HiDRA-seq: High-Throughput SARS-CoV-2 Detection by RNA Barcoding and Amplicon Sequencing — Source link [↗]

Emilio Yángüez, Griffin White, Susanne Kreutzer, Lennart Opitz ...+5 more authors

Institutions: University of Zurich

Related papers:

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

- LAMP-Seq: Population-Scale COVID-19 Diagnostics Using Combinatorial Barcoding

- A rapid, low cost, and highly sensitive SARS-CoV-2 diagnostic based on whole genome sequencing

- CRISPR-Cas12-based detection of SARS-CoV-2.

- Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands.

1 HiDRA-seq: High-Throughput SARS-CoV-2 Detection by RNA Barcoding and

2 Amplicon Sequencing

3   Emilio Yánguez[*,1,#], Griffin White[*,1], Susanne Kreutzer[1], Lennart Opitz[1], Lucy Poveda[1], Timothy Sykes[1], Maria

4   Domenica Moccia[1], Catharine Aquino[1] and Ralph Schlapbach[1].

5

6   [1] Functional Genomics Center Zurich (ETH/University of Zurich), Zurich, Switzerland.

7   * These authors contributed equally to this work.

8   [#] Correspondence to: emilio.yanguez@fgcz.ethz.ch.

9

10 Abstract

11   The recent outbreak of a new coronavirus that causes a Severe Acute Respiratory Syndrome in humans

12   (SARS-CoV-2) has developed into a global pandemic with over 6 million reported cases and more than

13   375,000 deaths worldwide. Many countries have faced a shortage of diagnostic kits as well as a lack

14   of infrastructure to perform necessary testing. Due to these limiting factors, only patients showing

15   symptoms indicating infection were subjected to testing, whilst asymptomatic individuals, who are

16   widely believed to be responsible for the fast dispersion of the virus, were largely omitted from the

17   testing regimes. The inability to implement high throughput diagnostic and contact tracing strategies

18   has forced many countries to institute lockdowns with severe economic and social consequences. The

19   World Health Organization (WHO) has encouraged affected countries to increase testing capabilities

20   to identify new cases, allow for a well-controlled lifting of lockdown measures, and prepare for future

21   outbreaks. Here, we propose HiDRA-seq, a rapidly implementable, high throughput, and scalable

22   solution that uses NGS lab infrastructure and reagents for population-scale SARS-CoV-2 testing. This

23   method is based on the use of indexed oligo-dT primers to generate barcoded cDNA from a large

24   number of patient samples. From this, highly multiplexed NGS libraries are prepared targeting SARS-

25   CoV-2 specific regions and sequenced. The low amount of sequencing data required for diagnosis

26   allows the combination of thousands of samples in a sequencing run, while reducing the cost to

27   approximately 2 CHF/EUR/USD per RNA sample. Here, we describe in detail the first version of the

28   protocol, which can be further improved in the future to increase its sensitivity and to identify other

29   respiratory viruses or analyze individual genetic features associated with disease progression.

30

31 Keywords

32   SARS-CoV-2, Coronavirus, Next-Generation Sequencing (NGS), diagnostics, testing.

33

Introduction

Over 350,000 deaths worldwide have resulted from complications resulting from Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection with another 6 million reported infected by this virus, causing a major pandemic. The overall public health and economic burden of this pandemic has yet to be realized and will only become apparent in the coming years. Switzerland alone has recorded one of the highest numbers of COVID-19 (the disease caused by caused by SARS-CoV-2) cases per capita in the world[1]. Thus, efficient and sensitive detection assays of SARS-CoV-2 are essential in managing this pandemic, as evidence suggests that the virus is most contagious on or before symptom onset. Furthermore, asymptomatic cases and the so-called "super spreaders" have broadly contributed to the dissemination of the virus, as reports from South Korea suggest[2]. Comprehensive contact tracing, tracking the spread of viral transmission, clearly requires an increase of mass testing to a population scale.

The vast majority of the currently available SARS-CoV-2 diagnostic assays are based on the amplification of specific loci in the viral genome through Real-Time Quantitative Polymerase Chain Reaction (RT-qPCR). RT-qPCR assays have been the gold standard in clinical diagnostics due to their high sensitivity. However, this outbreak has demonstrated that there exists a general lack of infrastructure for such population-scale testing, in addition to a limited supply of reagents for RT-qPCR tests. Furthermore, the outcome of RT-qPCR is a binary result (positive or negative) for the loci interrogated. The main drawback being the lack of genotypic information in RT-qPCR assays, which could enable the mapping of the spread and transmission, as well as the monitoring of the evolution of the etiological agent, which is crucial for vaccine development.

NGS has revolutionized biomedical research in the last 15 years and is increasingly impacting clinical diagnostics and the practice of medicine. Our aim was to develop a diagnostic assay that could profit from the power and sensitivity of these technologies, not only for viral detection but also for viral classification. In this study, we present HiDRA-seq, a low-cost, high-throughput targeted approach for SARS-CoV-2 infection diagnosis and for potentially tracing outbreak origin and tracking transmission. In recent months, several protocols have been developed for SARS-CoV-2 diagnosis using NGS. However, most of the proposed methods are based on the amplification of the entire viral genome, which is time consuming, rather expensive, and require specific kits for enrichment (amplicon or hybridization based) as well as a comprehensive de novo pipeline to analyze the data and advanced bioinformatic pipelines[3]. We propose a midway and rapidly implementable option that uses genomics lab infrastructure and reagents available in large NGS facilities, bypassing the need for commercially available kits and the limitations in the global chain of production. HiDRA-seq combines reverse transcription using barcoded oligo-dT primers, adapted from mcSCRB-seq[4], with the addition of virus-

2

68 specific amplicon generation and sequencing. The protocol targets a region of the putative ORF10,

69 which is highly conserved in the different SARS-CoV-2 isolates sequenced to date. The targeted region

70 is located near the 3'-end of the genome[5,6], so HiDRA-seq can capture both the viral genomic RNA

71 (gRNA) as well as all subgenomic RNA transcripts (sgRNA) generated in infected cells.

72 We would like to emphasize that our approach consists of a viral enrichment, followed by the

73 generation of a small amount of short read sequencing data and the use of a basic bioinformatics

74 pipeline for downstream mapping and diagnosis. The small amount of short read sequencing data

75 required to correctly diagnose an individual allows the multiplexing of hundreds to thousands of

76 patients in one sequencing run and is an affordable reality at a price of approximately 2 CHF/EUR/USD

77 per sample (from extracted RNA to diagnosis). Furthermore, HiDRA-seq can be adapted by a wide

78 variety of short read sequencers with diverse outputs across the clinics. The nature of the enrichment

79 step in this protocol offers an enormous versatility, as it can be tailored to any other respiratory virus

80 or organism of interest that produces poly-adenylated transcripts. The implementation of a rapid and

81 versatile approach such as HiDRA-seq would definitely enable a more efficient outbreak

82 management.

83

84 **Results**

85 **Protocol description**

86 Since the beginning of the SARS-CoV-2 pandemic, various research groups have been working on the

87 development of alternative testing protocols to bypass the shortage of standard diagnostic kits and

88 enable population scale testing. With this idea, we have developed HiDRA-seq, a high throughput,

89 rapidly implementable solution that uses standard lab infrastructure and reagents in medium sized

90 NGS facilities (**Figure 1A**). The reverse transcription and cDNA pooling strategies are adapted from the

91 mcSCBR-seq protocol[4]. Using a low volume liquid handling robot, the patients' RNA (previously

92 extracted from swab samples) is distributed in 384-well plates containing indexed oligo-dT primers. A

93 short reverse transcription is performed to generate barcoded cDNA and the contents of each well

94 are pooled into a single tube. Following bead purification and exonuclease treatment to digest the

95 excess of unbound primers, the pooled cDNA is used as template in a PCR reaction with a forward

96 primer specific to SARS-CoV-2 and a reverse primer binding to a sequence incorporated with the oligo-

97 dT in the reverse transcription. Both primers contain the sequences required to generate an Illumina

98 sequencer compatible library in a final PCR reaction, in which the amplicon pool can be barcoded,

99 which allows the multiplexing of multiple 384-well plates in a single sequencing run. After sequencing,

100 the reads are de-multiplexed based on both the plate and the sample barcodes and used in an

101   automated analysis pipeline to distinguish samples containing SARS-CoV-2-derived reads, which are

102   diagnosed as positive.

103

104   **SARS-CoV-2-specific amplicon design**

105   We designed three different partially over-lapping amplicons in order to compare their performance

106   and, simultaneously, simulate the combination of three different patient plates that need to be de-

107   multiplexed upon sequencing. For the amplicon design, highly conserved regions between SARS-CoV-

108   2 isolates were identified by creating a whole genome alignment of the European SARS-CoV-2

109   sequences (n = 1435, sequence identity = 98.96%). The identified sequences are not conserved in

110   other human coronavirus. As cDNA is barcoded using indexed oligo-dT, the forward primers must be

111   located close to the poly-A sequence in the viral genome to obtain amplicons of a reasonable size.

112   Moreover, primers binding to the 3' of the genome can efficiently capture both the viral genomic RNA

113   (gRNA) as well as all subgenomic RNA transcripts (sgRNAs) generated in infected cells, which may

114   increase the sensitivity of the method. With these premises, we identified a 100% conserved region

115   in the putative ORF10 of SARS-CoV-2. The GC content in this region is highly variable, so primers were

116   designed to target sequences of lower GC contents (<50%). We designed three forward primers

117   targeting this region that are used to generate three SARS-CoV-2 specific amplicons (**Figure 1B**). The

118   three amplicons were tested separately in order to compare their performance. For the human

119   internal control, we followed a similar strategy and designed a forward primer located in close

120   proximity to the poly-A sequence of human GAPDH (Figure 1B). This ensures that a fragment that can

121   be sequenced is always generated even in SARS-CoV-2 negative samples. The reverse primer is

122   common to all amplicons and it anneals to the 3'-end of the barcoded oligo-dT used to generate the

123   cDNA in the reverse transcription.

124

125   **Initial data quality control**

126   For the implementation of the protocol, 91 anonymized clinical samples containing RNA extracted

127   from patients' swaps were kindly provided by Dr. Michael Huber and Dr. Jürg Böni (Institute of Medical

128   Virology, UZH, Switzerland). These clinical samples were transferred to four identical quadrants in a

129   384-well plate, such that each sample would be processed in quadruplicate for each of the three SARS-

130   CoV-2 amplicons designed and tested for the study. The samples were processed using a Mosquito HV

131   liquid pipetting robot (SPT Labtech), as indicated in materials and methods, and sequenced in both an

132   Illumina MiniSeq (R1=16 bp, i7=8 bp, R2=50 bp) and NovaSeq6000 sequencers (data not shown, R1=16

133   bp, i7=8 bp, R2=150 bp). The 8 bases barcode was used, in this case, to discriminate the different

134   amplicons but it could alternatively be used to de-multiplex several patient plates. The first 6 bases in

135 read one were used to demultiplex the reads generated from the different patient samples. UMI

136 correction was not applied in this version of the protocol. The 50 bases in read 2 were used to

137 distinguish between positive and negative patients by mapping reads against the SARS-CoV-2 genome.

138 After demultiplexing and prior to mapping, the dataset was subjected to standard quality control

139 checks and data filtering. As the input for the protocols was not normalized, a wide distribution in the

140 number of reads per sample was observed (**Figure 2A**). Reads were filtered by abundance per

141 sample/amplicon combination (n < 250 for one sample/amplicon combination). 250 reads per

142 sample/amplicon combination (750 reads per sample) was sufficient to minimize the minor effects of

143 index hopping on the SARS-CoV-2 alignment rate for a given sample, and accurately represent the

144 abundance of SARS-CoV-2-mapping reads in a given samples' data. Based on the RT-qPCR diagnostics

145 results, samples for which GAPDH was undetectable after 40 cycles of RT-qPCR were removed, and

146 finally 83 patient samples in technical quadruplicates remained.

147 The alignment of reads to SARS-CoV-2 and GAPDH was performed using Bowtie2 using end-to-end

148 mode. More than 99.9% of mapped reads mapped uniquely to the targeted locus, indicating successful

149 primer design (**Figure 2B**). The global alignment rate of the reads was of 26.96%, 43.47% and 36.69%

150 for amplicons 29652, 29691, and 29709, respectively.

151

152 **Diagnostic capability of the assay**

153 Ct values for sets of RT-qPCR technical replicates were normalized and correlated to the alignment

154 rate using the Wilcoxon-Signed-Rank test for matched pairs (p<0.001 for each amplicon; see **Figure**

155 **3A**). The per sample alignment rate is shown to be highly correlated with RT-qPCR Ct value (see **Figure**

156 **3B**). Samples which returned a Ct value > 25 for SARS-CoV-2 in the diagnostic test, had their alignment

157 rates fall, typically within one standard deviation of the mean alignment rate, for the set of samples

158 diagnosed "negative" with RT-qPCR. This demonstrates that HiDRA-seq is thus far, incapable of

159 identifying with certainty positive samples with a RT-qPCR Ct value > 25. However, there were two

160 exceptions in which the HiDRA-seq system was able to successfully diagnose samples with Ct values >

161 37. Amplicon 29691 generated the results most consistent with the RT-qPCR diagnosis for Ct values

162 within the range 29 – 40 (see **Figure 3C**).

163 We generated potential diagnostic thresholds, based upon the alignment rate (AR) of a sample's reads

164 to the SARS-CoV-2 genome, by iterating through values on [0,1] in increments of 0.01. We counted for

165 each value, the number of positively diagnosed samples (via RT-qPCR) that had ARs below this value,

166 and the number of negatively diagnosed samples (via RT-qPCR) that had ARs above this value. We

167 then selected intervals for which the number of false diagnoses were minimized as a basis for a

168 theoretical diagnostic threshold. Potential diagnostic thresholds were set individually for each

169 amplicon as follows (see **Figure 4**): Amplicon 29652: AR ε [0.02 , 0.11]; Amplicon 29691: AR ε [0.05 ,

170 0.24]; Amplicon 29709: AR ε [0.04 , 0.18]. On the intervals of AR values for which the number of false

171 diagnoses was minimized, amplicons 29652 and 29709 mis-diagnosed 8 patients and Amplicon 29692

172 mis-diagnosed 7 patients. Furthermore, amplicons 29652, 29709 and 29692 had a successful

173 diagnostic rate with respect to RT-qPCR diagnosis of 90.4%, 90.4% and 91.6%, respectively.

174

175 ## Discussion

176 The surge in cases of SARS-CoV-2 infections around the world has created an urgent need for accurate

177 and fast diagnostic testing solutions. Despite the recent increase in available SARS-CoV-2 diagnostic

178 testing kits in the past few months, the majority of tests still rely on real-time quantitative PCR (RT-

179 qPCR). Whilst RT-qPCR reactions are generally very sensitive (*i.e.* able to detect true positive cases)

180 and specific**,** the technology has inherent limitations with regard to large scale population screening,

181 which has become increasingly important during this pandemic. Additionally, RT-qPCR does not

182 provide any genotypic information regarding a patient's infection beyond the causal organism.

183 Another advantage of NGS over RT-qPCR is that NGS provides a direct and functional measurement

184 SARS-CoV-2. RT-qPCR generates florescent measurements for individual plate wells, indirectly

185 quantifying the presence of genetic material in relative concentrations. Alternatively, NGS generates

186 thousands of reads, directly measuring the specific sequences present in a sample. The sequence

187 information generated could provide insight into the specific infecting isolate and aid in tracing

188 transmission within communities.

189 HiDRA-seq, built on Next Generation Sequencing technology, has the ability to multiplex thousands of

190 barcoded patient samples, significantly increasing current testing capacity. Our method is designed to

191 be partially performed on a small automated liquid handling machine, so that a single person is able

192 process more than 2,000 RNA samples per day with ease. This results in an overall shorter diagnostic

193 turnaround time, with library preparation and sequencing data obtained in as little as 1.5 days. Our

194 system does not match the speed of RT-qPCR for individual samples. However, it outperforms

195 standard diagnostics methods in scale, allowing one to process hundreds of thousands of samples per

196 week if extended and automated at scale. The faster a test can be administered, the sooner the results

197 can be received, and the quicker measures can be put in place to mitigate further spread or to evaluate

198 the impact of loosened containment measures. This system also minimizes the errors in sample

199 handling by fast-tracking sample preparation, an integral part of the workflow. The achievement of

200 consistent data across samples verifies the reproducibility and reliability of our system. Additionally,

201 the miniaturization of our reactions results in a much more affordable solution compared to other

202   methodologies available in the market. Our estimated price for sample screening from extracted RNA

203   to diagnostic result with our approach is 2 CHF/EUR/USD.

204   By comparing our results to the RT-qPCR-based clinical diagnostic test, we show that the mapping rate

205   is a strong predictor of Ct values (p<0.001). Our method is sensitive, as we have been able to correctly

206   recall positive samples in >90% of the cases, which is comparable to other NGS-based methods use

207   for virus detection in clinical samples[7]. The samples that escaped detection are characterized by

208   having high Ct values from the diagnostic RT-qPCR (Ct>25). In parallel to HiDRA-seq, we prepared

209   libraries using the highly sensitive Smart-seq2 protocol[8], which captures all poly-adenylated RNA in

210   the sample, and we were similarly unable to detect a significant number of virus-derived reads. The

211   sensitivity of HiDRA-seq could be improved by slightly increasing the number of PCR cycles used in the

212   amplicon generation and using UMI correction for detection of PCR duplicates to increase the

213   quantitative accuracy of the method. Although we were able to successfully diagnose the majority of

214   samples using 50 bp reads, the designed primers allowed us to access 71 bp of a highly variable region

215   (following Amplicon 29709) by generating reads of 150 bp, raising the possibility of using this method

216   for basic phylogenetic and epidemiological studies of isolates differing at this genomic position.

217   HiDRA-seq will be optimized to enable direct lysis from saliva collected by gargling. Given that the

218   mcSCRB-seq[4] method, from which this protocol derives, is designed to work with direct lysis from

219   single cells, this approach will be adapted for HiDRA-seq, since RNA extraction is one of the biggest

220   bottle-necks for large scale testing. The current sensitivity of our method makes it compatible with

221   direct testing from saliva, as suggested in a recent publication[9].

222   Our method uses barcoded oligo-dT primers to generate cDNA in the reverse transcription and this

223   feature leaves the door open to expanding the amplicon panel. To achieve this, additional PCR primers

224   could be added to generate amplicons that are specific for other human coronaviruses (hCoVs) or

225   other respiratory virus that produce polyadenylated transcripts. Such viruses include influenza viruses

226   (IAVs), respiratory syncytial viruses (RSVs), parainfluenza viruses (PIVs) or human metapneumoviruses

227   (MPVs), which would create a multi-viral identification test at almost no extra cost. Potentially, this

228   approach could also be implemented to specifically detect the expression of virtually any human

229   mRNA identified as a biomarker for estimating disease susceptibility and progression, or for designing

230   host group-specific COVID-19 treatment regimens. This is especially relevant in a clinical research

231   setup, in which the importance of a test that could both identify an infection and give information on

232   how to best treat that infection cannot be overstated.

233   This method has been designed with the practical necessities of large scale, affordable, adaptable and

234   rapid testing in mind. To these ends, we have developed a first version of a method that reuses

235   relatively common sets of barcoding primers available in NGS facilities, can scale effectively, does not

236 involve exotic reagents and relies on NGS to multiplex samples for both cost and time savings, allowing

237 any well-equipped sequencing lab in the world to quickly begin testing.

238

Materials and Methods

240 **Primer sequences**

Primer name	Sequence
Barcoded Oligo-dT	5'-Bio-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNNNNNT$_{30}$VN-3'
PCR	5'-Bio-ACACTCTTTCCCTACACGACGC-3'
SCoV2_29691	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCTCACATAGCAATCTTTAATCAGTG-3'
SCoV2_29709	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGTGTGTAACATTAGGGAGGAC-3'
SCoV2_29652	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGTAACTACATAGCACAAGTAGATG-3'
GAPDH_1127	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTCATTTCCTGGTATGACAACG-3'
i5 primer	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
i7 primer (Nextera XT)	5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTCTCGTGGGCTCGG-3'

241

242 **Reverse transcription with barcoded oligo-dT**

243 This part of the protocol is based on the reverse transcription strategy used in mcSCRB-seq (Bagnoli

244 et al., 2018). A 384-well barcoding plate was prepared using a Mosquito HV liquid handling robot (SPT

245 Labtech), each of the wells containing 0.5 µl of lysis solution (0.2% Triton X-100 [Roche], 0.8 units of

246 RNasin Plus [Promega], 4 mM dNTPs [Promega] and 1 µM of barcoded oligo-dT primers [E3V6NEXT,

247 IDT]). The plate was divided into four quadrants (91 samples each) with four identical copies of a 96-

248 well plate containing RNA extracted from anonymous patients' swaps (kindly provided by Dr. Michael

249 Huber and Dr. Jürg Böni, Institute of Medical Virology, UZH, Switzerland) by transferring 0.5 µl of RNA

250 to the corresponding well positions. The plate was heated at 65° C for 5 minutes and transferred to

251 ice prior to the addition of 1 ul of 2X Reverse transcription mix (15% PEG 8000 [Sigma Aldrich], 2X

252 Maxima RT buffer [Thermo Fischer] and 4 units of Maxima H Minus RT [Thermo Fisher]). cDNA

253 synthesis was performed for 15 min at 50°C followed by 5 min at 85° C for inactivation. The Mosquito

254 HV liquid handling robot (SPT Labtech) was used to pool the whole 384-well plate containing the

255 barcoded cDNA into a single 2 ml DNA LoBind tubes (Eppendorf) and cleaned up using Sera-Mag Select

256 beads (GE Healthcare) with a ratio 1:0.8 (pooled cDNA:beads). Purified cDNA was eluted in 17 µl and

257 residual primers digested with Exonuclease I (Thermo Fisher) for 20 min at 37 °C.

8

258 **SARS-CoV-2-specific amplicon generation**

259 Three different SARS-CoV-2-specific primers (SCoV2_29691, SCoV2_29709, SCoV2_29652) were used

260 in individual PCR reactions, in combination with a human GAPDH-specific primer (GAPDH_1127) and

261 a PCR primer binding to the barcoding sequence, to generate three virus specific amplicons. These

262 primers also contain the adaptor sequences needed to incorporate the Illumina compatible flow cell

263 binding sequences in a subsequent PCR reaction. Briefly, each PCR reaction was assembled by

264 combining 5 ul of the pooled barcoded cDNA with 20 µl of PCR master mix (1.25X KAPA HiFi HotStart

265 ReadyMix [Roche], 0.375 uM of SARS-CoV-2- and GAPDH-specific primers [Microsynth AG] and 0.375

266 uM of the PCR primer [Microsynth AG]). PCR was performed using the following program: 3 min at

267 98 °C for initial denaturation followed by 25 cycles of 20 sec at 98 °C, 15 sec at 60 °C, 15 sec at 72 °C.

268 Final elongation was performed for 5 min at 72 °C. Once the PCR was concluded, the amplicons were

269 cleaned up using Sera-Mag Select beads (GE Healthcare) with a ratio 1:0.8 (DNA:beads). The size and

270 concentration of the amplicons was analyzed in a 4200 TapeStation System (Agilent) using a D1000

271 ScreenTape.

272

273 **Library generation and plate barcoding incorporation**

274 A final PCR was performed from the amplicon to generate libraries compatible with Illumina

275 sequencer. In parallel, three different barcodes were assigned for the aforementioned amplicons in

276 order to combine them in a single sequencing run. This strategy can be used to combine different

277 plates with patient samples, allowing one to easily increase the throughput of the protocol. Each PCR

278 reaction was assembled by combining, in individual tubes, 5 ul of the three amplicons with 20 µl of

279 PCR master mix containing 1.25X KAPA HiFi HotStart ReadyMix (Roche), 0.375 uM of the i5 primer

280 (Microsynth AG) and 0.375 uM of the barcoded i7 primer (Illumina). PCR was performed using the

281 following program: 3 min at 98 °C for initial denaturation followed by 5 cycles of 20 sec at 98 °C, 15 sec

282 at 55 °C, 15 sec at 72 °C. A final elongation was performed for 5 min at 72 °C. Once the PCR was

283 completed, the libraries were cleaned up using Sera-Mag Select beads (GE Healthcare) with a ratio

284 1:0.8 (DNA:beads). The size and concentration of the libraries were analyzed in a 4200 TapeStation

285 System (Agilent) using a D1000 ScreenTape.
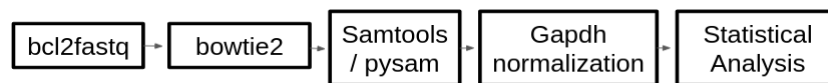
286

287 **Sequencing**

288 The three libraries generated from three different viral amplicons, all in combination with a GAPDH-

289 derived library, were paired-end sequenced together in both an Illumina MiniSeq (R1=16 bp, i7=8 bp,

290 R2=50 bp) and NovaSeq6000 sequencers (data not shown, R1=16 bp, i7=8 bp, R2=150 bp). PhiX was

291 added to account for 25% of the total library, to increase library diversity and subsequently,

9

292   sequencing performance. The 6 first bases in read one were used to demultiplex the reads generated

293   from the different patient. The 8 bases in the barcode allowed, in this case, to discriminate the

294   different amplicons but it could alternatively be used to multiplex several patient plates for

295   sequencing. The 50/150 bases in read 2 were used to distinguish between positive and negative

296   patients by mapping reads against the SARS-CoV-2 genome.

297

298   **Bioinformatic analysis**

299   Reads were segregated by patient, amplicon and plate. The demultiplexed reads were then processed

300   using the standard tools displayed below. A consensus sequence for SARS-CoV-2 genome was

301   generated from the set of published genomes on NCBI using the Bio.Align Python 3 package (data not

302   shown). Bowtie2[10], samtools[11,12] and pysam were used to generate pileup columns and calculate

303   alignment rates for a subset of our samples, on all loci in the SARS-CoV-2 transcriptome to verify the

304   locus-specificity of our primers. After verifying the quality of our reads, Bowtie2 was used to map all

305   of our samples against the SARS-CoV-2 genomic region of interest (from 29600 bp to 29900 bp,

306   inclusive), and GAPDH. Alignment rates were calculated as a ratio of total reads for a given sample, to

307   the number of reads aligning to our region of interest on the SARS-CoV-2 genome.

308

```
bcl2fastq → bowtie2 → Samtools → Gapdh → Statistical
                        / pysam   normalization   Analysis
```

309   Samples for which the initial diagnostic PCR failed (*i.e.* GAPDH was undetectable after 40 cycles of

310   PCR) were filtered from the dataset. Additionally, samples for which we recovered less than 250

311   sequencing reads per amplicon were filtered out of the dataset. After filtering had been applied, our

312   dataset contained 83 sets of patient quadruplicates. Undetected Ct values of SARS-CoV-2 were

313   imputed and values were then normalized to GAPDH using the Livak and Schmittgen method[13]. The

314   Wilcoxon Signed-Rank Test was applied to the set of patient replicates, comparing the SARS-CoV-2

315   alignment rate and the normalized Ct value recovered from RT-qPCR ($p < 0.0001$ for each amplicon).

316

References

318   1.   Marcel, S. et al. COVID-19 epidemic in Switzerland: On the importance of testing, contact tracing

319   and isolation. Swiss Med. Wkly. (2020). doi:10.4414/smw.2020.20225

320   2.   Ki, M. Epidemiologic characteristics of early cases with 2019 novel coronavirus (2019-nCoV)

321   disease in Korea. Epidemiol. Health (2020). doi:10.4178/epih.e2020007

322   3.   St Hilaire, B. G. et al. A rapid, low cost, and highly sensitive SARS-CoV-2 diagnostic based on whole

323   genome sequencing. bioRxiv 2020.04.25.061499 (2020). doi:10.1101/2020.04.25.061499

324   4.   Bagnoli, J. W. et al. Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. Nat.

325   Commun. 9, 2937 (2018).

326 5. Sola, I., Almazán, F., Zúñiga, S. & Enjuanes, L. Continuous and Discontinuous RNA Synthesis in

327   Coronaviruses. Annu. Rev. Virol. 2, 265–288 (2015).

328 6. Kim, D. et al. The Architecture of SARS-CoV-2 Transcriptome. Cell 181, 914-921.e10 (2020).

329 7. Huang, B. et al. Illumina sequencing of clinical samples for virus detection in a public health

330   laboratory. Sci. Rep. 9, 5409 (2019).

331 8. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat.

332   Methods 10, 1096–1098 (2013).

333 9. Wyllie, A. L. et al. Saliva is more sensitive for SARS-CoV-2 detection in COVID-19 patients than

334   nasopharyngeal   swabs.   medRxiv   2020.04.16.20067835   (2020).

335   doi:10.1101/2020.04.16.20067835

336 10. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–

337   359 (2012).

338 11. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079

339   (2009).

340 12. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and

341   population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993

342   (2011).

343 13. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time

344   Quantitative PCR and the 2−ΔΔCT Method. Methods 25, 402–408 (2001).

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359   **Figures**
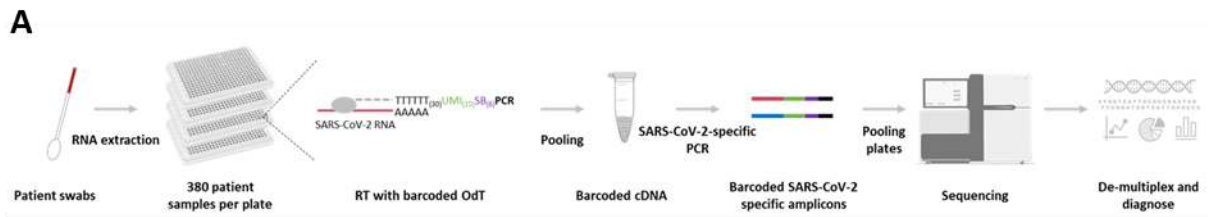
Label	Amplicon insert size	5'-3' Primer sequence	GC %	Lentgh bp	Tm °C	Amplicon Size PCR 1
SCoV2_29709	161	CAGTGTGTAACATTAGGGAGGAC	47.8	23	60	298
SCoV2_29691	179	CTCACATAGCAATCTTTAATCAGTG	36	25	57.7	318
SCoV2_29652	218	CGTAACTACATAGCACAAGTAGATG	40	25	58.7	357
GAPDH_1124	259	GCTCATTTCCTGGTATGACAACG	47.8	23	61.4	396

360

361

362   **Figure 1. (A) Schematic representation of the protocol.** Using a low volume liquid handling robot, the

363   patients' RNA is distributed in 384-well plates containing indexed oligo-dT primers. Barcoded cDNA is

364   generated by reverse transcription and pooled into a single tube. Libraries are produced from SARS-

365   CoV-2-specific amplicons and sequenced. Reads are de-multiplexed based on both a plate and a

366   sample barcode and used for downstream diagnostic analysis. **(B) Amplicon primer design.** The table

367   contains the sequences of the forward primers used to generate the different amplicons. The reverse

368   primer is common to all amplicons and it anneals to the 3'-end of the barcoded oligo-dT primers used

369   to generate cDNA in the reverse transcription. The primer labels contain the position of the first

370   sequenced base. Insert length was calculated using the start of the poly-A sequence. SARS-CoV-2

371   NC_045512_2 sequence and GAPDH NM_002046.4 sequence were used as reference. For the

372   theoretical size of the amplicon after PCR 1 the Nextera Tag (34bp), the length of the barcoded Oligo-

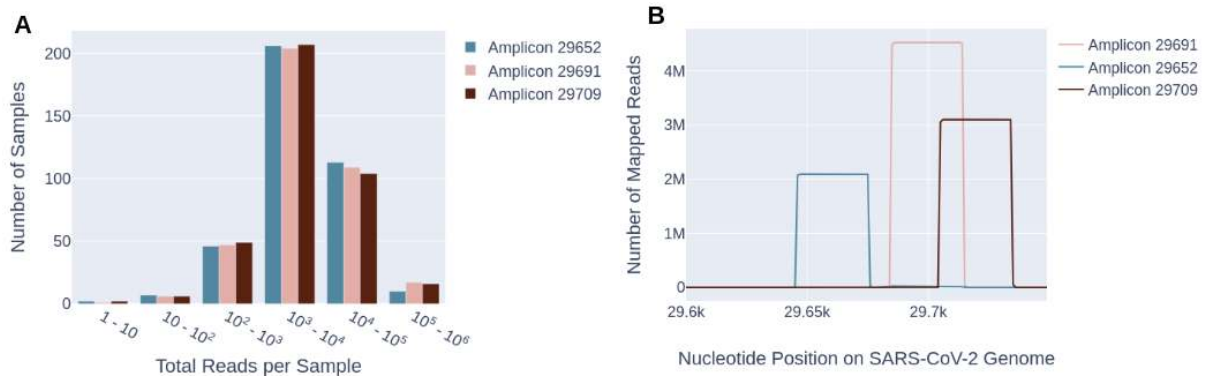373   dT-primer (80bp) and the primer sequence was added.

374

375

376

377

378

379

380

381

382

Figure 2. (A) Read distribution in the different wells of the plate. The histogram shows the read numbers obtained in the different wells of the plate for the different amplicons. As the input for the protocols was not normalized, a wide distribution in the number of reads is observed. Wells with <250 reads were discarded for further analysis. **(B) Read alignment to SARS-CoV-2 genome.** The number of reads aligning to the 3'-end of viral genome is shown for the different amplicons. More than 99.9% of reads aligning to SARS-CoV-2 mapped to the locus targeted in the amplicon design.
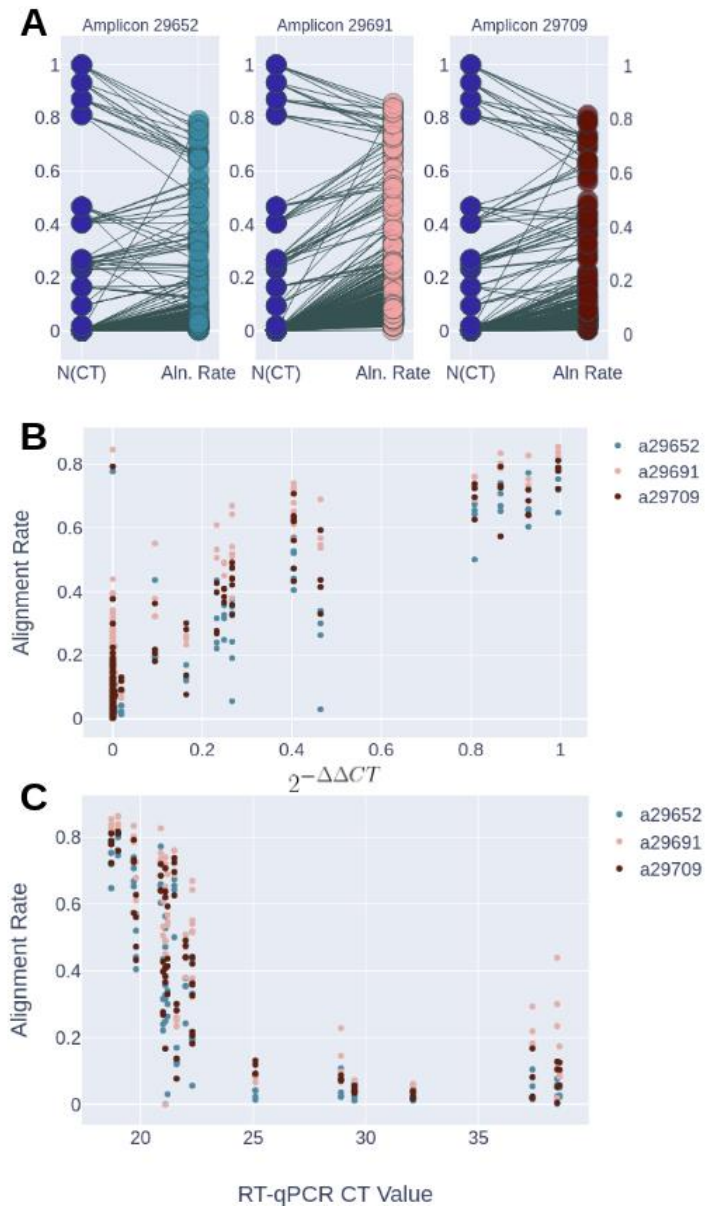
411

412

Figure 3. (A) Matched pairs of normalized Ct values and alignment rate. Ct values (y-axis-left) for sets of RT-qPCR technical replicates were normalized (N(Ct) = $2^{-\Delta\Delta Ct}$) and correlated to the alignment rate (y-axis-right) for the different amplicons using the Wilcoxon-Signed-Rank test for matched pairs ($p <$ 0.001). **(B) Correlation of normalized Ct values and alignment rate.** One point is shown for each sample and amplicon combination, coloured by amplicon type. The normalised Ct value on the x-axis, mapped against its corresponding alignment rate on the y-axis. **(C) Alignment rate vs. raw Ct values for samples diagnosed positive via RT-qPCR.** This figure shows one point for each sample and amplicon combination, and displays only those samples that were diagnosed positive via RT-qPCR. The x-axis shows the Ct value associated with these positive samples, with their corresponding alignment rate shown on the y-axis. This demonstrates a sensitivity threshold for HiDRA-seq in terms of RT-qPCR Ct values (Ct ≈ 25).
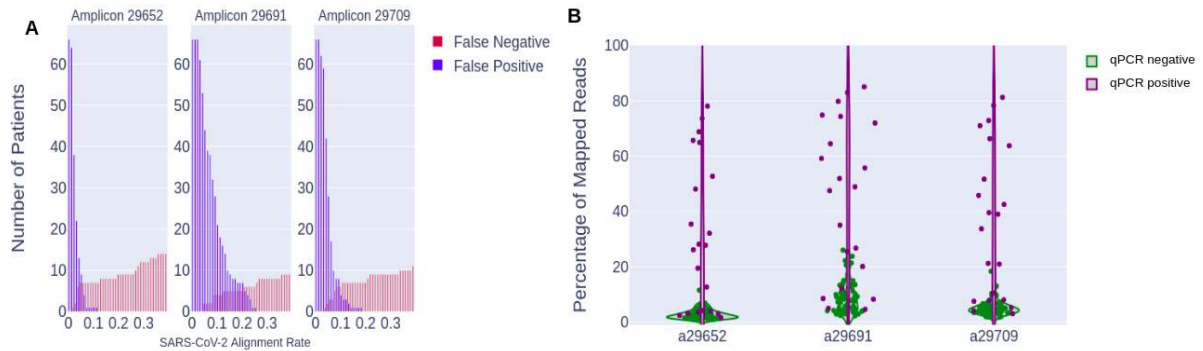
14

Figure 4. (A) Histogram of false diagnoses for potential diagnostic thresholds: For each amplicon used, the number of potential mischaracterized diagnoses (false-positives in blue or false-negatives in red), in numbers of patients (y-axis), are shown for a given alignment rate threshold (x-axis). Original diagnoses are given via RT-qPCR. Amplicon 29691 is shown to have the fewest number of mis-characterized diagnoses (7) when said diagnoses are minimized, and Amplicon 29652 is shown to have the shortest span of intersection between both distributions (0.9). **(B) Distributions of alignment rates for positive and negative patient quadruplicates, by amplicon:** For each amplicon, the mean percentage (across four replicates) of reads that aligned to the SARS-CoV-2 genome (y-axis) is shown as a point, with patients diagnosed positive via RT-qPCR shown in purple and patients diagnosed negative via RT-qPCR shown in green. The distribution of alignment rates for patients diagnosed positive with RT-qPCR is shown to be much wider (spanning roughly 87 percentiles on average) than those diagnosed negative (spanning roughly 14 percentiles on average).

15

Author contributions

CA and RS conceived the study. EY, GW, SK, LP and CA designed the protocol, prepared the libraries and sequenced them. Sequencing data was processed and analysed by GW and LO. EY, GW, SK, LO, LP, TS, MDM, CA and RS wrote the manuscript.

Acknowledgments

We thank Dr. Michael Huber and Dr. Jürg Böni (Institute of Medical Virology, University of Zurich, Switzerland) for providing the RNA extracted from patients' samples used for the implementation of the protocol. We are grateful to Prof. Andreas Moor (D-BSSE, ETH Basel, Switzerland) for providing the set of 384 barcoded oligo-dT primers used in this protocol.

Competing interests

The authors declare no competing interests.

Data availability

Sequencing data generated here are available at ENA under accession PRJEB38511.