Identification of a polymorphism in the N gene of SARS-CoV-2 that adversely impacts detection by RT-PCR.

Manu Vanaerschot[1], Sabrina A. Mann[1,2], James T. Webber[1], Jack Kamm[1], Sidney M. Bell[3], John Bell[4], Si Noon Hong[5], Minh Phuong Nguyen[5], Lienna Y. Chan[1], Karan D. Bhatt[1], Michelle Tan[1], Angela M. Detweiler[1], Alex Espinosa[4], Wesley Wu[1], Joshua Batson[1], David Dynerman[1], CLIAHUB Consortium, Debra A. Wadford[4], Andreas S. Puschnik[1], Norma Neff[1], Vida Ahyong[1], Steve Miller[2], Patrick Ayscue[1], Cristina M. Tato[1], Simon Paul[5], Amy Kistler[1], Joseph L. DeRisi[1,2], Emily D. Crawford[1,2,#]

[1]Chan Zuckerberg Biohub, San Francisco, California, USA

[2]University of California, San Francisco, California, USA

[3]Chan Zuckerberg Initiative, Redwood City, California, USA (Sidney M. Bell)

[4]California Department of Public Health, Richmond, California, USA

[5]Department of Public Health, Madera, California, USA

Running Head: N gene variant reduces SARS-CoV-2 test sensitivity

Address correspondence to Emily D. Crawford, emily.crawford@czbiohub.org .

New data letter

Since April 7, 2020, our COVID-19 diagnostic laboratory (CLIAHUB) has received samples from multiple counties in California – our RT-PCR protocol (1) employs N-gene (NIID_2019-nCov_N_F2/R2ver3/P2 (Japan) (2)) and E-gene (E_Sarbeco_F/R/P1 (Germany) (3)) simplex assays. In July 2020, we identified 40+ samples from Madera County with poor N-gene assay performance relative to the E-gene assay.

Figure 1A shows the concordance of C_t values for both assays in the 3,958 positive tests conducted during May 27 — August 7, 2020. For samples with positive E and N-gene results (n=3629), the N and E-gene C_t value difference (ΔC_t(N-E)) was 0.40±1.18 (mean±standard deviation).

Sequencing of the detected N-gene fragment of 57 samples with a ΔC_t(N-E) ≥2.96 (2.5 standard deviations above the mean) identified 46 samples (45 from Madera) to have a G29140U mutation located in the forward primer binding site (16th of 20 nucleotides) of the N-gene assay (Figure 1B). In 5 mutant samples the N-gene was undetectable by RT-PC but these cases were still recognized as positive for SARS-CoV-2 by the E-gene assay. The 11 wild-type samples with an increased ΔCt(N-E) are considered to be rare artefacts.

When the RT-PCR was repeated using a forward primer with full complementarity to the mutant sequence (Figure 1C), the mean ΔC_t(N-E) of 16 randomly-selected mutant samples dropped from 5.44 with the canonical primer to 0.19 with the mutated primer. This trend was inverted for the 14 randomly-selected wild-type samples where the ΔC_t(N-E) increased from 0.46 to 7.34

with the canonical and mutated primer, respectively. These data validate causality of G29140U for the observed aberrant C_t values of the N-gene assay, reducing its sensitivity by 67-fold.

G29140U encodes a Q289H amino acid mutation in the N-gene that was also found in 27 other sequences available on GISAID (4), showing world-wide occurrence of these mutants. Q289H is located within the dimerization domain of the nucleocapsid protein but is not involved in any known dimer interface interactions, though tertiary structure level interactions could be impacted by mutations at this position (5).

Whole genome sequencing of randomly-selected mutant (n=20) and wild-type samples (n=11) from Madera showed little genetic diversity between our mutant samples, and revealed that a GISAID-sequence from San Diego was identical by descent (Figure 1D). The remaining 26 mutants from GISAID fell on different clades of the tree, with 11 estimated recurrent mutation events at the locus.

Epidemiological data from Madera County indicated that the G29140U variant is replication-competent, retains its virulence, and adequately transmits within and between different communities (Supplementary Text).

Our data show that even in areas of high SARS-CoV-2 community spread, replication-competent mutations that impair RT-PCR performance can emerge and spread, leading to reduced test sensitivity and potentially under-diagnosis if only one viral target would have been used. Since mutations have been described in primer/probe-binding regions of all published SARS-CoV-2 diagnostic assays (6), our findings strongly support continuous monitoring for mismatches and

Journal of Clinical Microbiology

the routine use of at least two targets for SARS-CoV-2 detection by RT-PCR to avoid false negative results.

Acknowledgements

We thank Peter Kim and Don Ganem for helpful discussions and editorial assistance on the manuscript. This work was supported by the Chan Zuckerberg Biohub.

Conflict of Interest

The authors have no conflict of interest to declare.

Data and code availability

All code and Ct data are available in our Github repository: https://github.com/czbiohub/polymorphism_sarscov2_diagnostics. Sequence data is available via GISAID; see also Supplementary Table S1 for a list of sequences used.

References

1. CLIAHUB Consortium. 2020. Rapid deployment of SARS-CoV-2 testing: the CLIAHUB. PLoS Pathog accepted.

2. Shirato K, Nao N, Katano H, Takayama I, Saito S, Kato F, Katoh H, Sakata M, Nakatsu Y, Mori Y, Kageyama T, Matsuyama S, Takeda M. 2020. Development of genetic diagnostic methods for detection for novel coronavirus 2019(nCoV-2019) in Japan. Jpn J Infect Dis 73:304–307.

3. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW, Bleicker T, Brünink S, Schneider J, Schmidt ML, Mulders DGJC, Haagmans BL, Van Der Veer B, Van Den Brink S, Wijsman L, Goderski G, Romette JL, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C, Koopmans MPG, Drosten C. 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance 25:2000045.

4. Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Challenges 1:33–46.

5. Ye Q, West AMV, Silletti S, Corbett KD. 2020. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. Protein Sci https://doi.org/10.1002/pro.3909.

6. Wang R, Hozumi Y, Yin C, Wei G-W. 2020. Mutations on COVID-19 diagnostic targets. Genomics https://doi.org/10.1016/j.ygeno.2020.09.028.

Figure Legends

Figure 1. Single mutation in forward N gene primer binding site, prevalent across the world, decreased SARS-CoV-2 RT-PCR sensitivity. (A) Potential SARS-CoV-2 mutants were identified by their increased ΔC_t between the N and E gene assays (>2.5 × standard deviation of average ΔC_t, cut-off indicated by black lines). Dotted lines indicate the average C_t value at the limit of detection (LOD) of each assay, above which more variation is expected. NOTD: not detected. (B) Diagram showing a fragment of the N gene, with the N gene primers and probe originally developed by the National Institute of Infectious Diseases (NIID) in Tokyo, Japan (*4*) and the identified G29140T mutation indicated. (C) The increased ΔC_t(N-E) of mutant lines using the conventional RT-PCR with wild-type primer was reversed when a primer incorporating the mutation was used. The opposite was observed for wild-type samples that showed an increased ΔCt(N-E) when the mutated primer was used, further validating causality of the G29140U mutation for reduced N gene RT-PCR performance. Error bars indicate the standard error of the mean. **** indicates a significant difference determined by a t-test ($p<0.0001$). (D) Phylogeny of SARS-CoV-2 isolates with N mutation, including those with the G29140U mutation. Inferred mutation events on the tree are annotated with an * that is colored depending on the allele. Both synonymous variants of the Q289H mutant are found, with the mutation estimated to have recurred 11 times on the tree, and only one of the mutant samples from GISAID was identical by descent to the Madera cluster. One of the wild-type Madera samples was closely related to the mutant cluster, with a common ancestor just before the mutation event. Sequence data is available in Supplementary Table S1. All code used for analyses and figure generation are described in the Supplementary Text.