SUPPLEMENTARY INFORMATION

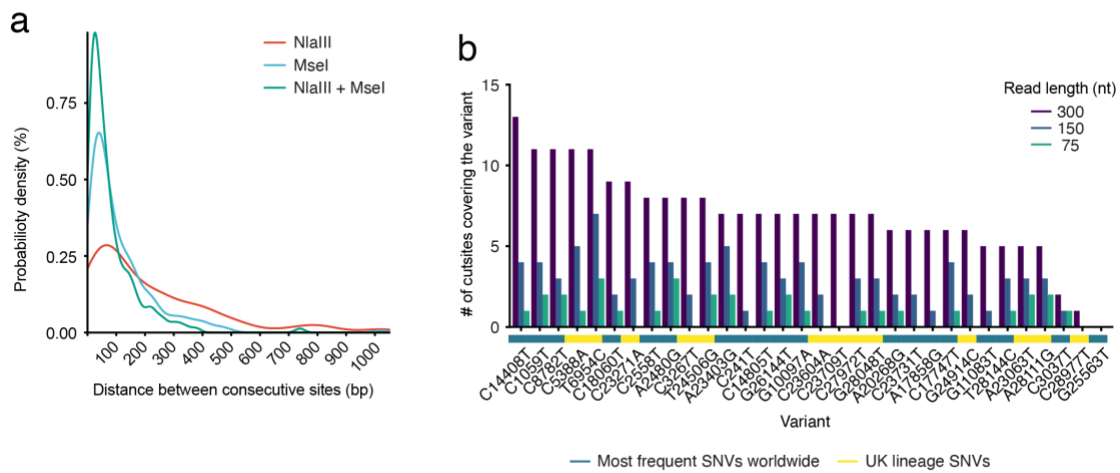
COVseq is a cost-effective workflow for mass-scale SARS-CoV-2 genomic surveillance

Michele Simonetti, Ning Zhang, Luuk Harbers, Maria Grazia Milia, Thi Thu Huong Nguyen, Silvia Brossa, Magda Bienko, Anna Sapino, Antonino Sottile, Valeria Ghisetti & Nicola Crosetto
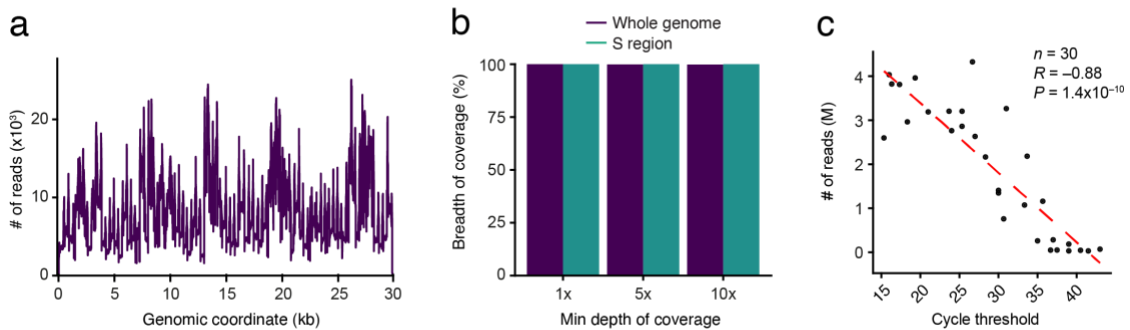
1. Supplementary Figures
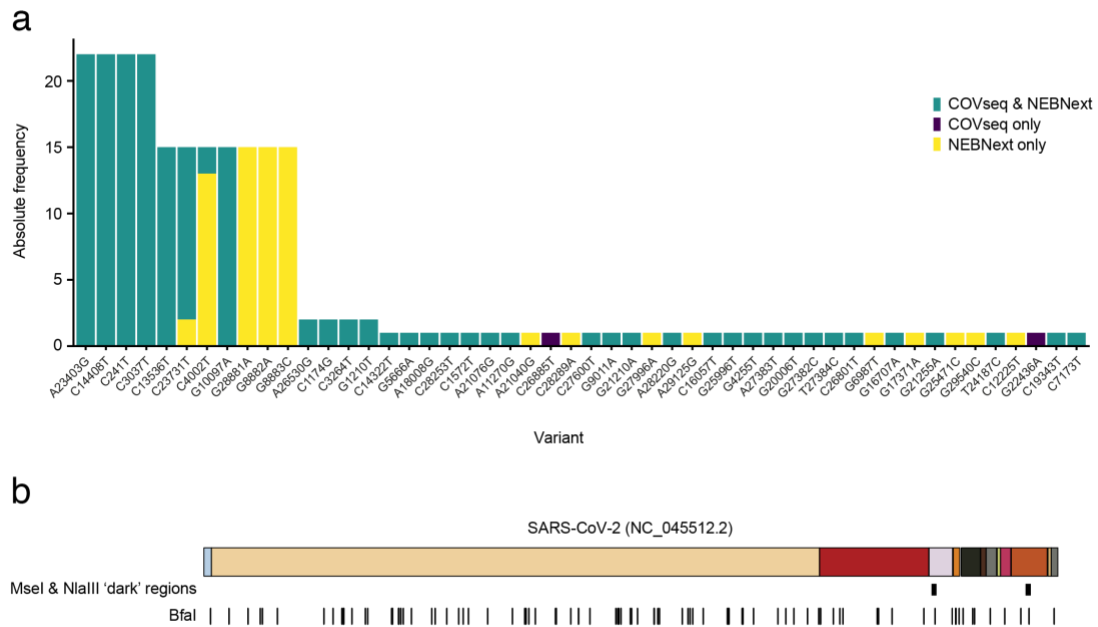
Supplementary Figure 1

Supplementary Fig. 1. Theoretical SARS-CoV-2 genome coverage by COVseq. (**a**) Distributions of distances between consecutive MseI and NlaIII recognition sites along the SARS-CoV-2 genome. Theoretically, 97.1% of the fragments generated by cutting with both MseI and NlaIII are less than 300 base pairs (bp). (**b**) Theoretical COVseq coverage of the most frequent SARS-CoV-2 SNVs detected worldwide[1] and of the SNVs in the recently emerged UK lineage[2], using single-end (SE) sequencing with different read lengths, as shown on the top right. The y-axis indicates the number of MseI or NlaIII recognition sites that are closer than the read length from each SNV.
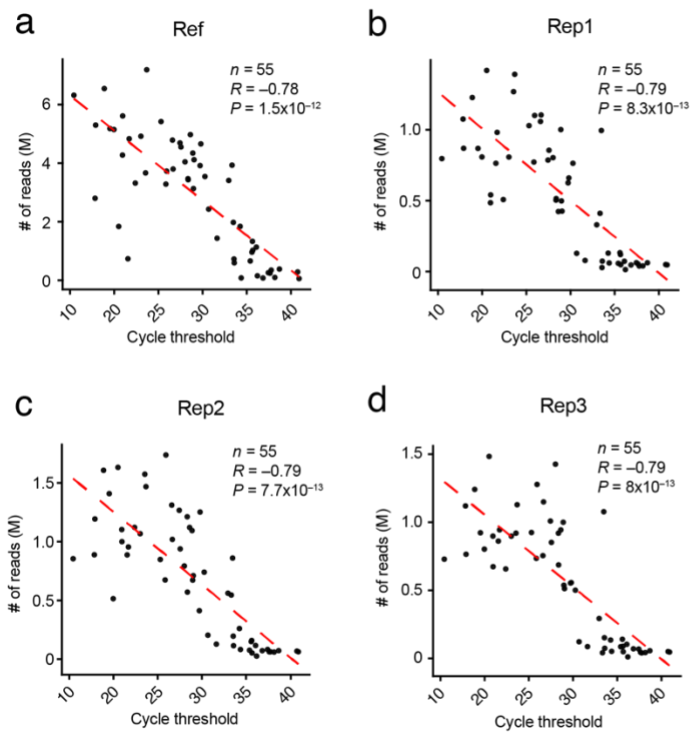
Supplementary Figure 2

Supplementary Fig. 2. SARS-CoV-2 genome coverage by the multiplexed PCR assay developed by the US CDC and NEBNext library preparation kit applied to one RNA sample extracted from the supernatant of a SARS-CoV-2 viral culture (see **Methods**). (**a**) Number of sequencing reads mapping at each base along the SARS-CoV-2 genome. The regions of the SARS-CoV-2 genome displaying the highest depth of coverage correspond to the regions covered by more amplicons, as shown in **Fig. 1a**. (**b**) Whole genome and S region coverage at various minimum sequencing depths, for the same NEBNext library shown in (a). (**c**) Inverse correlation between the cycle threshold determined by RT-PCR and the number of reads, for 30 (*n*) samples (samples 1–30 in **Supplementary Table 4**) sequenced by NEBNext (see **Methods**). M, millions. Each dot represents a sample. Dashed red line: linear regression fit. *R*, Pearson's correlation coefficient. *P*, t-test, two-tailed.

Supplementary Fig. 3. Validation of COVseq by a standard library preparation method (NEBNext). (**a**) Number of SNVs detected in 30 samples (samples 1–30 in **Supplementary Table 4**) sequenced by COVseq, NEBNext or both. (**b**) Ideogram showing the regions in the SARS-CoV-2 genome ('dark regions') that are more than 300 nt away from the closest MseI or NlaIII site and therefore cannot be covered by SE150. Adding an extra restriction enzyme (BfaI) is predicted to result in coverage of these 'dark regions' using SE150. Vertical black bars indicate BfaI recognition sites. The colors of the genes in the SARS-CoV-2 genome are the same as in **Fig. 1a**.

Supplementary Figure 4

Supplementary Fig. 4. (**a-d**) Inverse correlation between the cycle threshold determined by RT-PCR and the total number of reads obtained by COVseq, in the reference (Ref) and three replicate (Rep) libraries prepared from 55 SARS-CoV-2 positive left-over RNA samples (samples 31–85 in **Supplementary Table 4**). M, millions. Each dot represents a sample. Dashed red line: linear regression fit. *R*, Pearson's correlation coefficient. *P*, t-test, two-tailed.
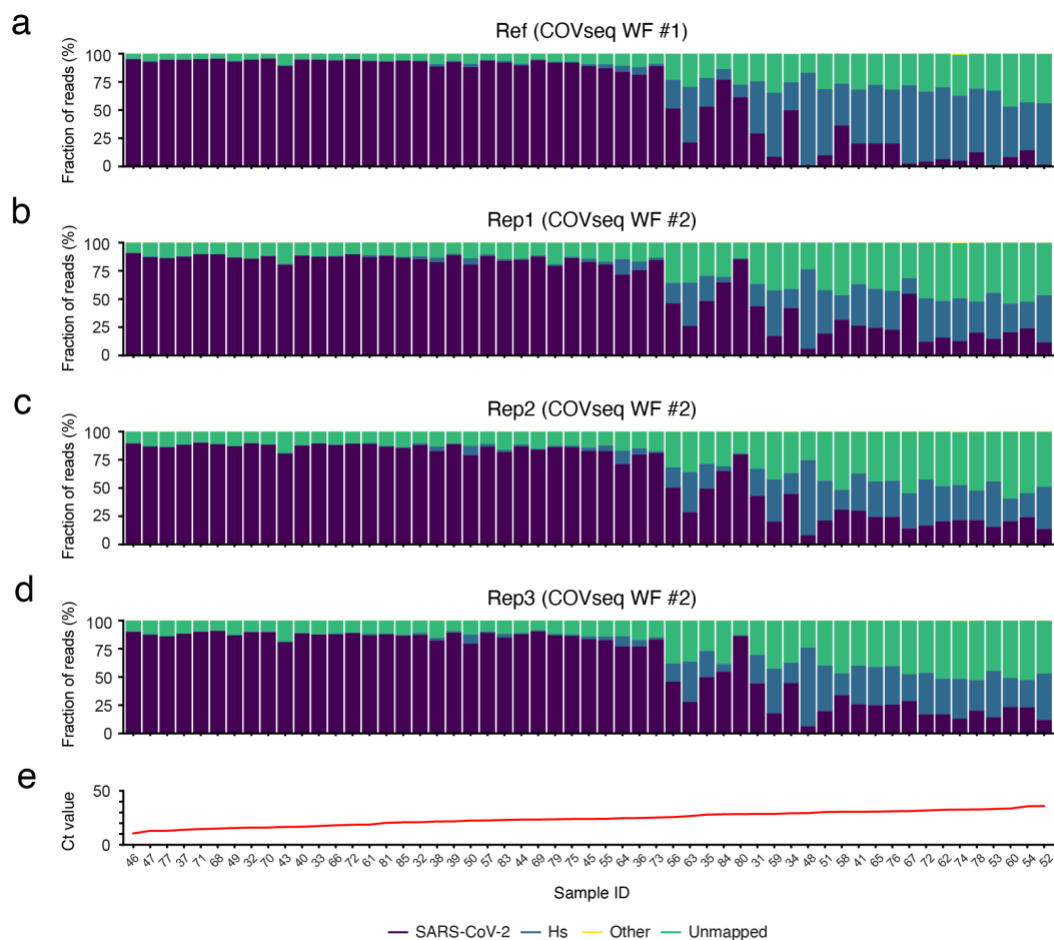
Supplementary Figure 5

Supplementary Fig. 5. (**a-d**) Percentage of sequencing reads aligned to the SARS-CoV-2 genome, human genome (Hs), other genomes (Other) or unmapped, in the reference (Ref) and three replicate (Rep) libraries prepared from 55 SARS-CoV-2 positive left-over RNA samples (samples 31–85 in **Supplementary Table 4**). WF, COVseq workflow (see **Methods**). (**e**) Cycle threshold (Ct) values of the 55 samples shown in (a-d). Sample IDs are the same as in **Supplementary Table 4**.

Supplementary Figure 6
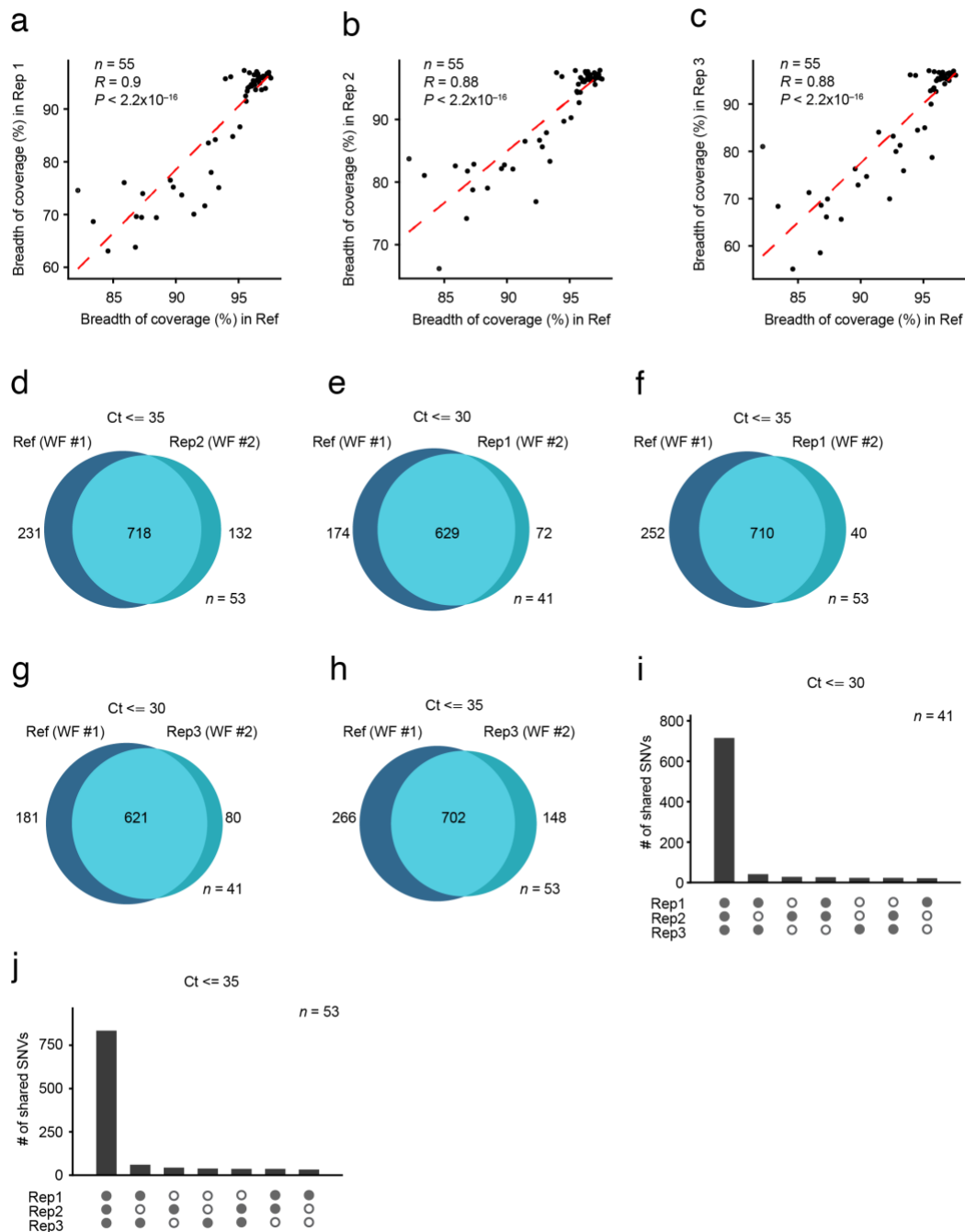
Supplementary Fig. 6. (**a-c**) Correlation between the breadth of coverage in the reference (Ref) and three replicate (Rep) libraries prepared from 55 SARS-CoV-2 positive left-over RNA samples (samples 31–85 in **Supplementary Table 4**). Each dot represents one sample. Dashed red lines: linear regression fit. *R*: Pearson's correlation coefficient. *P*: t-test, two-tailed. (**d-h**) Venn diagram showing the extent of overlap between the SNVs identified in samples with Ct either ≤ 30 or ≤ 35 included in the same Ref and Rep libraries shown in (a-c). (**i-j**) Barplot showing the number of SNVs shared between different Rep libraries.

Supplementary Figure 7
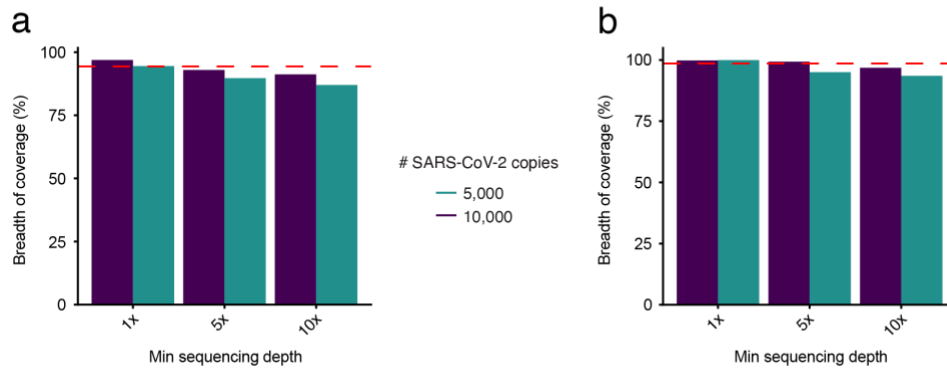
Supplementary Fig. 7. Proof-of-principle experiment using the COVseq workflow #3 described in the **Methods** and MseI and NlaIII in combination. (**a**) Breadth of coverage of the SARS-CoV-2 reference genome at varying sequencing depths, for two different input amounts of synthetic SARS-CoV-2 RNA (5,000 and 10,000 copies). The dashed red line represents the theoretical coverage at 1×. (**b**) Same as in (a), but for the S gene.

Supplementary Figure 8

a Influenza A

b Influenza B
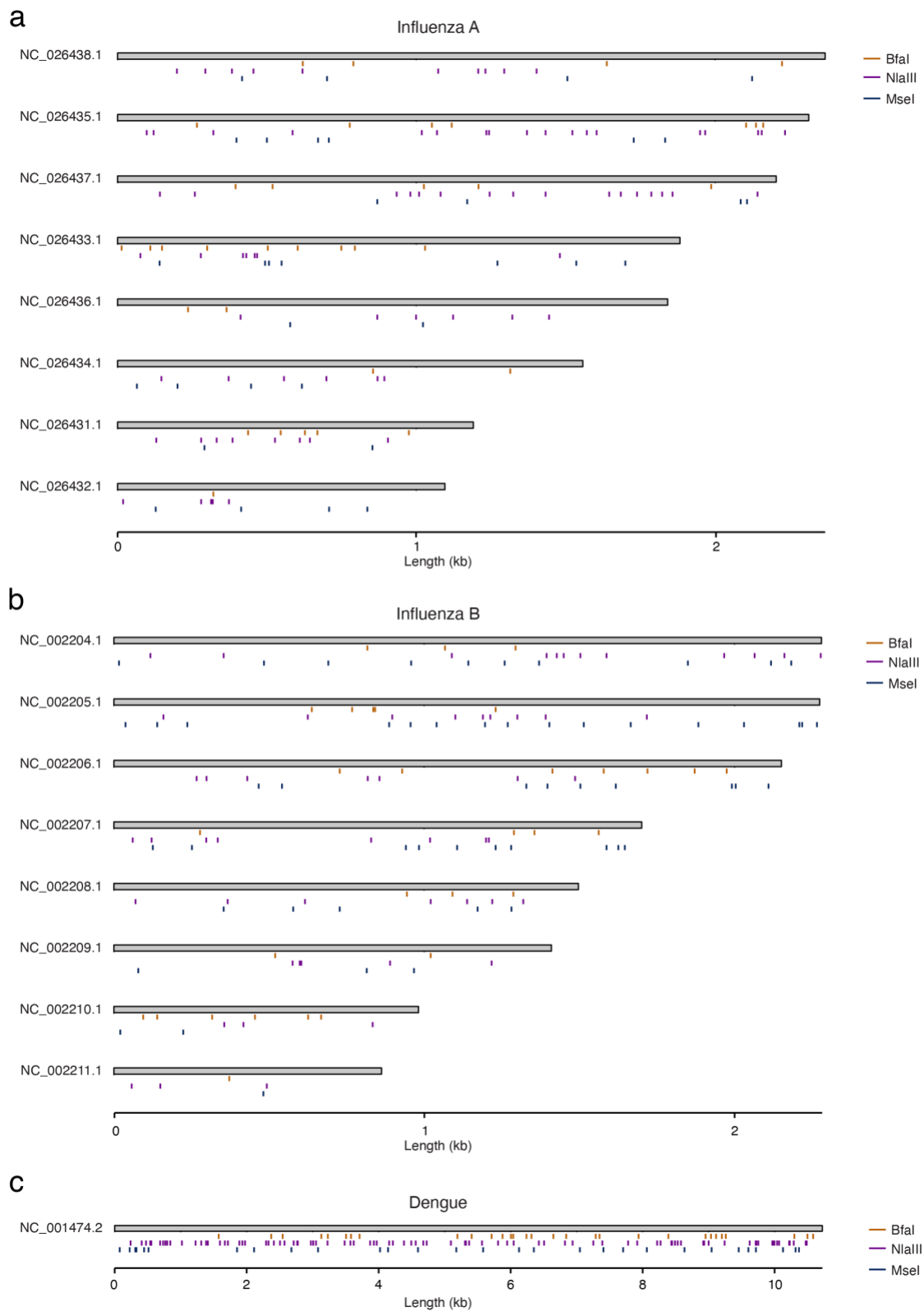
c Dengue

Supplementary Fig. 8. Locations of BfaI, MseI, NlaIII recognition sites (vertical colored bars) along the genome of (**a**) the H1N1 strain of Influenza type A, (**b**) Influenza type B and (**c**) Dengue virus. Reference NCBI accession numbers are indicated near each plot. Combining

these enzymes with virus-specific multiplexed PCR assays (see **Supplementary Table 6**), would expand the applications of COVseq in genomic epidemiological surveillance.

2. Supplementary Methods

Step-by-step COVseq protocol

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)

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

- 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)

- COVseq oligonucleotide adapters (see **Supplementary Table 2**)

- 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 with Scalable throughput (Beckman Coulter, cat. no. A63987)

- Agencourt AMPure XP (Beckman Coulter, cat. no. A6388§)

- 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)

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. 25 °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

Multiplex PCR

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.

7. 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

8. Add 3 µl of cDNA to each tube

9. Perform the following steps in PCR thermocycler with the lid set @ 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

Note: This is an optional step. For the continuation even unpurified material can be used.

10. Pool 20 µl from each of the six amplicon pools into a 1.5 ml LoBind tube

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

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

13. Place the sample on a magnetic stand

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

15. Remove and discard the supernatant

16. Wash the beads twice with 200 µl of freshly prepared 80% ethanol

17. 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

18. Remove the sample from the magnetic stand

19. Resuspend the beads in 80 µl of nuclease-free water

20. Incubate for 2 min at room temperature

21. Place the sample back on the magnetic stand

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

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

24. Check the library concentration using Qubit dsDNA BR kit

Note: Samples can be stored for long time @ -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 volumes of each reagent. For this step one can use unpurified material coming from the multiplex PCR after pooling an equal volume from each of the six pools. Other dispensing systems may also be used; however, volumes might have to be adjusted depending on the technical specifications of each instrument.

DNA digestion

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

26. Dispense 50 nL of purified or 100 nL of non-purified PCR amplicons

27. 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

28. Mix the following components:

NlaIII enzyme	50 nL

MseI enzyme	50 nL
10x CutSmart Buffer	50 nL

29. Dispense 150 nL per well

30. Perform the following incubation steps:

1. 37 °C	1 h
2. 65 °C	20 min
3. 4 °C	Hold

Ligation of COVseq adapters

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

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

33. Dispense 700 nL of ligation mix per well:

Nuclease-free water	250 nL
10x T4 ligase buffer	150 nL
ATP 10 mM	120 nL
BSA 50 mg/ml	30 nL
T4 standard ligase	150 nL

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

35. Dispense manually 5 µl of Nuclease-Free Water per well

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

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

Note: Pooling step can be performed by centrifuging the plate upside down into a collection plate placed at the bottom at 800 rpm for 1min.

DNA cleanup

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

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

40. Place the sample on a magnetic stand

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

42. Remove and discard the supernatant

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

44. 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

45. Remove the sample from the magnetic stand

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

47. Incubate for 2 min at room temperature

48. Place the sample back on the magnetic stand

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

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

51. Check the library concentration using Qubit dsDNA HS kit

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

In vitro **transcription**

52. Start with 8 μl from the previous step

53. 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

54. 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

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

56. Incubate for 15 min @ 37 °C

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

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

59. Place the sample on a magnetic stand

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

61. Remove and discard the supernatant

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

63. 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

64. Remove the sample from the magnetic stand

65. Resuspend the beads in 9 μl of nuclease-free water

66. Incubate for 2 min at room temperature

67. Place the sample back on the magnetic stand

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

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

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

RA3 adapter ligation

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

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

73. Add 3.2 µl of the following mix:

RNA ligase buffer	1.2 µl
RNaseOUT™ Recombinant Ribonuclease Inhibitor	1 µl
T4 RNA ligase truncated	1 µl

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

Reverse transcription

75. Add 2 µl per sample of RTP primer

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

77. Quickly transfer the sample to ice

78. Add 11 µl of the following mix:

5x SSIV buffer	5 µl
12.5mM dNTPs	1 µl
0.1M DTT	2 µl
RNaseOUT™ Recombinant Ribonuclease Inhibitor	1 µl
SuperScript IV reverse transcriptase	2 µl

79. 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

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

81. Add 359 µl of the following mix:

Nuclease-free water	143 µl
NEBNext® Ultra™ II PCR Master Mix	200 µl
RP1 primer	16 µl

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

83. 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

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

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

86. Place the sample on a magnetic stand

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

88. Remove and discard the supernatant

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

90. 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

91. Remove the sample from the magnetic stand

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

93. Incubate for 2 min at room temperature

94. Place the sample back on the magnetic stand

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

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

97. Check the library concentration using Qubit dsDNA HS kit

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

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

3. Supplementary Tables

Supplementary Table 1. List of primers used in the CDC SARS-CoV-2 multiplexed PCR assay. Due to its large size, the table is provided as a separate Excel file.

Supplementary Table 2. Summary of sequencing results. Due to its large size, the table is provided as a separate Excel file.

Supplementary Table 3. List of oligonucleotides that can be used to prepare COVseq adapters. Due to its large size, the table is provided as a separate Excel file.

Supplementary Table 4. List of samples and corresponding Ct values. AS: samples collected at the 'Amedeo di Savoia' Hospital in Turin, Italy during the Phase I of the 2020 pandemic (diagnostic samples). CCI: samples collected at the Candiolo Cancer Institute in Turin, Italy during the Phase II of the pandemic (screening samples).

Sample ID	Institution	Sampling date	Target region and corresponding Ct value			
			ORF1ab	S	E	N
1	AS	31.03.2020	40	/	/	43
2	AS	31.03.2020	/	/	/	37
3	AS	31.03.2020	39	/	/	/
4	AS	01.04.2020	25	/	23	23
5	AS	31.03.2020	23	/	20	20
6	AS	31.03.2020	31	/	29	30
7	AS	31.03.2020	26	/	23	23
8	AS	31.03.2020	30	/	30	30
9	AS	31.03.2020	21	/	17	17
10	AS	31.03.2020	17	/	15	14
11	AS	31.03.2020	27	/	24	25
12	AS	31.03.2020	34	/	32	34
13	AS	31.03.2020	29	/	26	26
14	AS	31.03.2020	27	/	24	25
15	AS	01.04.2020	32	/	30	31
16	AS	01.04.2020	35	/	32	34
17	AS	01.04.2020	39	/	/	/
18	AS	31.03.2020	37	/	34	36
19	AS	01.04.2020	21	/	18	19
20	AS	01.04.2020	28	/	25	27
21	AS	31.03.2020	/	/	43	/
22	AS	08.04.2020	19	/	15	15
23	AS	08.04.2020	31	/	27	27

24	AS	08.04.2020	20	/	16	16
25	AS	08.04.2020	18	/	15	15
26	AS	08.04.2020	33	/	29	30
27	AS	08.04.2020	39	/	35	36
28	AS	08.04.2020	38	/	33	34
29	AS	07.04.2020	40	/	/	35
30	AS	07.04.2020	40	/	/	41
31	CCI	17.10.2020	28	28	/	29
32	CCI	17.10.2020	15	15	/	16
33	CCI	17.10.2020	17	16	/	17
34	CCI	17.10.2020	29	29	/	29
35	CCI	17.10.2020	27	26	/	28
36	CCI	17.10.2020	24	24	/	25
37	CCI	17.10.2020	13	14	/	14
38	CCI	17.10.2020	20	19	/	22
39	CCI	17.10.2020	20	19	/	22
40	CCI	17.10.2020	16	16	/	17
41	CCI	17.10.2020	39	32	/	31
42	CCI	17.10.2020	19	19	/	21
43	CCI	17.10.2020	16	19	/	17
44	CCI	17.10.2020	22	22	/	23
45	CCI	17.10.2020	23	23	/	24
46	CCI	17.10.2020	13	13	/	13
47	CCI	17.10.2020	9	10	/	10
48	CCI	17.10.2020	29	29	/	29
49	CCI	17.10.2020	14	14	/	16
50	CCI	17.10.2020	22	22	/	22
51	CCI	02.11.2020	36	28	/	30
52	CCI	02.11.2020	/	/	/	36
53	CCI	02.11.2020	35	36	/	33
54	CCI	02.11.2020	/	/	/	36
55	CCI	02.11.2020	22	22	/	24
56	CCI	02.11.2020	25	25	/	26
57	CCI	02.11.2020	22	22	/	23
58	CCI	02.11.2020	30	29	/	31
59	CCI	02.11.2020	29	22	/	29
60	CCI	02.11.2020	/	/	/	34
61	CCI	02.11.2020	18	18	/	19
62	CCI	02.11.2020	33	37	/	32
63	CCI	02.11.2020	27	27	/	27
64	CCI	02.11.2020	23	22	/	25
65	CCI	02.11.2020	30	30	/	31
66	CCI	02.11.2020	18	18	/	18
67	CCI	02.11.2020	32	31	/	31
68	CCI	02.11.2020	14	15	/	15
69	CCI	02.11.2020	22	22	/	23
70	CCI	02.11.2020	15	15	/	16

71	CCI	02.11.2020	13	14	/	15
72	CCI	02.11.2020	37	37	/	32
73	CCI	02.11.2020	25	24	/	25
74	CCI	02.11.2020	36	32	/	33
75	CCI	02.11.2020	22	22	/	24
76	CCI	02.11.2020	30	29	/	31
77	CCI	02.11.2020	13	13	/	13
78	CCI	02.11.2020	37	37	/	33
79	CCI	02.11.2020	23	23	/	24
80	CCI	02.11.2020	28	28	/	28
81	CCI	02.11.2020	20	20	/	20
82	CCI	02.11.2020	18	18	/	19
83	CCI	02.11.2020	22	22	/	23
84	CCI	02.11.2020	27	27	/	28
85	CCI	02.11.2020	20	20	/	21

Supplementary Table 5. List of reagents and relative costs for sequencing SARS-CoV-2 samples using COVseq *vs.* three commercially available library preparation kits. Three separate lists are provided for COVseq, based on the three workflows discussed in the **Supplementary Note 1**. Due to its large size, the table is provided as a separate Excel file.

Supplementary Table 6. Available multiplexed PCR assays for WGS of other viruses in addition to SARS-CoV-2.

Assay	Target	Weblink
CDC Influenza SARS-CoV-2 (Flu SC2) Multiplex Assay	SARS-CoV-2, Influenza A and Influenza B viruses	https://www.fda.gov/media/139743/download
CDC DENV-1-4 rRT-PCR multiplex assay	Dengue virus	https://www.cdc.gov/dengue/healthcare-providers/testing/molecular-tests/realtime.html#anchor_1556657754547

Supplementary Table 7. List of reference sequences used for alignment and phylogenetic analyses.

Genome	Reference
SARS-CoV-2	NC_045512.2
Human	GRCh38
H1N1, Influenza type A	NC_026431.1 - NC_026438.1
Influenza type B	NC_002204.1 - NC_002211.1
Dengue	NC_001474.2
Adapters	Default FastQ-Screen reference
Arabidopsis	Default FastQ-Screen reference
Drosophila	Default FastQ-Screen reference
E Coli.	Default FastQ-Screen reference

Lambda	Default FastQ-Screen reference
Mitochondria	Default FastQ-Screen reference
Mouse	Default FastQ-Screen reference
PhiX	Default FastQ-Screen reference
Rat	Default FastQ-Screen reference
RRNA	Default FastQ-Screen reference
Vectors	Default FastQ-Screen reference
Worm	Default FastQ-Screen reference
Yeast	Default FastQ-Screen reference

4. Supplementary Notes

Cost analysis

To demonstrate the cost effectiveness of COVseq, we examine three different COVseq workflows and compare them with three widely used commercial kits for preparing libraries directly from SARS-CoV-2 RNA (TruSeq Stranded Total RNA Library Prep kit, Illumina cat. no. 20020596) or from purified amplicons (NEBNext Ultra II FS DNA Library Prep Kit, NEB cat. no. E7805S; Nextera XT DNA Library Preparation Kit, Illumina cat. no. FC-131-1024). For commercial kits, we assume to purchase the kit with the highest number available of library indexes (96 for TruSeq and NEBNext, 384 for Nextera), so that many samples can be sequenced together (see considerations on sequencing below). A detailed list of reagents, volumes per reaction, number of reactions and current prices (as of Dec 2020) is available in **Supplementary Table 5**. For simplicity, we omit the cost of plasticware and other consumables (pipette tips, gloves, etc.) from our analysis. To simulate mass-scale SARS-CoV-2 sequencing performed by a centralized laboratory or a public health agency, we compute how the cumulative reagent cost grows with increasing the number of samples processed up to 100,000. The simulation does not include sequencing costs, which are discussed separately below. The simulation is run through a custom script written in MATLAB, which we make available upon request.

1. Comparison between different COVseq workflows and formats

We consider three different scenarios, depending on which step in the COVseq protocol is performed using standard reaction volumes (microliter range) or nanoliter volumes. For the latter, we assume to use the I-DOT One contactless nanodispensing device (Dispendix GmBH), which we previously described for high-throughput CUTseq[1]. However, any other contactless device with similar characteristics should be equally effective.

- Workflow #1: (i) reverse transcription (RT) and multiplexed PCR are done using standard volumes; (ii) purification of the PCR products; (iii) CUTseq using I-DOT.
- Workflow #2: (i) RT and multiplexed PCR using standard volumes; (ii) no purification of the PCR products; (iii) CUTseq using I-DOT.
- Workflow #3: all reactions until IVT are done in nanoliter volumes using I-DOT. Since, due to logistic reasons, we could not transfer our I-DOT machine into a biosafety level-2 (BSL-2) laboratory, we could only test this scenario using synthetic SARS-CoV-2

RNA, which can be handled in a standard BSL-1 lab. The results of this test are shown in **Supplementary Fig. 6a and b**.

In all aforementioned workflows, we assume to use a combination of two restriction enzymes, MseI and NlaIII, to digest the SARS-CoV-2 genome. For each workflow, we consider two scenarios: (i) one case in which 96 samples are pooled into the same library; and (ii) one case in which 384 samples are pooled together. In the former, 96 MseI and 96 NlaIII adapters, each containing a different sample barcode sequence, must be purchased upfront, whereas in the latter case a total of $384 \times 2 = 768$ adapters are needed. As shown in **Fig. 2c**, the cumulative cost curves are very similar for workflow #1 and #2, with the latter workflow in combination with the use of 384 adapters being more cost-effective ($16.54 per sample to prepare libraries from 100,000 samples with workflow #2 and 384 adapters *vs. vs.* $20.28 per sample using workflow #1 and 96 adapters, respectively). The highest cost effectiveness is achieved with workflow #3 and 384 adapters provides ($2.76 and $1.90 per sample for 10,000 and 100,000 samples, respectively). Thus, if a nanodispensing device can be placed into a BSL-2 environment, COVseq workflow #3 provides the most cost-effective solution for mass-scale sequencing of SARS-CoV-2 genomes. In principle, a larger repertoire of adapters with different sample barcodes could be designed in order to pool more samples into the same library. Such high-level multiplexing is routinely performed in single-cell RNA sequencing assays, such as DROP-seq[2]. Therefore, we do not see any obstacle to adapting the barcodes used in DROP-seq or other highly multiplexed sequencing assays to expand the repertoire of COVseq adapters.

2. Comparison between COVseq and available commercial solutions

As shown in **Fig. 2d**, the cumulative cost curves for preparing sequencing libraries from SARS-CoV-2 samples using existing commercial kits are all above cumulative COVseq curves, independently of the workflow used. Illumina's TruSeq kit represents the most expensive solution, with an estimated cost per sample three orders of magnitude higher than the corresponding COVseq cost assuming to use workflow #3, 384 adapters and processing 100,000 samples ($3,103 *vs.* $1.90 per sample). The main reason for this dramatic cost discrepancy in favor of COVseq (independently of the workflow chosen) is that commercial kits are specifically designed to make a single library from each individual sample, whereas in COVseq multiple samples are pooled together into the same library, therefore allowing for a drastic reduction of the reagent costs required for library preparation. In principle, 'nano' versions of commercial library preparation kits could be implemented on I-DOT or other similar nanodispensing devices, in order to reduce the volume, of each reagent, and thus the cost per

sample. However, even by drastically reducing the volume of reagents used in commercial kits, the cost per sample would remain considerably higher compared to COVseq, due to the fact that only a limited number of libraries (and therefore samples) can be sequenced together in the same run, as discussed below.

3. Considerations on the sequencing platform and number of samples sequenced together

In this work, we have sequenced all our samples on the NextSeq 500 platform from Illumina. Therefore, for the purpose of this cost analysis, we assume to use this platform to compare the cost of sequencing COVseq libraries *vs.* libraries obtained using commercial kits, as discussed above. In general, the number of samples that can be sequenced together depends: (i) on the desired depth of coverage per sample, which in turn influences the subsequent ability to confidently call SNVs and other variants; and (ii) on the number of libraries that can be sequenced together. Assuming to use a commercial kit with 96 library indexes, at most 96 libraries can be sequenced together in the same sequencing run. Assuming to use the NextSeq 500/550 High Output Kit v2.5 (150 cycles) (Illumina, cat. no. 20024907)—which typically yields around 400 million reads—this would result in approx. 4 million reads per sample (assuming all the libraries to be well balanced) with a sequencing cost of ~\$40 per sample. Based on our results, this figure is one order of magnitude larger than the minimum number of reads (~400,000) required to cover the SARS-CoV-2 genome at an average depth (~1500×) sufficient to confidently call SNVs. Notably, when preparing libraries from individual samples, the concentration of each library must be accurately quantified before pooling multiple libraries into the same sequencing run, which further increases the cost per sample. In contrast, thanks to the fact that multiple samples are pooled into the same library, the number of samples that can be effectively sequenced together drastically increases in COVseq. For example, using 384 COVseq adapters, up to 1,152 samples (3 libraries) can be sequenced together in the same NextSeq 500 run, yielding ~347,000 reads per sample at a cost of ~\$4 per sample. This cost could be further lowered by combining higher multiplexing (using more adapters as discussed above) with higher throughput platforms, such as Illumina's NovaSeq 6000. In conclusion, based on our cost estimates, using COVseq workflow #3 with 384 adapters and 1,152 samples sequenced per NextSeq 500 SE150 run would allow sequencing 100,000 samples at a cost of less than \$6. To our knowledge, this makes COVseq the most cost-efficient solution for mass-scale genomic surveillance of SARS-CoV-2 currently available.

5. Supplementary References

1. Zhang, X. *et al.* CUTseq is a versatile method for preparing multiplexed DNA sequencing libraries from low-input samples. *Nat. Commun.* **10**, 4732 (2019).

2. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).