COVseq is a cost-effective workflow for mass-scale SARS-CoV-2 genomic surveillance

Michele Simonetti
Science for Life Laboratory, Karolinska Institutet https://orcid.org/0000-0003-3322-1697
Ning Zhang
Science for Life Laboratory, Karolinska Institutet and Qilu Hospital of Shandong University
https://orcid.org/0000-0002-6430-4236
Luuk Harbers
Science for Life Laboratory, Karolinska Instituet https://orcid.org/0000-0003-3910-6497
Maria Grazia Milia
Ospedale Amedeo di Savoia
Silvia Brossa
Instituto di Candiolo FPO - IRCCS
Thi Thu Huong Nguyen
Science for Life Laboratory, Karolinska Instituet
Francesco Cerruti
Ospedale Amedeo di Savoia
Enrico Berrino
Instituto di Candiolo FPO - IRCCS
Anna Sapino
Instituto di Candiolo FPO - IRCCS
Magda Bienko
Science for Life Laboratory, Karolinska Institutet https://orcid.org/0000-0002-6499-9082
Antonino Sottile
Instituto di Candiolo FPO - IRCCS
Valeria Ghisetti (✉ valeria.ghisetti@gmail.com)
Ospedale Amedeo di Savoia
Nicola Crosetto (✉ nicola.crosetto@ki.se)
Science for Life Laboratory, Karolinska Institutet https://orcid.org/0000-0002-3019-6978

Method Article

Keywords: SARS-CoV-2, Next-generation sequencing, genomic surveillance, COVseq, virology

Posted Date: May 26th, 2021

DOI: https://doi.org/10.21203/rs.3.pex-1338/v2

Abstract

While mass-scale vaccination campaigns are ongoing worldwide, genomic surveillance of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is critical to monitor the emergence and global spread of viral variants of concern (VOC). Here, we present a streamlined workflow—COVseq—which can be used to generate highly multiplexed sequencing libraries compatible with Illumina platforms from hundreds of SARS-CoV-2 samples in parallel, in a rapid and cost-effective manner. We benchmarked COVseq against a standard library preparation method (NEBNext) on 29 SARS-CoV-2 positive samples, reaching 95.4% of concordance between single-nucleotide variants detected by both methods. Application of COVseq to 245 additional SARS-CoV-2 positive samples demonstrated the ability of the method to reliably detect emergent VOC as well as its compatibility with downstream phylogenetic analyses. A cost analysis showed that COVseq could be used to sequence thousands of samples at less than 15 USD per sample, including library preparation and sequencing costs. We conclude that COVseq is a versatile and scalable method that is immediately applicable for SARS-CoV-2 genomic surveillance and easily adaptable to other pathogens such as influenza viruses.

Introduction

Since the identification of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as the causative agent of coronavirus disease 2019 (COVID-19)[1], thousands of SARS-CoV-2 genomes have been sequenced worldwide and the sequences have been made publicly available on the global initiative on sharing all influenza data (GISAID)[2] (https://www.gisaid.org/). This has enabled a phylogenetic reconstruction of the viral spread and evolution across different countries and continents at an unprecedented scale[3], allowing the rapid identification of new viral variants, including the UK[4], South African[5] and Brazilian[6] variants of concern (VOC). SARS-CoV-2 whole genome sequencing (WGS) based on next-generation sequencing (NGS) has been used for genomic surveillance to track infections in hospitals and community settings[7,8], as well as to monitor viral outbreaks in breeding farms, such as mink farms[9]. More recently, with the advent of mass-scale vaccination campaigns worldwide, SARS-CoV-2 genomic surveillance has become vital to rapidly trace the emergence and spread of VOC with potentially reduced susceptibility to the current vaccines[10,11]. In this context, the availability of scalable and cost-effective methods for centralized sequencing of hundreds or potentially thousands of samples per week would be greatly beneficial. Various NGS based approaches have been developed to perform SARS-CoV-2 WGS using different sequencing platforms. These include direct RNA sequencing and metagenomics[12–14], amplicon-based methods[13,15,16], and oligonucleotide capture-based methods[13,16–18]. Recently, numerous commercial kits based on the above methods have become available on the market and are being deployed in SARS-CoV-2 genomic surveillance[19]. However, existing commercial solutions are very costly and/or difficult to scale up largely because, typically, one sequencing library must be prepared for each individual sample and multiple indexed libraries must be carefully quantified and balanced before pooling them together prior to sequencing. This limits the number of samples that can be sequenced on a weekly basis, increasing the risk of missing emerging variants of potential

concern. In addition to its applications for WGS, NGS has also been used for mass-scale SARS-CoV-2 testing, such as in the SwabSeq method[20]; however, the latter does not provide full genome coverage. To counteract these limitations, here we present a versatile, scalable and cost-effective workflow—COVseq—that can be used to prepare multiplexed WGS libraries from many SARS-CoV-2 samples in parallel. We validate COVseq using RNA extracted from a SARS-CoV-2 viral culture as well as 274 diagnostic samples collected in three different phases of the ongoing pandemic at two hospitals in Italy. We demonstrate that the genome sequences obtained by COVseq are compatible with downstream phylogenomic analyses, including detailed phylogenetic reconstruction of a COVID-19 nosocomial outbreak that occurred in January 2021 at a single hospital in Italy. Lastly, we perform a real-life cost analysis based on our experience from the genomic surveillance program that we recently initiated for the Piemonte Region in North-West Italy, demonstrating that COVseq is a highly cost-effective method for SARS-CoV-2 genomic surveillance, including in low-income countries.

Reagents

Reagents:

- Absolute Ethanol (VWR, cat. no. 20816.367)

- Nuclease-Free Water (Thermo Fisher Scientific, cat. no. AM9932)

- Mineral oil (Sigma, cat. no. M5904)

- Random hexamers (50 µM) (Thermo Fisher Scientific, cat. no. N80800127)

- dNTPs (10 mM) (Thermo Fisher Scientific, cat.no. R0191)

- NEBNext Q5 Hot Start HiFi PCR Master Mix (NEB, cat.no. M0543L)

- SYBR Green (Thermo Fisher Scientific, cat.no. S7563)

- Primer pools 1,2,3,4,5 and 6 (IDT custom @ 50 µM)

- IDT ARTIC nCoV-2019 V3 panel pools 100 µM (IDT, cat.no. 10006788)

- MseI (NEB, cat.no. R0525L)

- NlaIII (NEB, cat.no. R0125L)

- CutSmart® buffer (NEB, cat. no. B7204S)

- T4 DNA Ligase (Thermo Fisher Scientific, cat. no. EL0011)

- T4 DNA rapid Ligase (Thermo Fisher Scientific, cat. no. K1423)

- COVseq oligonucleotide adapters

- UltraPure™ BSA (50 mg/ml) (Thermo Fisher Scientific, cat. no. AM2616)

- ATP Solution (100 mM) (Thermo Fisher Scientific, cat. no. R0441)

- dNTPs (25mM) (Thermo Fisher Scientific, cat.no. R1121)

- MEGAscript® T7 Transcription Kit (Thermo Fisher Scientific, cat. no. AM1334)

- DNase I, RNase-free (Thermo Fisher Scientific, cat. no. AM2222)

- RA3 adaptor and RTP, RP1 and RPI primers (custom-made by Integrated DNA Technologies Inc. based on the sequences in the TruSeq Small RNA Library Preparation kit, Illumina)

- RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen, cat. no. 10777-019)

- T4 RNA ligase 2, truncated (NEB, cat. no. M0242L)

- SuperScript® IV Reverse Transcriptase (Thermo Fisher Scientific, cat. no. 18090050)

- NEBNext UltraII Q5 PCR MasterMix (NEB, cat. no. M0544S)

- Agencourt RNAClean XP (Beckman Coulter, cat. no. A63987)

- Agencourt AMPure XP (Beckman Coulter, cat. no. A63881)

- Qubit™ RNA BR Assay Kit (Thermo Fisher Scientific, cat. no. Q10211)

- Qubit™ dsDNA BR Assay Kit (Thermo Fisher Scientific, cat. no. Q32850)

- Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32851)

- Bioanalyzer High Sensitivity DNA Kit (Agilent, cat. no. 5067-4627)

- Optional: RNase H (Thermo Fisher Scientific, cat.no. 18021014)

Consumables:

- Eppendorf® DNA LoBind microcentrifuge tubes 0.5 ml (Sigma, cat. no. EP0030108035-250EA)

- Eppendorf® DNA LoBind microcentrifuge tubes 1.5 ml (Sigma, cat. no. EP0030108051-250EA)

- Sapphire Filter tips, low retention (Greiner Bio-One, cat. no. 771265, 773265, 738265, 750265)

- microTUBE-50 AFA Fiber Screw-Cap (25) (Covaris, cat. no. 520166)

- # 96-well plates (Thermo Fisher Scientific, cat. no. 4316813)

- # 384-well plates (Thermo Fisher Scientific, cat. no. 4483320)

- Qubit™ Assay Tubes (Thermo Fisher Scientific, cat. no. Q32856)

- Bioanalyzer High-sensitivity DNA kit (Chips) (Agilent, cat. no. 5067-4626)

Equipment

- Incubator (for example, Binder incubator, Model KB 53 or Boekel Scientific InSlide Out, cat. no. 05-450-50)

- Tabletop centrifuge (for example, Eppendorf® Microcentrifuge 5424)

- I-DOT One (Dispendix GmbH, Stuttgart, Germany)

- Thermoshaker (for example, Eppendorf® Thermomixer Compact)

- PCR thermocycler (for example, Biometra TRIO)

- Sonication device (for example, ME220 Focused-ultrasonicator, Covaris)

- DynaMag™-2 Magnet (Thermo Fisher Scientific, cat. no. 12321D)

- Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, cat. no. Q32866)

- Bioanalyzer 2100 (Agilent, cat. no. G2943CA)

Procedure

DAY 1

First strand synthesis

1. Mix the following components:

RNA 5 µL

50 µM random hexamers 1 µL

10 mM dNTPs 1 µL

Nuclease-Free Water 6 µL

2. Incubate for 5 min @ 65 °C

3. Place the tube immediately on ice for 5 min

4. Add the following components:

5x SSIV buffer 4 µL

0.1 M DTT 1 µL

RNaseOUT 1 µL

SSIV Reverse Transcriptase enzyme 1 µL

5. Perform the following steps in a PCR thermocycler with the lid set @ 85 °C:

1. 23 °C 10 min

2. 50 °C 10 min

3. 85 °C 10 min

4. 4 °C Hold

6. Add 1 µL RNase H to the tube and incubate for 20 min at 37 °C

7. Note: RNase H is an optional step

Option 1: Multiplex PCR using the CDC primer pools

Note: Prepare primers as 50 µM primer stocks. Add an equal amount of each 50 µM primer stock to six different Eppendorf tubes labeled as pool 1,2,3,4,5 and 6. Prepare 10 µM working concentration by diluting each pool 1:5 with Nuclease-Free Water.

8. Mix the following components:

NEBNext Q5 Hot Start HiFi PCR Master Mix 15 µL

Nuclease-Free Water 9.2 µL

Primer pool 1,2,3,4,5 or 6 (10 µM) 1.8 µL

4x SYBR Green 1 µL

9. Add 3 µL of cDNA to each tube/well

10. Perform the following steps in PCR thermocycler with the lid set to 105 °C:

1. 98 °C 30 sec

2. 98 °C 15 sec

3. 65 °C 5 min

GOTO step 2, 40 times

4. 4 °C Hold

DNA purification

11. Pool 20 µL from each of the six amplicon pools into a 1.5 ml Eppendorf DNA LoBind tube or 96 well plate

12. Add 1 vol/vol ratio of AMPure XP beads pre-warmed at room temperature

13. Mix thoroughly and incubate for 10 min at room temperature

14. Place the sample on a magnetic stand

15. Incubate for at least 5 min until the liquid appears clear

16. Remove and discard the supernatant

17. Wash the beads twice with 200 µL of freshly prepared 80% ethanol

18. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

19. Remove the sample from the magnetic stand

20. Resuspend the beads in 80 µL of Nuclease-Free Water

21. Incubate for 2 min at room temperature

22. Place the sample back on the magnetic stand

23. Incubate for at least 5 min until the liquid appears clear

24. Transfer the supernatant to a new 1.5 µL Eppendorf DNA LoBind tube

25. Check the library concentration using Qubit dsDNA BR kit

Note: Samples can be stored for a long time at –20 °C

Option 2: Multiplex PCR using the ARTIC V3 primer pools

Note: If using IDT ARTIC nCoV-2019 V3 panel pools, prepare 10 µM working concentration by diluting each pool 1:10 with Nuclease-Free Water. The average concentration of each primer in the final reaction is 15 nM.

8. Mix the following components:

NEBNext Q5 Hot Start HiFi PCR Master Mix 12.5 µL

Nuclease-Free Water 2.9 µL

Primer pool 1 or 2 (10 µM) 3.6 µL

9. Add 6 µL of cDNA to each tube/well

10. Perform the following steps in PCR thermocycler with the lid set to 105 °C:

1. 98 °C 30 sec

2. 98 °C 15 sec

3. 63 °C 5 min

GOTO step 2, 35 times

4. 4 °C Hold

DNA purification

11. Pool 20 µL from each of the two amplicon pools into a 1.5 ml Eppendorf DNA LoBind tube or 96 well plate

12. Add 0.8 vol/vol ratio of AMPure XP beads pre-warmed at room temperature

13. Mix thoroughly and incubate for 10 min at room temperature

14. Place the sample on a magnetic stand

15. Incubate for at least 5 min until the liquid appears clear

16. Remove and discard the supernatant

17. Wash the beads twice with 200 µL of freshly prepared 80% ethanol

18. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

19. Remove the sample from the magnetic stand

20. Resuspend the beads in 40 µL of Nuclease-Free Water

21. Incubate for 2 min at room temperature

22. Place the sample back on the magnetic stand

23. Incubate for at least 5 min until the liquid appears clear

24. Transfer the supernatant to a new 1.5 µL Eppendorf DNA LoBind tube

25. Check the library concentration using Qubit dsDNA BR kit

Note: Samples can be stored for a long time at –20 °C

DAY 2

Note: To process multiple samples in parallel, we performed all reactions until IVT in 384-well plates. We used the I-DOT One nanodispensing device (Dispendix GmBH) to reduce the volumes of each reagent. Other dispensing systems may also be used; however, volumes might have to be adjusted depending on the technical specifications of each instrument.

DNA digestion

26. Dispense manually 5 µL of mineral oil per well in the targeted region of 384-well plates

27. Dispense 10 nL to 50 nL of purified PCR amplicons

Note: The amount of purified PCR amplicons can be increased or reduced depending on the Ct value of the samples

28. Dispense Nuclease-Free Water to a total volume of 350 nL

Note: From now, after dispensing for each step, we shake the plate in a ThermoMixer at 1,000 rpm for 1 min and centrifuge at 3,220 g for 5 min before each incubation

29. Mix the following components:

NlaIII enzyme 50 nL

MseI enzyme 50 nL

10x CutSmart Buffer 50 nL

30. Dispense 150 nL per well

31. Perform the following incubation steps:

1. 37 °C 1 h

2. 65 °C 20 min

3. 4 °C Hold

Ligation of COVseq adapters

32. Dispense 150 nL of 33 nM COVseq adaptor of NlaIII per well

33. Dispense 150 nL of 33 nM COVseq adaptor of MseI per well

34. Dispense 700 nL of ligation mix per well (standard ligase):

Nuclease-free water 250 nL

5x T4 ligase buffer 150 nL

ATP 10 mM 120 nL

BSA 50 mg/ml 30 nL

T4 standard ligase 150 nL

Alternatively, when using rapid ligase:

Nuclease-free water 50 nL

5x T4 ligase buffer 300 nL

ATP 10 mM 120 nL

BSA 50 mg/ml 30 nL

T4 rapid ligase 200 nL

35. Perform incubation at 22° C for 1h followed by inactivation at 70° C for 5 min

Note: When using rapid ligase decrease the incubation time to 30 min

36. Dispense manually 5 µl of Nuclease-Free Water/33 nM EDTA (for a final concentration of 25 nM) per well

37. Pool the content of multiple wells manually in a 1.5 mL or 5 mL eppendorf tube

38. Spin down the tube and carefully remove the upper phase containing mineral oil

Note: The pooling step can be performed by centrifuging the plate upside down at 117 g for 1 min into a collection plate placed at the bottom of the centrifuge.

DNA cleanup

39. Add 1.2 vol/vol ratio of AMPure XP beads pre-warmed at room temperature

40. Mix thoroughly and incubate for 10 min at room temperature

41. Place the sample on a magnetic stand

42. Incubate for at least 5 min until the liquid appears clear

43. Remove and discard the supernatant

44. Wash the beads twice with freshly prepared 80% ethanol (the ethanol should be enough to cover the beads)

45. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

46. Remove the sample from the magnetic stand

47. Resuspend the beads in 10 µl of nuclease-free water

48. Incubate for 2 min at room temperature

49. Place the sample back on the magnetic stand

50. Incubate for at least 5 min until the liquid appears clear

51. Transfer the supernatant to a new 1.5 µl DNA LoBind tube

52. Check the library concentration using Qubit dsDNA HS kit

Note: Samples can be stored for long time @ -20 °C

In vitro transcription

53. Start with 8 µl from the previous step

54. Add the following reagents on ice:

rATP+rUTP+rGTP+rCTP* 8 µl

10x T7 polymerase buffer 2 µl

T7 polymerase 1.5 µl

RNaseOUT™ Recombinant Ribonuclease Inhibitor 0.5 µl

*Prepared from separate rNTP solutions provided with the MEGAscript® T7 Transcription Kit

55. Incubate for 14 hours at 37 °C in a PCR thermocycler with the lid set @ 70 °C

Note: IVT can also be performed at 37°C for 2 hours to save time.

DAY 3

RNA cleanup

56. Add 1 µl of DNase I (RNase-free) to the IVT product

57. Incubate for 15 min @ 37 °C

58. Bring up the volume to 50 µl by adding 29 µl Nuclease-Free Water, then mix with 90 µl (1.8 vol/vol) of RNAClean XP beads pre-warmed at room temperature

59. Mix thoroughly and incubate for 10 min at room temperature

60. Place the sample on a magnetic stand

61. Incubate for at least 5 min until the liquid appears clear

62. Remove and discard the supernatant

63. Wash the beads twice with 200 µl of freshly prepared 70% ethanol

64. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

65. Remove the sample from the magnetic stand

66. Resuspend the beads in 9.5 µl of nuclease-free water

67. Incubate for 2 min at room temperature

68. Place the sample back on the magnetic stand

69. Incubate for at least 5 min until the liquid appears clear

70. Transfer 8.8 µl of supernatant to a new 0.5 µl DNA LoBind tube

71. Check the library concentration with 1 µl using Qubit dsDNA BR kit

RA3 adapter ligation

72. Add 1 µl of 10 µM RA3 adapter to 7.8 µl obtained after RNA cleanup

73. Incubate for 2 min @ 70 °C in a PCR thermocycler, then immediately place the sample on ice

74. Add 3.2 µl of the following mix:

RNA ligase buffer 1.2 µl

RNaseOUT^{TM} Recombinant Ribonuclease Inhibitor 1 µl

T4 RNA ligase truncated 1 µl

75. Incubate for 2 hours @ 25 °C in a PCR thermocycler with the lid set @ 30 °C

Reverse transcription

76. Add 2 µl per sample of 10 µM RTP primer

77. In a PCR thermocycler, incubate for 2 min @ 70 °C

78. Quickly transfer the sample to ice

79. Add 11 µl of the following mix:

5x SSIV buffer 5 µl

25mM dNTPs 1 µl

0.1M DTT 2 µl

RNaseOUT™ Recombinant Ribonuclease Inhibitor 1 µl

SuperScript IV reverse transcriptase 2 µl

80. Incubate for 20 min @ 50°C followed by inactivation for 10 min @ 80 °C in a PCR thermocycler with the lid set @ 80 °C

Library indexing and amplification

81. Add 16 µl per sample of the desired indexed Illumina primer

82. Add 359 µl of the following mix:

Nuclease-free water 143 µl

NEBNext® Ultra™ II PCR Master Mix 200 µl

10 µM RP1 primer 16 µl

83. Divide the final mix in 8 PCR tubes with each containing 50 µl

84. In a PCR thermocycler perform the following cycles:

1. 98 °C 30 sec

2. 98 °C 10 sec

3. 60 °C 30 sec

4. 65 °C 45 sec

GOTO step 2, 10 times

5. 65 °C 5 min

6. 4 °C Hold

Note: 10 PCR cycles are used for an input to the IVT of ≈ 200 ng. Please, adjust PCR cycles accordingly to the input to the IVT step.

Final library cleanup

85. Pool the 8 tubes for each sample and add 0.8 vol/vol ratio of AMPure XP beads pre-warmed at room temperature

86. Mix thoroughly and incubate for 10 min at room temperature

87. Place the sample on a magnetic stand

88. Incubate for at least 5 min until the liquid appears clear

89. Remove and discard the supernatant

90. Wash the beads twice with 1 ml of freshly prepared 80% ethanol

91. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

92. Remove the sample from the magnetic stand

93. Resuspend the beads in 50 µl of nuclease-free water

94. Incubate for 2 min at room temperature

95. Place the sample back on the magnetic stand

96. Incubate for at least 5 min until the liquid appears clear

97. Transfer the supernatant to a new 1.5 µl DNA LoBind tube

98. Check the library concentration using Qubit dsDNA HS kit

99. Check the fragment distribution on a Bioanalyzer 2100 using DNA HS chip

Note: Libraries can be stored for long time @ -20 °C

Troubleshooting

Time Taken

Anticipated Results

References

1. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).

2. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **1**, 33–46 (2017).

3. Mercatelli, D. & Giorgi, F. M. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front. Microbiol.* **11**, 1800 (2020).

4. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Virological* https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (2020).

5. Tegally, H. *et al.* Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* 2020.12.21.20248640 (2020) doi:10.1101/2020.12.21.20248640.

6. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology. *Virological* https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586 (2021).

7. Harilal, D. *et al.* SARS-CoV-2 Whole Genome Amplification and Sequencing for Effective Population-Based Surveillance and Control of Viral Transmission. *Clin. Chem.* **66**, 1450–1458 (2020).

8. Meredith, L. W. *et al.* Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.* **20**, 1263–1272 (2020).

9. Oude Munnink, B. B. *et al.* Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* (2020) doi:10.1126/science.abe5901.

10. Cyranoski, D. Alarming COVID variants show vital role of genomic surveillance. *Nature* **589**, 337–338 (2021).

11. Abdool Karim, S. S. & de Oliveira, T. New SARS-CoV-2 Variants - Clinical, Public Health, and Vaccine Implications. *N. Engl. J. Med.* (2021) doi:10.1056/NEJMc2100362.

12. Pillay, S. *et al.* Whole Genome Sequencing of SARS-CoV-2: Adapting Illumina Protocols for Quick and Accurate Outbreak Investigation during a Pandemic. *Genes* **11**, (2020).

13. Nasir, J. A. *et al.* A Comparison of Whole Genome Sequencing of SARS-CoV-2 Using Amplicon-Based Sequencing, Random Hexamers, and Bait Capture. *Viruses* **12**, (2020).

14. Chen, C. *et al.* MINERVA: A Facile Strategy for SARS-CoV-2 Whole-Genome Deep Sequencing of Clinical Samples. *Mol. Cell* **80**, 1123-1134.e4 (2020).

15. Tyson, J. R. *et al.* Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* (2020) doi:10.1101/2020.09.04.283077.

16. Xiao, M. *et al.* Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Med.* **12**, 57 (2020).

17. Paden, C. R. *et al.* Rapid, Sensitive, Full-Genome Sequencing of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg. Infect. Dis.* **26**, 2401–2405 (2020).

18. Doddapaneni, H. *et al.* Oligonucleotide capture sequencing of the SARS-CoV-2 genome and subgenomic fragments from COVID-19 individuals. *bioRxiv* (2020) doi:10.1101/2020.07.27.223495.

19. Chiara, M. *et al.* Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. *Brief. Bioinform.* **22**, 616–630 (2021).

20. Bloom, J. S. *et al.* Swab-Seq: A high-throughput platform for massively scaled up SARS-CoV-2 testing. *MedRxiv Prepr. Serv. Health Sci.* (2020) doi:10.1101/2020.08.04.20167874.

Acknowledgements

We gratefully acknowledge the authors from the Originating Laboratories and the Submitting Laboratories who generated and shared via GISAID the data on which this research is based[2]. We thank Britta Bouwman (Bienko-Crosetto lab) for critically reading the manuscript and for helping during the final revision. We acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure and the NGS Facility at the Candiolo Cancer Institute for their support with sequencing. This work was supported by funds from the National Natural Science Foundation of China (no. 81972475) and Chinese Postdoctoral Science Foundation (2019T120593, 2018M630787) to N.Z.; by a PhD fellowship under the funding of Dipartimenti di Eccellenza 2018–2022 (no. D15D18000410001) to E.B.; by funds from the Fondazione Piemontese per la Ricerca sul Cancro (INTEGRAZIONI DIAGNOSTICA IN ONCOLOGIA – INTERONC FPRC 5x1000 MIUR 2017) to A.S. and A.SO.; by a SciLifeLab/KAW National COVID-19 Research Program project grant, Research Area Viral Sequence Evolution to N.C.; by a grant from the Swedish Foundation for Strategic Research (SSF BD15-0095) to N.C., through which the I-DOT system was purchased; and by private donations for COVID-19 research from Chiesi Pharma AB and Tetra Pak also to N.C.