

1 Multiplex Fragment Analysis Identifies SARS-CoV-2 Variants

2 Clark, A.E.,¹ Wang, Z.,¹ Cantarel, B.,^{1,2,3} Kanchwala, M.,⁴ Xing, C.,⁴ Chen, L.,¹ Irwin, P.,¹ Xu, Y.,¹ Oliver, D.,¹ Lee,
3 F.,¹ Gagan, J.R.,¹ Filkins, L.,^{1,4} Muthukumar, A.,¹ Park, J.Y.,^{1,4,5}, Sarode, R.,¹ and SoRelle, J.A.^{1*}

4
5 ¹ Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas

6 ² Bioinformatics Core Facility, University of Texas Southwestern Medical Center, Dallas, Texas

7 ³ Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, Texas

8 ⁴ Department of Pathology and Laboratory Medicine, Children's Health System of Texas, Dallas, Texas

9 ⁵ McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center,
0 Dallas, Texas

1
2 *Corresponding Author

3 Jeffrey A. SoRelle, MD

4 Department of Pathology

5 University of Texas Southwestern Medical Center

6 5323 Harry Hines Blvd

7 Dallas, TX, 75290

8 Office: 214-645-5155

9 Email: Jeffrey.SoRelle@utsouthwestern.edu
0

1 Word Count: 1,206

2 Figures: 2

3 Supplemental Tables: 5
4

5 Commercial Interest Disclosure: Park is a scientific advisory board member of Miraca Holdings; Baylor

6 Genetics, Fujirebio and SRL Labs are subsidiaries. SoRelle owns stock in Myriad Genetics.
7
8
9
0

1 Abstract

2 The rapid spread of SARS-CoV-2 Variants of Concern (VOC) necessitates systematic efforts for epidemiological
3 surveillance. The current method for identifying variants is viral whole genome sequencing (WGS). Broad
4 clinical adoption of sequencing is limited by costly equipment, bioinformatics support, technical expertise, and
5 time for implementation. Here we describe a scalable, multiplex, non-sequencing-based capillary
6 electrophoresis assay to affordably screen for SARS-CoV-2 VOC.
7
8
9

0 Main

1 Novel strains of SARS-CoV-2 have emerged with clinically significant genetic mutations. These “variants
2 of concern” VOCs impact transmission kinetics,¹ vaccine responses,^{2,3} mortality,^{4,5} and drug resistance
3 (monoclonal antibodies).⁶⁻⁹ Therefore, methods for real-time identification of SARS-CoV-2 VOCs are of exigent
4 need in both clinical and public health arenas. Whole-genome sequencing (WGS) is the current gold standard
5 for SARS-CoV-2 variant identification; however, broad WGS adoption is hampered by cost and requirements
6 for specialized equipment and bioinformatics expertise.¹⁰ While these limitations may be mitigated through
7 centralized high-volume testing, the consequent increased turnaround time for batching, data deconvolution,
8 and sequence analysis exceeds what is clinically actionable for most patient assessments. Therefore, simple,
9 on-site, positive specimen reflex testing for VOCs represents a practical solution.

0 Here, we describe a rapid, cost-effective, and scalable method to detect known VOCs utilizing fragment
1 analysis by capillary electrophoresis in a routine clinical setting. In this procedure, extracted RNA is amplified
2 to create fluorescently-labeled amplicons separated by capillary electrophoresis on a Sanger sequencer. VOCs
3 are then identified by their abnormal amplicon sizes compared to a wild-type reference. By measuring the
4 resultant changes in fragment length, this assay permits a broader range of variant detection compared to
5 primer-probe based RT-qPCR. The reagent and instrumentation requirements for this assay are modest:
6 fluorescently labeled primers, a thermocycler, and Sanger sequencer. Moreover, the time requirement (4
7 hours, Fig. 1B), simplicity, and cost-effective use of 96-well format support the assay’s scalability. Amplicons
8 can be designed up to 500bp in length with up to four fluorescent labels, making the assay adaptable to detect
9 additional SARS-CoV-2 variants as they are identified (Fig. 1C, potential to multiplex to 20-40 targets). This
0 assay therefore allows routine clinical laboratories to rapidly screen SARS-CoV-2 positive specimens for VOCs
1 by a mechanism that can be easily and broadly implemented with limited financial burden or bioinformatic
2 expertise. Real-time monitoring of SARS-CoV-2 VOCs could impact epidemiological tracking and guide clinical
3 management prior to administering monoclonal antibody therapy.

4 For proof of concept, we targeted three deletion mutations initially described in the UK B.1.1.7 VOC. The
5 B.1.1.7 strain was originally discovered due to S-gene target failure (SGTF) secondary to the Spike gene
6 HV69_70del variant in a SARS-CoV-2 commercial assay.^{1,5,11} While present in B.1.1.7, the HV69_70del is not
7 specific to this strain (55% of northeast U.S. HV69_70del mutants are not B.1.1.7);¹² therefore, we targeted the
8 S-gene F144 deletion mutation, to improve specificity for B.1.1.7 to >99%.¹³ Frequent in-frame deletions near
9 the 69/70 and 144 amino acid positions suggest this is a mutational hotspot, which could arise in other emerging
0 strains.¹⁴ An ORF1A deletion (SDF3675_3677del) is present in B.1.1.7 (UK variant),⁴ B.1.351 (South African,
1 originally N150Y.2),¹⁵ P.1 (Brazil),¹⁶ and B.1.526 (New York) variants. Therefore, detecting an ORF1A deletion

without S gene deletions is suggestive of the B.1.351, P.1, or B.1.526 VOCs.¹⁷ (Fig. 1A, 2G). Lastly, the CDC N1 primer pair was added as an internal positive control.

We validated this method focusing on the principle SARS-CoV-2 VOC impacting European countries and the United States: B.1.1.7.¹¹ An initial screen of 182 retrospective (12/1/20- 1/20/21) and 466 prospective SARS-CoV-2 positive nasopharyngeal specimens (1/25/21-3/11/21) was performed. No VOCs were identified among retrospective specimens (0/182). On the third week in January 2021, the first VOC was identified with characteristic shortened amplicons reflecting B.1.1.7 deletions (Fig 1D, E, Interpretive criteria in Supplemental Methods). This deletion pattern was confirmed in control RNA containing the B.1.1.7 sequence. For verification, the suspected B.1.1.7 isolate and positive control sample were sequenced by WGS (Supplemental methods), and classified by PANGOLIN as belonging to the B.1.1.7 lineage (Fig. 1G-I).¹⁸

Overall, we detected 32 specimens containing SARS-CoV-2 consistent with B.1.1.7 (n=23 patient specimens, n=9 control RNA, 31 WGS verified, 1 apparent false positive). All specimens had low SARS-CoV-2 RT-PCR CT values (CT <30), and each displayed an identical deletion pattern by fragment analysis. Strikingly, B.1.1.7 was absent in our patient population before January 2021, but prevalence has steadily increased each week. Using this method, we estimate the weekly prevalence rate to be 28.5% (8/28) in the last week of the study (3/8/21-3/12/21, Fig. 1F) despite declining rates of COVID-19 cases in north Texas. This is consistent with the trend currently observed in the United Kingdom.¹⁹

During validation studies, we realized the reverse primer for S144 overlapped with the W152C mutation present in the emerging B.1.429/ B.1.427 variants (California, CAL.20C).²⁰ The primer was thus redesigned to bind the wild type position preferentially and fail to amplify when the S:W152C (22018G>T) variant occurs. To further destabilize S144R binding to the 22018G>T variant, nucleotides at the 2nd and 3rd positions upstream from 3' terminus were mutated. (Fig. 2C, D). Primer S144_R2'T>G was tested against 4 WGS confirmed B.1.429 variants with W152C mutations (Fig. 2H). However, S144 dropout was not detected for any primer pair. Instead, we detected a single base pair deletion (Fig. 2A, B). This was clearest in the primer with the 2nd upstream nucleotide mutated from T>G (Supplemental Table 1, S144_R2'T>G). An alternative primer binding configuration skipping the 22018G>T variant could explain the 1 base pair deletion. This aberration provides specificity without loss in sensitivity for California variants, thus demonstrating the assay's flexibility to adapt to emerging strains.

When the assay was designed, an isolated ORF1A:del3675_3677 (without S gene deletions) was thought to differentiate P.1 and B.1.351 variants from B.1.1.7 (Fig. 2E, G). Therefore, when we detected isolated 9 nucleotide deletions in the ORF1A target, these VOCs were suspected (n=5). Interestingly, although WGS confirmed the ORF1A:3675_3677 deletion (Fig. 2I), neither P.1 nor B.1.351 were identified. Rather, three strains were classified as B.1.526 (N501Y absent and E484K present, Fig.2 J,K) and the other two cases as

B.1.241 (N501Y and E484K absent). This recurrence of ORF1A:del3675_3677 in multiple VOC indicates ORF1A:DEL3675_3677 may represent convergent evolution arising in geographically disparate locations. This assay can therefore detect mutations consistent with clinically important SARS-CoV-2 strains, as well as monitor for emerging strains with related genetic alterations.

The limit of detection varies by target but ranges from 2,000 copies/ reaction for ORF1A to 62.5 copies/ reaction for all additional targets (Supplemental Table 5). Furthermore, as contamination is a concern for PCR reactions with high levels of amplification, this assay can detect the presence of a mixture of WT and mutant DNA by identifying the existence of multiple peaks.

Epidemiologic data indicates co-occurrence of the three deletions targeted are present in 95% of B.1.1.7 strains. Thus, fragment analysis of these three loci may be sufficient for B.1.1.7 classification. A similar principle may be true for other strains with unique genetic mutations. Additionally, the high rate of mutational change within the SARS-CoV-2 spike protein may lead to additional serendipitous findings, similar to the original SGTF.

In summary, we have applied fragment analysis to SARS-CoV-2 VOC detection and epidemiology for the first time. This assay can detect characteristic viral genomic changes found in B.1.1.7 (UK), P.1 (Brazil), B.1.351 (South African), B.1.529 (New York), and B.1.427/ B.1.429 (California) lineages. This approach is cost-effective, multiplexable (up to 20-40 primer pairs can be added), scalable (up to 96 specimens), and easily adaptable to routine clinical workflow in laboratories with molecular biology expertise and instrumentation. Our future directions include adding additional characteristic markers of P.1/ B.1.351/ B.1.526/ B.1.429 (i.e. S:N501Y, S:E484K, S:242_244del, S:L452R, and Ins28269_28273) to distinguish strains with 95% accuracy.²¹

Acknowledgements:

The authors gratefully acknowledge the technical assistance of Hui Zheng, Huiyu Yao, Emily Ostman, Vanessa Torres and Ginger Koetting. Additionally, we thank Prithvi Raj and Vanessa Schmid for sequencing assistance, and all members of the UTSW COVID diagnostic laboratory for archiving and retrieving specimens.

2 Figures:

Figure 1

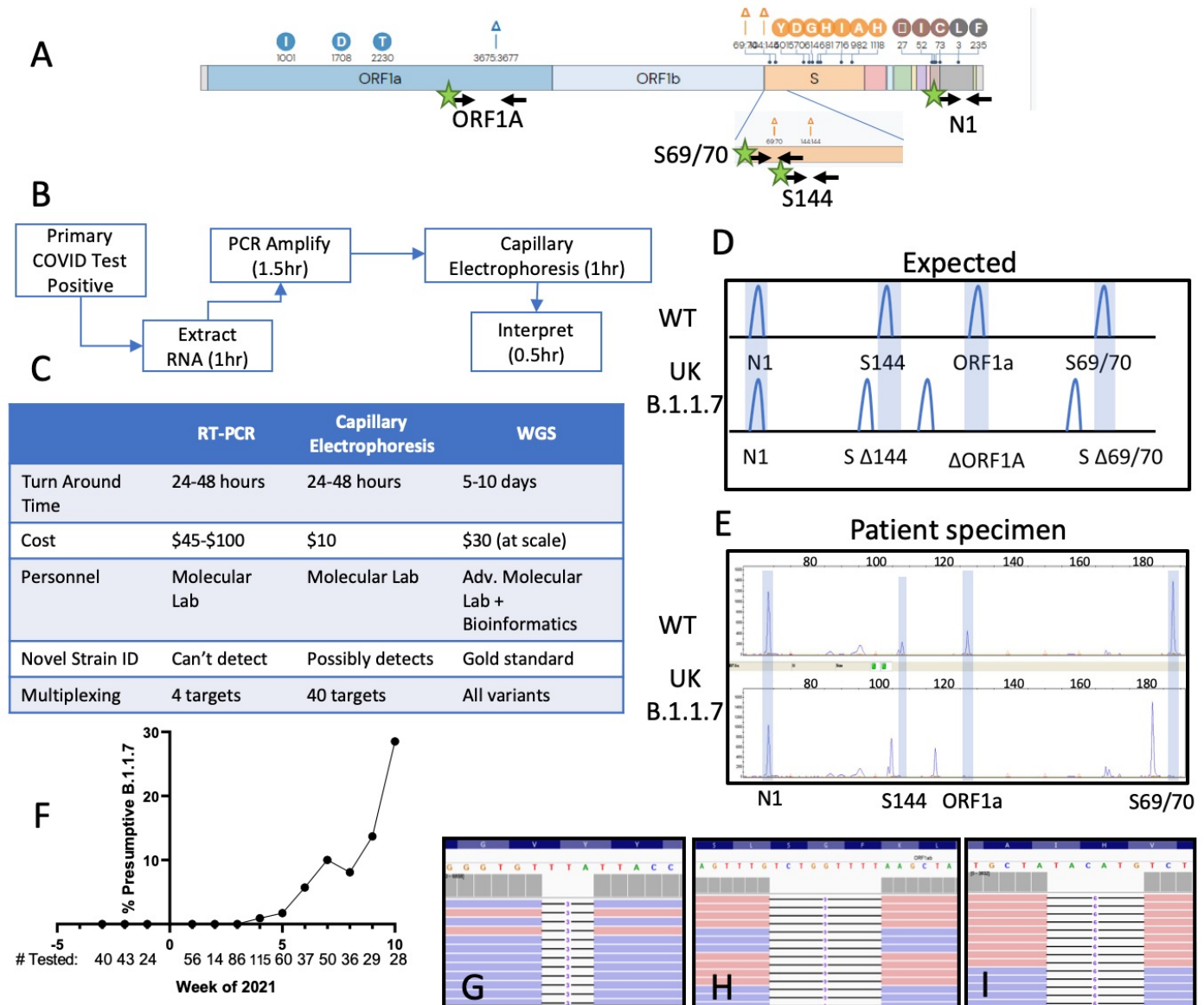


Figure 1. **Fragment analysis identifies the SARS-CoV-2 B.1.1.7 Variant of Concern.**

Genetic variations of the SARS-CoV-2 B.1.1.7 (UK) VOC. Black arrows indicate primer binding sites. Green stars indicating fluorescent dye (FAM) primer labels. (B) Routine clinical workflow and timetable of the SARS-CoV-2 fragment analysis assay. (C) Comparison of analytical and operational characteristics of real time PCR (RT-qPCR), fragment analysis by capillary electrophoresis and WGS (whole genome sequencing). (D) Expected deletions present in the SARS-CoV-2 B.1.1.7 strain compared to wild type as visualized by fragment analysis. (E) Electrophoretogram representative of two patients with WT and B.1.1.7 SARS-CoV-2 infections. (F) Estimated prevalence of presumptive B.1.1.7 per week with number of SARS-CoV-2 positive specimens tested below. (G-I) Integrated genome viewer visualization of the whole genome sequencing reads confirming the presence of the ORF1A:3675_3677del (G), S:69_70del (H), and S:144del (I) initially identified by capillary electrophoresis.

Figure 2

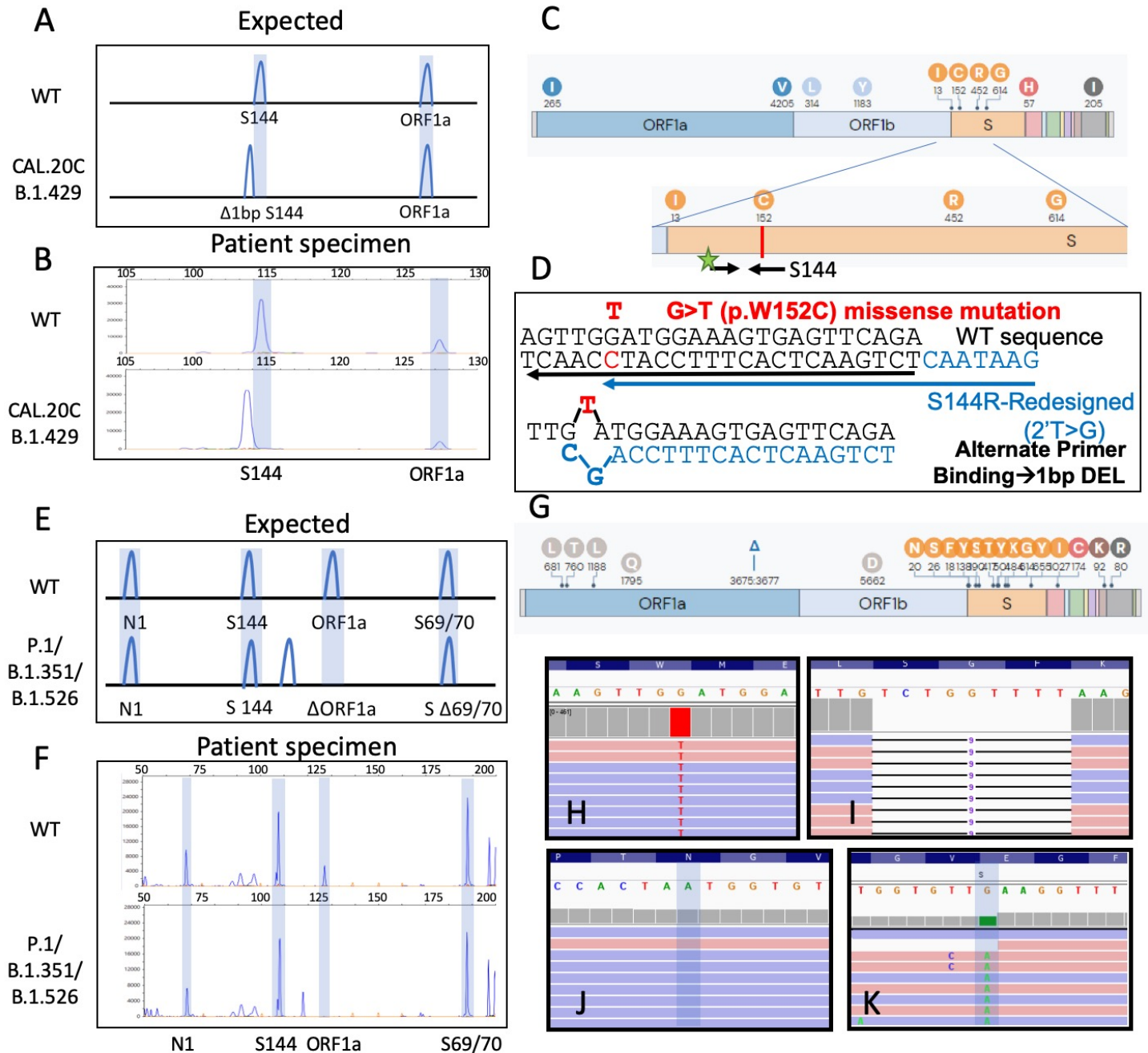


Figure 2. Detection of other SARS-CoV-2 variants by fragment analysis.

(A) Expected deletions present in B.1.429 (California) strain compared to wild type SARS-CoV-2. (B) Electrophoretogram representative of two patients with wild-type and B.1.429 SARS-CoV-2 infections. (C) Genetic variation of the SARS-CoV-2 B.1.429 strain. Black arrows indicate the primer binding sites for S144 target primers. The red line indicates the location of the W152C genetic variant overlapping with the original S114 reverse primer. (D) Detailed view of the variant causing W152C and proposed alternative primer binding site of S114 redesigned primer (blue arrow and blue text), which accounts for the single nucleotide deletion observed. (E) Expected deletions present in P.1, B.1.351, and B.1.526 strains compared to WT (wild type) SARS-CoV-2. (F) Electrophoretograms representative of two patients: one with wild type SARS-CoV-2 infection, the second with a virus that exhibits a nine base pair deletion in ORF1A. (G) Genetic variation of the SARS-CoV-2 P.1 strain. (H) Integrated genome viewer visualization of the whole genome sequencing reads confirming the presence of the SW152C variant characteristic of the B.1.429 strain (H) and ORF1A:3675_3677del (I), which is

- 8 found in each Variant of Concern. Lack of the N501Y (J) and E484K (K) mutations precludes classification of this
- 9 SARS-CoV-2 strain as P.1/ B.1.351.

References:

1. Volz E, Mishra S, Chand M, et al. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv*. January 2021:2020.12.30.20249034. doi:10.1101/2020.12.30.20249034
2. Muik A, Wallisch A-K, Sanger B, et al. Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine-elicited human sera. *bioRxiv*. January 2021:2021.01.18.426984. doi:10.1101/2021.01.18.426984
3. De Marco A, ATM Ferreira I, Meng B, et al. SARS-CoV-2 B.1.1.7 sensitivity to mRNA vaccine-elicited, convalescent and monoclonal 2 antibodies 3 4 Dami A. *Lourdes Ceron-Gutierrez L*. 2021;10(2):12. doi:10.1101/2021.01.19.21249840
4. *Investigation of Novel SARS-CoV-2 Variant Variant of Concern 202012/01 Technical Briefing 5*. www.gov.uk/government/publications/nervtag-paper-on-covid-19-variant-of-concern-b117. Accessed February 22, 2021.
5. Iacobucci G. Covid-19: New UK variant may be linked to increased death rate, early data indicate. *BMJ*. 2021;372. doi:10.1136/bmj.n230
6. Cele S, Gazy I, Jackson L, et al. Escape of SARS-CoV-2 501Y.V2 variants from neutralization by convalescent plasma. *medRxiv*. January 2021:2021.01.26.21250224. doi:10.1101/2021.01.26.21250224
7. Greaney AJ, Loes AN, Crawford KH, et al. Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies. *bioRxiv*. January 2021:2020.12.31.425021. doi:10.1101/2020.12.31.425021
8. Hoffmann M, Arora P, Gro R, et al. SARS-CoV-2 variants B.1.351 and B.1.1.248: Escape from therapeutic 1 antibodies and antibodies induced by infection and vaccination 2 3. doi:10.1101/2021.02.11.430787
9. Collier DA, De Marco A, Ferreira IATM, et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature*. March 2021. doi:10.1038/s41586-021-03412-7
10. Filkins L, SoRelle JA, Schoggins J, Park JY. Laboratory action plan for emerging SARS-CoV-2 variants. *Clin Chem*. January 2021. doi:10.1093/clinchem/hvab020
11. Solutions for surveillance of the S gene mutation in the B.1.1.7 (501Y.V1) SARS-CoV-2 strain lineage - Behind the Bench. <https://www.thermofisher.com/blog/behindthebench/solutions-for-surveillance-of-the-s-gene-mutation-in-the-b117-501yv1-sars-cov-2-strain-lineage/>. Accessed February 22, 2021.
12. Washington NL, Gangavarapu K, Zeller M, et al. Genomic epidemiology identifies emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. doi:10.1101/2021.02.06.21251159
13. Washington NL, White S, Schiabor Barrett KM, Cirulli ET, Bolze A, Lu JT. S gene dropout patterns in

- 3 SARS-CoV-2 tests suggest spread of the H69del/V70del mutation in the US. *medRxiv*. December
- 4 2020:2020.12.24.20248814. doi:10.1101/2020.12.24.20248814
- 5 14. McCarthy KR, Rennick LJ, Nambulli S, et al. Natural deletions in the SARS-CoV-2 spike glycoprotein drive
- 6 antibody escape. *Science (80-)*. February 2021. doi:10.1101/2020.11.19.389916
- 7 15. Tegally H, Wilkinson E, Lessells RJ, et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat Med*.
- 8 February 2021. doi:10.1038/s41591-021-01255-3
- 9 16. Nonaka CK V., Franco MM, Gräf T, et al. Early Release - Genomic Evidence of SARS-CoV-2 Reinfection
- 0 Involving E484K Spike Mutation, Brazil - Volume 27, Number 5—May 2021 - Emerging Infectious
- 1 Diseases journal - CDC. doi:10.3201/EID2705.210191
- 2 17. Kudo E, Israelow B, Vogels CBF, et al. Detection of SARS-CoV-2 RNA by multiplex RT-qPCR. *bioRxiv*. June
- 3 2020. doi:10.1101/2020.06.16.155887
- 4 18. Rambaut A, Holmes EC, Hill V, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic
- 5 epidemiology. *bioRxiv*. April 2020:2020.04.17.046086. doi:10.1101/2020.04.17.046086
- 6 19. Phe. *Investigation of Novel SARS-COV-2 Variant Variant of Concern 202012/01*.
- 7 20. Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E. Emergence of a Novel SARS-CoV-2
- 8 Variant in Southern California. *JAMA*. February 2021. doi:10.1001/jama.2021.1612
- 9 21. Long SW, Olsen RJ, Christensen PA, et al. Sequence Analysis of 20,453 SARS-CoV-2 Genomes from the
- 0 Houston Metropolitan Area Identifies the Emergence and Widespread Distribution of Multiple Isolates
- 1 of All Major Variants of Concern. *medRxiv*. March 2021:2021.02.26.21252227.
- 2 doi:10.1101/2021.02.26.21252227
- 3