REMBRANDT: A High-Throughput Barcoded Sequencing Approach for COVID-19 Screening.

Dario Palmieri
The Ohio State University
Amanda Javorina
United States Air Force
Jalal Siddiquii
The Ohio State University
Anne Gardner
The Ohio State University
Anthony Fries
United States Air Force
Richard Chapleau
United States Air Force
Clarise Starr
United States Air Force
Richard Fishel
The Ohio State University
wayne Miles (wayne.miles@osumc.edu)
The Ohio State University

Research Article

Keywords: coronavirus, sars, infectious, patients,covid, acute, respiratory, symptoms

Posted Date: July 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-630472/v1

REMBRANDT: A high-throughput barcoded sequencing approach for COVID-19 screening.

Dario Palmieri[1,*,#], Amanda Javorina[2,#], Jalal Siddiqui[1], Anne Gardner[1], Anthony Fries[2], Richard R Chapleau[2], Clarise Starr[2], Richard Fishel[1,*] and Wayne O Miles[1,*]

[1]Department of Cancer Biology and Genetics, College of Medicine and Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA.

[2]Public Health and Preventive Medicine Department, US Air Force School of Aerospace Medicine, Wright-Patterson Air Force Base, OH 45433, USA

[#]Equal contributing authors

[*]Corresponding authors: DP: Dario.Palmieri@osumc.edu, RF: rfishel@osu.edu and WOM: wayne.miles@osumc.edu

Abstract: The Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), also known as 2019 novel coronavirus (2019-nCoV), is a highly infectious RNA virus. A percentage of patients develop coronavirus disease 2019 (COVID-19) after infection, whose symptoms include fever, cough, shortness of breath and fatigue. Acute and life-threatening respiratory symptoms are experienced by 10-20% of symptomatic patients, particularly those with underlying medical conditions. One of the main challenges in the containment of COVID-19 is the identification and isolation of asymptomatic/pre-symptomatic individuals. A number of molecular assays are currently used to detect SARS-CoV-2. Many of them can accurately test hundreds or even thousands of patients every day. However, there are presently no testing platforms that enable more than 10,000 tests per day. Here, we describe the foundation for the REcombinase Mediated BaRcoding and AmplificatioN Diagnostic Tool (REMBRANDT), a high-throughput Next Generation Sequencing-based approach for the simultaneous screening of over 100,000 samples per day. The REMBRANDT protocol includes direct two-barcoded amplification of SARS-CoV-2 and control amplicons using an isothermal reaction, and the downstream library preparation for Illumina sequencing and bioinformatics analysis. This protocol represents a potentially powerful approach for community screening of COVID 19 that may be modified for application to any infectious or non-infectious genome.

Introduction

COVID-19 (coronavirus disease 2019) is an infectious disease whose etiopathogenic agent is the Severely Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) RNA virus[1-4]. As of April 2021, this viral infection has globally affected more than 130 million individuals and claimed almost 3 million lives, according to the World Health Organization (https://covid19.who.int/). The enormous volume of COVID-19 patients has placed significant strain on healthcare systems around the world and led to the confinement of over half of the world's population. Accommodating state- and country-wide populations resuming pre-pandemic activities will require serial monitoring of 1-100 million residents. To address this screening need, a scalable diagnostic test is required that may be used with widely available and existing equipment and infrastructure. In addition, as the proportion of vaccinated individuals grow, COVID-19 testing will need to be incorporated into panel clinical testing rather than the specialized PCR based assay used effectively during large scale pandemic testing.

Here, we describe, and provide reduction-to-practice data, for a mass COVID-19 screening platform: Recombinase Mediated BaRcoding/AmplificatioN Diagnostic Tool (REMBRANDT). This protocol contains a number of key advances over PCR-based and other parallel LAMP and PCR based approaches that maximize output and sensitivity whilst retaining speed and efficiency. Specifically, REMBRANDT uses recombination and repair enzymes to detect and amplify the viral genomic RNA[5]. By incorporating dual barcoded primers during this process, the REMBRANDT platform enables the amplification of simultaneously tagged individual samples. This step allows the rapid generation of DNA products directly from SARS-CoV-2 in the sample, bypassing the need for first extracting RNA. Utilizing two independent barcoded primers per well enables a single patient sample within a 96-well plate to be independently marked with a well-specific barcode *and* a plate-specific barcode. Once barcoded, the patient samples from

multiple 96-well plates can be pooled and purified, ultimately minimizing reagent usage, time and sample-to-sample variation.

Processing these samples together enables rapid library construction using any one of the barcoded Illumina kits. Utilizing Illumina barcodes, allows further sample multiplexing to maximize patient numbers that can be screened using Next Generation (NextGen) sequencing. The smaller numbers of barcodes require less time-consuming computational processing, as trimmed sequences can be quickly divided based on barcode, and mapped to the SARS-CoV-2 N gene. Importantly, this protocol introduces a unique, synthetic SARS-CoV-2 N gene sequence into every 96-well plate. This control has identical primer annealing regions to the SARS-CoV-2 gene but contains 6 engineered base pair substitutions that distinguishes it from native sequence; providing an internal quality control for the process and a measure of batch effects. The laboratory and computational framework are designed to maximize SARS-CoV-2 detection efficiency while minimizing reagent usage, processing and turn-around time.

Results

The REMBRANDT platform and *in vitro* substrate testing.

To evaluate this strategy for SARS-CoV-2 testing, we first tested whether this approach could identify synthetic SARS-CoV-2 sequences. For these assays, we generated T7-flanked PCR products of the SARS-CoV-2 N gene. These were used to make and purify RNA for pilot experiments (Figure 1A). These RNA fragments were then used in an isothermal amplification reaction using UvsX and UvsY (Figure 1B) followed by strand invasion and second strand synthesis mediated by GP32 and BSU DNA polymerase (Figure 1B). This highly efficient reaction catalyzes the rapid synthesis of dual barcoded SARS-CoV-2 regions (Figure 1B). To test whether this approach works on *in vitro* generated RNA, we measured the isothermal reaction and DNA production from forward and reverse primers of SARS-CoV-2 that contain different barcodes. As

shown in Figure 1C, each of the primer pairs tested generated a product of the correct size (Figure 1C: numbered), however the negative control sample for each only produced non-specific bands. Using this approach, we then generated a computational pipeline that would allow the rapid identification of SARS-CoV-2 positive patients from NextGen sequencing (Figure 1D).

REMBRANDT is able to detect COVID-19 in patient samples

Following these proof-of-principle *in vitro*-based assays, we next evaluated the capacity of REMBRANDT to identify SARS-CoV-2 patients. These experiments were conducted in a Clinical Laboratory Improvement Amendment (CLIA) certified laboratory to accurately mirror current SARS-CoV-2 screening conditions. For this analysis, we utilized two negative controls (RNAse P and non-templated) and the T7 SARS-CoV-2 N RNA as a positive control. To test the efficiency of the REMBRANDT pipeline to detect SARS-CoV-2 positive patients, we profiled 6 patient samples, 2 negative and 4 positive samples (Figure 2A), from Raw and RNA extracted specimens. The clinical diagnosis of each patient was blinded for the computational analysis. Following the REMBRANDT reaction, library construction and RNA-sequencing on the Illumina MiSeq, we utilized the REMBRANDT barcode calling pipeline to separate patient specific barcodes and mapped the number of reads found to SARS-CoV-2 amplicon. Using this approach, REMBRANDT can clearly segregate SARS-CoV-2 positive patients from controls in Raw (Figure 2B) and RNA extracted (Figure 2C) samples. To evaluate the necessity for RNA extraction in the correct identification of patients, we compared SARS-CoV-2 read depth and quality and found little difference between inputs (Figure 2D). These findings suggest that raw (unextracted) patient samples can be utilized in the REMBRANDT approach, which speeds up processing and overcomes existing reagent bottlenecks.

Discussion

SARS-CoV-2 is a highly infectious single-stranded RNA virus. However, increasing evidence suggests that the vast majority of infected individuals display few or very mild symptoms[6]. These people may still spread the virus for over 10 days after the initial contagion[7]. Even as the number of people vaccinated against SARS-CoV-2 grows, the rapid identification and quarantining of infected individuals will remain a key measure to containing the spread of SARS-CoV-2. One of the major current limitations of population testing is the availability of reagents. The current clinical standard for COVID-19 diagnosis (qRT-PCR) requires suitable equipment for the amplification of the viral RNA and the detection of the infection. Each of these steps significantly slows sample processing and limits the number of tests that can be performed in a day. For this reason, new diagnostic tools are needed that are fast and efficient as well as scalable to population sized numbers and based on readily available reagents.

To address this need, the scientific community has delivered a remarkable and unprecedented number of assays to diagnose COVID-19[8, 9-10]. One particularly promising approach uses viral RNA reverse transcription and patient-specific barcoding of the single strand of cDNA, followed by cDNA amplification and NGS analysis[8]. This approach, although interesting, requires 1 barcode per patient rather than multiplexing barcodes. That means to screen 10,000 patients one requires 10,000 barcodes. Furthermore, these large primers with large barcodes are likely to vary significantly in their amplification efficiency, making the individual testing of these primer sets essential. To overcome the issues related with the high number of barcodes and potentially limited sensitivity, Schmid-Burg and colleagues developed a highly-scalable Reverse-Transcription Loop-mediated Isothermal Amplification (RT-LAMP) method for population sequencing[10]. However, this approach relies on the use of 3 proprietary enzymes, whose availability might limit its scalability.

REcombinase Mediated BaRcoding and AmplificatioN Diagnostic Tool (REMBRANDT) builds on these principles to double barcode each patient sample using an isothermal RT-RPA reaction. The combinatorial use of multiple forward and reverse barcodes, one per patient and one per plate, enables 192 primers to generate 9,216 patient-specific combinations. This number can then be further multiplexed and amplified with 12 Illumina barcodes utilized during library construction. This simplified and straightforward approach does not require specialized equipment for patient detection and library construction. We therefore have confidence that this system could be readily implemented in most communities, including those with limited resources, provided the availability of partners able to perform NGS analysis and computational analysis (Figure 2E). As REMBRANDT uses an isothermal RNA reverse transcription and amplification reaction, it does not require PCR amplification. Moreover, since pairing of the template and primer during the amplification step relies on the activity of recombinases UvsX and UvsY, it is minimally affected by different T_m or T_a of the primers. The REMBRANDT pipeline also offers flexibility and can be readily adapted to detect other viral genes and/or species, by switching the amplification regions of the primers. Based on our laboratory testing on raw and extracted patient samples, our results show that the REMBRANDT approach can accurate amplify and identify SARS-CoV-2 positive patients.

Methods

In vitro RNA transcription: To generate positive control RNAs, T7-flanked forward primers and standard reverse primers (Supplementary Information) were used in a PCR reaction to apply a region of each target gene with Q5 polymerase. Templates used in this work were COVID-19-N (10006625, IDT) and RNPase P (10006626, IDT). Gel purified PCR products (200ng) are purified and used in T7 (Roche 10881775001) *in vitro* transcription reaction. DNase is then added for 15mins, before the RNA is precipitated using standard ethanol RNA purification method (8).

Oligo Plate Preparation: Forward barcoding primers (plate-specific barcode) and barcoding primers (well-specific barcode) were dissolved at a 10 µM concentration in RNase-free water. Aliquot 1 µl of every well-specific reverse primer into each well of a 96-well plate then aliquot 1 µl of a single plate specific forward barcoding primer into each well.

Primers for REMBRANDT

Forward primer	Sequence
COVID N FOR, BC 1	TGT AAA ACG GCC AGT TAT CTG TGG ACC CCA AAA TCA GCG AAA TGC ACC CCG
COVID N FOR, BC 2	TGT AAA ACG GCC AGT GCC GAA TGG ACC CCA AAA TCA GCG AAA TGC ACC CCG
COVID N FOR, BC 3	TGT AAA ACG GCC AGT TAC TGC AGG ACC CCA AAA TCA GCG AAA TGC ACC CCG

Reverse Primer	Sequence
COVID N REV BC1	CAG GAA ACA GCT ATG ACC TTA GTG GGC GTT CTC CAT TCT GGT TAC TGC CAG TTG
COVID N REV BC2	CAG GAA ACA GCT ATG ACC GCA TAG TGC GTT CTC CAT TCT GGT TAC TGC CAG TTG
COVID N REV BC3	CAG GAA ACA GCT ATG ACG AAG CGA TGC GTT CTC CAT TCT GGT TAC TGC CAG TTG

REMBRANDT Steps: Prepare Isothermal Amplification Buffer 2X (IAB2X) (Table 1). Keep the 2X IAB2X mix on ice. Prepare the UvsX/UvsY mix by combining 10 ml of recombinant UvsX (5 mg/ml) and 10 ml of recombinant UvsY (2mg/ml). Add 2 µl of the resulting mix to each well of the 96 well

plates. We recommend gentle trituration 2-3 times of the UvsX/UvsY mix with the primers. This premixed plate should be kept on ice for all subsequent additions. Add 1 µl of template RNA to each well of the 96-well plates. Prepare the Enzyme Mix (Table 2). Keep the mix on ice. Proceed immediately to step. Dispense 14 µl of Enzyme Mix into the all wells of the of the 96-well plates. Keep on ice. Add 1 µl/well of 280 mM MgOAc to start the reaction. Place all the 96-well plates at 38^{o} C for 2h. Collect the REMBRANDT products from all the 96-well plates into a single container (each 96-well plate will yield 1.92 ml). Add 5 volumes (9.6 ml) of DNA Binding Buffer. Mix well and load onto a Zymo-Spin VI column. Place the column(s) on a vacuum manifold, turn on the vacuum source and let the sample clear from the column (*repeat this step up to 5 times*). Add 5 ml of DNA wash Buffer. Repeat the wash step. Leave the vacuum source on for additional 5 minutes to remove all the wash buffer. Transfer the column into a 50 ml conical tube. Add 2 ml of water to the column and wait 1 minute. Centrifuge at 3,000 x g for 3 minutes and collect the eluted barcoded DNA. If performing multiple plates/columns, combine the eluted barcoded DNA and label the combination as “Batch 1”. Subsequent Batches are individually used for library preparation, and SHOULD NOT be combined until AFTER LIGATION (Library construction) with batch specific adapters.

Table 1: Isothermal Amplification Buffer 2X (IAB2X)

Reagent	Stock concentration	Final Concentration	Required volume 96 samples[#]
Tris Acetate pH 7.8	1 M	80 mM	92.8 µl
K-acetate	5 M	200 mM	46.4 µl
DTT	1 M	10 mM	11.6 µl
ATP	100 mM	5 mM	58 µl
dNTPs	10 mM	480 µM	55.68 µl

PEG-3500	44%	11 %	290 µl
Trehalose	40%	10 %	290 µl
Phosphocreatine	0.5 M	50 mM	116 µl
Creatine Kinase	2 µg/µl	200 ng/µl	116 µl
Acetylated BSA	10mg/ml	200 µg/ml	23.2 µl
H2O	/	/	60.32 µl
Final volume			1.16 ml

Table 2: Enzyme Mix

Reagent	Stock concentration	Final Concentration	Required volume 96 samples#
T4 Gene 32 Protein	10 mg/ml	0.25 mg/ml	58 µl
Bsu DNA Polymerase	5000 U/ml	0.25 U/µl	116 µl
RevertAid Reverse Transcriptase	200 U/µl	10 U/µl	116 µl
SUPERase RNase Inhibitor	20 U/µl	1 U/µl	116 µl
IAB2X	2X	1X	1.16 ml
Water##			58 µl
Total volume			1.624 ml

Library construction: Before library construction, a fraction of each purified pool of barcoded products should be examined using the Bioanalyzer to determine concentration and product size. The respective size of the COVID-19 product is 124bp. Using the Illumina library construction kit (e7490), start at protocol 1.6 and conduct End repair on each pool of DNA fragments. Proceed to

Adaptor Ligation Step. Proceed to Purification of Ligation of Reaction step. This is critical to remove any free adaptor that will add noise to the library construction. PCR Enrichment of Ligation Reaction. We recommend using 8 cycles to minimize over-amplification. Purification of PCR reaction. The overall library quality should be checked using a Bioanalyzer to confirm the size of the DNA fragments. Libraries that contain detectable levels of Primer-Primer annealing events should be repeated. Determine the absolute DNA concentration of each library using Qubit.

Sequencing: The libraries are then loaded onto the Illumina sequencing platform and run as per manufacturer's instructions. Once complete, the dataset is downloaded onto a suitable storage server for analysis.

Bioinformatics: De-multiplexing reads: Create CSV files listing of all 3 sets of barcodes (Illumina, plate, well). Create CSV file listing combinations of barcode matched to a patient sample. Build connection using the Fastq streamer function for file. Loop over all chunks of reads, extract a chunk of reads using yield function and loop over set of barcodes. Use *grepl* search for barcode, and if barcode present, write file to 'fastq' file. End loop over barcodes and end loop over all reads after all chunks are extracted. Perform multiplexing iteration on initial file using Illumina barcodes followed by multiplexing Iteration on resulting fastq files from Illumina barcode de-multiplexing using plate-specific barcodes. Perform multiplexing interaction on resulting files from de-multiplexing plate-specific barcodes using well-specific barcodes and then match files to patient samples and perform mapping via Rsubread, see below.

Build index for mapping reads using Rsubread: Create reference fasta file from nCoV, and SARS-CoV-2 sequences. Use buildIndex function to construct index, ref being location of fasta file: buildindex(basename="./reference_index",reference=ref). Construct SAF annotation file for RSubread alignment and featureCounts formatted: GeneID Chr Start End Strand. Map de-

multiplexed reads to reference index using align() function: align.stat <- align(index = "./reference_index", readfile1 = reads1, readfile2 = reads2, output_file = "./Rsubread_alignment.BAM"). Obtain counts using *featureCounts(),* featureCounts function on resulting BAM files with SAF annotation file.

Additional information

The authors have no competing interests.

Author contributions

DP, RF and WOM conceived the project. DP, AJ, AG and WOM conducted the experiments. JS, AF, RRC and CS ran the sequencing and computational analysis of the data. DP, RF and WOM wrote the manuscript.

Acknowledgments

This work was supported by The Ohio State University Comprehensive Cancer Center funds (RF). The authors thank Dr. Kristine Yoder, Laura Miles and Anna Tessari for their critical support. The full detailed protocol, including primer sequences and barcodes is available at https://www.biorxiv.org/content/10.1101/2020.05.16.099747v1. Figures were prepared using Biorender (biorender.com).

Figure Legends

Figure 1: *In vitro* REMBRANDT testing.

(A) Generation of *in vitro* transcribed control and SARS-CoV-2 RNA. **(B)** Schematic detailing the protocol utilized for the isothermal amplification and dual-barcoding of target RNAs. **(C)** Agarose gels using 8 different forward and 12 different reverse barcoded primers for the isothermal amplification of the SARS-CoV-2 N gene. Numbered lanes represent unique barcode

combinations; – represents water control for each primer pair. **(D)** Schematic detailing the protocol utilized for the bioinformatics analysis of the REMBRANDT RNA-seq data.

Figure 2: REMBRANDT can identify SARS-CoV-2 positive patient samples from raw and extracted patient samples. (A) Patient information from tested individuals including age, gender, days of test following symptoms and COVID and RNaseP values from CDC COVID assay. **(B)** Number of SARS-CoV-2 reads mapped to negative control, negative patients and SARS-CoV-2 positive patients from raw samples. **(C)** Number of SARS-CoV-2 reads mapped to negative control, negative patients and SARS-CoV-2 positive patients from RNA extracted samples. **(D)** Comparison of the number of SARS-CoV-2 reads from each patient comparing raw and extracted patient sample. **(E)** Schematically detailing the protocol utilized for the REMBRANDT pipeline for testing for COVID-19.

References

1 Yuki, K., Fujiogi, M. & Koutsogiannaki, S. COVID-19 pathophysiology: A review. *Clin Immunol* **215**, 108427, doi:10.1016/j.clim.2020.108427 (2020).

2 Guan, W. J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* **382**, 1708-1720, doi:10.1056/NEJMoa2002032 (2020).

3 Gandhi, R. T., Lynch, J. B. & Del Rio, C. Mild or Moderate Covid-19. *N Engl J Med* **383**, 1757-1766, doi:10.1056/NEJMcp2009249 (2020).

4 Berlin, D. A., Gulick, R. M. & Martinez, F. J. Severe Covid-19. *N Engl J Med* **383**, 2451-2460, doi:10.1056/NEJMcp2009575 (2020).

5 Piepenburg, O., Williams, C. H., Stemple, D. L. & Armes, N. A. DNA detection using recombination proteins. *PLoS Biol* **4**, e204, doi:10.1371/journal.pbio.0040204 (2006).

6 Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489-493, doi:10.1126/science.abb3221 (2020).

7 He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* **26**, 672-675, doi:10.1038/s41591-020-0869-5 (2020).

8 Long, Q. X. *et al.* Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat Med* **26**, 845-848, doi:10.1038/s41591-020-0897-1 (2020).

9. A Massively Parallel COVID-19 Diagnostic Assay for Simultaneous Testing of 19200 Patient Samples - Google Docs. https://docs.google.com/document/d/1kP2w_uTMSep2UxTCOnUhh1TMCjWvHEY0sUUpkJHPYV4/edit.

10. Schmid-Burgk, J. L. *et al.* LAMP-Seq: Population-Scale COVID-19 Diagnostics Using Combinatorial Barcoding. doi:10.1101/2020.04.06.025635.

Figures

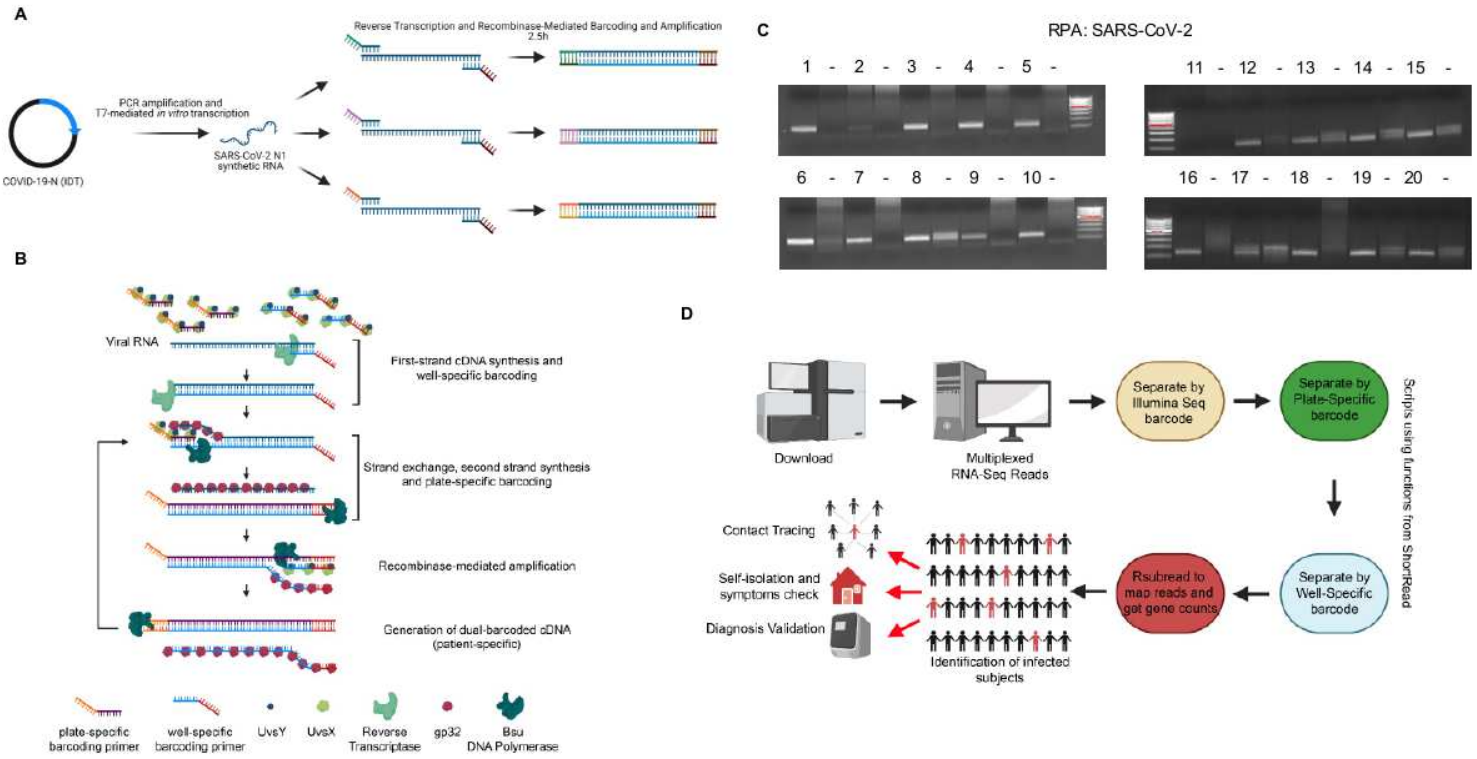
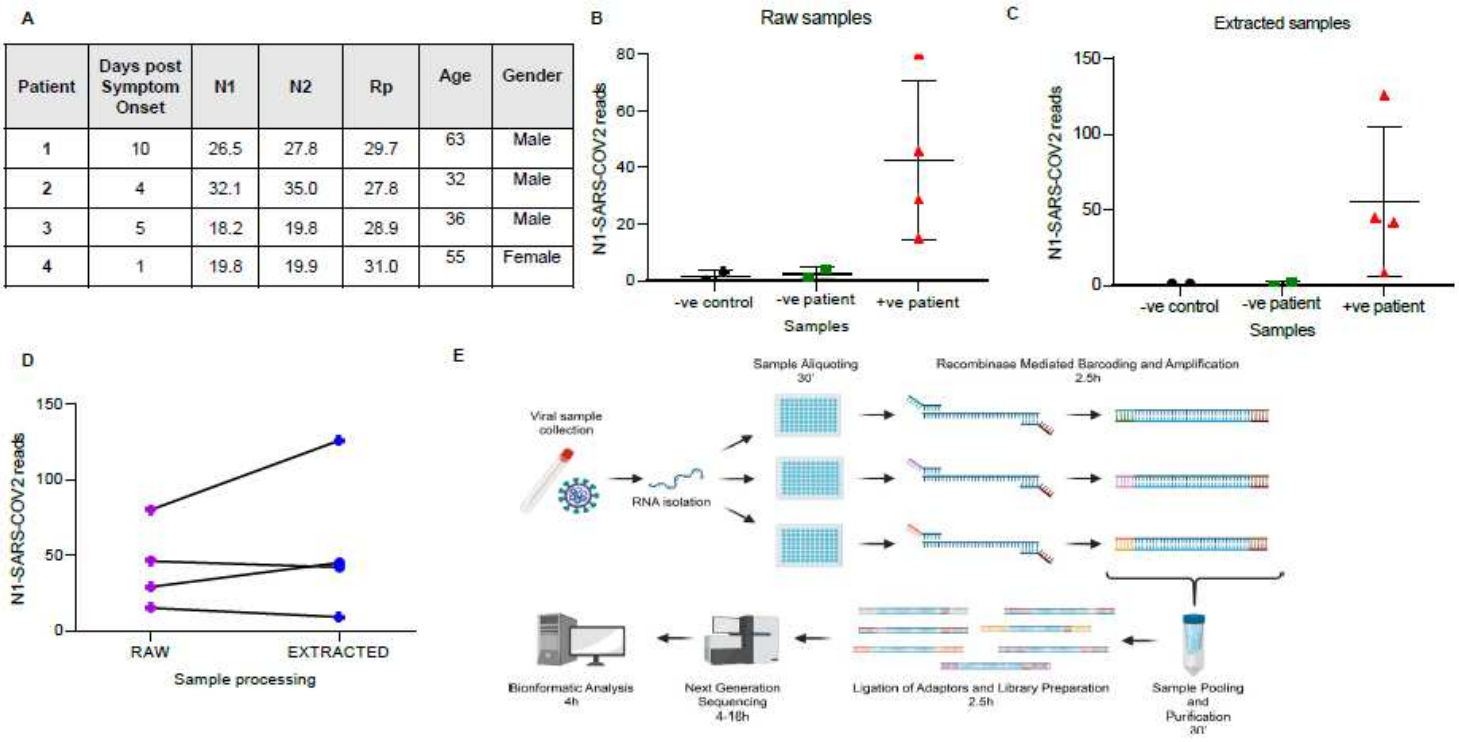


Figure 1

In vitro REMBRANDT testing. (A) Generation of in vitro transcribed control and SARS-CoV-2 RNA. (B) Schematic detailing the protocol utilized for the isothermal amplification and dual-barcoding of target RNAs. (C) Agarose gels using 8 different forward and 12 different reverse barcoded primers for the isothermal amplification of the SARS-CoV-2 N gene. Numbered lanes represent unique barcode combinations; – represents water control for each primer pair. (D) Schematic detailing the protocol utilized for the bioinformatics analysis of the REMBRANDT RNA-seq data.



Patient	Days post Symptom Onset	N1	N2	Rp	Age	Gender
1	10	26.5	27.8	29.7	63	Male
2	4	32.1	35.0	27.8	32	Male
3	5	18.2	19.8	28.9	36	Male
4	1	19.8	19.9	31.0	55	Female

Figure 2

REMBRANDT can identify SARS-CoV-2 positive patient samples from raw and extracted patient samples. (A) Patient information from tested individuals including age, gender, days of test following symptoms and COVID and RNaseP values from CDC COVID assay. (B) Number of SARS-CoV-2 reads mapped to negative control, negative patients and SARS-CoV-2 positive patients from raw samples. (C) Number of SARS-CoV-2 reads mapped to negative control, negative patients and SARS-CoV-2 positive patients from RNA extracted samples. (D) Comparison of the number of SARS-CoV-2 reads from each patient comparing raw and extracted patient sample. (E) Schematically detailing the protocol utilized for the REMBRANDT pipeline for testing for COVID-19. References