CRISPRclean® Powered Host Depletion Improves Pathogen Detection in a Shotgun RNA-Sequencing Based Universal Infectious Disease Test

Keith Brown*, Jon Armstrong*, Azeem Siddique*, Sridhar Ranganathan*, K. Sonal Choudhary*, Faizan Khalid*, Nicholas Schork**, Agnes Chan**, Yongwook Choi**, Andrew Boddicker***, Bryan Lajoie***, Semyon Kruglyak***, Junhua Zhao***, Scott Im***, Andrew Price***, Matthew Kellinger***, Kerry Gunning****, Amanda Cramer****

* Jumpcode Genomics, ** Translational Genomics Research Institute, *** Element Biosciences, **** PerkinElmer

Introduction

It took 42 days from the first reported clinical case of COVID-19 until the viral genome sequence was provided to the public. This critical information release acts as the starting point for diagnostic test development and mRNA vaccine generation. The first EUA approved RT-PCR tests in the United States were made available after 70 days on February 8th, 2020. Shotgun metagenomic NGS sequencing is the only technology capable of unbiased detection from previously unknown pathogens available on day ZERO for future pandemics. Beyond detection, meta-transcriptomic sequencing enables full genome sequence assembly, the ability to catalog co-infections, and provides valuable data on host response through host transcriptional profiling. Sensitivity of shotgun mNGS is hindered by abundant eukaryotic (host) and prokaryotic rRNA sequences from nasal swabs. Using a guide set of over 12,500 sgRNAs targeting abundant human and bacterial ribosomal RNA sequences in the CRISPRclean® Plus workflow, we show that sequence specific cleavage results in read space shifting from ribosomal "contaminants" to provide an average 10-fold increase in sensitivity to detect non-human sequences. In this study, 55 PCR positive samples and 17 negative samples by RT-PCR, from labs in California and Arizona, were subjected to amplicon-based sequencing and shotgun metagenomic sequencing with and without rRNA depletion using CRISPRclean. We report detection sensitivity comparisons to RT-PCR, variant concordance to amplicon sequencing and an early look at host response through transcriptional profiling across technical replicates. In addition to HCoV2, this study has identified several co-infections in both RT-PCR positive and negative samples as well as host response signatures. The protocol, including a strand specific library prep with unique dual indexes (UDIs) has been fully automated from extraction through library construction and depletion. The full solution represents a viable approach for pandemic preparedness by enabling immediate detection of pathogens at the site of an outbreak or points of entry into or out of the country. Bridging the time gap from initial outbreak to the availability of low cost, specific pathogen detection at scale could save lives in the future.

Methods

A total of 78 samples collected at two clinical sites (PerkinElmer, CA, TGen AZ) were used in this study. Samples include 55 RT-PCR confirmed positive samples (Ct values ranging from 14-39), 17 RT-PCR negative samples, and 6 contrived samples (Hu.,SARs-2, and 6 additional viruses from ATCC). All samples were subjected to next generation sequencing library construction (PerkinElmer) and CRISPRclean Plus depletion workflow using 1-10ng of starting total RNA input (**Figures 1 and 2**). Sequencing was performed on the Illumina and Element Biosciences sequencing platforms (Element AVITI™ System). Illumina data was analyzed using STAR for alignment to the SARS-CoV-2 genome and genome coverage calculations (**Figure 3 and 4**), CZ-ID with uploads to Nextclade for clade/strain ID (**Figure 5**), Kraken2 for microbial composition (**Figure 6**), Bowtie/RSEM/DGenR for human transcriptome analysis and differential gene expression and custom pipelines for ribosomal depletion rates (data not shown).

Sixteen libraries, consisting of 14 RT-PCR positive and two negative samples were resequenced on the AVITI™ System in order to compare Element data outputs to Illumina data. Analysis pipelines used to characterize Element data for comparison to Illumina data, included: STAR for alignment to the SARS-CoV-2 genome for coverage and human genome for gene expression levels (**Figure 7**), CZ-ID with uploads to Nextclade for clade/strain ID and Kraken2 for microbial composition (data not shown).

Technology Schematic

Figure 1: Overview of CRISPRclean technology. Double stranded cDNA library products are cleaved, following adapter ligation, using CAS9 and specifically designed guide RNAs. Cleaved fragments do not contain adapters on both ends and will not amplify in subsequent PCR reactions. The cleaved fragments are removed from the library pool during short fragment cleanup.

Study Design

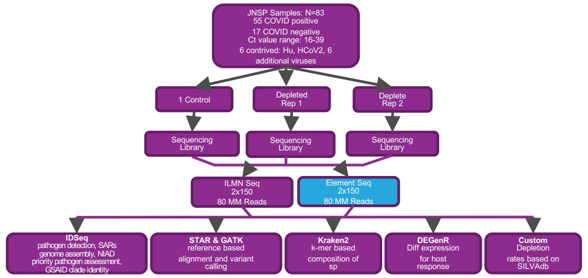


Figure 2. COVID Study Design. Schematic of samples used, libraries generated, sequencing platforms, and analysis approach.

Depth & Breadth of SARS-CoV-2 Genome Coverage versus C_t

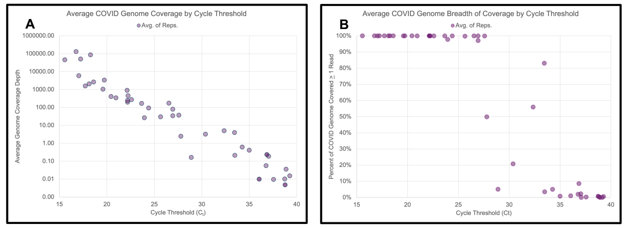


Figure 3. Depth and breadth of COVID genome coverage in CRISPRclean Plus samples over a range of cycle thresholds. (A) Average COVID genome coverage depth (y-axis, $\log_{10}$) from two replicate depleted samples, plotted against RT-PCR cycle threshold in 45 COVID positive sample (purple circles) ranging in C_t from 16 – 39. A linear relationship is observed between depth of genome coverage and C_t value. (B) Average percent COVID genome coverage breadth (y-axis) from two replicate depleted samples, plotted against RT-PCR cycle threshold in 45 COVID positive sample (purple circles) ranging in C_t from 16 – 39. Percent of genome coverage is 100% up to a C_t of approximately 27.

Example Coverage Using STAR/GATK

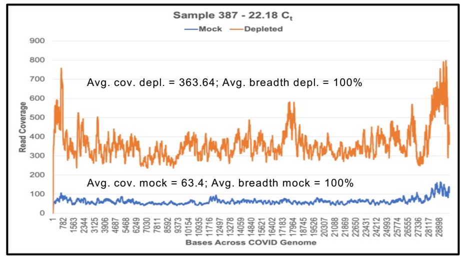


Figure 4. Comparison of COVID genome coverage between control and depleted sample 387. The genome coverage (y-axis) for sample 387 control is shown (blue line) across the COVID genome (x-axis). Genome coverage for sample 387 depleted is shown (orange line) across the COVID genome. Approximately a 6x increase in genome coverage is observed in the depleted sample.

Comparison of Detection & Strain Calling of SARS-CoV-2 Across C_t Values

$C_t < 30$		RT-PCR Detection Pos	RT-PCR Detection Neg
Jumpcode Detected	Pos	26	0
	Neg	1	15

$C_t \leq 35$		RT-PCR Detection Pos	RT-PCR Detection Neg
Jumpcode Detected	Pos	32	0
	Neg	1	15

$C_t < 40$		RT-PCR Detection Pos	RT-PCR Detection Neg
Jumpcode Detected	Pos	34	0
	Neg	11	15

$C_t < 30$		Variant-Seq Strain Call Pos	Variant-Seq Strain Call Neg
Jumpcode Strain Call	Pos	25	0
	Neg	1	15

$C_t \leq 35$		Variant-Seq Strain Call Pos	Variant-Seq Strain Call Neg
Jumpcode Strain Call	Pos	27	0
	Neg	5	15

$C_t < 40$		Variant-Seq Strain Call Pos	Variant-Seq Strain Call Neg
Jumpcode Strain Call	Pos	28	0
	Neg	16	15

	RT-PCR Detection			Variant-Seq Strain Call		
	Ct<30	Ct<35	Ct<40	C_t<30	$C_t \leq$35	C_t<40
Sens	96%	97%	76%	96%	84%	64%
Spec	100%	100%	100%	100%	100%	100%
PPV	100%	100%	100%	100%	100%	100%
NPV	94%	94%	58%	94%	75%	48%

Figure 5. Contingency table and resulting statistics for CRISPRclean Plus method. 45 COVID positive and 15 negative samples were analyzed for virus detection and strain calling using multiple methods at increasing cycle thresholds (C_t). The performance of the CRISPRclean Plus method (Jumpcode) versus CZ-ID for COVID virus detection, CZ-ID - Nextclade for COVID strain calling and PerkinElmer Variant-Seq™ strain call is shown in the contingency tables. Sensitivity (Sens), specificity (Spec), positive predictive value (PPV) and negative predictive value (NPV) was calculated for each technology and at increasing cycle thresholds.

Viral Heatmap

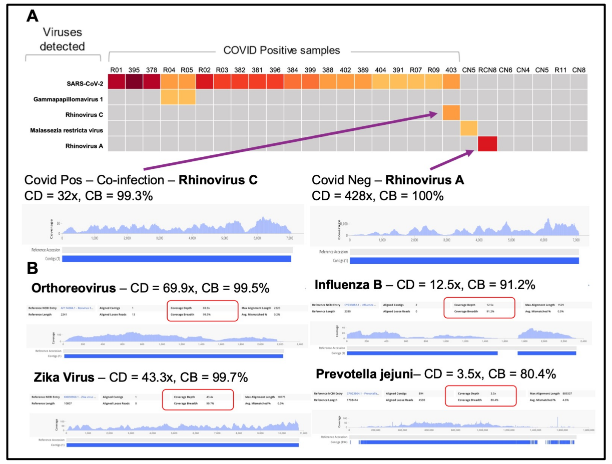
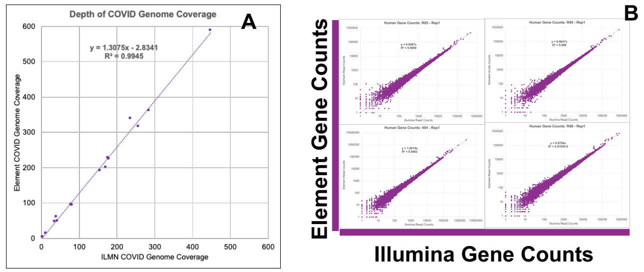


Figure 6. Heatmap of COVID positive and negative samples showing viral infection and co-infection with coverage plots. Heatmaps and coverage depth (CD) and coverage breadth (CB) plots were generated for viruses only in a subset of COVID positive and negative samples using CZ-ID. Heatmap color is on a log scale and determined from number of reads aligning to the taxon in the NCBI NT/NR database, per million reads sequenced (NT rPM) with a threshold cutoff of >= 10. The CRISPRclean Plus assay is able to identify the presence of viral material other than COVID due to the hypothesis neutral nature of the assay. Examples include co-infection of Rhinovirus C in COVID positive samples 403 and Rhinovirus A infection in COVID negative sample RCN8. Additional viral species are observed above threshold, Gammapapillomavirus 1 (COVID positive samples R04 and R05) and Malassezia restricta (COVID negative samples CN5). (B) Genome coverage of of ATCC viral pathogens included as positive controls. The bacterial species Prevotella jejuni showing a near-full length genome coverage was detected in one of the clinical specimens Coverage plots from CZ-ID are shown for 3 virus (Orthoreovirus, Zika and Influenza B) and 1 bacteria (Prevotella jejuni). Genome coverage can be generated for any organisms in a sample due to the hypothesis neutral aspects of the CRISPRclean Plus assay.

Data Comparison Between Illumina & Element



C		Element Strain		ILMN Strain	
Sample	Ct	Rep1	Rep2	Rep1	Rep2
R05	20.95	20G	20G	20G	20G
388	20.95	20G	20G	20G	20G
R04	23.66	20G	20G	20G	20G
R06	23.95	20G	20G	20G	20G
403	26.53	20A	20A	20A	20A
404	26.94	20G	20G	20G	20G
391	27.58	20G	20G	20G	20G
R08	27.79	20G	20G	20G	20G

Figure 7. Data comparison between the Illumina and Element AVITI™ System. (A) Comparison of the depth of COVID genome coverage between Illumina (x-axis) and Element (y-axis) platforms. Coverage depth is concordant between both platforms (R^2 = 0.9945) with a slight increase in genome coverage on the Element platform (slope = 1.30). (B) Correlation of reads counts, in four samples, to human genes between Illumina (x-axis) and Element (y-axis) platforms. Read counts to human genes are concordant between platforms (R^2 and slope ~ 1). (C) COVID strains calls across a range of cycle thresholds on both platforms. High concordance is seen between strain calls. The CRISPRclean Plus assay is platform agnostic.

Conclusions

Shotgun meta-transcriptomic sequencing, improved by the in-vitro use of the CRISPR-CAS system, is a viable solution for pandemic preparedness enabling ZERO-Day diagnostic capabilities at population scale. A single test enables pathogen detection, variant calling, co-infection and human host response.

Highlights of Jumpcode CRISPRclean Plus

1. Low sample inputs
2. Single day protocol
3. Automated and high throughput
4. 10 – fold increase in sensitivity of detection
5. 1 assay for multiple results: detection, variant calling, co-infection and immune response

Highlights of Element AVITI™ System

1. Compatibility with existing libraries
2. Concordance with Illumina results
3. Highly accurate data

To learn more visit jumpcodegenomics.com