A Massively Parallel COVID-19 Diagnostic Assay for Simultaneous Testing of 19200 Patient Samples

Ayaan Hossain[1], Alexander C. Reis[2], Sarthok Rahman[5], and Howard M. Salis[1-4 *]

[1]Bioinformatics and Genomics, [2]Department of Chemical Engineering, [3]Department of Biological Engineering, [4]Department of Biomedical Engineering, [5]Department of Biology, Pennsylvania State University, University Park, PA 16802, USA.

*Corresponding author: Howard M. Salis, salis@psu.edu

The COVID-19 pandemic requires advanced diagnostic assays that can be scaled to many thousands of patient samples. Here, we developed a massively parallel diagnostic assay (MPDA) that combines Reverse Transcriptase (RT), Polymerase Chain Reaction (PCR), and Next-generation Sequencing (NGS) for diagnostic COVID-19 testing on up to 19200 patient samples per workflow. To do this, we applied our Non-Repetitive Parts Calculator to computationally design 57600 high-performance RT primers with highly unique barcode sequences that link patient samples to cDNA products across 3 amplicons (N1 and N2 SARS-CoV-2, RNAseP human control). After the RT reaction, synthesized cDNA is pooled together and amplified in a single PCR, followed by next-generation sequencing of the barcoded amplicon library. Within a highly optimized computational pipeline, patient-specific barcodes and viral amplicon sequences are then mapped and counted, yielding verified read counts indicative of viral RNA levels. The proposed approach also includes internal spike-in controls for improved quantitation. Each workflow requires one PCR thermocycler and one next-generation sequencer, yielding positive/negative indications on up to 19200 patient samples. Our approach is highly synergistic with improved automation of viral RNA extraction from nasal swabs. Protocol instructions and primer sequences are released for immediate usage.

Google Doc: https://docs.google.com/document/d/1kP2w_uTMSep2UxTCOnUhh1TMCjWvHEY0sUUpkJHPYV4

Google Sheet: https://docs.google.com/spreadsheets/d/1y_Bf0Zz4FJRx53oSkX59u0kfouDOdrtr8E26M5NINFs/edit#gid=1974209521

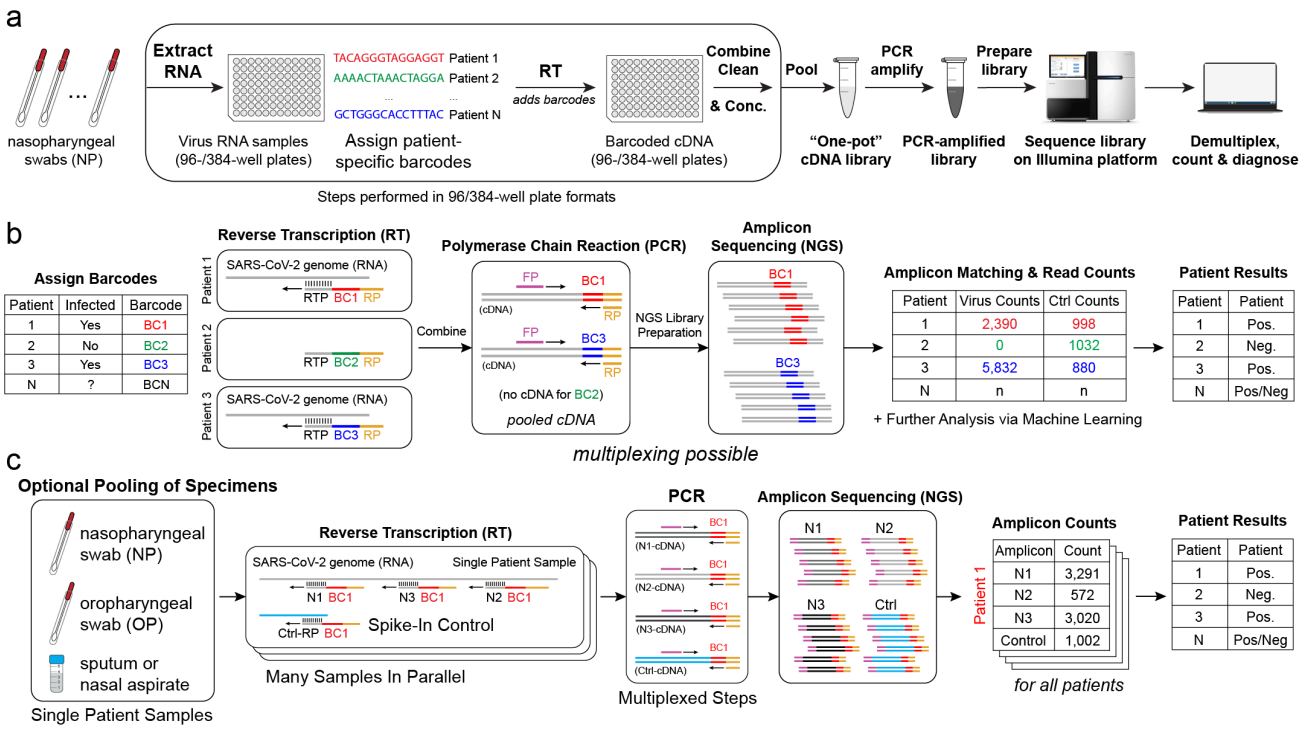

Figure 1: A massively parallel diagnostic assay for COVID-19. **(a)** A schematic showing the overall workflow, including viral RNA extraction, reverse transcription using barcoded transcript-specific primers, PCR amplification of the pooled cDNA synthesis products, and next-generation sequencing to quantify the amounts of barcoded cDNA synthesis products. **(b)** Barcodes are first assigned to patient samples, followed by Reverse Transcription and cDNA synthesis using the corresponding barcoded RT primer. The RT primer contains a constant region where the reverse primer binds. The following PCR amplification uses the same forward and reverse primer for all barcoded cDNA variants. After next-generation sequencing of the amplicon library, read sequences are compared against the expected barcode and amplicon sequences, and the numbers of each barcode and amplicon are counted. **(c)** The proposed approach is naturally extended to simultaneously measuring cDNA levels from multiple amplicons and spike-in controls. Amplicon matching enables re-use of the same patient-specific barcode to produce different cDNA products, followed by mapping and counting.

Introduction

The COVID-19 pandemic has revealed several deficiencies in our diagnostic capabilities, particularly when running the same test on thousands of patient samples. As of March 20th, the CDC has reported successful COVID-19 tests on only 4,524 and 49,681 patients across the U.S. in CDC and public health laboratories, respectively[1]. As a result of inadequate testing, it is likely that there are substantial number of infected persons who can unknowingly transmit COVID-19, amplifying the pandemic conditions[2]. Advanced diagnostic techniques with massively increased throughput are essential to limiting the transmission from asymptomatic persons and preventing the occurrence of follow-on waves of infection after suppression social tactics have been suspended.

The current approach to testing for SARS-CoV-2 – the virus that causes COVID-19 – consists of a widely used assay that combines Reverse Transcriptase with Quantitative Polymerase Chain Reaction (RT-qPCR) to measure viral RNA levels from patient nasal swab samples. The throughput for these assays is limited by several factors, including the rate of viral RNA extraction and the availability of qPCR thermocyclers. For example, a

single ABI 7500 DX in Fast mode can measure SARS-CoV-2 RNA levels from only 480 samples per day, yielding diagnoses for about 120 patients per day when including multiple sites and controls. Simply scaling up RT-qPCR assays to many thousands of patient samples per day remains extremely difficult; supplies of these qPCR thermocyclers are limited and the technical labor needed to handle plates – even with automated liquid handlers – does not adequately scale to regularly testing millions of patients.

Instead, it is now possible to develop Massively Parallel Diagnostic Assays (MPDA) that leverage next-generation sequencing to simultaneously carry out thousands of viral RNA measurements in a single assay. Notably, next-generation sequencers are now widely available across the world, including inside hospitals, clinics, and commercial service providers. Here, we propose a MPDA to detect COVID-19 viral RNA in up to 19200 patient samples per workflow, combining viral RNA extractions, Reverse Transcriptase (RT) using patient-specific primers with unique barcodes, a single pooled Polymerase Chain Reaction (PCR), followed by Next-generation Sequencing (NGS) (**Figure 1A**). The key novelty of our approach is the unique labeling of patient samples at the RT step, the pooling of thousands of synthesized cDNA samples into a single PCR amplification step, the application of next-generation sequencing in amplicon mode to measure thousands of variant cDNA levels, and the rapid demultiplexing of the NGS data to deliver thousands of diagnoses with patient-specific internal controls (**Figure 1B**). Our approach naturally extended to including multiple spike-in controls and RNA extraction controls to yield a more accurate and quantitative diagnosis (**Figure 1C**).

Importantly, considerable computational design of primer sequences was essential to ensuring that the pooled PCR amplification is equally efficient across all patient-specific cDNA variants and that all patient-specific cDNA levels are uniquely identifiable by NGS amplicon sequencing. As the first step, we applied our newly developed algorithm, the Non-Repetitive Parts Calculator, as part of a computational pipeline to design the 57600 patient-specific RT primer sequences, the 3 pairs of pooled PCR primer sequences, and the 3 spike-in control DNA fragment sequences (**Methods**). We are publicly releasing these sequences and calculations for immediate usage (**Supplementary Data**).

In the following, we describe a preliminary protocol that illustrates how to carry out the proposed Massively Parallel Diagnostic Assay for COVID-19 detection. Notably, the protocol is highly similar to existing amplicon NGS workflows, which are commonly performed. However, the proposed protocol still relies on viral RNA extraction from patient samples as the first step, which is similar to the current testing regime. We anticipate that – through automated handling of nasal swabs and automated liquid handling of RT reagents in 384-well plate format – it is possible to greatly expand the throughput of viral RNA extraction, feeding into our MPDA approach and matching its throughput.

Results

A Massively Parallel Diagnostic Assay for SARS-CoV-2

Preliminary Steps for Preparing High-Throughput RT in Multi-well Format:

1. Chemically synthesize up to 57600 patient-specific RT primers in 96-well or 384-well format for the 3 amplicons (2 coronavirus, 1 human control), according to the oligonucleotide synthesis provider's specifications. The **Supplementary Data** also lists sets of 288 (3 x 96) and 2880 (3 x 960) patient-specific RT primers, which can be used as part of the scale-up process. Each well will contain 3 barcoded RT primers (3 amplicons), each at a stock concentration of 10 µM. This can be achieved by adding 33.3 µL of 30 µM primer stock solutions to each well. Primer stock solutions can be formulated by oligonucleotide service providers at the desired concentration of 30 µM (this service can be carried out by the service provider by request). Keep cold. [**RTP solution plates**]
2. Prepare a reservoir containing multiples of 400 µL of 25 mM $MgCl_2$, 200 µL of 10 mM dNTP mixture, and 200 µL of 10X Reverse Transcription Buffer. Keep cold. [**RT Master Mix reservoir**]
3. Prepare a reservoir containing multiples of 50 µL of recombinant ribonuclease inhibitor (e.g. RNasin, 20 to 40 units per µL) and 150 µL of AMV Reverse Transcriptase (10 units per µL). Keep cold. [**RT Enzyme Mix**]
4. Prepare a reservoir containing multiples of 100 µL of 10 nM spike-in control DNA fragment N1, 100 µL of 10 nM spike-in control DNA fragment N2, and 100 µL of 10 nM spike-in control DNA fragment RNAseP. This can be accomplished by adding 33.3 µL of 30 nM DNA fragment N1, 33.3 µL of 30 nM DNA fragment N2, and 33.3 µL of 30 nM DNA fragment RnaseP together. [**Control Fragment Mix**]

Optimization Note: The final concentrations of the spike-in control DNA fragments in the RT reaction should be similar to the expected viral RNA concentration in a positive diagnosis scenario. This protocol uses a final concentration of 10 fmoles per 20 uL RT reaction (0.50 nM) for the spike-in DNA fragment controls.

Step 1: Carry out Viral RNA Extraction on Patient Sample (e.g. nasal swab)

Follow the manufacturer's protocol, e.g. using the Qiagen viral RNA extraction kit, to produce a **Viral RNA solution**. The proposed approach is agnostic towards the viral RNA extraction protocol.

We anticipate the development of high throughput viral RNA extraction protocols leveraging a many-well plate format, automated liquid handlers, and vacuum manifolds.

Optimization Note: According to Bruce et. al. "RT-qPCR DETECTION OF SARS-CoV-2 RNA FROM PATIENT NASOPHARYNGEAL SWAB USING QIAGEN RNEASY KITS OR DIRECTLY VIA OMISSION OF AN RNA EXTRACTION STEP" bioRxiv *submitted* [https://www.biorxiv.org/content/10.1101/2020.03.20.001008v1], it may be possible to omit the viral RNA extraction step and proceed directly with Reverse Transcription. Based on

preliminary data using RT-qPCR comparisons, the deltaCt is reduced by about 4, compared to using the Qiagen QIAamp Viral RNA extraction kit.

Alt Step 1:

1A. Add 2 mL of M6 transport media (Remel) to nasopharyngeal swab container. Incubate during sample transit to laboratory.

Step 2: Carry out Reverse Transcription using Patient-Specific RT Primers.

2A. Into each plate well, dispense 8 µL of **RT Master Mix reservoir**, 2 µL of **RT Enzyme Mix**, 1 µL of the corresponding well from the **RTP solution plates**, 8 µL of **Viral RNA solution**, and 1 µL of **Control Fragment Mix**. (in this order)

2B. Incubate the plate for 15 to 60 minutes at 42°C to carry out first-strand cDNA synthesis.

2C. Into each well, add 1 µL of Exonuclease I (e.g. NEB M0293L or a thermolabile Exonuclease I, NEB M0568L) to degrade leftover RT primer. Incubate at 37°C for 30 minutes.

2D. Incubate the plate for 5 minutes at 85°C to deactivate enzymes.

Step 3: Combine, Clean, and Concentrate the RT Product

3A. Transfer 20 µL from each plate well in Step 2 into the same container for a total combined volume of 1.92 mL.

3B. Add 9.6 mL of DNA binding buffer (e.g. Zymo D4004-1-L). Mix for 1 minute.

3C. Transfer mixture to a high capacity spin column (e.g. Zymo C1013-20). Place into 50 mL conical tube for waste. Centrifuge for 5 to 10 minutes at 3000 x g. Discard the flow through.

3D. Add 10 mL of diluted DNA Wash Buffer (e.g. from stock solution of Zymo D4003-2-48) to high capacity spin column. Centrifuge for 5 to 10 minutes at 3000 x g. Discard flow through. Repeat again if desired.

3E. Transfer high capacity spin column to a new 50 mL conical tube. Add 2 mL DNA elution buffer (e.g. Zymo D3004-4-50) or ultra-pure water to column. Incubate for 1 minute at room temperature. Centrifuge for 5 minutes at 3000 x g. The eluted volume is your **pooled cDNA product**.

Optimization Note: **Steps 3B to 3E** may not be necessary if only 10% of the combined RT product (e.g. 0.192 mL from 1.92 mL) is used in **Step 4** for PCR amplification.

Optimization Note: Step 3 can be carried out with a lower capacity spin column with smaller elution volumes. Optimal spin column capacity and minimally possible elution volume will depend on the cDNA synthesis amounts and the number of pooled samples.

Step 4: One-Pot PCR Amplification of Pooled Patient-derived cDNA Variants

This protocol uses Hot-start Q5 High-fidelity DNA polymerase. Adjust as needed.

4A. In one PCR tube, combine 10 µL of Q5 5X Reaction Buffer, 1 µL of 10 mM dNTPs, 1 µL of 10 µM Forward Primer **FP1**, 1 µL of 10 µM Forward Primer **FP2**, 1 µL of 10 µM Forward Primer **FP3**, 3 µL of 10 µM Reverse Primer **RP**, 32.5 µL of **pooled cDNA product**, and 0.5 µL of Q5 Hot Start High-fidelity DNA polymerase (in order)

4B. Place the PCR tube into a PCR thermocycler and carry out the program:

Initial Denaturation at 98°C for 10 minutes.
35 Cycles of the following steps:
Denaturation: 98°C for 5 seconds.
Annealing: 53°C for 15 seconds.
Extension: 72°C for 30 seconds.
Final Extension at 72°C for 2 minutes.
Hold at 4°C.

4C. Remove the **Pooled PCR product**. Purify it using clean-and-concentrate spin-column protocol, which also removes leftover primer.

The melting temperatures of **FP1**, **FP2**, **FP3**, and **RP** are 52.5°C, 52.7°C, 53.7°C, and 51.9°C, calculated using a thermodynamic nearest neighbor model with Sugimoto 1996 parameters.

Optimization Note: Size-selection of **Pooled PCR product** may improve amplicon NGS. Consider using gel electrophoresis to extract the cDNA amplicons at lengths **64 to 146 bp**.

Step 5: Next-generation Amplicon Sequencing

We recommend usage of a chemistry and platform for amplicon sequencing that yields at least 200 million reads per sample. Amplicon lengths are at most 100 bp. Read lengths must be at least 50 bp to capture both the patient-specific barcode and a sufficiently large portion of the amplicon region to achieve unique mapping. Paired-end 50 bp reads are recommended for improved amplicon matching. These requirements are satisfied by Illumina platform sequencers, including the HiSeq 3000/4000, NextSeq 550, and NextSeq 2000.

For library preparation, follow the manufacturer's protocol to ligate the provided adapters to the **Pooled PCR product**. At this stage, it is possible to further scale the proposed approach by utilizing adapters with unique indexes (barcodes). For example, when using a modern sequencing that yields 400 million reads per lane, it is possible to utilize four adapter indexes and maintain a high depth of about 10000 reads per patient-derived cDNA, yielding diagnoses for 38400 patients from a single lane. After library preparation, QA/QC is performed using a Bioanalyzer and TapeStation to verify amplicon sizes.

For Illumina platforms, amplicon sequencing requires the addition of 30% PhiX genomic DNA to library samples to prevent color-dye imbalances during image recognition. This will reduce the total number of mapped reads by 30%. The augmented library sample is then loaded into a lane of a flow cell, which is then loaded into the sequencer. The desired sequencing program is then initiated. Sequencing run times depend on read length; for example, a NextSeq 2000 requires about 13 hours to deliver 400 million paired-end 50 bp reads.

Step 6: Demultiplexing, Amplicon Mapping, and Barcode Counting

We've previously developed a computational pipeline that carries out demultiplexing, mapping, and counting of reads with expected barcode and/or amplicon sequences. The pipeline accepts FASTQ files containing single-end or merged paired-end reads along with the lists of expected reference sequences. The mapping process uses highly optimized data structures and a novel approach called read-packing to process about 75000 reads per second per core. Analysis requires about 10 minutes per sequencing lane on an 8-core system. We anticipate that the analysis step will not be a rate-limiting step in the diagnostic process. However, it is beneficial if there is an HPC system with a dedicated network connection to the next-generation sequencer to maximize data throughput and minimize file transfer times. **Python source code is available on request. To the many computational bioinformaticians who have asked, we will add the code to a GitHub repo soon. But any amplicon sequencing pipeline will do. We developed our own approach, using some fancy data structures to make it faster, but it is *not* the rate-limiting step to diagnostic testing!**

Additional analysis of our RT primer sets and methodological details will become available in the coming days. The purpose of this report is to immediately facilitate massively parallel diagnostic assays.

References

1. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/testing-in-us.html. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/testing-in-us.html.
2. Li, R.; Pei, S.; Chen, B.; Song, Y.; Zhang, T.; Yang, W.; Shaman, J., Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **2020**, eabb3221.