High-throughput detection of SARS-CoV-2

Marie-Ming Aynaud

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital <https://orcid.org/0000-0003-1788-3180>

J. Javier Hernandez

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital

Seda Barutcu

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital <https://orcid.org/0000-0003-2739-4153>

Ulrich Braunschweig

Donnelly Centre for Cellular and Biomolecular Research, University of Toronto

Kin Chan

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital

Joel D. Pearson

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital

Daniel Trcka

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital

Suzanna L. Prosser

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital

Jaeyoun Kim

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital

Miriam Barrios-Rodiles

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital

Mark Jen

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital

Siyuan Song

Donnelly Centre for Cellular and Biomolecular Research, University of Toronto

Jess Shen

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital

Christine Bruce

Mount Sinai Hospital/University Health Network

Bryn Hazlett

Mount Sinai Hospital/University Health Network

Susan Poutanen

Mount Sinai Hospital/University Health Network

Liliana Attisano

Donnelly Centre for Cellular and Biomolecular Research, University of Toronto

Rod Bremner

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital

Benjamin J. Blencowe

Donnelly Centre for Cellular and Biomolecular Research, University of Toronto

Tony Mazzulli

Mount Sinai Hospital/University Health Network

Hong Han

Donnelly Centre for Cellular and Biomolecular Research, University of Toronto

Laurence Pelletier

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital <https://orcid.org/0000-0003-1171-4618>

Jeffrey L. Wrana (✉ wrana@lunenfeld.ca)

Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital <https://orcid.org/0000-0003-0932-0644>

Method Article

Keywords: SARS-CoV-2, C19-SPAR-Seq, High Throughput screening, variant detection

Posted Date: March 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.pex-1356/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

Population scale sweeps of viral pathogens, such as SARS-CoV-2, require high intensity testing for effective management. This protocol describes the “Systematic Parallel Analysis of RNA coupled to Sequencing for Covid-19 screening” (C19-SPAR-Seq), a multiplexed, scalable, readily automated platform for SARS-CoV-2 detection that is capable of analyzing tens of thousands of patient samples in a single run. To address strict requirements for control of assay parameters and output demanded by clinical diagnostics, we employed a control-based Precision-Recall and Receiver Operator Characteristics (coPR) analysis to assign run-specific quality control metrics. C19-SPAR-Seq coupled to coPR on a trial cohort of several hundred patients performed with a specificity of 100% and sensitivity of 91% on samples with low viral loads, and a sensitivity of > 95% on high viral loads associated with disease onset and peak transmissibility. This study establishes the feasibility of employing C19-SPAR-Seq for the large-scale monitoring of SARS-CoV-2 and other pathogens.

Introduction

Reagents

Equipment

Procedure

RNA purification (format 96 well plates)

For one reaction:

- 1- Prepare 14.5 μ L of 10X SDS lysis buffer (1% SDS, 10mM EDTA) + 48 μ L of 6M GuHCl, + 7.25 μ L proteinase K (20 mg/mL, ThermoFisher, 4333793)
- 2- Add 75.25 μ L of patient swabs in transfer buffer
- 3- Incubate at room temperature for 10' and heated at 65°C for 10'
- 4- Add 145 μ L of RNAClean XP beads (Beckman, A66514)
- 5- Wash twice in 70% ethanol using a magnetic stand
- 6- Elute RNA into 30 μ L Resuspension buffer or Water RNase/DNase free

Reverse Transcription

- 1- Each reaction includes: 0.5 μ L Oligo-dT, 0.5 μ L hexamers, 4 μ L purified Total RNA, 1 μ L dNTP (2.5 mM each dATP, dGTP, dCTP and dTTP), *quantum satis* (qs) 13 μ L RNase/DNase free water.
- 2- Incubate samples at 65°C for 5', and then placed on ice for at least for 1'.
- 3- Add the following to each reaction: 4 μ L 5X First-Strand Buffer, 1 μ L 0.1 M DTT, 1 μ L Ribolock RNase Inhibitor, 1 μ L of SuperScript™ III RT (200 units/ μ L)
- 4- Mix by gently pipetting.
- 5- Incubate samples at:
 - 25°C for 5 minutes
 - 50°C for 60 minutes
 - 70°C for 15 minutes
 then stored at 4°C

Multiplex PCR

- 1- For each reaction (Total volume 25 μ L):

5 μ L	5X Phusion buffer – Thermo #F530L
0.5 μ L	dNTP mix (2.5mM each) – (FroggaBio #DN001025-5)
0.125 μ L	014- <i>S-PBS</i> For (10 μ M)
0.125 μ L	014- <i>S-PBS</i> Rev (10 μ M)
0.125 μ L	013- <i>S-RBD</i> For (10 μ M)
0.125 μ L	013- <i>S-RBD</i> Rev (10 μ M)
0.125 μ L	023- <i>RdRP</i> For (10 μ M)
0.125 μ L	023- <i>RdRP</i> Rev (10 μ M)
0.125 μ L	019- <i>ACTB/G</i> For (10 μ M)
0.125 μ L	019- <i>ACTB/G</i> Rev (10 μ M)
0.25 μ L	Phusion polymerase (50% glycerol, 2U/ μ L) – Thermo #F530L

2 µL cDNA
13.25 µL RNase/DNase free water

Primer sequences are described below:

013-S-RBD-For: acactctttccctacacgacgctcttccgatctATCAGGCCGGTAGCACACCT

013-S-RBD-Rev: gtgactggaggttcagacgctgtgctcttccgatctACTCTGTATGGTTGGTAACCAACAC

014-S-PBS-For: acactctttccctacacgacgctcttccgatctTATGCGCTAGTTATCAGACTCAGAC

014-S-PBS-Rev: gtgactggaggttcagacgctgtgctcttccgatctGTAAGCAACTGAATTTTCTGCACCA

023-RdRP-For: acactctttccctacacgacgctcttccgatctGATGCCACAACTGCTTATGC

023-RdRP-Rev: gtgactggaggttcagacgctgtgctcttccgatctTTGCGGACATACTTATCGGC

019-ACTB/G-For: acactctttccctacacgacgctcttccgatctTCACCATTGGCAATGAGCGGTTC

019-ACTB/G-Rev: gtgactggaggttcagacgctgtgctcttccgatctCCACGTCACACTTCATGATGGAG

2- The thermal cycling conditions are as follows:

98°C for 2 minutes

30 cycles: 98°C for 15 seconds, 60°C for 15 seconds, 72°C for 20 seconds

72°C for 5 minutes

4°C for ∞

Barcoding PCR

Unique reverse 8-nucleotide barcodes are used for each sample, while forward 8-based barcodes were used to mark each half (48) of the samples in 96-well plate to provide additional redundancy¹.

1- For each reaction (Total volume 20µL):

4μL 5X Phusion buffer – Thermo #F530L

0.4μL dNTP mix (2.5mM each) – (FroggaBio #DN001025-5)

2μL Barcode Primers For +Rev pre-mix Barcodes primers

0.2μL Phusion polymerase (50% glycerol, 2U/μl) – Thermo #F530L

4 μL cDNA

9.6μL RNase/DNase free water

2- The thermal cycling conditions are as follows:

98°C for 30 seconds

15 cycles: 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds

72°C for 5 minutes

4°C for ∞

Library preparation and Sequencing

1- For all libraries, pool 7 μL/sample and purify (2 rounds of 1:1; beads:library) library PCR products with SPRIselect beads (A66514, Beckman Coulter).

2- Assess Library quality with the 5200 Agilent Fragment Analyzer (Agilent), and quantify with Qubit 2.0 Fluorometer (ThermoFisher) (Note: If non-specific amplicon abundance is greater than specific amplicons after purification, re-purify as explained in 1- in this section). Run library quantification by qPCR using Colibri Library Quantification Kit (ThermoFisher) on a BioRad CFX96 Touch Real-Time PCR Detection System.

3- Quality checked library had 30% spike-in PhiX DNA (Illumina cat#FC-110-3001) mixed in prior to loading and then sequenced with MiSeq or NextSeq500 (illumina) following manufacturer's instructions using paired-end 75 read lengths and a typical target read depth of 100K reads per sample.

Analysis Pipeline

- 1- De-multiplexing: We used MiSeq Reporter v 2.6.2.3, and Illumina bcl2fastq v2.17 to demultiplex Illumina MiSeq sequencing output based on the unique combinations of the forward and reverse 8 nucleotide barcodes without any mismatch allowance.
- 2- Mapping: Full-length (75 base) forward and reverse reads were separately aligned to the expected amplicon sequence library using bowtie 2² with parameters `-best -v 3 -k 1 -m 1`. Read counts per amplicon were represented as reads per million or absolute read counts. The scripts for these steps are available at <https://github.com/UBrau/SPARpipe>³.
- 3- Filtering out low-input samples: Before assessing the viral content of the samples, we set a threshold for defining low RNA quality. For this, we computed precision-recall curves for classifying control samples into 'low amplification' and 'high amplification' based on reads mapped to RNA amplicons but ignoring mapping to genomic sequence, if applicable. The 'low amplification' group contained all negative controls (H2O controls) and the 'high amplification' group comprised HEK293T and synthetic SARS-CoV-2 RNA controls. For each run, we obtained the total mapped read threshold (including reads mapping to both human and viral amplicons) associated with the highest F1 score, representing the point with optimal balance of precision and recall. Samples with reads lower than this threshold were removed from subsequent steps due to insufficient total amplicon read count. The scripts related to this step can be found at <https://github.com/UBrau/ModelPerformance>⁴.
- 4- Determining positive and negative samples: We used viral read counts (total reads mapping to all three viral amplicons) from negative (H2O and HEK293T) and positive (synthetic SARS-CoV-2 RNA dilutions) internal controls for each run to calculate optimum cut-off for viral reads by PROC algorithm⁵, which defines the threshold for optimum PPV (positive predictive value) and NPV (negative predictive value) for diagnostic tests. Thus, a sample was labelled positive if it had viral reads above the viral read threshold; negative if it had viral reads below the viral read threshold and human reads above the mapped read threshold; and inconclusive if it had both viral and human reads below the respective thresholds.
- 5- Sample classification by heatmap clustering: We used the 'pheatmap' R package (<https://cran.r-project.org/web/packages/pheatmap>) to generate heatmap and hierarchical clustering of the samples. Viral and control amplicons $\log_{10}(\text{mapped reads}+1)$, were used to analyze and classify all samples

including patient samples, negative and positive controls. Samples that did not pass the RNA QC threshold (analysis steps 3-4) were excluded from the analysis.

6- Viral mutation assessment: To assess full length amplicons, paired end reads were stitched together by matching the last 12 nucleotides of read1 sequences to the reverse complement of read2 sequences. The number of full length reads per unique sequence variation were counted for each amplicon per sample by matching the 10 nucleotides from the 3' and 5' end of the sequence with gene-specific primers. (scripts are available at <https://github.com/seda-barutcu/FASTQstitch>⁶, and <https://github.com/seda-barutcu/MultiplexedPCR-DeepSequence-Analysis>⁷). The top enriched sequences (WT or variant) from each sample are then aligned to the reference sequence using CLUSTALW V2.1.

7- Non-specific amplicon assessment: Single-end reads that contain the first 10 nucleotides of the illumina adaptor sequence were counted and binned into relevant forward and reverse gene specific primer pools by matching the first 10 nt of the reads with primer sequences. Relative abundance of the non-specific amplicons was quantified as percentage of the reads corresponding to non-specific amplicon per forward or reverse primer (Scripts are available at <https://github.com/seda-barutcu/MultiplexedPCR-DeepSequence-Analysis>⁷).

8- Computational requirements for the demultiplexing step is 32 GB RAM and minimum 1GB network infrastructure, with a Linux operating system.

Code Availability

We provided the code for demultiplexing and mapping at <https://github.com/UBrau/SPARpipe>³, quality filtering at <https://github.com/UBrau/ModelPerformance>⁴, viral mutation assessment and non-specific amplicon assessment at <https://github.com/seda-barutcu/FASTQstitch>⁶ and <https://github.com/seda-barutcu/MultiplexedPCR-DeepSequence-Analysis>⁷.

Ethical declarations

Ethical regulation

Patient samples were obtained from the Department of Microbiology at Mount Sinai Hospital under MSH REB Study #20-0078-E, 'Use of known COVID-19 status tissue samples for development and validation of novel detection methodologies'. Patient samples were obtained as part of routine diagnostic testing.

Competing Interests statement

J. Wrana is founder and CEO of iTP Biomedica Inc, which employs whole transcriptome NGS tests in cancer, and he is founder and consultant for Fibrocor LP, which is developing therapeutics for fibrotic disease. The other authors declare no competing interests.

Troubleshooting

Time Taken

Anticipated Results

References

1. Han H, et al. Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Mol Cell* **65**, 539-553 e537 (2017).
2. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
3. Braunschweig U. A multiplexed next generation sequencing platform for high-throughput detection of SARS-CoV-2. <https://githubcom/UBrau/SPARpipe>, DOI: 10.5281/zenodo.4454116 (2021).
4. Braunschweig U. A multiplexed next generation sequencing platform for high-throughput detection of SARS-CoV-2. <https://githubcom/UBrau/ModelPerformance>, DOI: 10.5281/zenodo.4463831 (2021).
5. López Ratón M, Rodríguez Álvarez MX, Cadarso Suárez CM, Gude Sampedro F. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests.). American Statistical Association (2014).
6. Barutcu S. A multiplexed next generation sequencing Platform for High-Throughput Detection of SARS-CoV-2. <https://githubcom/seda-barutcu/FASTQstitch>, DOI: 10.5281/zenodo.4453811 (2021).
7. Barutcu S. A multiplexed next generation sequencing platform for high-throughput detection of SARS-CoV-2. <https://githubcom/seda-barutcu/MultiplexedPCR-DeepSequence-Analysis>, DOI: 10.5281/zenodo.4453805 (2021).

Acknowledgements

The authors thank Drs. Rita Kandel and Jim Woodgett for discussions. The authors are grateful to Tanja Durbic and Kyle Turner in the Donnelly Sequencing Center for sequencing samples and Kathy Fung at the Network Biology Collaboration Center.