LAMP-Seq: Population-Scale COVID-19 Diagnostics Using a Compressed Barcode Space

Jonathan L. Schmid-Burgk[1,2,3,4,8,13,*], David Li[1,2,3,4], David Feldman[9], Mikołaj Słabicki[1,10,11], Jacob Borrajo[1,4], Jonathan Strecker[1,2,3,4], Brian Cleary[1], Aviv Regev[1,5,6,7,12], Feng Zhang[1,2,3,4,12]

NOTE: This protocol has not been validated with clinical samples. To facilitate collaborations with interested parties to jointly advance the fight against the current coronavirus pandemic, we have set up a public forum on www.LAMP-Seq.org.

[1] Broad Institute of MIT and Harvard
Cambridge, MA 02142, USA
[2] McGovern Institute for Brain Research
[3] Department of Brain and Cognitive Sciences
[4] Department of Biological Engineering
[5] Department of Biology
[6] Klarman Cell Observatory
[7] Koch Institute for Integrative Cancer Research
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[8] Institute of Clinical Chemistry and Clinical Pharmacology, University Hospital Bonn, 53127 Bonn, Germany
[9] Department of Biochemistry and Institute for Protein Design, University of Washington, Seattle, WA 98195, USA
[10] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[11] Division of Translational Medical Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), 69120 Heidelberg, Germany
[12] Howard Hughes Medical Institute, Cambridge, MA 02139, USA
[13] Lead Contact

*Correspondence should be addressed to J.S.-B. (jschmid@broadinstitute.org)

Summary

The ongoing COVID-19 pandemic has already caused devastating losses. Early evidence shows that the exponential spread of COVID-19 can be slowed by restrictive isolation measures, but these place a tremendous burden on society. Moreover, once these restrictions are lifted, the exponential spread is likely to re-emerge. It has been suggested that population-scale testing can help break the cycle of isolation and spread, but current detection methods are not capable of such large-scale processing. Here we propose LAMP-Seq, a barcoded Reverse-Transcription Loop-mediated Isothermal Amplification (RT-LAMP) protocol that could dramatically reduce the cost and complexity of population-scale testing. In this approach, individual samples are processed in a single heat step, producing barcoded amplicons that can be shipped to a sequencing center, pooled, and analyzed *en masse*. Using unique barcode combinations per sample from a compressed barcode space enables extensive pooling, significantly reducing cost and organizational efforts. Given the low cost and scalability of next-generation sequencing, we believe that this method can be affordably scaled to analyze millions of samples per day using existing sequencing infrastructure.

Introduction

As of April 2020, the recent global spread of a novel coronavirus, SARS-CoV-2, has already resulted in over 1,200,000 confirmed cases and 65,000 deaths (Johns Hopkins CSEE Covid tracker accessed April 4, 2020 (Dong et al., 2020)) since its emergence several months ago. Early epidemiological studies indicate that the exponential spread of COVID-19, the disease caused by SARS-CoV-2, can be slowed by restrictive isolation measures (Chinazzi et al., 2020), but these measures place an enormous burden on societies and economies. Moreover, once isolation measures are lifted, exponential spread is predicted to resume (Li et al., 2020). To exacerbate the problem, many infected individuals do not show symptoms, are unlikely to seek testing, and unknowingly contribute to the spread of infection. To effectively combat the transmission of SARS-CoV-2, we need population-scale testing that can enable rapid identification of all infected individuals, pinpoint key potential outbreak areas to support epidemiological interventions that maximize human health, while reducing the extent of restrictive isolation measures.

Currently, the majority of COVID-19 testing is performed using viral RNA extraction followed by RT-qPCR to amplify and detect one of several highly conserved regions of the SARS-CoV-2 genome, or by detecting serum antibodies specific for viral proteins. The global capacity for testing using these approaches, however, has been limited by a combination of access and supply issues for reagents and instruments.

Here, we propose LAMP-Seq, a novel protocol that would allow for population-scale testing using massively parallel RT-LAMP (Nagamine et al., 2002; Notomi et al., 2000) by employing

sample-specific barcodes. This approach requires only a single heating step for each individual sample (i.e., nasal swab, oropharyngeal swab, nasal wash, fecal sample, or potentially sputum), followed by pooled processing, parallelizable deep sequencing, and standard computational analysis. By using a simple thermal protocol for individual samples and pooling many samples prior to resource-intensive steps, the requirement for specialized reagents, equipment, and labor is greatly reduced as compared to established RT-qPCR protocols. Unique tracking of tens of millions of samples as well as asynchronous testing logistics can be obtained by additionally employing a compressed barcode space. Below, we describe the design and initial validation of the method and simulated barcoding strategies. We estimate that the cost per sample would be < 7 USD based on the list-price of off-the-shelf products (exclusive of labor and instrument costs), with a potential for a further 4-fold cost reduction through optimized production of a single enzyme. Most importantly, this approach is predicted to be scalable to hundreds of thousands of samples per day per sequencing facility.

Method Design

We propose the following approach for population-scale testing for SARS-CoV-2 infection: a barcoded RT-LAMP reaction is performed on an unpurified swab sample with primers specific for the SARS-CoV-2 genome, which is followed by large-scale pooling of samples, PCR amplification with additional barcoding, deep sequencing, and data analysis to identify positive individuals (Fig. 1A) (see below for detailed suggested protocol). Advantages of performing the RT-LAMP reaction at the site of sample collection include eliminating RNA extraction and

concerns about RNA stability, providing sterilization before shipment, and allowing parallel logistics of large numbers of samples through remote sample pooling.

RT-LAMP reactions have been demonstrated to be highly sensitive for sequence-specific viral nucleic acid detection (Lamb et al., 2020; Yang et al., 2020; Zhang et al., 2020), even from unpurified samples (Estrela et al., 2019). To establish a barcoded RT-LAMP reaction, we inserted barcode sequences into the forward inner primer (FIP), which enables generation of repeatedly barcoded palindromic amplification products (Fig. 1B). To limit the number of unique barcode primers needed for testing a large number of samples, a compressed barcode space can be used (Fig. 1C) (see below for details) as long as a small fraction of samples is expected to be positive during population scale testing.

Results

We designed three barcoded primer sets based on validated RT-LAMP amplicons (Suppl. Table 1, (Broughton et al., 2020; Lamb et al., 2020; Zhang et al., 2020)) perfectly matching 97.5% (amplicon A), 96.9% (amplicon B), and 95.6% (amplicon C) of 183 SARS-CoV-2 genomes available in the NCBI database (April 1st, 2020). 10-nt barcodes with GC content of 30%-70% and no homopolymer repeats of more than 4 nucleotides were inserted into the FIP primer. Barcodes were made robust to sequencing errors by ensuring a minimum Levenshtein edit distance between any barcode pair sufficient to detect either one (10,000 barcode set) or two (1,000 barcode set) insertion, deletion or substitution errors (Suppl. Table 1; the code is available at https://github.com/feldman4/dna-barcodes).

Comparing barcoded LAMP reactions to non-barcoded controls using a dsDNA surrogate template for SARS-CoV-2, we confirmed that the presence of a 10-nt barcode within the FIP primer does not affect LAMP sensitivity, product amounts, or downstream PCR amplification (Fig. 1D). Templating two individually barcoded LAMP reactions with 100-fold differing amounts of dsDNA template, combining them for PCR amplification, and sequencing the products resulted in read numbers within two-fold between the two samples (Fig. 1E), indicating that LAMP saturation can effectively compress the dynamic range of input viral loads. Furthermore, the expected sequence of barcoded LAMP-PCR products was confirmed by sequencing (Fig. 1F).

A standard Illumina NextSeq run can generate 200 million sequencing reads in 14 hours and we predict that this is sufficient for 100,000 patient samples per run, even accounting for library skewing due to differences in viral loads, largely because the vast majority of samples will be negative (for modeling see Supplementary Note 1). To explore parameters for using a compressed barcoding space, we conservatively assume that 1% of synthesized barcode primers systematically fail to work ($\Delta_{synth} = 0.01$), while additionally 5% of all sample-specific barcodes are not detected due to varying sequencing depth ($\Delta_{stoch} = 0.05$; this is independent of dropout due to low viral load). For automated assembly of testing reactions with unique barcode combinations, we anticipate that up to m = 10,000 barcode primers can be handled by available pipetting robots. Under these assumptions, we investigated for 100,000 samples what number of barcodes per sample (k), number of barcode primers total (m), and number of pools per run (m_2) would allow for minimal false-positive and false-negative rates of detection (Fig. 2A-C).

Interpreting the compressive barcoding problem as a modified Bloom filter (Supplementary Note 1), we predict that when using k = 5 barcodes per sample, requesting k' = 3 barcodes to be detected per sample, and splitting samples into m_2 = 10 pools per run, both the false-negative and false-positive rates of detection using a compressed barcode space will be less than 0.2% as long as the global frequency of positive samples is below 1.3% (Fig. 2A), at an estimated cost of < 7 USD per sample (Fig. 2D). We must emphasize, however, that the suggested barcoding method has not yet been experimentally validated, and that additional false negatives may arise during the RT-LAMP reaction stage.

Suggested Protocol

1. A fresh swab sample is inserted into a 500 μl RT-LAMP reaction, containing the following components:
    a. 1x Isothermal Amplification buffer (NEB),
    b. 6 mM $MgSO_4$,
    c. 1.4 mM dNTP mix,
    d. 0.5 μl Triton X-100 (amount to be optimized),
    e. 1.6 µM total of a unique set of one to five barcoded FIP primers (B-FIP-Barcode, TCTGGCCCAGTTCCTAGGTAGTNNNNNNNNNNCCAGACGAATTCGTGGTGG), where Ns denote a specific barcode sequence,
    f. 1.6 µM B-BIP primer (AGACGGCATCATATGGGTTGCACGGGTGCCAATGTGATCT),
    g. 0.2 µM B-F3 primer (TGGCTACTACCGAAGAGCT),
    h. 0.2 µM B-B3 primer (TGCAGCATTGTTAGCAGGAT),

i. 0.4 μM B-LF primer (GGACTGAGATCTTTCATTTTACCGT),

j. 0.4 μM B-LB primer (ACTGAGGGAGCCTTGAATACA),

k. 160 units Bst 3.0 DNA polymerase (NEB),

l. optionally, a dilute control template DNA or RNA differing from the target viral sequence, but sharing all primer binding sites,

a. water.

2. The RT-LAMP reaction with the swab is heated to 65 °C for 30 minutes to react, and to 95 °C for 10 minutes to sterilize.

3. Up to 100,000 reactions are pooled in batches of 1,000 to 10,000 samples per batch.

4. For each pool, a 12-cycle 100 μl PCR reaction is performed:

   a. 50 μl NEBNext 2x Master Mix (NEB),

   b. 0.5 μM PCR-B-fwd-10 primer (ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCGTCTTTGTTAGCACCATAGGG),

   c. 0.5 μM PCR-B-rev-12 primer (TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTACCATCTTGGACTGAGATCTTTC),

   d. 10 μl of pooled RT-LAMP reactions pre-diluted 100-fold in water,

   e. water.

5. For each pool, a secondary 12-cycle 100 μl PCR reaction is performed with:

   a. 50 μl NEBNext 2x Master Mix (NEB),

b. 0.5 μM pool-specific fwd barcoding primer (AATGATACGGCGACCACCGAGATCTACACNNNNNNNNNNACACTCTTTCCCTACACGACGCT), where Ns denote a specific barcode sequence,

c. 0.5 μM pool-specific rev barcoding primer (CAAGCAGAAGACGGCATACGAGATNNNNNNNNNNGTGACTGGAGTTCAGACGTGTGCT), where Ns denote a specific barcode sequence,

d. 5 μl of previous PCR reaction,

e. water.

6. The PCR products are pooled, gel-purified, and sequenced on an Illumina NextSeq sequencer, or similar device (A MiSeq sequencer can be used for testing the method, or when screening smaller numbers of samples).
7. Computationally, the barcodes co-occurring with the correct viral genome sequence excluding sequence portions covered by primers are determined.
8. Positive samples are determined using a database of barcode combinations assigned to sample IDs, requiring either one (single barcoding scenario) or at least three out of five sample barcodes (compressed barcode space) being positive.

Discussion

LAMP and RT-LAMP have been previously established for use as highly sensitive methods for pathogen detection from unpurified human samples with detection limits below 100 nucleic acid molecules. While colorimetric or turbidimetric (Mori and Notomi, 2009) readouts of LAMP reactions can suffer from false positive results (Estrela et al., 2019), we demonstrate that a sequencing-based readout detects correct fusions of barcode sequences with two stretches of

viral sequence. To further increase specificity, viral sequences can be filtered for sequence portions that are non-overlapping with primer sequences. In addition, we expect that this novel multiplexing-LAMP strategy is unlikely to suffer from barcode cross-contamination originating from template switching events at the PCR stage, as two template switching events would be required in order to create a sequencing-compatible amplicon.

A major advantage of the proposed method is that early pooling allows for scalable logistics (a hypothetical deployment protocol is provided in Suppl. Note 2). To that end, one or several barcode sequences need to be inserted in the forward inner primer (FIP) or backward inner primer (BIP) primer used during the RT-LAMP reaction, which we experimentally validated. We consider that one limitation of the presented approach is that skewing of sample representation at the pooling stage may affect testing sensitivity. Although in initial tests, the LAMP reaction saturated in positive samples largely independently of template concentrations, thus potentially equalizing the representation across positive samples in an advantageous manner (Fig. 1E), the reaction might also add random skewing to pooled samples; however, preliminary modeling suggested that pooling 100,000 samples per NextSeq run offers robust detection even assuming linear amplification of anticipated spread of viral loads (Supplementary Note 1). Remaining critical parameters to establish are the sensitivity of the RT-LAMP reaction using unpurified swab samples, the stability of RT-LAMP reactions during shipment, and, most importantly, the timely regulatory and ethical assessment of this novel testing approach for clinical validation.

LAMP-Seq requires low amounts of consumables with the exception of Bst 3.0 polymerase (5.7 USD per sample), which, however, could be mass-produced in *E. coli*, titrated down in

concentration, or replaced by a more cost-effective enzyme. The synthesis cost of the barcode primer library is negligible (5,000 USD total), leaving point-of-test infrastructure and logistics as putatively cost driving items. Once successfully established, however, this infrastructure could be rapidly adapted to counter future outbreaks; in a future scenario, multiplexing several targets might even enable scalable routine differential diagnostics.

Home testing might offer the advantage of higher public acceptance, lower potential for viral spread at diagnostic centers, and lower probability of sample cross-contamination, however, it is not clear if reliable swab acquisition at home is realistic. In addition, the RT-LAMP reaction would require a broadly available controlled heat source, which hypothetically could be a kitchen oven, a cooling down glass of boiling water, a USB-powered heating device, or a chemical process. Of note, the temperature ramp in an oven between 45 °C and 72 °C might allow Bst 3.0 polymerase to be active for several minutes, and offer the added benefit of sterilizing closed sample containers before shipment. Nevertheless, the issue of guaranteeing safety of postal workers during shipping of the inactivated samples has to be thoroughly addressed to allow implementation of a home testing scenario.

LAMP-Seq has the potential to enable scalable logistics through early-stage barcoding, while providing a low false-positive rate. Given the urgency of the present need, we hope that sharing this protocol early, even in the absence of clinical validation data, will spur additional development and parallel testing. We welcome collaboration from any interested parties so we can join together and rapidly develop a solution to advance the fight against the coronavirus pandemic.

Acknowledgements

We would like to thank Rhiannon Macrae for help with preparing the manuscript, and Hendrik Streeck, Michael Hölzel, Kerstin Ludwig, Ronja Hollstein, Gunther Hartmann, Markus Nöthen, Sophie Strobel, Samantha Laber, Amy Guo, Julia Reinhardt, André Heimbach, Eric Zhang, Vincent Huang, Daniel Liu, Sebastian Virreira Winter, Per Hoffmann, Johannes Schmid-Burgk, Stefan Frank, and José Miguel Zapata Rolón for helpful discussions. Support in protocol establishment was provided by the Institute of Human Genetics, University Hospital Bonn.

Declaration of Interests

J.S.-B., D.L., and F.Z. are inventors on a patent application filed by the Broad Institute related to this work with the specific aim of ensuring this technology can be made freely, widely, and rapidly available for research and deployment. F.Z. is a co-founder of Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies, and Sherlock Biosciences. A.R. is a founder of Celsius Therapeutics, equity holder in Immunitas, and an SAB member for ThermoFisher Scientific, Syros Pharmaceuticals, Asimov, and Neogene Therapeutics.

Figure Legends:

Figure 1 | Scalable deep-sequencing based approach for SARS-CoV-2 detection. (A) Schematic outline of a proposed scalable testing procedure. (B) Schematic of anticipated enzymatic reactions and reaction products. (C) Schematic illustration of a compressed barcode space allowing unique identification of millions of samples while minimizing barcode primer logistics. (D) Experimental validation of LAMP-Seq. All steps were performed as described in the Suggested Protocol section, with the exception that plasmid DNA containing the SARS-CoV-2 N-gene (IDT) was used as template instead of a swab sample, 1 ng/μl pX330 plasmid DNA was present as unspecific decoy DNA, 1x WarmStart LAMP Master Mix (NEB) was used instead of buffer, $MgSO_4$, dNTPs, Triton X-100, and polymerase, and the reaction was scaled down to a volume of 25 μl. Samples were run on an 1% agarose gel and visualized using ethidium bromide. (E) Barcoded LAMP reactions templated with either 100 or 10,000 dsDNA molecules were combined after heat inactivation, PCR amplified, and sequenced on an Illumina MiSeq sequencer. Relative read counts with respect to template amounts are shown as mean and standard deviation from two experimental replicates. (F) Base frequencies observed by sequencing a barcoded LAMP-Seq amplicon on a MiSeq without applying any read filtering are shown as a color-coded sequence logo.

Figure 2 | Modeling of compressed barcoding parameters and per-sample cost of LAMP-Seq. (A-C) Simulation of False Positive Probability and False Negative Probability depending on m_2 = the number of sub-pools (A), m = the complexity of the barcode library (B), and k = the number of barcodes per sample (C) when utilizing a compressed barcode space accounting for

barcode loss. Dashed grey lines indicate a probability threshold of 0.2%. (D) Cost estimation per sample using LAMP-Seq.

References

Broughton, J.P., Deng, X., Yu, G., Fasching, C.L., Singh, J., Streithorst, J., Granados, A., Sotomayor-Gonzalez, A., Zorn, K., Gopez, A., et al. (2020). Rapid Detection of 2019 Novel Coronavirus SARS-CoV-2 Using a CRISPR-based DETECTR Lateral Flow Assay. 1–27.

Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., et al. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science eaba9757–12.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis.

Estrela, P.F.N., de Melo Mendes, G., de Oliveira, K.G., Bailão, A.M., de Almeida Soares, C.M., Assunção, N.A., and Duarte, G.R.M. (2019). Ten-minute direct detection of Zika virus in serum samples by RT-LAMP. Journal of Virological Methods *271*, 113675.

Lamb, L.E., Bartolone, S.N., Ward, E., and Chancellor, M.B. (2020). Rapid Detection of Novel Coronavirus (COVID-19) by Reverse Transcription-Loop-Mediated Isothermal Amplification. 1–17.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., and Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science *6*, eabb3221–eabb3229.

Mori, Y., and Notomi, T. (2009). Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases. Journal of Infection and Chemotherapy *15*, 62–69.

Nagamine, K., Hase, T., and Notomi, T. (2002). Accelerated reaction by loop-mediated isothermal amplification using loop primers. Molecular and Cellular Probes *16*, 223–229.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., and Hase, T. (2000). Loop-mediated isothermal amplification of DNA. Nucl. Acids Res. *28*, E63–E63.

Yang, W., Dang, X., Wang, Q., Xu, M., Zhao, Q., Zhou, Y., Zhao, H., Wang, L., Xu, Y., Wang, J., et al. (2020). Rapid Detection of SARS-CoV-2 Using Reverse transcription RT-LAMP method. 1–25.

Zhang, Y., Odiwuor, N., Xiong, J., Sun, L., Nyaruaba, R.O., Wei, H., and Tanner, N.A. (2020). Rapid Molecular Detection of SARS-CoV-2 (COVID-19) Virus RNA Using Colorimetric LAMP. 1–14.

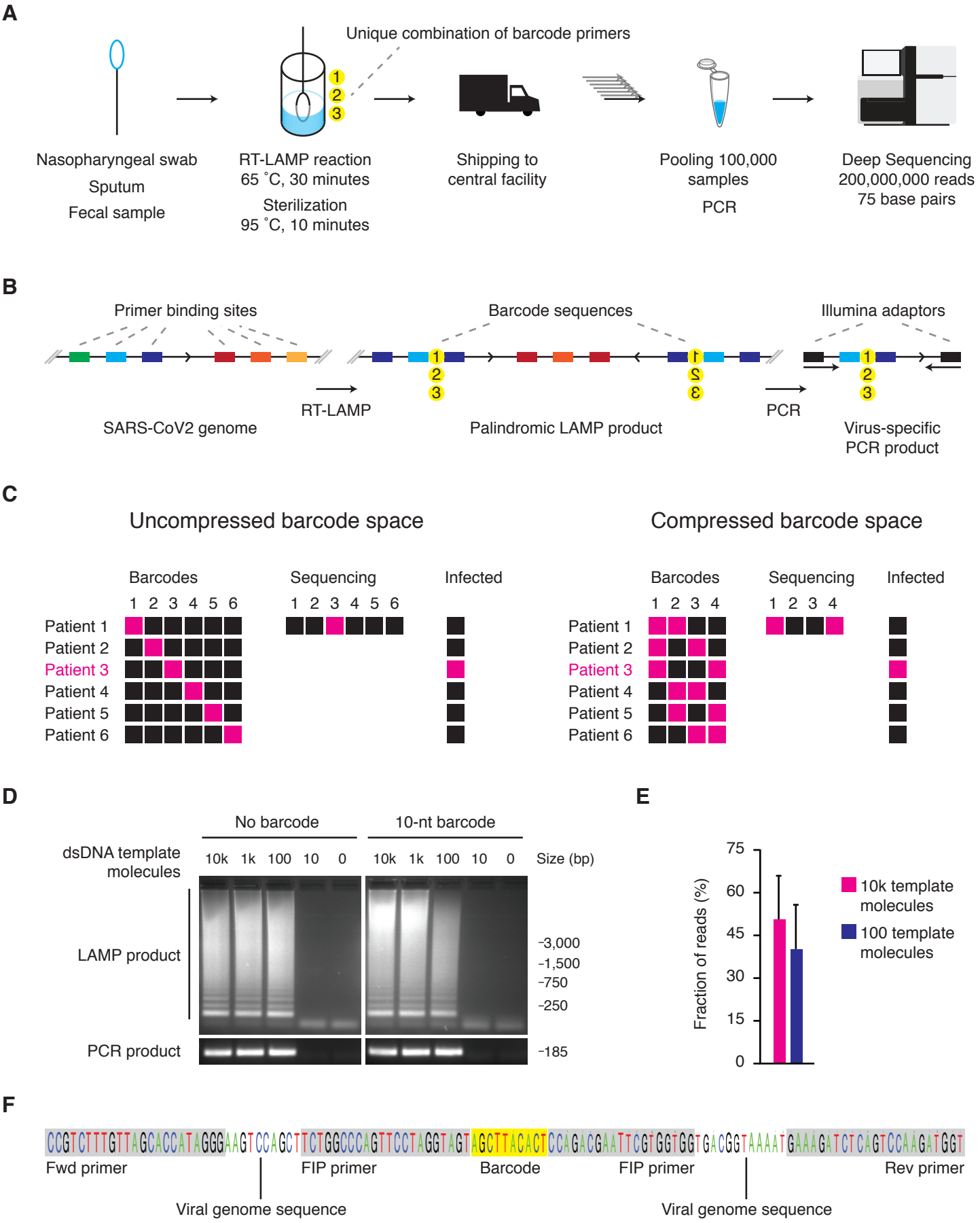
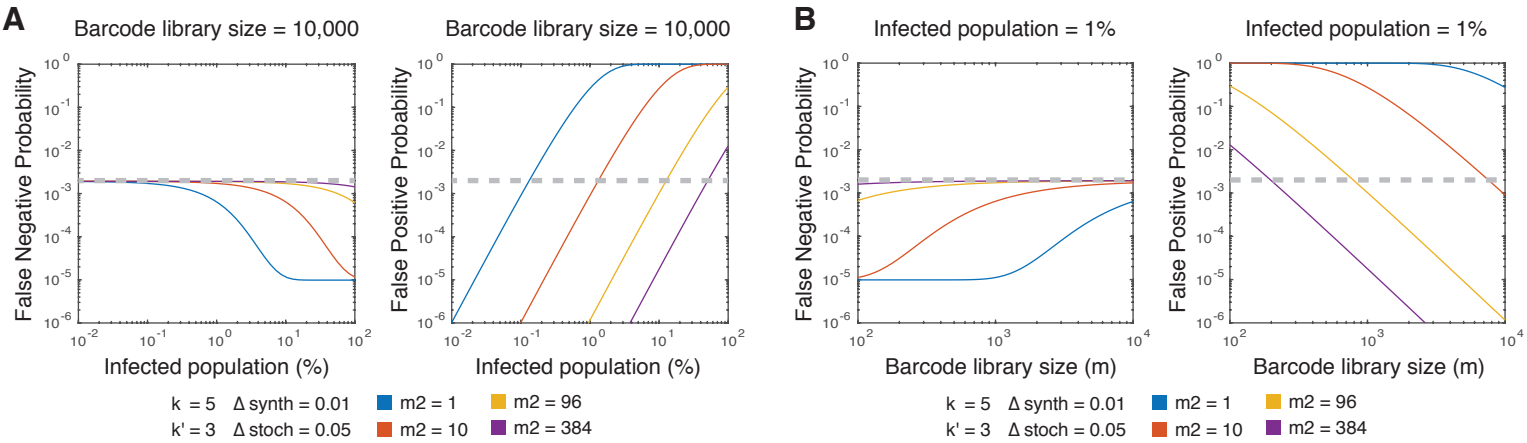
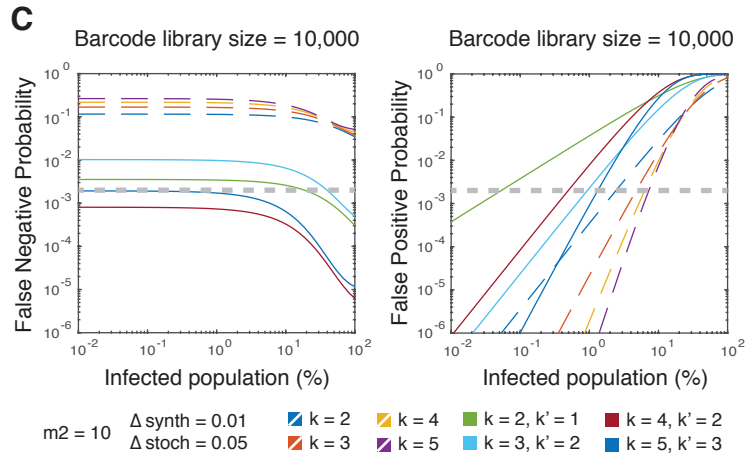

A
Unique combination of barcode primers
1
2
3
Nasopharyngeal swab
Sputum
Fecal sample
RT-LAMP reaction
65 ˚C, 30 minutes
Sterilization
95 ˚C, 10 minutes
Shipping to
central facility
Pooling 100,000
samples
PCR
Deep Sequencing
200,000,000 reads
75 base pairs
B
Primer binding sites
Barcode sequences
Illumina adaptors
SARS-CoV2 genome
RT-LAMP
Palindromic LAMP product
PCR
Virus-specific
PCR product
C
Uncompressed barcode space
Compressed barcode space
Barcodes
Sequencing
Infected
1 2 3 4 5 6
Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
1 2 3 4
D
No barcode
10-nt barcode
dsDNA template
molecules
10k 1k 100 10 0
Size (bp)
LAMP product
PCR product
-3,000
-1,500
-750
-250
-185
E
Fraction of reads (%)
75
60
45
30
15
0
10k template
molecules
100 template
molecules
F
CCGTCTTTGTTAGCACCATAGGGAAGTCCAGCTTCTGGCCCAGTTCCTAGGTAGTAGCTTACACTCCAGACGAATTCGTGGTGGTGACGGTAAAATGAAAGATCTCAGTCCAAGATGGT
Fwd primer
FIP primer
Barcode
FIP primer
Rev primer
Viral genome sequence
Viral genome sequence

Figure 1





D

Item	Cost per sample
Matrix tubes	0.5 $
Bst 3.0 polymerase	5.68 $
Pipette tips	0.5 $
Pooling containers	0.02 $
PCR reagents	0.006 $
Sequencing kits	0.02 $
Computation power	0.1 $
TOTAL	6.83 $ / sample

Figure 2