bioRxiv preprint doi: https://doi.org/10.1101/2022.11.07.515545; this version posted November 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Optimization of the Illumina COVIDSeq™ protocol for decentralized, cost-effective**
2 **genomic surveillance**
3

4 Rob E. Carpenter, PhD[1,2,5], Vaibhav K. Tamrakar, PhD[3,4], Sadia Almas, PhD[1], Chase Rowan,
5 MS[1], Rahul Sharma PhD[1,4,5*]
6 [1]Advanta Genetics, 10935 CR 159, Tyler, Texas 75703, USA
7 [2]University of Texas at Tyler, 3900 University Boulevard, Tyler, Texas 75799, USA
8 [3]ICMR-National Institute of Research in Tribal Health, Jabalpur, MP 482003, INDIA
9 [4]RetroBioTech LLC, 838 Dalmalley Ln, Coppell, TX 75019, USA
10 [5]Scienetix, 10935 CR 159, Tyler, Texas 75703, USA
11
12
13 Corresponding author:
14 Rahul Sharma, PhD
15 Advanta Genetics, 10935 CR 159 Tyler, Texas 75703, USA
16 Phone: +1(225) 573-2690
17 Email: rahuldnadx@gmail.com; rsharma@aalabs.com
18
19
20
21
22
23
24
25

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.07.515545; this version posted November 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

26 **Abstract**
27
28 A decentralized surveillance system to identify local outbreaks and monitor SARS-CoV-2
29 Variants of Concern is one of the primary strategies for the pandemic's containment. Although
30 next-generation sequencing (NGS) is a gold standard for genomic surveillance and variant
31 discovery, the technology is still cost-prohibitive for decentralized sequencing, particularly in
32 small independent labs with limited resources. We have optimized the Illumina COVID-seq
33 protocol to reduce cost without compromising accuracy. 90% of genomic coverage was achieved
34 for 142/153 samples analyzed in this study. The lineage was correctly assigned to all samples
35 (152/153) except for one. This modified protocol can help laboratories with constrained
36 resources contribute to decentralized SARS-CoV-2 surveillance in the post-vaccination era.

37

38

39 **Keywords:** Illumina COVID-seq, SARS-CoV-2 variants, genomic surveillance, transmission
40 dynamics, cost-effective approach

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.07.515545; this version posted November 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

41 The severe acute respiratory syndrome coronavirus (SARS-CoV-2) pandemic has
42 claimed millions of lives globally, highlighting the need for systemic de-centralized scientific
43 capacities. Although currently, there is a downward progression of global COVID-19 cases
44 (Murray et al., 2022), community-level surveillance by de-centralized genomic sequencing
45 remains vital for monitoring the transmission dynamics of the pandemic and SARS-CoV-2
46 mutagenic capacities. Keeping SARS-CoV-2 in the foreground, new lineages will likely emerge,
47 and monitoring evolving variants is epidemiologically critical (Aleem et al., 2022; Anderson et
48 al., 2021). While next-generation sequencing (NGS) has emerged as the gold standard
49 technology for genomic surveillance and variant discovery (Berno et al., 2022), the technology
50 remains cost prohibitive for de-centralized sequencing, particularly in small independent and
51 resource-limited laboratories (Umunnakwe et al., 2022). Illumina COVIDSeq™ is one of the
52 most utilized methods for COVID surveillance, but the application has been limited mainly to
53 centralized surveillance programs.

54 As part of a surveillance program sponsored by the Centers for Disease Control and
55 Prevention (CDC), 95 clinical samples from Advanta Genetics were sent to Fulgent Genetics for
56 sequencing. Our samples were sequenced at an average depth of 20,071.41x using Illumina
57 COVIDSeq assay on NovaSeq 6000 instrument (Cost ~ 1 million USD), which requires pooling
58 of 1000s of samples. The standard protocol doesn't require the normalization of libraries; thus,
59 sequencing depth among 95 samples ranges from 286x to 287,4419x (Supplemental Table). This
60 over-sequencing approach is acceptable in high throughput laboratories, where samples are
61 sequenced at higher depths to achieve sufficient coverage for each sample. However, using low
62 throughput instruments, the standard protocol is cost-prohibitive for de-centralized sequencing
63 facilities.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.07.515545; this version posted November 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

64 Although Illumina has introduced a COVIDSeq™ 96 sample kit (Cat#20049393;
65 cost/sample: $40.00), the protocol directs preparing a 50µl library from each sample and pooling
66 only 5µl for sequencing. The standard protocol doesn't recommend quantification and
67 normalization of libraries before pooling, and 90% of the library volume is not used for pooling
68 and sequencing.

69 Accordingly, this study describes the optimization of the Illumina COVIDSeq™
70 Research Use Only (RUO) assay protocol for the decentralized implementation of SARS-CoV-2
71 genomic surveillance. The protocol is modified to reduce cost and enhance efficiency on low
72 throughput sequencing instruments like the Illumina MiniSeq® or MiSeq®. We optimized the
73 protocol by using 50% reagent volume at each step during library preparation, resulting in a
74 25µL library from each sample, cutting the library preparation cost by half ($20/sample). Next,
75 the individual library was quantified using a Qubit™ Flex Fluorometer (Invitrogen, Inc.), and
76 representative libraries (N=12) prepared from the reference strains were also analyzed on a
77 fragment analyzer. The average library size (347.66±22.82 bp) and its concentration in (ng/µL)
78 were used to calculate the molarity of the individual library. Individual libraries were also
79 quantified using the KAPA Library Quantification Kits (Roche Cat # 07960140001). Individual
80 libraries were diluted to achieve a normalized concentration of 10nM, and the normalized
81 libraries were pooled in equimolar concentration instead of equal volume as recommended by
82 the Illumina COVIDSeq™ kit. These additional normalization steps allowed us to achieve
83 uniform coverage of all the libraries in the pool and efficiently use a low-throughput sequencing
84 instrument. The final library pool was again quantified using a Qubit™ Flex Fluorometer
85 (Invitrogen, Inc.) and a PCR-based library quantification kit (Scienetix, USA). The final library
86 pool was denatured and diluted to a 2 pM loading concentration. Dual indexed paired-end

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.07.515545; this version posted November 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

87 sequencing with 75bp read length was carried out using the high output flow cell (list price:
88 $1102.00) on the Illumina MiniSeq® instrument. This approach allowed us to reduce the cost of
89 the COVIDSeq™ reagent (library preparation + sequencing) to $56.00 per sample with a batch
90 size of only ~30 samples. Pre-pooling normalization allowed us to achieve uniform coverage
91 (median depth 595x) across the samples in the pool and higher efficiency. We could sequence
92 ~30 clinical samples (diagnostic PCR Ct<30), positive control (Wuhan-Hu-1), and no template
93 control (NTC) with each batch of libraries on a MiniSeq high-output flow cell.

94 The lowest sequencing depth of >200x and 90% genome coverage were found adequate
95 for accurate variant detection while validating this assay for clinical testing according to the CAP
96 (College of American Pathologists) guidelines for NGS-based Laboratory Developed Test (LDT)
97 (Carpenter et al., 2022). Importantly, we sequenced 153 samples using this modified approach,
98 and 152/153 (99.34%) samples (PCR Ct<30) resulted in the correct variant using DRAGEN
99 COVID Lineage (v3.5.4), and 142/153 (92.81%) samples achieved 90% genome coverage with
100 an average depth of >200x (Supplemental Table). Only 11/154 (7%) samples are below 200x
101 coverage; out of 11 samples < 200x coverage (Table 1), the DRAGEN COVID Lineage analysis
102 pipeline failed in assigning the lineage of only one sample with the lowest depth. We also re-
103 sequenced six samples already sequenced by another reference laboratory (Fulgent Genetics,
104 Inc.) at extremely high coverage (>30,000X) and compared the variant identities from this
105 modified sequencing protocol. The results confirmed that all six samples were identified to carry
106 identical variants by both laboratories, implicating 100% accuracy in the inter-laboratory testing
107 of this modified approach. Interestingly, 3 of the six split samples were sequenced at >50,000x
108 coverage by Fulgent Genetics, whereas we were able to sequence the same samples at only 200X
109 coverage with identical variant detection. This illustrates higher sequencing efficiency with pre-

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.07.515545; this version posted November 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

110 pooling quantification without compromising the test accuracy. Such higher efficiency is
111 essential for the cost-effective application of this test in limited-resourced and de-centralized
112 laboratory settings and for reference laboratories that do not have access to high throughput
113 instruments such as Illumina NextSeq® or NovSeq® instruments. (Thomas et al., 2021)

114 This modified protocol can potentially empower small resource-limited laboratories to
115 contribute to local genomic surveillance. We have adopted this modified protocol for sequencing
116 153 genomes from East Texas, USA, and compared the results with PCR-based variant detection
117 (Carpenter et al., 2022). High accuracy and reproducibility of this approach have been
118 demonstrated in validating the COVIDSeq™ assay for clinical application according to Clinical
119 Laboratory Improvement Amendments and College of American Pathologists guidelines. This
120 cost-effective approach can be widely adopted for low-throughput sequencing and monitoring of
121 emerging variants of SARS-CoV-2 and further support de-centralized genomic surveillance.

122 **Author Contributions**

123 Conceptualization RC and RS; Data curation SA, CR and RS; Formal analysis; RS and SA
124 Validation RS and SA; Roles/Writing: original draft RS, SA, and VT; Writing: review & editing
125 RC.

126 **Statement of Informed Consent**

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

128 **Declaration of Competing Interest**

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

132

133

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.07.515545; this version posted November 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

134 Table 1: Comparative sequencing using the modified COVIDSeq™ protocol
135

Sequencing Parameter	Fulgent Genetics (n=95*)	Advanta Genetics (n=153*)
Mean Coverage	30257.05	1132.30
Median Coverage	32148.50	595.00
Standard Deviation (SD)	20071.41	1232.19
Highest Coverage (X times)	2874419.50	6913.00
Lowest Coverage (X times)	268.00	56.00
Accurately Variant Calls	100% (95/95)	99.34 (152/153)

* 6 samples were split and sequenced by both approaches.

136

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.07.515545; this version posted November 8, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

137 **References**

138 Aleem, A., Samad, A. B. A., & Slenker, A. K. (2022). Emerging variants of SARS-CoV-2 and
139 novel therapeutics against coronavirus (COVID-19). In *StatPearls [Internet]*. StatPearls
140 Publishing. https://www.ncbi.nlm.nih.gov/books/NBK570580/
141 Anderson, R. M., Vegvari, C., Hollingsworth, T. D., Pi, L., Maddren, R., Ng, C. W., &
142 Baggaley, R. F. (2021). The SARS-CoV-2 pandemic: remaining uncertainties in our
143 understanding of the epidemiology and transmission dynamics of the virus, and
144 challenges to be overcome. *Interface Focus*, *11*(6), 20210008.
145 https://doi.org/10.1098/rsfs.2021.0008
146 Berno, G., Fabeni, L., Matusali, G., Gruber, C. E. M., Rueca, M., Giombini, E., & Garbuglia, A.
147 R. (2022). SARS-CoV-2 Variants Identification: Overview of Molecular Existing
148 Methods. *Pathogens*, *11*(9), 1058. https://doi.org/10.3390/pathogens11091058
149 Carpenter, R. E., Tamrakar, V., Chahar, H., Vine, T., & Sharma, R. (2022). Confirming
150 multiplex q-PCR use in COVID-19 with next generation sequencing: Strategies for
151 epidemiological advantage. *Global Health, Epidemiology and Genomics*, 1-10.
152 https://doi.org/10.1155/2022/2270965
153 Carpenter, R. E., Tamrakar, V., Almas, S., Brown, E., & Sharma, R. (2022). COVID Seq as
154 Laboratory Developed Test (LDT) for diagnosis of SARS-CoV-2 Variants of Concern
155 (VOC). *Archives of Virology*, in review (ARVI-D-22-00875)
156 Murray, C. J. (2022). COVID-19 will continue, but the end of the pandemic is near. *The Lancet*,
157 *399*(10323), 417-419. https://doi.org/10.1016/S0140-6736(22)00100-3

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.07.515545; this version posted November 8, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

158 Thomas, E., Delabat, S., Carattini, Y. L., & Andrews, D. M. (2021). SARS-CoV-2 and variant
159 diagnostic testing approaches in the United States. *Viruses*, *13*(12), 2492.
160 https://doi.org/10.3390/v13122492

161 Umunnakwe, C. N., Makatini, Z. N., Maphanga, M., Mdunyelwa, A., Mlambo, K. M., Manyaka,
162 P., & Tempelman, H. A. (2022). Evaluation of a commercial SARS-CoV-2 multiplex
163 PCR genotyping assay for variant identification in resource-scarce settings. *Plos one*,
164 *17*(6), e0269071. https://doi.org/10.1371/journal.pone.0269071