

1 *Higher Relative Viral Load Excretion Determined by Normalised Threshold Crossing Value in Acute*
2 *Cases infected with the B.1.1.7 Lineage VOC 202012/01 (Using S gene target failure as a Proxy) When*
3 *Compared to other Circulating Lineages in Wales*

4 *Anastasia Couzens, Isa Murrell, Laura Gifford, Ben Johns, Kathleen Pheasant, Luke Turner, Jonathan*
5 *Evans and Catherine Moore**

6 *Wales Specialist Virology centre, Public Health Wales Microbiology Cardiff, University Hospital of*
7 *Wales, Heath Park, Cardiff, CF14 4XW*

8 **corresponding author*

9

10 Abstract

11 Since the emergence of SARS-CoV-2, global monitoring of the virus using whole genome sequencing
12 has identified mutations occurring across the viral genome. Whilst the majority have little impact on
13 the virus, they are used effectively to monitor the movement of the virus globally and to inform
14 locally on transmission chains.

15 In late 2020, a variant of SARS-CoV-2 (B.1.1.7 - VOC 202012/01) was identified in the UK with a
16 distinct constellation of mutations, including in the spike gene that increased transmissibility. A
17 deletion in spike also affected one of the screening qPCR tests being used in the UK outside of
18 Wales, causing a failure to detect the target. This quickly became a surrogate marker for the variant
19 to allow rapid monitoring of the virus as it seeded into new regions of the UK.

20 A screening study using this assay as a proxy marker, was undertaken to understand the prevalence
21 of the variant in Wales. Secondary analysis of a screening qPCR that didn't target the S gene and also
22 included an endogenous control, was also performed to understand viral load excretion in those
23 infected with the variant compared to other circulating lineages. Using a combination of analytical
24 methods based on the C_t values of two gene targets normalised against the endogenous control,

25 there was a difference in the excreted viral load. Those with the variant excreting more virus than
26 those not infected with the variant. Supporting not only increased infectivity but offering a plausible
27 reason why increased transmission was associated with this particular variant.

28 Whilst there are limitations in this study, the method using C_t as a proxy for viral load can be used at
29 the population level to determine differences in viral excretion kinetics associated with different
30 variants.

31 Introduction

32 Severe acute respiratory syndrome coronavirus (SARS-CoV-2) was first detected within the United
33 Kingdom (UK) in January 2020, with the first confirmed case identified in Wales at the end of
34 February 2020 (Welsh Government, 2020; Lillie *et al.*, 2020). Variants of viruses commonly arise due
35 to mutation in the viral genome, and are mostly identified via whole-genome sequencing (Burki,
36 2021). Mutations and variants are often of little consequence with little effect on the virus or
37 naturally burn-out over time, however can become concerning if the change causes a competitive
38 advantage over the ancestral phenotype (Yurkovetskiy *et al.*, 2020). A variant of lineage B.1.1.7 (1)
39 was first identified in the South East of England, and was designated Variant of Concern (VOC)
40 202012/01 in December 2020 (PHE, 2020). Preliminary analysis suggested an approximate growth
41 rate 71% higher than other variants, and reported estimated increase of between 40% and 70% in
42 transmissibility (ECDC, 2020; NERVTAG, 2020).

43 VOC 202012/01 is characterised by an unprecedented number of mutations including an out of
44 frame six nucleotide deletion causing the removal of histidine and valine at positions 69 and 70
45 respectively (Δ H69/V70) from the S1 protein subunit (Rambaut *et al.*, 2020; Gloubchik *et al.*, 2021).
46 In total there are eight lineage-defining amino acid changes in Spike, in addition to pre-existing
47 D614G, which are of particular concern due to the antigenicity of the peplomer (Rambaut *et al.*,
48 2020; Yurkovetskiy *et al.*, 2020).

49 'Lighthouse laboratories' provide mass testing for 'Pillar 2' of the UK government's '5-pillar plan' for
50 the upscaling of SARS-CoV-2 testing, whereas the Public Health England and Wales (PHE/PHW) labs
51 and National Health Service (NHS) hospital laboratories cover 'Pillar 1' (UK Department of Health and
52 Social care, 2020). In 3 of these Lighthouse laboratories, a phenomenon of S-gene target failure
53 (SGTF) was noticed in Applied Biosystems™ TaqPath™ SARS-CoV-2 qualitative reverse transcriptase
54 polymerase chain reaction (RT-qPCR) assay (PHE, 2020). SGTF was found to indicate the presence of
55 VOC 202012/01 in samples due to Δ H69/V70, with further study showing that, in samples where
56 both sequence and SGTF status are known, 99.3% of samples with Δ H69/V70 sequences are SGTF;
57 furthermore, 98% of all Δ H69/V70 samples in England were identified as being VOC202012/01 by
58 the beginning of week 50/2020 (PHE, 2020). With this reasoning, the TaqPath™ assay was used as a
59 supplementary test in PHW's laboratory in University Hospital of Wales (UHW) on samples that had
60 tested positive via other RT-PCR assays from across Wales in order to provide a crude screen for VOC
61 202012/01 in Pillar 1 samples. An incidental observation of such was the ability to compare original
62 cycle threshold (C_t) values of samples with and without SGTF using both Livak and Schmittgen's $\Delta\Delta C_t$
63 comparative method (2001), and a relative quantification method in which the ratio of target gene C_t
64 compared to internal control (IC) C_t was calculated. Using SGTF as a crude indicator of the presence
65 of VOC 202012/01 and representing a unique insight into Pillar 1 testing, this analysis sought to
66 determine if there is a statistically significant difference in the relative quantification of virus, and in
67 the relative viral gene expression in VOC 202012/01 vs non-VOC 202012/01 when normalised C_t
68 values obtained by independent means are used as a proxy in lieu of viral loads.

69 Methods

70 Between the 21st December 2020 and 26th January 2021, all dry oropharyngeal/nasopharyngeal
71 swabs from across Wales that returned a positive result. These samples were split into those with C_t
72 values of ≤ 27 for any gene target, and those with C_t values > 27 for all gene targets, as the Pathogen

73 Genomics Unit of UHW were to have priority access to samples with $C_t \leq 27$ for pre-existing whole-
74 genome sequencing efforts.

75 Samples were largely received as 96-deep-well extraction plates containing extracts with a mix of
76 both positive and negative results. Positive samples were manually picked based on plate maps
77 constructed using RT-qPCR results and stored at 5°C until further testing. Sample-specific NHS
78 barcodes and episode numbers were utilised for record keeping of processed samples, and to allow
79 for future clinical and epidemiological investigation using the Laboratory Information Management
80 System (LIMS) TrakCare[®] Lab by InterSystems™ (Cambridge, Massachusetts, United States) used in
81 PHW labs.

82 The TaqPath™ assay was performed on all sample extracts, according to kit instructions, and results
83 were split into 3 qualitative result categories independent of the C_t results obtained through the
84 original diagnostic testing method; ‘Negative’ (RNA not detected by TaqPath but originally reported
85 as positive), ‘SGTF’ and ‘Wild-type (WT)’. These completed results were forwarded to Public Health
86 Wales Communicable Disease Surveillance Centre for analysis to understand the wider epidemiology
87 of SGTF in Wales.

88 To analyse the data further locally, those samples that were tested using the PerkinElmer[®] (Waltham,
89 Massachusetts, United States) (PE) SARS-CoV-2 RT-qPCR assay, were selected due to the use of an
90 endogenous control to normalise C_t values against and because S was not a target, the kinetics of
91 the assay would likely be not affected by absence of amplification or detection in this region.

92 The original PCR C_t values for the PE Nucleocapsid (N) and Open-Reading Frame 1ab (ORF1/ab) gene
93 targets, plus the undisclosed human-originating endogenous internal control (IC), were then
94 recorded for each sample. Together with the corresponding TaqPath™ result, were then split into
95 populations based on the aforementioned categories. Samples with a complete target failure of
96 either N gene or ORF/1ab in the PE assay were then excluded from the final analysis. ΔC_t was then

97 calculated for each sample by subtracting the value received for IC from the value received from the
98 gene target.

$$\Delta C_t = [\textit{target gene } C_t] - [IC C_t]$$

99 This was completed for both the N gene, and then ORF1/ab. Mean and median for these ΔC_t values
100 were then calculated in order to show crude differences in averages.

101 Livak and Schmittgen's (2001) $\Delta\Delta C_t$ method was conducted for each target gene; with all calculations
102 taking place in Microsoft® (Redmond, Washington, United States) Excel® 2016 (V 16.0.5095.100). The
103 results of which then underwent log transformations for ease of statistical analysis, and subject to
104 one-way ANOVA and Tukey's multiple comparisons post-hoc test in Prism® (V 9.0.1) by GraphPad
105 Software (San Diego, California, United States). Data was also plotted for normality in order to satisfy
106 an assumption of ANOVA. In order to perform the $\Delta\Delta C_t$ method, the C_t values for N gene, ORF1/ab
107 and IC for the positive control (PC) reported for every PerkinElmer® plate tested at the UHW PHW
108 site between 21st December 2020 and 26th January 2021 inclusively were recorded retroactively.

$$2^{-\Delta\Delta C_t} = 2^{-[(\textit{target gene } C_t - IC C_t) - \textit{Avg.}(PC \textit{ target gene } C_t - PC IC C_t)]}$$

109 Normalisation with regards to the IC was also performed in which the ratio of target gene C_t to IC C_t
110 was determined. Normality was plotted, and all data was subject to one-way ANOVA and Tukey's
111 multiple comparisons post-hoc test.

112 Results

113 In total, 849 samples that tested positive in the PerkinElmer® assay were also tested using the
114 TaqPath™ assay, 65 samples were excluded from the final analyses due to target failure in the PE
115 assay. Of those analysed, 119 were 'Negative' according to the TaqPath™ assay, 275 were 'SGTF',
116 and 390 were 'WT'.

117 The mean ΔC_t for 'SGTF' samples were lower than 'WT' by 3.43 (3s.f.) data points for N gene, and
118 3.78 (3s.f.) for ORF1/ab; and lower than 'Negative' samples by 8.28 (3s.f.) and 10.4 (3s.f.) for N gene

119 and ORF1/ab, respectively. Similar differences are retained for the median average, with median ΔC_t
120 for 'SGTF' samples dwarfing those for 'WT' by 5.00 (3s.f.) for N gene, and 4.00 (3s.f.) for ORF1/b; and
121 those for 'Negative' by 10.0 (3s.f.) and 11.0 (3s.f.) for N gene and ORF1/ab, respectively (Table 1).

122 The distribution of data point values, with statistical significance between groups indicated. QQ
123 normal plots for C_t values compared by both the $\Delta\Delta C_t$ comparative method and relative
124 quantification by ratio proved that both sets of data satisfied the normality (figure 1). Furthermore,
125 all samples were independent with a continuous dependent variable so, assuming equal variance,
126 both datasets were deemed suitable for analysis by one-way ANOVA. Results presented in Fig.1A
127 indicated that there was a statistically significant difference in the log-fold difference relative to the
128 IC C_t between each paired gene target in the 'SGTF', 'WT' and 'Negative' groups ($F(3, 781) = 113$ 3s.f.,
129 $p < 0.0001$).

130 A Tukey post hoc test of multiple comparisons revealed that the relative gene expression in 'SGTF'
131 samples was higher than in 'WT' (1.03 ± 0.401 3s.f., $p < 0.0001$; 1.14 ± 0.401 3s.f., $p < 0.0001$) and
132 'Negative' (-2.49 ± 0.559 3s.f., $p < 0.0001$; -3.12 ± 0.559 3s.f., $p < 0.0001$) samples, for N gene and
133 ORF1/ab respectively. No statistical significance was detected between the log fold-difference
134 relative to IC C_t of gene targets within the same qualitative TaqPath™ group. Similar results were
135 observed when normalising to the IC using ratio data, shown in Fig.1B; whereby one-way ANOVA
136 revealed a statistically significant difference in the relative quantification between each paired gene
137 target in the 3 groups ($F(3, 781) = 98.8$ 3s.f., $p < 0.0001$). A Tukey test showed that the log-fold
138 difference was higher in 'SGTF' samples than in 'WT' (-1.03 ± 0.374 3s.f., $p < 0.0001$; -1.14 ± 0.374
139 3s.f., $p < 0.0001$) and 'Negative' (-2.49 ± 0.521 3s.f., $p < 0.0001$; -3.12 ± 0.521 3s.f., $p < 0.0001$)
140 samples, for N gene and ORF1/ab respectively. Furthermore statistical significance was detected in
141 the difference between the log-fold relative C_t values of N gene and ORF1/ab in the 'Negative' group
142 (0.718 ± 0.679 3s.f., $p = 0.0115$), and no statistically significant difference was detected between the
143 differences in relative quantification of the paired gene targets within the remaining 2 groups.

144 Discussion

145 These results indicate a trend in which viral target gene fold-difference relative to human
146 endogenous IC is higher in samples that result in SGTF than in those that do not. In this analysis,
147 rather than assuming and using the negative correlation between raw C_t values and viral load,
148 normalisation of the target gene to the internal control was taken into account, which builds upon C_t
149 analysis carried out in other studies (Calistri *et al.*, 2021; Walker *et al.*, 2021).

150 When using SGTF as an indicator of VOC 202012/01, and a significantly higher calculated log fold-
151 difference relative to the IC as a proxy for viral load in these samples, the results support other
152 studies that show patients infected with VOC 202012/01 excrete higher viral loads than patients
153 with earlier circulating lineages of SARS-CoV-2 (Kidd *et al.*, 2020; Golubchik *et al.*, 2021). Supporting
154 the suggested increased infectivity and transmissibility reported for this variant.

155 A limitation of this analyses is as this was performed retrospectively, the efficiency of the
156 amplification kinetics of the PE assay was not determined. Because of this, the original iteration of
157 the $\Delta\Delta C_t$ method in which efficiency is assumed to be 100%, was used.

158 Using a single assay and gene target failure to determine the presence or not of a particular variant
159 also has limitations, as despite genomics showing that in regions where the variant might dominate
160 98% or more of the target failure samples are positive for the VOC, the change that leads to the
161 failure may also be sporadic. For example, $\Delta H69/V70$ was reported in some circulating lineages
162 before the emergence of VOC 202012/01, In Wales however, the molecular epidemiology generated
163 by systematic whole genome sequencing in pillar 1 and latterly in pillar 2, largely support the data
164 generated by the proxy assay for the arrival and establishment of VOC 202012/01 across Wales.

165 This study does demonstrate that whilst C_t values generated in a qualitative PCR test can't be used to
166 firmly determine infectivity, they can be used to indicate changes in viral load excretion patterns

167 associated with emerging variants at a population level, when a common testing platform with an
168 endogenous control is used to normalise target gene against.

169 Acknowledgements

170 The laboratory staff across the Public Health Wales and NHS Wales laboratory network for providing
171 the extracts for testing. Tom Connor, Sally Corden and the Staff of the Pathogen Genomics Unit for
172 supporting the validation of the data generated by the SGTF screening study. Chris Williams and
173 Clare Sawyer and the team in the PHW variant cell for feedback on the manuscript and for
174 supporting the wider epidemiological data analysis of the SGTF screening study. Public Health Wales
175 IMT for providing the funding to support the SGTF screening study.

176

177

178

179

180

181

182

183

184

185

186

187

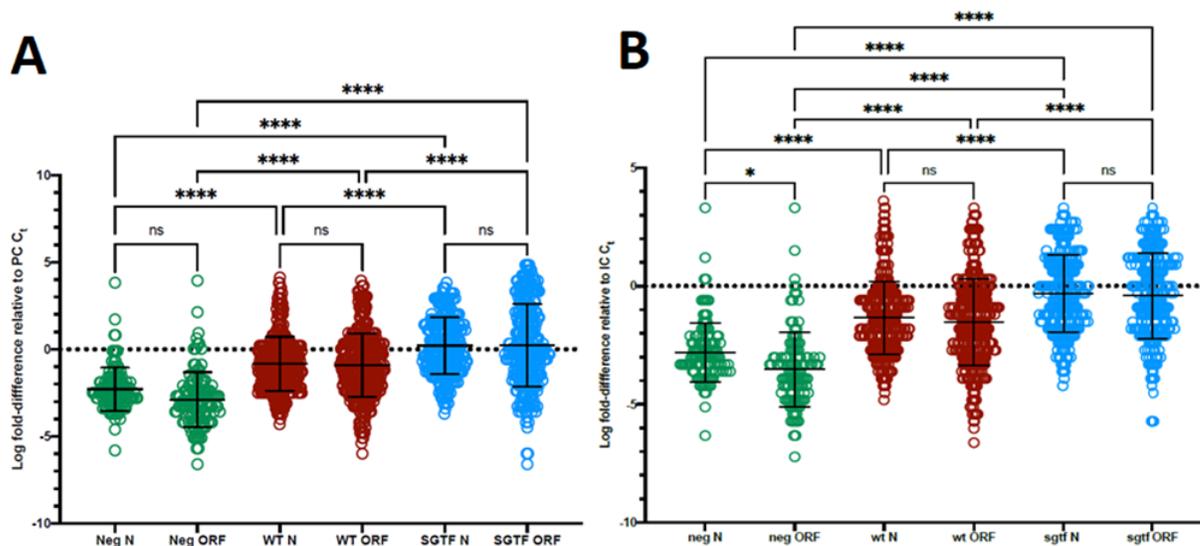
188

	Negative		SGTF		WT	
ΔC_t	N gene	ORF1/ab	N gene	ORF1/ab	N gene	ORF1/ab
Mean	9.31	10.7	1.03	0.331	4.46	4.11
Median	10.0	11.0	1.00	0.00	5.00	4.00

Table 1: shows the mean and median ΔC_t values calculated for both N gene and ORF1/ab, as determined by PerkinElmer® SARS-COV-2 RT-qPCR. ΔC_t results are split into 3 groups ('Negative', 'SGTF', 'WT') which are based on independent qualitative results from TaqPath™ SARS-COV-2 RT-qPCR.

189 Table 1

190 Figure 1



191

192 Figure 1: shows the distribution of data C_t analysed by Livak and Schmittgen's (2001) $\Delta\Delta C_t$ comparative method for relative
 193 gene expression (Fig. 1A) and by relative quantification by ratio (Fig. 1B) for both N gene and ORF/ab for samples grouped
 194 together as 'Negative' (green), 'WT' (red) and 'SGTF' (blue) – determined by TaqPath™ assay. Data analysed by one-way
 195 ANOVA and Tukey post-hoc test, the degree of significance of which are indicated by asterisks, whereby 4 asterisks (****)
 196 indicate a p value <0.0001, 1 (*) indicates a p value <0.05, and 'ns' denotes a lack of statistical significance.

197

198

199

200 *References*

201 Burki, T., 2021. Understanding variants of SARS-CoV-2. *The Lancet*, 397(10273), p.462.

202 Calistri, P., Amato, L., Puglia, I., Cito, F., Di Giuseppe, A., Danzetta, M., Morelli, D., Di Domenico, M.,

203 Caporale, M., Scialabba, S., Portanti, O., Curini, V., Perletta, F., Cammà, C., Ancora, M., Savini, G.,

204 Migliorati, G., D'Alterio, N. and Lorusso, A., 2021. Infection sustained by lineage B.1.1.7 of SARS-CoV-

205 2 is characterised by longer persistence and higher viral RNA loads in nasopharyngeal

206 swabs. *International Journal of Infectious Diseases*,.

207 Department of Health and Social Care, 2020. *Coronavirus (COVID-19) Scaling up our testing*

208 *programmes*. London: UK Government, pp.9-11.

209 European Centre for Disease Prevention and Control (ECDC), 2020. *Risk related to spread of new*

210 *SARSCoV-2 variants of concern in the EU/EEA*. 1st ed. [ebook] Solna, Sweden: European Union, p.11.

211 Available at: <[https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-risk-related-to-](https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-risk-related-to-spread-of-new-SARS-CoV-2-variants-EU-EEA.pdf)

212 [spread-of-new-SARS-CoV-2-variants-EU-EEA.pdf](https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-risk-related-to-spread-of-new-SARS-CoV-2-variants-EU-EEA.pdf)> [Accessed 18 March 2021].

213 Golubchik, Lythgoe, Hall, Ferretti, Fryer, MacIntyre-Cockett, de Cesare, Trebes, Piazza, Buck, Todd,

214 COVID-19 Genomics UK (COG-UK) consortium, Fraser and Bonsall, 2020. Early analysis of a potential

215 link between viral load and the N501Y mutation in the SARS-COV-2 spike protein. *Preprint*, [online]

216 Available at: <<https://www.medrxiv.org/content/10.1101/2021.01.12.20249080v1.full>> [Accessed

217 18 March 2021].

218 Kidd, M., Richter, A., Best, A., Cumley, N., Mirza, J., Percival, B., Mayhew, M., Megram, O., Ashford,

219 F., White, T., Moles-Garcia, E., Crawford, L., Bosworth, A., Atabani, S., Plant, T. and McNally, A., 2021.

220 S-variant SARS-CoV-2 lineage B1.1.7 is associated with significantly higher viral loads in samples

221 tested by ThermoFisher TaqPath RT-qPCR. *The Journal of Infectious Diseases*,.

222 Lillie, P., Samson, A., Li, A., Adams, K., Capstick, R., Barlow, G., Easom, N., Hamilton, E., Moss, P.,
223 Evans, A., Ivan, M., PHE Incident Team, Taha, Y., Duncan, C., Schmid, M. and
224 the Airborne HCID Network, 2020. Novel coronavirus disease (Covid-19): The first two patients in the
225 UK with person to person transmission. *Journal of Infection*, 80(5), pp.578-606.

226 Livak, K. and Schmittgen, T., 2001. Analysis of Relative Gene Expression Data Using Real-Time
227 Quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods*, 25(4), pp.402-408.

228 New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG), 2020. *NERVTAG meeting*
229 *on SARS-CoV-2 variant under investigation VUI-202012/01*. London: UK Government, p.2.

230 Public Health England (PHE), 2020. *Investigation of novel SARS-CoV-2 variant: Variant of Concern*
231 *202012/01 Technical briefing 2*. London: UK Government, pp.2, 3.

232 Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T.,
233 Robertson, D. and Volz, E., 2021. *Preliminary genomic characterisation of an emergent SARS-CoV-2*
234 *lineage in the UK defined by a novel set of spike mutations*. [online] Virological. Available at:
235 <[https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-](https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563)
236 [in-the-uk-defined-by-a-novel-set-of-spike-mutations/563](https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563)> [Accessed 18 March 2021].

237 Walker, A., Vihta, K., Gethings, O., Pritchard, E., Jones, J., House, T., Bell, I., Bell, J., Newton, J., Farrar,
238 J., Diamond, I., Studley, R., Rourke, E., Hay, J., Hopkins, S., Crook, D., Peto, T., Matthews, P., Eyre, D.,
239 Stoesser, N. and Pouwels, K., 2021. Increased infections, but not viral burden, with a new SARS-CoV-
240 2 variant. *Preprint*, [online] Available at:
241 <<https://www.medrxiv.org/content/10.1101/2021.01.13.21249721v1>> [Accessed 18 March 2021].

242 Welsh Government, 2020. *Wales confirms first positive case of coronavirus (COVID-19) |*
243 *GOV.WALES*. [online] GOV.WALES. Available at: <[https://gov.wales/wales-confirms-first-positive-](https://gov.wales/wales-confirms-first-positive-case-coronavirus-covid-19)
244 [case-coronavirus-covid-19](https://gov.wales/wales-confirms-first-positive-case-coronavirus-covid-19)> [Accessed 18 March 2021].

245 Yurkovetskiy, L., Wang, X., Pascal, K., Tomkins-Tinch, C., Nyalile, T., Wang, Y., Baum, A., Diehl, W.,
246 Dauphin, A., Carbone, C., Veinotte, K., Egri, S., Schaffner, S., Lemieux, J., Munro, J., Rafique, A.,
247 Barve, A., Sabeti, P., Kyratsous, C., Dudkina, N., Shen, K. and Luban, J., 2020. Structural and
248 Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell*, 183(3), pp.739-751.