Key Summary of Conference Abstract

Analytical Validation of a SARS-CoV-2 Whole-Genome Sequencing Method by Amplicon-based NGS

Background

- Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the pathogen that causes coronavirus disease 2019 (COVID-19).
- Adaptive changes in the SARS-CoV-2 genome could affect virus detection, transmissibility, and pathogenicity. Whole-genome sequencing (WGS) could help monitor the changes, identify potential therapeutic targets, follow transmission, and contribute to sustained vaccine clinical utility.[1]
- Several WGS methods for SARS-CoV-2 have been developed, but clinical performance of the assays is uncertain.
- **Objective:** The investigators of this study developed an amplicon-based SARS-CoV-2 WGS method that uses next-generation sequencing (NGS) and assessed the analytical performance of the assay on clinical specimens.

Methods

- The validation study included remnant extracted RNA from deidentified clinical specimens consecutively collected in March 2020 for SARS-CoV-2 testing: 141 that were positive for SARS-CoV-2 (cycle threshold [Ct] ranged 31 to 9, indicating ~40 copies to 163 million copies) from unique patients; 24 pools of positive samples (pooled RNA from 3 replicates with similar Ct); and 24 negatives.
- Remnant extracted RNA was reverse transcribed to cDNA and PCR amplified using ARTIC Network primers. NGS libraries were prepared from the resulting amplicons and sequenced using an Illumina MiSeq sequencer.
- An in-house bioinformatics pipeline was used to generate consensus genomes and identify variants relative to the MN908947.3 reference genome, Wuhan-Hu-1.

Results

- For both inter- and intra-assay precision studies, 96% (66 of 69) of specimens with a Ct ≤30 had 100% consensus sequence coverage.
  - For 22 inter- and 22 intra-assay replicates of pooled positives, amino acid variants present in ≥15% of the reads were 100% concordant in all 3 replicates.
- Of the 141 positive patient specimens, 127 (>90%) provided high-quality sequence data that could be used for clade classification: 60% were clade G, 25% were clade S, 13% were clade O, and 2% were clade V.
- Of 127 specimens, 117 generated ≥99% consensus sequence coverage, which was used for variant analysis.
  - A median of 7 (IQR: 6-8) amino acid substitutions per genome was identified, though it differed by clade; the median was highest in clade S (8.5; IQR: 7-11; *P*<0.0001).
  - Coding regions *orf3a* and *orf8* had a higher proportion of variants than did other coding regions (*P*<0.0001).

Conclusions

- The amplicon-based SARS-CoV-2 WGS method described by the investigators produced near-complete genome coverage of the virus from clinical specimens.
- This method may help classify SARS-CoV-2 subspecies and track changes in the virus genome.

Poster presentation at the Association for Molecular Pathology annual meeting (AMP2020)

Authors
Sun Hee Rosenthal, Ron M Kagan, Anna Gerasimova, Yan Liu, Michael Hua, Ben Anderson, Joseph Catanese, Felicitas Lacbawan, Renius Owen

Affiliation
Quest Diagnostics, San Juan Capistrano, CA USA

AMP 2020 Virtual Meeting
Date: November 16-20, 2020

Webpage
https://www.jmdjournal.org/article/S1525-1578(20)30513-4/pdf

Reference

1. Nasir JA, Kozak R, Aftanas P, et al. A comparison of whole genome sequencing of SARS-CoV-2 using amplicon-based sequencing, random hexamers, and bait capture. *Viruses*. 2020;12:895. doi: 10.3390/v12080895