "other" lineage, expected outcome = detection with all three primer/probe sets), 41 samples from lineages B.1.375, B.1.2, and B.1.1.50 (none are current variants of concern) that only have spike Δ69-70 (expected outcome = target failure with the spike set), and 16 samples from B.1.1.7 that have both ORF1a Δ3675-3677 and spike Δ69-70 (expected outcome = target failure with both the ORF1a and spike sets). We found that the expected outcomes were in 100% agreement with the sequence classification. Importantly, unlike the TaqPath assay SGTF results, we could differentiate between B.1.1.7 and other variants that only have the spike deletion, such as B.1.375 that is not currently a variant of concern. Thus, our clinical results demonstrate how our multiplex RT-qPCR assay can detect potential SARS-CoV-2 variants of concern and can be used to prioritize samples for sequencing.

Table 2: Validation of the multiplexed RT-qPCR assay with clinical samples. Listed are cycle threshold (Ct) values for the three primer-probe sets targeting the SARS-CoV-2 nucleocapsid (N1-FAM), ORF1a 3675-3766 deletion (ORF1a-Cy5), and spike 69-70 deletion (Spike-HEX).

Yale ID	GISAID ID	Lineage	N1-FAM	Orf1a-Cy5	Spike-HEX	PCR classification
Yale-521	EPI_ISL_729768	Other	18.1	17.4	19.4	Correct
Yale-522	EPI_ISL_729767	Other	23.2	22.2	23.6	Correct

medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Yale-523	EPI_ISL_729764	Other	20.3	19.1	20.7	Correct
Yale-525	EPI_ISL_729742	Other	22.3	21.2	22.6	Correct
Yale-526	EPI_ISL_729751	Other	22.7	21.6	25.5	Correct
Yale-528	EPI_ISL_729777	Other	23.5	22.0	23.8	Correct
Yale-568	EPI_ISL_730053	Other	24.3	23.1	24.8	Correct
Yale-570	EPI_ISL_730054	Other	18.4	17.2	18.5	Correct
Yale-571	EPI_ISL_730055	Other	24.3	24.2	25.7	Correct
Yale-572	EPI_ISL_730056	Other	24.2	24.9	26.6	Correct
Yale-573	EPI_ISL_730057	Other	23.4	24.2	26.0	Correct
Yale-574	EPI_ISL_730058	Other	22.6	21.3	22.6	Correct
Yale-575	EPI_ISL_730059	Other	26.3	25.3	26.6	Correct
Yale-576	EPI_ISL_730060	Other	29.7	28.2	29.7	Correct
Yale-577	EPI_ISL_730061	Other	27.7	27.1	28.6	Correct
Yale-642	EPI_ISL_802502	Other	25.7	23.9	25.9	Correct
Yale-643	EPI_ISL_802503	Other	22.4	21.0	22.3	Correct
Yale-644	EPI_ISL_802504	Other	33.5	32.3	33.7	Correct
Yale-645	EPI_ISL_802505	Other	34.7	32.7	34.6	Correct
Yale-S001	EPI_ISL_779137	B.1.375	30.4	29.2	ND	Correct
Yale-S002	EPI_ISL_779138	B.1.375	27.2	26.1	ND	Correct
Yale-S003	EPI_ISL_779139	B.1.375	29.3	28.2	ND	Correct
Yale-S004	EPI_ISL_779140	B.1.375	27.2	26.3	ND	Correct
Yale-S005	EPI_ISL_779141	B.1.375	23.1	22.7	ND	Correct
Yale-S006	EPI_ISL_779142	B.1.375	24.1	22.7	ND	Correct
Yale-S007	EPI_ISL_779143	B.1.375	28.4	26.1	ND	Correct
Yale-S008	EPI_ISL_779144	B.1.375	34.7	33.6	ND	Correct
Yale-S009	EPI_ISL_779145	B.1.375	28.9	27.4	ND	Correct
Yale-S010	EPI_ISL_779146	B.1.375	28.1	27.4	ND	Correct
Yale-S011	EPI_ISL_779147	B.1.375	35.9	33.9	ND	Correct
Yale-S012	EPI_ISL_779148	B.1.375	24.9	24.3	ND	Correct
Yale-S014	EPI_ISL_779149	B.1.375	26.7	25.3	ND	Correct
Yale-S015	EPI_ISL_779150	B.1.375	35.0	33.2	ND	Correct
Yale-S016	EPI_ISL_779151	B.1.375	20.2	18.5	ND	Correct
Yale-S017	EPI_ISL_779152	B.1.375	21.0	19.5	ND	Correct
Yale-S021	EPI_ISL_847830	B.1.375	25.4	23.7	ND	Correct
Yale-S022	EPI_ISL_847839	B.1.375	31.4	29.9	ND	Correct
Yale-S023	EPI_ISL_847836	B.1.375	31.4	30.1	ND	Correct
Yale-S025	EPI_ISL_847829	B.1.375	22.1	20.3	ND	Correct
Yale-S049	EPI_ISL_847835	B.1.375	23.9	22.5	ND	Correct
Yale-S100	EPI_ISL_861752	B.1.375	31.8	29.7	ND	Correct
Yale-S101	EPI_ISL_861750	B.1.375	24.9	23.5	ND	Correct

medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Yale-S102	EPI_ISL_861753	B.1.375	31.2	29.5	ND	Correct
Yale-S103	EPI_ISL_861771	B.1.375	23.5	22.1	ND	Correct
Yale-S111	EPI_ISL_861747	B.1.375	18.1	17.0	ND	Correct
Yale-S118	EPI_ISL_861733	B.1.375	28.2	26.9	ND	Correct
Yale-S119	EPI_ISL_861755	B.1.375	18.1	17.1	ND	Correct
Yale-S135	EPI_ISL_861749	B.1.375	17.3	16.6	ND	Correct
Yale-S137	EPI_ISL_861735	B.1.375	15.3	14.5	ND	Correct
Yale-S139	EPI_ISL_861738	B.1.375	20.0	19.6	ND	Correct
Yale-S144	EPI_ISL_861760	B.1.375	12.9	12.3	ND	Correct
Yale-S145	EPI_ISL_861746	B.1.375	18.5	18.2	ND	Correct
Yale-S148	EPI_ISL_861756	B.1.375	18.5	18.1	ND	Correct
Yale-S149	EPI_ISL_861751	B.1.375	24.8	23.1	ND	Correct
Yale-S150	EPI_ISL_861772	B.1.375	20.8	20.0	ND	Correct
Yale-S020	EPI_ISL_847840	B.1.2	26.1	23.2	ND	Correct
Yale-S109	EPI_ISL_861744	B.1.2	32.2	30.3	ND	Correct
Yale-S110	EPI_ISL_861742	B.1.2	17.0	15.8	ND	Correct
Yale-S123	EPI_ISL_861765	B.1.2	23.2	21.0	ND	Correct
Yale-S122	EPI_ISL_861763	B.1.1.50	31.1	29.0	ND	Correct
Yale-S018	EPI_ISL_779153	B.1.1.7	25.1	ND	ND	Correct
Yale-S019	EPI_ISL_779154	B.1.1.7	26.4	ND	ND	Correct
Yale-S112	EPI_ISL_861739	B.1.1.7	23.4	ND	ND	Correct
Yale-S114	EPI_ISL_861743	B.1.1.7	19.3	ND	ND	Correct
Yale-S115	EPI_ISL_861762	B.1.1.7	21.5	ND	ND	Correct
Yale-S116	EPI_ISL_861740	B.1.1.7	21.7	ND	ND	Correct
Yale-S120	EPI_ISL_861748	B.1.1.7	17.2	ND	ND	Correct
Yale-S121	EPI_ISL_861766	B.1.1.7	21.0	ND	ND	Correct
Yale-S132	EPI_ISL_861732	B.1.1.7	23.9	ND	ND	Correct
Yale-S133	EPI_ISL_861761	B.1.1.7	14.3	ND	ND	Correct
Yale-S140	EPI_ISL_861769	B.1.1.7	17.3	ND	ND	Correct
Yale-S141	EPI_ISL_861741	B.1.1.7	19.3	ND	ND	Correct
Yale-S142	EPI_ISL_861757	B.1.1.7	20.6	ND	ND	Correct
Yale-S143	EPI_ISL_861759	B.1.1.7	22.0	ND	ND	Correct
Yale-S146	EPI_ISL_861737	B.1.1.7	24.1	ND	ND	Correct
Yale-S147	EPI_ISL_861754	B.1.1.7	25.7	ND	ND	Correct

N1-FAM = CDC N1 primer-probe set targeting the nucleocapsid with FAM fluorophore[17]; ORF1a-Cy5 = Yale primer-probe set targeting the ORF1a gene 3675-3677 deletion with Cy5 fluorophore[8]; Spike-HEX = Yale primer-probe set targeting the spike gene 69-70 deletion with HEX fluorophore[8]. Lineages were assigned using Pangolin[18]. "Other" is used for all samples that fall outside of the B.1.1.7 and B.1.375 lineages and did not produce the "target failure" profile.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

There are some limitations to our study as presented here. First, we have observed autofluorescence of the N1 primer-probe set when testing negative template controls (average Ct = 39.4, with outliers of Ct 33.4). This could potentially lead to a false B.1.1.7 drop-out profile, and therefore we are continuing to optimize our RT-qPCR conditions. Importantly, this PCR assay should only be used to screen known SARS-CoV-2 positive clinical samples for the presence of key deletions found in variants of concern (where autofluorescence will not be a factor), and it should not be used as a primary clinical diagnostic. We also suggest using a N1 threshold Ct of 35 for calling target failures in the ORF1a and spike sets and performing whole genome sequencing to confirm the identity of variants.

Second, although our assay is suitable for detection of ORF1a Δ3675-3677 found in the B.1.351 and P.1 variants, we have not been able to empirically test clinical samples with these variants due to access limitations. We are actively seeking additional clinical samples and laboratory partners to expand our clinical validation.

Third, our assay will not be able to detect all B.1.351 viruses. There is a monophyletic clade within the B.1.351 lineage that has ORF1a Δ3675-3677 filled back in, perhaps due to recombination with viruses that did not have the deletion. How often this is expected to occur is not known, but it demonstrates that continuous monitoring for the presence of ORF1a Δ3675-3677 and spike Δ69-70 within the variants of concern will be necessary to ensure that our assay will still be effective.

The rapid emergence of the SARS-CoV-2 variants of concern necessitates an immediate roll out of surveillance tools. Although whole genome sequencing is required to definitively identify specific variants, resource and capacity constraints can limit the number of samples that can be sequenced. The ThermoFisher TaqPath assay has demonstrated the value of PCR for variant surveillance, but it is limited to B.1.1.7 and cannot differentiate between other viruses containing spike Δ69-70. By targeting two different large nucleotide deletions, ORF1a Δ3675-3677 and spike Δ69-70, our multiplex PCR can rapidly screen for B.1.1.7, B.1.351, and P.1 variants and differentiate between non-variants of concern. Thus, our multiplex RT-qPCR variant screening assay can be used to prioritize samples for sequencing and as a surveillance tool to help monitor the distribution and population frequency of suspected variants.

Methods

Ethics

The Institutional Review Board from the Yale University Human Research Protection Program determined that the RT-qPCR testing and sequencing of de-identified remnant COVID-19 clinical samples conducted in this study is not research involving human subjects (IRB Protocol ID: 2000028599). The “Yale ID” numbers displayed in Table 2 are not known outside the research group and cannot be used to re-identify any subject.

Analysis of public SARS-CoV-2 genomes

All available SARS-CoV-2 data (402,899 genomes) were downloaded on 2021-01-22 from GISAID and evaluated for the presence of ORF1a Δ3675-3677 and spike Δ69-70. Phylogenetic analysis of a subset

medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

of 4,046 SARS-CoV-2 genomes was performed using Nextstrain[15], downsampled as shown using the "global build" on 2021-01-22 (https://nextstrain.org/ncov/global). A list of SARS-CoV-2 genomes used in the analysis is available in **Source Data Fig. 1**.

Multiplex RT-qPCR with probes

A detailed protocol of our multiplexed RT-qPCR to screen for SARS-COV-2 B.1.1.7, B.1.351, and P.1 variants of concern can be found on protocols.io[8]. In brief, our multiplex RT-qPCR assay consists of the CDC N1[17], and the newly designed Yale ORF1a Δ3675-3677 and Yale spike Δ69-70 primer-probe sets (**Supplementary Table 2**). We used the NEB Luna universal probe one step RT-qPCR kit with 400 nM of primers, 200 nM of probes, and 5 µL of nucleic acid in a total reaction volume of 20 µL. Thermocycler conditions were reverse transcription for 10 minutes at 55°C, initial denaturation for 1 minute at 95°C, followed by 40 cycles of 10 seconds at 95°C and 30 seconds at 55°C. During validation we ran the PCR for 45 cycles. Differentiation between variants of concern is based on drop-out of the Yale ORF1a and/or Yale spike primer-probe sets (**Supplementary Table 3**).

Limit of detection

We used Twist synthetic SARS-CoV-2 RNA controls 2 (Genbank ID: MN908947.3; GISAID ID: Wuhan-Hu-1) and control 14 (Genbank ID: EPI_ISL_710528; GISAID ID: England/205041766/2020) to determine the limit of detection of the screening RT-qPCR assay. We tested a two-fold dilution series from 100 copies/µL to 1 copy/µL for both RNA controls in triplicate, and confirmed the lowest concentration that was detected in all three replicates by 20 additional replicates.

Validation and sequence confirmation

We validated our approach using known SARS-CoV-2 positive clinical samples. Briefly, we extracted nucleic acid from 300 µL viral transport medium from nasopharyngeal swabs and eluted in 75 µL using the MagMAX viral/pathogen nucleic acid isolation kit (ThermoFisher Scientific). Extracted nucleic acid was tested by our multiplexed RT-qPCR assay and then sequenced using a slightly modified ARTIC Network nCoV-2019 sequencing protocol for the Oxford Nanopore MinION[19,20]. These modifications include extending incubation periods of ligation reactions and including a bead-based clean-up step following dA-tailing. MinION sequencing runs were monitored using RAMPART[21]. Consensus sequences were generated using the ARTIC Network bioinformatics pipeline and lineages were assigned using Pangolin v.2.0[18,22]. GISAID accession numbers for all SARS-CoV-2 genomes used to validate our approach are listed in **Table 2.**

Data availability

Genomic data are available on GISAID (see **Table 2** for accession numbers). All RT-qPCR data are included in this article, supplementary files, and source data.

References

1. Davies, N. G. *et al.* Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. *bioRxiv* (2020) doi:10.1101/2020.12.24.20248822.
2. Rambaut, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in

medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

the UK defined by a novel set of spike mutations. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (2020).

3. Lauring, A. S. & Hodcroft, E. B. Genetic variants of SARS-CoV-2 - What do they mean? *JAMA* (2021) doi:10.1001/jama.2020.27124.
4. ThermoFisher Scientific. Solutions for surveillance of the S gene mutation in the B.1.1.7 (501Y.V1) SARS-CoV-2 strain lineage. https://www.thermofisher.com/blog/behindthebench/solutions-for-surveillance-of-the-s-gene-mutation-in-the-b117-501yv1-sars-cov-2-strain-lineage/ (2020).
5. Kidd, M. *et al.* S-variant SARS-CoV-2 is associated with significantly higher viral loads in samples tested by ThermoFisher TaqPath RT-QPCR. *bioRxiv* (2020) doi:10.1101/2020.12.24.20248834.
6. Tegally, H. *et al.* Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *bioRxiv* (2020) doi:10.1101/2020.12.21.20248640.
7. Faria, N. R. *et al.* Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586 (2021).
8. Vogels, C., Fauver, J. & Grubaugh, N. Multiplexed RT-qPCR to screen for SARS-COV-2 B.1.1.7, B.1.351, and P.1 variants of concern v2. *protocols.io* (2021) doi:10.17504/protocols.io.brrhm536.
9. Borges, V. *et al.* Tracking SARS-CoV-2 VOC 202012/01 (lineage B.1.1.7) dissemination in Portugal: insights from nationwide RT-PCR Spike gene drop out data. https://virological.org/t/tracking-sars-cov-2-voc-202012-01-lineage-b-1-1-7-dissemination-in-portugal-insights-from-nationwide-rt-pcr-spike-gene-drop-out-data/600 (2021).
10. Volz, E. *et al.* Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *bioRxiv* (2021) doi:10.1101/2020.12.30.20249034.
11. Galloway, S. E. *et al.* Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020-January 12, 2021. *MMWR Morb. Mortal. Wkly. Rep.* **70**, 95–99 (2021).
12. Moreno, G. *et al.* Detection of non-B.1.1.7 spike Δ69/70 sequences (B.1.375) in the United States. https://virological.org/t/detection-of-non-b-1-1-7-spike-69-70-sequences-b-1-375-in-the-united-states/587/2 (2021).
13. Larsen, B. B. & Worobey, M. Identification of a novel SARS-CoV-2 Spike 69-70 deletion lineage circulating in the United States. https://virological.org/t/identification-of-a-novel-sars-cov-2-spike-69-70-deletion-lineage-circulating-in-the-united-states/577 (2020).
14. Greaney, A. J. *et al.* Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies. *Cold Spring Harbor Laboratory* 2020.12.31.425021 (2021) doi:10.1101/2020.12.31.425021.
15. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
16. Kudo, E. *et al.* Detection of SARS-CoV-2 RNA by multiplex RT-qPCR. *bioRxiv* 2020.06.16.155887 (2020) doi:10.1101/2020.06.16.155887.
17. Lu, X. *et al.* US CDC real-time reverse transcription PCR panel for detection of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **26**, (2020).
18. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403–1407 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

19. Quick, J. nCoV-2019 sequencing protocol v3. (2020) doi:10.17504/protocols.io.bbmuik6w.
20. Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).
21. Artic Network. https://artic.network/ncov-2019/ncov2019-using-rampart.html (2020).
22. Artic Network. https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html (2020).

Acknowledgements

We thank A. Brito, A. Altajar, and D. Comstock for data or clinical support. A list of acknowledgements for the SARS-CoV-2 data used in Fig. 1 can be found in the **Source Data Fig. 1**. This work was funded by CTSA Grant Number TL1 TR001864 (TA and MEP), Fast Grant from Emergent Ventures at the Mercatus Center at George Mason University (NDG), and CDC Contract # 75D30120C09570 (NDG).

Author information

Affiliations

Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT 06510, USA
Chantal B.F. Vogels, Mallery Breban, Tara Alpert, Mary E. Petrone, Anne E. Watkins, Joseph R. Fauver, Nathan D. Grubaugh

Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06510, USA
Nathan D. Grubaugh

Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
Emma B. Hodcroft

Tempus Labs, Chicago, IL 60654, USA
Chris Mason, Gaurav Khullar, Joel T. Dudley, Matthew MacKay

Department of Pathology, Yale University School of Medicine, New Haven, CT 06510, USA
Jianhui Wang, Chen Liu, Pei Hui

Murphy Medical Associates, Greenwich, CT 06614, USA
Steven Murphy, Caleb Neal, Eva Laszlo

Departments of Laboratory Medicine and Medicine, Yale School of Medicine, New Haven, CT 06510, USA
Marie L. Landry

Connecticut State Department of Public Health, Rocky Hill, CT 06067, USA
Anthony Muyombwe, Randy Downing, Jafar Razeq

Biozentrum, University of Basel, 4056 Basel, Switzerland
Richard A. Neher

medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
Richard A. Neher

Contributions

CBFV, RAN, JRF, and NDG designed the study; CEM, GK, JD, MM, JW, CL, PH, SM, CN, EL, MLL, AM, RD, and JR collected and provided clinical samples; CBFV, MB, TA, MEP, AEW, EBH, RAN, JRF, and NDG collected and analyzed data; JRF and NDG supervised the project; CBFV and NDG wrote and edited the manuscript; all authors read and approved the final manuscript.

Corresponding authors

Correspondence to Chantal B.F. Vogels, Joseph R. Fauver, and Nathan D. Grubaugh.

Ethics declarations

Competing interests

The authors declare no competing interests.

Source Data

Source Data Fig. 1: Supporting data for Fig. 1.

medRxiv preprint doi: https://doi.org/10.1101/2021.01.28.21250486; this version posted February 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary information

Supplementary Table 1: Summary of SARS-CoV-2 genomes with the ORF1a 3675-3677 and/or Spike 69-70 deletions (GISAID on 2020-01-21).

Clade	Total sequences	ORF1a: Δ3675-3677	Spike: Δ69-70	% ORF1a: Δ3675-3677	% Spike: Δ69-70
19A	16,034	5	9	0.0%	0.1%
19B	8,869	1	4	0.0%	0.0%
20A	81,139	22	6,494	0.0%	8.0%
20A.EU2	10,380	5	13	0.0%	0.1%
20B	95,315	21	1,366	0.0%	1.4%
20C	45,041	43	274	0.1%	0.6%
20D	4,804	0	4	0.0%	0.1%
20E.EU1	91,540	2	7	0.0%	0.0%
20F	12,613	4	0	0.0%	0.0%
20G	11,276	0	22	0.0%	0.2%
B.1.351	582	474	0	81.4%	0.0%
B.1.1.7	25,262	25,178	25,142	99.7%	99.5%
P.1	34	34	0	100.0%	0.0%

Supplementary Table 2: Primers and probes used in the multiplexed RT-qPCR variant screening assay.

Set name	Nt positions	TM	Primer/probe	Sequence
CDC N1	28,287-28,306	53.6	Fwd primer	GACCCCAAAATCAGCGAAAT
	28,335-28,358	57.7	Rev primer	TCTGGTTACTGCCAGTTGAATCTG
	28,309-28.332	63.3	Probe	**FAM**-ACCCCGCATTACGTTTGGTGGACC-**BHQ1**
Yale ORF1a Δ3675-3677	11,229-11,248	60	Fwd primer	TGCCTGCTAGTTGGGTGATG
	11,332-11,356	57.8	Rev primer	TGCTGTCATAAGGATTAGTAACACT
	11,283-11,312	61.9	Probe	**Cy5**-GTTTGTCTGGTTTTAAGCTAAAAGACTGTG-**BHQ2**
Yale Spike Δ69-70	21,710-21,733	59.3	Fwd primer	TCAACTCAGGACTTGTTCTTACCT
	21,796-21,817	57.4	Rev primer	TGGTAGGACAGGGTTATCAAAC
	21,755-21,779	61.2	Probe	**HEX**-TTCCATGCTATACATGTCTCTGGGA-**BHQ1**

Supplementary Table 3: Interpretation of results from the multiplexed RT-qPCR variant screening assay.

Result	N1-FAM	Orf1a-Cy5	Spike-HEX
Potentially B.1.1.7	CT ≤ 35	Undetected	Undetected
Potentially B.1.351 or P.1	CT ≤ 35	Undetected	CT ≤ 35
Potentially B.1.375	CT ≤ 35	CT ≤ 35	Undetected
Other lineages	CT ≤ 35	CT ≤ 35	CT ≤ 35
Inconclusive	CT > 35 or undetected	Any value	Any value