Neutralising antibodies in Spike mediated SARS-CoV-2 adaptation

Kemp SA[1]*, Collier DA[1,2,3]*, Datir R*[1,2,3], Ferreira IATM[2,3], Gayed S[4], Jahun A[5], Hosmillo M[5], Rees-Spear C[1], Mlcochova P[2,3], Ines Ushiro Lumb[6], David J 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[13], Doffinger R[15], Ceron-Gutierrez L[15], Barcenas-Morales G[15], Pollock DD[16], Goldstein RA[1], Smielewska A[5,13], Skittrall JP [4,14,17], Gouliouris T[4], Goodfellow IG[5], Gkrania-Klotsas E[4], Illingworth CJR[14,18], McCoy LE[1], Gupta RK[2,3,19]

[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] NHS Blood and Transplant, Oxford and BRC Haematology Theme, University of Oxford, UK

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

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

[9]The COVID-19 Genomics UK (COG-UK) Consortium, https://www.cogconsortium.uk. 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] Department of Clinical Biochemistry and Immunology, Addenbrookes Hospital

[16]Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, Colorado, USA;
[17]Clinical Microbiology and Public Health Laboratory, Addenbrookes' Hospital, Cambridge, UK
[18] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
[19]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 chronic SARS-CoV-2 with reduced sensitivity to 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. Although little change was observed in the overall viral population structure following two courses of remdesivir over the first 57 days, N501Y in Spike was transiently detected at day 55

and V157L in RdRp emerged. However, following convalescent plasma we observed large, dynamic virus population shifts, with the emergence of a dominant viral strain bearing D796H in S2 and ΔH69/ΔV70 in the S1 N-terminal domain NTD of the Spike protein. As passively transferred serum antibodies diminished, viruses with the escape genotype diminished in frequency, before returning during a final, unsuccessful course of convalescent plasma. *In vitro*, the Spike escape double mutant bearing ΔH69/ΔV70 and D796H conferred decreased sensitivity to convalescent plasma, whilst maintaining infectivity similar to wild type. D796H appeared to be the main contributor to decreased susceptibility, but incurred an infectivity defect. The ΔH69/ΔV70 single mutant had two-fold higher infectivity compared to wild type and appeared to compensate for the reduced infectivity of D796H. Consistent with the observed mutations being outside the RBD, monoclonal antibodies targeting the RBD were not impacted by either or both mutations, but a non RBD binding monoclonal antibody was less potent against ΔH69/ΔV70 and the double mutant. These data reveal strong selection on SARS-CoV-2 during convalescent plasma therapy associated with emergence of viral variants with reduced susceptibility to neutralising antibodies.

Introduction

SARS-CoV-2 is an RNA betacoronavirus, with closely related viruses identified in pangolins and bats[1,2]. RNA viruses have inherently higher rates of mutation than DNA viruses such as Herpesviridae[3]. However, amongst RNA viruses coronaviruses have a relatively modest mutation rate at around 23 nucleotide substitutions per year[4], likely due to proof reading capability of coronavirus RNA dependent RNA polymerase[5]. The capacity for successful adaptation is exemplified by the Spike D614G mutation, that arose in China and rapidly spread worldwide[6], now accounting for more than 90% of infections. The mutation appears to increase infectivity and transmissibility in animal models[7]. Although the SARS-CoV-2 Spike protein is critical for virus infection via engagement of ACE2, substantial Spike amino acid variation has been observed in circulating viruses[8]. Mutations in the receptor binding domain (RBD) of Spike are of particular concern because the RBD is an important target of neutralising antibodies and therapeutic monoclonal antibodies.

Deletions in the N-terminal domain (NTD) of Spike S1 have also been increasingly recognised, both within hosts[9] and across individuals[10]. The evolutionary basis for the emergence of deletions is unclear at present but could be related to escape from immunity or to enhanced fitness/transmission. The most notable deletion in terms of frequency is ΔH69/ΔV70. This double deletion has been detected in multiple unrelated lineages, including the recent 'Cluster 5' mink related strain in the North Jutland region of Denmark (https://files.ssi.dk/Mink-cluster-5-short-report_AFO2). There it was associated with the RBD mutation Y453F in almost 200 individuals. Another European cluster in GISAID includes ΔH69/ΔV70 along with the RBD mutation N439K.

Although ΔH69/ΔV70 has been detected multiple times, within-host emergence remains undocumented and the reasons for its selection are unknown. Here we document real time *SARS-CoV-2* emergence of ΔH69/ΔV70 in combination with the S2 mutation D796H following convalescent plasma therapy in an immunocompromised human host, demonstrating selection and reduced phenotypic susceptiblility of selected mutations.

Results

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

Clinical data are available from the corresponding author.

Given the history of B cell depletion therapy and hypogammaglobulinemia we measured serum SARS-CoV-2 specific antibodies over the course of the admission. Total serum antibodies to SARS-CoV-2 were tested at days 44 and 50 by S protein immunoassay (Siemens). Results were negative. Three units (200mL each) of convalescent plasma (CP) from three independent donors were obtained through the NHS Blood and Transplant Clearance Registry and administered on compassionate named patient basis. These had been assayed for SARS-CoV-2 IgG antibody titres (Supplementary figure 4). Patient serum was subsequently positive for SARS-

CoV-2 specific antibodies by S protein immunoassay (Siemens) in the hospital diagnostic laboratory on days 68, 90 and 101.

Virus genomic comparative analysis of 23 sequential respiratory samples over 101 days

The majority of samples were respiratory samples from nose and throat or endotracheal aspirates during the period of intubation. Ct values ranged from 16-34 and all 23 respiratory samples were successfully sequenced by standard single molecule sequencing approach as per the ARTIC protocol implemented by COG-UK; of these 20 additionally underwent short-read deep sequencing using the Illumina platform (Supplementary Table 3). There was generally good agreement between the methods (Supplementary Figure 5, Supplementary table 4). Additionally, Single genome amplification and sequencing of Spike using extracted RNA from respiratory samples was used as an independent method to detect mutations observed (Supplementary Table 5). Finally, we detected no evidence of recombination, based on two independent methods.

Maximum likelihood analysis of patient-derived whole genome consensus sequences demonstrated clustering with other local sequences from the same region (Figure 2A). The infecting strain was assigned to lineage 20B bearing the D614G Spike variant. Environmental sampling showed evidence of virus on surfaces such as telephone and call bell. Sequencing of these surface viruses showed clustering with those derived from the respiratory tract (Figure 2B). All samples were consistent with having arisen from a single viral population. In our phylogenetic analysis, we included sequential sequences from three other local patients identified with persistent viral RNA shedding over a period of 4 weeks or more as well as two long term immunosuppressed SARS-CoV-2 'shedders' recently reported[9,11], (Figure 2B and supplementary table 2). While the sequences from the three local patients as well as from Avanzato et al[11] showed little divergence with no amino acid changes in Spike over time, the case patient showed significant diversification. The Choi[9] et al report showed similar degree of diversification as the case patient. Further investigation of the sequence data suggested the existence of an underlying structure to the viral population in our patient, with samples

collected at days 93 and 95 being rooted within, but significantly divergent from the original population (Figure 2B and 3). The relationship of the divergent samples to those at earlier time points argues against superinfection.

SARS-CoV-2 antibodies in serum, and convalescent plasma associated changes in viral diversity

We measured longitudinal serum and convalescent plasma SARS-CoV-2 IgG to SARS-CoV-2 trimeric Spike (S), Spike receptor binding domain (RBD) and Nucleocapsid protein (N) using a Luminex based assay. Levels of antibodies were very low at day 39, consistent with the standard lab based assay result above. Following CP1 at day 63 and CP2 at day 65, antibody levels were significantly increased against all three protein targets. As expected antibody titres fell over time before increasing after the third unit of CP on day 93.

All samples tested positive by RT-PCR and there was no sustained change in Ct values throughout the 101 days following the first two courses of remdesivir (days 41 and 54), or the first two units of convalescent plasma (days 63 and 65). Consensus sequences from short read deep sequence Illumina data revealed dynamic population changes after day 65, as shown by a highlighter plot (Figure 3). In addition, we were also able to follow the dynamics of virus populations down to low frequencies during the entire period (Figure 4). Following remdesivir at day 41 the low frequency variant analysis allowed us to observe transient amino acid changes in populations at below 50% abundance in Orf 1b, 3a and Spike, with aT39I mutation in ORF7a reaching 77% on day 45 (Figure 4). At day 66 we noted I513T in NSP2 and V157L in RdRp had emerged from undetectable at day 54 to 100% frequency (Figure 4 orange line), with the polymerase being the more plausible candidate for driving this sweep. Notably, spike variant N501Y, which can increase the ACE2 receptor affinity[12], and which is present in the new UK B1.1.7 lineage[13], was observed on day 55 at 33% frequency, but was eliminated by the sweep of the NSP2/RdRp variant.

173 In contrast to the early period of infection, between days 66 and 82, following the first two
174 administrations of convalescent sera, a dramatic shift in the virus population was observed,
175 with a variant bearing D796H in S2 and ΔH69/ΔV70 in the S1 N-terminal domain (NTD)
176 becoming the dominant population at day 82. This was identified in a nose and throat swab
177 sample with high viral load as indicated by Ct of 23 (Figure 5). The deletion was not detected at
178 any point prior to the day 82 sample, even as minority variants by short read deep sequencing.
179
180 On Days 86 and 89, viruses obtained from upper respiratory tract samples were characterised
181 by the Spike mutations Y200H and T240I, with the deletion/mutation pair observed on day 82
182 having fallen to frequencies of 10% or less (Figure 4 and 5). The Spike mutations Y200H and
183 T240I were accompanied at high frequency by two other non-synonymous variants with similar
184 allele frequencies, coding for I513T in NSP2, V157L in RdRp and N177S in NSP15 (Figure 4). Of
185 these, the former was previously observed at 100% frequency in the sample on day 66 (Figure
186 3, orange line), arguing that this new lineage emerged out of a previously-existing population.
187
188 Sequencing of a nose and throat swab sample at day 93 identified viruses characterised by
189 Spike mutations P330S at the edge of the RBD and W64G in S1 NTD at close to 100%
190 abundance, with D796H along with ΔH69/ΔV70 at <1% abundance and the variants Y200H and
191 T240I at frequencies of <2%. Viruses with the P330S variant were detected in two independent
192 samples from different sampling sites, arguing against the possibility of contamination. The
193 divergence of these samples from the remainder of the population (Figure 4, 5B and
194 Supplementary Figure 6) suggests the possibility that they represent the emergence of a
195 previously unobserved subpopulation. Following the third course of remdesivir (day 93) and
196 third CP (day 95), we observed a re-emergence of the D796H + ΔH69/ΔV70 viral population.
197 The inferred linkage of D796H and ΔH69/ΔV70 was maintained as evidenced by the highly
198 similar frequencies of the two variants.
199
200 Patterns in the variant frequencies suggest competition between virus populations carrying
201 different mutations, viruses with the D796H/ΔH69/ΔV70 deletion/mutation pair rising to high

frequency during CP therapy, then being outcompeted by another population in the absence of therapy. Specifically, these data are consistent with a lineage of viruses with the NSP2 I513T and RdRp V157L variant, dominant on day 66, being outcompeted during therapy by the mutation/deletion variant. With the lapse in therapy, the original strain, having acquired NSP15, N1773S and the Spike mutations, regained dominance, followed by the emergence of a separate population with the W64G and P330S mutations.

In a final attempt to reduce the viral load, a third course of remdesivir (day 93) and third CP (day 95) were administered. We observed a re-emergence of the D796H + ΔH69/ΔV70 viral population. The inferred linkage of D796H and ΔH69/ΔV70 was maintained as evidenced by the highly similar frequencies of the two variants, suggesting that the third unit of CP led to the re-emergence of this population under renewed positive selection. In further support of our proposed idea of competition, noted above, frequencies of these two variants appeared to mirror changes in the NSP2 I513T mutation (Figure 4), suggesting these as markers of opposing clades in the viral population. Ct values remained low throughout this period with hyperinflammation, eventually leading to multi-organ failure and death at day 102. The repeated increase in frequency of the viral population with CP therapy strongly supports the hypothesis that the deletion/mutation combination conferred selective advantage.

Spike mutants emerging post convalescent plasma impair neutralisating antibodies

Using lentiviral pseudotyping we generated wild type, ΔH69/ΔV70 + D796H and single mutant Spike proteins in enveloped virions in order to measure neutralisation activity of CP against these viruses (Figure 6). This system has been shown to give generally similar results to replication competent virus[14,15]. Spike protein from each mutant was detected in pelleted virions (Figure 6A). We also probed with an HIV-1 p24 antibody to monitor levels of lentiviral particle production. We then measured infectivity of the pseudoviruses, correcting for virus input, and found that ΔH69/ΔV70 appeared to have two-fold higher infectivity over a single round of infection compared to wild type (Figure 6B, supplementary figure 7). By contrast, the

D796H single mutant had significantly lower infectivity as compared to wild type and the double mutant had similar infectivity to wild type (Figure 6B, supplementary figure 7).

We found that D796H alone and the D796H + ΔH69/ΔV70 double mutant were less sensitive to neutralisation by convalescent plasma samples (Figure 6C-E). By contrast the ΔH69/ΔV70 single mutant did not impact neutralisation. In addition, patient derived serum from days 64 and 66 (one day either side of CP2 infusion) similarly showed lower potency against the D796H + ΔH69/ΔV70 mutants (Figure 6F, G).

A panel of nineteen monoclonal antibodies (mAbs) isolated from three donors was previously identified to neutralize SARS-CoV-2. To establish if the mutations incurring in vivo (D796H and ΔH69/ΔV70) resulted in a global change in neutralization sensitivity we tested neutralising mAbs targeting the seven major epitope clusters previously described (excluding non-neutralising clusters II, V and small [n =<2] neutralising clusters IV, X). The seven RBD-specific mAbs (Supplementary table 6) exhibited no major change in neutralisation potency and non-RBD specific COVA1-21 showing 3-5 fold reduction in potency against ΔH69/ΔV70+D796H and ΔH69/ΔV70, but not D796H alone[15] (Figure 7). We observed no differences in neutralisation between single/double mutants and wild type, suggesting that the mechanism of escape was likely outside these epitopes in the RBD. These data confirm the specificity of the findings from convalescent plasma and suggest that mutations observed are related to antibodies targeting regions outside the RBD.

In order to understand the mechanisms that might confer resistance to antibodies we examined a published Spike structure and annotated it with our residues of interest (Figure 8). This analysis showed that ΔH69/ΔV70 is in a disordered, glycosylated loop at the very tip of the NTD, and therefore could alter binding of antibodies. ΔH69/V70 is close to the binding site of the polyclonal antibodies derived from COV57 plasma[16,17]. D796H is in an exposed loop in S2

(Figure 8) and appears to be in a region frequently targeted by antibodies[18], despite mutations at position 796 being rare (Supplementary table 7).

Discussion

Here we have documented a repeated evolutionary response by SARS-CoV-2 in the presence of antibody therapy during the course of a persistent infection in an immunocompromised host. The observation of potential selection for specific variants coinciding with the presence of antibodies from convalescent plasma is supported by the experimental finding of two-fold reduced susceptibility of these viruses to plasma. Further, we were able to document real-time emergence of a variant ΔH69/ΔV70 in the NTD of Spike that has been increasing in frequency in Europe and is present in the new UK variant B1.1.7[19]. In this case the emergence of the variant was not the primary reason for treatment failure. However, given that both vaccines and therapeutics are aimed at Spike, our study raises the possibility of virus evasion, particularly in immune suppressed individuals where prolonged viral replication occurs.

Our observations represent a very rare insight, possible due to poor T cell responses and a lack of antibodies in the individual, and an intensive sampling course. Persistent viral replication and the failure of antiviral therapy allowed us to define the viral response to convalescent plasma. Our findings follow those of Choi et al[9], who reported persistent infection in an immune suppressed individual; they noted significant virus evolution, including NTD deletions and RBD mutations in the absence of SARS-CoV-2 specific antibody therapy. A second paper reported asymptomatic long term shedding with four sequences over 105 days[11], demonstrated similarly dramatic shifts in genetic composition of the viral population without phenotypic impact. A common finding with our study was the very low neutralisation activity in serum post transfusion of CP with waning over time as expected. Apart from the difference in the outcome of infection (severe, fatal disease versus asymptomatic disease and clearance), critically important differences in our study include: 1. The intensity of sampling and use of both long and short read sequencing to verify variant calls, thereby providing a unique scientific resource for longitudinal population genetic analysis. 2. The close alignment between the genetic

288 composition of the viral population and CP administration, with an experimentally verified
289 variant with reduced susceptibility emerging, falling to low frequency, and then rising again
290 under CP selection. 3. Real time detection of emergence of a variant, ΔH69/ΔV70, that is
291 increasing in frequency in Europe, and present in the new UK multiply mutated variant B1.1.7.
292
293 An interesting observation is that in the two cases of chronic infection highlighted, the
294 Avanzato case where CP was used for asymptomatic shedding[11] exhibited lower diversification
295 of virus as compared to the fatal Choi et al case where CP was not used[9]. There are clearly a
296 number of factors that could account for these differences, though this highlights the fact that
297 use of CP does not necessarily lead to rapid adaptation. Intriguingly, deletions spanning amino
298 acids 139-145 emerged in both these cases in contrast to the ΔH69/ΔV70 observed in the
299 present study. Deletion of amino acid 144 and ΔH69/ΔV70 is observed in the UK lineage
300 B1.1.7[19], and is therefore of concern.
301
302 We have noted in our analysis the potential influence of compartmentalised viral replication
303 upon the sequences recovered in upper respiratory tract samples. Both population genetic and
304 small animal studies have shown a lack of reassortment between influenza viruses within a
305 single host during an infection, suggesting that acute respiratory viral infection may be
306 characterised by spatially distinct viral populations[20,21]. In the analysis of data, it is important to
307 distinguish genetic changes which occur in the primary viral population from apparent changes
308 that arise from the stochastic observation of spatially distinct subpopulations in the host. While
309 the samples we observe on days 93 and 95 of infection are genetically distinct from the others,
310 the remaining samples are consistent with arising from a consistent viral population. We note
311 that Choi et al reported the detection in post-mortem tissue of viral RNA not only in lung tissue,
312 but also in the spleen, liver, and heart[9]. Mixing of virus from different compartments, for
313 example via blood, or movement of secretions from lower to upper respiratory tract, could lead
314 to fluctuations in viral populations at particular sampling sites. Experiments in animal models
315 with sampling of different replication sites could allow a better understanding of SARS-CoV-2

population genetics and enable prediction of escape variants following antibody based therapies.

This is a single case report and therefore limited conclusions can be drawn about generalisability.

In addition to documenting the emergence of SARS-CoV-2 Spike ΔH69/ΔV70 *in vivo*, we show that this mutation increases infectivity of the Spike protein in a pseudotyping assay. The deletion was observed contemporaneously with the rare S2 mutation D796H after two separate courses of CP, with other viral populations emerging. D796H, but not ΔH69/ΔV70, conferred reduction in susceptibility to polyclonal antibodies in the units of CP administered, though we cannot speculate as to their individual impacts on sera from other individuals. Importantly, neither of these mutations alone or in combination affected susceptibility of virus to a set of RBD-targeting monoclonal antibodies. The reduced sensitivity of ΔH69/ΔV70 to a non RBD binding antibody may hint at a role in antigen recognition, for example to polyclonal sera from COV-57.

The effects of CP on virus evolution seen here are unlikely to apply in immune competent hosts where viral diversity is likely to be lower due to better immune control. Our data highlight that infection control measures may need to be tailored to the needs of immunocompromised patients and also caution in interpretation of CDC guidelines that recommend 20 days as the upper limit of infection prevention precautions in immune compromised patients who are afebrile[22]. Due to the difficulty with culturing clinical isolates, use of surrogates are warranted[23]. However, where 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.

Clinical efficacy of convalescent plasma in severe COVID-19 has not been demonstrated[24], and its use in different stages of infection and disease remains experimental; as such, we suggest

that it should be reserved for use within clinical trials, with rigorous monitoring of clinical and virological parameters. The data from this single case report might warrant caution in use of convalescent plasma 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 theoretically raising the potential for escape mutations. Whilst we await further data, where clinical trial enrolment is not possible, convalescent plasma administered for clinical need in immune suppression should ideally only be considered as part of observational studies, undertaken preferably in single occupancy rooms with enhanced infection control precautions, including SARS-CoV-2 environmental sampling and real-time sequencing. Understanding of viral dynamics and characterisation of viral evolution in response to different selection pressures in the immunocompromised host is necessary not only for improved patient management but also for public health benefit.

Ethics

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

Acknowledgements

We are immensely grateful to the patient and his family. We would also like to thank the staff at CUH and the NIHR Cambridge Clinical Research Facility. We would like to thank Dr Ruthiran Kugathasan and Professor Wendy Barclay for helpful discussions and Dr Martin Curran, Dr William Hamilton, and Dr. Dominic Sparkes. We would like to thank Prof Andres Floto and Prof Ferdia Gallagher. We thank Dr James Voss for the kind gift of HeLa cells stably expressing ACE2. COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research

Limited, operating as the Wellcome Sanger Institute. RKG is supported by a Wellcome Trust Senior Fellowship in Clinical Science (WT108082AIA). LEM is supported by a Medical Research Council Career Development Award (MR/R008698/1). SAK is supported by the Bill and Melinda Gates Foundation via PANGEA grant: OPP1175094. DAC is supported by a Wellcome Trust Clinical PhD Research Fellowship. CJRI acknowledges MRC funding (ref: MC_UU_00002/11). This research was supported by the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre, the Cambridge Clinical Trials Unit (CCTU) and by the UCL Coronavirus Response Fund and made possible through generous donations from UCL's supporters, alumni, and friends (LEM). JAGB is supported by the Medical Research Council (MC_UP_1201/16). IG is a Wellcome Senior Fellow and supported by the Wellcome Trust (207498/Z/17/Z). DDP is supported by NIH GM083127.

Author contributions

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

Designed experiments: RKG, SAK, DAC, LEM, JAGB, EGK, AC, NT, AC, CS, RD, RG, DDP

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

Interpreted data: RKG, SAK, DAC, PM, LEM, JAGB, PM, SG, KS, TG, JB, KGCS, IG, CJRI, JAGB, IUL, DR, JS, BB, RAG. DDP, RD, LCG, GBM

Methods

Clinical Sample Collection and Next generation sequencing

Serial samples were collected from the patient periodically from the lower respiratory tract (sputum or endotracheal aspirate), upper respiratory tract (throat and nasal swab), and from stool. Nucleic acid extraction was done from 500µl of sample with a dilution of MS2 bacteriophage to act as an internal control, using the easyMAG platform (Biomerieux, Marcy-l'Étoile) according to the manufacturers' instructions. All samples were tested for presence of SARS-CoV-2 with a validated one-step RT q-PCR assay developed in conjunction with the Public Health England Clinical Microbiology [25]. Amplification reaction were all performed on a

403 Rotorgene™ PCR instrument. Samples which generated a CT of ≤36 were considered to be
404 positive.
405
406 Sera from recovered patients in the COVIDx study[26] were used for testing of neutralisation
407 activity by SARS-CoV-2 mutants.
408
409 **SARS-CoV-2 serology by multiplex particle-based flow cytometry (Luminex)**:
410 Recombinant SARS-CoV-2 N, S and RBD were covalently coupled to distinct carboxylated
411 bead sets (Luminex; Netherlands) to form a 3-plex and analyzed as previously described
412 (Xiong et al. 2020). Specific binding was reported as mean fluorescence intensities (MFI).
413
414 **Whole blood T cell and innate stimulation assay**
415 Whole blood was diluted 1:5 in RPMI into 96-well F plates (Corning) and activated by single
416 stimulation with phytohemagglutinin (PHA; 10 µg/ml; Sigma-Aldrich), or LPS (1 µg/ml, List
417 Biochemicals) or by co-stimulating with anti-CD3 (MEM57, Abcam, 200 ng/ml,) and IL-2
418 (Immunotools, 1430U/ml,). Supernatants were taken after 24 hours. Levels (pg/ml) are shown
419 for IFNg, IL17, IL2, TNFa, IL6, IL1b and IL10. Cytokines were measured by multiplexed particle
420 based Flow cytometry on a Luminex analyzer (Bio-Plex, Bio-Rad, UK) using an R&D Systems
421 custom kit (R&D Systems, UK).
422
423 For viral genomic sequencing, total RNA was extracted from samples as described. Samples
424 were sequenced using MinION flow cells version 9.4.1 (Oxford Nanopore Technologies)
425 following the ARTICnetwork V3 protocol (https://dx.doi.org/10.17504/protocols.io.bbmuik6w)
426 and BAM files assembled using the ARTICnetwork assembly pipeline
427 (https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html). A representative set of
428 10 sequences were selected and also sequenced using the Illumina MiSeq platform. Amplicons
429 were diluted to 2 ng/µl and 25 µl (50 ng) were used as input for each library preparation
430 reaction. The library preparation used KAPA Hyper Prep kit (Roche) according to manufacturer's
431 instructions. Briefly, amplicons were end-repaired and had A-overhang added; these were then

432 ligated with 15mM of NEXTflex DNA Barcodes (Bio Scientific, Texas, USA). Post-ligation products
433 were cleaned using AMPure beads and eluted in 25 µl. Then, 20 µl were used for library
434 amplification by 5 cycles of PCR. For the negative controls, 1ng was used for ligation-based
435 library preparation. All libraries were assayed using TapeStation (Agilent Technologies,
436 California, USA) to assess fragment size and quantified by QPCR. All libraries were then pooled
437 in equimolar accordingly. Libraries were loaded at 15nM and spiked in 5% PhiX (Illumina,
438 California, USA) and sequenced on one MiSeq 500 cycle using a Miseq Nano v2 with 2x 250
439 paired-end sequencing. A minimum of ten reads were required for a variant call.
440
441 *Bioinformatics Processes*
442 For long-read sequencing, genomes were assembled with reference-based assembly and a
443 curated bioinformatics pipeline with 20x minimum coverage across the whole-genome [27]. For
444 short-read sequencing, FASTQs were downloaded, poor-quality reads were identified and
445 removed, and both Illumina and PHiX adapters were removed using TrimGalore v0.6.6 [28].
446 Trimmed paired-end reads were mapped to the National Center for Biotechnology Information
447 SARS-CoV-2 reference sequence MN908947.3 using MiniMap2-2.17 with arguments -ax and sr
448 [29]. BAM files were then sorted and indexed with samtools v1.11 and PCR optical duplicates
449 removed using Picard (http://broadinstitute.github.io/picard). A consensus sequences of
450 nucleic acids with a minimum whole-genome coverage of at least 20× were generated with
451 BCFtools using a 0% majority threshold.
452
453 *Single Genome Amplification and sequencing*
454
455 Viral RNA extracts were reverse transcribed from each sample to sufficiently capture the
456 diversity of the viral population without introducing resampling bias. SuperScript IV
457 (Thermofisher Scientific) and the gene specific primers were used for reverse transcription.
458 Template RNA was degraded with RNAse H (Thermofisher Scientific). All primers used were 'in-
459 house' primers designed using the multiple sequence alignment of the patient's consensus NGS
460 sequences. Partial Spike (amino acids 21- 800) was amplified as 1 continuous length of DNA
461 (Spike ~ 1.8 kb) by nested PCR. Terminally diluted cDNA was PCR- amplified using Platinum®

462 Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA) so that 30% of reactions were
463 positive[30]. By Poisson statistics, sequences were deemed ≥80% likely to be derived from HIV-1
464 single genomes. We obtained between 20–60 single genomes at each sample time point to
465 achieve 90% confidence of detecting variants present at ≥8% of the viral population in vivo[31,32].
466 Partial spike amplicons obtained from terminal dilution PCR amplification were Sanger
467 sequenced to form a contiguous sequence using another set of 8 in-house primers. Sanger
468 sequencing was provided by Genewiz UK and manual sequence editing was performed using
469 DNA Dynamo software (Blue Tractor Software Ltd, UK).
470
471 *Phylogenetic Analysis*
472 All available full-genome SARS-CoV-2 sequences were downloaded from the GISAID database
473 (http://gisaid.org/) [33] on 16th December. Duplicate and low-quality sequences (>5% N regions)
474 were removed, leaving a dataset of 212,297 sequences with a length of >29,000bp. All
475 sequences were sorted by name and only sequences sequenced with United Kingdom / England
476 identifiers were retained. From this dataset, sequecnes were de-duplicated and where
477 background sequences were required in figures, randomly subsampled using seqtk
478 (https://github.com/lh3/seqtk). All sequences were aligned to the SARS-CoV-2 reference strain
479 MN908947.3, using MAFFT v7.475 with automatic flavour selection[34]. Major SARS-CoV-2 clade
480 memberships were assigned to all sequences using both the Nextclade server v0.9
481 (https://clades.nextstrain.org/) and Phylogenetic Assignment Of Named Global Outbreak
482 Lineages (pangolin)[35].
483
484 Maximum likelihood phylogenetic trees were produced using the above curated dataset using
485 IQ-TREE v2.1.2[36]. Evolutionary model selection for trees were inferred using ModelFinder[37] and
486 trees were estimated using the GTR+F+I model with 1000 ultrafast bootstrap replicates[38]. All
487 trees were visualised with Figtree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/), rooted
488 on the SARS-CoV-2 reference sequence and nodes arranged in descending order. Nodes with
489 bootstraps values of <50 were collapsed using an in-house script.
490

In-depth allele frequency variant calling

The SAMFIRE package[39] was used to call allele frequency trajectories from BAM file data. Reads were included in this analysis if they had a median PHRED score of at least 30, trimming the ends of reads to achieve this if necessary. Nucleotides were then filtered to have a PHRED score of at least 30; reads with fewer than 30 such reads were discarded. Distances between sequences, accounting for low-frequency variant information, was also conducted using SAMFIRE. The sequence distance metric, described in an earlier paper [40], combines allele frequencies across the whole genome. Where L is the length of the genome, we define q(t) as a 4 x L element vector describing the frequencies of each of the nucleotides A, C, G, and T 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.

Western blot analysis.

517
518 Forty-eight hours after transfection of cells with plasmid preparations, the culture supernatant
519 was harvested and passed through a 0.45-µm-pore-size filter to remove cellular debris. The
520 filtrate was centrifuged at 15,000 rpm for 120 min to pellet virions. The pelleted virions were
521 lysed in Laemmli reducing buffer (1 M Tris-HCl [pH 6.8], SDS, 100% glycerol, β-mercaptoethanol,
522 and bromophenol blue). Pelleted virions were subjected to electrophoresis on SDS–4 to 12%
523 bis-Tris protein gels (Thermo Fisher Scientific) under reducing conditions. This was followed by
524 electroblotting onto polyvinylidene difluoride (PVDF) membranes. The SARSCOV-2 Spike
525 proteins were visualized by a ChemiDoc® MP imaging system (Biorad) using anti-Spike S2
526 (Invitrogen) and anti-p24 Gag antibodies.
527
528 *Recombination Detection*
529 All sequences were tested for potential recombination, as this would impact on evolutionary
530 estimates. Potential recombination events were explored with nine algorithms (RDP, MaxChi,
531 SisScan, GeneConv, Bootscan, PhylPro, Chimera, LARD and 3SEQ), implemented in RDP5 with
532 default settings [41]. To corroborate any findings, ClonalFrameML v1.12 [42] was also used to infer
533 recombination breakpoints. Neither programs indicated evidence of recombination in our data.
534
535
536 *Structural Viewing*
537 The Pymol Molecular Graphics System v2.4.0 (https://github.com/schrodinger/pymol-open-
538 source/releases) was used to map the location of the four spike mutations of interested onto a
539 SARS-CoV-2 spike structure visualised by Wrobel et al (PDB: 6ZGE)[43].
540
541 *Testing of convalescent plasma for antibody titres*
542 The Anti-SARS-CoV-2 ELISA (IgG) assay used to test CP for *antibody titres* was Euroimmun
543 Medizinische Labordiagnostika AG. This indirect ELISA based assay uses a recombinant
544 structural spike 1 (S1) protein of SARS-CoV-2 expressed in the human cell line HEK 29 for the
545 detection of SARS-CoV2 IgG.
546

547 *Generation of Spike mutants*

548 Amino acid substitutions were introduced into the D614G pCDNA_SARS-CoV-2_Spike plasmid
549 as previously described[44] using the QuikChange Lightening Site-Directed Mutagenesis kit,
550 following the manufacturer's instructions (Agilent Technologies, Inc., Santa Clara, CA).

551

552 *Pseudotype virus preparation*

553 Viral vectors were prepared by transfection of 293T cells by using Fugene HD transfection
554 reagent (Promega). 293T cells were transfected with a mixture of 11ul of Fugene HD, 1µg of
555 pCDNAΔ19Spike-HA, 1ug of p8.91 HIV-1 gag-pol expression vector[45,46], and 1.5µg of pCSFLW
556 (expressing the firefly luciferase reporter gene with the HIV-1 packaging signal). Viral
557 supernatant was collected at 48 and 72h after transfection, filtered through 0.45um filter and
558 stored at -80˚C. The 50% tissue culture infectious dose ($TCID_{50}$) of SARS-CoV-2 pseudovirus
559 was determined using Steady-Glo Luciferase assay system (Promega).

560

561 *Standardisation of virus input by SYBR Green-based product-enhanced PCR assay (SG-PERT)*

562 The reverse transcriptase activity of virus preparations was determined by qPCR using a SYBR
563 Green-based product-enhanced PCR assay (SG-PERT) as previously described[47].
564 Briefly, 10-fold dilutions of virus supernatant were lysed in a 1:1 ratio in a 2x lysis solution
565 (made up of 40% glycerol v/v 0.25% Trition X-100 v/v 100mM KCl, RNase inhibitor 0.8 U/ml,
566 TrisHCL 100mM, buffered to pH7.4) for 10 minutes at room temperature.

567

568 12µl of each sample lysate was added to thirteen 13µl of a SYBR Green master mix (containing
569 0.5µM of MS2-RNA Fwd and Rev primers, 3.5pmol/ml of MS2-RNA, and 0.125U/µl of Ribolock
570 RNAse inhibitor and cycled in a QuantStudio. Relative amounts of reverse transcriptase activity
571 were determined as the rate of transcription of bacteriophage MS2 RNA, with absolute RT
572 activity calculated by comparing the relative amounts of RT to an RT standard of known activity.

573
574

575 *Serum/plasma pseudotype neutralization assay*

576 Spike pseudotype assays have been shown to have similar characteristics as neutralisation
577 testing using fully infectious wild type SARS-CoV-2[14].Virus neutralisation assays were performed
578 on 293T cell transiently transfected with ACE2 and TMPRSS2 using SARS-CoV-2 Spike
579 pseudotyped virus expressing luciferase[48]. Pseudotyped virus was incubated with serial dilution
580 of heat inactivated human serum samples or convalescent plasma in duplicate for 1h at 37˚C.
581 Virus and cell only controls were also included. Then, freshly trypsinized 293T ACE2/TMPRSS2
582 expressing cells were added to each well. Following 48h incubation in a 5% CO_2 environment at
583 37°C, the luminescence was measured using Steady-Glo Luciferase assay system (Promega).
584
585 *mAb pseudotype neutralisation assay*
586 Virus neutralisation assays were performed on HeLa cells stably expressing ACE2 and using
587 SARS-CoV-2 Spike pseudotyped virus expressing luciferase as previously described[49].
588 Pseudotyped virus was incubated with serial dilution of purified mAbs[15] in duplicate for 1h at
589 37˚C. Then, freshly trypsinized HeLa ACE2- expressing cells were added to each well. Following
590 48h incubation in a 5% CO_2 environment at 37°C, the luminescence was measured using Bright-
591 Glo Luciferase assay system (Promega) and neutralization calculated relative to virus only
592 controls. IC50 values were calculated in GraphPad Prism.
593
594 **Data Availability**
595 Long-read sequencing data that support the findings of this study have been deposited in the
596 NCBI SRA database with the accession codes SAMN16976824 - SAMN16976846 under
597 BioProject PRJNA682013 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA682013). Short reads
598 and data used to construct figures were deposited at https://github.com/Steven-
599 Kemp/sequence_files.
600
601 **References**
602 1 Zhang, T., Wu, Q. & Zhang, Z. Probable pangolin origin of SARS-CoV-2 associated with the
603 COVID-19 outbreak. *Current Biology* (2020).
604 2 Xiao, K. *et al.* Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* **583**,
605 286-289, doi:10.1038/s41586-020-2313-x (2020).

606 3 Sanjuán, R. & Domingo-Calap, P. Mechanisms of viral mutation. *Cell Mol Life Sci* **73**, 4433-4448,
607 doi:10.1007/s00018-016-2299-6 (2016).
608 4 Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121-
609 4123 (2018).
610 5 Robson, F. *et al.* Coronavirus RNA proofreading: molecular basis and therapeutic targeting.
611 *Molecular cell* (2020).
612 6 Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases
613 Infectivity of the COVID-19 Virus. *Cell*, doi:10.1016/j.cell.2020.06.043 (2020).
614 7 Yurkovetskiy, L. *et al.* Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein
615 Variant. *Cell* **183**, 739-751 e738, doi:10.1016/j.cell.2020.09.032 (2020).
616 8 Korber, B. *et al.* Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity
617 of the COVID-19 virus. *Cell* **182**, 812-827. e819 (2020).
618 9 Choi, B. *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N Engl*
619 *J Med*, doi:10.1056/NEJMc2031364 (2020).
620 10 McCarthy, K. R. *et al.* Natural deletions in the SARS-CoV-2 spike glycoprotein drive antibody
621 escape. *bioRxiv*, 2020.2011.2019.389916, doi:10.1101/2020.11.19.389916 (2020).
622 11 Avanzato, V. A. *et al.* Case Study: Prolonged infectious SARS-CoV-2 shedding from an
623 asymptomatic immunocompromised cancer patient. *Cell* (2020).
624 12 Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals
625 Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310.e1220,
626 doi:10.1016/j.cell.2020.08.012 (2020).
627 13 Rambaut A., L. N., Pybus O, Barclay W, Carabelli A. C., Connor T., Peacock T., Robertson D. L.,
628 Volz E., on behalf of COVID-19 Genomics Consortium UK (CoG-UK). *Preliminary genomic*
629 *characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike*
630 *mutations*, <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-
631 sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563> (2020).
632 14 Schmidt, F. *et al.* Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and
633 chimeric viruses. 2020.2006.2008.140871, doi:10.1101/2020.06.08.140871 %J bioRxiv (2020).
634 15 Brouwer, P. J. M. *et al.* Potent neutralizing antibodies from COVID-19 patients define multiple
635 targets of vulnerability. *Science* **369**, 643-650, doi:10.1126/science.abc5902 (2020).
636 16 Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent individuals.
637 *Nature* **584**, 437-442, doi:10.1038/s41586-020-2456-9 (2020).
638 17 Barnes, C. O. *et al.* Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common
639 Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828-842 e816,
640 doi:10.1016/j.cell.2020.06.025 (2020).
641 18 Shrock, E. *et al.* Viral epitope profiling of COVID-19 patients reveals cross-reactivity and
642 correlates of severity. *Science*, doi:10.1126/science.abd4250 (2020).
643 19 Kemp, S. *et al.* Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion
644 ΔH69/V70. *bioRxiv*, 2020.2012.2014.422555, doi:10.1101/2020.12.14.422555 (2020).
645 20 Sobel Leonard, A. *et al.* The effective rate of influenza reassortment is limited during human
646 infection. *PLoS pathogens* **13**, e1006203, doi:10.1371/journal.ppat.1006203 (2017).
647 21 Richard, M., Herfst, S., Tao, H., Jacobs, N. T. & Lowen, A. C. Influenza A Virus Reassortment Is
648 Limited by Anatomical Compartmentalization following Coinfection via Distinct Routes. *Journal*
649 *of virology* **92**, doi:10.1128/JVI.02063-17 (2018).
650 22 CDC. *Discontinuation of Transmission-Based Precautions and Disposition of Patients with COVID-*
651 *19 in Healthcare Settings (Interim Guidance)*, <https://www.cdc.gov/coronavirus/2019-
652 ncov/hcp/disposition-hospitalized-patients.html> (2020).

23 Boshier, F. A. T. *et al.* Remdesivir induced viral RNA and subgenomic RNA suppression, and evolution of viral variants in SARS-CoV-2 infected patients. *medRxiv*, 2020.2011.2018.20230599, doi:10.1101/2020.11.18.20230599 (2020).

24 Simonovich, V. A. *et al.* A Randomized Trial of Convalescent Plasma in Covid-19 Severe Pneumonia. *N Engl J Med*, doi:10.1056/NEJMoa2031304 (2020).

25 Meredith, L. W. *et al.* Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *The Lancet Infectious Diseases* **20**, 1263-1272, doi:10.1016/S1473-3099(20)30562-4 (2020).

26 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).

27 Loman, N., Rowe, W. & Rambaut, A. (v1, 2020).

28 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10-12 (2011).

29 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)* **34**, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).

30 Jordan, M. R. *et al.* Comparison of standard PCR/cloning to single genome sequencing for analysis of HIV-1 populations. *J Virol Methods* **168**, 114-120, doi:10.1016/j.jviromet.2010.04.030 (2010).

31 Palmer, S. *et al.* Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *Journal of clinical microbiology* **43**, 406-413, doi:10.1128/JCM.43.1.406-413.2005 (2005).

32 Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7552-7557 (2008).

33 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).

34 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).

35 Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology* **5**, 1403-1407, doi:10.1038/s41564-020-0770-5 (2020).

36 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).

37 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).

38 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).

39 Illingworth, C. J. SAMFIRE: multi-locus variant calling for time-resolved sequence data. *Bioinformatics* **32**, 2208-2209, doi:10.1093/bioinformatics/btw205 (2016).

40 Lumby, C. K., Zhao, L., Breuer, J. & Illingworth, C. J. A large effective population size for established within-host influenza virus infection. *Elife* **9**, doi:10.7554/eLife.56915 (2020).

41 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).

42 Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* **11**, e1004041 (2015).

43 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).

44 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).

45 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).

46 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).

47 Vermeire, J. *et al.* Quantification of reverse transcriptase activity by real-time PCR as a fast and accurate method for titration of HIV, lenti- and retroviral vectors. *PloS one* **7**, e50859-e50859, doi:10.1371/journal.pone.0050859 (2012).

48 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).

49 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: Clinical time line of events with longitudinal respiratory sample CT values. CT- cycle threshold

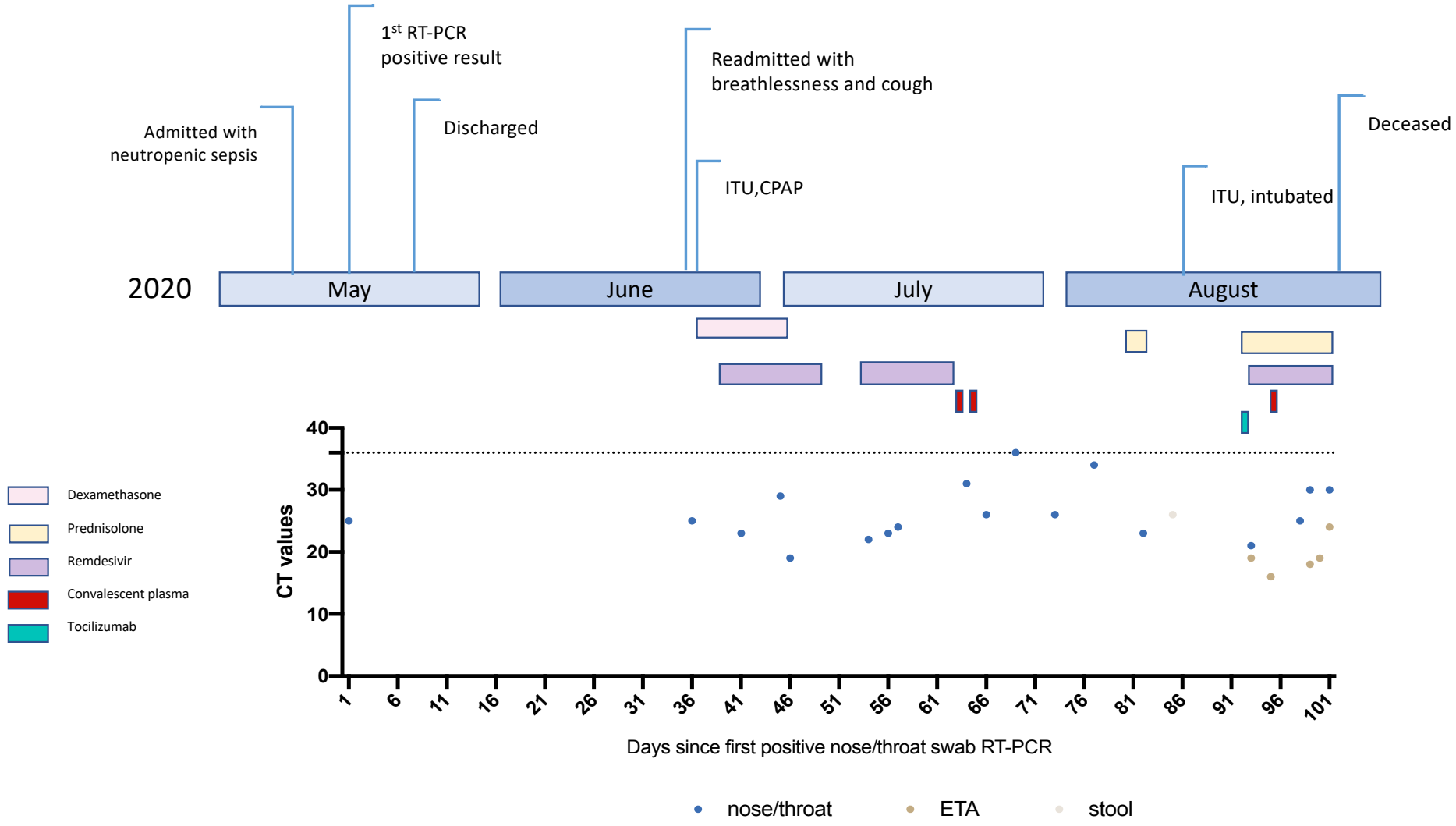


Figure 2

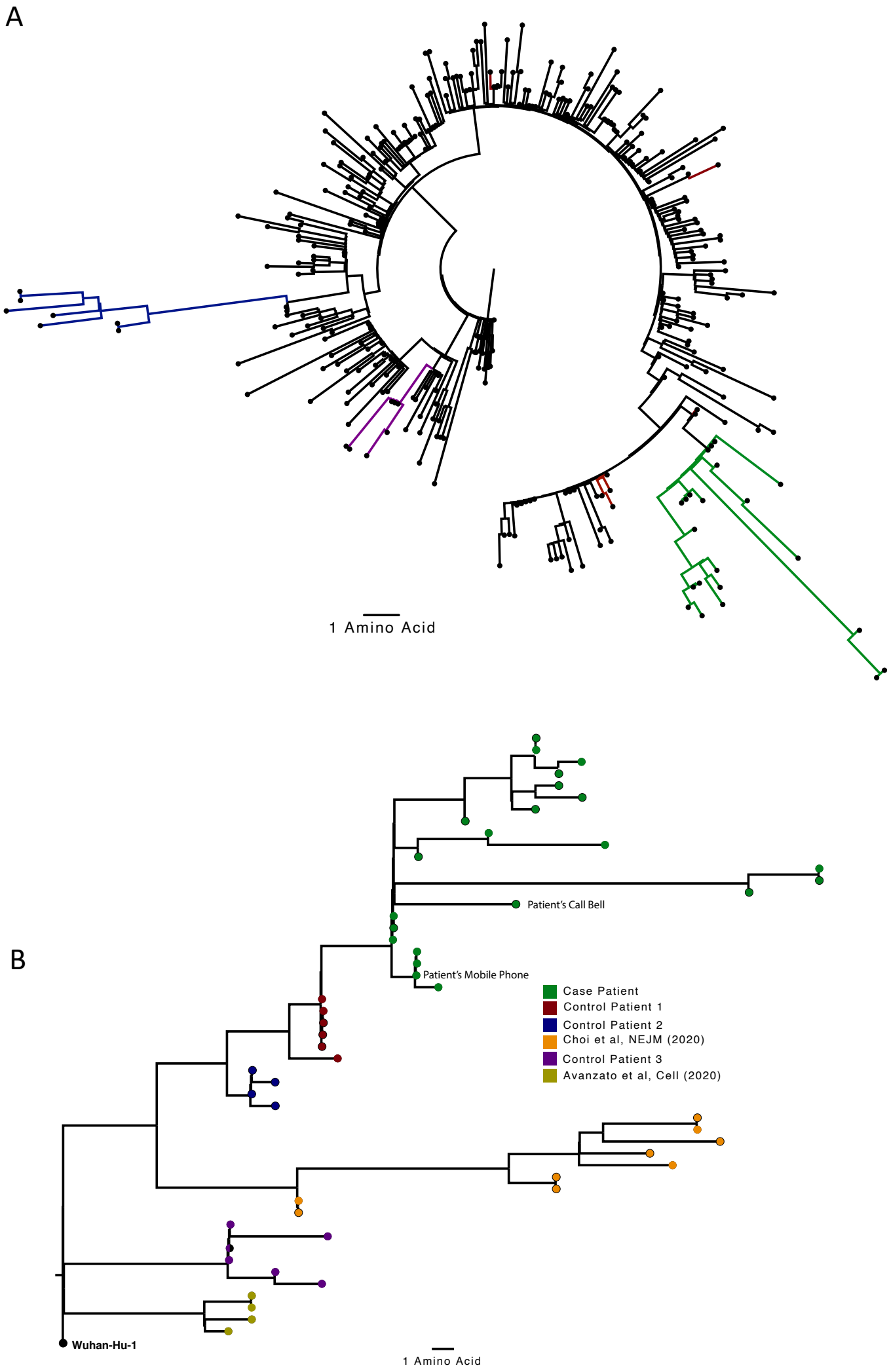


Figure 2. Analysis of 23 Patient derived whole SARS-CoV-2 genome sequences in context of national sequences and other cases of chronic SARS-CoV-2 shedding. A. Circularised maximum-likelihood phylogenetic tree rooted on the Wuhan-Hu-1 reference sequence, showing a subset of 250 local SARS-CoV-2 genomes 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 "United Kingdom / English" SARS-CoV-2 genomes were downloaded from the GISAID database and a random subset of 250 selected as background. **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 (patient's call bell, and patient's mobile phone) are indicated.

Figure 3

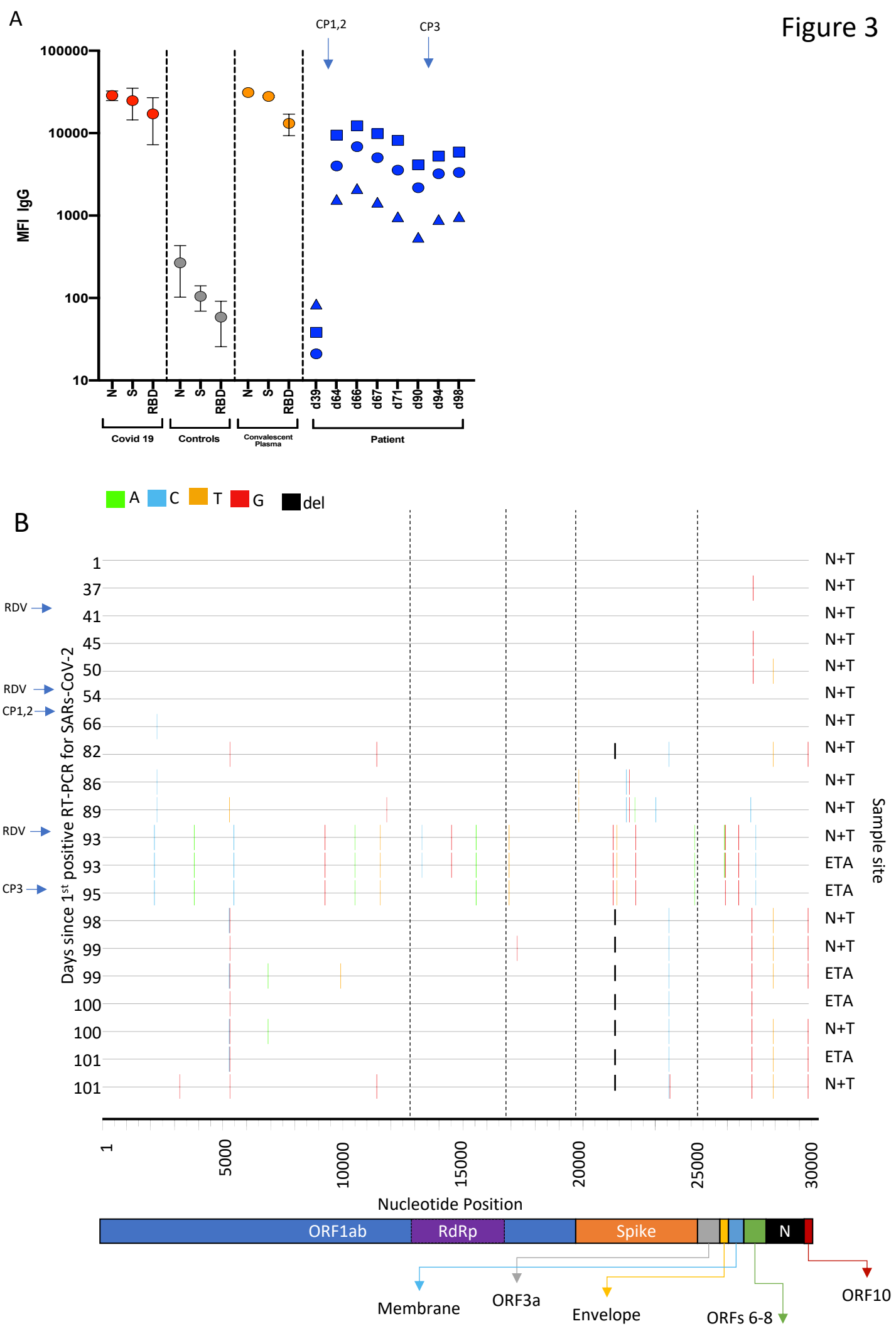
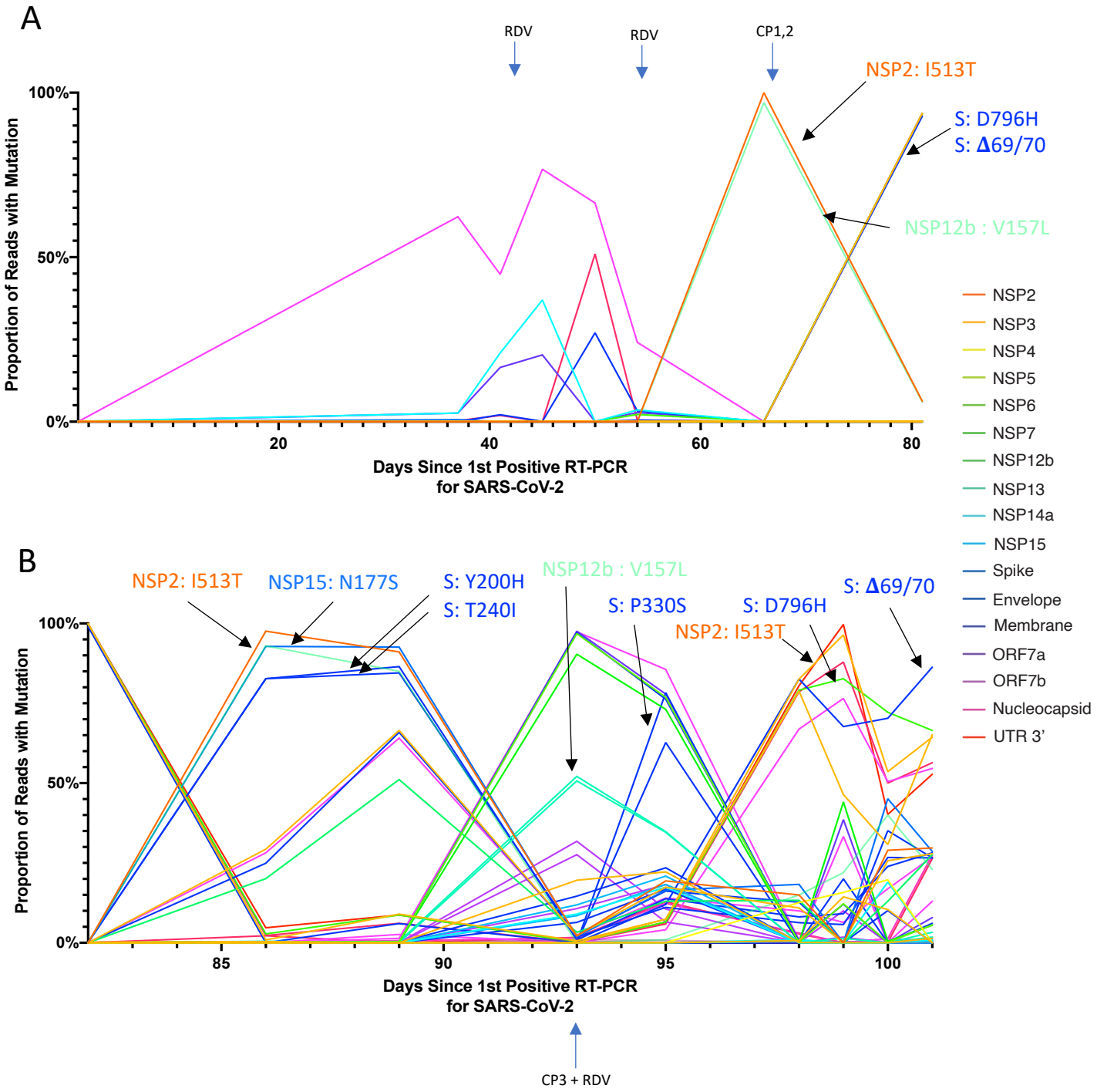


Figure 3. Serum SARS-CoV-2 antibody levels and virus population changes in chronic SARS-CoV-2 infection. A. Anti SARS-CoV2 IgG antibodies in patient and pre/post convalescent plasma compared to RNA+ Covid19 patients and prepandemic healthy controls: Red, grey and gold: IgG antibodies to SARS-CoV2 nucleocapsid protein (N), trimeric S protein (S) and the receptor binding domain (RBD) were measured by multiplexed particle based flow cytometry (Luminex) in RNA+ Covid 19 patients (N=20, red dots), Pre-pandemic healthy controls (N=20, grey dots) and in the convalescent donor plasma (orange dots); Results are shown as mean fluorescent intensity (MFI) +/- SD. **Patient sera over time in blue:** Anti SARS-CoV2 IgG to N (blue squares), S (blue circles) and RBD (blue triangles). Timing of CP units is also shown **B. 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. There were few nucleotide substitutions between days 1-54, 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 amino acid deletion in S1, ΔH69/V70 is indicated by the black lines. Sites: Endotracheal aspirate (ETA) or Nose/throat swabs (N+T).

Figure 4



C

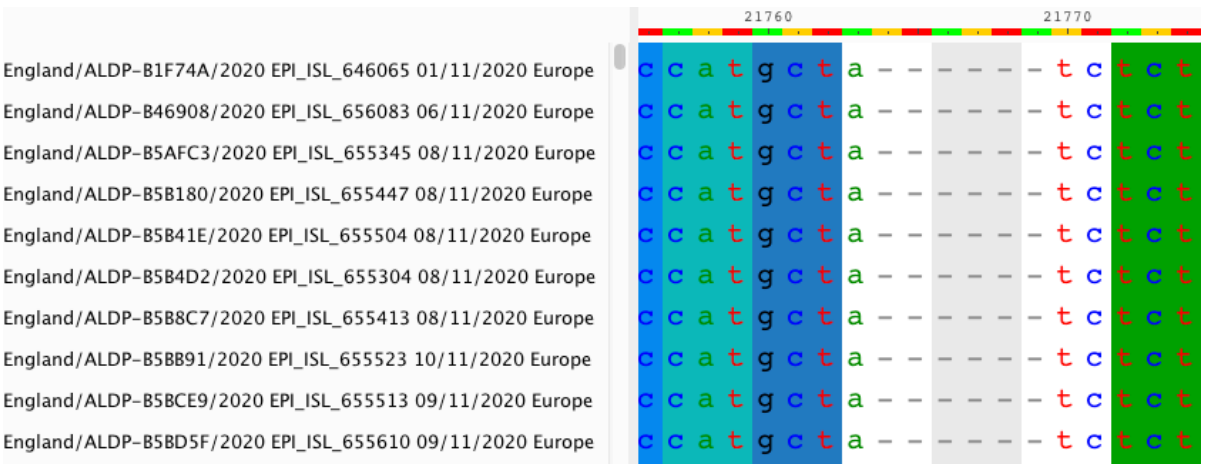
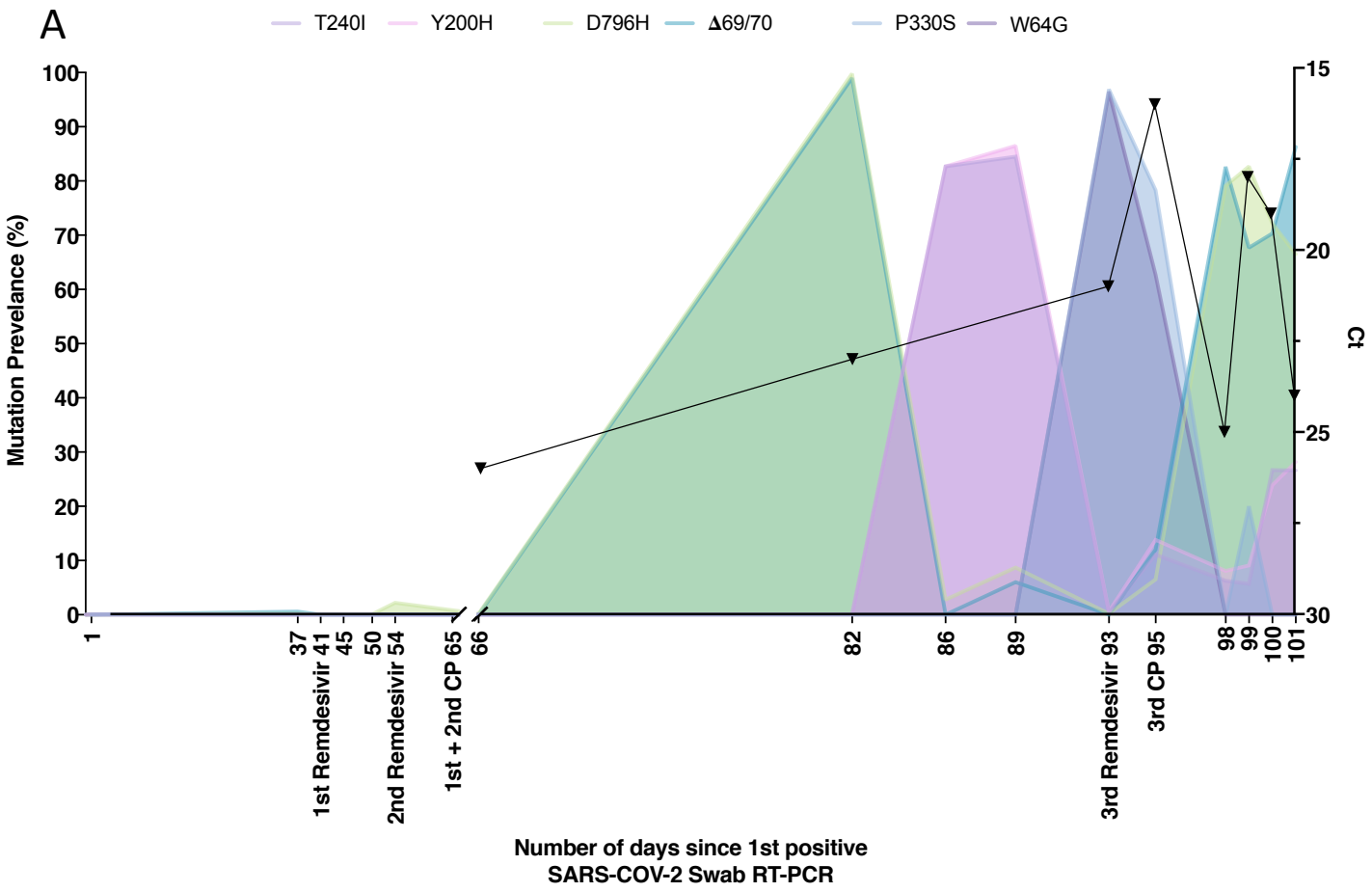


Figure 4. Whole genome variant trajectories showing amino acids and relationship to treatments. Data based on Illumina short read ultra deep sequencing at 1000x coverage. Variants shown reached a frequency of at least 10% in at least 2 samples. Treatments indicated are convalescent plasma (CP) and Remdesivir (RDV). Variants described in the text are designated by labels using the same colouring as the position in the genome. **A.** Variants detected in the patient from days 1-81. **B.** Variants detected in the patient from day 81-101. **C.** Multiple sequence alignment of 10 randomly selected SARS-CoV-2 Genomes containing the **Δ**69/70 deletion. In all cases, six nucleotides are deleted that are out of frame.

Figure 5



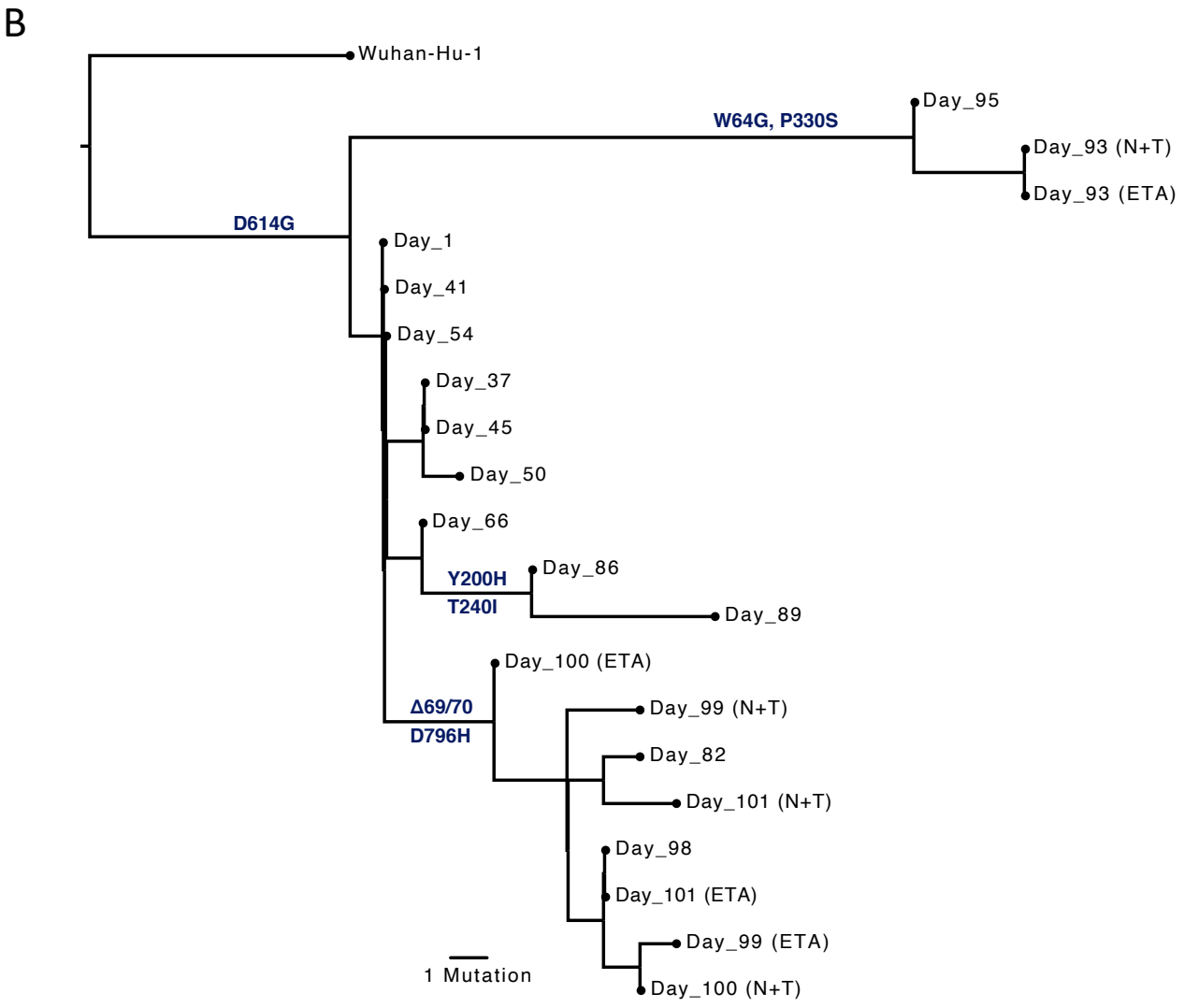


Figure 5. Longitudinal variant frequencies and phylogenetic relationships for virus populations bearing six Spike (S) mutations A. At baseline, all four S variants (Illumina sequencing) were absent (<1% and <20 reads). Approximately two weeks after receiving two units of convalescent plasma (CP), viral 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 6 days. These viral populations were then replaced by virus carrying W64G and P330S mutations in Spike, which both dominated 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. CT values from respiratory samples are indicated on the right y-axis (black triangles) **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. On dates where multiple samples were collect, these are indicated as endotracheal aspirate (ETA) and Nose + throat swabs (N+T).

Figure 6

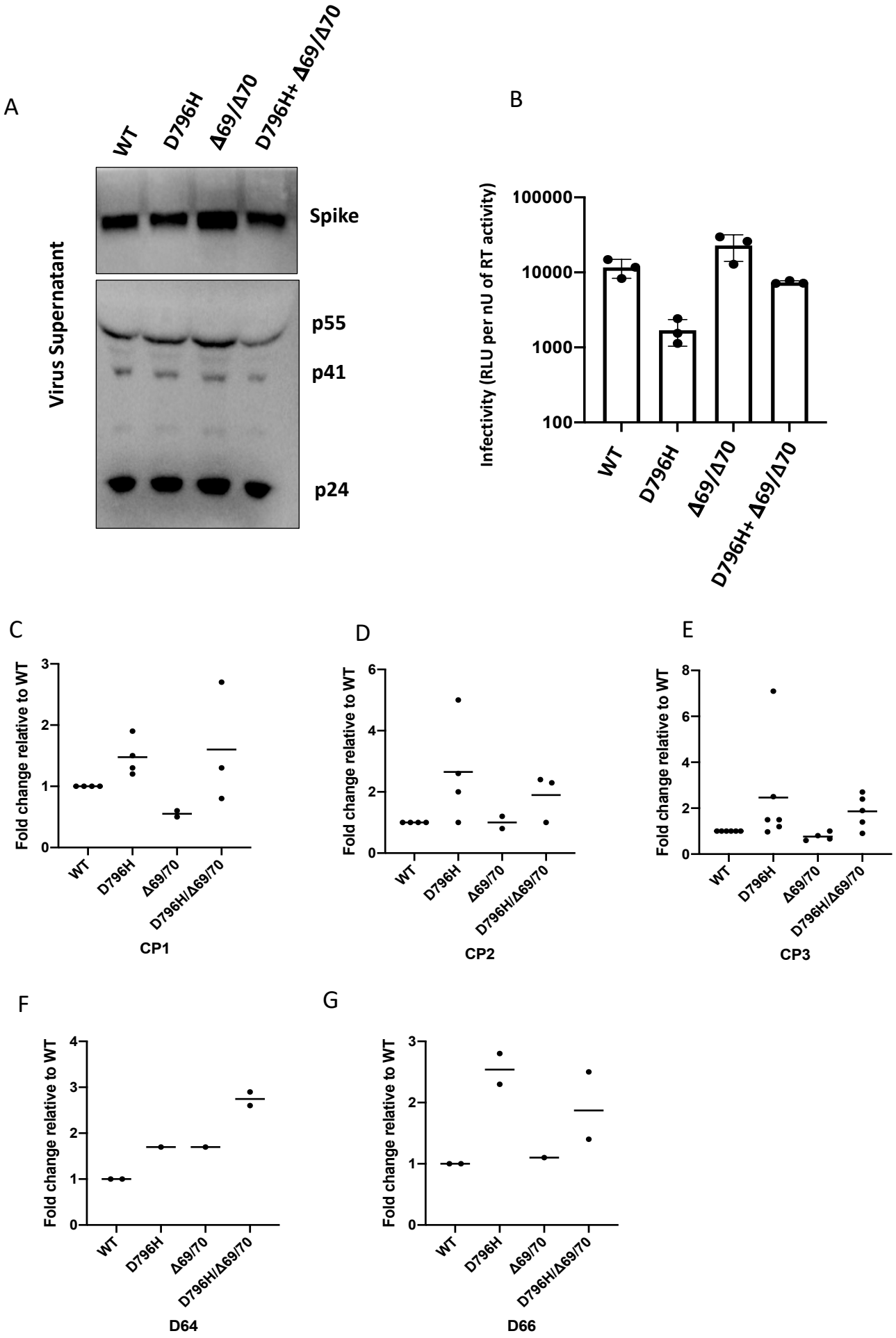


Figure 6: Spike mutant D796H + ΔH69/V70 has infectivity comparable to wild type but is less sensitive to multiple units of convalescent plasma (CP). A. western blot of virus pellets after centrifugation of supernatants from cells transfected with lentiviral pseudotyping plasmids including Spike protein **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. Infectivity is corrected for reverse transcriptase activity in virus supernatant as measured by real time PCR **C-E.** convalescent plasma (CP units 1-3) neutralization potency. against pseudovirus virus bearing Spike mutants D796H, ΔH69/V70 and D796H + ΔH69/V70 F, G patient serum neutralisation potency against pseudovirus virus bearing Spike mutants D796H, ΔH69/V70 and D796H + ΔH69/V70. Patient serum was taken at indicated Day (D). Indicated is serum dilution required to inhibit 50% of virus infection (ID50). Data points represent means of technical replicates and each data point is an independent exp.

Figure 7

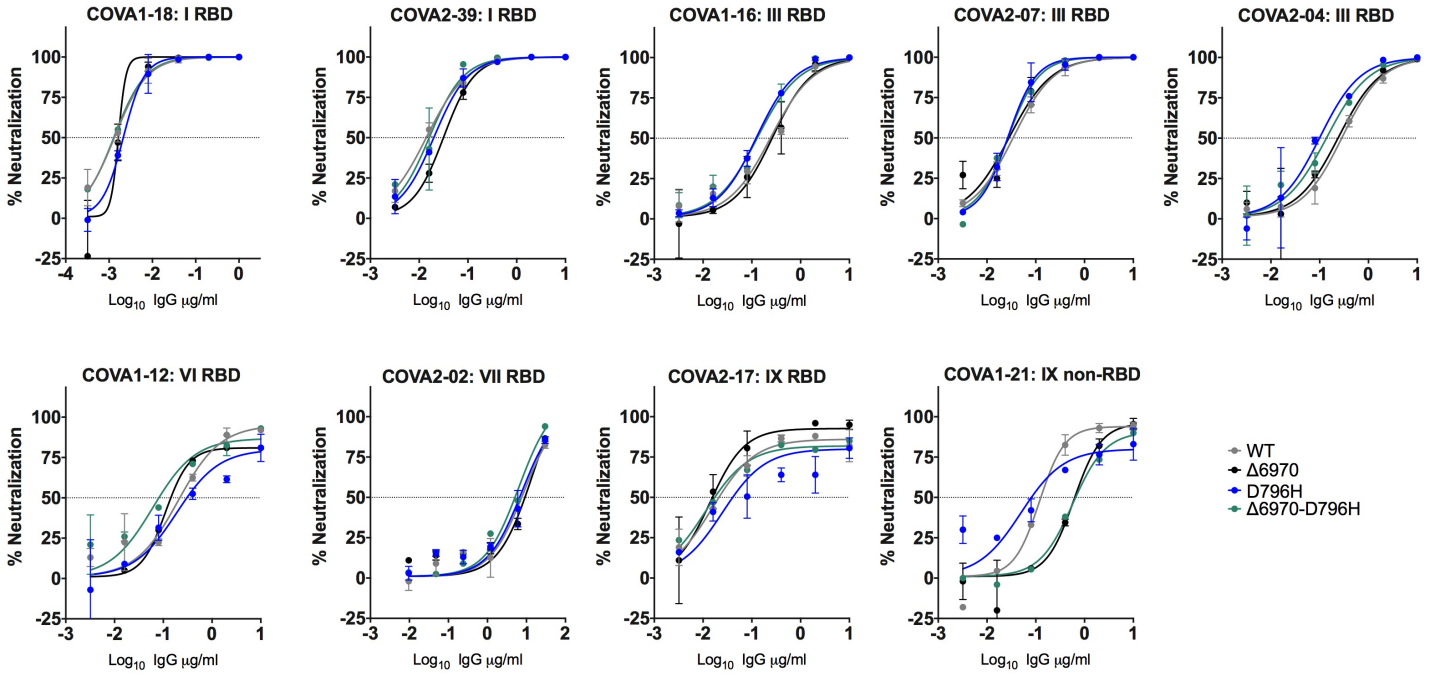


Figure 7: Neutralization potency of a panel of monoclonal antibodies targeting the RBD is not impacted by Spike mutations. Lentivirus pseudotyped with SARS-CoV-2 Spike protein: WT (D614G background), D796H, ΔH69/V70, D796H+ΔH69/V70 were produced in 293T cells and used to infect target Hela cells stably expressing ACE2 in the presence of serial dilutions of indicated monoclonal antibodies. Data are representative of at least two independent experiments. RBD: receptor binding domain. Classes of RBD binding antibodies are indicated based Bouwer et al.

Figure 8

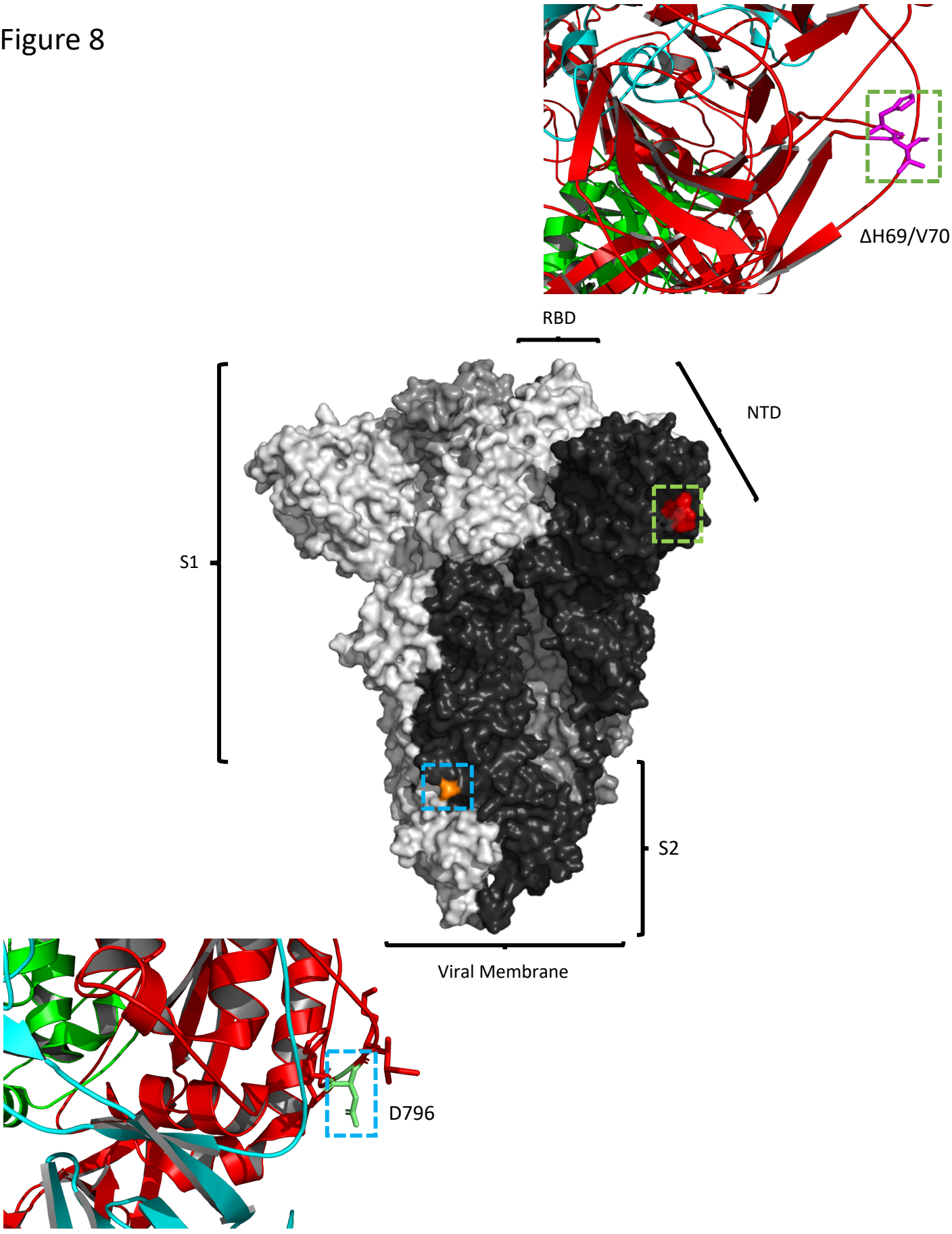


Figure 8. Location of Spike mutations ΔH69/Y70 in S1 and D796H in S2. Amino acid residues H69 and Y70 deleted in the N-terminal domain (red) and D796H in subunit 2 (orange) are highlighted on a SARS-CoV-2 spike trimer (PDB: 6ZGE Wrobek et al., 2020). Each of the three protomers making up the Spike homotrimer 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.

Component	Reference range	Day 49	Day 89
Lymphocyte count	1.00 – 2.80 X 10^9/l	0.16	1.97
CD3 %	%	49	22
CD3 total	0.70 – 2.10 X 10^9/l	0.08	0.43
CD4 %	%	29	11
CD4 total	0.30 – 1.40 X 10^9/l	0.05	0.22
CD8 %	%	21	10
CD8 total	0.20 – 0.90 X 10^9/l	0.03	0.2
CD19 %	%	21	58
CD19 total	0.10 – 0.50 X 10^9/l	0.03	1.14
CD56 %	%	23	19
CD56 total	0.09 – 0.60 X 10^9/l	0.04	0.37
IgG	6.00 – 16.00 g/l	<1.40	1.9
IgA	0.80 – 4.00 g/l	<0.30	<0.30
IgM	0.50 – 2.00 g/l	1.27	1.92
Sars-CoV-2 Total antibody	(Siemens)	Negative	Positive

Supplemental Table 1: Immunological parameters of patient pre- and post- receipt of convalescent plasma (CP) demonstrating profound lymphopenia with reduction in total serum IgG and IgA. The patient continued to have normal serum IgM which either reflects his underlying lymphoproliferative state or ongoing infection. Following receipt of CP the presence of SARS-CoV-2 antibodies is evident.

Supplementary Table 2. Characteristics of genomes used in the phylogenetic analysis in Figure 2B. Pangolin lineages were identified using Pangolin COVID-19 Lineage Assigner v2.0.8 (https://pangolin.cog-uk.io/).. CP- convalescent plasma; mAb – monoclonal antibody.

Patient	Underlying Clinical Diagnosis	Pangolin Lineage	Treatments
Global GSAID Average	-	-	NA
Choi et al	Antiphospholipid syndrome	B.1	Dual mAb, 3 x remdesivir
Avanzato et al	Chronic lymphocytic leukemia and HLA	A.1	2x CP units
Control patient 1 (purple)	X-linked Agammaglobulinemia	B.2.6	2 x remdesevir, CP after virus clearance
Control Patient 2 (blue)	Chronic renal impairment	B.1.1	None
Control Patient 3 (red)	Heart disease	B.1.1.35	None
Study patient	Marginal B Cell Lymphoma	B.1.1.35/1.1.1	3x CP units, 3 x remdesivir

Timepoint (Day)	sequence	CT value	Sample type	Nanopore	Illumina	SGA
1	NB16_CAMB-1B19D5	25	NT	Y	Y	Y
37	NB23_CAMB-1B4FB2	NA	NT	Y	Y	Y
41	NB01_CAMB-1B5124	23	NT	Y	Y	-
45	NB08_CAMB-1B529A	29	NT	Y	Y	-
50	NB13_CAMB-1B54A3	NA	NT	Y	Y	-
54	NB10_CAMB-1B5643	22	NT	Y	-	-
55	NB08_CAMB-1B5467	NA	NT	Y	Y	Y
56	NB11_CAMB-1B5616	23	NT	Y	-	-
57	NB14_CAMB-1B5607	24	NT	Y	-	-
66	NB05_CAMB-1B55FB	26	NT	Y	Y	-
82	NB06_CAMB-1B55CE	23	NT	Y	Y	-
89	NB01_CAMB-1B5ADE	NA	NT	Y	Y	-
93	NB16_CAMB-1B5BAE	21	NT	Y	Y	-
93	NB17_CAMB-1B5B71	19	ETA	Y	Y	-
95	NB02_CAMB-1B5BCC	16	ETA	Y	Y	-
98	NB01_CAMB-1B5BDB	25	NT	Y	Y	Y
99	NB09_CAMB-1B5C50	30	NT	Y	Y	-
99	NB10_CAMB-1B5C6F	18	ETA	Y	Y	-
100	NB13_CAMB-1B5CE7	NA	NT	Y	Y	-
100	NB14_CAMB-1B5CBA	19	ETA	Y	Y	-
101	NB15_CAMB-1B5CC9	24	ETA	Y	Y	-
101	NB16_CAMB-1B5C9C	30	NT	Y	Y	-

Supplementary table 3 : Samples and sequencing methods. Timepoint indicates the day since 1st positive qPCR for SARS-COV-2 that the sample was taken. CT values are reported where available. Y- Yes, - Not done, NA not available

Supplementary Table 4. Prevalence of selected Spike glycoprotein mutations at sequential time points and sequencing depth (number of reads covering the amino acid position) measured by both short-read (Illumina) and long-read (Oxford Nanopore) methods. There was low coverage of S:P330S and S:D796H in the final three timepoints, measured by both short- and long-read methods.

				Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore
			Day	1	1	37	37	41	41	45	45	50	50	54	54	66	66
Locus	Position	From	To														
2343	NSP2	T	C	0.0%	0.2%	0.0%	0.1%	0.0%	0.2%	0.0%	0.6%	0.0%	0.2%	0.0%	0.9%	100.0%	0.0%
3057	NSP3	A	G	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%
5393	NSP3	T	C	0.0%	1.4%	0.2%	0.7%	0.0%	1.1%	0.0%	0.0%	0.2%	1.0%	0.0%	0.9%	0.0%	1.0%
5406	NSP3	T	G	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.1%	0.0%
5425	NSP3	G	T	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	0.0%	0.0%
8389	NSP3	C	T	0.0%	0.9%	0.0%	0.4%	0.1%	0.5%	0.0%	0.0%	0.0%	0.9%	0.0%	0.4%	0.0%	0.5%
9130	NSP4	G	T	0.0%	0.1%	0.0%	0.1%	0.0%	0.2%	0.0%	0.0%	0.0%	0.2%	0.0%	0.3%	0.0%	0.4%
10097	NSP5	A	G	30.8%	45.7%	43.6%	56.1%	17.4%	65.1%	23.3%	66.7%	34.0%	46.4%	19.8%	67.8%	18.1%	77.4%
10700	NSP5	G	A	0.0%	0.5%	0.0%	0.7%	0.1%	0.7%	0.0%	0.0%	0.0%	0.7%	0.1%	0.9%	0.0%	0.8%
11770	NSP6	A	G	0.0%	0.7%	0.0%	0.4%	0.1%	0.2%	0.0%	2.2%	0.1%	0.7%	0.0%	0.2%	0.0%	0.4%
12043	NSP7	C	T	0.0%	0.7%	0.1%	0.5%	0.0%	0.7%	0.0%	0.0%	0.1%	0.5%	0.0%	0.9%	0.0%	0.7%
13527	NSP12b	T	C	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	0.0%	0.1%
14776	NSP12b	G	T	0.0%	0.5%	0.0%	0.3%	0.0%	0.2%	0.1%	0.0%	0.2%	0.1%	0.0%	0.1%	0.1%	0.1%
16733	NSP13	C	T	0.1%	0.2%	2.6%	1.7%	20.9%	13.2%	37.0%	29.8%	0.0%	0.3%	3.6%	1.3%	0.0%	2.0%
17320	NSP13	G	T	0.1%	0.3%	0.3%	0.1%	0.1%	0.2%	0.0%	0.7%	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%
17703	NSP13	C	T	0.0%	0.6%	0.1%	0.8%	0.0%	1.3%	0.0%	1.9%	0.0%	1.5%	0.0%	1.0%	0.0%	1.0%
18488	NSP14a	T	C	0.0%	0.3%	0.0%	0.2%	0.1%	0.4%	0.0%	0.0%	0.0%	0.3%	0.0%	0.2%	0.0%	0.3%
19388	NSP14a	G	A	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	3.1%	0.0%	0.4%	0.0%	0.2%	0.0%	0.5%
20150	NSP15	A	G	0.0%	0.0%	0.0%	0.5%	0.0%	0.2%	0.0%	0.0%	0.2%	0.2%	0.0%	0.8%	0.0%	0.2%
21600	S	G	T	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%
21635	S	C	A	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	0.0%	0.0%	0.0%	1.4%	0.0%	0.0%	0.2%	0.8%
21753	S	T	G	0.0%	0.4%	0.0%	0.8%	0.0%	0.0%	0.0%	0.7%	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%
21768	S		-	0.0%	8.4%	0.6%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	4.4%

22088	S	C	T	0.1%	1.0%	0.0%	0.3%	0.0%	0.4%	0.0%	0.0%	0.0%	0.6%	0.5%	0.7%	0.0%	0.9%
22160	S	T	C	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.2%
22281	S	C	T	0.1%	0.2%	0.0%	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%	0.4%	0.0%	0.1%
22551	S	C	T	0.0%	1.1%	0.0%	0.8%	0.0%	0.0%	0.0%	0.7%	0.0%	0.0%	0.1%	1.6%	0.0%	9.0%
23398	S	T	C	0.0%	0.3%	0.0%	0.2%	0.2%	0.1%	0.0%	0.7%	0.0%	0.0%	0.1%	0.1%	0.1%	0.2%
23948	S	G	C	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	0.0%	2.2%	3.2%	0.2%	1.9%
24257	S	G	T	0.0%	0.0%	0.0%	0.0%	2.1%	2.3%	0.1%	2.7%	27.0%	25.7%	3.3%	0.6%	0.0%	2.6%
25537	E	G	A	0.0%	0.3%	2.6%	2.8%	16.5%	15.4%	20.3%	20.8%	0.1%	0.2%	2.8%	1.0%	0.0%	2.4%
26333	E	C	T	0.0%	0.6%	0.0%	0.1%	0.0%	0.4%	0.0%	0.0%	0.0%	0.2%	0.2%	0.5%	0.0%	0.3%
26634	M	G	T	0.1%	0.4%	0.0%	0.2%	0.0%	0.2%	0.0%	0.8%	0.0%	0.3%	0.0%	0.2%	0.0%	0.2%
26647	M	G	A	0.0%	0.0%	0.0%	0.2%	0.0%	0.2%	0.0%	0.9%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%
27087	M	G	A	0.0%	0.6%	0.0%	0.3%	0.1%	0.1%	0.0%	0.0%	0.0%	0.1%	0.0%	0.2%	0.0%	0.1%
27408	ORF7a	T	C	0.0%	0.0%	0.1%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%
27459	ORF7a	G	T	0.1%	0.4%	0.1%	0.0%	0.1%	0.1%	0.0%	1.4%	0.1%	0.1%	0.0%	0.2%	0.0%	0.4%
27509	ORF7a	C	T	0.0%	0.0%	62.3%	59.8%	44.8%	41.6%	76.7%	65.6%	66.5%	65.3%	24.1%	26.5%	0.0%	22.1%
27618	ORF7a	T	C	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.8%
28209	ORF7b	G	C	0.0%	0.2%	0.0%	0.0%	0.1%	0.1%	0.0%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
28356	N	A	G	0.0%	0.2%	0.3%	0.6%	1.9%	3.0%	0.0%	0.7%	51.0%	48.1%	0.5%	0.5%	0.0%	0.5%
28748	N	C	T	0.0%	0.4%	0.0%	0.1%	0.1%	0.3%	0.0%	0.0%	0.0%	0.3%	0.2%	0.3%	0.0%	0.2%
29825	3'UTR	G	T	0.0%	1.3%	0.2%	0.3%	0.0%	0.6%	0.0%	0.0%	0.2%	0.4%	0.0%	0.2%	0.0%	0.7%

				Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore
			Day	82	82	86	86	89	89	93	93	93	93	95	95	98	98
Locus	Position	From	To														
2343	NSP2	T	C	0.0%	0.1%	97.6%	2.2%	91.1%	98.5%	0.1%	0.2%	2.1%	95.5%	19.4%	90.7%	15.0%	0.4%
3057	NSP3	A	G	0.0%	0.1%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	0.1%
5393	NSP3	T	C	0.2%	1.3%	0.4%	0.9%	0.0%	1.3%	0.2%	1.2%	0.0%	1.3%	7.4%	1.2%	78.8%	1.2%
5406	NSP3	T	G	0.0%	0.0%	29.3%	0.1%	66.5%	0.0%	0.3%	0.1%	2.3%	15.8%	17.4%	56.2%	10.9%	0.3%
5425	NSP3	G	T	99.9%	0.3%	0.9%	0.0%	9.0%	0.0%	0.0%	97.4%	0.6%	0.5%	7.4%	7.9%	82.6%	0.2%
8389	NSP3	C	T	0.0%	0.7%	0.0%	0.3%	0.0%	0.3%	20.6%	0.5%	19.6%	0.7%	22.2%	0.4%	0.1%	21.9%
9130	NSP4	G	T	0.0%	0.4%	0.0%	0.2%	0.0%	0.3%	0.1%	0.3%	0.0%	0.3%	0.0%	0.3%	12.6%	0.4%
10097	NSP5	A	G	41.8%	74.4%	17.2%	82.3%	9.5%	70.0%	21.5%	49.0%	13.9%	64.3%	12.4%	73.6%	20.9%	73.7%
10700	NSP5	G	A	0.0%	0.4%	0.0%	0.6%	0.3%	0.8%	99.5%	0.6%	96.8%	0.6%	76.9%	0.5%	0.1%	98.8%
11770	NSP6	A	G	0.1%	0.5%	0.3%	0.4%	0.1%	0.4%	97.9%	0.3%	90.4%	0.4%	73.1%	0.3%	0.1%	96.8%
12043	NSP7	C	T	0.0%	1.1%	20.1%	0.4%	51.1%	0.6%	0.2%	0.7%	3.3%	20.7%	13.0%	53.3%	13.3%	0.6%
13527	NSP12b	T	C	0.0%	0.2%	0.3%	0.1%	0.4%	0.0%	57.3%	0.0%	52.1%	0.1%	34.8%	0.1%	0.0%	54.6%
14776	NSP12b	G	T	0.1%	0.2%	0.2%	0.1%	0.0%	0.1%	54.1%	0.3%	50.7%	0.6%	34.6%	0.2%	0.0%	57.7%
16733	NSP13	C	T	0.0%	5.0%	0.0%	2.6%	0.0%	0.2%	0.0%	0.1%	0.0%	0.2%	0.0%	0.2%	0.4%	0.1%
17320	NSP13	G	T	0.0%	0.1%	0.5%	0.0%	0.0%	0.1%	0.0%	0.1%	0.0%	0.5%	0.0%	0.2%	1.0%	0.1%
17703	NSP13	C	T	0.0%	1.4%	0.0%	1.5%	0.0%	0.8%	15.5%	1.4%	9.0%	1.0%	17.4%	1.1%	0.0%	16.1%
18488	NSP14a	T	C	0.0%	0.2%	0.0%	0.2%	0.0%	0.4%	16.2%	0.2%	8.5%	0.3%	18.3%	0.3%	0.0%	15.7%
19388	NSP14a	G	A	0.0%	0.2%	0.0%	0.2%	0.0%	0.2%	17.8%	0.5%	11.9%	0.7%	20.9%	0.0%	0.0%	17.4%
20150	NSP15	A	G	0.0%	0.4%	92.8%	0.6%	92.6%	0.1%	0.6%	0.7%	1.9%	94.9%	16.5%	90.8%	18.3%	0.7%
21600	S	G	T	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%	99.2%	0.0%	97.1%	0.0%	76.3%	0.0%	0.7%	99.1%
21635	S	C	A	0.0%	0.5%	0.0%	0.0%	0.2%	1.3%	27.9%	0.8%	14.6%	0.4%	23.5%	0.0%	0.0%	26.8%
21753	S	T	G	0.0%	0.7%	0.0%	0.8%	0.0%	0.5%	96.5%	80.2%	0.0%	0.0%	62.7%	60.3%	0.0%	0.3%
21768	S		-	99.0%	93.0%	0.0%	6.9%	6.0%	1.0%	0.0%	9.2%	0.0%	0.0%	12.0%	10.4%	82.6%	93.5%
22088	S	C	T	0.1%	0.9%	0.0%	0.9%	0.0%	1.1%	15.6%	1.1%	6.4%	2.8%	16.8%	1.2%	0.0%	6.8%
22160	S	T	C	0.1%	0.0%	82.7%	0.0%	86.5%	0.0%	0.6%	0.1%	0.9%	68.7%	13.8%	85.6%	8.1%	0.6%
22281	S	C	T	0.1%	0.3%	82.7%	0.2%	84.5%	0.1%	0.4%	0.2%	1.8%	54.0%	11.1%	73.3%	6.3%	0.3%
22551	S	C	T	0.0%	0.9%	0.0%	0.9%	0.0%	0.9%	96.9%	92.8%	0.0%	0.0%	78.3%	73.3%	0.0%	1.3%

23398	S	T	C	0.1%	0.0%	24.9%	0.0%	65.9%	0.1%	0.4%	0.1%	2.2%	22.1%	16.2%	62.1%	12.7%	0.4%
23948	S	G	C	99.8%	2.3%	2.8%	4.8%	8.7%	0.0%	0.2%	99.2%	0.2%	2.9%	6.5%	8.7%	79.1%	0.1%
24257	S	G	T	0.0%	0.4%	0.0%	0.3%	0.0%	0.4%	0.0%	0.3%	0.0%	0.1%	0.0%	0.2%	0.0%	0.0%
25537	E	G	A	0.0%	3.5%	0.0%	1.4%	0.0%	0.1%	0.0%	0.1%	0.0%	0.1%	0.0%	0.2%	0.0%	0.1%
26333	E	C	T	0.0%	0.5%	0.0%	0.2%	0.2%	0.2%	99.3%	0.5%	97.6%	0.3%	77.9%	0.1%	0.0%	97.8%
26634	M	G	T	0.1%	0.2%	0.1%	0.2%	0.0%	0.1%	43.0%	0.3%	31.8%	0.2%	10.6%	0.2%	0.0%	40.2%
26647	M	G	A	0.0%	0.1%	0.0%	0.1%	0.0%	0.1%	22.5%	0.1%	27.6%	0.1%	6.5%	0.1%	0.0%	18.5%
27087	M	G	A	0.1%	0.3%	0.0%	0.2%	0.0%	0.1%	16.4%	0.3%	10.7%	0.3%	18.2%	0.2%	0.0%	18.6%
27408	ORF7a	T	C	0.0%	0.1%	28.2%	0.0%	64.1%	0.0%	0.3%	0.1%	2.5%	23.3%	14.0%	49.5%	10.0%	0.3%
27459	ORF7a	G	T	0.2%	0.3%	0.0%	0.7%	2.6%	0.1%	0.1%	0.3%	0.1%	0.0%	4.1%	3.5%	66.9%	0.2%
27509	ORF7a	C	T	0.0%	27.3%	0.0%	10.8%	0.0%	0.6%	0.0%	0.5%	0.4%	0.4%	0.1%	0.1%	0.2%	0.5%
27618	ORF7a	T	C	0.1%	0.0%	0.0%	0.0%	1.4%	0.0%	99.5%	0.3%	97.4%	0.0%	85.6%	0.0%	0.8%	97.5%
28209	ORF7b	G	C	0.0%	0.0%	0.0%	0.0%	0.7%	0.0%	0.2%	0.0%	1.5%	0.0%	12.4%	0.6%	3.1%	0.3%
28356	N	A	G	99.9%	0.6%	2.3%	0.6%	6.1%	0.6%	0.2%	98.3%	0.2%	2.7%	5.9%	6.6%	78.6%	0.6%
28748	N	C	T	0.0%	0.4%	2.2%	0.2%	0.0%	0.3%	0.1%	0.3%	1.7%	2.6%	12.9%	0.1%	2.8%	0.4%
29825	3'UTR	G	T	99.8%	0.5%	4.7%	0.4%	8.7%	0.1%	0.3%	98.5%	0.5%	3.0%	6.4%	9.4%	80.5%	0.7%

				Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore
			Day	99	99	99	99	100	100	100	100	101	101	101	101
Locus	Position	From	To												
2343	NSP2	T	C	18.2%	81.4%	5.4%	67.1%	0.0%	0.4%	2.8%	11.9%	7.8%	3.1%	0.0%	0.6%
3057	NSP3	A	G	0.0%	2.4%	38.4%	14.5%	29.0%	13.1%	32.4%	0.2%	29.7%	28.6%	0.0%	28.2%
5393	NSP3	T	C	14.4%	0.0%	0.0%	0.0%	10.4%	0.1%	0.0%	12.3%	0.0%	0.0%	0.0%	9.5%
5406	NSP3	T	G	46.4%	1.4%	65.3%	7.4%	30.8%	77.0%	65.3%	39.3%	65.2%	65.5%	16.0%	30.8%
5425	NSP3	G	T	0.0%	1.4%	29.8%	10.7%	25.7%	7.0%	29.9%	0.0%	27.9%	19.4%	0.0%	20.2%
8389	NSP3	C	T	96.4%	0.2%	65.6%	5.2%	53.6%	79.5%	65.5%	97.8%	64.4%	60.8%	98.9%	42.4%
9130	NSP4	G	T	0.0%	16.9%	1.3%	21.9%	0.0%	0.3%	0.6%	0.0%	1.4%	1.1%	0.0%	0.7%
10097	NSP5	A	G	15.8%	0.1%	0.0%	0.3%	19.7%	10.9%	0.0%	16.0%	0.0%	0.3%	0.0%	18.9%
10700	NSP5	G	A	1.7%	63.8%	87.6%	73.1%	16.7%	67.0%	17.9%	87.5%	25.6%	9.7%	0.0%	58.9%
11770	NSP6	A	G	12.2%	96.9%	4.4%	79.1%	0.1%	0.8%	3.3%	13.6%	5.6%	4.8%	0.1%	0.7%
12043	NSP7	C	T	44.1%	93.2%	5.2%	78.4%	0.5%	0.2%	3.3%	44.2%	5.5%	4.4%	0.2%	0.4%
13527	NSP12b	T	C	0.0%	2.6%	27.6%	15.1%	13.0%	12.1%	26.0%	0.9%	27.6%	27.9%	18.7%	12.9%
14776	NSP12b	G	T	1.4%	50.2%	1.5%	32.6%	0.0%	0.2%	1.0%	1.3%	3.3%	2.3%	0.0%	0.0%
16733	NSP13	C	T	0.0%	51.4%	2.3%	31.3%	0.1%	0.2%	1.5%	0.0%	1.8%	0.9%	0.0%	0.0%
17320	NSP13	G	T	0.0%	0.1%	0.0%	0.2%	0.0%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
17703	NSP13	C	T	0.0%	0.1%	0.2%	0.1%	19.0%	1.1%	0.0%	0.1%	0.1%	0.2%	14.2%	18.9%
18488	NSP14a	T	C	0.0%	9.7%	0.7%	17.1%	0.0%	1.2%	1.0%	0.0%	0.8%	2.0%	0.0%	1.7%
19388	NSP14a	G	A	0.0%	7.6%	1.0%	16.3%	0.0%	0.4%	0.2%	0.0%	0.6%	1.3%	0.0%	0.2%
20150	NSP15	A	G	0.0%	9.8%	0.9%	20.0%	0.0%	1.0%	1.0%	0.0%	1.6%	1.4%	0.0%	0.0%
21600	S	G	T	0.0%	2.0%	31.6%	16.8%	45.1%	17.1%	35.5%	0.0%	28.6%	31.3%	2.1%	40.9%
21635	S	C	A	12.1%	96.5%	16.9%	78.6%	0.0%	0.4%	18.9%	20.0%	6.1%	0.0%	0.0%	0.0%
21753	S	T	G	0.0%	20.2%	5.3%	23.4%	0.0%	0.7%	12.2%	7.7%	0.0%	0.0%	0.0%	0.0%
21768	S		-	0.0%	3.6%	0.0%	0.0%	0.0%	3.1%	0.0%	0.0%	0.0%	0.7%	0.0%	0.0%
22088	S	C	T	67.7%	90.0%	0.0%	0.0%	70.3%	96.5%	0.0%	0.0%	86.3%	91.0%	0.0%	0.0%
22160	S	T	C	0.0%	3.0%	0.0%	5.9%	0.0%	1.4%	0.0%	0.0%	0.2%	0.0%	0.0%	6.3%
22281	S	C	T	9.1%	1.4%	0.0%	13.9%	23.9%	8.1%	35.7%	0.0%	28.2%	16.7%	0.8%	33.3%
22551	S	C	T	5.7%	1.0%	0.0%	10.7%	26.7%	8.0%	6.1%	0.0%	26.6%	3.0%	0.8%	35.0%

23398	S	T	C	20.0%	0.0%	0.0%	0.0%	0.0%	4.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
23948	S	G	C	0.0%	1.8%	36.9%	13.0%	35.1%	10.0%	30.5%	0.0%	26.6%	28.5%	0.1%	28.1%
24257	S	G	T	82.7%	0.0%	58.0%	6.8%	72.2%	81.2%	71.8%	85.6%	66.5%	70.8%	100.0%	76.7%
25537	E	G	A	0.0%	0.7%	0.0%	0.2%	10.1%	0.3%	0.0%	0.0%	0.1%	0.2%	0.0%	7.4%
26333	E	C	T	0.0%	0.3%	0.0%	0.1%	0.0%	0.3%	0.0%	0.1%	0.0%	0.1%	0.0%	0.0%
26634	M	G	T	38.6%	96.3%	0.9%	76.9%	0.0%	0.1%	3.7%	40.9%	7.9%	7.7%	0.0%	1.0%
26647	M	G	A	1.8%	30.9%	0.2%	9.2%	0.0%	0.4%	0.0%	2.0%	0.6%	0.4%	0.0%	0.0%
27087	M	G	A	0.0%	25.4%	0.2%	5.3%	0.0%	0.1%	0.1%	0.1%	0.0%	0.4%	0.1%	0.0%
27408	ORF7a	T	C	10.0%	10.9%	1.0%	19.4%	0.0%	0.2%	0.8%	7.6%	1.2%	0.9%	0.0%	0.0%
27459	ORF7a	G	T	6.4%	2.0%	29.7%	10.3%	16.9%	7.1%	32.3%	3.6%	26.9%	22.5%	0.0%	14.0%
27509	ORF7a	C	T	76.5%	0.4%	64.1%	6.9%	50.5%	74.7%	59.6%	80.9%	54.6%	65.1%	79.7%	55.9%
27618	ORF7a	T	C	0.1%	0.3%	0.0%	0.5%	1.2%	0.5%	0.1%	0.3%	0.8%	0.3%	0.0%	1.9%
28209	ORF7b	G	C	33.3%	96.4%	34.8%	82.8%	0.0%	0.0%	7.2%	0.0%	13.0%	0.0%	0.0%	0.0%
28356	N	A	G	0.0%	1.8%	25.7%	13.2%	1.3%	2.6%	28.5%	0.1%	27.5%	25.0%	0.0%	0.8%
28748	N	C	T	87.9%	0.6%	55.1%	7.2%	50.0%	78.6%	58.9%	86.5%	56.4%	59.0%	94.6%	45.0%
29825	3'UTR	G	T	0.0%	2.4%	26.1%	13.5%	0.1%	3.3%	27.7%	0.0%	26.4%	30.6%	0.0%	0.5%

Supplementary Table 5: Single genome sequencing (SGS) data from respiratory samples at indicated days. Indicated are the number of single genomes obtained at each time point with the mutations of interest (identified by deep sequencing). *denominator is 19 as for 2 samples the primer reads were poor quality at amino acid 796 at day 98.

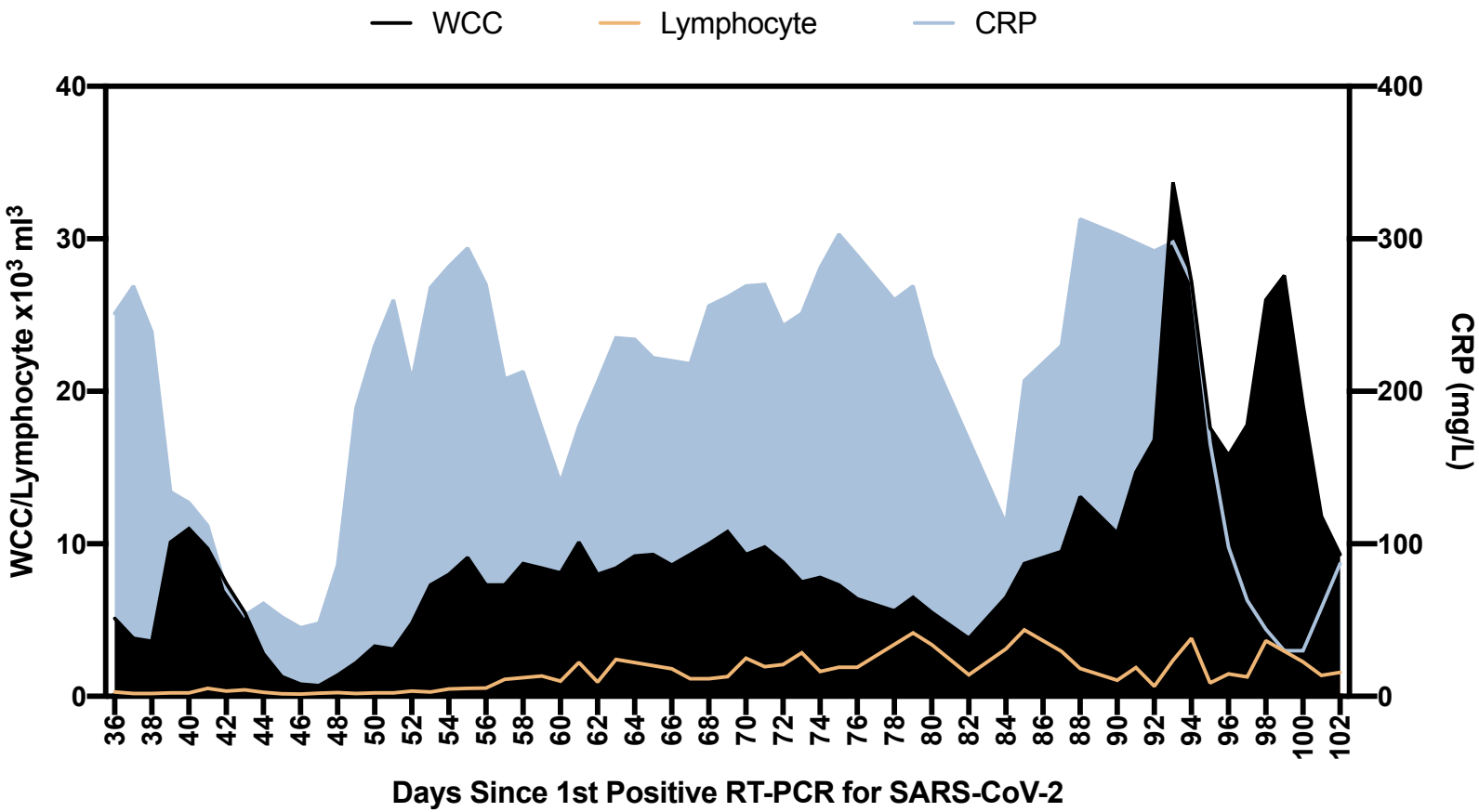
	W64G	P330S	69/70	D796H	T200I	Y240H
Day 1 (n=7)	0	0	0	0	0	0
Day 37 (n=38)	0	0	0	0	0	0
Day 98 (n=21)	1 (4.8%)	1 (4.8%)	17 (81.0%)	13*(68.4%)	3 (14.3%)	3 (14.3%)

Supplementary table 6: Neutralisation of mutants by *Seven RBD-specific mAbs (from Bauwer et al. in Figure 7).* Clusters II, V contain only non-neutralising mAbs, smaller neutralising mAb clusters IV (n=2) and X (n=1) were not tested. Red indicates significant fold changes.

mAb	Cluster	Target		WT	D796H	Δ6970	Δ6970-D796H	Fold decrease		
COVA1-18	I	RBD		0.0014	0.0022	0.0016	0.0013	1.6	1.2	1.0
COVA2-39	I	RBD	Structure	0.0143	0.0203	0.0319	0.0163	1.4	2.2	1.1
COVA1-16	III	RBD	Structure	0.2441	0.1242	0.2651	0.1308	0.5	1.1	0.5
COVA2-07	III	RBD		0.0349	0.0269	0.0288	0.0272	0.8	0.8	0.8
COVA2-04	III	RBD	Structure	0.2887	0.1009	0.2425	0.1401	0.3	0.8	0.5
COVA2-17	IX	RBD		0.0156	0.0248	0.0139	0.0113	1.6	0.9	0.7
COVA2-02	VII	RBD		5.8590	4.9670	6.5680	3.5380	0.8	1.1	0.6
COVA1-12	VI	RBD		0.2007	0.1863	0.1105	0.0611	0.9	0.6	0.3
COVA1-21	XI	Non-RBD		0.1189	0.0498	0.6035	0.5682	0.4	5.1	4.8

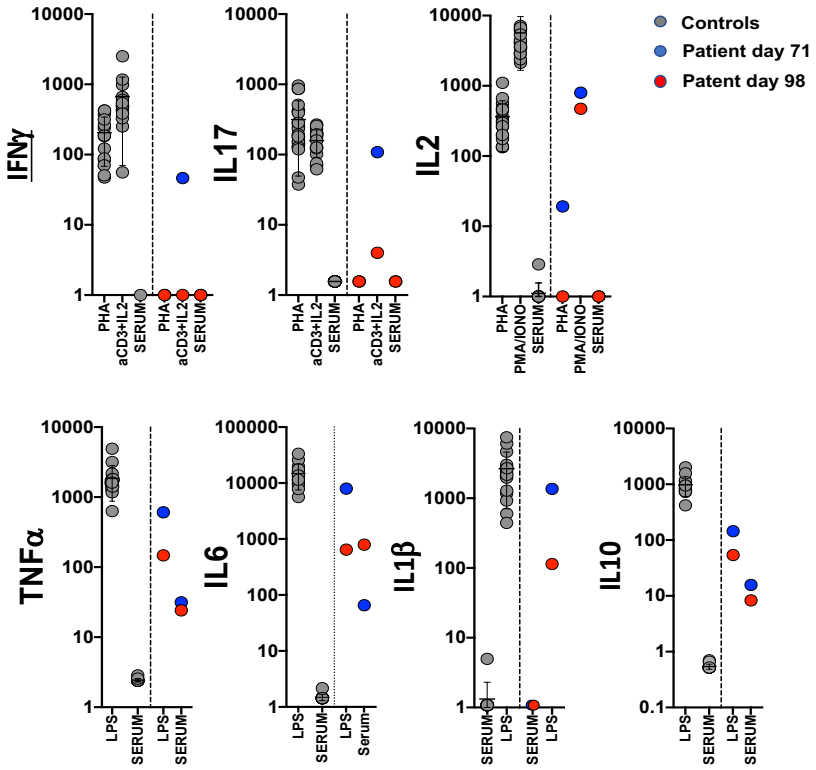
Supplementary Table 7. Global prevalence of selected spike mutations detailed in this paper. All high coverage sequences were downloaded from the GISAID database on 11th November and aligned using MAFFT. The global prevalence of each of the six spike mutations W64G, ΔH69/V70, Y200H, T240I, P330S and D796H were assessed by viewing the multiple sequence alignment in AliView, sorting by the column of interest, and counting the number of mutations.

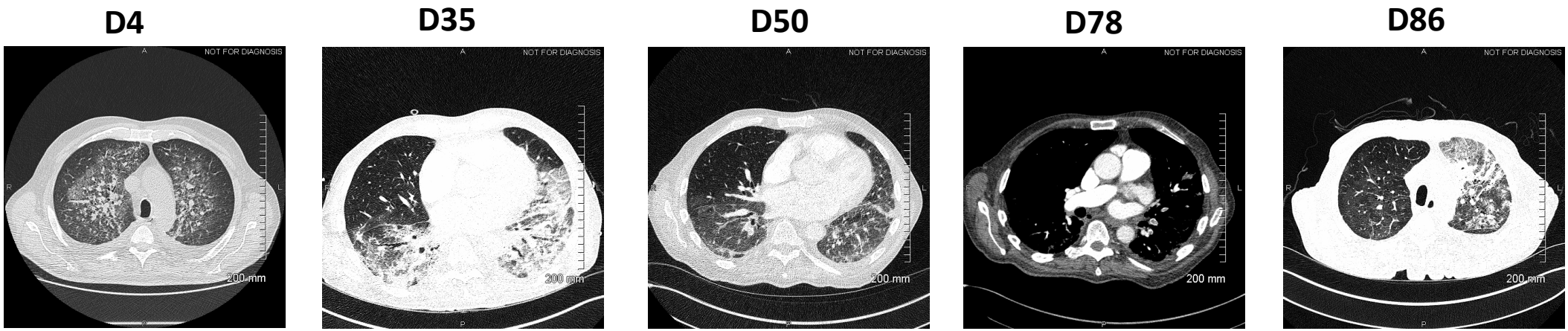
Mutation	Number of Sequences	Global Prevalence
W64G	0	0.00%
ΔH69/V70	2920	1.38%
Y200H	7	0.00%
T240I	34	0.02%
P330S	130	0.06%
D796H	25	0.01%



Supplementary Figure 1: Blood parameters over time in patient case: White cell count (WCC) and lymphocyte counts are expressed as $x10^3$ Cells/mm^3. CRP: C reactive protein

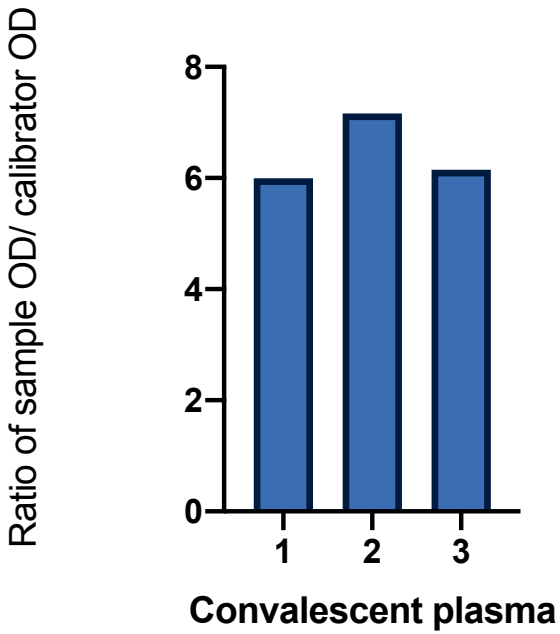
Supplementary Figure 2: Assessment of T cell and innate function. Whole blood cytokines were measured in whole blood after 24 hours stimulation either after T-cell stimulation with PHA or anti CD3/IL2 or innate stimulation with LPS . Healthy controls are shown as grey circles (N=15), Patient at d71 and d98 is shown as blue circles or red circles respectively. Cytokine levels are shown as pg/ml. stimulation.

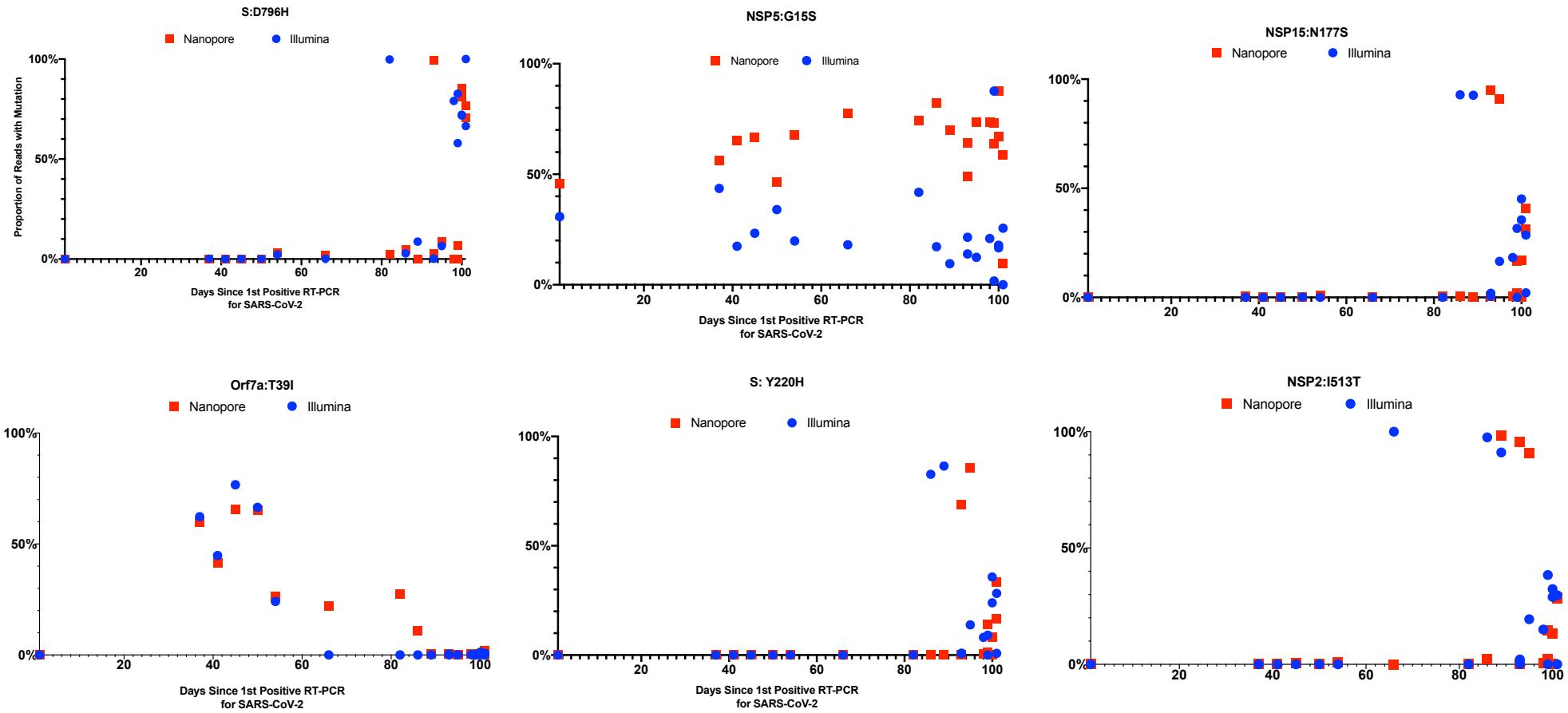




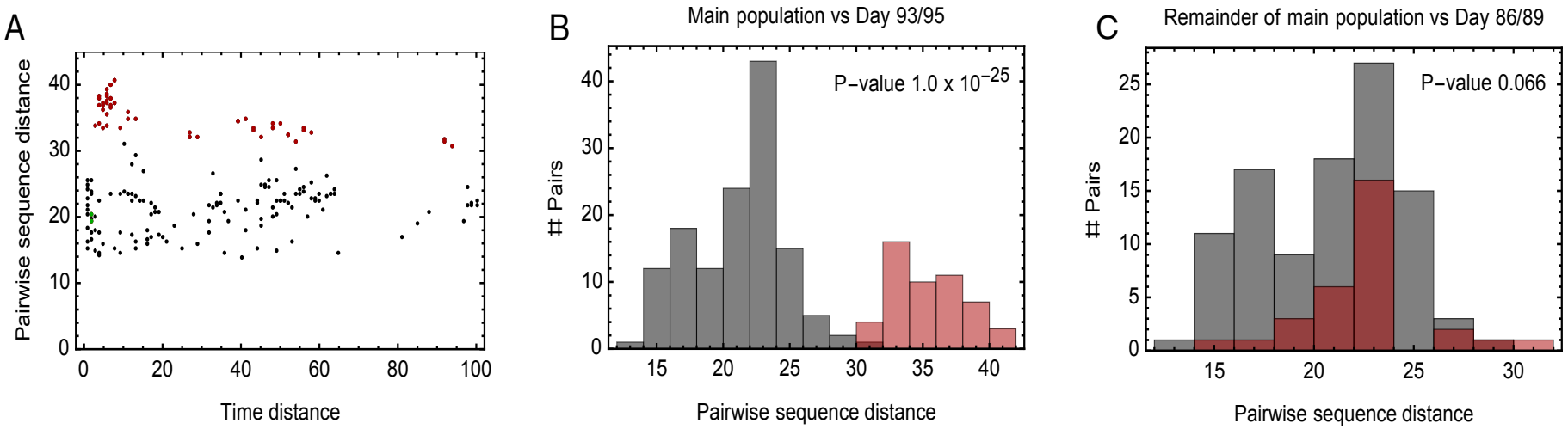
Supplementary Figure 3: Serial CT images following detection of SARS-CoV-2.
The patient initially presented with ground glass and peribroncho-vascular consolidation with associated intralobular septal thickening/reticulation and architectural distortion and interlobular septal thickening. By day 50 there is some improvement with evidence of resolving pnuemonitis, however, his condition deteriorated following the detection of bilateral pulmonary emboli, a well-recognized complication of SARS-CoV-2. Despite multiple therapeutic interventions, the patient's condition deteriorated with worsening of inflammatory changes and chronic organizing pneumonia (COP), particularly on the left, and ongoing changes compatible with persistent SARS-CoV2 infection.

Supplementary Figure 4: SARS-CoV-2 antibody titres in convalescent plasma. Measurement of SARS-CoV-2 specific IgG antibody titres in three units of convalescent plasma (CP) by Euroimmun assay.

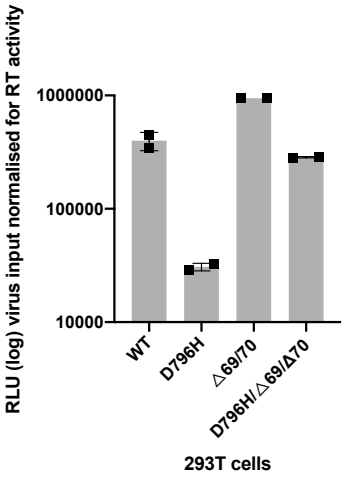


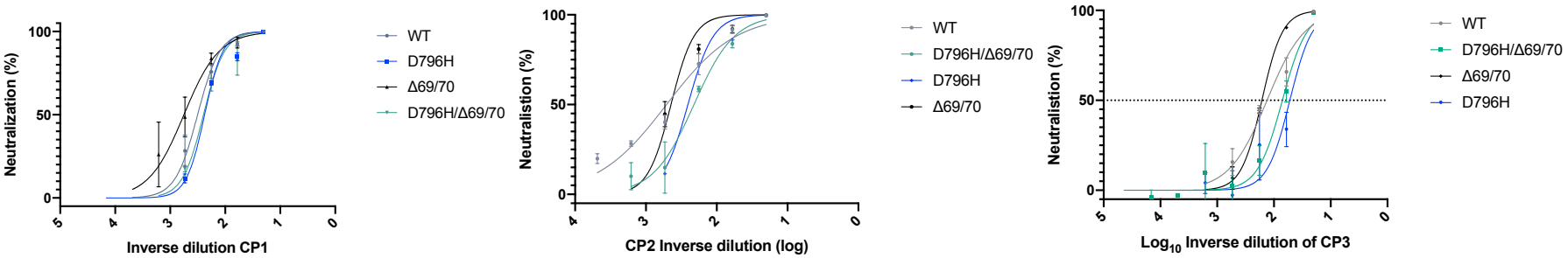


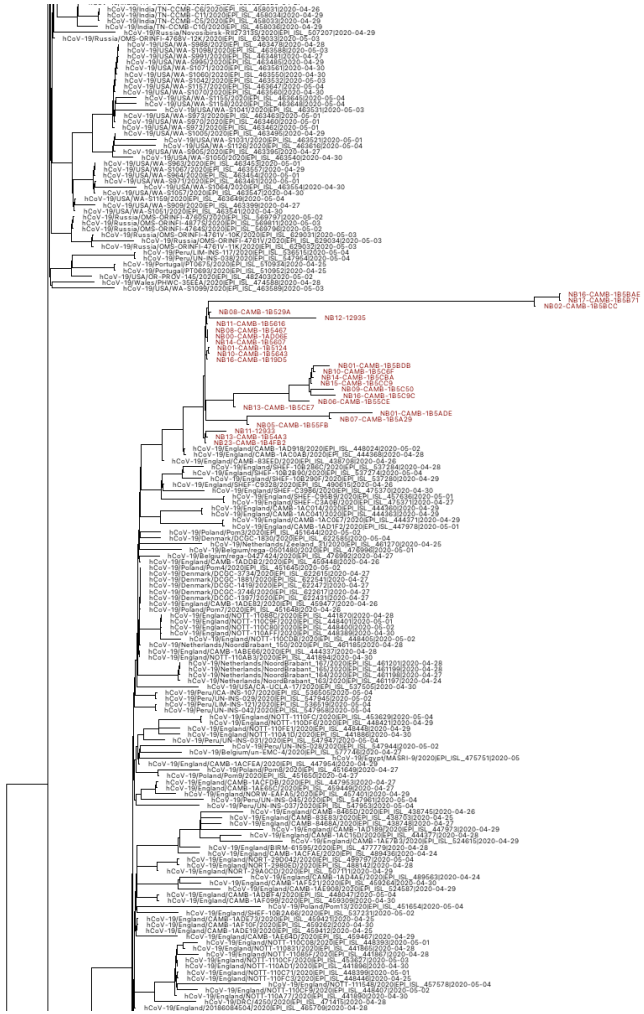
Supplementary Figure 5. Concordance between short-read (Illumina) and long-read single molecule (Oxford Nanopore) sequencing methods for twenty samples. Points are the frequency of the variant at each timepoint; only 20 samples were sequenced using the Illumina method. Boxes represent inter-quartile ranges. Error bars are 95% CI. There was good concordance for mutations between the two methods, and no significant difference between the proportion of reads measured by both methods.



Supplementary Figure 6: Additional evidence for within-host cladal structure. A. Pairwise distances between samples measured using the all-locus distance metric plotted against pairwise distances in time (measured in days) between samples being collected. Internal distances between samples in the proposed main clade are shown in black, distances between samples in the main clade and samples collected on days 93 and 95 are shown in red, and internal distances between samples collected on days 93 and 95 are shown in green. **B.** Pairwise distances between samples in the larger clade (black) and between these samples and those collected on days 93 and 95 (red). The median values of the distributions of these values are significantly different according to a Mann Whitney test. **C.** Pairwise distances between samples in the main clade, once those collected on days 86, 89, 93, 95 have been removed (black) and between these samples and those collected on days 86 and 89 (red). The median values of the distributions of these values are not significantly different at the 5% level according to a Mann Whitney test.





Supplementary Figure 7: In vitro infectivity and neutralisation sensitivity of Spike pseudotyped lentiviruses. Top: infection of target 293T cells expressing TMPRSS2 and ACE2 receptors using equal amounts of virus as determined by reverse transcriptase activity. Bottom: Representative Inverse dilution plots for Spike variants against convalescent plasma units 1-3. Data points represent mean % neutralisation and error bars represent standard error of the mean

Supplementary Figure 8. Maximum-Likelihood global phylogeny of SARS-CoV-2. To ensure that the patient case was not resultant of a superinfection event, all patient sequences were aligned with a snapshot of global SARS-CoV-2 sequences downloaded from the GISAID database between 1st April and 5th May (26472 sequences, only full sequences and excluding all low coverage sequences). Sequences were aligned using MAFFT and a maximum-likelihood tree inferred with IQTREE v2.1.3. All 23 sequenced from the patient case (red) formed a distinct clade, suggesting that all viral populations diversified within-host.