University of Texas at Tyler

Scholar Works at UT Tyler

Spring 2-1-2023

Optimization of the Illumina COVIDSeq™ protocol for decentralized, cost-effective genomic surveillance

Rob E. Carpenter
University of Texas at Tyler, rcarpenter@uttyler.edu

Vaibhav Tamrakar
ICMR-National Institute of Research in Tribal Health, vaibhav.ys90@gmail.com

Sadia Almas
Advanta Genetics, salmas@aalabs.com

Aditya Sharma
RetroBioTech, asharma@aalabs.com

Chase Rowan
Advanta Genetics, crowan@aalabs.com

See next page for additional authors

Follow this and additional works at: https://scholarworks.uttyler.edu/hrd_fac

Part of the Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons

Author
Rob E. Carpenter, Vaibhav Tamrakar, Sadia Almas, Aditya Sharma, Chase Rowan, and Rahul Sharma
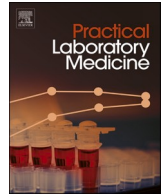
Optimization of the Illumina COVIDSeq™ protocol for decentralized, cost-effective genomic surveillance

Rob E. Carpenter [a,b,e], Vaibhav K. Tamrakar [c,d], Sadia Almas [a], Aditya Sharma [d], Chase Rowan [a], Rahul Sharma [a,d,e,*]

[a] *Advanta Genetics, 10935 CR 159, Tyler, TX, 75703, USA*
[b] *University of Texas at Tyler, 3900 University Boulevard, Tyler, TX, 75799, USA*
[c] *ICMR-National Institute of Research in Tribal Health, Jabalpur, MP, 482003, India*
[d] *RetroBioTech LLC, 838 Dalmalley Ln, Coppell, TX, 75019, USA*
[e] *Scienetix, 10935 CR 159, Tyler, TX, 75703, USA*

ABSTRACT

A decentralized surveillance system to identify local outbreaks and monitor SARS-CoV-2 Variants of Concern is one of the primary strategies for the pandemic's containment. Although next-generation sequencing (NGS) is a gold standard for genomic surveillance and variant discovery, the technology is still cost-prohibitive for decentralized sequencing, particularly in small independent labs with limited resources. We have optimized the Illumina COVIDSeq™ protocol for the Illumina MiniSeq instrument to reduce cost without compromising accuracy. We slashed the library preparation cost by half by using 50% of recommended reagents at each step and normalizing the libraries before pooling to achieve uniform coverage. Reagent-only cost (~ $43.27/sample) for SARS-CoV-2 variant analysis with this normalized input protocol on MiniSeq instruments is comparable to what is achieved on high throughput instruments such as NextSeq and NovaSeq. Using this modified protocol, we tested 153 clinical samples, and 90% of genomic coverage was achieved for 142/153 samples analyzed in this study. The lineage was correctly assigned to all samples (152/153) except for one. This modified protocol can help laboratories with constrained resources to contribute in decentralized COVID-19 surveillance in the post-vaccination era.

1. Introduction

The severe acute respiratory syndrome coronavirus (SARS-CoV-2) pandemic has claimed millions of lives globally, accentuating the need for systemic decentralized scientific breadths. Although currently, there is a downward progression of global COVID-19 cases [1], community-level surveillance by decentralized genomic sequencing is paramount for monitoring the transmission dynamics of the pandemic. Keeping SARS-CoV-2 in the foreground, new lineages will likely emerge, and monitoring the evolving variants is epidemiologically critical [2,3]. New SARS-CoV-2 variants may also have different clinical manifestations and vaccine efficacy. Thus, determining the infecting variant may also be used for clinical decision-making for an individual patient. While next-generation sequencing (NGS) has arisen as the gold standard technology for genomic surveillance and variant discovery [4], the technology

* Corresponding author. Advanta Genetics, 10935 CR 159, Tyler, TX, 75703, USA.
E-mail addresses: rahuldnadx@gmail.com, rsharma@aalabs.com (R. Sharma).

remains cost prohibitive for decentralized sequencing, particularly in small independent and resource-limited laboratories [5]. Illumina COVIDSeq™ is one of the most adopted methods for COVID-19 surveillance. However, the application has been limited to centralized surveillance programs because of high initial capital investment and reoccurring maintenance costs. Still, high throughput instruments and protocols often mandate sequencing a large batch of samples to attain efficiency and presumed low cost. Accordingly, this study describes the optimization of the Illumina COVIDSeq™ Research Use Only (RUO) assay protocol for the decentralized implementation of SARS-CoV-2 genomic surveillance.

2. Material and methods

As part of a Centers for Disease Control and Prevention (CDC) sponsored COVID surveillance program, 95 clinical samples from Advanta Genetics were sent to Fulgent Genetics for sequencing. These samples were sequenced at an average depth of 20,071.41x using Illumina COVIDSeq assay on NovaSeq 6000 instrument (Cost ~ 1 million USD), which requires pooling of 1000s of samples for efficient use of sequencing reagents and achieves the lowest sequencing cost (~$30/sample). The standard COVIDSeq™ protocol does not require the normalization of libraries; thus, the sequencing depth among 95 samples ranges from 268x to 77,387x (Supplementary Table). This over-sequencing approach is acceptable in high throughput laboratories, where samples are sequenced at higher depths to achieve sufficient coverage for each sample. However, using low throughput instruments, the standard protocol is cost-prohibitive for decentralized sequencing facilities.

Illumina has introduced a COVIDSeq™ 96 sample kit (Library preparation cost ~39.18/sample), and the protocol directs preparing a 50 μl library from each sample and pooling only 5 μl for sequencing. The standard protocol does not recommend quantification and normalization of libraries before pooling, and 90% of the library volume is not used for pooling and sequencing. We optimized the COVIDSeq™ 96 sample kit protocol to reduce cost and enhance efficiency on low throughput sequencing instruments like the Illumina MiniSeq® or MiSeq®. We optimized the protocol for using 50% reagent volume at each step during library preparation, resulting in a 25 μL library from each sample, cutting the library preparation reagent cost by half ($20/sample). Next, the individual library was quantified using a Qubit™ Flex Fluorometer (Invitrogen, Inc.). We also analyzed representative libraries (N = 12) prepared from the reference strains by capillary electrophoresis on the 5300 Fragment Analyzer system (Agilent Inc USA) and the average library size was 347.66 ± 22.82 bp. The final molarity of individual libraries was calculated using concentration (ng/μl) from Qubit™ Flex Fluorometer, but fragment analysis was not performed for clinical samples used in the study. The actual concentration of each library and universal library size of 400bp were used to compute each library's molarity. Individual libraries were pooled in equimolar concentration instead of equal volume, as the Illumina COVIDSeq™ kit recommended. These additional normalization steps allowed us to achieve uniform coverage of all the libraries in the pool and efficiently use a low-throughput sequencing instrument. The final library pool was again quantified using a Qubit™ Flex Fluorometer, denatured, and diluted to a 2 pM loading concentration following the manufacturer's instruction. Dual indexed paired-end sequencing with 75bp read length was carried out using the high output flow cell (list price: $1102.00) on the Illumina MiniSeq® instrument. This approach allowed us to reduce the material cost of SARS-CoV-2 variant detection (library preparation + normalization + sequencing) to ~$43.27/sample for a batch of 48 samples (Table 2). This protocol can be optimized to sequence 96 samples and potentially reduce the cost by $10/sample. Pre-pooling normalization using the individual library Qubit readings allowed us to achieve uniform coverage (median depth 595x) across the samples in the pool and higher efficiency.

Data were analyzed using free and easy-to-use resources. Illumina Basespace (https://basespace.illumina.com) bioinformatics pipeline was used for sequencing QC, FASTQ generation, and genome assembly, followed by SARS-CoV-2 variant determination using the DRAGEN COVID Lineage (Version: 3.5.4) application. Finally, the genome assembly FASTA file was also analyzed for lineage assignment using the web version of Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN) software (https://pangolin.cog-u).

3. Results

The lowest sequencing depth of >200x and 90% genome coverage were found adequate for accurate variant detection while validating this assay for clinical testing according to the CAP (College of American Pathologists) guidelines for NGS-based Laboratory Developed Test (LDT) [6]. Importantly, we sequenced 153 samples using this modified approach, and 152/153 (99.34%) samples (PCR Ct < 30) resulted in the correct variant using DRAGEN COVID Lineage (v3.5.4), and 142/153 (92.81%) samples achieved 90% genome

Table 1
Comparative sequencing using the modified COVIDSeq™ protocol.

Sequencing Parameter	Fulgent Genetics (n = 95[a])	Advanta Genetics (n = 153[a])
Mean Coverage	30257.05	1132.30
Median Coverage	32148.50	595.00
Standard Deviation (SD)	20177.89	1232.19
Highest Coverage (X times)	77387.80	6913.00
Lowest Coverage (X times)	268.00	56.00
Accurately Variant Calls	100% (95/95)	99.34 (152/153)

[a] 6 samples were split and sequenced by both approaches.

coverage with an average depth of >200x (Supplementary Table-1). Only 11/154 (7%) samples were below 200x coverage; out of 11 samples <200x coverage (Table 1), the DRAGEN COVID Lineage analysis pipeline failed in assigning the lineage of only one sample with the lowest depth. We also re-sequenced six samples already sequenced by another reference laboratory (Fulgent Genetics, Inc.) at extremely high coverage (>30,000X) and compared the variant identities from this modified sequencing protocol. The results confirmed that all six samples were identified to carry identical variants by both laboratories, implicating 100% accuracy in the inter-laboratory testing of this modified approach. Interestingly, 3 of the six split samples were sequenced at >50,000x coverage by Fulgent Genetics, whereas we were able to sequence the same samples at only 200X coverage with identical variant detection. This illustrates higher sequencing efficiency with pre-pooling quantification without compromising the test accuracy. Such higher efficiency is essential for the cost-effective application of this test in limited-resourced and decentralized laboratory settings and for reference laboratories that do not have access to high throughput instruments such as Illumina NextSeq® or NovSeq® instruments [7].

4. Discussion

COVID incident rate is sloping downward globally so is the financial support for high throughput genomic surveillance. However, persistent surveillance of circulating genomic variants in the community is crucial for scrutinizing the transmission dynamics of existing variants and tracking the emergence of new variants. We have optimized the SARS-CoV-2 variant detection assays for sequencing small batches (n = 48) of samples on a low-cost instrument (Illumina MiniSeq) which is ideal for small laboratories often not supported by centralized funding sources. Furthermore, surveillance data from such laboratories is critical for broadening the representation of the under-served population in global databases such as GSAID.

Simultaneous sequencing of ~3000 samples using COVIDSeq™ EUA kit on Illumina NovaSeq (Price~$1.0 million + $100,000/year maintenance cost) instrument is the most cost-effective (~$21.20/sample) option for SARS-CoV-2 genome sequencing often applied for mass surveillance. However, this published low cost is unattainable for independent clinical laboratories because of high capital investment and large batch size. Standard COVIDSeq™ protocol in Illumina's EUA test allows the sequencing of up to 384 samples on the NextSeq 550 at a lower cost ($25.33/sample), which is slightly lower than the estimated cost ($43.27/sample) of this modified protocol for MinSeq. However, cost-effective testing will still require pooling of >300 samples to achieve ~$25/sample reagent cost. Capital investment for NextSeq 550 instrument (Cost ~$300,000 + ~$30,000/year maintenance cost) is still significantly higher than MiniSeq (Cost ~$50,000 + ~$5,000/year maintenance cost). Pre-pooling normalization has reduced the library preparation and sequencing cost on MiniSeq instruments close to the NextSeq or NovaSeq. Therefore, the modified protocol could empower small resource-limited laboratories to contribute to local genomic surveillance. We have adopted this modified protocol for sequencing 153 genomes from East Texas, USA, and compared the results with PCR-based variant detection [8]. High accuracy and reproducibility of this approach have been demonstrated in validating the COVIDSeq™ RUO assay for clinical application according to Clinical Laboratory Improvement Amendments (CLIA) and College of American Pathologists (CAP) guidelines [6]. We have only scrutinized this procedure for accurate detection of circulating variants rather than detecting new mutations and variant frequency in the mixed variant population. Secondly, the cost estimates presented in this study are for the core reagents (library preparation and sequencing) only. Personnel cost is a significant expense in processing NGS samples, which can vary significantly according to geographical location. For example, capital investment and the cost of imported reagents are often the limiting factors in low-income countries, not so much the trained personnel. Therefore, this cost-effective approach can still benefit the low-throughput sequencing for monitoring emerging variants of SARS-CoV-2 and support decentralized genomic surveillance, particularly in resource-limited settings.

Author statement

Rob E. Carpenter: Conceptualization, review and editing. **Vaibhav K. Tamrakar:** Data curation, writing. **Sadia Almas:** Experiment planning and execution Investigation. **Aditya Sharma:** Data analysis and curation. **Chase Rowan**: Sample preparation and experimentation. **Rahul Sharma:** Conceptualization, writing-reviewing and editing.

Statement of informed consent

No conflicts, informed consent, or human or animal rights apply to this work.

Table 2
Estimated cost (reagents only) for decentralized sequencing of SARS-Cov-2 virus genome for epidemiological purpose.

Method	List price	Samples	Cost/sample
COVIDSeq Assay RUO with Index (Ct#20051773)	$ 3,840.00	192	$ 20.00
Qubit dsDNA Quantitation reagents (Invitrogen™ Q32851)	$ 361.00	500	$ 0.72
Sequencing Reagent (Illumina MiniSeq)	$ 1,102.00	30	$ 36.73
COVIDSeq library preparation and sequencing cost*/sample			$ 43.27
*Material only cost is estimated for processing a batch of 48 samples			

Declaration of competing interest

RS and RC declare that they have a financial interest in Scienetix, Inc. None of the other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plabm.2023.e00311.

References

[1] C.J. Murray, COVID-19 will continue, but the end of the pandemic is near, Lancet 399 (10323) (2022) 417–419, https://doi.org/10.1016/S0140-6736(22)00100-3.
[2] A. Aleem, A.B.A. Samad, A.K. Slenker, Emerging Variants of SARS-CoV-2 and Novel Therapeutics against Coronavirus (COVID-19), StatPearls Publishing, 2022. *StatPearls [Internet]*, https://www.ncbi.nlm.nih.gov/books/NBK570580/.
[3] R.M. Anderson, C. Vegvari, T.D. Hollingsworth, L. Pi, R. Maddren, C.W. Ng, R.F. Baggaley, The SARS-CoV-2 pandemic: remaining uncertainties in our understanding of the epidemiology and transmission dynamics of the virus, and challenges to be overcome, Interface Focus 11 (6) (2021), 20210008, https://doi.org/10.1098/rsfs.2021.0008.
[4] G. Berno, L. Fabeni, G. Matusali, C.E.M. Gruber, M. Rueca, E. Giombini, A.R. Garbuglia, SARS-CoV-2 variants identification: overview of molecular existing methods, Pathogens 11 (9) (2022) 1058, https://doi.org/10.3390/pathogens11091058.
[5] C.N. Umunnakwe, Z.N. Makatini, M. Maphanga, A. Mdunyelwa, K.M. Mlambo, P. Manyaka, H.A. Tempelman, Evaluation of a commercial SARS-CoV-2 multiplex PCR genotyping assay for variant identification in resource-scarce settings, PLoS One 17 (6) (2022), e0269071, https://doi.org/10.1371/journal.pone.0269071.
[6] R.E. Carpenter, V. Tamrakar, H. Chahar, T. Vine, R. Sharma, Confirming multiplex q-PCR use in COVID-19 with next generation sequencing: strategies for epidemiological advantage, Global Health, Epidemiology and Genomics (2022) 1–10, https://doi.org/10.1155/2022/2270965.
[7] E. Thomas, S. Delabat, Y.L. Carattini, D.M. Andrews, SARS-CoV-2 and variant diagnostic testing approaches in the United States, Viruses 13 (12) (2021) 2492, https://doi.org/10.3390/v13122492.
[8] R.E. Carpenter, V. Tamrakar, S. Almas, E. Brown, R. Sharma, COVID Seq as laboratory developed test (LDT) for diagnosis of SARS-CoV-2 variants of concern (VOC), Arch Clin Biomed Res 6 (6) (2022) 954–970, https://doi.org/10.26502/acbr.50170309.Epub.2022.Nov.28, 2022.