

S-variant SARS-CoV-2 lineage B1.1.7 is associated with significantly higher viral loads in samples tested by ThermoFisher TaqPath RT-qPCR.

Michael Kidd^{1*}, Alex Richter², Angus Best³, Nicola Cumley³, Jeremy Mirza³, Benita Percival³, Megan Mayhew³, Oliver Megram³, Fiona Ashford³, Thomas White³, Emma Moles-Garcia³, Liam Crawford³, Andrew Bosworth⁴, Sowsan F Atabani¹, Tim Plant³, Alan McNally⁵

1. Public Health England and University Hospitals Birmingham NHS Foundation Trust UK.
2. Institute of Cancer and Genomic Science, College of Medical and Dental Science, University of Birmingham, Birmingham, UK.
3. Clinical Immunology Service, Institute of Immunology and Immunotherapy, College of Medical and Dental Science, University of Birmingham, Birmingham, UK.
4. University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK.
5. Institute of Microbiology and Infection, College of Medical and Dental Science, University of Birmingham, Birmingham, UK.

* Corresponding author: Michael Kidd, Public Health England, PHE Laboratory Birmingham, Birmingham Heartlands Hospital, Bordesley Green East, West Midlands, B9 5SS, UK

Email: Michael.Kidd@phe.gov.uk

Telephone: +44 (0) 7515 880 947

Abstract Text Article 40-word summary: This study looked for evidence of the B1.1.7 variant SARS-CoV-2 in routine testing results at a UK diagnostic centre. We demonstrated its presence and showed that a significantly greater proportion of subjects with the variant had very high viral loads.

Footnote Page:

None of the authors have a commercial or other association that might pose a conflict of interest.

The diagnostic work that led to the generation of these results was funded by the UK Department of Health and Social Care.

None of this data has been presented at any meeting prior to submission.

Corresponding author information:

Dr Michael Kidd, Public Health England, PHE Laboratory Birmingham, Birmingham Heartlands Hospital, Bordesley Green East, West Midlands, B9 5SS

Email: Michael.Kidd@phe.gov.uk

Telephone: +44 (0) 7515 880 947

Fax: +44 (0) 121 772 6229

Abstract:

A SARS-CoV-2 variant B.1.1.7 containing a mutation $\Delta 69/70$ has spread rapidly in the UK and shows an identifiable profile in ThermoFisher TaqPath RTqPCR (S-gene target failure; SGTF). We analysed recent test data for trends and significance. Linked Ct values for respiratory samples showed that a low Ct for ORF1ab and N were clearly associated with SGTF. Significantly more SGTF samples had higher inferred viral loads between 1×10^7 and 1×10^8 . Our conclusion is that patients whose samples exhibit the SGTF profile are more likely to have high viral loads, which may explain higher infectivity and rapidity of spread.

Keywords: SARS-CoV-2; COVID-19; RT-PCR; TaqPath; viral load; Ct values; $\Delta 69/70$; S-variant; B.1.1.7; VUI-202012/01; SGTF; S-gene dropout.

Background:

The UK response to the SARS-CoV-2 pandemic has involved the setup of high-throughput diagnostic centres [1] operationally standardised using commercial reverse-transcriptase qualitative polymerase chain reaction (RT-qPCR) testing [2]. The ThermoFisher 'TaqPath' test co-amplifies three SARS-CoV-2 viral gene targets from a single clinical sample (ORF1ab, N, and S). Test accuracy is verified between centres using the Qnostics external quality assurance (EQA) panel [3], subsequent quality assurance by UK-NEQAS [4], and has National Health Service (NHS-E/I) accreditation for the routine diagnostic service.

In December 2020, UK authorities were alerted to the emergence of a SARS-CoV-2 variant of genetic lineage B.1.1.7 [5], synonymous with the 'variant under investigation' (VUI-202012/01) by Public Health England, which appeared to be spreading rapidly through the UK [6]. The variant possesses several mutations, of which the $\Delta 69/70$ deletion results in failure to detect the S-gene target (SGTF) of the TaqPath test, with the other two gene targets ORF1ab and N not affected. We undertook to review past positive results for evidence that the $\Delta 69/70$ variant had been present in these respiratory samples, that overall test accuracy had still been maintained, and investigated its relationship with the other viral gene targets.

Methods:

The data used for this analysis are RT-qPCR threshold-crossing (Ct) values originating from laboratory testing of respiratory samples in the UK Department of Health and Social Care 'Test and Trace' network. A dataset of 641 SARS-CoV-2 positive results received during the period 25th October to 25th November 2020 was used for this analysis (available in [7]).

All positive results had amplifiable MS-2 internal control, with no evidence of general inhibition in the RT-PCR reaction. Raw Ct values were analysed with respect to the presence of ORF1ab, N, or S-gene single-target signals. For the purposes of this analysis, gene-target 'negative' signals are separated into either (a) target signal detected but above the threshold of Ct 37, as stipulated for clinical interpretation of results in the TaqPath IFU, and (b) targets having no signal detected, which were assigned a nominal Ct value of '45'. This latter group of S-gene target failures are defined as SGTF.

Frequency comparisons, Chi-squared, and Mann-Whitney 'U' tests for significance of non-Gaussian distributions between SGTF and S-gene positive RT-qPCR results were performed using GraphPad Prism version 5.03.

The process of inferring relative viral loads was based on the laboratory performance for the Qnostics EQA panel results (see figure 1A), which has quantitative information relating copies per mL of whole virus lysate derived from cell culture of SARS-CoV-2. Relative viral loads were inferred by standard efficiency calculations of RT-PCR tests, where 3.3 Ct difference between targets approximates to a 10-fold change in substrate.

Results:

Frequency analysis for all three gene targets

Figure 1B shows a significantly higher proportion of SGTF samples (178 of 641; 27.7%), arrowed, far right at Ct 45, compared with either ORF or N-gene undetectable positive profiles (both 13 of 641; 2.0%; $\chi^2 (1, N = 641) = 165.46, p < .00001$). Further non-statistical observation suggests that for lower Cts (9 – 22), in each bar the proportion of S-gene appears at a lower frequency than the corresponding ORF1ab and N genes. At Cts from 25 -33 and onwards, this trend is less obvious with similar frequencies of all three viral genes detected. In particular, towards the limit of test sensitivity range (Ct 35 – 40), ORF1ab, N, and S are detected at approximately equal frequencies with no apparent consistent loss of sensitivity for any gene target.

SGTF samples had significantly lower median Ct of corresponding ORF1ab- and N-gene targets.

To determine whether, on a population basis, SGTF samples were significantly associated with lower Ct values of either ORF1ab or N gene target in positive samples, the distribution of SGTF and S-detected samples was compared within all ORF1ab and N positives (Figure 2A). In both, the median Ct value of SGTF and S-detected was significantly different (ORF1ab gene 18.16 vs. 22.30; N gene 19.39 vs. 23.16; both $p < .0001$ (Mann-Whitney 'U')). Clustering of SGTF results around very low Ct values of ORF1ab and N can be clearly observed, probably accounting for the lowering of the group median Ct.

Linked Ct values for individual samples show a propensity for SGTF samples to be associated with a lower Ct for ORF1ab and N gene targets.

Since it is possible that a low Ct of ORF1ab or N-gene targets in SGTF could occur in different samples, we further investigated the likelihood for the SGTF profile to be associated with low Ct values of both other viral targets in the same sample (figure 2B). Such SGTF profiles appeared more

likely to be associated with low Ct, and a Chi-squared analysis of the number of samples below a Ct value of 15 in both categories of SGTF (63/178; 35.4%) and non-SGTF (46/450; 10.2%) was performed for the ORF-positive group. Results showed a highly significant difference between proportions ($\chi^2 (1, N = 628) = 36.61, p < .0001$). Therefore, our observed cluster of SGTF samples corresponds to a significantly larger population of infectious subjects having an increased viral load; which can be up to 10,000-fold higher than the non-SGTF median (Ct 9 vs. Ct 23). Conservative extrapolation from the upper end of the Qnostics EQA panel data suggests that the larger population has viral loads of between 1×10^7 (\approx Ct 12) and 1×10^8 (\approx Ct 9) copies per mL.

Discussion

The SARS-CoV-2 variant of concern (VOC-202012/01; Public Health England) has spread rapidly throughout the south-east of the UK, and latterly to other regions [6]. More detailed published evidence shows that the presence of the $\Delta 69/70$ mutation in the viral genome, causing the SGTF phenomenon in TaqPath RT-qPCR tests [8], strongly correlates [9] with presence of the VOC / B1.1.7 in clinical samples as determined by sequencing [10], and is now used as an epidemiological proxy for presence of the variant.

We analysed positive results from samples submitted to the Birmingham Turnkey laboratory between October 25th and November 25th 2020, from both local and distant UK geographical areas at a time where the incidence of VOC / B1.1.7 was increasing sharply [6]. Our analysis shows a high proportion of SGTF, with other gene targets ORF1ab and N being clearly detected in the same sample. Further analysis shows that a significantly higher number of SGTF samples are associated with these lower Ct values of ORF1ab and N; from which it is possible to infer conservatively a high viral load in this larger population of between 1×10^7 and 1×10^8 copies per mL. Although in our comparisons, similar viral loads were seen in non-SGTF samples, the significantly higher number of subjects having an extreme viral load is of great concern.

The capability of increased transmission has been ascribed to the VOC: epidemiological tracking, as either its SGTF proxy or sequenced as B1.1.7, has shown that its secondary attack rate is higher than wild-type virus [10] and that it has a significant 50-75% multiplicative increase in reproductive number compared to non-SGTF variants [11]. Our finding in laboratory data that a significantly larger proportion of subjects whose samples show SGTF have an inferred viral load at the extreme end of the range, may represent an explanation for this. We recommend that further investigations should include the possibility that either short-term very high viral load, or an extension of period of infectious viral excretion during the symptomatic phase of illness [12], would

both increase the likelihood of the virus to transmit onwards and still be compatible with our findings

At a technical level, we considered alternative explanations for SGTF at very low ORF1ab and N Ct values. These include the argument that chemical components in an individual RT-qPCR reaction become limited when amplifying multiple targets at high viral load input, and possibly the S-gene target is first to become non-amplifiable. However, the TaqPath test contains an internal control provided by co-amplification of non-human bacteriophage MS-2, the target RNA of which is included at a concentration that is more likely to become undetectable under adverse reaction conditions than any of the three specific gene targets; and provides reassurance that amplification of all specific targets is not being inhibited. All positive TaqPath results in the data table were passed as valid, determined by the presence of MS-2 amplification, making it unlikely that SGTF are due to a general reaction chemistry bias.

During the laboratory verification of the TaqPath test, using an EQA standard dilution series derived from cultured SARS-CoV-2, there were no observed SGTF at Ct levels of approximately 15 (see supplement); which is lower than the median Ct value at which SGTF were seen to occur in our data. This lends further weight to discount the influence of general RT-qPCR reaction inefficiencies in our observations.

Another alternative, but theoretical, explanation for the low Ct of ORF1ab and N with SGTF, is that inhibition of S-gene amplification results in greater availability of in vitro reagents and less competition for enzymatic activity, which enables preferential increases in ORF1ab and N. However, we believe this is unlikely, since the deletion mutation interfering with S-gene detection only affects the probe-binding, with silent yet authentic co-amplification of the S-gene target [6]. We have confirmed this to be the case by colleagues sequencing our SGTF samples (personal communication to M.K. from A.M.), in which the S-gene amplicons are clearly present and have the $\Delta 69/70$ deletion.

It should be emphasised that the authenticity of positive results is not affected by the presence of the SGTF phenomenon, as the TaqPath test result is classed as 'positive' when two gene targets are detected. Thus, the ability of the TaqPath test to detect three viral targets provides a degree of robustness to the 'Test and Trace' programme, even when a viral mutation renders one of them undetectable.

As a parallel observation, some types of commercial test for SARS-CoV-2 rely on reactivity to the S-gene or its protein product, for example the lateral flow devices. A recent preliminary assessment of the performance of five of these tests has shown that they are not affected by the VOC / B.1.1.7 [13].

For clarity, we do not anticipate that the mutation causing SGTF is necessarily responsible for higher viral load in patients. There are additional mutations in B.1.1.7 which could have a direct contribution, with SGTF as an indirect marker in the TaqPath RT-PCR for the presence of the B.1.1.7 variant. Whole genome sequencing of individual samples will prove to be valuable in strengthening the association with changes in the viral genome.

Finally, we also observed dropouts for the ORF1ab and N genes (see Figure 1B) – albeit at a much lower frequency and not apparently associated with high viral loads – and we believe these should be similarly investigated for mutations in the corresponding genes that could have affected their detection. We also note that double-dropouts – where two viral genes are not amplified in a sample – by implication will not be represented in the original data as they would be classed as negative. A more exhaustive analysis would involve reviewing all negative results where a single viral gene was amplified.

Limitations of this data are:

- (a) Our analysis may provide additional evidence to explain why VOC-202012/01 (B.1.1.7) may be transmitting more rapidly amongst populations, but it does not provide an explanation of how an increased viral load could occur. If verified by others, the biological plausibility of its higher infectivity, whether through evolutionary viral replication advantages or evasion of the host immune system, is yet to be determined.
- (b) Although we have made broad inferences in relative viral load in the samples, the TaqPath is not designed as a quantitative assay for SARS-CoV-2 and our observations should be repeated by a dilution series or a validated quantitative method.

References

1. Richter A, Plant T, Kidd M, et al. How to establish an academic SARS-CoV-2 testing laboratory. *Nat Microbiol.* **2020**; 5(12):1452–1454.
2. ThermoFisher. TaqPath™ COVID-19 CE-IVD RT-PCR Kit [Internet]. [cited 2020 Dec 28]. Available from: <https://www.thermofisher.com/order/catalog/product/A48067>
3. Molecular Controls – Qnostics [Internet]. [cited 2020 Dec 28]. Available from: <https://www.qnostics.com/our-products/?filter=analytical-q-panels-en>
4. Home - UK NEQAS | External Quality Assessment Services [Internet]. UK NEQAS. [cited 2020 Dec 30]. Available from: <https://ukneqas.org.uk/>
5. COG Consortium UK data reports [Internet]. 2020 Dec p. 9. Report No.: 1. Available from: https://www.cogconsortium.uk/wp-content/uploads/2020/12/Report-1_COG-UK_20-December-2020_SARS-CoV-2-Mutations_final_updated2.pdf
6. Investigation of novel SARS-COV-2 variant: Variant of Concern 202012/01 #1 [Internet]. Public Health England; Report No.: 1. Available from: <https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201>
7. Richter A, Best A, Mirza J, et al. Extrapolation of United Kingdom Pillar 2 Care home Covid-19 test data to ascertain effectiveness of lateral flow testing in low prevalence settings. *medRxiv.* **2020**; :2020.12.01.20237784.
8. Genetic Variants of SARS-CoV-2 May Lead to False Negative Results with Molecular Tests for Detection of SARS-CoV-2 - Letter to Clinical Laboratory Staff and Health Care Providers [Internet]. US Food and Drug Administration; 2021 Aug. Available from: <https://www.fda.gov/medical-devices/letters-health-care-providers/genetic-variants-sars-cov-2-may-lead-false-negative-results-molecular-tests-detection-sars-cov-2>
9. Emerging SARS-CoV-2 Variants [Internet]. 2021 Jan. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/scientific-brief-emerging-variants.html>
10. Investigation of novel SARS-CoV-2 variant: Variant of Concern 202012/01 #3 [Internet]. Public Health England; Report No.: 3. Available from: <https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201>
11. Rambaut A. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data [Internet]. *ARTIC*; 2020 Dec. Available from: <https://virological.org/>
12. Woelfel R, Corman VM, Guggemos W, et al. Clinical presentation and virological assessment of hospitalized cases of coronavirus disease 2019 in a travel-associated transmission cluster [Internet]. *Infectious Diseases (except HIV/AIDS)*; 2020 Mar. Available from: <http://medrxiv.org/lookup/doi/10.1101/2020.03.05.20030502>

13. SARS-CoV-2 lateral flow antigen tests: evaluation of VUI-202012/01 [Internet]. Public Health England; 2020 Dec. Available from: <https://www.gov.uk/government/publications/sars-cov-2-lateral-flow-antigen-tests-evaluation-of-vui-20201201/sars-cov-2-lateral-flow-antigen-tests-evaluation-of-vui-20201201>

Accepted Manuscript

Figure 1 legend

(A) Verification data for TaqPath RT-PCR test using the Qnostics external quality assurance dilution series 'SCV2AQP01-A'. (i) – (iii) linear regression plots for ORF1ab, N, and S-gene targets respectively. (iv) Table of regression data including efficiency calculations for each gene target. (B) Stacked bar frequency diagram showing the proportional relationship between all three viral gene targets detected by RT-qPCR, at each Ct value across the range. Right Y-axis is identical to the left and both are divided into a low sample frequency range of 0-75 and a high range of 130-240.

Figure 2 legend

(A) Scatter plot of the population of S-gene negative (SGTF) / S-positive Cts within corresponding ORF1ab and N-gene positive samples. Median Ct is shown by a black horizontal bar. Above both plots the results of Mann-Whitney U tests for significant differences are shown with conventional notation ('***' for $p < .0001$). (B) Scatter plot of viral gene Ct values for individual positive samples. The horizontal dotted line at Ct 37 represents the threshold for a 'target negative' result as defined for clinical interpretation in the TaqPath instructions for use. The class of 'target undetectable' results have a nominal value of Ct 45 assigned (described in Methods section). Each set of 3 Ct values for a sample is linked by a grey dotted line.

Figure 1

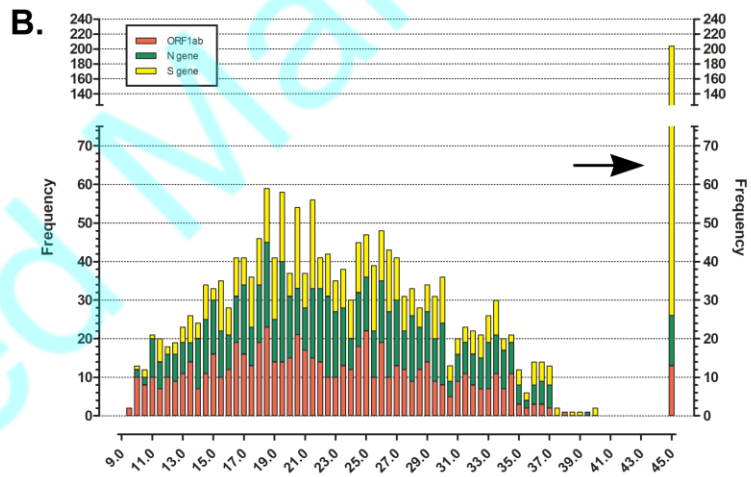
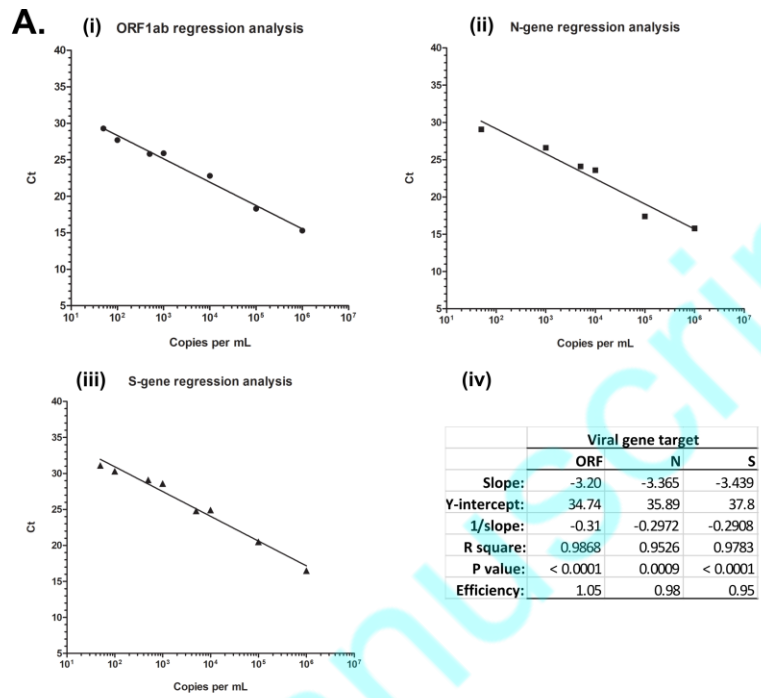


Figure 2

