A Universal Day Zero Infectious Disease Testing Strategy Leveraging CRISPR-based Sample Depletion and Metagenomic Sequencing

Agnes P. Chan[2†], Azeem Siddique[1†], Yvain Desplat[1], Yongwook Choi, Sridhar Ranganathan[1], Kumari Sonal Choudhary[1], Josh Diaz[1], Jon Bezney[1], Dante DeAscanis[1], Zenas George[1], Shukmei Wong[2], William Selleck[2], Jolene Bowers[2], Victoria Zismann[2], Lauren Reining[2], Sarah Highlander[2], Yaron Hakak[1], Jon R. Armstrong[1], Keith Brown[1], Nicholas J. Schork[2,3,4*], Jumpcode Genomics, San Diego, CA 92121,
The Translational Genomics Research Institute (TGen), An Affiliate of the City of Hope National Medical Center, Phoenix, AZ 85004
The University of California San Diego, San Diego CA 92093, The Scripps Research Institute, San Diego, CA 92037
†These authors contributed equally to the work as co-first authors
*Nicholas J. Schork, 445 N. Fifth Street, Phoenix, AZ 85004, 602-343-8429, nschork@tgen.org

Abstract

The lack of preparedness for detecting the highly infectious SARS-CoV-2 pathogen, the pathogen responsible for the COVID-19 disease, has caused enormous harm to public health and the economy. It took ~60 days for the first reverse transcription quantitative polymerase chain reaction (RT-qPCR) tests for SARS-CoV-2 infection developed by the United States Centers for Disease Control (CDC) to be made publicly available. It then took >270 days to deploy 800,000 of these tests at a time when the estimated actual testing needs required over 6 million tests per day.

Testing was therefore limited to individuals with symptoms or in close contact with confirmed positive cases. Testing strategies deployed on a population scale at 'Day Zero' i.e., at the time of the first reported case, would be of significant value. Next Generation Sequencing (NGS) has such Day Zero capabilities with the potential for broad and large-scale testing. However it has limited detection sensitivity for low copy numbers of pathogens which may be present.

Here we demonstrate that by using CRISPR-Cas9 to remove abundant sequences that do not contribute to pathogen detection, NGS detection sensitivity of COVID-19 is equivalent to RT-qPCR. In addition, we show that this assay can be used for variant strain typing, co-infection detection, and individual human host response assessment, all in a single workflow using existing open-source analysis pipelines. This NGS workflow is pathogen agnostic, and therefore has the potential to transform how both large-scale pandemic response and focused clinical infectious disease testing are pursued in the future.

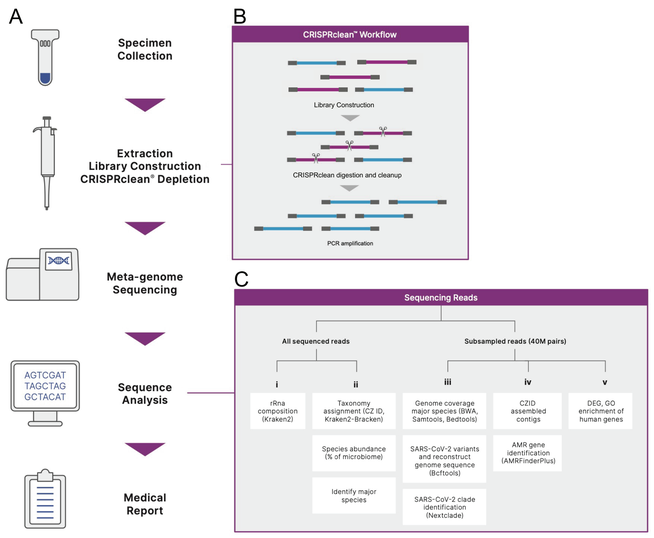

Figure 1: Study workflow and design.
(A) From clinical specimens to reporting.
(B) CRISPclean workflow.
(C) Data analysis workflow:
I. Estimating rRNA composition
II. Reporting taxonomic classification and the abundance of pathogens and coinfections as percentage of microbiome (non-human) reads
III. Reporting pathogen genome coverage metrics,
IV. Reporting AMR, and v. Reporting host gene expression.

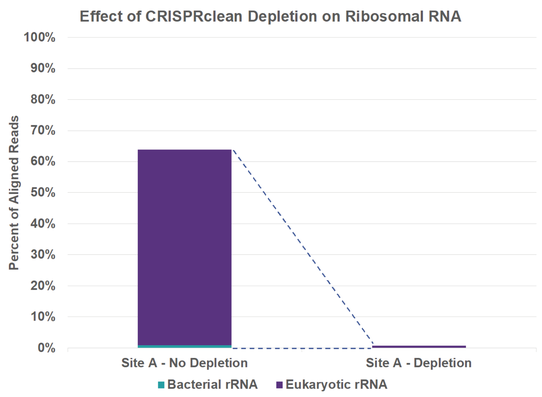

Figure 2: Ribosomal RNA composition before and after CRISPRClean depletion at Site A.
The average percent of ribosomal aligned reads (y-axis) was determined for bacterial (blue) and eukaryotic ribosomal (purple) in all sample libraries from site A (n=180). Percent of aligned reads is shown with and without CRISPRclean depletion. CRISPRclean depletion removes nearly all bacterial and eukaryotic RNA.

Study design and samples used for sequencing studies

Two types of samples were analyzed in this study: clinical specimens and contrived samples. For clinical specimens, human nasal swabs with COVID-19 infection status determined by RT-qPCR were previously collected from two locations, one in California and one in Arizona (referred to as site A and site B, respectively), and then processed and sequenced at separate sites (Jumpcode Genomics, San Diego, CA for Site A samples and TGen, Phoenix, AZ for Site B samples). In total, 56 patient specimens with confirmed positive COVID-19 status and 16 specimens with no detectable SARS-CoV-2 were analyzed in this study.

The contrived samples were generated by combining known viral RNA pathogens with human RNA. A viral reference genome mix, consisting of four RNA viruses (Zika virus, mammalian orthoreovirus, influenza B virus, human orthopneumovirus (i.e., respiratory syncytial virus, RSV) and SARS-CoV-2) was spiked at various concentrations into human lung total RNA samples. The samples contained a 10-fold dilution series of the viral reference mix, with titers ranging from an estimated 20 copies of each viral RNA genome to 20,000 copies. Various negative controls were also prepared, including human lung total RNA only (a no-viral RNA
control) and a water-only sample (no-template control). The latter was included to monitor background contaminants that may originate from molecular reagents and NGS workflows. An overview of the workflow is shown in Figure 1.

Both mock depletions (depletion without Cas9 and guide RNA) and Cas9 depletions were performed with site A specimens. Mock-depletion was not performed with site B specimens. A summary of sequencing statistics is provided in Supplementary Table 1.

A

CRISPRclean NGS \ RT-qPCR (C_t < 35)	Positive	Negative	Total		
Site A (n = 93)					
Positive	63	0	63	100%	PPV
Negative	1	29	30	96.7%	NPV
Total	64	29	93		
	98.4%	100%			
	Sensitivity	Specificity			

B

CRISPRclean NGS \ RT-qPCR (C_t < 35)	Positive	Negative	Total		
Site B (n = 24)					
Positive	20	0	20	100%	PPV
Negative	1	3	4	75%	NPV
Total	21	3	24		
	95%	100%			
	Sensitivity	Specificity			

Table 1: Contingency tables comparing the performance of CRISPRclean NGS and RT-qPCR for Site A and Site B. Samples with Ct < 35 were processed with the CRISPRclean NGS assay and positive/negative results compared to RT-qPCR results from the same samples.
Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated for both Site A and B. CRISPRclean results are comparable those seen from RT-qPCR.

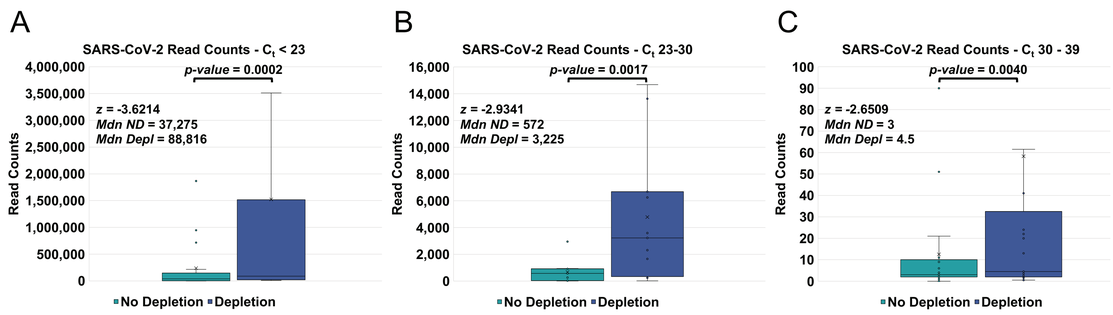

Figure 3: Sequencing read counts for SARS-CoV-2 in clinical specimens across Ct values.
The sequencing read counts, shown on the y-axis, from Kraken[2] workflow were calculated for non-depleted (blue) and depleted (purple) samples. Box and whisker plots were generated for three cycle threshold (C_t) bins.
A. C_t <23 (non-depleted - n = 17, depleted – n = 34). B. C_t 23-30 (non-depleted – n = 11, depleted – n = 22). C. C_t 30 – 39 (non-depleted – n = 17, depleted – n = 34).
Values for the two depleted sample replicates were averaged and compared to single non-depleted samples in order to provide paired vales for Wilcoxon Singed-Rank test. The Wilcoxon Signed-Rank Test indicated that sequence read counts to SARS-CoV-2 genome were statistically significantly higher with CRISPRclean depletion than without depletion. The z value (z), median of non depleted (Mdn ND) and depleted (Mdn Depl) samples are shown in the upper left of the graph for each C_t bin.

Conclusion / Highlights

- Agnostic: Whole genomes or Transcriptomes from all pathogens (viral, bacterial or fungal)
- Detect co-infections
- Predict human host response for severity or tolerance
- Day Zero Capabilities

To learn more visit **jumpcodegenomics.com**

Download the poster