A CRISPR-Powered Universal Infectious Disease Assay

Keith Brown[5] (keith@Jumpcodegenomics.com), Azeem Siddique,[1,3] , Dhanya Ramachandran,[1,3] Gaia Suckow,[1,3] Nils Homer,[2] Jay Carey,[2] Phillip Ordoukhanian,[1,3] Steve Head,[1,3] Stan Nelson[3] | 1 The Scripps Research Institute, La Jolla, CA; 2 Fulcrum Genomics, Somerville, MA; 3 Jumpcode Genomics, San Diego, CA

Highlights

- A CRISPR-based method of depletion of uninformative NGS sequences
- Allows detection of all viral, bacterial and fungal pathogens in a single test
- High detection sensitivity

Abstract

The COVID-19 pandemic has brought intense awareness to the dangers of zoonotic diseases and the real danger they pose to global human health and welfare. The pandemic highlights how easily an infectious agent can spread through a population, particularly when a significant proportion of affected individuals is asymptomatic and those who are symptomatic have symptoms that can be easily mistaken for any number of diseases. As such, a sensitive and specific means of detecting the virus has been essential to tracking it and implementing effective public health measures. PCR testing is rapid and sensitive but has significant limitations: (1) by its nature, it is not available when the initial outbreak occurs, limiting its effectiveness when tracking is most needed to prevent an epidemic, (2) it is of little use for genome surveillance and tracking viral evolution, and (3) it does not provide any information on co-infections and the human host response to the infection.

Next generation sequencing (NGS) approaches, in which the entire sample is sequenced to obtain total genetic information, circumvent the need for a priori knowledge of the infectious agent and address the shortcomings of PCR listed above. Here we present an RNA metagenomic next generation sequencing (mNGS) protocol that enables detection of the SARS-CoV-2 genome sequence, the source of any co-infection, and the host transcriptional response in a single workflow. The critical component enabling this approach is JumpCode's CRISPRclean technology, which removes abundant human and bacterial ribosomal RNA sequences from mNGS libraries.

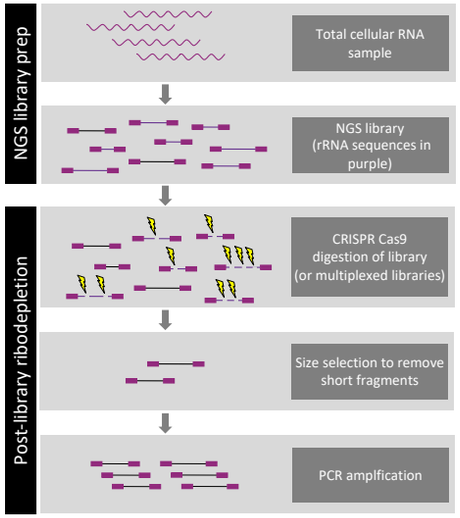

Figure 1: Schematic of the Jumpcode CRISPRclean™ protocol.

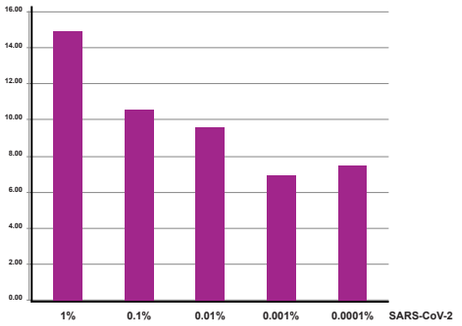

Figure 2: Fold enrichment of SARS-CoV-2 reads after CRISPRclean depletion of libraries prepared from contrived samples.

SARs COV-2 fraction of total RNA	% of genome covered at 1x	% genome covered at 10x
1.0000%	100%	100%
0.1000%	100%	100%
0.0100%	100%	100%
0.0010%	80%	48%
0.0001%	30%	7%

Table 1: Coverage of the SARS-CoV-2 genome at 50 million reads.

Methods:

JumpCode CRISPRclean depletion performance in meta-transcriptomic applications was evaluated by preparing contrived and semi-contrived samples. Contrived samples consist of 1 ng of lung-derived total RNA (Thermo Fisher Scientific, Waltham, MA) combined with different fractions of Staphylococcus aureus total RNA (kindly provided to us by Laurey Steinke, University of Nebraska Medical Center) and a synthetic SARS-CoV-2 control (control consists of six 5-kb fragments of the "Wuhan" strain produced by Twist Biosciences (San Francisco, CA)). Both *S. aureus* and SARS-CoV-2 were added to the samples at 1%, 0.1%, 0.01%, 0.001% and 0.0001% of total RNA. For SARS-CoV-2, this amounts to a maximum of 600,000 and a minimum of 60 viral copies. The semi-contrived set of samples was prepared from 100 ng of total RNA from two patient stool samples kindly provided by DNA Genotek (Ottawa, Canada), spiked with 600 or 60 copies of the same synthetic SARS-CoV-2 control (0.00001% or 0.000001% of total RNA, respectively).

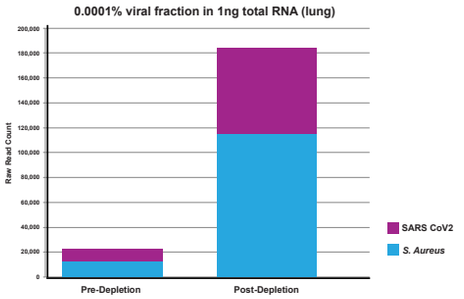

Figure 3: Number of reads aligning to the *S. aureus* and SARS-CoV-2 genomes before and after CRISPRclean depletion.

NEBNext Ultra II Directional RNA libraries were prepared from the samples. The libraries were subsequently treated with JumpCode CRISPRclean by following the protocol described in Figure 1. Libraries were sequenced on NovaSeq 6000 (contrived samples) and NextSeq 500 (semi-contrived samples) instruments. Data analysis was performed using JumpCode's proprietary software to measure alignment and depletion rates, the Silva database for ribosomal RNA read alignment, and the Kraken2 and the CosmosID pipelines for k-mer based metagenomic investigation.

Results:

CRISPRclean treatment of the fully contrived samples increases the fraction of reads that map to the SARS-CoV-2 genome by an average of ~10-fold (Figure 2). For the sample containing 0.0001% SARS-CoV-2, (60 viral copies), the number of reads mapping to the SARS-CoV-2 genome increases from ~10,000 reads to ~70,000 reads, as shown in Figure 3. A similar increase in reads occurs for *S. aureus*. Consistent with the increase in aligned reads, the percentage of SARS-CoV-2 genome covered at 1X and 10X also increases as a consequence of rRNA depletion (Table 1).

The sequencing data were also evaluated using the CosmosID shotgun metagenomic sequencing analysis pipeline, which employs a k-mer based approach to interrogate a database of bacterial, viral and fungal genome sequences (CosmosID, Rockville, MD). A heat map summarizing the results of a query of the viral database is shown in Figure 4. Similar results were achieved even after downsampling the datasets to 5 million reads.

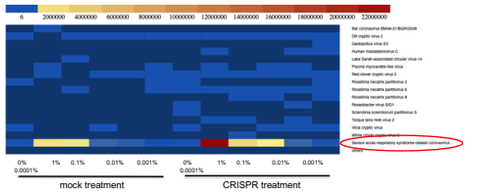

Figure 4: A heat map showing read alignments to viral genomes. The yellow color indicates high read counts. The CosmosID shotgun metagenomic analysis software was used to analyze the sequencing data, classify the sequences and generate the heat map.

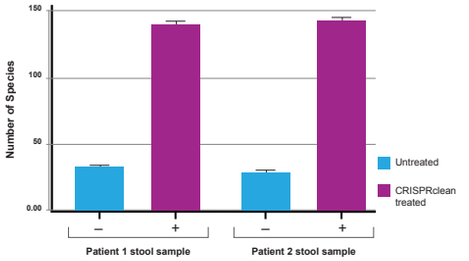

Figure 5: Bacterial species composition of patient stool samples before and after CRISPRclean depletion. Sequencing data downsampled to 20 million reads.

Data from the semi-contrived samples were analyzed on the CosmosID platform. Figure 5 shows that there is a ~4-fold increase in bacterial species detection in these stool samples after CRISPRclean treatment. SARS-CoV-2 was detectable by sequencing even when present at only 60 copies in the sample (i.e., 0.000001% of the sample) [data not shown].

Conclusions:

- mNGS powered by CRISPR-mediated depletion of abundant ribosomal RNA species offers a robust methodology for the detection and understanding of infectious disease. It enables clinicians and researchers to acquire viral genomic data, microbiome composition, including information on any co-infection, and the transcriptional status of the host immune response in a single workflow.
- This sequencing-based approach to infectious disease diagnosis could be available on the 1st day of the next viral outbreak and should be considered as a first-line test for novel zoonotic virus detection.

To learn more, visit Jumpcodegenomics.com/