Neutralising antibodies drive Spike mediated SARS-CoV-2 evasion

Kemp SA[1]*, Collier DA[1,2,3]*, Datir R[1,2,3], Gayed S[4], Jahun A[5], Hosmillo M[5], Ferreira IATM[2,3], Rees-Spear C[1], Mlcochova P[2,3], Ines Ushiro Lumb[6], David Roberts[6], Anita Chandra[2,3], Temperton N[7] , The CITIID-NIHR BioResource COVID-19 Collaboration[8], The COVID-19 Genomics UK (COG-UK) Consortium[9], Sharrocks K[4], Blane E[3], Briggs JAG[10], van Gils MJ[11], Smith KGC [2,3], Bradley JR[3,12], Smith C[14], Goldstein RA[1], Goodfellow IG[5], Smielewska A[5,13], Skittrall JP [4,14,15], Gouliouris T[4], Gkrania-Klotsas E[4], Illingworth CJR[14,16], McCoy LE[1], Gupta RK[2,3,16,17]

[1]Division of Infection and Immunity, University College London, London, UK.

[2] Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Cambridge, UK.

[3]Department of Medicine, University of Cambridge, Cambridge, UK.

[4]Department of Infectious Diseases, Cambridge University NHS Hospitals Foundation Trust, Cambridge, UK.

[5]Department of Pathology, University of Cambridge, Cambridge

[6]Public Health England, Colindale, London, UK

[7]Viral Pseudotype Unit, Medway School of Pharmacy, University of Kent, UK

[8]The CITIID-NIHR BioResource COVID-19 Collaboration, see appendix 1 for author list

[9]The COVID-19 Genomics UK (COG-UK) Consortium, Full list of consortium names and affiliations are in the appendix

[10]Medical Research Council Laboratory of Molecular Biology, Cambridge, UK.

[11]Department of Medical Microbiology, Academic Medical Center, University of Amsterdam, Amsterdam Institute for Infection and Immunity, Amsterdam, Netherlands

[12] NIHR Cambridge Clinical Research Facility, Cambridge, UK.

[13]Department of Virology, Cambridge University NHS Hospitals Foundation Trust

[14]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK

[15]Clinical Microbiology and Public Health Laboratory, Addenbrookes' Hospital, Cambridge, UK

[16] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

[17]Africa Health Research Institute, Durban, South Africa

*equal contribution

Address for correspondence:
Ravindra K. Gupta
Cambridge Institute for Therapeutic Immunology and Infectious Diseases
Jeffrey Cheah Biomedical Centre
Puddicombe Way
Cambridge CB2 0AW, UK
Tel: +44 1223 331491
rkg20@cam.ac.uk

Key words: SARS-CoV-2; COVID-19; antibody escape, Convalescent plasma; neutralising antibodies; mutation; evasion; resistance; immune suppression

Abstract

SARS-CoV-2 Spike protein is critical for virus infection via engagement of ACE2, and amino acid variation in Spike is increasingly appreciated. Given both vaccines and therapeutics are designed around Wuhan-1 Spike, this raises the theoretical possibility of virus escape, particularly in immunocompromised individuals where prolonged viral replication occurs. Here we report fatal SARS-CoV-2 escape from neutralising antibodies in an immune suppressed individual treated with convalescent plasma, generating whole genome ultradeep sequences by both short and long read technologies over 23 time points spanning 101 days. Little evolutionary change was observed in the viral population over the first 65 days despite two courses of remdesivir. However, following convalescent plasma we observed dynamic virus population shifts, with the emergence of a dominant viral strain bearing D796H in S2 and **Δ**H69/**Δ**V70 in the S1 NTD of the Spike protein. As serum neutralisation waned, viruses with the escape genotype diminished in frequency, before returning during a final, unsuccessful course of convalescent plasma. *In vitro*, the Spike escape variant conferred decreased sensitivity to multiple units of convalescent plasma/sera from different recovered patients, whilst maintaining infectivity similar to wild type. These data reveal strong positive selection on SARS-CoV-2 during convalescent plasma therapy and identify the combination of Spike mutations D796H and **Δ**H69/**Δ**V70 as a broad antibody resistance mechanism against commonly occurring antibody responses to SARS-CoV-2.

66 **Introduction**

67 SARS-CoV-2 is an RNA betacoronavirus, with closely related viruses identified in pangolins and
68 bats [1,2]. RNA viruses have inherently higher rates of mutation than DNA viruses such as
69 Herpesviridae [3]. The capacity for successful adaptation is exemplified by the Spike D614G
70 mutation, that arose in China and rapidly spread worldwide [4], now accounting for more than
71 90% of infections. The mutation appears to increase infectivity and transmissibility in animal
72 models [5]. Although the SARS-CoV-2 Spike protein is critical for virus infection via engagement
73 of ACE2, substantial Spike amino acid variation is being observed in circulating viruses [6].
74 Logically, mutations in the receptor binding domain (RBD) of Spike are of particular concern
75 due to the RBD being targeted by neutralising antibodies and therapeutic monoclonal
76 antibodies.

77

78 Deletions in the N-terminal domain (NTD) of Spike S1 are also being increasingly recognised,
79 both within hosts [7] and across individuals [8]. The evolutionary basis for the emergence of
80 deletions is unclear at present, and could be related to escape from immunity or to enhanced
81 fitness/transmission. The most notable deletion in terms of frequency is **Δ**H69/**Δ**V70. This
82 double deletion has been detected in multiple unrelated lineages, including the recent
83 'Cluster 5' mink related strain in the North Jutland region of Denmark
84 (https://files.ssi.dk/Mink-cluster-5-short-report_AFO2). There it was associated with the RBD
85 mutation Y453F in almost 200 individuals. Another European cluster in GISAID includes
86 **Δ**H69/**Δ**V70 along with the RBD mutation N439K.

87

88 Although **Δ**H69/**Δ**V70 has been detected multiple times, within-host emergence remains
89 undocumented and the reasons for its selection are unknown. Here we document real time
90 *SARS-CoV-2* emergence of **Δ**H69/**Δ**V70 in response to convalescent plasma therapy in an
91 immunocompromised human host treated with the B cell depletion agent rituximab,
92 demonstrating broad antibody escape in combination with the S2 mutation D796H.

93

94 **Results**

95 **Clinical case history of SARS-CoV-2 infection in setting of immune-compromised host**

96 These details are available on request from the authors.

97

98 **Negative SARS-CoV-2 specific antibodies and requirement for convalescent plasma (CP)**
99 We measured blood parameters including serum SARS-CoV-2 specific antibodies over time
100 (Supplementary table 1, Supplementary Figure 1, 2). Imaging was consistent with COVID-19
101 disease (Supplementary Figure 3). Total serum antibodies to SARS-CoV-2 were tested at days
102 44 and 50 by S protein immunoassay (Siemens). Results were negative. Three units (200mL
103 each) of convalescent plasma (CP) from three independent donors were obtained from the
104 NHS Blood and Transfusion Service and administered on compassionate use basis. These had
105 been assayed for antibody titres using the validated Euroimmun assay (Supplementary figure
106 3). Patient serum was subsequently positive for SARS-CoV-2 specific antibodies by S protein
107 immunoassay (Siemens) in the hospital diagnostic laboratory on days 68, 90 and 101.
108
109 **Virus genomic comparative analysis of 23 sequential respiratory samples over 101 days**
110 The majority of samples were respiratory samples from nose and throat or endotracheal
111 aspirates during the period of intubation. Ct values ranged from 16-34 and all 23 respiratory
112 samples were successfully sequenced by standard long read approach as per the ARTIC
113 protocol implemented by COG-UK; of these 20 additionally underwent short-read deep
114 sequencing using the Illumina platform. There was generally good agreement between the
115 methods, though as expected nanopore demonstrated greater error at low variant
116 frequencies (<10%) (Supplementary Figure 4). We detected no evidence of recombination,
117 based on two independent methods.
118
119 Maximum likelihood analysis of patient-derived whole genome consensus sequences
120 demonstrated clustering with other local sequences from the same region (Figure 1A). The
121 infecting strain was assigned to lineage 20B bearing the D614G Spike variant. Environmental
122 sampling showed evidence of virus on surfaces such as telephone and call bell but not in air
123 on days 59, 92 and 101. Sequencing of these surface viruses showed clustering with those
124 derived from the respiratory tract (Figure 1B). All samples were consistent with having arisen
125 from a single viral population. In our phylogenetic analysis, we included sequential sequences
126 from three other local patients identified with persistent viral RNA shedding over a period of
127 4 weeks or more (Supplementary Table 2). Viruses from these individuals showed very little
128 divergence in comparison to the case patient (Figure 1B) and none showed amino acid
129 changes in Spike over time. We additionally inferred a maximum likelihood phylogeny

130 comparing sequences from these three local individuals and two long term
131 immunosuppressed SARS-CoV-2 'shedders' recently reported[7,9], (Figure 1B). While the
132 sequences from Avanzato et al showed a pattern of evolution more similar to two of the three
133 other local patients, the case patient showed significant diversification with a mutation rate
134 of 30 per year (Supplementary table 2).
135
136 Further investigation of the sequence data suggested the existence of an underlying structure
137 to the viral population in our patient, with samples collected at days 93 and 95 being rooted
138 within, but significantly divergent from the original population (Figure 1B and 2A). The
139 relationship of the divergent samples to those at earlier time points rules out the possibility
140 of superinfection. The increased divergence of sequences does not necessarily indicate
141 selection; a spatially compartmentalised subset of viruses, smaller in number than the main
142 viral population, would be expected to evolve more quickly than the main population due to
143 the increased effect of genetic drift [10,11].
144
145 **Virus population structure changes following convalescent plasma and remdesivir**
146 All samples tested positive by RT-PCR and there was no sustained change in Ct values
147 throughout the 101 days following the first two courses of remdesivir (days 41 and 54), or the
148 first two units of convalescent plasma (days 63 and 65). According to nanopore data, no
149 polymorphisms occurred over the first 60 days at consensus level (Figure 2A). However, short
150 read deep sequence Illumina data revealed minority polymorphisms in the viral population
151 during this period (Figure 2B). For example, T39I in ORF7a reached a majority frequency of
152 77% on day 44, arising *de novo* and increased in frequency during the first period of the
153 infection (Supplementary Figure 6).
154
155 In contrast to the early period of infection, between days 66 and 82 a dramatic shift in virus
156 population structure was observed, with the near-fixation of D796H in S2 along with
157 **Δ**H69/**Δ**V70 in the S1 N-terminal domain (NTD) at day 82. This was identified in a nose and
158 throat swab sample with high viral load as indicated by Ct of 23 (Figure 3). The deletion was
159 not detected at any point prior to the day 82 sample, even as minority variants by short read
160 deep sequencing.
161

162 On Days 86 and 89, viruses collected were characterised by the Spike mutations Y200H and
163 T240I, with the deletion/mutation pair observed on day 82 having fallen to very low
164 frequency. Sequences collected on these days formed a distinct branch at the bottom of the
165 phylogeny in Figure 3, but were clearly associated with the remainder of the samples,
166 suggesting that they did not result from superinfection (Supplementary Figure 7), and further
167 were not significantly divergent from the bulk of the viral population (Supplementary Figure
168 5).
169
170 Sequencing of a nose and throat swab sample at day 93 again showed D796H along with
171 **Δ**H69/**Δ**V70 at <10% abundance, along with an increase in a virus population characterised
172 by Spike mutations P330S at the edge of the RBD and W64G in S1 NTD. This new lineage
173 reached near 100% abundance at day 93. Viruses with the P330S variant were detected in
174 two independent samples from different sampling sites, ruling out the possibility of
175 contamination. The divergence of these samples from the remainder of the population
176 (Figure 3B), noted above, suggests the possibility of their resulting from the stochastic
177 emergence, in the upper respiratory tract, of a previously unobserved subpopulation of
178 viruses (Supplementary Figure 5).
179
180 Following the third course of remdesivir (day 93) and third CP (day 95), we observed a re-
181 emergence of the D796H + **Δ**H69/**Δ**V70 viral population. The inferred linkage of D796H and
182 **Δ**H69/**Δ**V70 was maintained as evidenced by the highly similar frequencies of the two
183 variants, suggesting that the third unit of CP led to the re-emergence of this viral strain under
184 renewed positive selection. Ct values remained low throughout this period with
185 hyperinflammation, eventually leading to multi-organ failure and death at day 102. The
186 repeated increase in frequency of the novel viral strain during CP therapy strongly supports
187 the hypothesis that the deletion/mutation combination conferred antibody escape
188 properties.
189
190 **ΔH69/ΔV70 + D796H confers impaired neutralisation by multiple convalescent plasma units**
191 **and sera from recovered COVID-19 patients**
192 Using lentiviral pseudotyping we expressed wild type and **Δ**H69/**Δ**V70 + D796H mutant Spike
193 protein in enveloped virions and compared neutralisation activity of CP against these viruses.

194 This system has been shown to give similar results to replication competent virus[12,13]. We first
195 tested infection capacity over a single round of infection and found that **Δ**H69/**Δ**V70 + D796H
196 had similar infectivity to wild type (both in a D614G background, Figure 4A, B). The
197 **Δ**H69/**Δ**V70 + D796H mutant was partially resistant to the first two CP units (Figure 4C, Table
198 1A). In addition, patient derived serum from days 64 and 66 (one day either side of CP2
199 infusion) similarly showed lower potency against the mutant (Figure 4C, Table 1A). The
200 repeated observation of D796H + **Δ**H69/**Δ**V70 emergence and positive selection strengthens
201 the hypothesis that these variants were the key drivers of antibody escape. Experimentally,
202 the D796H + **Δ**H69/**Δ**V70 mutant also demonstrated reduced susceptibility to the CP3
203 administered on day 95, explaining its re-emergence (Figure 4C, Table 1A).
204
205 Given reduced susceptibility of the mutants to at least two units of CP, and the expansion of
206 sequences bearing **Δ**H69/**Δ**V70, we hypothesised that this represented a broad escape
207 mechanism. We therefore screened antiviral neutralisation activity in sera from five
208 recovered patients against the mutant and wild type viruses (Figure 4D). We observed that
209 the mutant was indeed significantly less susceptible to four of five randomly selected sera,
210 with the fifth showing reduced susceptibility that did not reach statistical significance (Table
211 1B). Fold change reductions in susceptibility of the mutant were as high as ten-fold compared
212 to wild type (Table 1B).
213
214 In order to probe the impact of the D796H and **Δ**H69/**Δ**V70 mutations on potency of
215 monoclonal antibodies (mAbs) targeting Spike, we screened neutralisation activity of a panel
216 of seven neutralizing mAbs across a range of epitope clusters[13] (Figure 4E). We observed no
217 differences in neutralisation between single mutants and wild type, suggesting that the
218 mechanism of escape was likely outside these epitopes.
219
220 In order to understand the mechanisms that might confer resistance to antibodies we
221 examined a published Spike structure and annotated it our residues of interest (Figure 5).
222 This analysis showed that **Δ**H69/**Δ**V70 is in a disordered, glycosylated loop at the very tip of
223 the NTD, and therefore could alter binding of antibodies. ΔH69/V70 is close to the binding
224 site of the polyclonal antibodies derived from COV57 plasma, indicating the sera tested here
225 may contain similar antibodies [14,15]. D796H is in an exposed loop in S2 (Figure 5), and appears

226 to be in a region frequently targeted by antibodies[16], despite mutations at position 796 being
227 rare (Supplementary table 4).
228
229 **Discussion**
230 Here we have documented a repeated evolutionary response by SARS-CoV-2 against antibody
231 therapy during the course of a persistent and eventually fatal infection in an
232 immunocompromised host. The observation of potential selection for specific variants
233 coinciding with the presence of antibodies from convalescent plasma is supported by the
234 experimental finding of reduced susceptibility of these viruses to plasma. Further, we were
235 able to document real-time emergence of a variant **Δ**H69/**Δ**V70 in the NTD of Spike that has
236 been increasing in frequency in Europe. In the case we report that it was not clear that the
237 emergence of the antibody escape variant was the primary reason for treatment failure.
238 However, given that both vaccines and therapeutics are aimed at Spike, our study raises the
239 possibility of virus evasion, particularly in immune suppressed individuals where prolonged
240 viral replication occurs.
241
242 Our observations represent a very rare insight, and only possible due to lack of antibody
243 responses in the individual following administration of the B cell depleting agent rituximab
244 for lymphoma, and an intensive sampling course undertaken due to concerns about
245 persistent shedding and risk of nosocomial transmission. Persistent viral replication and the
246 failure of antiviral therapy allowed us to define the viral response to convalescent plasma,
247 similar to a recent report on asymptomatic long term shedding with four sequences over 105
248 days[9], although the reported shifts in genetic composition of the viral population could not
249 be explained phenotypically. Another common finding is the very low neutralisation activity
250 in serum post transfusion of CP with waning as expected. Apart from the difference in the
251 outcome of infection (severe, fatal disease versus asymptomatic disease and clearance),
252 critically important differences in our study include: 1. The intensity of sampling and use of
253 both long and short read sequencing to verify variant calls, thereby providing a unique
254 scientific resource for longitudinal population genetic analysis. 2. The close alignment
255 between the genetic composition of the viral population and CP administration, with an
256 experimentally verified resistant strain emerging, falling to low frequency, and then rising
257 again under CP selection. 3. Real time detection of emergence of a variant, **Δ**H69/**Δ**V70, that

258 is increasing in frequency in Europe, and shown here to affect neutralization by multiple
259 COVID-19 patient derived sera.
260
261 We have noted in our analysis the potential influence of compartmentalised viral replication
262 upon the sequences recovered in upper respiratory tract samples. Both population genetic
263 and small animal studies have shown a lack of reassortment between influenza viruses within
264 a single host during an infection, suggesting that acute respiratory viral infection may be
265 characterised by spatially distinct viral populations[17,18]. In the analysis of data it is important
266 to distinguish genetic changes which occur in the primary viral population from apparent
267 changes that arise from the stochastic observation of spatially distinct subpopulations in the
268 host. While the samples we observe on days 93 and 95 of infection are genetically distinct
269 from the others, the remaining samples are consistent with arising from a consistent viral
270 population, supporting the finding of a reversion and subsequent regain of antibody
271 resistance. We note that in a study of SARS-CoV-2, Choi et al reported the detection in post
272 mortem tissue of viral RNA not only in lung tissue, but also in the spleen, liver and heart[7].
273 Mixing of virus from different compartments, for example via blood, or movement of
274 secretions from lower to upper respiratory tract, could lead to fluctuations in viral populations
275 at particular sampling sites. Experiments in animal models with sampling of different
276 replication sites could allow a better understanding of SARS-CoV-2 population genetics and
277 enable prediction of escape variants following antibody based therapies.
278
279 This is a single case report and therefore limited conclusions can be drawn about
280 generalisability.
281 In addition to documenting the emergence of SARS-CoV-2 Spike **Δ**H69/**Δ**V70 + D796H *in vivo*,
282 conferring broad reduction in susceptibility to serum/plasma polyclonal, but no effect on a
283 set of predominantly RBD-targeting monoclonal antibodies, these data highlight that
284 infection control measures need to be specifically tailored to the needs of
285 immunocompromised patients. The data also highlight caution in interpretation of CDC
286 guidelines that recommend 20 days as the upper limit of infection prevention precautions in
287 immune compromised patients who are afebrile[19]. Due to the difficulty with culturing clinical
288 isolates, use of surrogates for infectious virus such as sgRNA are warranted[20]. However, where
289 detection of ongoing viral evolution is possible, this serves as a clear proxy for the existence

of infectious virus. In our case we detected environmental contamination whilst in a single occupancy room and the patient was moved to a negative-pressure high air-change infectious disease isolation room.

The clinical efficacy of CP has been called into question recently[21], and our data suggest caution in use of CP in patients with immune suppression of both T cell and B cell arms. In such cases, the antibodies administered have little support from cytotoxic T cells, thereby reducing chances of clearance and raising the potential for escape mutations. Whilst we await further data, CP administered for clinical need in immune suppression, should ideally be undertaken in single occupancy rooms with enhanced infection control precautions, including SARS-CoV-2 environmental sampling and real-time sequencing.

321 **Table 1: Neutralisation activity of convalescent plasma and patient sera taken in between**
322 **CP doses against SARS-CoV-2 viral variants observed in A. patient case. B. five patients with**
323 **recovered COVID-19.** Shown are IC50 dilution titres - the CP or serum reciprocal dilution at
324 which 50% of virus infectivity is lost against wild type (baseline) or D796H + Δ69/70 mutant
325 Spike protein expressed on the surface of a lentiviral particle. Fold change is the difference
326 between the two viruses, and values below 1 indicate that the D796H/Δ69/70 mutant is less
327 sensitive to the CP or serum. Only the D66 serum sample failed to demonstrate significantly
328 higher resistance of the D796H/Δ69/70 mutant.
329

A.	IC50 as reciprocal dilution (95% CI)		
	WT	D796H + Δ69/70	Fold Change IC50 (95% CI)
CP1	2480.0 (2104.0 - 2965.0)	924.2 (634.1 - 1316.0)	0.37 (0.24 - 0.56)
D64 serum	163.7 (106.6 - 256.4)	55.6 (29.5 - 100.6)	0.27 (0.12 - 0.77)
CP2	447.1 (305.2 - 670.1)	196 (129.8 - 294.7)	0.47 (0.25 - 0.86)
D66 serum	175.7 (90.7 - 410.4)	85.5 (48.2 - 149.8)	0.82 (0.29 - 1.96)
CP3	280.2 (172.8 - 480.3)	115 (75.0 - 175.5)	0.47 (0.24 - 0.91)

330
331
332
333
334

B.	IC50 as reciprocal dilution (95% CI)		
	WT	D796H + Δ69/70	Fold Change IC50 (95% CI)
Serum 1	994.1 (852.7-1159.0)	755.8 (426.7-1269.0)	0.71 (0.39-1.33)
Serum 2	2107.0 (1680.0-2662.0)	1045.0 (745.9-1387.0)	0.42 (0.27-0.70)
Serum 3	3241.0 (2377.0-4643.0)	347.4 (234.4-505.4)	0.11 (0.07-0.17)
Serum 4	17368.0 (14357.0-22130.0)	2965.0 (2287.0-3767.0)	0.26 (0.20-0.32)
Serum 5	52735.0 (34276.0-93957.0)	3871.0 (3525.0-4252.0)	0.20 (0.17-0.24)

335
336
337
338
339
340
341
342

343 **Ethics**

344 The study was approved by the East of England – Cambridge Central Research Ethics
345 Committee (17/EE/0025). The patient and family provided written informed consent.
346 Additional controls with COVID-19 were enrolled to the NIHR BioResource Centre Cambridge
347 under ethics review board (17/EE/0025).

348

349 **Acknowledgements**

350 We are immensely grateful to the patient and family. We would also like to thank the staff at
351 CUH and the NIHR Cambridge Clinical Research Facility. We would like to thank Dr Ruthiran
352 Kugathasan and Professor Wendy Barclay for helpful discussions and Dr Martin Curran, Dr
353 William Hamilton and Dr. Dominic Sparkes. We would like to thank Prof Andres Floto and Prof
354 Ferdia Gallagher. We thank Dr James Voss for the kind gift of HeLa cells stably expressing
355 ACE2. COG-UK is supported by funding from the Medical Research Council (MRC) part of UK
356 Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome
357 Research Limited, operating as the Wellcome Sanger Institute. RKG is supported by a
358 Wellcome Trust Senior Fellowship in Clinical Science (WT108082AIA). LEM is supported by a
359 Medical Research Council Career Development Award (MR/R008698/1). SAK is supported by
360 the Bill and Melinda Gates Foundation via PANGEA grant: OPP1175094. DAC is supported by
361 a Wellcome Trust Clinical PhD Research Fellowship. CJRI acknowledges MRC funding (ref:
362 MC_UU_00002/11). This research was supported by the National Institute for Health
363 Research (NIHR) Cambridge Biomedical Research Centre, the Cambridge Clinical Trials Unit
364 (CCTU) and by the UCL Coronavirus Response Fund and made possible through generous
365 donations from UCL's supporters, alumni and friends (LEM). JAGB is supported by the Medical
366 Research Council (MC_UP_1201/16).

367

368 **Author contributions**

369 Conceived study: RKG, SAK, DAC, AS, TG, EGK

370 Designed experiments: RKG, SAK, DAC, LEM, JAGB, EGK, AC, NT, AC, CS

371 Performed experiments: SAK, DAC, LEM, RD, CRS, AJ, IATMF, KS, TG, CJRI, BB, JS, MJvG

372 Interpreted data: RKG, SAK, DAC, PM, LEM, JAGB, PM, SG, KS, TG, JB, KGCS, IG, CJRI, JAGB,
373 IUL, DR, JS, BB

374

375 **Methods**

376 *Clinical Sample Collection and Next generation sequencing*

377 Serial samples were collected from the patient periodically from the lower respiratory tract
378 (sputum or endotracheal aspirate), upper respiratory tract (throat and nasal swab), and from
379 stool. Nucleic acid extraction was done from 500µl of sample with a dilution of MS2
380 bacteriophage to act as an internal control, using the easyMAG platform (Biomerieux, Marcy
381 LEtoile) according to the manufacturers' instructions. All samples were tested for presence of
382 SARS-CoV-2 with a validated one-step RT q-PCR assay developed in conjunction with the
383 Public Health England Clinical Microbiology [22]. Amplification reaction were all performed on
384 a Rotorgene™ PCR instrument. Samples which generated a CT of ≤36 were considered to be
385 positive.

386

387 Sera from recovered patients in the COVIDx study[23] were used for testing of neutralisation
388 activity by SARS-CoV-2 mutants.

389

390 For viral genomic sequencing, total RNA was extracted from samples as described. Samples
391 were sequenced using MinION flow cells version 9.4.1 (Oxford Nanopore Technologies)
392 following the ARTICnetwork V3 protocol
393 (https://dx.doi.org/10.17504/protocols.io.bbmuik6w) and BAM files assembled using the
394 ARTICnetwork assembly pipeline (https://artic.network/ncov-2019/ncov2019-
395 bioinformatics-sop.html). A representative set of 10 sequences were selected and also
396 sequenced using the Illumina MiSeq platform. Amplicons were diluted to 2 ng/µl and 25 µl
397 (50 ng) were used as input for each library preparation reaction. The library preparation used
398 KAPA Hyper Prep kit (Roche) according to manufacturer's instructions. Briefly, amplicons
399 were end-repaired and had A-overhang added; these were then ligated with 15mM of NEXTflex
400 DNA Barcodes (Bio Scientific, Texas, USA). Post-ligation products were cleaned using AMPure
401 beads and eluted in 25 µl. Then, 20 µl were used for library amplification by 5 cycles of PCR.
402 For the negative controls, 1ng was used for ligation-based library preparation. All libraries
403 were assayed using TapeStation (Agilent Technologies, California, USA) to assess fragment
404 size and quantified by QPCR. All libraries were then pooled in equimolar accordingly. Libraries
405 were loaded at 15nM and spiked in 5% PhiX (Illumina, California, USA) and sequenced on one

406 MiSeq 500 cycle using a Miseq Nano v2 with 2x 250 paired-end sequencing. A minimum of
407 ten reads were required for a variant call.
408
409 *Bioinformatics Processes*
410 For long-read sequencing, genomes were assembled with reference-based assembly and a
411 curated bioinformatics pipeline with 20x minimum coverage across the whole-genome [24]. For
412 short-read sequencing, FASTQs were downloaded, poor-quality reads were identified and
413 removed, and both Illumina and PHiX adapters were removed using TrimGalore v0.6.6 [25].
414 Trimmed paired-end reads were mapped to the National Center for Biotechnology
415 Information SARS-CoV-2 reference sequence MN908947.3 using MiniMap2-2.17 with
416 arguments -ax and sr [26]. BAM files were then sorted and indexed with samtools v1.11 and PCR
417 optical duplicates removed using Picard (http://broadinstitute.github.io/picard). Single
418 nucleotide polymorphisms (SNPs) were called using Freebayes v1.3.2 [27] with a ploidy setting
419 of 1, minimum allele frequency of 0.20 and a minimum depth of five reads. Finally, a
420 consensus sequences of nucleic acids with a minimum whole-genome coverage of at least
421 20× were generated with BCFtools using a 0% majority threshold.
422
423 *Phylogenetic Analysis*
424 All available full-genome SARS-CoV-2 sequences were downloaded from the GISAID database
425 (http://gisaid.org/) [28] on 17th September. Duplicate and low-quality sequences (>5% N
426 regions) were removed, leaving a dataset of 138,472 sequences with a length of >29,000bp.
427 All sequences were sorted by name and only sequences sequenced with United Kingdom /
428 England identifiers were retained. From this dataset, a subset of 250 sequences were
429 randomly subsampled using seqtk (https://github.com/lh3/seqtk). These 250 sequences were
430 aligned to the 23 patient sequences, as well as three other control patients (persistent long-
431 term shedders from the same hospital) (Supplementary Table 2) and the SARS-CoV-2
432 reference strain MN908947.3, using MAFFT v7.473 with automatic flavour selection [29]. Major
433 SARS-CoV-2 clade memberships were assigned to all sequences using the Nextclade server
434 v0.8 (https://clades.nextstrain.org/).
435
436 Maximum likelihood phylogenetic trees were produced using the above curated dataset using
437 IQ-TREE v2.1.2 [30]. Evolutionary model selection for trees were inferred using ModelFinder [31]

438 and trees were estimated using the GTR+F+I model with 1000 ultrafast bootstrap replicates
439 [32]. All trees were visualised with Figtree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/),
440 rooted on the SARS-CoV-2 reference sequence and nodes arranged in descending order.
441 Nodes with bootstraps values of <50 were collapsed using an in-house script.
442
443 A time tree indicating temporal divergence was inferred using TreeTime v0.75 [33]. For this a
444 further subsample of 100 sequences from the multiple sequence alignment used to make the
445 maximum likelihood tree were selected, which represented sequences belonging to
446 NextStrain clade 20B. This included the 23 patient genomes, all genomes from two of the
447 three long-term shedders (one was excluded as it belonged to NextStrain clade 19A), 75
448 subsampled Cambridge sequences, and the SARS-CoV-2 reference strain MN908947.3. All
449 sequences were re-aligned using MAFFT v7.473 with automatic flavour selection [29] and used
450 as input.
451
452 Molecular substitution (clock) rates for the index patient, as well as three long-term shedders
453 and two recently described immunocompromised patients from literature, were estimated
454 using BEAST v2.6.3 [34] using a HKY substitution model with 4 rate categories drawn from a
455 gamma distribution, a strict clock and a coalescent exponential population tree prior. MCMC
456 was run for 100 million iterations excluding a 15% burn-in. Tracer v1.7.1 was used to analyse
457 the BEAST trace in order extract the clock rate and ensure convergence had occurred.
458
459 *In-depth allele frequency variant calling*
460 The SAMFIRE package[35] was used to call allele frequency trajectories from BAM file data.
461 Reads were included in this analysis if they had a median PHRED score of at least 30, trimming
462 the ends of reads to achieve this if necessary. Nucleotides were then filtered to have a PHRED
463 score of at least 30; reads with fewer than 30 such reads were discarded. Distances between
464 sequences, accounting for low-frequency variant information, was also conducted using
465 SAMFIRE. The sequence distance metric, described in an earlier paper [11], combines allele
466 frequencies across the whole genome. Where L is the length of the genome, we define q(t)
467 as a 4 x L element vector describing the frequencies of each of the nucleotides A, C, G, and T
468 at each locus in the viral genome sampled at time t. For any given locus i in the genome we

calculate the change in allele frequencies between the times t_1 and t_2 via a generalisation of the Hamming distance

$$d\big(q_i(t_1), q_i(t_2)\big) = \frac{1}{2} \sum_{a \in \{A,C,G,T\}} |q_i^a(t_1) - q_i^a(t_2)|$$

where the vertical lines indicate the absolute value of the difference. These statistics were then combined across the genome to generate the pairwise sequence distance metric

$$D\big(\boldsymbol{q}(t_1), \boldsymbol{q}(t_2)\big) = \sum_i d\big(q_i(t_1), q_i(t_2)\big)$$

The Mathematica software package was to conduct a regression analysis of pairwise sequence distances against time, leading to an estimate of a mean rate of within-host sequence evolution. In contrast to the phylogenetic analysis, this approach assumed the samples collected on days 93 and 95 to arise via stochastic emission from a spatially separated subpopulation within the host, leading to a lower inferred rate of viral evolution for the bulk of the viral population.

Inference of selection

Under the assumption of a large effective population size, a deterministic one-locus model of selection was fitted to genome sequence data describing changes in the frequency of the variant T39I. Where q(t) represents the frequency of a single variant at time t, we used a maximum likelihood method to infer the initial variant frequency at time t=1, and the selection coefficient s, for times from the initial time point to the disappearance of the variant from the population. Specifically, where $n(t_i)$ and $N(t_i)$ were the number of observations of the variant and the total read depth at that locus at time t_i, we fitted the model

$$q(t) = \frac{q(1)e^{st}}{1-q(1)+q(1)e^{st}}$$

so as to maximise the binomial likelihood

498

499 $$L = \sum_i \log \binom{N_i}{n_i} q(t_i)^{n_i} \left(1 - q(t_i)\right)^{N-n_i}$$

500 A similar calculation was performed to estimate the mean effective selection (incorporating
501 intrinsic selection for the variant plus linkage with other selected alleles) acting upon the
502 variant D796H during the final period of CP therapy. In this case selection was modelled as
503 being time-dependent, kicking in with the administration of therapy. We fitted the model:
504

505 $$q(t) = q \; if \; t < \tau$$
$$q(t) = \frac{qe^{s(t-\tau)}}{1 - q + qe^{s(t-\tau)}} if \; t \geq \tau$$

506

507 to the data, setting $\tau = 95$.
508

509 *Recombination Detection*
510 All sequences were tested for potential recombination, as this would impact on evolutionary
511 estimates. Potential recombination events were explored with nine algorithms (RDP, MaxChi,
512 SisScan, GeneConv, Bootscan, PhylPro, Chimera, LARD and 3SEQ), implemented in RDP5 with
513 default settings [36]. To corroborate any findings, ClonalFrameML v1.12 [37] was also used to infer
514 recombination breakpoints. Neither programs indicated evidence of recombination in our
515 data.
516

517 *Structural Viewing*
518 The Pymol Molecular Graphics System v2.4.0 (https://github.com/schrodinger/pymol-open-
519 source/releases) was used to map the location of the four spike mutations of interested onto
520 a SARS-CoV-2 spike structure visualised by Wrobel et al (PDB: 6ZGE) [38].
521

522 *Generation of Spike mutants*
523 Amino acid substitutions were introduced into the D614G pCDNA_SARS-CoV-2_Spike plasmid
524 as previously described[39] using the QuikChange Lightening Site-Directed Mutagenesis kit,
525 following the manufacturer's instructions (Agilent Technologies, Inc., Santa Clara, CA).
526

527 *Pseudotype virus preparation*

528 Viral vectors were prepared by transfection of 293T cells by using Fugene HD transfection
529 reagent (Promega). 293T cells were transfected with a mixture of 11ul of Fugene HD, 1μg of
530 pCDNAΔ19Spike, 1ug of p8.91 HIV-1 gag-pol expression vector[40,41], and 1.5μg of pCSFLW
531 (expressing the firefly luciferase reporter gene with the HIV-1 packaging signal). Viral
532 supernatant was collected at 48 and 72h after transfection, filtered through 0.45um filter
533 and stored at -80˚C. The 50% tissue culture infectious dose ($TCID_{50}$) of SARS-CoV-2
534 pseudovirus was determined using Steady-Glo Luciferase assay system (Promega).
535
536 RT activity in virion containing supernatant was measured as previously described.
537
538 *Serum/plasma pseudotype neutralisation assay*

539 Spike pseudotype assays have been shown to have similar characteristics as neutralisation
540 testing using fully infectious wild type SARS-CoV-2[12].Virus neutralisation assays were
541 performed on 293T cell transiently transfected with ACE2 and TMPRSS2 using SARS-CoV-2
542 Spike pseudotyped virus expressing luciferase[42]. Pseudotyped virus was incubated with
543 serial dilution of heat inactivated human serum samples or convalescent plasma in duplicate
544 for 1h at 37˚C. Virus and cell only controls were also included. Then, freshly trypsinized 293T
545 ACE2/TMPRSS2 expressing cells were added to each well. Following 48h incubation in a 5%
546 CO_2 environment at 37°C, the luminescence was measured using Steady-Glo Luciferase
547 assay system (Promega).
548
549 *mAb pseudotype neutralisation assay*

550 Virus neutralisation assays were performed on HeLa cells stably expressing ACE2 and using
551 SARS-CoV-2 Spike pseudotyped virus expressing luciferase as previously described[43].
552 Pseudoyped virus was incubated with serial dilution of purified mAbs[13] in duplicate for 1h at
553 37˚C. Then, freshly trypsinized HeLa ACE2- expressing cells were added to each well.
554 Following 48h incubation in a 5% CO_2 environment at 37°C, the luminescence was
555 measured using Bright-Glo Luciferase assay system (Promega) and neutralization calculated
556 relative to virus only controls. IC50 values were calculated in Graphpad Prism.
557

558 **References**

559 1 Zhang, T., Wu, Q. & Zhang, Z. Probable pangolin origin of SARS-CoV-2 associated with the
560 COVID-19 outbreak. *Current Biology* (2020).
561 2 Xiao, K. *et al.* Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature*
562 **583**, 286-289, doi:10.1038/s41586-020-2313-x (2020).
563 3 Sanjuán, R. & Domingo-Calap, P. Mechanisms of viral mutation. *Cell Mol Life Sci* **73**, 4433-
564 4448, doi:10.1007/s00018-016-2299-6 (2016).
565 4 Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases
566 Infectivity of the COVID-19 Virus. *Cell*, doi:10.1016/j.cell.2020.06.043 (2020).
567 5 Yurkovetskiy, L. *et al.* Structural and Functional Analysis of the D614G SARS-CoV-2 Spike
568 Protein Variant. *Cell* **183**, 739-751 e738, doi:10.1016/j.cell.2020.09.032 (2020).
569 6 Korber, B. *et al.* Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases
570 infectivity of the COVID-19 virus. *Cell* **182**, 812-827. e819 (2020).
571 7 Choi, B. *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N*
572 *Engl J Med*, doi:10.1056/NEJMc2031364 (2020).
573 8 McCarthy, K. R. *et al.* Natural deletions in the SARS-CoV-2 spike glycoprotein drive antibody
574 escape. *bioRxiv*, 2020.2011.2019.389916, doi:10.1101/2020.11.19.389916 (2020).
575 9 Avanzato, V. A. *et al.* Case Study: Prolonged infectious SARS-CoV-2 shedding from an
576 asymptomatic immunocompromised cancer patient. *Cell* (2020).
577 10 Wright, S. Size of population and breeding structure in relation to evolution. *Science*, 430-431
578 (1938).
579 11 Lumby, C. K., Zhao, L., Breuer, J. & Illingworth, C. J. A large effective population size for
580 established within-host influenza virus infection. *Elife* **9**, doi:10.7554/eLife.56915 (2020).
581 12 Schmidt, F. *et al.* Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and
582 chimeric viruses. 2020.2006.2008.140871, doi:10.1101/2020.06.08.140871 %J bioRxiv (2020).
583 13 Brouwer, P. J. M. *et al.* Potent neutralizing antibodies from COVID-19 patients define multiple
584 targets of vulnerability. *Science* **369**, 643-650, doi:10.1126/science.abc5902 (2020).
585 14 Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent
586 individuals. *Nature* **584**, 437-442, doi:10.1038/s41586-020-2456-9 (2020).
587 15 Barnes, C. O. *et al.* Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal
588 Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828-842 e816,
589 doi:10.1016/j.cell.2020.06.025 (2020).
590 16 Shrock, E. *et al.* Viral epitope profiling of COVID-19 patients reveals cross-reactivity and
591 correlates of severity. *Science*, doi:10.1126/science.abd4250 (2020).
592 17 Sobel Leonard, A. *et al.* The effective rate of influenza reassortment is limited during human
593 infection. *PLoS pathogens* **13**, e1006203, doi:10.1371/journal.ppat.1006203 (2017).
594 18 Richard, M., Herfst, S., Tao, H., Jacobs, N. T. & Lowen, A. C. Influenza A Virus Reassortment Is
595 Limited by Anatomical Compartmentalization following Coinfection via Distinct Routes.
596 *Journal of virology* **92**, doi:10.1128/JVI.02063-17 (2018).
597 19 CDC. *Discontinuation of Transmission-Based Precautions and Disposition of Patients with*
598 *COVID-19 in Healthcare Settings (Interim Guidance)*,
599 <https://www.cdc.gov/coronavirus/2019-ncov/hcp/disposition-hospitalized-patients.html>
600 (2020).
601 20 Boshier, F. A. T. *et al.* Remdesivir induced viral RNA and subgenomic RNA suppression, and
602 evolution of viral variants in SARS-CoV-2 infected patients. *medRxiv*,
603 2020.2011.2018.20230599, doi:10.1101/2020.11.18.20230599 (2020).
604 21 Simonovich, V. A. *et al.* A Randomized Trial of Convalescent Plasma in Covid-19 Severe
605 Pneumonia. *N Engl J Med*, doi:10.1056/NEJMoa2031304 (2020).
606 22 Meredith, L. W. *et al.* Rapid implementation of SARS-CoV-2 sequencing to investigate cases of
607 health-care associated COVID-19: a prospective genomic surveillance study. *The Lancet*
608 *Infectious Diseases* **20**, 1263-1272, doi:10.1016/S1473-3099(20)30562-4 (2020).

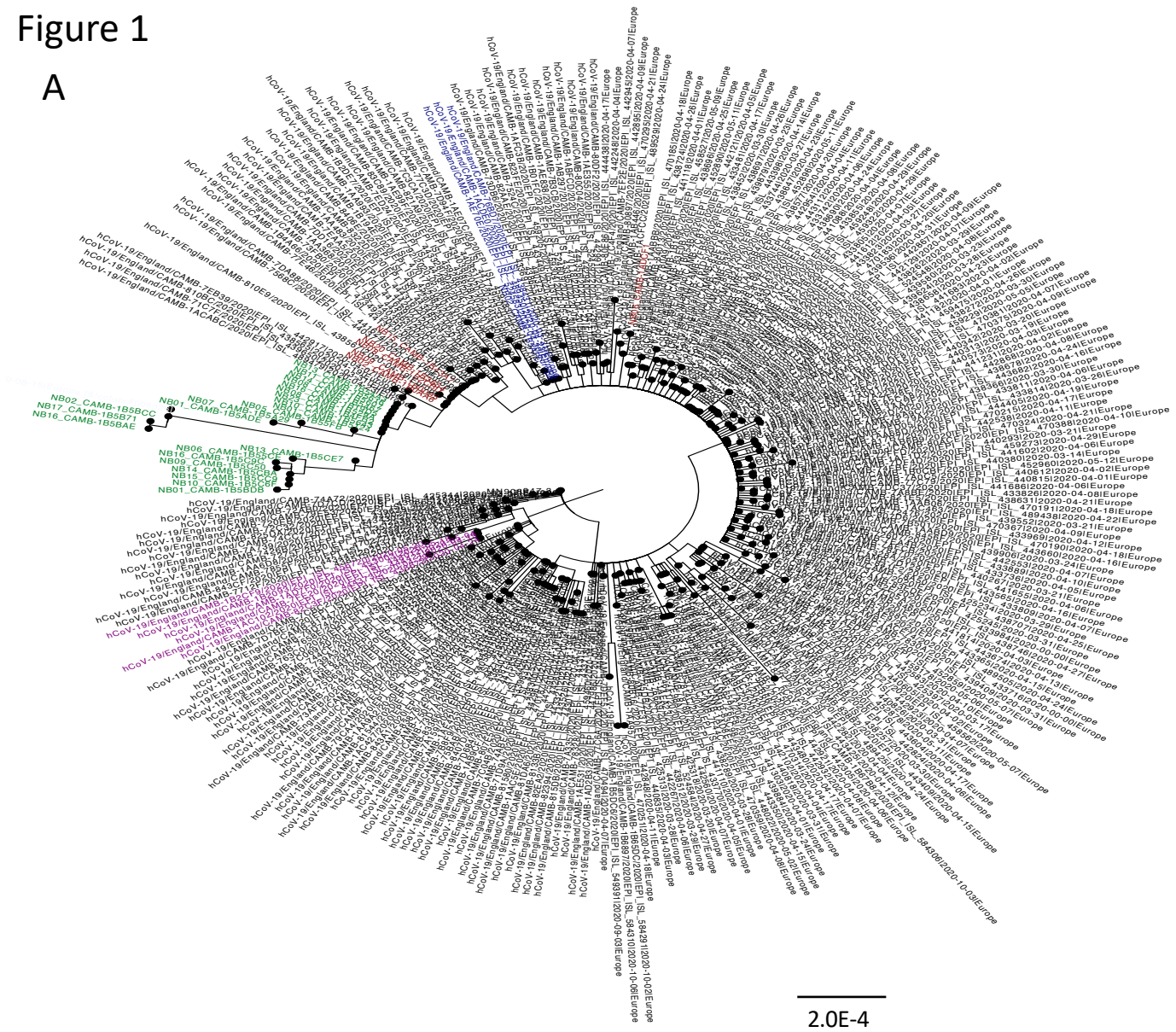
23 Collier, D. A. *et al.* Point of Care Nucleic Acid Testing for SARS-CoV-2 in Hospitalized Patients: A Clinical Validation Trial and Implementation Study. *Cell Rep Med*, 100062, doi:10.1016/j.xcrm.2020.100062 (2020).
24 Loman, N., Rowe, W. & Rambaut, A. (v1, 2020).
25 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10-12 (2011).
26 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)* **34**, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).
27 Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907* (2012).
28 Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **22**, 30494, doi:10.2807/1560-7917.ES.2017.22.13.30494 (2017).
29 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
30 Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530-1534, doi:10.1093/molbev/msaa015 (2020).
31 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587-589, doi:10.1038/nmeth.4285 (2017).
32 Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution* **30**, 1188-1195, doi:10.1093/molbev/mst024 (2013).
33 Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus evolution* **4**, vex042-vex042, doi:10.1093/ve/vex042 (2018).
34 Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **10**, e1003537 (2014).
35 Illingworth, C. J. SAMFIRE: multi-locus variant calling for time-resolved sequence data. *Bioinformatics* **32**, 2208-2209, doi:10.1093/bioinformatics/btw205 (2016).
36 Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus evolution* **1** (2015).
37 Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* **11**, e1004041 (2015).
38 Wrobel, A. G. *et al.* SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nature Structural & Molecular Biology* **27**, 763-767, doi:10.1038/s41594-020-0468-7 (2020).
39 Gregson, J. *et al.* HIV-1 viral load is elevated in individuals with reverse transcriptase mutation M184V/I during virological failure of first line antiretroviral therapy and is associated with compensatory mutation L74I. *The Journal of infectious diseases*, doi:10.1093/infdis/jiz631 (2019).
40 Naldini, L., Blomer, U., Gage, F. H., Trono, D. & Verma, I. M. Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 11382-11388 (1996).
41 Gupta, R. K. *et al.* Full-length HIV-1 Gag determines protease inhibitor susceptibility within in vitro assays. *Aids* **24**, 1651-1655, doi:10.1097/QAD.0b013e3283398216 (2010).

42 Mlcochova, P. *et al.* Combined point of care nucleic acid and antibody testing for SARS-CoV-2 following emergence of D614G Spike Variant. *Cell Rep Med*, 100099, doi:10.1016/j.xcrm.2020.100099 (2020).

43 Seow, J. *et al.* Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nat Microbiol* **5**, 1598-1607, doi:10.1038/s41564-020-00813-8 (2020).

Figure 1

A



B

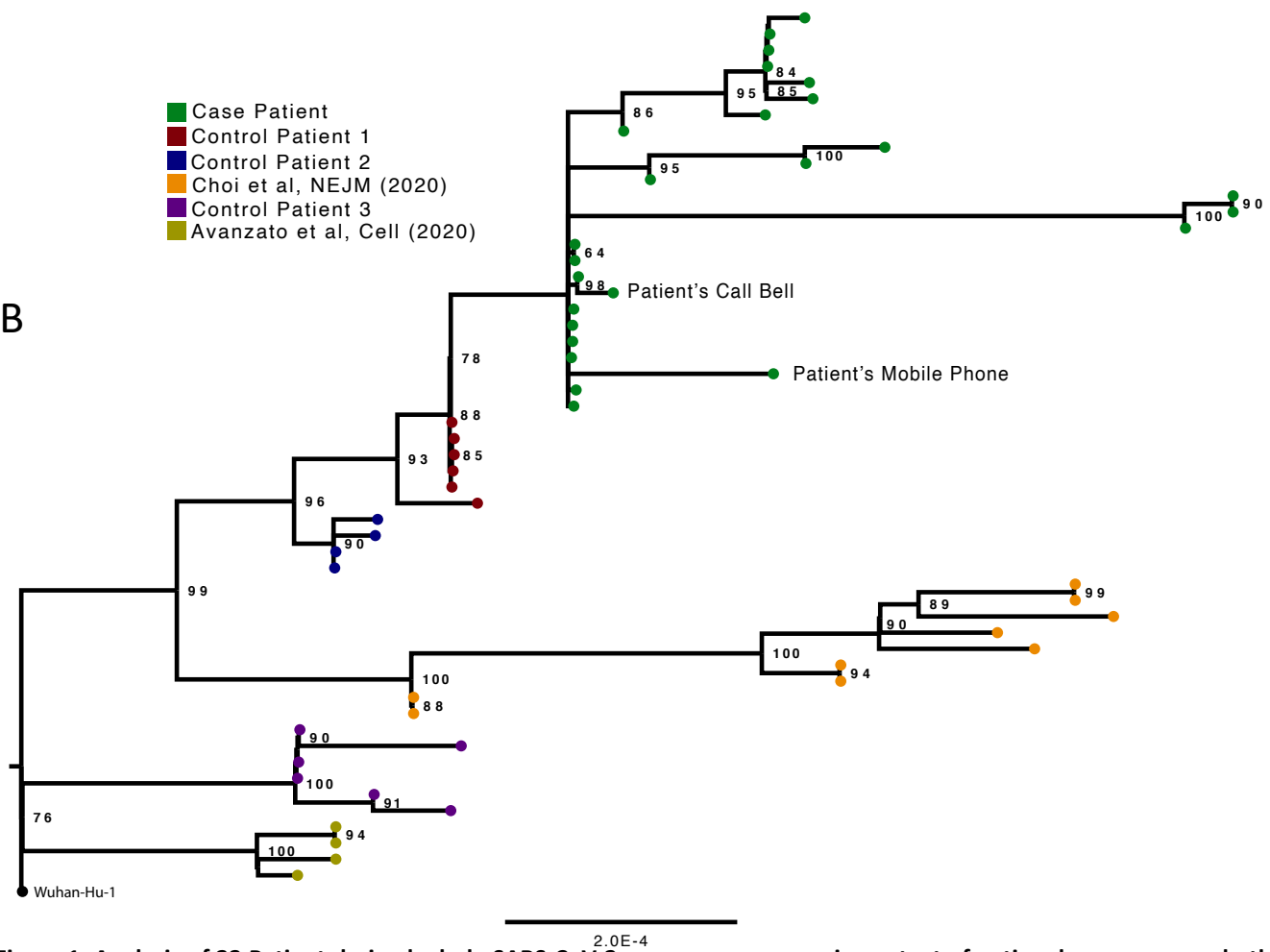


Figure 1. Analysis of 23 Patient derived whole SARS-CoV-2 genome sequences in context of national sequences and other cases of chronic SARS-CoV-2 shedding. A. Circularised maximum-likelihood phylogenetic tree rooted on the Wuhan-Hu-1 reference sequence, showing a subset of 250 local SARS-CoV-2 genomes from GISAID. This diagram highlights significant diversity of the case patient (green) compared to three other local patients with prolonged shedding (blue, red and purple sequences). All SARS-CoV-2 genomes were downloaded from the GISAID database and a random subset of 250 local sequences selected. **B.** Close-view maximum-likelihood phylogenetic tree indicating the diversity of the case patient and three other long-term shedders from the local area (red, blue and purple), compared to recently published sequences from Choi et al (orange) and Avanzato et al (gold)**.** Control patients generally showed limited diversity temporally, though the Choi et al sequences were found to be even more divergent than the case patient. Environmental samples are indicated. 1000 ultrafast bootstraps were performed and support at nodes is indicated.

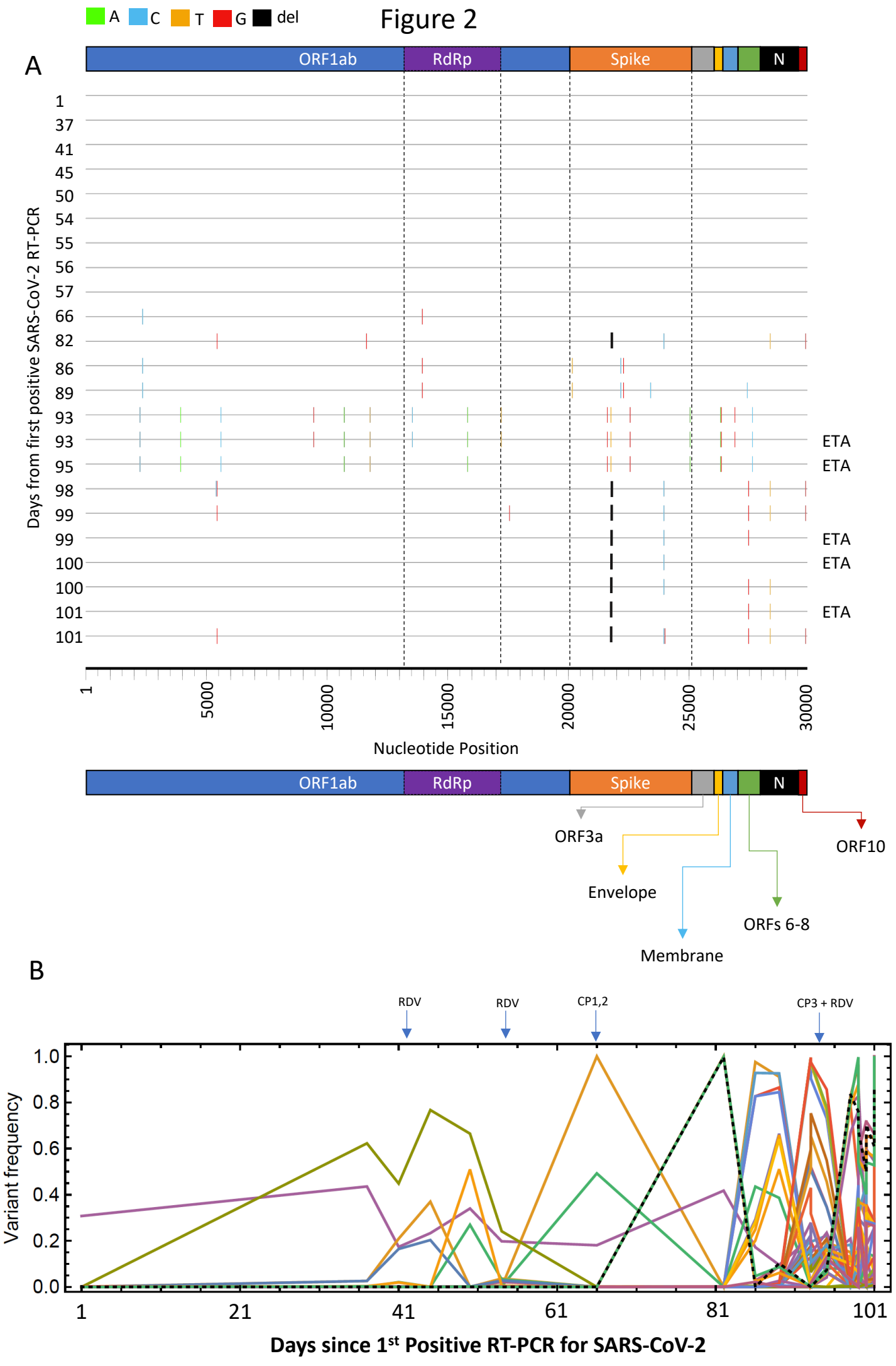
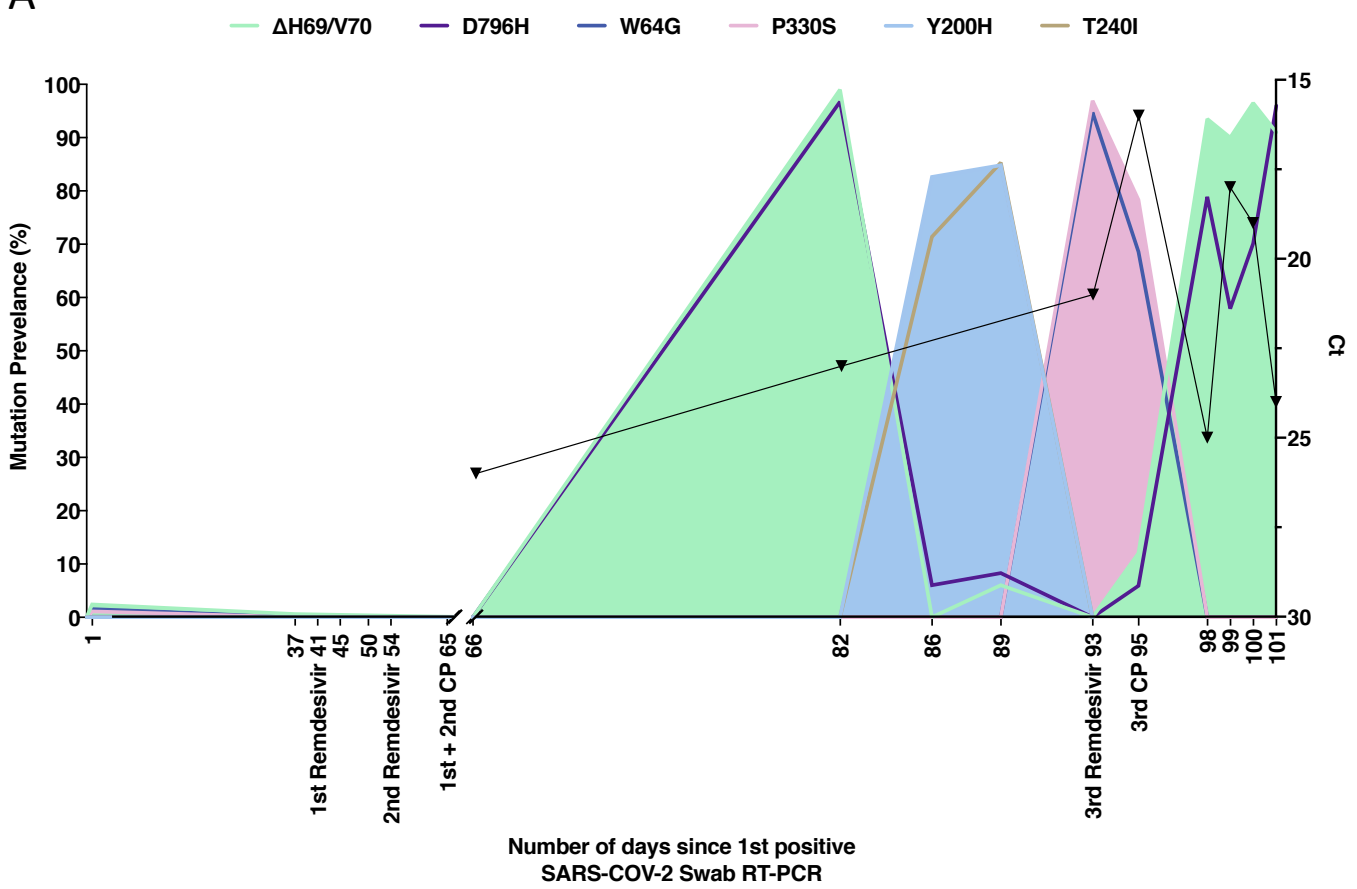


Figure 2. Virus genetics and population structure in chronic SARS-CoV-2 infection. A. Highlighter plot indicating nucleotide changes at consensus level in sequential respiratory samples compared to the consensus sequence at first diagnosis of COVID-19. Each row indicates the timepoint the sample was collected (number of days from first positive SARS-CoV-2 RT-PCR). Black dashed lines indicate the RNA-dependent RNA polymerase (RdRp) and Spike regions of the genome. Of particular interest, at consensus level, there were no nucleotide substitutions on days 1-57, despite the patient receiving two courses of remdesivir. The first major changes in the spike genome occurred on day 82, following convalescent plasma given on days 63 and 65. The double amino acid deletion in S1, ΔH69/V70 is indicated by black lines. All samples are nose and throat swabs unless indicated with ETA (endotracheal aspirate). **B.** Whole genome amino acid variant trajectories based on Illumina short read ultra deep sequencing at 1000x coverage. All variants which reached a frequency of at least 10% in at least two samples were plotted. Black dashed line represents Δ69/70. CP, convalescent plasma; RDV, Remdesivir.

Figure 3

A



B

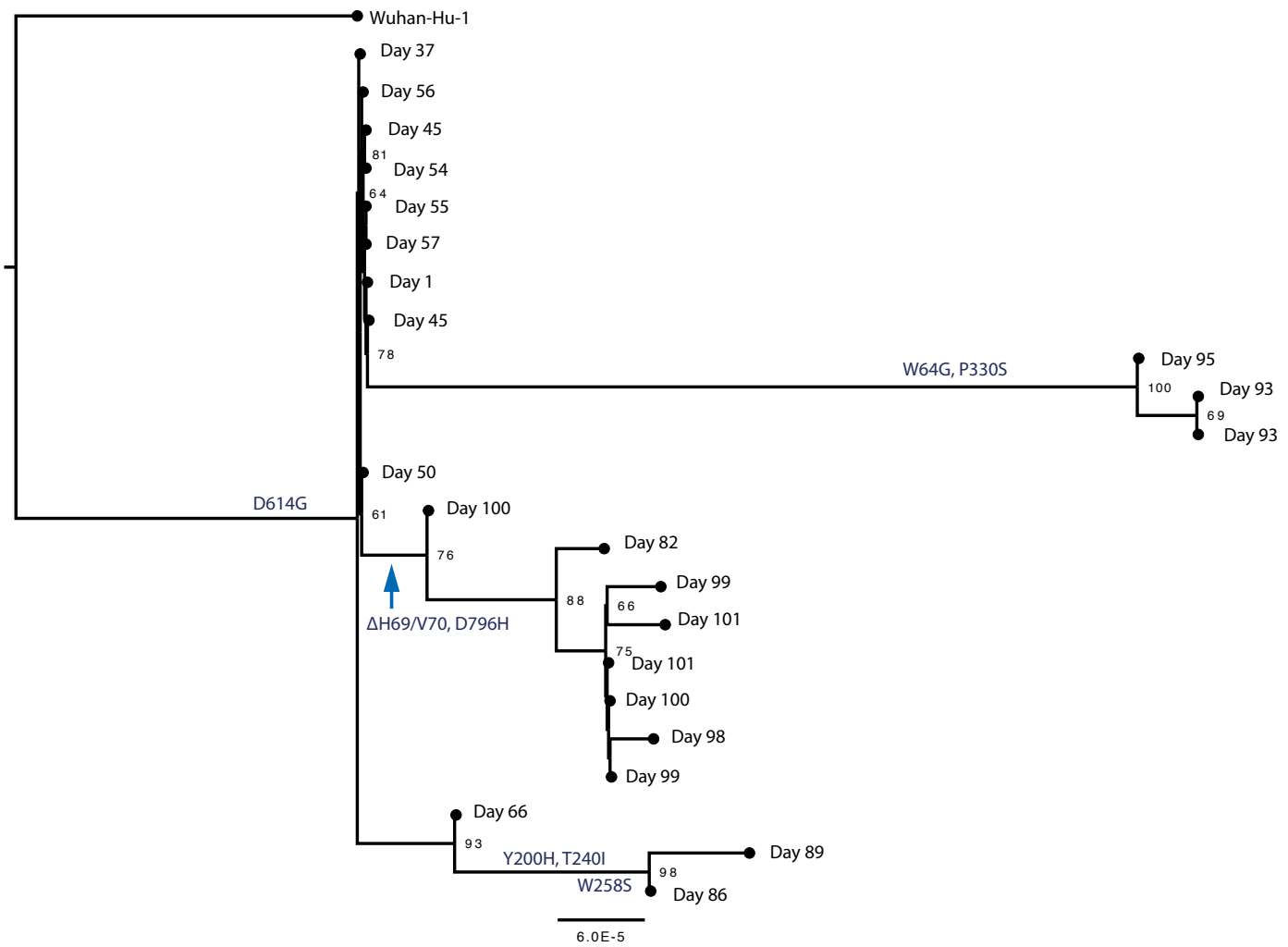


Figure 3. Longitudinal variant frequencies and phylogenetic relationships for virus populations bearing six Spike (S) mutations A. At baseline, all four S variants (Illumina sequencing) were absent (<1% and <20 reads). Approximately two weeks after receiving two units of convalescent plasma (CP), viral populations carrying ΔH69/V70 and D796H mutants rose to frequencies >90% but decreased significantly four days later. This population was replaced by a population bearing Y200H and T240I, detected in two samples over a period of six days. These viral populations were then replaced by virus carrying W64G and P330S mutations in Spike, which both reached near fixation at day 93. Following a 3rd course of remdesivir and an additional unit of convalescent plasma, the ΔH69/V70 and D796H virus population re-emerged to become the dominant viral strain reaching variant frequencies of >90%. Pairs of mutations arose and disappeared simultaneously indicating linkage on the same viral haplotype. **B.** Maximum likelihood phylogenetic tree of the case patient with day of sampling indicated. Spike mutations defining each of the clades are shown ancestrally on the branches on which they arose. Number at node denotes support by ultrafast bootstrapping consisting of 1000 replicates.

Figure 4

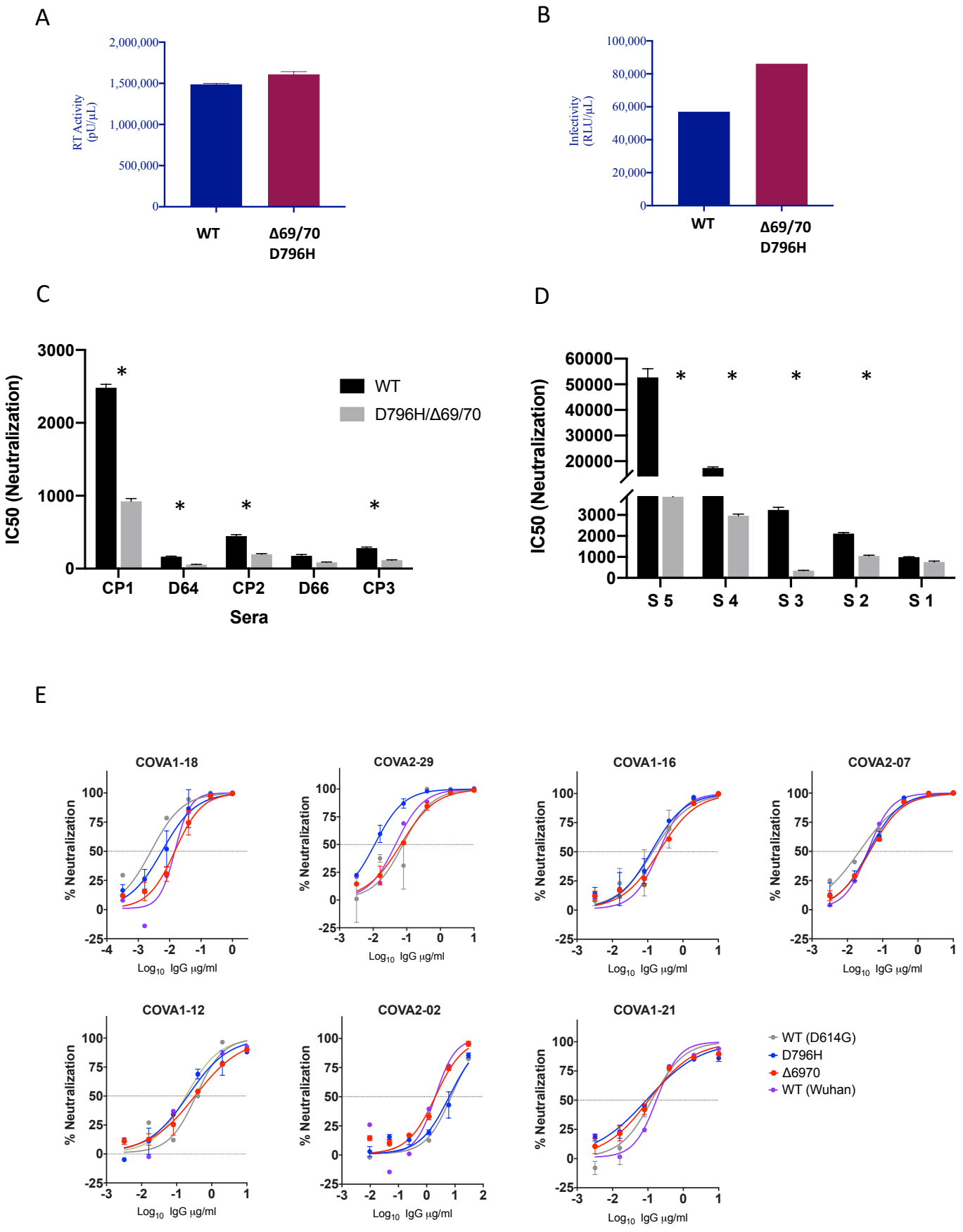


Figure 4: Spike mutant D796H + ΔH69/V70 has infectivity comparable to wild type but is less sensitive to multiple units of convalescent plasma (CP) and sera from recovered individuals. A. Reverse transcriptase activity of virus supernatants containing lentivirus pseudotyped with SARS-CoV-2 Spike protein (WT versus mutant) **B.** Single round Infectivity of luciferase expressing lentivirus pseudotyped with SARS-CoV-2 Spike protein (WT versus mutant) on 293T cells co-transfected with ACE2 and TMPRSS2 plasmids. **C.** Patient serum (taken at indicated Day (D)) and convalescent plasma (CP units 1-3) neutralization potency against Spike mutant D796H + ΔH69/V70 measured using lentivirus pseudotyped with SARS-CoV-2 Spike protein (WT versus mutant). Indicated is serum dilution required to inhibit 50% of virus infection. **D.** Neutralization potency of sera from 5 unselected convalescent patients (with previous confirmed SARS-CoV-2 infection) against WT versus mutant virus as in panel C. **E** Neutralization potency of a panel of monoclonal antibodies against lentivirus pseudotyped with SARS-CoV-2 Spike protein (WT Wuhan versus single mutants D614G, D796H, ΔH69/V70). Data are representative of at least two independent experiments * $p<0.05$

Figure 5

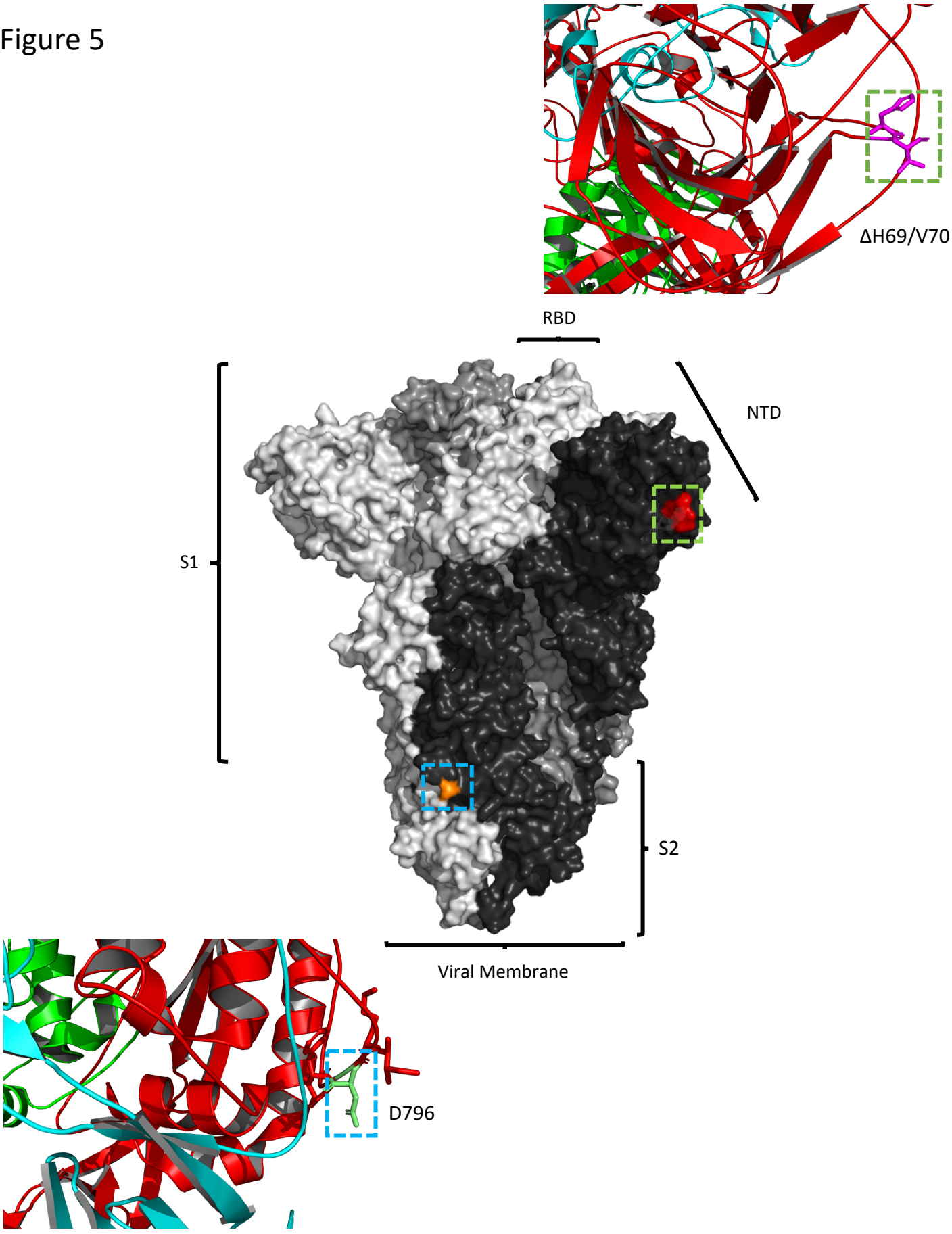


Figure 5. Location of Spike mutations ΔH69/Y70 in S1 and D796H in S2. Amino acid residues H69 and Y70 deleted in the N-terminal domain (red) and D796H in subunit 2 (orange) are highlighted on a SARS-CoV-2 spike trimer (PDB: 6ZGE Wrobek et al., 2020). Each of the three protomers making up the Spike homotrimer are coloured separately in shades of grey (centre). Close-ups of ΔH69/Y70 (above) and D796H (below) are shown in cartoon, stick representation. Both mutations are in exposed loops.