

1 **In-House, Rapid, and Low-Cost SARS-CoV-2 Spike Gene Sequencing Protocol to Identify**  
2 **Variants of Concern Using Sanger Sequencing**

3 Fatimah S. Alhamlan<sup>1,2,3</sup>, Dana M. Bakheet<sup>2,4</sup>, Marie F. Bohol<sup>1</sup>, Madain S. Alsanea<sup>1</sup>, Basma M.  
4 Alahideb<sup>1</sup>, Faten M. Alhadeq<sup>1</sup>, Feda A. Alsuwairi<sup>1</sup>, Maha Al-Abdulkareem<sup>1</sup>, Mohamed S. Asiri<sup>1</sup>,  
5 Reem S. Almaghrabi<sup>5</sup>, Sarah A. Altamimi<sup>3</sup>, Maysoun S. Mutabagani<sup>3</sup>, Sahar I. Althawadi<sup>3</sup>,  
6 Ahmed A. Alqahtani<sup>1,2\*</sup>

7

8 <sup>1</sup>Department of Infection and Immunity, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

9 <sup>2</sup>College of Medicine, Alfaisal University, Riyadh, Saudi Arabia

10 <sup>3</sup>Department of Pathology and Laboratory Medicine, King Faisal Specialist Hospital and Research Center, Riyadh,  
11 Saudi Arabia

12 <sup>4</sup>Center for Genomic Medicine, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

13 <sup>5</sup>Organ Transplant Center of Excellence, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

14

15

16

17

18

19

20

21

22

23

24

25

26 **\*Corresponding author.** Ahmed A. Al-Qahtani. Tel.: 966 11 442 4550, Fax: 966 11 442 4519.

27 Email address: [aqahatani@kfshrc.edu.sa](mailto:aqahatani@kfshrc.edu.sa)

28 **Abstract**

29 **Background:** The need for active genomic sequencing surveillance to rapidly identify  
30 circulating SARS-CoV-2 variants of concern (VOCs) is critical. However, increased global  
31 demand has led to a shortage of commercial SARS-CoV-2 sequencing kits, and not every  
32 country has the technological capability or the funds for high-throughput sequencing platforms.  
33 Therefore, this study aimed to develop and validate a rapid, cost-efficient genome sequencing  
34 protocol that uses supplies, equipment, and methodologic expertise available in standard  
35 molecular or diagnostic laboratories to identify circulating SARS-CoV-2 variants of concern.

36 **Methods:** Sets of primers flanking the SARS-CoV-2 spike gene were designed using SARS-  
37 CoV-2 genome sequences retrieved from the Global Initiative on Sharing Avian Influenza Data  
38 (GISAID) Database and synthesized in-house. Primer specificity and final sequences were  
39 verified using online prediction analyses with BLAST. The primers were validated using 282  
40 nasopharyngeal samples collected from patients assessed as positive for SARS-CoV-2 at the  
41 diagnostic laboratory of the hospital. The patient samples were subjected to RNA extraction  
42 followed by cDNA synthesis, conventional polymerase chain reaction, and Sanger sequencing.  
43 Protocol specificity was confirmed by comparing these results with SARS-CoV-2 whole genome  
44 sequencing of the same samples.

45 **Results:** Sanger sequencing using the newly designed primers and next-generation whole  
46 genome sequencing of 282 patient samples indicated identical VOCs results: 123 samples  
47 contained the alpha variant (B.1.1.7); 78, beta (B.1.351), 0, gamma (P.1), and 13, delta  
48 (B.1.617.2). The remaining samples were not 100% identical to the reference genome; however,  
49 99.97% identity indicated that there was minimal variation as the virus spread throughout the

50 nation. Only four samples had poor sequence quality by Sanger sequencing owing to a low RNA  
51 count (Ct value >38). Therefore, mutation calls were >98% accurate.

52 **Conclusions:** Sanger sequencing method using in-house primers is an alternative approach that  
53 can be used in facilities with existing equipment to mitigate limitations in high throughput  
54 supplies required to identify SARS-CoV-2 variants of concern during the COVID-19 pandemic.  
55 This protocol is easily adaptable for detection of emerging variants.

56

57 **Keywords:** Genomic Surveillance, Sanger Sequencing, Oligonucleotide Synthesis, SARS-CoV-  
58 2, VOCs, Delta

59

## 60 **Background**

61 In March 2020, the World Health Organization (WHO) declared a global pandemic caused by  
62 SARS-CoV-2. As of July 26, 2021, the virus had infected more than 194 million people  
63 worldwide and caused the deaths of 4.16 million individuals [1]. As of July 2021, there were  
64 four SARS-CoV-2 variants of concern and nine variants under investigation. The variants of  
65 concern were the alpha (B.1.1.7), beta (B.1.351), gamma (P.1), and delta (B.1.617.2) variants  
66 where variant alpha had been reported in 178 countries, beta in 123 countries, gamma in 75  
67 countries, and delta in 111 countries [2].

68

69 SARS-CoV-2 is a single-stranded RNA-enveloped virus. An RNA-based metagenomic next-  
70 generation sequencing (NGS) approach was applied to characterize its entire genome, which is

71 29,881 □base pairs in length (GenBank No. MN908947), encoding 9860 amino acids [3]. The  
72 spike (S) gene accounts for 1273 amino acids and consists of a signal peptide (amino acids 1–13)  
73 located at the N-terminus, an S1 subunit (14–685 residues), and an S2 subunit (686–1273  
74 residues); the last two regions are responsible for receptor binding and membrane fusion,  
75 respectively [4]. Thus, the spike gene plays an important role in the attachment of the viral  
76 particle to the receptor cells in the host, range determination, and membrane fusion [5, 6]. We  
77 focused on the spike gene for this study because of its many implications in disease  
78 transmissibility, infectiousness, and immune escape as well as for its potential as a therapeutic or  
79 diagnostic target.

80  
81 Genomic sequencing is the primary tool for revealing the virus blueprint and genetic code. Given  
82 the automation and commercialization of high-throughput DNA sequencing, conducting whole  
83 genome sequencing using NGS technology would be ideal under normal situations with  
84 sufficient supplies. However, owing to low supplies during the COVID-19 pandemic and a lack  
85 of expertise in this technology in some countries, scientists worldwide have turned to the gold  
86 standard in genomic sequencing: Sanger sequencing [7]. Sanger sequencing (with polymerase  
87 chain reaction [PCR]) can be conducted using equipment that typically exists in most standard  
88 molecular or diagnostic laboratories worldwide. We therefore aimed in this work to provide a  
89 rapid and cost-efficient approach to detect circulating variants of concern (VOCs) using a  
90 protocol that can be applied in any standard molecular or diagnostic laboratory. The present  
91 study successfully designed primers in-house targeting the SARS-CoV-2 spike gene and  
92 developed and validated a rapid, cost-efficient protocol for using these primers and conventional

93 PCR combined with Sanger genomic sequencing to timely report the circulating SARS-CoV-2  
94 variants.

95

## 96 **Methods**

### 97 **Design and Synthesis of Oligonucleotide Primers**

98 More than 200 complete SARS-CoV-2 genome sequences were retrieved from the Global  
99 Initiative on Sharing Avian Influenza Data (GISAID) database  
100 (<https://www.epicov.org/epi3/cfrontend#lightbox-814829872>) December 2020 and were aligned  
101 using the Clustal W algorithm of the MegAlign module to identify the conserved regions using  
102 DNASTar software (DNASTAR; Madison, WI). Oligonucleotide primers were designed for the  
103 SARS-CoV-2 spike genes to ensure maximal efficiency and sensitivity. The desired primers  
104 were designed using the consensus sequences from different SARS-CoV-2 isolates from around  
105 the world, and the genome sequence from the first virus detected in Wuhan was used as a  
106 reference (accession No. MN908947). Primer3 online tools (version 4.1.0)  
107 (<https://primer3.ut.ee/>) was used for primer design. Because a successful PCR assay requires  
108 efficient and specific amplification, the primers were assessed for several properties, including  
109 melting temperature, secondary structure, and complementarity. Primers required a GC content  
110 of 50%-60%, a melting temperature between 50 °C and 65 °C, a salt concentration of 50 mM,  
111 and an oligonucleotide concentration of 300 nM. The specificity of the primers and final  
112 sequences were verified using in silico prediction analyses with the online Basic Local  
113 Alignment Search Tool (BLAST) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). None of our  
114 designed primers showed genomic cross-reactivity with other viruses, the human genome, or

115 other probable interfering genomes in the BLAST database analysis [8]. Primers were  
116 synthesized in-house at the Oligonucleotide Synthesis Unit of the Center for Genomic Medicine  
117 at King Faisal Specialist Hospital and Research Centre.

### 118 **Sample Collection and Ethical Considerations**

119 This study was performed in compliance with all applicable national and international ethical  
120 guidelines for conducting research on human participants, including in accordance with the Code  
121 of Ethics of the World Medical Association (Declaration of Helsinki), and was approved by the  
122 institutional review board at King Faisal Specialist Hospital and Research Centre (KFSHRC)  
123 (IRB #220 0021). This board also granted a waiver for obtaining informed consent owing to the  
124 use of deidentified samples for this study.

125 In total, deidentified 282 nasopharyngeal swabs were collected from the CAP-accredited  
126 diagnostic laboratory in the Pathology and Laboratory Medicine Department at KFSHRC. These  
127 samples were confirmed positive for the presence of SARS-CoV-2 with Ct values ranging from  
128 14 to 40.

129

### 130 **Nucleic Acid Extraction and cDNA Synthesis**

131 In total, 282 samples obtained by nasopharyngeal swab and collected in universal viral transport  
132 medium were used in the nucleic acid extraction with a MagMAX Viral/Pathogen Nucleic Acid  
133 Isolation Kit (catalog No. A42352, Thermo Fisher Scientific) using a KingFisher Flex system  
134 (catalog No. 5400610, Thermo Fisher Scientific). Because cDNA synthesis is required to prepare  
135 patient samples for subsequent PCR, SuperScript IV VILO Master Mix (catalog No. 11766050,  
136 Thermo Fisher Scientific) was used. Briefly, a mixture of 8  $\mu$ L of RNA extract (approximately

137 50 ng/ $\mu$ L), 4  $\mu$ L of Master Mix, and 8  $\mu$ L of nuclease-free water was incubated at 25 °C for 10  
138 min. The RNA was reverse transcribed by incubating the mixture at 50 °C for 10 min, and the  
139 reaction was terminated at 85 °C for 5 min, and then place on ice.

140 The quality and the quantity of the RNA and cDNA samples were determined using a NanoDrop  
141 spectrophotometer. The ratio of sample absorbance at wavelengths of 260 and 280 nm was  
142 obtained to assess the nucleic acid purity (approximately 2.0 for RNA and approximately 1.8 for  
143 cDNA).

144

#### 145 **PCR and Sanger Sequencing**

146 The primer sets used are given in **Table 1**. For a total reaction volume of 25  $\mu$ L, 12.5  $\mu$ L of 2 $\times$   
147 GoTaq Master Mix (Promega; Madison, WI, USA), 2  $\mu$ L of cDNA, 0.25  $\mu$ L of sense primer (20  
148  $\mu$ M), 0.25  $\mu$ L of anti-sense primer (20  $\mu$ M), and 10  $\mu$ L of nuclease-free water were used. The  
149 PCR cycle was conducted using a Veriti 96-Well Fast Thermal Cycler (catalog No. 4375305,  
150 Thermo Fisher Scientific) as follows: 95 °C for 5 min; then 45 cycles of 95 °C for 30 s, 58° C for  
151 30 s, and 72 °C for 30 s; followed by 72 °C for 5 min. The positive control was a sample  
152 containing SARS-CoV-2, and the negative control was water.

153 The PCR products were separated using 1.5% agarose gel electrophoresis and visualized using  
154 ultraviolet light (Gel Doc EZ System; Bio-Rad). The retained PCR products were purified using  
155 AMPure XP and then sent for Sanger sequencing to the Core Facility of the Center for Genomic  
156 Medicine at KFSHRC. For Sanger sequencing, sequence chromatograms of 282 spike gene  
157 sequences were aligned using the Lasergene suite for sequence analysis (DNASTAR), with the  
158 WH-human1 SARS-CoV-2 sequence (MN908947) used as a reference [9].

159

## 160 SARS-CoV-2 Whole Genome Sequencing

161 Portions of the same 282 nasopharyngeal samples used for nucleic acid extraction and cDNA  
162 synthesis were also subjected to nucleic acid extraction using a MagMAX™ Viral/Pathogen  
163 Nucleic Acid Isolation Kit (Cat No. A42352, Thermo Fisher Scientific; MA, USA) for whole  
164 genome sequencing using the Ion AmpliSeq SARS-CoV-2 Research Panel with the Ion Torrent  
165 S5 platform by following the manufacturer's instructions (Thermo Fisher Scientific). For SARS-  
166 CoV-2 whole genome sequencing, variants were detected using the S5 Torrent Suite™ software  
167 Variant Caller (TVC), version 5.12.

168 **Table 1.** PCR primer sets for detection of SARS-CoV-2 variants of concern

Primer Name	Position	Forward Sequence 5'–3'	Position	Reverse Sequence 5'–3'
G1_KFSHRC	21645- 21664	ACACTAATTCTTTCACACGT	22101- 22082	TCAAGGTCCATAAGAAAAGG
G2(A)_KFSHRC	22733- 22754	TGCTTTACTAATGTCTATGCAG	23226- 23207	GACTCAGTAAGAACACCTGT
G2(B)_KFSHRC	22882- 22903	TCTTGATTCTAAGGTTGGTGGT	23519- 23500	CCCCTATTAAACAGCCTGCA
G2(C)_KFSHRC	23296- 23315	TCCACAGACACTTGAGATTC	23873- 23854	CTATTCCAGTTAAAGCACGG
G3_KFSHRC	24368- 24387	GACTCACTTCTCCACAGC	24993- 24974	TCAGGTTGCAAAGGATCATA

169

## 170 **Results**

171 Our primers and protocol were validated by first determining that the bands with the  
172 corresponding amplicon size were visualized for positive control samples, whereas for negative  
173 samples, the band was absent. Of samples from 282 patients, 282 samples showed the correct  
174 band of the corresponding size, however four samples with high Ct values ( $\geq 38$ ) showed faint  
175 bands. These results indicated that our primers were specific and provided positive results with  
176 samples that were previously confirmed being positive. Our results were 100% concordant with  
177 those of the diagnostic laboratory at KFSHRC, which used the DiaPlexQ<sup>TM</sup> Novel Coronavirus  
178 (2019-nCoV) Detection Kit. The PCR product of 282 samples, including the samples with faint  
179 bands, were sent to the sequencing core facility to perform Sanger sequencing.

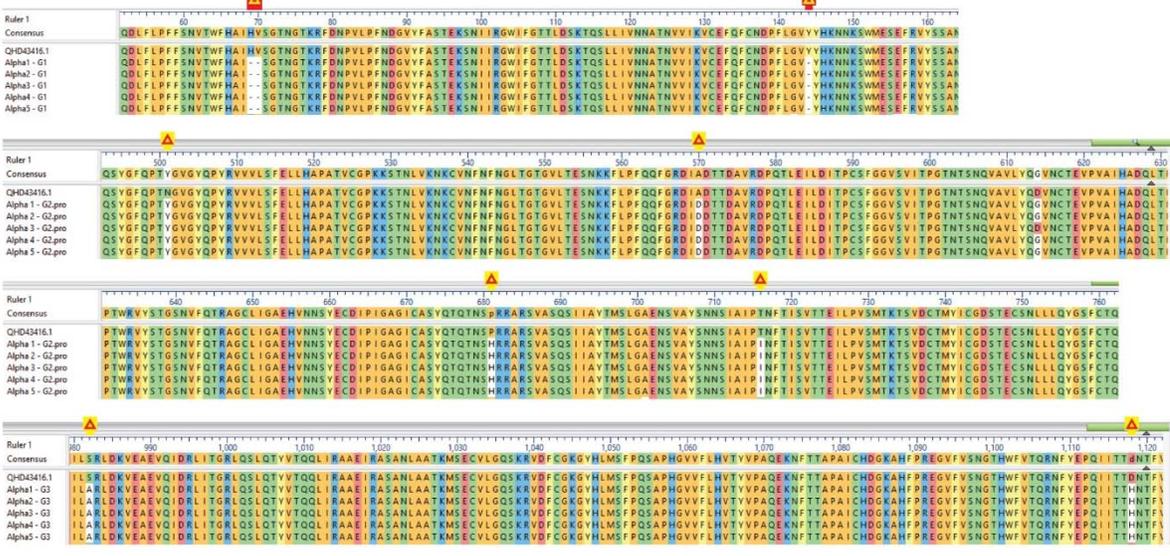
180

181 The output sequences of 282 patient samples were analyzed using DNASTar Lasergene Suite.  
182 Multiple sequence alignment of the spike gene was conducted with the reference genome  
183 (GenBank Protein Accession: QHD43416.1). The sequencing results were as follows: 123  
184 samples contained the alpha variant, 78 samples contained the beta variant, no sample contained  
185 the gamma variant, 13 samples contained the delta (B.1.617.2), 64 samples were non-VOCs that  
186 belonged to none of these variants and 4 samples had poor sequence. The major mutations  
187 detected were as follows: 69-70Del, 144Del, N501Y, A570D, P681H, T716I, S982A and  
188 D1118H for the alpha variant; D80A, K417N, E484K, N501Y and A701V for the beta variant;  
189 and G142D, 156-157Del, R158G, L452R, T478K, P681R, and D950N for the delta variant  
190 (Figure 1). Moreover, the remaining samples were non-VOCs that belonged to none of these

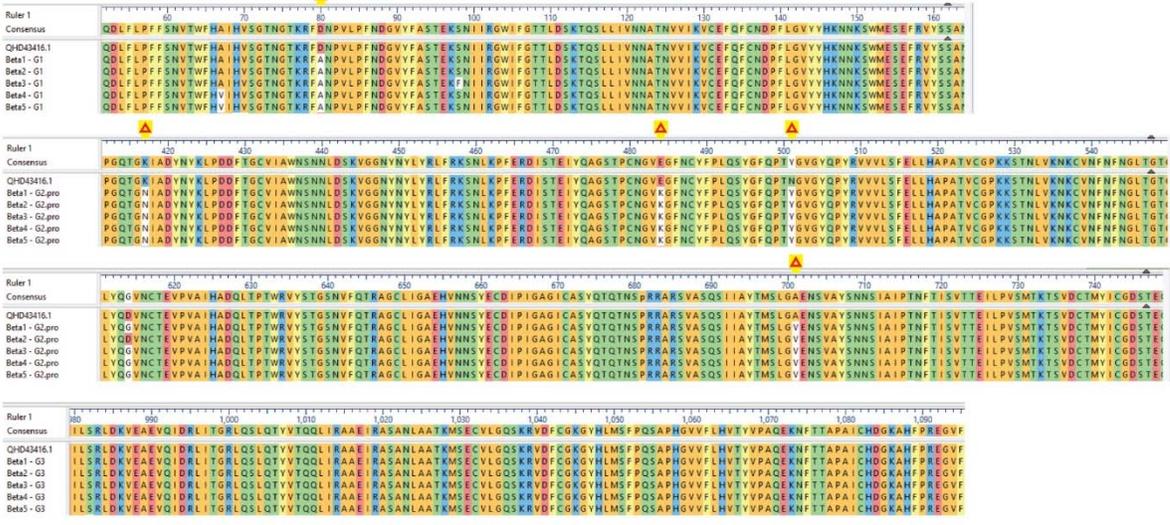
191 variants and had 99.97% identity with the reference genome. These mutations are not currently  
192 considered variants of concern nor variants under investigation. Of note, the D614G mutation  
193 was the most predominant in this cohort. Using the generated primers and developed protocol,  
194 we were able to sequence >98% of the positive samples with accurate mutation calls that  
195 matched the NGS data.

196

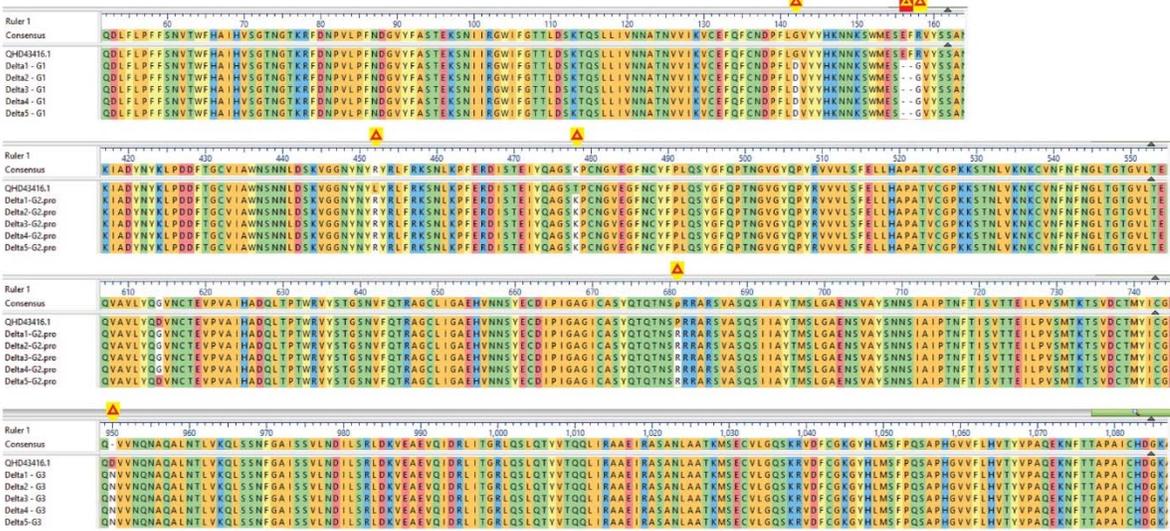
**A**



**B**



**C**



198 **Figure 1.** Aligned amino acid residues of the spike protein of SARS-CoV-2 from representative  
199 patient samples against the reference genome. Yellow triangles indicate deletions while red  
200 triangles indicate major mutations at different positions. **A:** represents alpha variant with major  
201 mutations such as 69-70Del, 144Del, N501Y, A570D, P681H, T716I, S982A, D1118H. **B:**  
202 represents beta variant with major mutations such as D80A, K417N, E484K, N501Y, A701V. **C:**  
203 represents delta with major mutations such as G142D, 156 -157 Del, R158G, L452R, T478K,  
204 P681R and D950N. Reference sequence: GenBank Protein Accession: QHD43416.1. [SARS-  
205 CoV-2 isolate Wuhan-Hu-1, complete genome]

206

207 To test the accuracy of and to validate the PCR and Sanger sequencing results, parallel tests to  
208 detect SARS-CoV-2 were conducted with whole genome sequencing using the Ion AmpliSeq  
209 SARS-CoV-2 Research Panel with the Ion Torrent S5 platform. For NGS data analysis, S5  
210 Torrent Server VM was used to check the quality parameters, summary, and unaligned reads  
211 with high ISP density. Our samples showed over 70% IPS density. To check the coverage  
212 analysis, mapped reads were around 500,000 reads which indicates over 3000 coverage (Mean  
213 Depth). Our 282 samples reported  $\geq 96\%$  On Target Scores. Consistent with the results of the in-  
214 house primers and Sanger sequencing assay, whole genome sequencing indicated that of 278  
215 samples, 123 samples contained the alpha variant, 78 samples contained the beta variant, no  
216 samples contained the gamma variant, and 13 samples contained the delta. For the four samples  
217 that failed to be sequenced by the Sanger approach, one sample contained the alpha variant, and  
218 three samples contained a few mutations that were not of concern.

219

## 220 **Discussion**

221 Given the global decreased availability of commercial SARS-CoV-2 sequencing kits, we  
222 designed, synthesized, and validated SARS-CoV-2 spike gene primer sets to identify current  
223 variants of concern in an efficient, cost-effective, and timely manner. We specifically designed  
224 and validated these primer sets to make them available for use in other laboratories worldwide.

225 This is a low-cost, easily performed protocol, and the reagents, laboratory supplies, and  
226 equipment are currently available as well as already found in most pathology or diagnostic  
227 laboratories.

228

229 The COVID-19 pandemic has underscored the importance of genomic surveillance as a valuable  
230 tool that quickly and effectively informs governments regarding public health decisions.

231 However, these surveillance efforts are scattered globally owing to limited supplies and lack of  
232 sequencing infrastructure and trained personnel, especially in low- and middle-income countries  
233 [10]. Indeed, a lack of accurate and fast SARS-CoV-2 genomic results may prolong the  
234 pandemic. According to the U. S. Centers for Disease Control and Prevention, genomic  
235 sequencing is an important approach that allows scientists to identify SARS-CoV-2 and its  
236 variants and monitor how they change over time into new variants, to understand how these  
237 changes affect the characteristics of the virus, and to use this information to better understand  
238 how it might impact public health [11]. According to GISAID, most countries are falling short  
239 on the global repository of SARS-CoV-2 genomic data. Of 180 million confirmed COVID-10  
240 cases, approximately 2 million have been submitted [9]. Thus, our sequencing protocol will help  
241 overcome the challenges of limited supplies and provide in-house sequencing primers and  
242 protocol that can be performed in many laboratories.

243 Our whole genome sequencing results showed high concordance with the PCR/Sanger  
244 sequencing results. Both platforms were successful in detecting the alpha, beta, and delta variants  
245 of concern. Recent studies have explored the possibility of using in-house primers and Sanger  
246 sequencing to identify SARS-CoV-2 variants. Testing with PCR has been considered the gold  
247 standard for viral gene detection, and it is widely used because of the ease in assay design and its

248 relatively low setup and running costs. During the pandemic, several other laboratories have  
249 leveraged PCR technology for detection or sequencing to diagnose COVID-19, with infected  
250 individuals identified by the successful amplification of the viral genome obtained using  
251 nasopharyngeal swabs. Some facilities, especially those in resource-constrained settings, may  
252 lack NGS equipment, but most facilities have Sanger technology. Indeed, Shaibu et al. in Nigeria  
253 sequenced the whole SARS-CoV-2 genome using Sanger sequencing technology [12]. When  
254 researchers in China mapped the genome of the coronavirus isolated from one of the first  
255 patients in the Wuhan outbreak in December 2019 and made it available to scientists worldwide,  
256 they enabled the global development of diagnostic tests, vaccines, and drugs. Using this  
257 information, we previously developed in-house detection assay that was approved for emergency  
258 use by the Saudi Food and Drug Authority [8].

259 In addition, when Public Health England announced in December 2020 that their national testing  
260 laboratories using a TaqPath detection assay were showing marked increases in spike gene target  
261 failures as of late November 2020 [13], we went back to analyze our previous TaqPath assay  
262 results that were conducted from June to August 2020 at KFSHRC using the same TaqPath  
263 assay. Out of 20,000 tests, we never found negative spike gene results (data not shown). This  
264 finding provides supporting evidence for the absence of Alpha variant in our community during  
265 that time. Since then, we established our epidemiology and genomic surveillance system in our  
266 hospital to monitor the emerging SARS-CoV-2 variants.

267

268 A major strength of this study was that by focusing on detecting only the spike gene, the  
269 turnaround time was faster than that for a whole genome sequencing platform, with less than 24  
270 hours needed to provide the results for 94 samples. However, major limitations of using

271 conventional PCR and Sanger sequencing are that this approach is not high throughput and only  
272 sequence short reads ( $\leq 1000$  bp) compared with NGS.

273

## 274 **Conclusions**

275 In conclusion, the ongoing COVID-19 pandemic poses one of the greatest global threats in  
276 modern history and has triggered severe human, social, and economic costs. Efficient, rapid, and  
277 cost-effective sequencing protocols will aid in genomic surveillance. Thus, we successfully  
278 developed such as assay for detecting SARS-CoV-2 variants of concerns that can be used by  
279 laboratories that cannot afford or cannot obtain commercial high throughput sequencing kits. We  
280 found  $\geq 98\%$  identity between the full genome sequence generated by NGS and by our developed  
281 Sanger sequencing platform. This in-house assay offers a viable alternative approach to increase  
282 sequencing capacity. KFSHRC will provide these primers free of charge to research laboratories  
283 in Saudi Arabia. Moreover, the developed method may be beneficial beyond the COVID-19  
284 pandemic because it can be easily adapted for use with any emerging microbe.

285

## 286 **Declarations**

287 Authors have nothing to declare.

## 288 **Ethics approval and consent to participate**

289 This study was approved by Research Advisory Council (RAC) of King Faisal Specialist  
290 Hospital and Research Centre (RAC No. 2200021).

## 291 **Consent for publication**

292 Not applicable.

### 293 **Availability of data and materials**

294 Raw data generated in this study and primers are available on request.

### 295 **Competing interests**

296 The authors declare that they have no competing interests.

### 297 **Funding**

298 This study was funded by the KFSHRC COVID-19 grant fund (RAC No. 2200021). The funder  
299 had no role in the design of the study and collection, analysis, and interpretation of data and in  
300 writing the manuscript.

### 301 **Authors' contributions**

302 F-Alhamlan supervised the project and wrote the manuscript. DB led the oligonucleotide  
303 synthesis work, MB, M-Alsanea, M-Asiri ran the in-house protocol and Sanger sequencing. BA,  
304 F-Alhadeq, F-Alsuwairi, and M-Al-Abdulkareem conducted the whole genome sequencing and  
305 data analysis. RA, S-Altamimi, MM and S-Althawadi provided the positive patient samples, ran  
306 the same samples at the diagnostic laboratory, and validated our assay. AA is the principal  
307 investigator and developed the study design. All authors have read and approved the final draft  
308 of the article and have approved its submission for publication.

309

### 310 **Acknowledgements**

311 We would like to extend our gratitude to the Sequencing Core Facility, Center for Genomic  
312 Medicine, King Faisal Specialist Hospital and Research Center. The support of the Research  
313 Center administration at King Faisal Specialist Hospital & Research Center is highly  
314 appreciated. We thank the Integrated Gulf BioSystems LLC team led by Shiva, Zaem,  
315 Balavenkatesh Mani and Udayaraja GK who helped with the NGS experiments.

316

## 317 **References**

- 318 1. WHO, *Coronavirus Disease (COVID-19) Pandemic*. 2020.
- 319 2. WHO, *Weekly epidemiological update on COVID-19 - 13 July 2021*. 2021.
- 320 3. Chan, J.F., et al., *Genomic characterization of the 2019 novel human-pathogenic*  
321 *coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan*. *Emerg*  
322 *Microbes Infect*, 2020. **9**(1): p. 221-236.
- 323 4. Huang, Y., et al., *Structural and functional properties of SARS-CoV-2 spike protein:*  
324 *potential antiviral drug development for COVID-19*. *Acta Pharmacol Sin*, 2020. **41**(9): p.  
325 1141-1149.
- 326 5. Heald-Sargent, T. and T. Gallagher, *Ready, set, fuse! The coronavirus spike protein and*  
327 *acquisition of fusion competence*. *Viruses*, 2012. **4**(4): p. 557-80.
- 328 6. Qiang, X.L., et al., *Using the spike protein feature to predict infection risk and monitor*  
329 *the evolutionary dynamic of coronavirus*. *Infect Dis Poverty*, 2020. **9**(1): p. 33.
- 330 7. Slatko, B.E., A.F. Gardner, and F.M. Ausubel, *Overview of Next-Generation Sequencing*  
331 *Technologies*. *Curr Protoc Mol Biol*, 2018. **122**(1): p. e59.

- 332 8. Alhamlan FA, A.-Q.A., Bakheet DM, Bohol MF, Althawadi SI, Mutabagani MS,  
333 Almaghrabi RS, Obeid D, *Development and Validation of an In-house, Low-Cost SARS-*  
334 *CoV-2 Detection Assay*. Journal of Infection and Public Health, 2021.
- 335 9. GISAID, *2,385,336 sequence entries with complete collection date information shared*  
336 *via GISAID by July 26 2021*. 2021.
- 337 10. Cyranoski, D., *Alarming COVID variants show vital role of genomic surveillance*.  
338 Nature, 2021. **ISSN 1476-4687 (online)**
- 339 11. CDC, *What is Genomic Surveillance?* 2021.
- 340 12. Shaibu, J.O., et al., *Full length genomic sanger sequencing and phylogenetic analysis of*  
341 *Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Nigeria*. PLoS One,  
342 2021. **16(1)**: p. e0243271.
- 343 13. PHE, *Investigation of novel SARS-COV-2 variant. Variant of Concern 202012/01*. 2020.