

1 **A Multiplexed, Next Generation Sequencing Platform for High-Throughput**

2 **Detection of SARS-CoV-2**

3 Marie-Ming Aynaud^{1**}, J. Javier Hernandez^{1,3,4**}, Seda Barutcu^{1**}, Ulrich
4 Braunschweig^{3,4}, Kin Chan¹, Joel D. Pearson¹, Daniel Trcka¹, Suzanna L. Prosser¹,
5 Jaeyoun Kim¹, Miriam Barrios-Rodiles¹, Mark Jen¹, Siyuan Song^{4,5}, Jess Shen¹, Christine
6 Bruce², Bryn Hazlett², Susan Poutanen², Liliana Attisano^{4,5}, Rod Bremner¹, Benjamin J.
7 Blencowe^{3,4}, Tony Mazzulli², Hong Han^{3,4}, Laurence Pelletier^{1,3*}, Jeffrey L. Wrana^{1,3*}.

8

9 **Authors affiliation**

10 ¹Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, M5G
11 1X5, Canada

12 ²Department of Microbiology, Mount Sinai Hospital/University Health Network, Toronto,
13 Ontario, M5G 1X5, Canada.

14 ³Department of Molecular Genetics, Donnelly Centre, University of Toronto, Toronto,
15 Ontario, M5S 3E1, Canada

16 ⁴Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto,
17 Ontario, M5S 3E1, Canada

18 ⁵Department of Biochemistry, Donnelly Centre for Cellular and Biomolecular Research,
19 University of Toronto, Toronto, Ontario, M5S 3E1, Canada.

20 **Equal contribution

21 ***Corresponding authors:**

22 Dr. Jeffrey L. Wrana, wrana@lunenfeld.ca and Dr. Laurence Pelletier,
23 pelletier@lunenfeld.ca

24 **Abstract**

25 Population scale sweeps of viral pathogens, such as SARS-CoV-2, that incorporate large
26 numbers of asymptomatic or mild symptom patients present unique challenges for public
27 health agencies trying to manage both travel and local spread. Physical distancing is the
28 current major strategy to suppress spread of the disease, but with enormous socio-
29 economic costs. However, modelling and studies in isolated jurisdictions suggest that
30 active population surveillance through systematic molecular diagnostics, combined with
31 contact tracing and focused quarantining can significantly suppress disease spread¹⁻³ and
32 has significantly impacted disease transmission rates, the number of infected people, and
33 prevented saturation of the healthcare system⁴⁻⁷. However, reliable systems allowing for
34 parallel testing of 10-100,000's of patients in larger urban environments have not yet been
35 employed. Here we describe "COVID-19 screening using Systematic Parallel Analysis of
36 RNA coupled to Sequencing" (C19-SPAR-Seq), a scalable, multiplexed, readily
37 automated next generation sequencing (NGS) platform⁸ that is capable of analyzing tens
38 of thousands of COVID-19 patient samples in a single instrument run. To address the
39 strict requirements in clinical diagnostics for control of assay parameters and output, we
40 employed a control-based Precision-Recall and predictive Receiver Operator
41 Characteristics (coPR) analysis to assign run-specific quality control metrics. C19-SPAR-
42 Seq coupled to coPR on a trial cohort of over 600 patients performed with a specificity of
43 100% and sensitivity of 91% on samples with low viral loads and a sensitivity of > 95%
44 on high viral loads associated with disease onset and peak transmissibility. Our study
45 thus establishes the feasibility of employing C19-SPAR-Seq for the large-scale monitoring
46 of SARS-CoV-2 and other pathogens.

47 **Main**

48 The current gold standard diagnostic for SARS-CoV-2 is Real-Time Quantitative
49 Polymerase Chain Reaction (RT-qPCR), which is not readily adaptable to large scale
50 population testing⁹. To establish a population-scale testing platform we designed a SPAR-
51 Seq multiplex primer mix v1 that targets RNA-dependent RNA Polymerase (*RdRP*),
52 Envelope (*E*), Nucleocapsid (*N*), and two regions of the Spike (*S*) gene that correspond
53 to the receptor binding domain (RBD) and the polybasic cleavage site (PBS) (**Fig. 1a and**
54 **Supplementary Table 1**). The latter two are SARS-CoV-2-specific regions that capture
55 five key residues necessary for ACE2 receptor binding (*Srbd*) and the furin cleavage site
56 (*Spbs*) that is critical for viral infectivity^{10,11}. For quality control, we targeted Peptidylprolyl
57 Isomerase B (*PPIB*). Current standard testing strategies for viral pathogens employ gene-
58 specific primers in “all-in-one” qRT-PCR reactions that could in principle be adapted to
59 incorporate barcodes into gene-specific primers. However, to allow for rapid adaptation
60 to test for novel and multiple pathogens, and/or profiling host responses we used a
61 generic oligo-dT and random hexamer primed reverse transcription step followed by
62 multiplex PCR and barcoding in a rapid, readily automated format we call “COVID-19
63 screening using Systematic Parallel Analysis of RNA coupled to Sequencing” or C19-
64 SPAR-Seq (**Fig. 1b and Supplementary Table 1**). Although cost is often cited as a
65 concern for NGS-based testing, our platform is cost effective with retail material costs
66 ranging from USD ~\$9 to \$6 for 500 *versus* 10,000 sample batch sizes, respectively
67 (**Supplementary Table 2**).

68

69 To assess C19-SPAR-Seq performance, we assembled a proof-of-concept (PoC) cohort
70 of 19 archival Nasopharyngeal (NASOP) swab eluents from the Toronto University Health
71 Network-Mount Sinai Hospital clinical diagnostics lab (**Supplementary Table 3**), 17 of
72 which were positive for SARS-CoV-2. Viral load in these archival samples was quantified
73 using the clinically approved TaqMan-based SARS-CoV-2 RT-qPCR detection kit¹²
74 ('BGI'), which identified 5 SARS-CoV-2^{low} (Ct > 25), 7 SARS-CoV-2^{medium} (Ct between 20
75 and 25), and 5 SARS-CoV-2^{high} (Ct < 20) patients (**Supplementary Table 3**). After
76 confirming the efficiency of multiplex v1 primer pairs using a SARS-CoV-2^{high} sample,
77 (LTRI-18, Ct < 20; **Extended data Fig. 1**), we performed C19-SPAR-Seq using HEK293T
78 RNA as a negative control (n = 2), and serial dilutions of synthetic SARS-CoV-2 RNA
79 (Twist) as positive controls (n = 5). Pooled sequence data was demultiplexed to individual
80 samples prior to mapping to amplicon sequences. C19-SPAR-Seq was sensitive in
81 detecting as low as 12.5 copies/ μ L of E, S_{rbd}, and S_{pbs} amplicons from Twist RNA (**Fig.**
82 **1c, left panel**). In patient samples, *PPIB* was present in all samples, and all viral targets
83 were robustly detected in high/medium load samples, with reduced detection of *E* and
84 *RdRP* genes in low samples (**Fig. 1c, right panel**).

85
86 To establish a diagnostic platform, we performed C19-SPAR-Seq on a larger test
87 development cohort of 24 COVID-19 positive and 88 negative archival patient samples (n
88 = 112; **Supplementary Table 4**). The SARS-CoV-2 RNA standard curve showed a linear
89 relationship between total viral reads and estimated viral copy numbers (**Extended data**
90 **Fig. 2a**). Negative patient samples had low viral reads (median of 4; range 0-55)
91 compared to positive samples (median of 5,899; range 2-253,956 corresponding to 18 to

92 705,960 amplicon reads per million reads per sample) (**Fig. 2a**). C19-SPAR-Seq read
93 counts tracked inversely with qRT-PCR Ct values for *RdRP*, *E* and *N* genes quantified in
94 the diagnostic lab using the Seegene Allplex™ assay¹³ (**Fig. 2b**). Unsupervised clustering
95 showed that the controls performed similarly to the PoC cohort (**Fig. 2c**), as did the
96 positive and negative patient samples, with two exceptions: clinical samples LTRI042 and
97 LTRI050, which displayed background signal, and corresponded to samples with extreme
98 Ct values in only one viral gene (*N* gene, Ct > 38; **Supplementary Table 4**). ROC analysis
99 using total viral reads (**Fig. 2d**) showed excellent performance with an area under the
100 ROC curve (AUC) of 0.969, sensitivity of 92%, specificity of 100%, and overall accuracy
101 of 98%. Thus, C19-SPAR-Seq robustly detects SARS-CoV-2 transcripts, correlates with
102 Ct values from clinical diagnostic tests, and displays excellent performance in
103 distinguishing positive and negative samples.

104
105 Robust application of C19-SPAR-Seq as a diagnostic tool requires assigning thresholds
106 for both viral RNA detection, as well as host RNA for filtering poor quality samples. In
107 qRT-PCR diagnostics, external validation studies and rigorous standard operating
108 procedures establish pre-defined cutoffs for sample quality and positive *versus* negative
109 assignment (Seegene¹³; BGI¹²). However, in scalable, massively parallel, multiplexed
110 NGS assays, variability in sample numbers and flow cell loading can create run-to-run
111 variations in read numbers, while index-mismatching¹⁴, as well as trace cross-
112 contamination events can create technical noise that are challenging to control.
113 Furthermore, external validation strategies create a laborious path to adapt and test new
114 multiplex designs to SARS-CoV-2, additional respiratory pathogens, or host responses.

115 We therefore exploited the throughput of C19-SPAR-Seq to include in every run a training
116 set of large numbers of controls that can be exploited to define cutoffs tailored to each
117 C19-SPAR-Seq run (**Fig. 3a**). To define quality metrics, we computed precision-recall
118 (PR) curves for classifying control samples as either negative (H₂O blanks), or positive
119 for any anticipated amplicon (HEK293T for PPIB or synthetic SARS-CoV-2 RNA for viral
120 amplicons) and calculated the highest F1 score, which is the harmonic mean of PR and
121 a common measure of classifier accuracy (**Fig. 3b**). When mapped onto a ROC curve
122 this corresponded to the region closest to perfect sensitivity and specificity (0, 1)
123 (**Extended data Fig. 2b**). To define the threshold for identifying SARS-CoV-2 positive
124 cases, we next analyzed the embedded standard curve of synthetic SARS-CoV-2 RNA.
125 This displays a linear relationship over 4 orders of magnitude and extends to lower limits
126 of detection indistinguishable from background reads from HEK293T cells (**Fig. 2a and**
127 **Extended data Fig. 2a**), thus allowing us to identify the viral read count in each C19-
128 SPAR-Seq run that most accurately distinguishes positive from negative (**Fig. 3a**). To
129 identify this threshold, we computed PROC01, which optimizes negative predictive value
130 (NPV) and positive predictive value (PPV)¹⁵ and defined a point (88 viral reads) close to
131 perfect PR (**Fig. 3c**) and sensitivity and specificity on the ROC curve (**Extended data**
132 **Fig. 2c**). Importantly, these methods control for run-specific variables by employing
133 training sets that are embedded in every C19-SPAR-Seq run.

134

135 We next mapped the control-based cutoffs onto patient SPAR-Seq data (**Fig. 3d**). This
136 showed 15 of these archival samples had low *PPIB* counts that may be due to lost RNA
137 integrity upon repeated freeze-thaw cycles (**Fig. 3d and Supplementary Table 4**), a

138 variability we also observed in the PoC cohort (**Fig. 1c**). Of note, C19-SPAR-Seq
139 performance was not affected by filtering poor quality samples (AUC = 0.970; **Extended**
140 **data Fig. 2d**). Furthermore, using PROC01 thresholding of viral reads identified 22/24
141 positives with no false positives (**Fig. 3d**). This yielded an overall test performance of 92%
142 sensitivity, 100% specificity, and 98% accuracy (**Fig. 3d** and **Supplementary Table 5**).
143 This is similar to the observed performance of C19-SPAR-Seq on clinical samples
144 quantified by ROC analysis (**Fig. 2d** and **Extended data Fig. 2d**, respectively). Thus, an
145 extensive array of internal reference samples is effective as an embedded training set for
146 implementing a control-based PR/pROC classifier (coPR) tailored to each C19-SPAR-
147 Seq run.

148
149 To validate our C19-SPAR-Seq platform we established a pilot cohort of 378 samples
150 that contains 89 positive samples collected in May of 2020. We first screened for positivity
151 using the clinically approved BGI SARS-CoV-2 kit¹² which showed 52 samples were
152 positive with > 4 viral copies/ μ L (**Supplementary Table 6,9**). Of the 37 failed samples,
153 86% had very low viral RNA (only 1 or 2 of the 3 genes detected and/or Ct > 35 on the
154 'Seegene' platform) that may have lost integrity upon storage. Indeed, comparison of Ct
155 values for RdRP detection showed an overall increase of 4 cycles in these archived
156 samples (**Extended data Fig. 3a**), despite the high sensitivity of the BGI platform¹⁶. The
157 cohort also contained 289 negative samples collected prior to Ontario's¹⁷ first confirmed
158 COVID-19 positive case in January 20, 2020, and 1 negative sample collected in May,
159 2020 (**Supplementary Table 6, Table 1**), and included broncho-alveolar lavages (BALs)
160 and NASOP swabs. Surprisingly, the detection of human RNA dropped substantially to a

161 median of 29 (range 0- 41,874), compared to 15,058 (range 2- 170,870) in the original
162 test cohort. coPR filtering (**Extended data Fig. 3b**), marked 50% of samples as
163 inconclusive compared to 13% in the test cohort (**Extended Data Fig. 3c**), despite similar
164 distribution of raw reads per sample (**Extended data Fig. 3d**), while mapping rates in the
165 PoC, test and pilot cohorts, progressively declined to as low as 0.1% (**Extended Data**
166 **Fig. 3e**). To understand this collapse we analyzed unmapped reads and found that > 90%
167 were consumed by non-specific amplification products (NSAs; **Extended Data Fig. 4a**)
168 that comprised complex chimeric combinations of many viral and human primers
169 (**Extended Data Fig. 4a, b**). For example, RdRP and PPIB contributed to 4 of the top 5
170 NSAs (NSA1-4), and 2 had a spurious sequence (NSA4,5). Indeed, analysis of C19-
171 SPAR-Seq PoC, test and pilot libraries using a Bioanalyzer, showed that as cohort size
172 and number of negatives increased, NSAs were more apparent, and dominated the pilot
173 library (**Extended data Fig. 4c**). This suggests that NSAs, enriched in negative samples
174 (3.7-fold increase in the pilot cohort), clog the NGS pipeline as sample numbers rise
175 (**Supplementary Table 9**). This has serious implications for deploying an NGS platform
176 in a population-scale COVID-19 surveillance strategy and highlights the importance of
177 using large-scale cohorts during the development of multiplex testing platforms.

178
179 SARS-CoV-2 RNA concentration spans a large dynamic range, such that spike-in mutant
180 amplicons which have been suggested to improve performance of NGS-based
181 strategies¹⁸ might interfere with detection of COVID-19 positive cases with low viral reads.
182 Therefore, we instead used our NSA data to create multiplex panel v2.0 (see Methods)
183 that removed primers yielding NSAs by targeting a distinct region of *RdRP*, removing *E*

184 and *N* genes, and switching to primers that amplify intron spanning regions of the *ACTB*
185 and *ACTG* genes (**Supplementary Table 1 and Extended data Fig. 1**). We extended
186 the pilot cohort to 663 samples that included 98 confirmed positives and performed C19-
187 SPAR-Seq, which showed targeted amplicons were the predominant product generated
188 by multiplex panel v2.0 (**Extended data Fig. 5a**), and mapping percentages were
189 restored to test cohort levels (**Extended data Fig. 5b**). Total viral read distributions for
190 multiplex panel v2.0 showed good separation in clinically positive samples (**Fig. 4a** and
191 **Extended data Fig. 5c**), while applying coPR thresholding (**Extended data Fig. 5d**)
192 identified 121 samples as inconclusive (**Fig. 4a**), all of which were older, pre-COVID19
193 material. Of these, 112 were BALs (40% of all BALs), 1 was a bronchial wash (BMSH),
194 and only 8 were NASOPs (1.8% of all NASOPs) (**Supplementary Table 7**). Furthermore,
195 analysis of 10 BAL samples below the QC threshold revealed little or no RNA, contrasting
196 BALs with moderate levels of *ACTB/G* transcripts (representative examples in **Extended**
197 **data Fig. 6a**), and BAL *ACTB/G* read distributions were much lower than NASOPs
198 (**Extended data Fig. 6b**). This suggests that archival BALs suffered from substantive
199 sample degradation and also highlights how coPR-based thresholding successfully
200 identifies poor quality samples and readily adapts to the use of distinct primer sets.

201
202 Next, we analyzed viral reads, which had a broad range in positive samples (median =
203 680.5 reads per sample, range 0-200,850; **Fig. 4a** and **Extended data Fig. 5c**). Two-
204 dimensional clustering showed background SARS-CoV-2 products in negative samples
205 were low to undetectable, and *ACTB* typically yielded higher reads than *ACTG*, likely
206 reflecting their differential expression (**Fig. 4b**). Positive samples were generally well

207 separated, although some distinct clusters with lower SARS-CoV-2 reads were apparent
208 (**Fig. 4b** and **Extended data Fig. 5e**). Indeed, total read distributions in positive samples
209 displayed biphasic distribution (**Extended data Fig. 5e**), similar to observations made
210 from RT-qPCR analyses of ~4000 positive patients¹⁹. Since the early rapid increase in
211 SARS-CoV-2 viral load at symptom onset is followed by a long tail of low viral load during
212 recovery^{20,21}, this biphasic distribution could reflect patients in distinct phases of the
213 disease. We also assessed viral amplicon sequences which matched the SARS-CoV-2
214 reference (MN908947.3²²) and found no variants (data not shown). Since neutralizing
215 antibodies are generally thought to target the critical region of the RBD analyzed here¹⁷,
216 these results suggest the emergence of variant strains that might bypass acquired
217 immunity is not a major feature of SARS-CoV-2. In addition, this supports the notion that
218 biologic therapies targeting the RBD may show broad activity in the population.

219
220 We next compared performance of multiplex panel v2.0 to v1.0 using the embedded
221 controls, which showed similar performance (AUC = 0.90, **Extended data Fig. 5f** versus
222 0.92, **Extended data Fig. 2c**, respectively), with coPR yielding an optimal read cutoff of
223 ≥ 16 total viral reads (**Extended data Fig. 5f**) that corresponded to a technical sensitivity
224 of 3 viral copies/ μ L (**Extended data Fig. 6c**). coPR thus identified 82 positive samples
225 (**Fig. 4a** and **Supplementary Table 7**) all of which were BGI-confirmed cases to give an
226 overall sensitivity of 86%, specificity of 100%, and accuracy of 97% (**Supplementary**
227 **Table 8**). Importantly, total viral reads tracked with BGI Ct values (**Fig. 4c**), and for
228 samples with Ct < 35 (corresponding to ~12 viral copies/ μ L of specimen), sensitivity was
229 similar to the test cohort at 91%. However, for samples with Ct between 35-37 (4-12 viral

230 copies/ μ L) sensitivity dropped markedly to 44% (**Supplementary Table 8**), whereas at
231 higher viral loads (Ct = 25 or \sim 8,400 viral copies/ μ L) sensitivity rose to 100% (**Fig. 4c**).
232 ROC analysis of actual C19-SAR-Seq performance yielded an AUC of 0.96, sensitivity of
233 87% and specificity of 100%, similar to coPR (**Fig. 4d**), while individual amplicons each
234 underperformed total viral reads (AUC: 0.85-0.94; **Extended data Fig. 6d**). Our cohort
235 was biased for samples with very low to low viral loads, which represents a small portion
236 of the COVID-19 population¹⁹. Therefore, we mapped our sensitivity data onto the
237 population distribution of viral load data¹⁹, which showed C19-SPAR-Seq sensitivity of
238 \sim 97% for patients displaying > 10,000 viral copies/mL (**Extended data Fig. 6e**), which
239 encompasses \sim 90% of the positive population. Altogether, these results demonstrate
240 that at high patient sample loads of predominantly negative samples, C19-SPAR-Seq
241 using coPR displays 100% specificity and > 95% sensitivity at viral loads typically
242 observed in populations.

243
244 Systematic population-scale testing has been identified as an important tool in managing
245 pandemics such as SARS-CoV-2, where large numbers of infected individuals display
246 mild or no symptoms yet transmit disease. The scalable throughput of C19-SPAR-Seq
247 combined with its excellent sensitivity and specificity at reasonable cost make it well-
248 suited for this role. Data generated by large-scale routine testing of local and larger
249 communities, with different interaction levels would provide valuable epidemiologic
250 information on mechanisms of viral transmission, particularly when coupled to multiplex
251 panels targeting regions of sequence variance currently development. In addition, the
252 C19-SPAR-Seq platform can be readily adapted to incorporate panels tracking multiple

253 pathogens, as well as host responses. C19-SPAR-Seq quantitation would also facilitate
254 real-time tracking of viral load dynamics in populations that may be associated with
255 COVID-19 expansion or resolution phases²⁰. Although C19-SPAR-Seq is dependent on
256 centralized regional facilities, it is readily coupled to saliva-based, at-home collection that
257 exploits extensive transport infrastructure and industrialized sample processing to enable
258 frequent widespread testing.

259

260

261 **Methods**

262 **Samples collection**

263 Patient samples (**Supplementary Table 2, 3, 5**) were obtained from the Department of
264 Microbiology at Mount Sinai Hospital under MSH REB Study #20-0078-E, 'Use of known
265 COVID-19 status tissue samples for development and validation of novel detection
266 methodologies'.

267

268 **Total RNA extraction**

269 Total RNA was extracted by using the Total RNA extraction kit (Norgen Biotek kit, Cat.
270 #7200) for the samples in Supplementary Table S2 following the manufacturers
271 guidelines. For all other samples (**Supplementary Table 3, 5**), total RNA was purified in
272 96 well plates using RNAClean XP beads (Beckman, A66514) and a customized protocol.
273 Briefly, 75.25 μ L of patient swabs in transfer buffer were mixed with 14.5 μ L of 10X SDS
274 lysis buffer (1% SDS, 10mM EDTA), 48 μ L of 6M GuHCl, and 7.25 μ L proteinase K (20
275 mg/mL, ThermoFisher, 4333793), incubated at room temperature for 10' and heated at
276 65°C for 10' prior to the addition of 145 μ L of beads. Beads were washed twice in 70%
277 ethanol using a magnetic stand and then RNA eluted into 30 μ L Resuspension buffer
278 supplied with the kit. RNA quality was assessed using a Bioanalyzer (5200 Agilent
279 Fragment Analyzer). HEK293T RNA was extracted using the Total RNA extraction kit
280 (Qiagen). Synthetic Twist SARS-CoV-2 RNA (Twist Bioscience #102024 - MN908947.3)
281 was used as positive control.

282

283

284 **Reverse Transcription (RT)**

285 Total RNA was reverse transcribed using SuperScript™ III Reverse Transcriptase
286 (Invitrogen) in 5X First-Strand Buffer containing DTT, a custom mix of Oligo-dT (Sigma)
287 and Hexamer random primers (Sigma), dNTPs (Genedirex) and Ribolock RNase inhibitor
288 (ThermoScientific). We followed the manufacturer's protocol. Each reaction included: 0.5
289 µL Oligo-dT, 0.5 µL hexamers, 4 µL purified Total RNA, 1 µL dNTP (2.5 mM each dATP,
290 dGTP, dCTP and dTTP), *quantum satis (qs)* 13 µL RNase/DNase free water. Samples
291 were incubated at 65°C for 5', and then placed on ice for at least for 1'. The following was
292 added to each reaction: 4 µl 5X First-Strand Buffer, 1 µl 0.1 M DTT, 1 µl Ribolock RNase
293 Inhibitor, 1 µl of SuperScript™ III RT (200 units/µl) and then mixed by gently pipetting.
294 Samples were incubated at 25°C for 5', 50°C for 60', 70°C for 15' and then store at 4°C.

295

296 **TaqMan-based RT-qPCR detection**

297 A Real-Time Fluorescent RT-PCR kit from 'BGI' was used according to manufacturer's
298 instructions. Experiments were carried out in a 10µl reaction volume in 384-well plates,
299 using 3 µl of sample (LTRI patient samples or Twist RNA), and were analyzed using a
300 Bio-Rad CFX384 detection system (**Supplementary Tables 3,6,7**). Real-time
301 Fluorescent RT-PCR results from 'Seegene' assay were provided by the Department of
302 Microbiology diagnostic lab at Mount Sinai Hospital (**Supplementary Tables 3,4,6,7**).

303

304

305

306

307 **C19-SPAR-Seq primer design and optimization**

308 Optimized multiplex PCR primers for SARS-CoV-2 (*N*, *S*, *E* and *RdRP*) and human genes
309 (*PPIB* and *ACTB/G*) were designed using the SPAR-Seq pipeline⁸, with amplicon size >
310 100 bases (see **Supplementary Table 1**). For *S* gene, two regions were monitored the
311 *S receptor binding domain* (*Srbd*), and *S polybasic cleavage site* (*Spbs*). The Universal
312 adapter sequences used for sequencing were F: 5'-acactctttccctacacgacgctcttccgatct and
313 R: 5'-gtgactggagttcagacgtgtgctcttccgatct). Primers were optimized to avoid primer-dimer
314 and non-specific multiplex amplification. To assess the primers sensitivity and specificity,
315 we performed qPCR (SYBR green master mix, BioApplied) on cDNA prepared from
316 patient samples. Each primer was used at 0.1 μM in qPCR reaction run on 384 well plates
317 using Biorad CFX 384 detection system. The thermal cycling conditions were as follows:
318 one cycle at 95°C for 2', and then 40 cycles of 95°C for 15'', 60°C for 15'', 72°C for 20'',
319 followed by a final melting curve step.

320

321 **Multiplexing PCR**

322 The multiplex PCR reaction was carried out using Phusion polymerase (ThermoFisher).
323 The manufacturer's recommended protocol was followed with the following primer
324 concentrations: all primers (*N*, *Spoly*, *Srbd*, *E*, *RdRP*, and *PPIB*) were at 0.1 μM for the
325 PoC cohort (**Supplementary Table 3**), SARS-CoV-2 primers (*N*, *Spoly*, *Srbd*, *E* and
326 *RdRP*) were at 0.05 μM, and *PPIB* primer was at 0.1 μM for the test and pilot cohort
327 (**Supplementary Table 4, 6**), all primers (*Spoly*, *Srbd*, *RdRP* and *ACTB/G*) were at 0.05
328 μM for the extended cohort (**Supplementary Table 7**). For each reaction: 5 μL 5X
329 Phusion buffer, 0.5 μL dNTP (2.5 mM each dATP, dGTP, dCTP and dTTP), 0.25 μL for

330 each human primers (10 μ M) , 0.125 μ L for each SARS-CoV2 primers (10 μ M), 2 μ L of
331 cDNA, 0.25 μ L Phusion Hot start polymerase, qs 25 μ L RNase/DNase free water. The
332 thermal cycling conditions were as follows: one cycle at 98°C for 2', and 30 cycles of 98°C
333 for 15'', 60°C for 15'', 72°C for 20'', and a final extension step at 72°C for 5' and then
334 stored at 4°C for the PoC and extended cohorts (**Supplementary Table 3, 7**), one cycle
335 at 98°C for 2', and 35 cycles of 98°C for 15'', 60°C for 15'', 72°C for 20'', and a final
336 extension step at 72°C for 5' and then stored at 4°C for the test and pilot cohorts
337 (**Supplementary Table 4, 6**).

338

339 **Barcoding PCR**

340 For multiplex barcode sequencing, dual-index barcodes were used⁸. The second PCR
341 reaction on multiplex PCR was performed using Phusion polymerase (ThermoFisher).
342 For each reaction: 4 μ L 5X Phusion buffer, 0.4 μ L dNTP (2.5 mM each dATP, dGTP,
343 dCTP and dTTP), 2 μ L Barcoding primers F+R (pre-mix), 4 μ L of multiplex PCR reaction,
344 0.2 μ L Phusion polymerase, qs 20 μ L RNase/DNase free water. The thermal cycling
345 conditions were as follows: one cycle at 98°C for 30'', and 15 cycles of 98°C for 10'', 65°C
346 for 30'', 72°C for 30'', and a final extension step at 72°C for 5' and stored at 4°C.

347

348 **Library preparation and Sequencing**

349 For all libraries, each sample was pooled (7 μ L/sample) and library PCR products were
350 purified with SPRIselect beads (A66514, Beckman Coulter). The PoC, test, and pilot
351 cohorts were purified as follows: ratio 0.8:1 (beads:library), and the extended cohort with
352 1:1 (beads:library) (Beckman Coulter). Due to NSA products in the fragment analyzer

353 profile (**Extended data Fig. 3c**) in the test cohort and pilot cohort, we performed size
354 selection purification (220-350 bp) using the Pippin Prep system (Pippin HT, Sage
355 Science). Library quality was assessed with the 5200 Agilent Fragment Analyzer
356 (ThermoFisher) and Qubit 2.0 Fluorometer (ThermoFisher). All libraries were sequenced
357 with MiSeq or NextSeq 500 (Illumina) using 75 bp paired-end sequencing.

358

359 **COVID-19 (C19-)SPAR-Seq platform**

360 Our Systematic Parallel Analysis of Endogenous RNA Regulation Coupled to Barcode
361 Sequencing (SPAR-Seq) system⁸ was modified to simultaneously monitor COVID-19 viral
362 targets and additional controls by multiplex PCR assays. For barcode sequencing,
363 unique, dual-index C19-SPAR-Seq barcodes were used. Unique reverse 8-nucleotide
364 barcodes were used for each sample, while forward 8-based barcodes were used to mark
365 each half (48) of the samples in 96-well plate to provide additional redundancy. These
366 two sets of barcodes were incorporated into forward and reverse primers, respectively,
367 after the universal adaptor sequences and were added to the amplicons in the second
368 PCR reaction.

369

370 **Demultiplexing and Mapping**

371 Illumina MiSeq sequencing data was demultiplexed based on perfect matches to unique
372 combinations of the forward and reverse 8 nucleotide barcodes. Full-length forward and
373 reverse reads were separately aligned to dedicated libraries of expected amplicon
374 sequences using bowtie²³ with parameters `-best -v 3 -k 1 -m 1`. Read counts per amplicon

375 were represented as reads per million or absolute read counts. The scripts for these steps
376 are available at <https://github.com/UBrau/SPARpipe>.

377

378 **Filtering of low-input samples**

379 To remove samples with low amplified product, likely reflecting low input due to inefficient
380 sample collection or degradation, before attempting to classify, we computed precision-
381 recall curves for classifying control samples into 'low amplification' and 'high amplification'
382 based on reads mapped to RNA amplicons but ignoring mapping to genomic sequence,
383 if applicable. The former group comprised all controls in which individual steps were
384 omitted (H2O controls) and the latter comprised HEK293T as well as synthetic SARS-
385 CoV-2 RNA controls. For each PoC, test and pilot runs, we obtained the mapped read
386 threshold associated with the highest F1 score, representing the point with optimal
387 balance of precision and recall. Samples with reads lower than this threshold were
388 removed from subsequent steps.

389

390 **SARS-CoV2 positive sample classification**

391 To assign positive and negative samples, we used negative (H2O and HEK293T) and
392 positive (synthetic SARS-CoV-2 RNA dilutions) internal controls for each run and
393 calculated optimum cut-offs for viral reads by PROC which defines the threshold for
394 optimum PPV (positive predictive value) and NPV (negative predictive value) for
395 diagnostic tests. Thus, a sample was labelled positive if it had viral reads above the viral
396 read threshold; negative if it had viral reads below the viral read threshold and human

397 reads above the mapped read threshold; and inconclusive if it had both viral and human
398 reads below the respective thresholds.

399

400 **Sample classification by heatmap clustering**

401 Heatmap and hierarchical clustering of viral and control amplicons, $\log_{10}(\text{mapped}$
402 $\text{reads}+1)$, was used to analyze and classify all samples. Samples with a total mapped
403 read count lower than the RNA QC threshold were labeled as inconclusive and removed
404 before the analysis. Known positive (high, medium, and low) and negative control
405 samples were used as references to distinguish different clusters. In addition, dilutions of
406 synthetic SARS-CoV-2 RNA were also included as controls and analyzed across different
407 PCR cycle and primer pool conditions.

408

409 **Viral mutation assessment**

410 To remove PCR and sequencing errors for the assessment of viral sequence variations,
411 we determined the top enriched amplicon sequence. For this, firstly, paired end reads
412 were stitched together to evaluate full length amplicons. The last 12 nucleotides of read1
413 sequence are used to join the reverse complement of read2 sequences. No mismatches
414 were allowed for stitching criteria. The number of full length reads per unique sequence
415 variation were counted for each amplicon per sample by matching the 10 nucleotides from
416 the 3' and 5' end of the sequence with gene-specific primers. (scripts are available at
417 <https://github.com/seda-barutcu/FASTQstitch> and [https://github.com/seda-](https://github.com/seda-barutcu/MultiplexedPCR-DeepSequence-Analysis)
418 [barutcu/MultiplexedPCR-DeepSequence-Analysis](https://github.com/seda-barutcu/MultiplexedPCR-DeepSequence-Analysis)) The top enriched sequence variant
419 from each sample is used for multiple alignment analysis using CLUSTALW.

420 **Non-specific amplicon assessment**

421 Single-end reads that contain the first 10 nucleotides of the illumina adaptor sequence
422 were counted and binned into relevant forward and reverse gene specific primer pools by
423 matching the first 10 nt of the reads with primer sequences. Relative abundance of the
424 non-specific amplicons was quantified as percentage of the reads corresponding to non-
425 specific amplicon per forward or reverse primer (Scripts are available at
426 <https://github.com/seda-barutcu/MultiplexedPCR-DeepSequence-Analysis>).

427

428 **Data Availability**

429 Data submitted to GEO (accession number pending).

430

431 **Code Availability**

432 We provided the code for demultiplexing and mapping at
433 <https://github.com/UBrau/SPARpipe>, viral mutation assessment and non-specific
434 amplicon assessment at <https://github.com/seda-barutcu/FASTQstitch> and
435 <https://github.com/seda-barutcu/MultiplexedPCR-DeepSequence-Analysis>.

436 **References**

- 437 1. Berger, D.W., Herkenhoff, K.F. & Mongey, S. An SEIR infectious disease model
438 with testing and conditional quarantine. (National Bureau of Economic Research, 2020).
- 439 2. Taipale, J., Romer, P. & Linnarsson, S. Population-scale testing can suppress the
440 spread of COVID-19. medRxiv (2020).
- 441 3. Peak, C.M., et al. Individual quarantine versus active monitoring of contacts for the
442 mitigation of COVID-19: a modelling study. *Lancet Infect Dis* 20, 1025-1033 (2020).
- 443 4. Helgason, D., et al. Beating the odds with systematic individualized care:
444 Nationwide prospective follow-up of all patients with COVID-19 in Iceland. *Journal of*
445 *Internal Medicine* (2020).
- 446 5. Giordano, G., et al. Modelling the COVID-19 epidemic and implementation of
447 population-wide interventions in Italy. *Nature Medicine*, 1-6 (2020).
- 448 6. Rotondi, V., Andriano, L., Dowd, J.B. & Mills, M.C. Early evidence that social
449 distancing and public health interventions flatten the COVID-19 curve in Italy. (2020).
- 450 7. Torri, E., et al. Italian Public Health Response to the COVID-19 Pandemic: Case
451 Report from the Field, Insights and Challenges for the Department of Prevention. *Int J*
452 *Environ Res Public Health* 17(2020).
- 453 8. Han, H., et al. Multilayered Control of Alternative Splicing Regulatory Networks by
454 Transcription Factors. *Mol Cell* 65, 539-553 e537 (2017).
- 455 9. Esbin, M.N., et al. Overcoming the bottleneck to widespread testing: A rapid review
456 of nucleic acid testing approaches for COVID-19 detection. *RNA*, rna. 076232.076120
457 (2020).

- 458 10. Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C. & Garry, R.F. The
459 proximal origin of SARS-CoV-2. *Nat Med* 26, 450-452 (2020).
- 460 11. Walls, A.C., et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike
461 Glycoprotein. *Cell* 181, 281-292 e286 (2020).
- 462 12. BGI Genomics Co. Ltd. Shenzhen, C. Real-Time Fluorescent RT-PCR Kit for
463 Detecting SARS-CoV-2. (2020).
- 464 13. Seegene Inc., S., South Korea. Allplex™ 2019-nCoV Assay (version 2.0). (2020).
- 465 14. Farouni, R., Djambazian, H., Ferri, L.E., Ragoussis, J. & Najafabadi, H.S. Model-
466 based analysis of sample index hopping reveals its widespread artifacts in multiplexed
467 single-cell RNA-sequencing. *Nat Commun* 11, 2704 (2020).
- 468 15. López Ratón, M., Rodríguez Álvarez, M.X., Cadarso Suárez, C.M. & Gude
469 Sampedro, F. Optimal Cutpoints: an R package for selecting optimal cutpoints in
470 diagnostic tests. (American Statistical Association, 2014).
- 471 16. Pearson, J.D., et al. Comparison of SARS-CoV-2 Indirect and Direct Detection
472 Methods. *bioRxiv* (2020).
- 473 17. Li, Q., et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and
474 antigenicity. *Cell* 182, 1284-1294. e1289 (2020).
- 475 18. Bloom, J.S., et al. Swab-Seq: A high-throughput platform for massively scaled up
476 SARS-CoV-2 testing. *medRxiv* (2020).
- 477 19. Jacot, D., Greub, G., Jatou, K. & Opota, O. Viral load of SARS-CoV-2 across
478 patients and compared to other respiratory viruses. *Microbes and Infection* (2020).
- 479 20. He, X., et al. Temporal dynamics in viral shedding and transmissibility of COVID-
480 19. *Nature medicine* 26, 672-675 (2020).

481 21. Mina, M.J., Parker, R. & Larremore, D.B. Rethinking Covid-19 Test Sensitivity - A
482 Strategy for Containment. *N Engl J Med* (2020).

483 22. Wu, F., et al. A new coronavirus associated with human respiratory disease in
484 China. *Nature* 579, 265-269 (2020).

485 23. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-
486 efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25
487 (2009).

488

489 **Acknowledgments**

490 The authors thank Drs. Rita Kandel and Jim Woodgett for discussions. The authors are
491 grateful to Tanja Durbic and Kyle Turner in the Donnelly Sequencing Center for
492 sequencing samples and Kathy Fung at the Network Biology Collaboration Center.

493

494 **Funding**

495 This work is supported by the Toronto COVID-19 Action Initiative (TCAI) Fund from the
496 University of Toronto awarded to J.L.W, L.P. and R.B., and by a donation from the Krembil
497 Foundation (SHSF Krembil SARS-COV-2) to J.L.W, L.P. and R.B.

498

499 **Author contributions**

500 J.L.W., M.M.A, J.J.H., H.H., U.B., B.B. and S.B. designed the study. M.M.A, J.J.H.
501 performed C19-SPAR-Seq experiments. S.B. performed non-specific amplicon/mutation
502 analysis, S.B. and U.B. performed NGS analysis and established C19-SparSeq
503 interpretation pipeline, H.H. performed unsupervised clustering, M.M.A. and J.J.H.

504 assisted with the rest of the analysis. D.T. collected the samples and purified the RNA.
505 J.P.D. performed qPCR control studies ('BGI') under supervision of R.B. K.C. performed
506 sequencing. M.B. and M.J. assisted with automation optimization. S.L.P., K.J., and S.S.
507 prepared control samples under supervision of L.P. and L.A.. T.M., S.P, C.B., and B.H.
508 provided access to patient samples, collection of diagnostics information, and assembly
509 of the cohorts. All experiments were carried out under the supervision of B.B, L.P and
510 J.L.W. The manuscript was written by M.M.A., J.J.H., S.B., U.B., L.P. and J.L.W. with
511 input from R.B., T.M., S.P., L.A., H.H., and B.B..

512

513 **Ethical declarations**

514 **Competing Interests statement**

515 J. Wrana is founder and CEO of iTP Biomedica Inc, which employs whole transcriptome
516 NGS tests in cancer, and he is founder and consultant for Fibrocor LP, which is
517 developing therapeutics for fibrotic disease. The other authors declare no competing
518 interests.

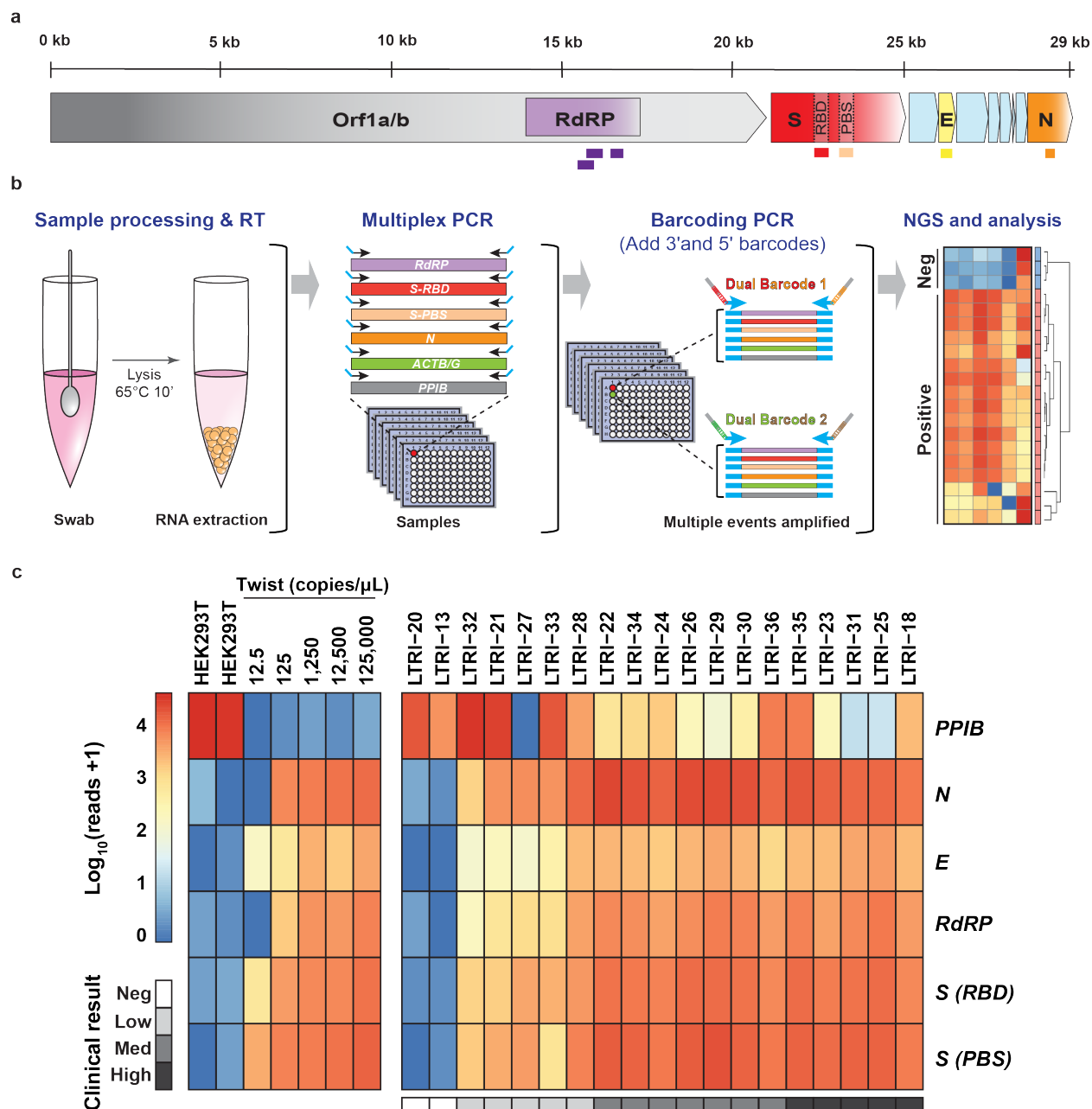


Fig. 1

519

520 **Fig. 1: Application of C19-SPAR-Seq to detect SARS-CoV-2.** a, Schematic

521 representation of the SARS-CoV-2 with the 5 regions targeted for multiplex C19-SPAR-

522 Seq indicated: *RdRP* (purple), *S receptor binding domain (Srbd)* (red), *S polybasic*

523 *cleavage site (Spbs)* (light red), *E* (yellow), and *N* (orange). b, Schematic of the C19-

524 SPAR-Seq strategy for detecting SARS-CoV-2. cDNA is synthesized using reverse
525 transcriptase (RT) from RNA extracted from clinical samples, subjected to multiplex PCR,
526 then barcoded, pooled and analyzed by next generation sequencing (NGS). **c**, Analysis
527 of archival NASOP swab eluents by C19-SPAR-Seq. A Proof-of-Concept (PoC) cohort (n
528 = 19) was analyzed by C19-SPAR-Seq, and read numbers for each of the indicated
529 amplicons are presented in a heatmap. Control samples (HEK293T, synthetic SARS-
530 CoV-2 RNA) are represented in the left panel, while the right panel shows unsupervised
531 2D hierarchical clustering of results from negative (blue) and positive (red) patients.

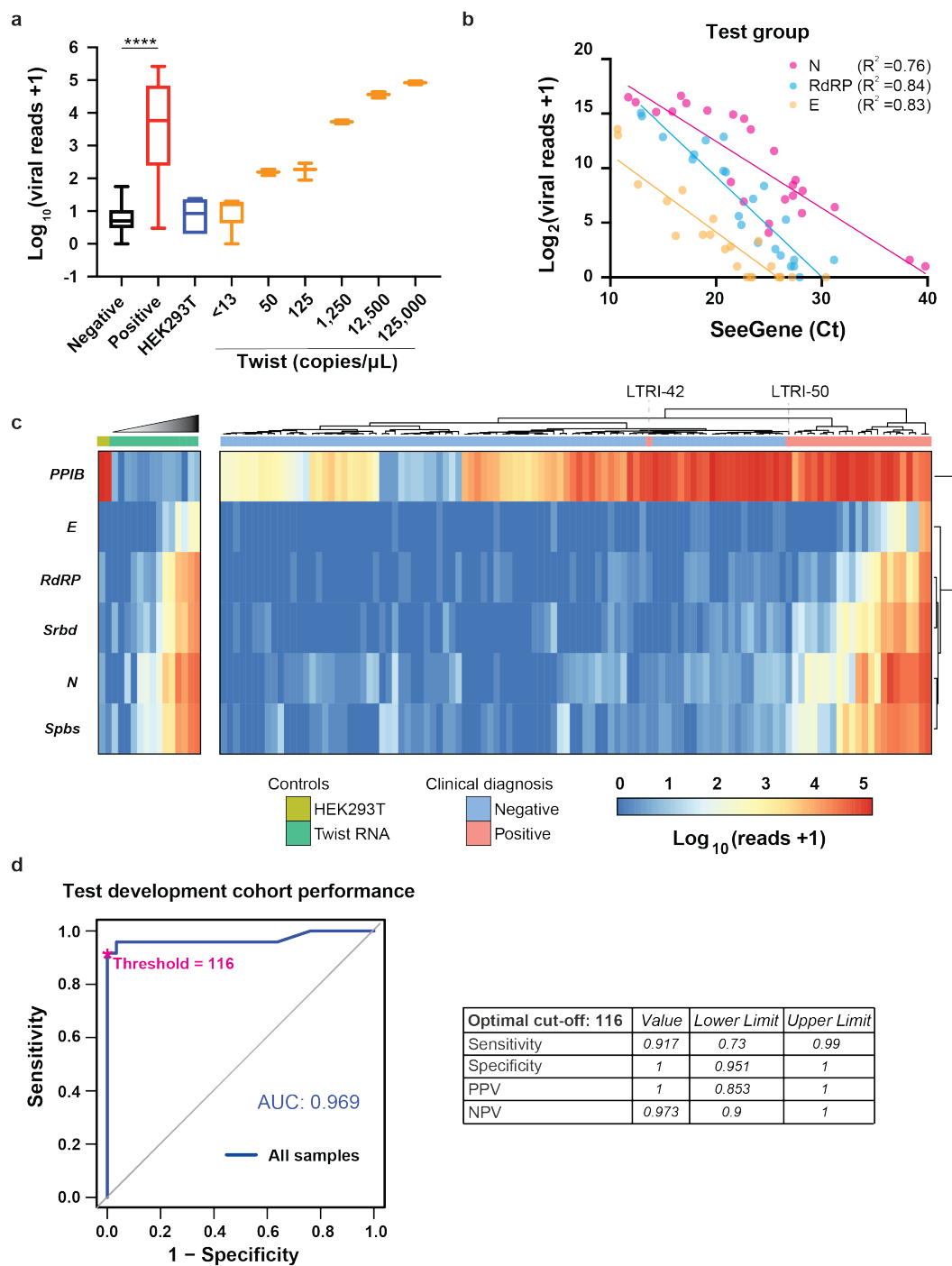


Fig. 2

532

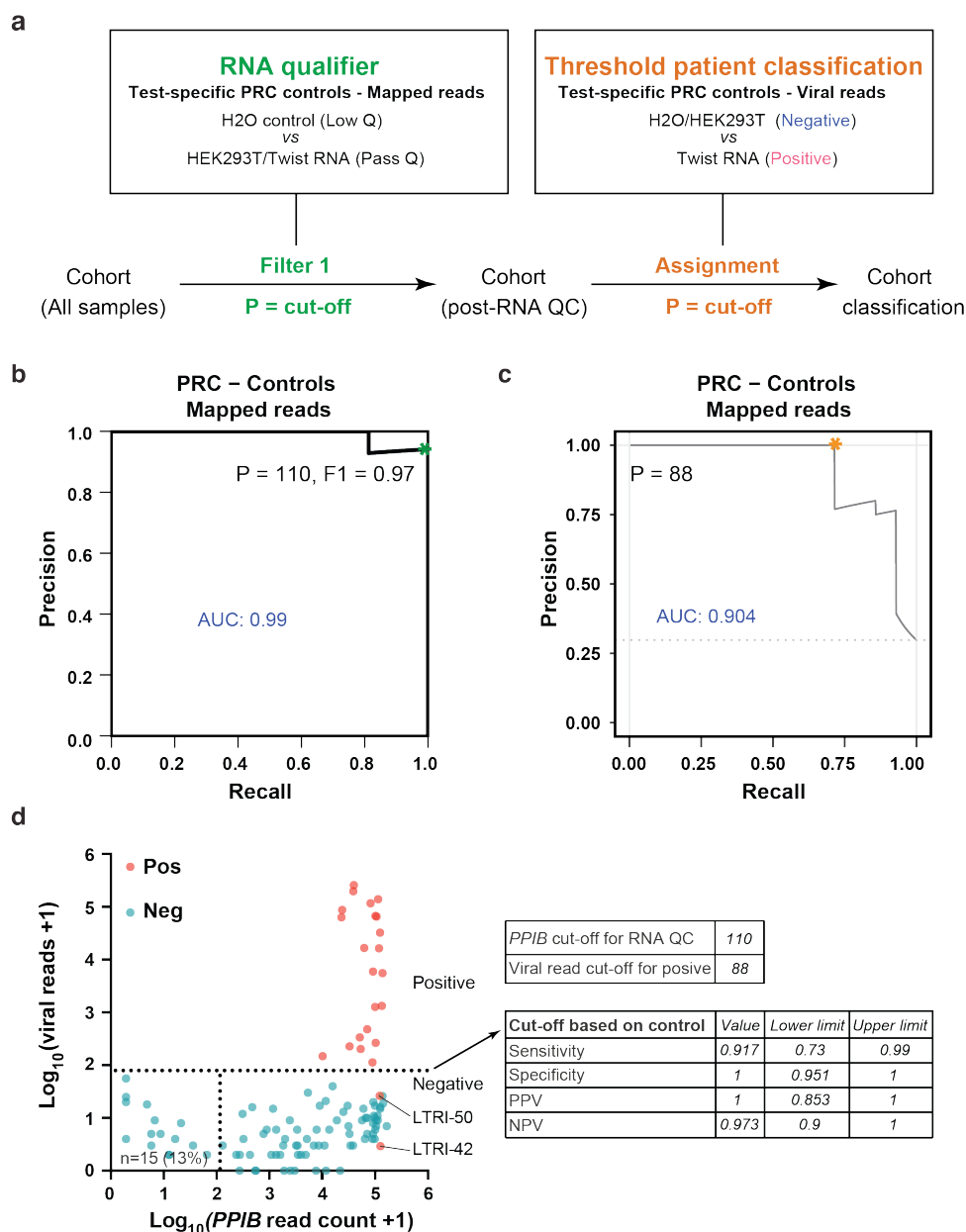
533 **Fig. 2: Performance of C19-SPAR-Seq in detecting SARS-CoV-2.** a, C19-SPAR-Seq

534 of the test development cohort was performed and total viral reads+1 (log₁₀) (Y-axis) are

535 plotted for negative (n = 88, black) and positive (n = 24, red) patient samples, HEK293T

536 RNA (n = 6, blue), and the indicated serial dilutions of synthetic SARS-CoV-2 RNA (n=2-

537 6, orange). For each group, the median, lower and upper confidence limit for 95% of the
538 median are plotted. Whiskers are minimum and maximum values. Unpaired *t*-test of
539 negative *versus* positive samples (****: $p < 0.0001$) **b**, C19-SPAR-Seq reads for the
540 indicated gene in each patient sample were compared to Ct values obtained by the clinical
541 diagnostics lab using the 'Seegene' Allplex assay. **c**, Heatmap of C19-SPAR-Seq results.
542 Read counts for the indicated target amplicons in control samples ($n = 16$; left) and patient
543 samples ($n = 112$; right) are plotted according to the scale, and sample types labelled as
544 indicated. Samples are arranged by hierarchical clustering with euclidean distance
545 indicated by the dendrogram on the top, which readily distinguishes positive from negative
546 samples. **d**, Performance of C19-SPAR-Seq. ROC analysis on patient samples was
547 performed using clinical diagnostic results (Seegene Allplex qRT-PCR assay,
548 **Supplementary Table 3**) and total viral reads for patient samples ($n = 112$). AUC (area
549 under the curve) scores are indicated on the graph (left), with statistics at the optimal
550 cutoff as indicated (right).



551 Fig. 3

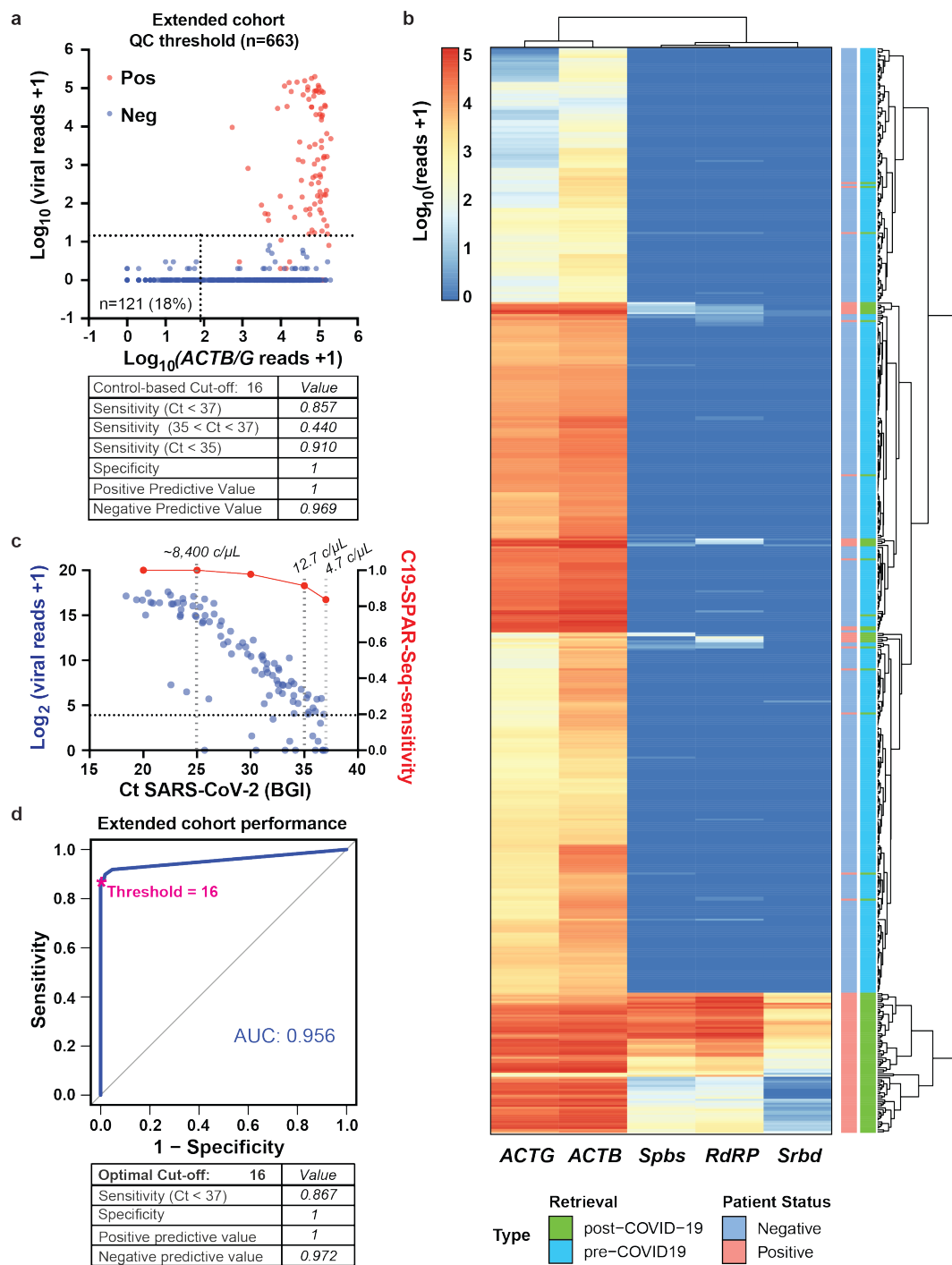
552 **Fig. 3: Performance of C19-SPAR-Seq in detecting SARS-CoV-2 using control**

553 **based classifier.** **a**, Schematic of the control based cut-off procedure for RNA quality

554 and viral threshold by coPR analysis. **b**, Thresholding sample quality. coPR analysis on

555 control samples: PRC of control samples for accurate detection of mapped reads are

556 plotted. The optimal precision and recall read cut-off associated ($P = 110$) with the highest
557 F1 (0.97) score, and AUC (area under the curve) are indicated in the PR plot. . **c**,
558 Threshold for classification of positives in the test cohort. Optimum cut-off for viral
559 threshold is calculated by PROC01 using clinical diagnosis and total viral reads, and
560 plotted on the precision-recall curve. **d**, Threshold assignments for sample quality and
561 classification. Total viral reads +1 (Y-axis) are plotted against *PPIB* reads +1 (X-axis) for
562 positive (red) and negative (blue) patient samples. coPR-based RNA-QC filter and viral
563 read filter are shown as indicated. Assay statistics using coPR thresholding are listed
564 (right).



570 in the table. **b**, Heatmap of C19-SPAR-Seq results. Read counts for the indicated target
571 amplicons in the filtered set of samples (n = 542) are plotted according to the scale, and
572 sample types labelled as indicated. Samples are arranged by hierarchical clustering with
573 euclidean distance indicated by the dendrogram on the right. **c**, Scatter plot of total viral
574 reads+1 (left Y-axis, blue) *versus* Ct values of positive samples (n = 98, BGI) (X-axis).
575 C19-SPAR-seq sensitivity at the indicated Ct values is overlaid (right Y-axis, red). Gray
576 dashed lines indicate average copies/ μ L (c/ μ L) **d**, ROC curve analysis. ROC curves were
577 processed on filtered samples (n = 542). AUC scores are indicated for filtered samples
578 (blue; left) with corresponding performance statistics for the optimal cut-off indicated
579 below.

1 **A Multiplexed, Next Generation Sequencing Platform for High-Throughput**

2 **Detection of SARS-CoV-2**

3 Marie-Ming Aynaud^{1**}, J. Javier Hernandez^{1,3**}, Seda Barutcu^{1**}, Ulrich Braunschweig^{3,4},

4 Kin Chan¹, Joel D. Pearson¹, Daniel Trcka¹, Suzanna L. Prosser¹, Jaeyoun Kim¹, Miriam

5 Barrios-Rodiles¹, Mark Jen¹, Siyuan Song^{4,5}, Jess Shen¹, Christine Bruce², Bryn Hazlett²,

6 Susan Poutanen², Liliana Attisano^{4,5}, Rod Bremner¹, Benjamin J. Blencowe^{3,4}, Tony

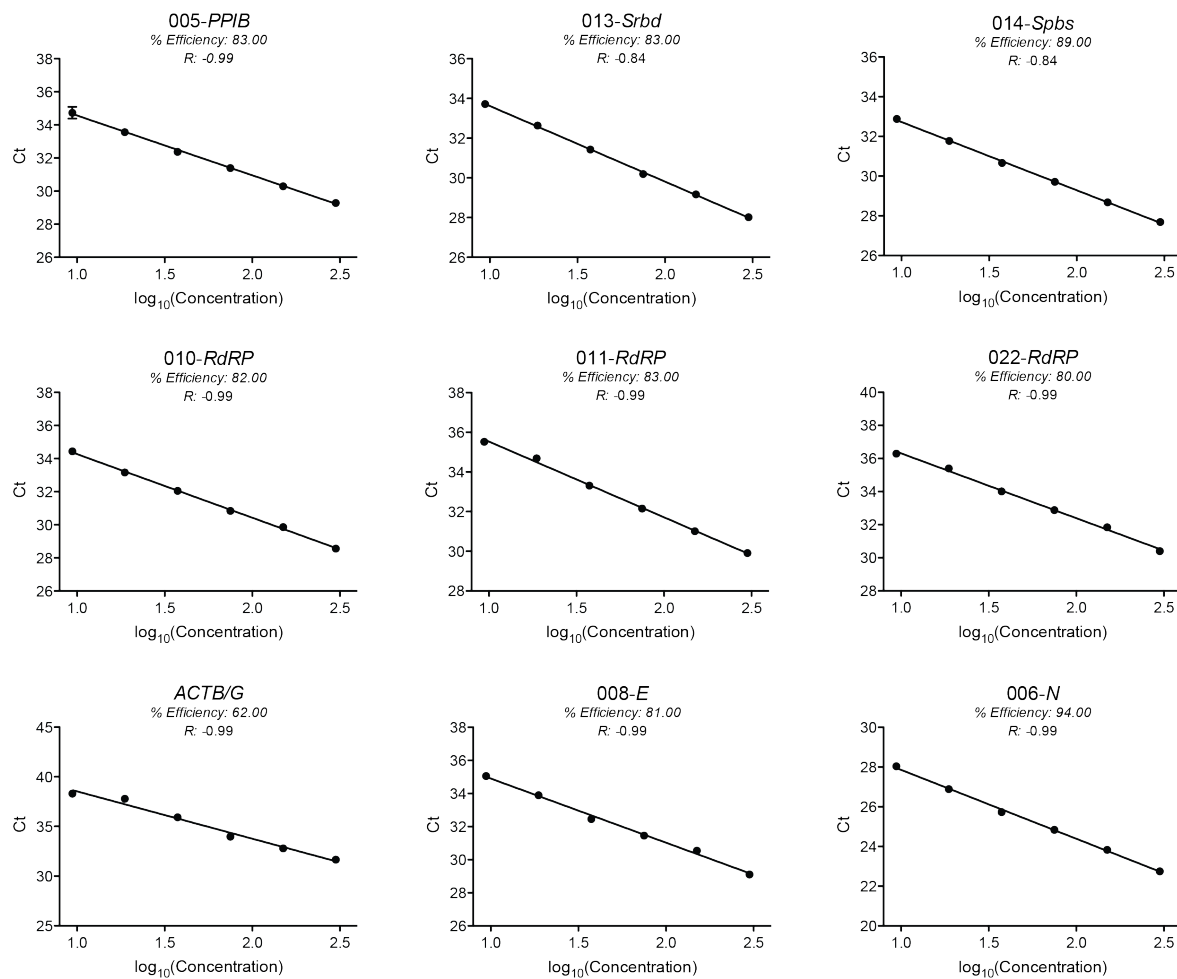
7 Mazzulli², Hong Han^{3,4}, Laurence Pelletier^{1,3*}, Jeffrey L. Wrana^{1,3*}.

8

9 **Extended data Figures**

10

11

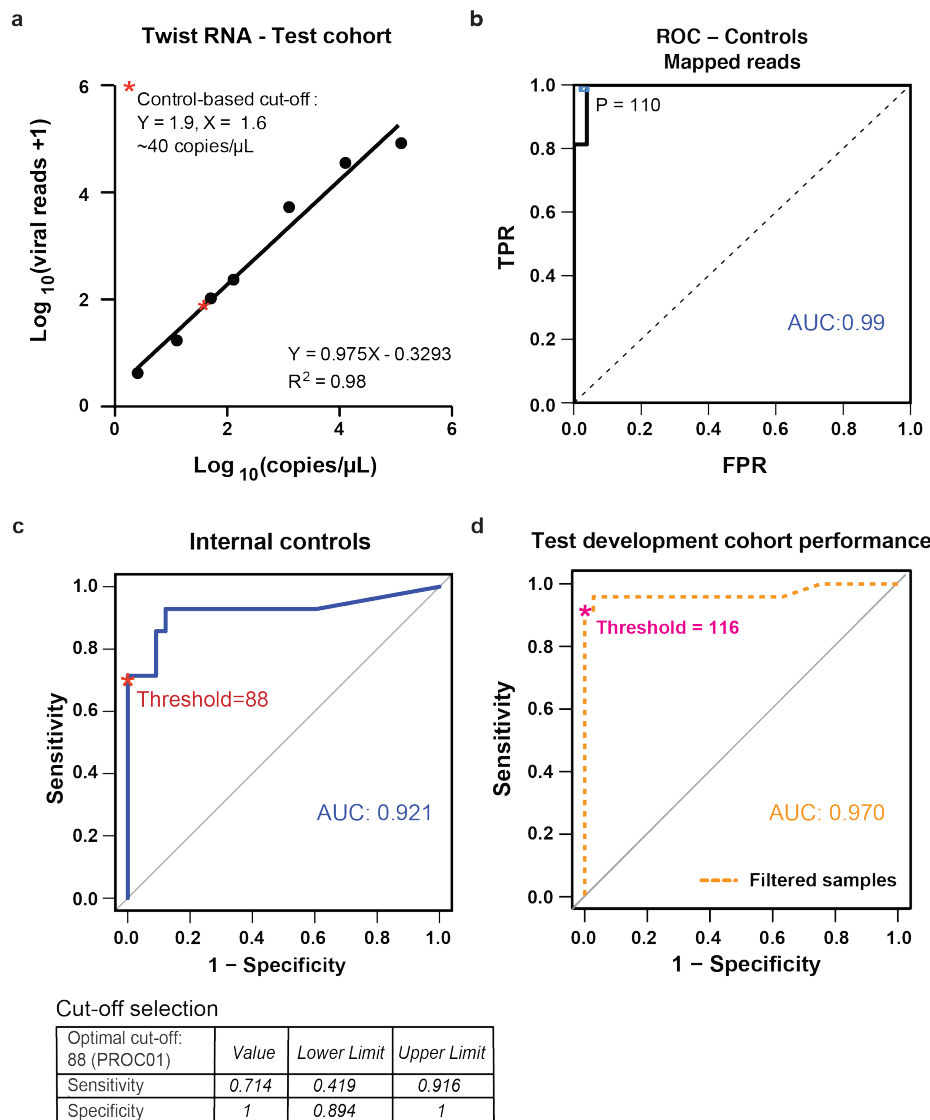


12

Extended data Fig. 1

13 **Extended data Fig. 1: Efficiency of multiplex primers.** Standard curve of Ct values (Y-
14 axis) and log₁₀(Concentration) (X-axis) of 6 limited dilutions of SARS-CoV-2^{high} sample
15 (LTRI-18) for 9 pairs of primers (see Supplementary Table S1). Each condition was tested

- 16 in duplicate. Means are plotted for each point. The percent efficiency and the correlation
- 17 (r) are calculated for each pair of primers after linear regression.



Extended data Fig. 2

18

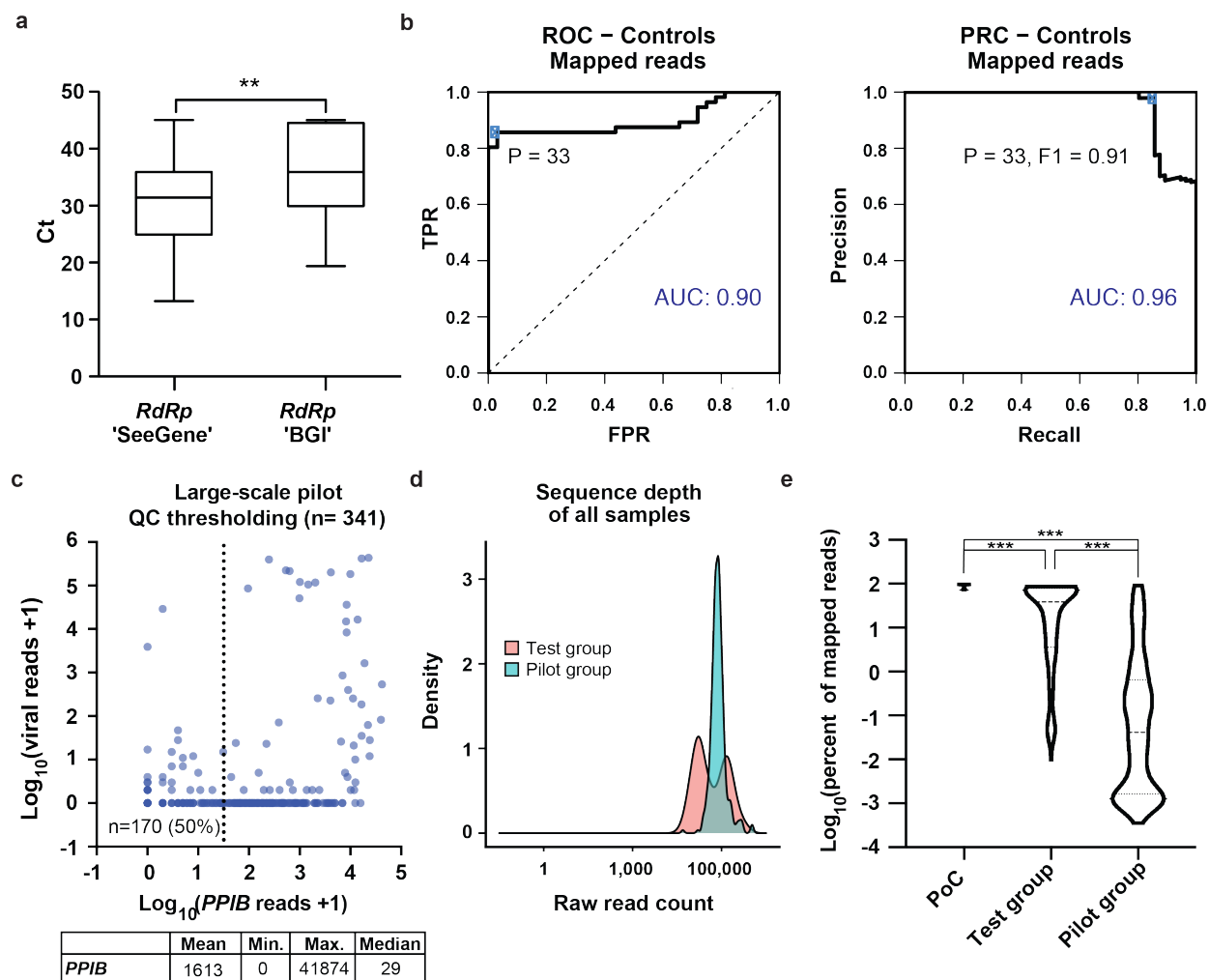
19 **Extended data Fig. 2: Using embedded controls as a training set for a control-based**

20 **PR and ROC classifier.** a, Total viral read counts are plotted against estimated viral

21 copies (copies/ μL) obtained using synthetic Twist SARS-CoV-2 RNA with statistics

22 indicated. The cutoff defined by PROC analysis (see panel c) is marked with a red

23 asterisk. **b**, Thresholding sample quality. coPR analysis on control samples: ROC of
24 control samples for accurate detection of mapped reads are plotted. The optimal precision
25 and recall read cut-off associated ($P = 110$) with the highest F1 (0.97) score, and AUC
26 (area under the curve) is indicated on the ROC plot. **c**, Threshold for classification of
27 positives in the test cohort. Total viral reads of negative (H₂O and HEK293T) and positive
28 (Twist dilutions) samples are used to calculate optimum cut-off by PROC and the defined
29 threshold ($P = 88$) is plotted on the ROC curve. Values of sensitivity, and specificity at
30 this cut-off are indicated (below). **d**, Performance of C19-SPAR-Seq. ROC analysis on
31 patient samples that passed RNA-QC threshold was performed using clinical diagnostic
32 results (Seegene Allplex qRT-PCR assay, **Supplementary Table 3**) and total viral reads
33 for patient samples ($n = 112$). AUC is indicated on the graph.



Extended data Fig. 3

34

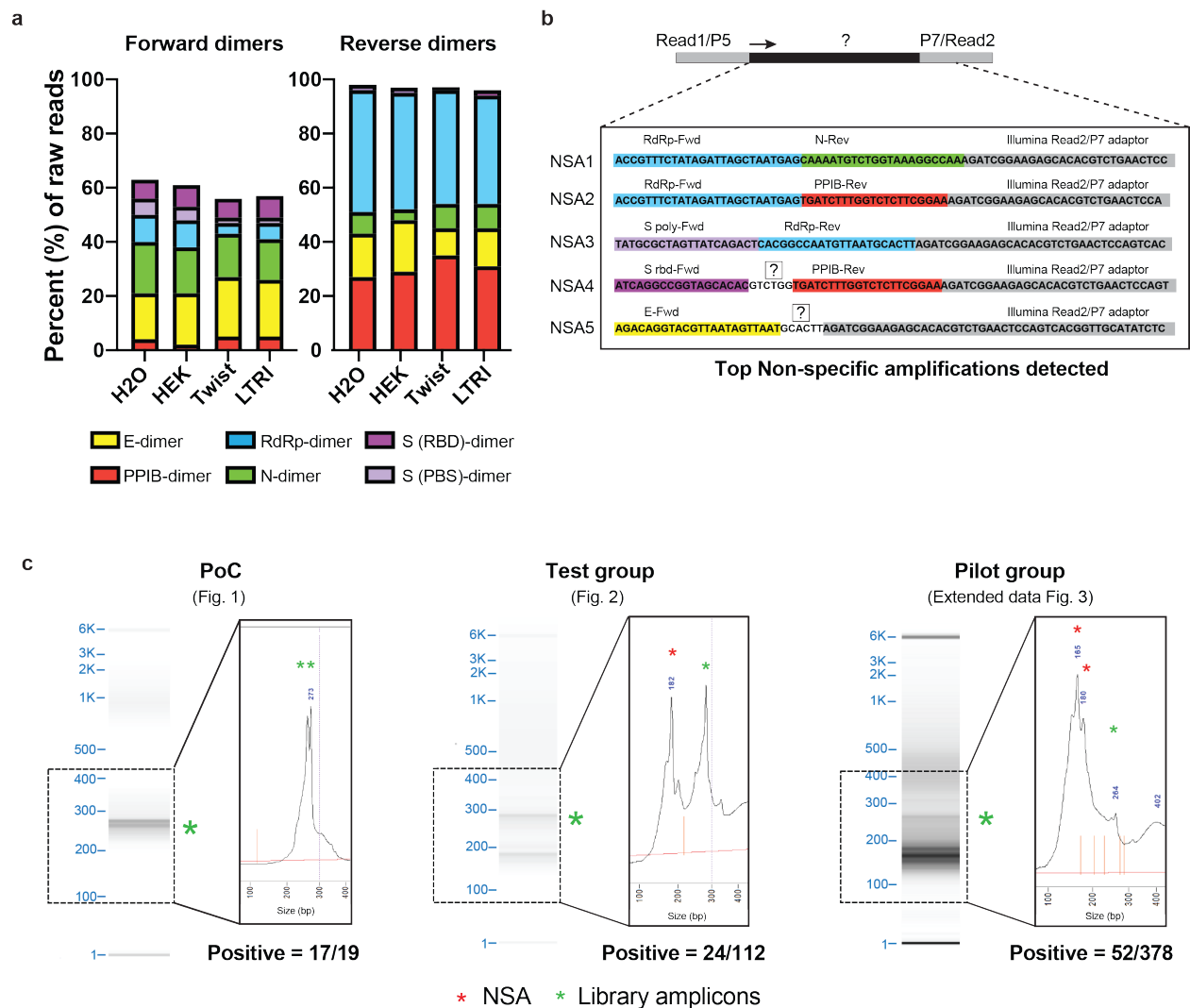
35 **Extended data Fig. 3: Quality metrics assignment for the pilot cohort.** a, Comparison

36 of Ct (*RdRP*) values in 'SeeGene' versus 'BGI' tests of the positive archival samples. b,

37 coPR analysis on control samples. ROC and PRC of control samples are plotted and the

38 optimal precision and recall cut-off ($P = 33$) associated with the highest F1 score (0.91)

39 was calculated, as indicated in the PRC plot. **c**, coPR thresholding of the pilot cohort. Plot
40 of total viral reads +1 (Y-axis) versus *PPIB* reads +1 (X-axis) of 341 patient samples in a
41 pilot cohort (see Methods) is shown with the threshold (*PPIB* read counts > 33) to filter
42 low-input samples marked. 170/341 (50%) samples were inconclusive (upper panel).
43 Mean, minimum, maximum, and median values of *PPIB* and total viral read counts are
44 indicated in the table (lower panel). **d**, Sequencing depth of test development and pilot
45 cohort. Distribution density of raw read counts for the test development (pink) and pilot
46 (turquoise) cohorts are shown. **e**, Read mapping percentages. Comparison of overall read
47 mapping percentages between the PoC (**Fig. 1**), test (**Fig. 2**) and pilot cohort (n = 341).
48 One way ANOVA - Tukey's multiple comparison test (****: $p < 0.0001$).



Extended data Fig. 4

49

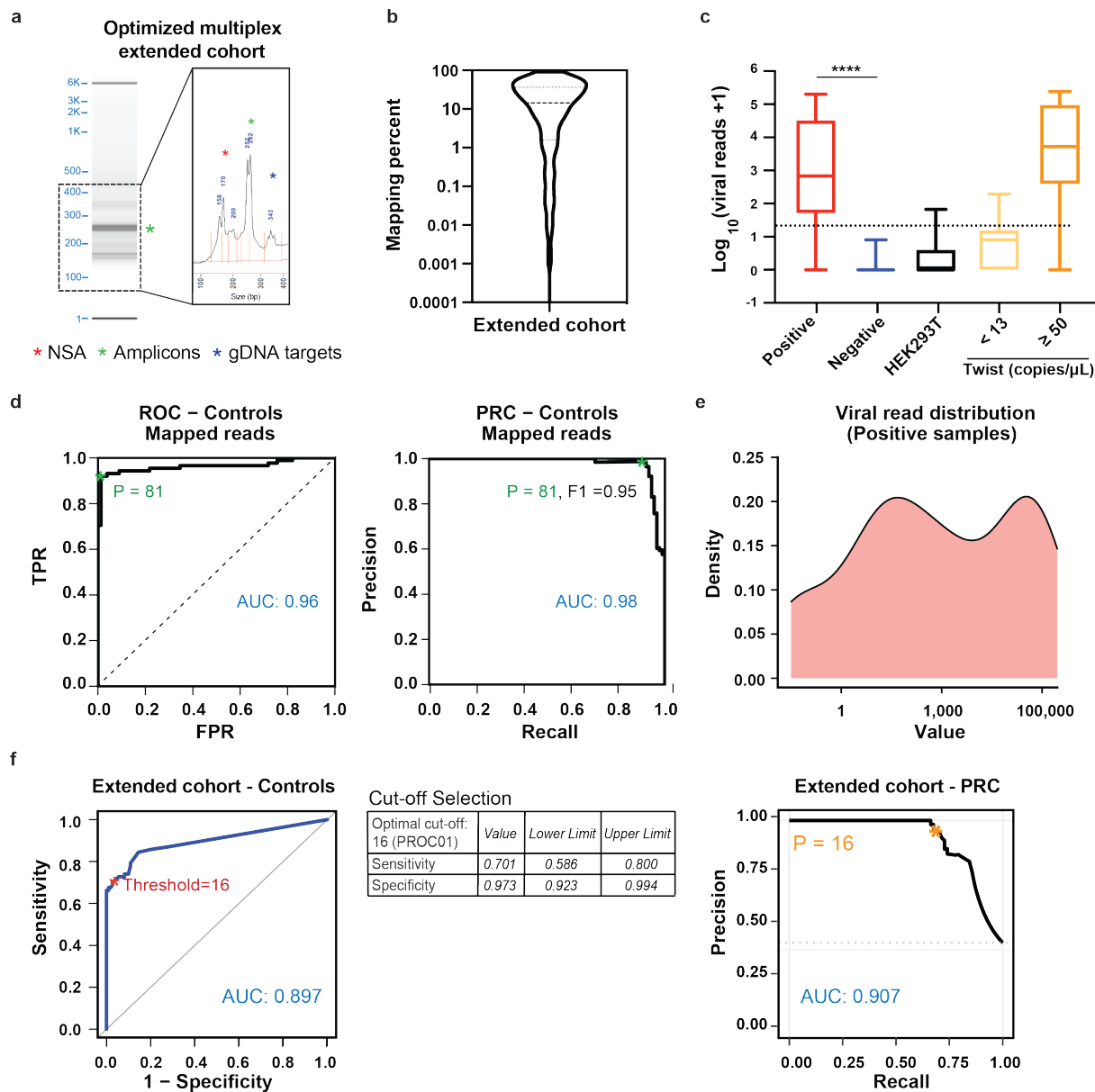
50 **Extended data Fig. 4: Non-specific amplification (NSA) in pilot cohort.** a, Analysis of

51 NSAs in the pilot cohort. NSAs contaminating the C19-SPAR-Seq library were quantified

52 and percentage of reads mapping to the indicated forward and reverse primers are

53 plotted. b, Schematic examples and sequences of the top 5 NSAs are shown. c,

54 Comparison of fragment analyzer profile of the PoC, test development, and pilot cohort
55 libraries after 0.8X SPRI bead purification. Fragment separation (DNA gel) and blow up
56 view of the product abundance (electropherogram) are shown. Expected library
57 amplicons (green stars) and non-specific amplicons (red stars).



Extended data Fig. 5

58

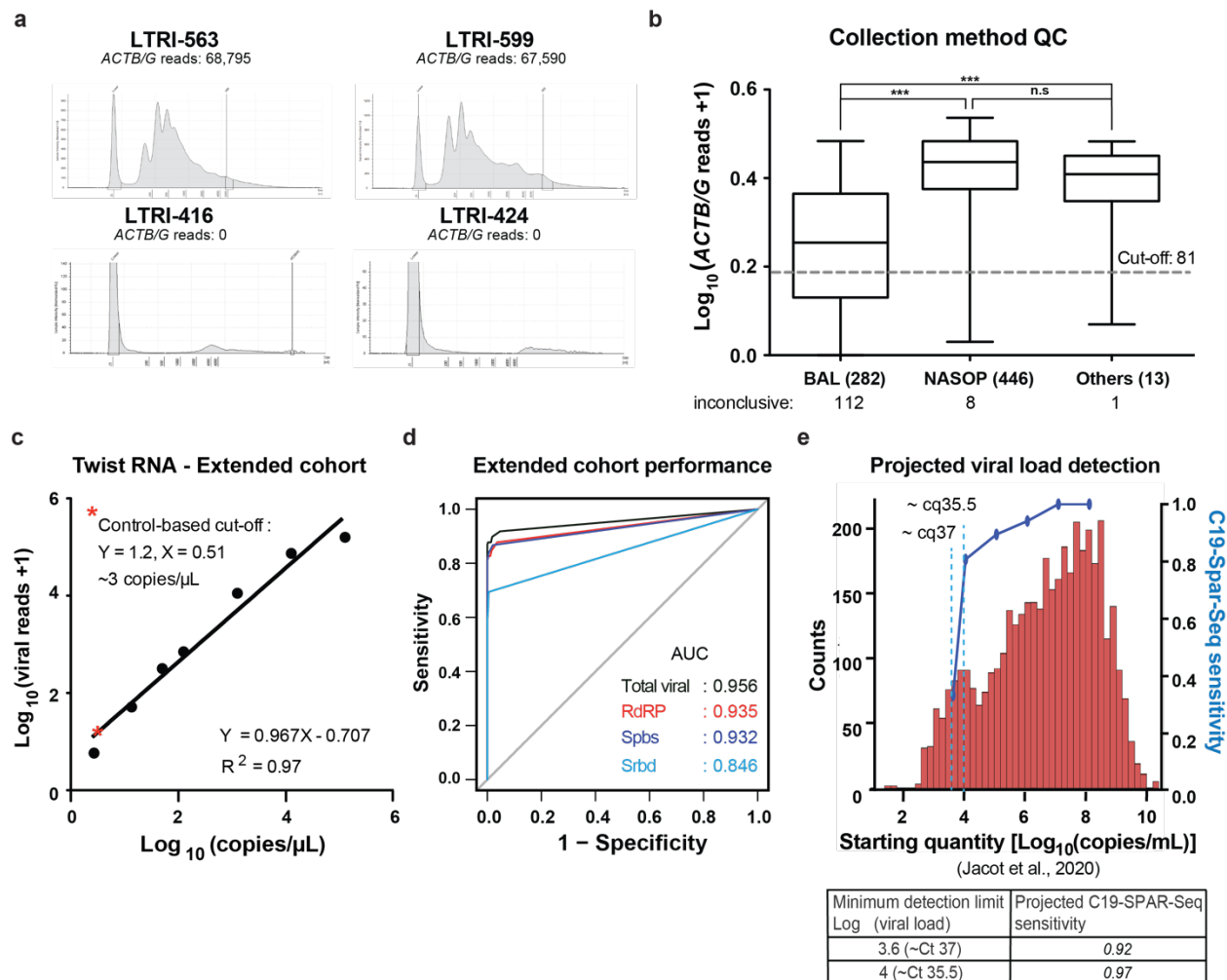
59 **Extended data Fig. 5: Suppressing non-specific amplicons and quality metrics**

60 **assignment for the extended cohort.** a, Fragment analyzer profile of the extended

61 cohort library using an optimized multiplex primer set targeting *ACTB/G*, *Spoly*, *Srbd*, and

62 *RdRP*. Fragment separation (DNA gel) and blow up view of the product abundance

63 (electropherogram) is shown. **b**, Mapping percentage of the extended cohort. **c**, Overall
64 distribution of total viral reads in the indicated positive samples (n = 98, red), negative
65 samples (n = 444, blue), HEK293T (n = 21, black), synthetic SARS-CoV-2-RNA (< 13.2
66 copies/ μ L, n = 6, yellow), and synthetic SARS-CoV-2-RNA (\geq 50 copies/ μ L, n = 30,
67 orange) are plotted. Unpaired *t*-test of negative *versus* positive samples (****: $p < 0.0001$).
68 **d**, coPR thresholding of sample quality and classification in the extended cohort. coPR
69 analysis on control samples for sample quality yielded an optimal precision and recall
70 read cut-off (P = 81) as indicated. **e**, Distribution of \log_{10} total reads +1 of the positive (n
71 = 98) samples. **f**, Threshold for classification of the extended cohort. ROC on control
72 samples (HEK293T and synthetic SARS-CoV-2 RNA control) was assessed to identify
73 an optimal cut-off (P = 16) for classifying patient samples. Performance on the controls
74 is summarized.



Extended data Fig. 6

75

76 **Extended data Fig. 6: C19-SPAR-Seq performance.** **a**, RNA profile of BALs. RNA

77 purified from ten BALs above and 10 below the QC threshold was profiled and two

78 representative traces of each group are shown. *ACTB/G* reads are indicated for each

79 sample. **b**, *ACTB/G* reads according to collection type. *ACTB/G* reads are plotted for each

80 collection type as a box and whisker (median \pm 95% confidence interval, and the
81 maximum and minimum values). The number of samples filtered by coPR (ACTB/G reads
82 < 81) are indicated for each group. 1way ANOVA - Tukey's multiple comparison test (****:
83 $p < 0.0001$, ns: non significative) **c**, Standard curve of total viral reads plotted against
84 synthetic SARS-CoV-2 RNA concentrations obtained from C19-SPAR-Seq analysis of
85 the extended cohort. **d**, ROC curve analysis was performed for each of the indicated viral
86 amplicons and the AUC is shown. **e**, Projection of our C19-SPAR-Seq sensitivity onto the
87 viral load data of ~4,000 patients from Jacot *et al.*, 2020 study¹⁹. Minimum detection limit
88 and C19-SPAR-Seq sensitivity values are indicated in the table below.

1 **A Multiplexed, Next Generation Sequencing Platform for High-Throughput**

2 **Detection of SARS-CoV-2**

3 Marie-Ming Aynaud^{1**}, J. Javier Hernandez^{1,3**}, Seda Barutcu^{1**}, Ulrich Braunschweig^{3,4},
4 Kin Chan¹, Joel D. Pearson¹, Daniel Trcka¹, Suzanna L. Prosser¹, Jaeyoun Kim¹, Miriam
5 Barrios-Rodiles¹, Mark Jen¹, Siyuan Song^{4,5}, Jess Shen¹, Christine Bruce², Bryn Hazlett²,
6 Susan Poutanen², Liliana Attisano^{4,5}, Rod Bremner¹, Benjamin J. Blencowe^{3,4}, Tony
7 Mazzulli², Hong Han^{3,4}, Laurence Pelletier^{1,3*}, Jeffrey L. Wrana^{1,3*}.

8
9 **Supplementary information**

10 **Supplementary Table 1: List of SARS-CoV-2 and human primers.** Primer sequences,
11 name of the targeted regions, size of amplicons after multiplex and barcode PCR are
12 indicated.

13
14 **Supplementary Table 2: Itemized cost of C19-SPAR-Sseq per sample**

15
16 **Supplementary Table 3: Description of the proof-of-concept cohort for C19-SPAR-**
17 **Seq detection of SARS-CoV-2.** Barcodes ID, sample identification (ID), date of retrieval,
18 collection method, diagnostic laboratory status, and 'BGI' qRT-PCR results are indicated.
19 These patient samples were used to develop C19-SPAR-Seq detection of SARS-CoV-2
20 (PoC cohort) (**Fig. 1**).

21
22 **Supplementary Table 4: Description of test development cohort.** Barcodes ID,
23 sample identification (ID), date of retrieval, collection method, diagnostic laboratory qRT-

24 PCR results ('Seegene') are indicated (n = 112). These patient samples were used to
25 establish SARS-CoV-2 clinical status assignment using diagnostic laboratory qRT-PCR
26 results ('Seegene') and to test C19-SPAR-Seq detection of SARS-CoV-2 (**Fig. 2,3**).

27

28 **Supplementary Table 5: Confusion matrix of the test development cohort.**

29

30 **Supplementary Table 6: Description of the pilot cohort.** Barcodes ID, sample
31 identification (ID), date of retrieval, collection method, diagnostic laboratory qRT-PCR
32 results ('Seegene'), 'BGI' qRT-PCR results are indicated. Filtered archival samples are
33 indicated. (Extended data **Fig. 3,4**).

34

35 **Supplementary Table 7: Description of the extended cohort.** Barcodes ID, sample
36 identification (ID), date of retrieval, collection method, diagnostic laboratory qRT-PCR
37 results ('Seegene'), and 'BGI' qRT-PCR results are indicated (**Fig. 4**).

38

39 **Supplementary Table 8: Confusion matrix of the extended cohort.**

40

41 **Supplementary Table 9: Group classifications**