SARS-CoV-2 evolution during treatment of chronic infection

Steven A Kemp[1]*, Dami A Collier [1,2,3]*, Rawlings P Datir*[2,3], Isabella ATM Ferreira [2,3] , Salma Gayed[4], Aminu Jahun[5], Myra Hosmillo[5], Chloe Rees-Spear[1], Petra Mlcochova [2,3], Ines Ushiro Lumb[6], David J Roberts[6], Anita Chandra[2,3], Nigel Temperton[7] , The CITIID-NIHR BioResource COVID-19 Collaboration°, The COVID-19 Genomics UK (COG-UK) Consortium°°, Katherine Sharrocks K[4], Elizabeth Blane[3], Yorgo Modis[8], Kendra Leigh[8], John Briggs[8], Marit van Gils[9], Kenneth GC Smith[2,3], John R Bradley[3,10], Chris Smith[11], Rainer Doffinger[13], Lourdes Ceron-Gutierrez[13], Gabriela Barcenas-Morales[13,14], David D Pollock[15], Richard A Goldstein[1], Anna Smielewska [5,11], Jordan P Skittrall [4,12,16], Theodore Gouliouris [4], Ian G Goodfellow[5], Effrossyni Gkrania-Klotsas[4], Christopher JR Illingworth[12,17], Laura E McCoy[1], Ravindra K Gupta[2,3,18]

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

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

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

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

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

[13] Department of Clinical Biochemistry and Immunology, Addenbrookes Hospital

[14] FES-Cuautitlán, UNAM, Mexico

[15]Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, Colorado, USA;
[16]Clinical Microbiology and Public Health Laboratory, Addenbrookes' Hospital, Cambridge, UK
[17] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
[18]Africa Health Research Institute, Durban, South Africa

°The CITIID-NIHR BioResource COVID-19 Collaboration, see appendix for author list
°°The COVID-19 Genomics UK (COG-UK) Consortium, see appendix for author list

*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

Summary

SARS-CoV-2 Spike protein is critical for virus infection via engagement of ACE2[1], and is a major antibody target. 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 over 23 time points spanning 101 days. Little change was observed in the overall viral population structure following two courses of remdesivir over the

first 57 days. However, following convalescent plasma therapy we observed large, dynamic virus population shifts, with the emergence of a dominant viral strain bearing D796H in S2 and **Δ**H69/**Δ**V70 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 **Δ**H69/**Δ**V70 and D796H conferred modestly 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 **Δ**H69/**Δ**V70 single mutant had two-fold higher infectivity compared to wild type, possibly compensating for the reduced infectivity of D796H. These data reveal strong selection on SARS-CoV-2 during convalescent plasma therapy associated with emergence of viral variants with evidence of reduced susceptibility to neutralising antibodies.

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

A septuagenarian male was admitted to a tertiary hospital in summer of 2020 and had tested positive for SARS-CoV-2 RT-PCR 35 days previously on a nasopharyngeal swab (Day 1) at a local hospital (Extended data 1 and 2). His past medical history was significant for marginal B cell lymphoma diagnosed in 2012, with previous chemotherapy including vincristine, prednisolone, cyclophosphamide and anti-CD20 B cell depletion with rituximab. It is likely that both chemotherapy and underlying lymphoma contributed to B and T cell combined immunodeficiency (Extended data 2 and 3, Supplementary Table 1). Computed tomography (CT) of the chest showed widespread abnormalities consistent with COVID-19 pneumonia (Supplementary Figure 1). Treatment included two 10-day courses of remdesivir with a five day gap in between (Extended data 1). Two units of convalescent plasma were administered on days 63 and 65 (Extended data 3). Following clinical deterioration, remdesivir and a unit of convalescent plasma were administered on day 95, but the individual unfortunately died on day 102 (Supplementary text).

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 (Supplementary Table 3). 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 4). There was general agreement between the two methods (Extended data 4). However due to the higher reliability of Illumina for low frequency variants, this was used for formal analysis[2,3]. Additionally, single genome amplification and sequencing of Spike using extracted RNA from respiratory samples was used as an independent method to detect mutations observed (Extended data 4). 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 1). 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 (Extended data 2). All samples were consistent with having arisen from a single underlying 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[4,5], (Extended data 2, Supplementary Table 2). While the sequences from the three local patients as well as from Avanzato et al[5] showed little divergence with no amino acid changes in Spike over time, the case patient showed significant diversification. The Choi et al report[4] 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 (Extended data 5 and 6). The relationship of the divergent samples to those at earlier time points argues against superinfection.

SARS-CoV-2 viral diversity

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 with polyclonal antibodies (days 63 and 65, Extended data 3). Of note we were not able to culture virus from stored swab samples. Consensus sequences from short read deep sequence Illumina data revealed dynamic population changes after day 65, as shown by a highlighter plot (Extended data 6). In addition, we were also able to follow the dynamics of virus populations down to low frequencies during the entire period (Figure 2, Supplementary Table 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 a T39I (C27509T) mutation in ORF7a reaching 79% on day 45 (Figure 2, pink, supplementary information). At day 66 we noted I513T in NSP2 (T2343C) and V157L (G13936T) in RdRp had emerged from undetectable at day 54 to almost 100% frequency (Figure 2, red and green dashed lines), with the polymerase being the more plausible candidate for driving this sweep. Notably, spike variant N501Y, which can increase the ACE2 receptor affinity[6], and which is present in the new UK B1.1.7 lineage[7], was observed on day 55 at 33% frequency, but was eliminated by the sweep of the NSP2/RdRp variant.

In contrast to the early period of infection, between days 66 and 82, following the first two administrations of convalescent sera, a shift in the virus population was observed, with a variant bearing D796H in S2 and **Δ**H69/**Δ**V70 in the S1 N-terminal domain (NTD) becoming the dominant population at day 82. This was identified in a nose and throat swab sample with high viral load as indicated by Ct of 23 (Figure 3A). The deletion was detected transiently at baseline according to short read deep sequencing. **Δ**H69/**Δ**V70 was due to an out of frame six nucleotide deletion resulting in the sequence of codon 68 changing from ATA to ATC.

On Days 86 and 89, viruses obtained from upper respiratory tract samples were characterised by the Spike mutations Y200H and T240I, with the deletion/mutation pair observed on day 82

having fallen to frequencies of 10% or less (Figure 2 and 3). The Spike mutations Y200H and T240I were accompanied at high frequency by two other non-synonymous variants with similar allele frequencies, coding for I513T in NSP2, V157L in RdRp and N177S in NSP15 (Figure 2A). Both of these were also previously observed at >98% frequency in the sample on day 66 (Figure 2A, red and green lines), arguing that this new lineage emerged out of a previously existing population.

Sequencing of a nose and throat swab sample at day 93 identified viruses characterised by Spike mutations P330S at the edge of the RBD and W64G in S1 NTD at close to 100% abundance, with D796H along with **Δ**H69/**Δ**V70 at <1% abundance and the variants Y200H and T240I at frequencies of <2%. Viruses with the P330S variant were detected in two independent samples from different sampling sites, arguing against the possibility of contamination. The divergence of these samples from the remainder of the population (Figure 2, 3B and Extended data 5 and 6) suggests the possibility that they represent a compartmentalised subpopulation.

Patterns in the variant frequencies suggest competition between virus populations carrying different mutations, viruses with the D796H/**Δ**H69/**Δ**V70 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 Y200H and T240I, 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 dose of CP (day 95) were administered. We observed a re-emergence of the D796H + **Δ**H69/**Δ**V70 viral population (Figure 2, 3). The inferred linkage of D796H and **Δ**H69/**Δ**V70 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 2), 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 neutralising antibody potency

Using lentiviral pseudotyping we generated wild type, **Δ**H69/**Δ**V70 + D796H and single mutant Spike proteins in enveloped virions in order to measure neutralisation activity of CP against these viruses (Figure 4). This system has been shown to give generally similar results to replication competent virus[8,9]. Spike protein from each mutant was detected in pelleted virions (Figure 4A). We also probed with an HIV-1 p24 antibody to monitor levels of lentiviral particle production (Figure 4A, Supplementary Figure 2). We then measured infectivity of the pseudoviruses, correcting for virus input using reverse transcriptase activity measurement, and found that **Δ**H69/**Δ**V70 appeared to have two-fold higher infectivity over a single round of infection compared to wild type (Figure 4B, Extended data 7). By contrast, the D796H single mutant had significantly lower infectivity as compared to wild type and double mutant had similar infectivity to wild type (Figure 4B, Extended data 7).

We found that D796H alone and the D796H + **Δ**H69/**Δ**V70 double mutant were less sensitive to neutralisation by convalescent plasma samples (Figure 4C-E, Extended data 7). By contrast the **Δ**H69/**Δ**V70 single mutant did not reduce neutralisation sensitivity. In addition, patient derived serum from days 64 and 66 (one day either side of CP2 infusion) similarly showed lower potency against the D796H + **Δ**H69/**Δ**V70 mutants (Figure 4F, 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 **Δ**H69/**Δ**V70) 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 eight RBD-specific mAbs (Extended data 8) exhibited no major change in neutralisation potency and non-RBD specific COVA1-21 showing 3-5 fold reduction in potency against **Δ**H69/**Δ**V70+D796H and **Δ**H69/**Δ**V70, but not D796H alone[9] (Extended data 8). 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. Interestingly , **Δ**H69/**Δ**V70 containing viruses showed reduced neutralisation sensitivity to the mAb COVA1-21, targeting an as yet undefined epitope outside the RBD. [10].

To understand how the **Δ**H69/**Δ**V70 and D796H might confer antibody resistance, we assessed how they might affect the Spike structure (Extended data 9). We based this analysis primarily on a structure lacking stabilising modifications (PDB 6xr8)[11], but also referred to stabilised structures determined at different pH values[12]. **Δ**H69/**Δ**V70 is located in a disordered, glycosylated loop at the distal surface of the NTD, near the binding site of polyclonal antibodies derived from COV57 plasma[13,14] (Extended data 9). As this loop is flexible and highly accessible, ΔH69/V70 could in principle affect antibody binding in this region. D796 is located near the base of Spike, in a surface loop that is structurally somewhat disordered in the prefusion conformation and becomes part of a large disordered region in the post fusion S2 trimer[11] (Extended data 9). The loop containing residue 796 is proposed to be targeted by antibodies[15], despite mutations at position 796 being relatively uncommon (Extended data 9). In the RBD-down Spike structures[11,12], D796 forms contacts with residues in the neighbouring protomer, including the glycosylated residue N709 (Extended data 9).

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 convalescent plasma containing polyclonal antibodies. In this case the emergence of the variant was not the primary reason for treatment failure.

We have noted in our analysis signs of compartmentalised viral replication based on the sequences recovered in upper respiratory tract samples. Both population genetic and small animal studies have shown a lack of reassortment between influenza viruses within a single host during an infection, suggesting that acute respiratory viral infection may be characterised by spatially distinct viral populations[16,17]. In the analysis of data, it is important to distinguish genetic changes which occur in the primary viral population from apparent changes that arise from the stochastic observation of spatially distinct subpopulations in the host. While the samples we observe on days 93 and 95 of infection are genetically distinct from the others, the remaining samples are consistent with arising from a consistent viral population. We note that Choi et al reported the detection in post-mortem tissue of viral RNA not only in lung tissue, but also in the spleen, liver, and heart[4]. Mixing of virus from different compartments, for example via blood, or movement of secretions from lower to upper respiratory tract, could lead to fluctuations in viral populations at particular sampling sites.

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

An important limitation is that the data were derived from sampling from the upper respiratory tract and not the lower tract, thus limiting the inferences that can be drawn regarding viral populations in this single case.

In addition to documenting the emergence of SARS-CoV-2 Spike **Δ**H69/**Δ**V70 *in vivo*, we show that this mutation modestly 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 **Δ**H69/**Δ**V70, conferred reduction in susceptibility to polyclonal antibodies in the units of CP administered,

261 though we cannot speculate as to their individual impacts on sera from other individuals. It is
262 intriguing that the **Δ**H69/**Δ**V70 + D796H double mutant diminished in between CP courses,
263 suggesting that there were other selective forces at play in the intervening period, possibly
264 driven by the inflammation observed in the individual. This includes the possibility that the
265 haplotype with **Δ**H69/**Δ**V70 + D796H may have carried mutations in other regions deleterious
266 during that intervening period. Although **Δ**H69/V70 is expanding at a high rate[18], D796
267 mutations are also increasing. D796H has been documented in 0.02% of global sequences and
268 D796Y appears in 0.05% of global sequences (Extended data 9).
269
270 The effects of CP on virus evolution seen here are unlikely to apply in immune competent hosts
271 where viral diversity is likely to be lower due to better immune control. Our data highlight that
272 infection control measures may need to be tailored to the needs of immunocompromised
273 patients and also caution in interpretation of CDC guidelines that recommend 20 days as the
274 upper limit of infection prevention precautions in immune compromised patients who are
275 afebrile[19]. Due to the difficulty with culturing clinical isolates, use of surrogates are warranted[20].
276 However, where detection of ongoing viral evolution is possible, this serves as a clear proxy for
277 the existence of infectious virus. In our case we detected environmental contamination whilst
278 in a single occupancy room and the patient was moved to a negative-pressure high air-change
279 infectious disease isolation room.
280
281 Clinical efficacy of convalescent plasma in severe COVID-19 has not been demonstrated[21], and
282 its use in different stages of infection and disease remains experimental; as such, we suggest
283 that it should be reserved for use within clinical trials, with rigorous monitoring of clinical and
284 virological parameters. The data from this single case report might warrant caution in use of
285 convalescent plasma in patients with immune suppression of both T cell and B cell arms; in such
286 cases, the antibodies administered have little support from cytotoxic T cells, thereby reducing
287 chances of clearance and theoretically raising the potential for escape mutations. Whilst we
288 await further data, where clinical trial enrolment is not possible, convalescent plasma
289 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.

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. We would like to thank James Nathan for RBD protein and Leo James for N protein. 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.

Competing interests: the authors declare no competing interests

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).

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, YM

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

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

References

1 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 e278, doi:10.1016/j.cell.2020.02.052 (2020).

2 Kim, K. W. *et al.* Respiratory viral co-infections among SARS-CoV-2 cases confirmed by virome capture sequencing. (2020).

3 Bull, R. A. *et al.* Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat Commun* **11**, 6272, doi:10.1038/s41467-020-20075-6 (2020).

4 Choi, B. *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *The New England journal of medicine* **383**, 2291-2293, doi:10.1056/NEJMc2031364 (2020).

5 Avanzato, V. A. *et al.* Case Study: Prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised cancer patient. *Cell* (2020).

6 Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310 e1220, doi:10.1016/j.cell.2020.08.012 (2020).

7 Rambaut A., L. N., Pybus O, Barclay W, Carabelli A. C., Connor T., Peacock T., Robertson D. L., Volz E., on behalf of COVID-19 Genomics Consortium UK (CoG-UK). *Preliminary genomic*

characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations, <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563> (2020).

8 Schmidt, F. *et al.* Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. 2020.2006.2008.140871, doi:10.1101/2020.06.08.140871 %J bioRxiv (2020).

9 Brouwer, P. J. M. *et al.* Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643-650, doi:10.1126/science.abc5902 (2020).

10 Zussman, M. E., Bagby, M., Benson, D. W., Gupta, R. & Hirsch, R. Pulmonary vascular resistance in repaired congenital diaphragmatic hernia vs. age-matched controls. *Pediatr Res* **71**, 697-700, doi:10.1038/pr.2012.16 (2012).

11 Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science*, doi:10.1126/science.abd4251 (2020).

12 Zhou, T. *et al.* Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host & Microbe* **28**, 867-879.e865, doi:10.1016/j.chom.2020.11.004 (2020).

13 Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437-442, doi:10.1038/s41586-020-2456-9 (2020).

14 Barnes, C. O. *et al.* Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828-842 e816, doi:10.1016/j.cell.2020.06.025 (2020).

15 Shrock, E. *et al.* Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science*, doi:10.1126/science.abd4250 (2020).

16 Sobel Leonard, A. *et al.* The effective rate of influenza reassortment is limited during human infection. *PLoS Pathog* **13**, e1006203, doi:10.1371/journal.ppat.1006203 (2017).

17 Richard, M., Herfst, S., Tao, H., Jacobs, N. T. & Lowen, A. C. Influenza A Virus Reassortment Is Limited by Anatomical Compartmentalization following Coinfection via Distinct Routes. *J Virol* **92**, doi:10.1128/JVI.02063-17 (2018).

18 Kemp, S. *et al.* Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion H69/V70. *bioRxiv*, 2020.2012.2014.422555, doi:10.1101/2020.12.14.422555 (2021).

19 CDC. *Discontinuation of Transmission-Based Precautions and Disposition of Patients with COVID-19 in Healthcare Settings (Interim Guidance)*, <https://www.cdc.gov/coronavirus/2019-ncov/hcp/disposition-hospitalized-patients.html> (2020).

20 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).

21 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).

22 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).

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 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).

28 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).

29 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).

30 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).

31 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).

32 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).

33 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).

34 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).

35 Minh, B. Q., Nguyen, M. A. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* **30**, 1188-1195, doi:10.1093/molbev/mst024 (2013).

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

37 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).

38 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).

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

40 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).

41 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. *Journal of Infectious Diseases* (2019).

42 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. *Proc Natl Acad Sci U S A* **93**, 11382-11388, doi:10.1073/pnas.93.21.11382 (1996).

43 Gupta, R. K. *et al.* Full length HIV-1 gag determines protease inhibitor susceptibility within in vitro assays. *AIDS* **24**, 1651 (2010).

44 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).

45 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).

46 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 legends

Figure 1. Analysis of 23 Patient derived whole SARS-CoV-2 genome sequences in context of local sequences and other cases of chronic SARS-CoV-2 shedding. 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.

Figure 2. 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. Variants labelled are represented by dashed lines. **A.** Variants detected in the patient from days 1-82. *D796H (light blue) is at the same frequency as NSP3 K902N (orange) therefore it is hidden beneath **B.** Variants detected in the patient from days 82-101.

Figure 3. Longitudinal variant frequencies and phylogenetic relationships for virus populations bearing six Spike (S) mutations A. At baseline, all six S variants (Illumina sequencing) except for **Δ**H69/V70 were absent (<1% and <20 reads). Approximately two weeks after receiving two units of convalescent plasma (CP), viral populations carrying **Δ**H69/V70 and

D796H mutants rose to frequencies >80% 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 **Δ**H69/V70 and D796H virus population re-emerged to become the dominant viral strain reaching variant frequencies of >75%. 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 dashed line and triangles). Where there were duplicate readings on the same day, to remain consistent, N+T samples were plotted **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 4: Spike mutant D796H + ΔH69/V70 infectivity and sensitivity convalescent plasma (CP). A. western blot of virus pellets after centrifugation of supernatants from cells transfected with lentiviral pseudotyping plasmids including Spike protein. Blots are representative of two independent transfections. **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. Data points represent technical replicates (n=3) with mean and error bars representing standard error of mean; data are representative of two independent experiments **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), expressed as fold change relative to WT. Data points represent means of technical replicates and each data point is an independent experiment (n=2-6). Mean of data points in C-G is shown by horizontal bars.

Extended data legends:

Extended Data Figure 1: Clinical time line of events with longitudinal respiratory sample CT values. CT – cycle threshold.

Extended data 2: A. 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. **B. 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. Mean is shown by line and whiskers representing standard deviation.

Extended Data Figure 3. **A. Serum SARS-CoV-2 antibody levels and virus population changes in chronic SARS-CoV-2 infection. 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. SARS-CoV-2 antibody titres in patient and in convalescent plasma.** Measurement of SARS-CoV-2 specific IgG antibody titres in three units of convalescent plasma (CP) by Euroimmun assay.

Extended data 4. Comparison between short-read (Illumina) and long-read single molecule (Oxford Nanopore) sequencing methods for the six observed Spike mutations. Concordance was generally good between the majority of timepoints, however due to large discrepancies in a number of timepoints, we suggest that due to the high base calling error rate, Nanopore is not yet suitable for calling minority variants. As such, all figures in the main paper were produced using Illumina data only. **B. 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. Amino acid variant and corresponding nucleotide position: S:W64G = 21752, S:**Δ**69 = 21765-21770, S:Y200H = 22160, S:T240I = 22281, S:P330S = 22550, S:D795H = 23948

Extended Data Figure 5: 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.

Extended Data Figure 6: A. 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 highly divergent. Environmental samples (patient's call bell, and patient's mobile phone) are indicated. Tree branched have been collapsed where bootstrap support was <60.

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).

Extended Data 7: In vitro infectivity and neutralisation sensitivity of Spike pseudotyped lentiviruses. A. infection of target 293T cells expressing TMPRSS2 and ACE2 receptors using equal amounts of virus as determined by reverse transcriptase activity. Data points represent technical replicates (n=2), with mean shown with error bars representing standard deviation. Data are representative of n=2 independent experiments (n=2). **B.** Representative Inverse dilution plots for Spike variants against convalescent plasma units 1-3. Data points represent mean neutralisation of technical replicates and error bars represent standard error of the mean of replicates. Data are representative of two independent experiments (n=2).

Extended Data Figure 8. A. Neutralization potency of a panel of monoclonal antibodies targeting the RBD is not impacted by Spike mutations D796H or ΔH69/V70. 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

means of technical replicates with error bars representing SD. Data are representative of at least two independent experiments. RBD: receptor binding domain. **B. Classes of RBD binding antibodies and fold changes for Spike mutations D796H or ΔH69/V70** are indicated based Bouwer et al. 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.

Extended Data 9. Location of Spike mutations ΔH69/Y70 and D796H. A. The SARS-CoV-2 spike trimer (PDB ID: 6xr8) with two protomers represented as surfaces and one protomer represented as a ribbon. The NTD is coloured in light blue, the RBD in light pink, the fusion peptide in dark pink, the HR1 domain in yellow, the CH domain in pale green, and the CD domain in brown. The location of D796 and H69 are indicated by red spheres. The loop connecting D796 to the fusion peptide is coloured magenta to improve visibility. The double grey lines provide orientation relative to the membrane. **B.** A close-up of the region defined by the box around H69 in panel A. H69 is highlighted in yellow. Residues containing atoms that are within 6 Å of H69 are highlighted in cyan. **C.** A close-up of the region defined by the box around D796 in panel A. D796 is highlighted in yellow. Residues containing atoms that are within 6 Å of D796 are highlighted in cyan. Hydrogen bonds are indicated by dashed yellow lines. Hydrophobic residues in the vicinity of D796 have been labelled. Y707 is from the neighbouring protomer. **D. Global prevalence of selected spike mutations detailed in this paper.** All high coverage sequences were downloaded from the GISAID database on 6th January and aligned using MAFFT; as of this date there were 298254 sequences available. 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.

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[22]. Amplification reaction were all performed on a Rotorgene™ PCR instrument. Samples which generated a CT of ≤36 were considered to be positive.

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

SARS-CoV-2 serology by multiplex particle-based flow cytometry (Luminex):

Recombinant SARS-CoV-2 N, S and RBD were covalently coupled to distinct carboxylated bead sets (Luminex; Netherlands) to form a 3-plex and analyzed as previously described (Xiong et al. 2020). Specific binding was reported as mean fluorescence intensities (MFI).

Whole blood T cell and innate stimulation assay

Whole blood was diluted 1:5 in RPMI into 96-well F plates (Corning) and activated by single stimulation with phytohemagglutinin (PHA; 10 µg/ml; Sigma-Aldrich), or LPS (1 µg/ml, List Biochemicals) or by co-stimulating with anti-CD3 (MEM57, Abcam, 200 ng/ml, 1:1000) and IL-2 (Immunotools, 1430U/ml, 1:1000). Supernatants were taken after 24 hours. Levels (pg/ml) are shown for IFNg, IL17, IL2, TNFa, IL6, IL1b and IL10. Cytokines were measured by multiplexed particle based Flow cytometry on a Luminex analyzer (Bio-Plex, Bio-Rad, UK) using an R&D Systems custom kit (R&D Systems, UK).

For viral genomic sequencing, total RNA was extracted from samples as described. Samples were sequenced using MinION flow cells version 9.4.1 (Oxford Nanopore Technologies) following the ARTICnetwork V3 protocol (https://dx.doi.org/10.17504/protocols.io.bbmuik6w) and BAM files assembled using the ARTICnetwork assembly pipeline

(https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html). A representative set of 10 sequences were selected and also sequenced using the Illumina MiSeq platform. Amplicons were diluted to 2 ng/µl and 25 µl (50 ng) were used as input for each library preparation reaction. The library preparation used KAPA Hyper Prep kit (Roche) according to manufacturer's instructions. Briefly, amplicons were end-repaired and had A-overhang added; these were then ligated with 15mM of NEXTflex DNA Barcodes (Bio Scientific, Texas, USA). Post-ligation products were cleaned using AMPure beads and eluted in 25 µl. Then, 20 µl were used for library amplification by 5 cycles of PCR. For the negative controls, 1ng was used for ligation-based library preparation. All libraries were assayed using TapeStation (Agilent Technologies, California, USA) to assess fragment size and quantified by QPCR. All libraries were then pooled in equimolar accordingly. Libraries were loaded at 15nM and spiked in 5% PhiX (Illumina, California, USA) and sequenced on one MiSeq 500 cycle using a Miseq Nano v2 with 2x 250 paired-end sequencing. A minimum of ten reads were required for a variant call.

Bioinformatics Processes

For long-read sequencing, genomes were assembled with reference-based assembly and a curated bioinformatics pipeline with 20x minimum coverage across the whole-genome[24]. For short-read sequencing, FASTQs were downloaded, poor-quality reads were identified and removed, and both Illumina and PHiX adapters were removed using TrimGalore v0.6.6[25]. Trimmed paired-end reads were mapped to the National Center for Biotechnology Information SARS-CoV-2 reference sequence MN908947.3 using MiniMap2-2.17 with arguments -ax and sr[26]. BAM files were then sorted and indexed with samtools v1.11 and PCR optical duplicates removed using Picard (http://broadinstitute.github.io/picard). A consensus sequences of nucleic acids with a minimum whole-genome coverage of at least 20× were generated with BCFtools using a 0% majority threshold.

Variant calling

Variant frequencies were validated using custom code as part of the *AnCovMulti* package (github.com/PollockLaboratory/AnCovMulti). The main idea behind this validation was to

identify and remove consistent potential amplification errors and mutability near the end of Illumina reads. Furthermore, stringent filtering was applied to remove biased amplification of early laboratory-induced mutations or very low copy variations.

Filtering consisted of requiring exact initiation at a primer within two bp of the start of a read, a minimum of 247 bp length read, fewer than four well-separated sites divergent from the reference sequence, a maximum insertion size of three nucleotides, a maximum deletion size of 11 bp, and resolution of conflicting signal from different primers.

Single Genome Amplification and sequencing

Viral RNA extracts were reverse transcribed from each sample to sufficiently capture the diversity of the viral population without introducing resampling bias. SuperScript IV (Thermofisher Scientific) and the gene specific primers were used for reverse transcription. Template RNA was degraded with RNAse H (Thermofisher Scientific). All primers used were 'in-house' primers designed using the multiple sequence alignment of the patient's consensus NGS sequences. Partial Spike (amino acids 21- 800) was amplified as 1 continuous length of DNA (Spike ~ 1.8 kb) by nested PCR. Terminally diluted cDNA was PCR- amplified using Platinum® Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA) so that 30% of reactions were positive[27]. By Poisson statistics, sequences were deemed ≥80% likely to be derived from HIV-1 single genomes. We obtained between 20–60 single genomes at each sample time point to achieve 90% confidence of detecting variants present at ≥8% of the viral population in vivo[28,29]. Partial spike amplicons obtained from terminal dilution PCR amplification were Sanger sequenced to form a contiguous sequence using another set of 8 in-house primers. Sanger sequencing was provided by Genewiz UK and manual sequence editing was performed using DNA Dynamo software (Blue Tractor Software Ltd, UK).

Phylogenetic Analysis

All available full-genome SARS-CoV-2 sequences were downloaded from the GISAID database (http://gisaid.org/)[30] on 16th December. Duplicate and low-quality sequences (>5% N regions) were removed, leaving a dataset of 212,297 sequences with a length of >29,000bp. All

sequences were sorted by name and only sequences sequenced with United Kingdom / England identifiers were retained. From this dataset, sequences were de-duplicated and where background sequences were required in figures, randomly subsampled using seqtk (https://github.com/lh3/seqtk). All sequences were aligned to the SARS-CoV-2 reference strain MN908947.3, using MAFFT v7.475 with automatic flavour selection[31]. Major SARS-CoV-2 clade memberships were assigned to all sequences using both the Nextclade server v0.9 (https://clades.nextstrain.org/) and Phylogenetic Assignment Of Named Global Outbreak Lineages (pangolin)[32].

Maximum likelihood phylogenetic trees were produced using the above curated dataset using IQ-TREE v2.1.2[33]. Evolutionary model selection for trees were inferred using ModelFinder[34] and trees were estimated using the GTR+F+I model with 1000 ultrafast bootstrap replicates[35]. All trees were visualised with Figtree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/), rooted on the SARS-CoV-2 reference sequence and nodes arranged in descending order. Nodes with bootstraps values of <50 were collapsed using an in-house script.

In-depth allele frequency variant calling

The SAMFIRE package version 1.06 [36] 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[37], 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.

All variants were indecently validated using custom code as part of the AnCovMulti package, found at https://github.com/PollockLaboratory/AnCovMulti.

Western blot analysis

Forty-eight hours after transfection of cells with plasmid preparations, the culture supernatant was harvested and passed through a 0.45-µm-pore-size filter to remove cellular debris. The filtrate was centrifuged at 15,000 rpm for 120 min to pellet virions. The pelleted virions were lysed in Laemmli reducing buffer (1 M Tris-HCl [pH 6.8], SDS, 100% glycerol, β-mercaptoethanol, and bromophenol blue). Pelleted virions were subjected to electrophoresis on SDS–4 to 12% bis-Tris protein gels (Thermo Fisher Scientific) under reducing conditions. This was followed by electroblotting onto polyvinylidene difluoride (PVDF) membranes. The SARS-CoV-2 Spike proteins were visualized by a ChemiDoc® MP imaging system (Biorad) using anti-Spike S2 (Invitrogen at 1:1000 dilution) and anti-p24 Gag antibodies (NIH AIDS Reagents 1:1000 dilution).

Recombination Detection

All sequences were tested for potential recombination, as this would impact on evolutionary estimates. Potential recombination events were explored with nine algorithms (RDP, MaxChi, SisScan, GeneConv, Bootscan, PhylPro, Chimera, LARD and 3SEQ), implemented in RDP5 with default settings[38]. To corroborate any findings, ClonalFrameML v1.12[39] was also used to infer recombination breakpoints. Neither programs indicated evidence of recombination in our data.

Structural Viewing

The Pymol Molecular Graphics System v2.4.0 (https://github.com/schrodinger/pymol-open-source/releases) was used to map the location of the four spike mutations of interested onto a SARS-CoV-2 spike structure visualised by Wrobel et al (PDB: 6ZGE)[40].

Testing of convalescent plasma for antibody titres

The Anti-SARS-CoV-2 ELISA (IgG) assay used to test CP for *antibody titres* was Euroimmun Medizinische Labordiagnostika AG. This indirect ELISA based assay uses a recombinant structural spike 1 (S1) protein of SARS-CoV-2 expressed in the human cell line HEK 293 for the detection of SARS-CoV2 IgG.

Generation of Spike mutants

Amino acid substitutions were introduced into the D614G pCDNA_SARS-CoV-2_Spike plasmid as previously described[41] using the QuikChange Lightening Site-Directed Mutagenesis kit, following the manufacturer's instructions (Agilent Technologies, Inc., Santa Clara, CA).

Pseudotype virus preparation

Viral vectors were prepared by transfection of 293T cells by using Fugene HD transfection reagent (Promega). 293T cells were transfected with a mixture of 11ul of Fugene HD, 1µg of pCDNAΔ19Spike-HA, 1ug of p8.91 HIV-1 gag-pol expression vector[42,43], and 1.5µg of pCSFLW (expressing the firefly luciferase reporter gene with the HIV-1 packaging signal). Viral supernatant was collected at 48 and 72h after transfection, filtered through 0.45um filter and

stored at -80˚C. The 50% tissue culture infectious dose ($TCID_{50}$) of SARS-CoV-2 pseudovirus was determined using Steady-Glo Luciferase assay system (Promega).

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

The reverse transcriptase activity of virus preparations was determined by qPCR using a SYBR Green-based product-enhanced PCR assay (SG-PERT) as previously described[44]. Briefly, 10-fold dilutions of virus supernatant were lysed in a 1:1 ratio in a 2x lysis solution (made up of 40% glycerol v/v 0.25% Trition X-100 v/v 100mM KCl, RNase inhibitor 0.8 U/ml, TrisHCL 100mM, buffered to pH7.4) for 10 minutes at room temperature.

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

Serum/plasma pseudotype neutralization assay

Spike pseudotype assays have been shown to have similar characteristics as neutralisation testing using fully infectious wild type SARS-CoV-2[8].Virus neutralisation assays were performed on 293T cell transiently transfected with ACE2 and TMPRSS2 using SARS-CoV-2 Spike pseudotyped virus expressing luciferase[45]. Pseudotyped virus was incubated with serial dilution of heat inactivated human serum samples or convalescent plasma in duplicate for 1h at 37˚C. Virus and cell only controls were also included. Then, freshly trypsinized 293T ACE2/TMPRSS2 expressing cells were added to each well. Following 48h incubation in a 5% CO_2 environment at 37°C, the luminescence was measured using Steady-Glo Luciferase assay system (Promega).

mAb pseudotype neutralisation assay

Virus neutralisation assays were performed on HeLa cells stably expressing ACE2 and using SARS-CoV-2 Spike pseudotyped virus expressing luciferase as previously described[46].

Pseudotyped virus was incubated with serial dilution of purified mAbs[9] in duplicate for 1h at 37˚C. Then, freshly trypsinized HeLa ACE2- expressing cells were added to each well. Following 48h incubation in a 5% CO_2 environment at 37°C, the luminescence was measured using Bright-Glo Luciferase assay system (Promega) and neutralization calculated relative to virus only controls. IC50 values were calculated in GraphPad Prism.

Data Availability

Long-read sequencing data that support the findings of this study have been deposited in the NCBI SRA database with the accession codes SAMN16976824 - SAMN16976846 under BioProject PRJNA682013 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA682013). Short reads and data used to construct figures were deposited at https://github.com/Steven-Kemp/sequence_files. All data are also available from the corresponding author.

Code Availability

The SAMFIRE package Version 1.06 was used for filtering and calling variants from the Illumina data. It is available at https://github.com/cjri/samfire/ for review. Additional code was used to validate the variant frequencies and can be found at https://github.com/PollockLaboratory/AnCovMulti.

The CITIID-NIHR BioResource COVID-19 Collaboration

Principal Investigators

Stephen Baker[2, 3], Gordon Dougan[2, 3], Christoph Hess[2,3,26,27], Nathalie Kingston[20, 12], Paul J. Lehner[2, 3], Paul A. Lyons[2,3], Nicholas J. Matheson[2,3], Willem H. Owehand[20], Caroline Saunders[19], Charlotte Summers[3,24,25,28], James E.D. Thaventhiran[2, 3, 22], Mark Toshner[3, 24, 25], Michael P. Weekes[2]

CRF and Volunteer Research Nurses

Ashlea Bucke[19], Jo Calder[19], Laura Canna[19], Jason Domingo[19], Anne Elmer[19], Stewart Fuller[19], Julie Harris[41], Sarah Hewitt[19], Jane Kennet[19], Sherly Jose[19], Jenny Kourampa[19], Anne Meadows[19], Criona O'Brien[41], Jane Price[19], Cherry Publico[19], Rebecca Rastall[19], Carla Ribeiro[19], Jane Rowlands[19], Valentina Ruffolo[19], Hugo Tordesillas[19]

Sample Logistics

Ben Bullman[2], Benjamin J. Dunmore[3], Stuart Fawke[30], Stefan Gräf [3,12,20], Josh Hodgson[3], Christopher Huang[3], Kelvin Hunter[2, 3], Emma Jones[29], Ekaterina Legchenko[3], Cecilia Matara[3], Jennifer Martin[3], Federica Mescia[2, 3], Ciara O'Donnell[3], Linda Pointon[3], Nicole Pond[2, 3], Joy Shih[3], Rachel Sutcliffe[3], Tobias Tilly[3], Carmen Treacy[3], Zhen Tong[3], Jennifer Wood[3], Marta Wylot[36],

Sample Processing and Data Acquisition

Laura Bergamaschi[2, 3], Ariana Betancourt[2, 3], Georgie Bower[2, 3], Chiara Cossetti[2, 3], Aloka De Sa[3], Madeline Epping[2, 3], Stuart Fawke[32], Nick Gleadall[20], Richard Grenfell[31], Andrew Hinch[2,3], Oisin Huhn[32], Sarah Jackson[3], Isobel Jarvis[3], Daniel Lewis[3], Joe Marsden[3], Francesca Nice[39], Georgina Okecha[3], Ommar Omarjee[3], Marianne Perera[3], Nathan Richoz[3], Veronika Romashova[2,3], Natalia Savinykh Yarkoni[3], Rahul Sharma[3], Luca Stefanucci[20], Jonathan Stephens[20], Mateusz Strzlecki[31], Lori Turner[2, 3],

Clinical Data Collection

Eckart M.D.D. De Bie[3], Katherine Bunclark[3], Masa Josipovic[40], Michael Mackay[3], Federica Mescia[2,3], Alice Michael[25], Sabrina Rossi[35], Mayurun Selvan[3], Sarah Spencer[15], Cissy Yong[35]

Royal Papworth Hospital ICU

Ali Ansaripour[25], Alice Michael[25], Lucy Mwaura[25], Caroline Patterson[25], Gary Polwarth[25]

Addenbrooke's Hospital ICU

Petra Polgarova[28], Giovanni di Stefano[28]

Cambridge and Peterborough Foundation Trust

Codie Fahey[34], Rachel Michel[34]

ANPC and Centre for Molecular Medicine and Innovative Therapeutics

Sze-How Bong[21], Jerome D. Coudert[33], Elaine Holmes[37]

NIHR BioResource

John Allison[20,12], Helen Butcher[12,38], Daniela Caputo[12,38], Debbie Clapham-Riley[12,38], Eleanor Dewhurst[12,38], Anita Furlong[12,38], Barbara Graves[12,38], Jennifer Gray[12,38], Tasmin Ivers[12,38], Mary Kasanicki[12,28], Emma Le Gresley[12,38], Rachel Linger[12,38], Sarah Meloy[12,38], Francesca Muldoon[12,38], Nigel Ovington[12,20], Sofia Papadia[12,38], Isabel Phelan[12,38], Hannah Stark[12,38], Kathleen E Stirrups[12,20], Paul Townsend[12,20], Neil Walker[12,20], Jennifer Webster[12,38].

19. Cambridge Clinical Research Centre, NIHR Clinical Research Facility, Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
20. Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK

21. Australian National Phenome Centre, Murdoch University, Murdoch, Western Australia WA 6150, Australia
22. MRC Toxicology Unit, School of Biological Sciences, University of Cambridge, Cambridge CB2 1QR, UK
23. R&D Department, Hycult Biotech, 5405 PD Uden, The Netherlands
24. Heart and Lung Research Institute, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK
25. Royal Papworth Hospital NHS Foundation Trust, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK
26. Department of Biomedicine, University and University Hospital Basel, 4031Basel, Switzerland
27. Botnar Research Centre for Child Health (BRCCH) University Basel & ETH Zurich, 4058 Basel, Switzerland
28. Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
29. Department of Veterinary Medicine, Madingley Road, Cambridge, CB3 0ES, UK
30. Cambridge Institute for Medical Research, Cambridge Biomedical Campus, Cambridge CB2 0XY, UK
31. Cancer Research UK, Cambridge Institute, University of Cambridge CB2 0RE, UK
32. Department of Obstetrics & Gynaecology, The Rosie Maternity Hospital, Robinson Way, Cambridge CB2 0SW, UK
33. Centre for Molecular Medicine and Innovative Therapeutics, Health Futures Institute, Murdoch University, Perth, WA, Australia
34. Cambridge and Peterborough Foundation Trust, Fulbourn Hospital, Fulbourn, Cambridge CB21 5EF, UK
35. Department of Surgery, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK
36. Department of Biochemistry, University of Cambridge, Cambridge, CB2 1QW, UK
37. Centre of Computational and Systems Medicine, Health Futures Institute, Murdoch University, Harry Perkins Building, Perth, WA 6150, Australia
38. Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK
39. Cancer Molecular Diagnostics Laboratory, Department of Oncology, University of Cambridge, Cambridge CB2 0AH, UK
40. Metabolic Research Laboratories, Wellcome Trust-Medical Research Council Institute of Metabolic Science, University of Cambridge, Cambridge CB2 0QQ, UK
41. Department of Paediatrics, University of Cambridge, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, UK

The COVID-62 Genomics UK (COG-UK) Consortium. https://www.cogconsortium.uk

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:

Samuel C Robson [54].

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Nicholas J Loman [82] and Thomas R Connor [51, 110].

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Tanya Golubchik [46].

Funding acquisition, Metadata curation, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Rocio T Martinez Nunez [83].

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:
Catherine Ludden [129].

Funding acquisition, Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:
Sally Corden [110].

Funding acquisition, Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:
Ian Johnston [140] and David Bonsall [46].

Funding acquisition, Leadership and supervision, Sequencing and analysis, Software and analysis tools, and Visualisation:
Colin P Smith [128] and Ali R Awan [69].

Funding acquisition, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:
Giselda Bucca [128].

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:
M. Estee Torok [63, 142].

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Visualisation:
Kordo Saeed [122, 151] and Jacqui A Prieto [124, 150].

Leadership and supervision, Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:
David K Jackson [140].

Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:

William L Hamilton [63].

Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Visualisation:
Luke B Snell [52].

Funding acquisition, Leadership and supervision, Metadata curation, and Samples and logistics:
Catherine Moore [110].

Funding acquisition, Leadership and supervision, Project administration, and Samples and logistics:
Ewan M Harrison [129,140].

Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:
Sonia Goncalves [140].

Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:
Derek J Fairley [44, 113], Matthew W Loose [59] and Joanne Watkins [110].

Leadership and supervision, Metadata curation, Samples and logistics, and Software and analysis tools:
Rich Livett [140].

Leadership and supervision, Metadata curation, Samples and logistics, and Visualisation:
Samuel Moses [66, 147].

Leadership and supervision, Metadata curation, Sequencing and analysis, and Software and analysis tools:
Roberto Amato [140], Sam Nicholls [82] and Matthew Bull [110].

Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:
Darren L Smith [78, 99, 146].

Leadership and supervision, Sequencing and analysis, Software and analysis tools, and Visualisation:
Jeff Barrett [140], David M Aanensen [55, 155].

Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:
Martin D Curran [106], Surendra Parmar [106], Dinesh Aggarwal [136, 140, 105] and James G Shepherd [89].

Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:
Matthew D Parker [134].

Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:
Sharon Glaysher [102].

Metadata curation, Sequencing and analysis, Software and analysis tools, and Visualisation:
Matthew Bashton [78, 99], Anthony P Underwood [55, 155], Nicole Pacchiarini [110] and Katie F Loveson [118].

Project administration, Sequencing and analysis, Software and analysis tools, and Visualisation:
Alessandro M Carabelli [129].

Funding acquisition, Leadership and supervision, and Metadata curation:
Kate E Templeton [94, 131].

Funding acquisition, Leadership and supervision, and Project administration:
Cordelia F Langford [140], John Sillitoe [140], Thushan I de Silva [134] and Dennis Wang [134].

Funding acquisition, Leadership and supervision, and Sequencing and analysis:
Dominic Kwiatkowski [140, 148], Andrew Rambaut [131], Justin O'Grady [111, 130] and Simon Cottrell [110].

Leadership and supervision, Metadata curation, and Sequencing and analysis:
Matthew T.G. Holden [109] and Emma C Thomson [89].

Leadership and supervision, Project administration, and Samples and logistics:
Husam Osman [77, 105], Monique Andersson [100], Anoop J Chauhan [102] and Mohammed O Hassan-Ibrahim [47].

Leadership and supervision, Project administration, and Sequencing and analysis:
Mara Lawniczak [140].

Leadership and supervision, Samples and logistics, and Sequencing and analysis:
Alex Alderton [140], Meera Chand [107], Chrystala Constantinidou [135], Meera Unnikrishnan [135], Alistair C Darby [133], Julian A Hiscox [133] and Steve Paterson [133].

Leadership and supervision, Sequencing and analysis, and Software and analysis tools:
Inigo Martincorena [140], David L Robertson [89], Erik M Volz [80], Andrew J Page [111] and Oliver G Pybus [64].

Leadership and supervision, Sequencing and analysis, and Visualisation:
Andrew R Bassett [140].

Metadata curation, Project administration, and Samples and logistics:
Cristina V Ariani [140], Michael H Spencer Chapman [129, 140], Kathy K Li [89], Rajiv N Shah [89], Natasha G Jesudason [89] and Yusri Taha [91].

Metadata curation, Project administration, and Sequencing and analysis:
Martin P McHugh [94] and Rebecca Dewar [94].

Metadata curation, Samples and logistics, and Sequencing and analysis:
Aminu S Jahun [65], Claire McMurray [82], Sarojini Pandey [125], James P McKenna [44], Andrew Nelson [99, 146], Gregory R Young [78, 99], Clare M McCann [99, 146] and Scott Elliott [102].

Metadata curation, Samples and logistics, and Visualisation:
Hannah Lowe [66].

Metadata curation, Sequencing and analysis, and Software and analysis tools:
Ben Temperton [132], Sunando Roy [123], Anna Price [51], Sara Rey [110] and Matthew Wyles [134].

Metadata curation, Sequencing and analysis, and Visualisation:
Stefan Rooke [131] and Sharif Shaaban [109].

Project administration, Samples and logistics, Sequencing and analysis:
Mariateresa de Cesare [139].

Project administration, Samples and logistics, and Software and analysis tools:
Laura Letchford [140].

Project administration, Samples and logistics, and Visualisation:
Siona Silveira [122], Emanuela Pelosi [122] and Eleri Wilson-Davies [122].

Samples and logistics, Sequencing and analysis, and Software and analysis tools:
Myra Hosmillo [65].

Sequencing and analysis, Software and analysis tools, and Visualisation:
Áine O'Toole [131], Andrew R Hesketh [128], Richard Stark [135], Louis du Plessis [64], Chris Ruis [129], Helen Adams [45] and Yann Bourgeois [117].

Funding acquisition, and Leadership and supervision:
Stephen L Michell [132], Dimitris Gramatopoulos [125], [153], Jonathan Edgeworth [53], Judith Breuer [71, 123], John A Todd [139] and Christophe Fraser [46].

Funding acquisition, and Project administration:
David Buck [139] and Michaela John [50].

Leadership and supervision, and Metadata curation:
Gemma L Kay [111].

Leadership and supervision, and Project administration:
Steve Palmer [140], Sharon J Peacock [129], [105] and David Heyburn [110].

Leadership and supervision, and Samples and logistics:
Danni Weldon [140], Esther Robinson [105, 77], Alan McNally [82, 127], Peter Muir [105], Ian B Vipond [105], John BoYes [70], Venkat Sivaprakasam [87], Tranprit Salluja [116], Samir Dervisevic [95] and Emma J Meader [95].

Leadership and supervision, and Sequencing and analysis:
Naomi R Park [140], Karen Oliver [140], Aaron R Jeffries [132], Sascha Ott [135], Ana da Silva Filipe [89], David A Simpson [113] and Chris Williams [110].

Leadership and supervision, and Visualisation:
Jane AH Masoli [114, 132].

Metadata curation, and Samples and logistics:
Bridget A Knight [114, 132], Christopher R Jones [114, 132], Cherian Koshy [42], Amy Ash [42], Anna Casey [112], Andrew Bosworth [105, 77], Liz Ratcliffe [112], Li Xu-McCrae [77], Hannah M Pymont [105], Stephanie Hutchings [105], Lisa Berry [125], Katie Jones [125], Fenella Halstead [87], Thomas Davis [62], Christopher Holmes [57], Miren Iturriza-Gomara [133], Anita O Lucaci [133], Paul Anthony Randell [79, 145], Alison Cox [79, 145], Pinglawathee Madona [79, 145], Kathryn Ann

Harris [71], Julianne Rose Brown [71], Tabitha W Mahungu [115], Dianne Irish-Tavares [115], Tanzina Haque [115], Jennifer Hart [115], Eric Witele [115], Melisa Louise Fenton [116], Steven Liggett [120], Clive Graham [97], Emma Swindells [98], Jennifer Collins [91], Gary Eltringham [91], Sharon Campbell [58], Patrick C McClure [138], Gemma Clark [56], Tim J Sloan [101], Carl Jones [56] and Jessica Lynch [43], [152].

Metadata curation, and Sequencing and analysis:
Ben Warne [49], Steven Leonard [140], Jillian Durham [140], Thomas Williams [131], Sam T Haldenby [133], Nathaniel Storey [71], Nabil-Fareed Alikhan [111], Nadine Holmes [59], Christopher Moore [59], Matthew Carlile [59], Malorie Perry [110], Noel Craine [110], Ronan A Lyons [121], Angela H Beckett [54], Salman Goudarzi [118], Christopher Fearn [118], Kate Cook [118], Hannah Dent [118] and Hannah Paul [118].

Metadata curation, and Software and analysis tools:
Robert Davies [140].

Project administration, and Samples and logistics:
Beth Blane [129], Sophia T Girgis [129], Mathew A Beale [140], Katherine L Bellis [140, 129], Matthew J Dorman [140], Eleanor Drury [140], Leanne Kane [140], Sally Kay [140], Samantha McGuigan [140], Rachel Nelson [140], Liam Prestwood [140], Shavanthi Rajatileka [140], Rahul Batra [53], Rachel J Williams [123], Mark Kristiansen [123], Angie Green [139], Anita Justice [100], Adhyana I.K Mahanama [122, 143] and Buddhini Samaraweera [122, 143].

Project administration, and Sequencing and analysis:
Nazreen F Hadjirin [129] and Joshua Quick [82].

Project administration, and Software and analysis tools:
Radoslaw Poplawski [82].

Samples and logistics, and Sequencing and analysis:
Leanne M Kermack [129], Nicola Reynolds [48], Grant Hall [65], Yasmin Chaudhry [65], Malte L Pinckert [65], Iliana Georgana [65], Robin J Moll [140], Alicia Thornton [107], Richard Myers [107], Joanne Stockton [82], Charlotte A Williams [123], Wen C Yew [99], Alexander J Trotter [111], Amy Trebes [139], George MacIntyre-Cockett [139], Alec Birchley [110], Alexander Adams [110], Amy Plimmer [110], Bree Gatica-Wilcox [110], Caoimhe McKerr [110], Ember Hilvers [110], Hannah Jones [110], Hibo Asad [110], Jason Coombes [110], Johnathan M Evans [110], Laia Fina [110], Lauren Gilbert [110], Lee Graham [110], Michelle Cronin [110], Sara Kumziene-SummerhaYes [110], Sarah Taylor [110], Sophie Jones [110], Danielle C Groves [134], Peijun Zhang [134], Marta Gallis [134] and Stavroula F Louka [134].

Samples and logistics, and Software and analysis tools:
Igor Starinskij [89].

Sequencing and analysis, and Software and analysis tools:
Chris Jackson [88], Marina Gourtovaia [140], Gerry Tonkin-Hill [140], Kevin Lewis [140], Jaime M Tovar-Corona [140], Keith James [140], Laura Baxter [135], Mohammad T. Alam [135], Richard J Orton [89], Joseph Hughes [89], Sreenu Vattipally [89], Manon Ragonnet-Cronin [80], Fabricia F. Nascimento [80], David Jorgensen [80], Olivia Boyd [80], Lily Geidelberg [80], Alex E Zarebski [64], Jayna Raghwani [64], Moritz UG Kraemer [64], Joel Southgate [51, 110], Benjamin B Lindsey [134] and Timothy M Freeman [134].

Software and analysis tools, and Visualisation:
Jon-Paul Keatley [140], Joshua B Singer [89], Leonardo de Oliveira Martins [111], Corin A Yeats [55], Khalil Abudahab [55], [155], Ben EW Taylor [55], [155] and Mirko Menegazzo [55].

Leadership and supervision:

John Danesh [140], Wendy Hogsden [87], Sahar Eldirdiri [62], Anita Kenyon [62], Jenifer Mason [84], Trevor I Robinson [84], Alison Holmes [79, 144], James Price [79, 144], John A Hartley [123], Tanya Curran [44], Alison E Mather [111], Giri Shankar [110], Rachel Jones [110], Robin Howe [110] and Sian Morgan [50].

Metadata curation:

Elizabeth Wastenge [94], Michael R Chapman [75, 129, 140], Siddharth Mookerjee [79, 144], Rachael Stanley [95], Wendy Smith [56], Timothy Peto [100], David Eyre [100], Derrick Crook [100], Gabrielle Vernet [74], Christine Kitchen [51], Huw Gulliver [51], Ian Merrick [51], Martyn Guest [51], Robert Munn [51], Declan T Bradley [104, 113] and Tim Wyatt [104].

Project administration:

Charlotte Beaver [140], Luke Foulser [140], Sophie Palmer [129], Carol M Churcher [129], Ellena Brooks [129], Kim S Smith [129], Katerina Galai [129], Georgina M McManus [129], Frances Bolt [79, 144], Francesc Coll [60], Lizzie Meadows [111], Stephen W Attwood [64], Alisha Davies [110], Elen De Lacy [110], Fatima Downing [110], Sue Edwards [110], Garry P Scarlett [117], Sarah Jeremiah [124] and Nikki Smith [134].

Samples and logistics:

Danielle Leek [129], Sushmita Sridhar [129, 140], Sally Forrest [129], Claire Cormie [129], Harmeet K Gill [129], Joana Dias [129], Ellen E Higginson [129], Mailis Maes [129], Jamie Young [129], Michelle Wantoch [48], Sanger Covid Team (www.sanger.ac.uk/covid-team) [140], Dorota Jamrozy [140], Stephanie Lo [140], Minal Patel [140], Verity Hill [131], Claire M Bewshea [132], Sian Ellard [114, 132], Cressida Auckland [114], Ian Harrison [107], Chloe Bishop [107], Vicki Chalker [107], Alex Richter [126], Andrew Beggs [126], Angus Best [127], Benita Percival [127], Jeremy Mirza [127], Oliver Megram [127], Megan Mayhew [127], Liam Crawford [127], Fiona Ashcroft [127], Emma Moles-Garcia [127], Nicola Cumley [127], Richard Hopes [105], Patawee Asamaphan [89], Marc O Niebel [89], Rory N Gunson [141], Amanda Bradley [93], Alasdair Maclean [93], Guy Mollett [93], Rachel Blacow [93], Paul Bird [57], Thomas Helmer [57], Karlie Fallon [57], Julian Tang [57], Antony D Hale [90], Louissa R Macfarlane-Smith [90], Katherine L Harper [90], Holli Carden [90], Nicholas W Machin [86, 105], Kathryn A Jackson [133], Shazaad S Y Ahmad [86, 105], Ryan P George [86], Lance Turtle [133], Elaine O'Toole [84], Joanne Watts [84], Cassie Breen [84], Angela Cowell [84], Adela Alcolea-Medina [73, 137], Themoula Charalampous [53, 83], Amita Patel [52], Lisa J Levett [76], Judith Heaney [76], Aileen Rowan [80], Graham P Taylor [80], Divya Shah [71], Laura Atkinson [71], Jack CD Lee [71], Adam P Westhorpe [123], Riaz Jannoo [123], Helen L Lowe [123], Angeliki Karamani [123], Leah Ensell [123], Wendy Chatterton [76], Monika Pusok [76], Ashok Dadrah [116], Amanda Symmonds [116], Graciela Sluga [85], Zoltan Molnar [113], Paul Baker [120], Stephen Bonner [120], Sarah Essex [120], Edward Barton [97], Debra Padgett [97], Garren Scott [97], Jane Greenaway [98], Brendan AI Payne [91], Shirelle Burton-Fanning [91], Sheila Waugh [91], Veena Raviprakash [58], Nicola Sheriff [58], Victoria Blakey [58], Lesley-Anne Williams [58], Jonathan Moore [68], Susanne Stonehouse [68], Louise Smith [96], Rose K Davidson [130], Luke Bedford [67], Lindsay Coupland [95], Victoria Wright [59], Joseph G Chappell [138], Theocharis Tsoleridis [138], Jonathan Ball [138], Manjinder Khakh [56], Vicki M Fleming [56], Michelle M Lister [56], Hannah C Howson-Wells [56], Louise Berry [56], Tim Boswell [56], Amelia Joseph [56], Iona Willingham [56], Nichola Duckworth [101], Sarah Walsh [101], Emma Wise [43, 152], Nathan Moore [43, 152], Matilde Mori [43, 149, 152], Nick Cortes [43, 152], Stephen Kidd [43, 152], Rebecca Williams [74], Laura Gifford [110], Kelly Bicknell [102], Sarah Wyllie [102], Allyson Lloyd [102], Robert Impey [102], Cassandra S Malone [47], Benjamin J Cogger [47], Nick Levene [103], Lynn Monaghan [103], Alexander J Keeley [134], David G Partridge [119, 134], Mohammad Raza [119, 134], Cariad Evans [119, 134] and Kate Johnson [119, 134].

Sequencing and analysis:

Emma Betteridge [140], Ben W Farr [140], Scott Goodwin [140], Michael A Quail [140], Carol Scott [140], Lesley Shirley [140], Scott AJ Thurston [140], Diana Rajan [140], Iraad F Bronner [140], Louise Aigrain [140], Nicholas M Redshaw [140], Stefanie V Lensing [140], Shane McCarthy [140], Alex Makunin [140], Carlos E Balcazar [131], Michael D Gallagher [131], Kathleen A Williamson [131], Thomas D Stanton [131], Michelle L Michelsen [132], Joanna Warwick-Dugdale [132], Robin Manley [132], Audrey Farbos [132], James W Harrison [132], Christine M Sambles [132], David J Studholme [132], Angie Lackenby [107], Tamyo Mbisa [107], Steven Platt [107], Shahjahan Miah [107], David Bibby [107], Carmen Manso [107], Jonathan Hubb [107], Gavin Dabrera [107], Mary Ramsay [107], Daniel Bradshaw [107], Ulf Schaefer [107], Natalie Groves [107], Eileen Gallagher [107], David Lee [107], David Williams [107], Nicholas Ellaby [107], Hassan Hartman [107], Nikos Manesis [107], Vineet Patel [107], Juan Ledesma [108], Katherine A Twohig [108], Elias Allara [105], [129], Clare Pearson [105], [129], Jeffrey K. J. Cheng [135], Hannah E. Bridgewater [135], Lucy R. Frost [135], Grace Taylor-Joyce [135], Paul E Brown [135], Lily Tong [89], Alice Broos [89], Daniel Mair [89], Jenna Nichols [89], Stephen N Carmichael [89], Katherine L Smollett [81], Kyriaki Nomikou [89], Elihu Aranday-Cortes [89], Natasha Johnson [89], Seema Nickbakhsh [89], [109], Edith E Vamos [133], Margaret Hughes [133], Lucille Rainbow [133], Richard Eccles [133], Charlotte Nelson [133], Mark Whitehead [133], Richard Gregory [133], Matthew Gemmell [133], Claudia Wierzbicki [133], Hermione J Webster [133], Chloe L Fisher [69], Adrian W Signell [61], Gilberto Betancor [61], Harry D Wilson [61], Gaia Nebbia [53], Flavia Flaviani [72], Alberto C Cerda [137], Tammy V Merrill [137], Rebekah E Wilson [137], Marius Cotic [123], Nadua Bayzid [123], Thomas Thompson [113], Erwan Acheson [113], Steven Rushton [92], Sarah O'Brien [92], David J Baker [111], Steven Rudder [111], Alp Aydin [111], Fei Sang [59], Johnny Debebe [59], Sarah Francois [64], Tetyana I Vasylyeva [64], Marina Escalera Zamudio [64], Bernardo Gutierrez [64], Angela Marchbank [51], Joshua Maksimovic [50], Karla Spellman [50], Kathryn McCluggage [50], Mari Morgan [110], Robert Beer [50], Safiah Afifi [50], Trudy Workman [51], William Fuller [51], Catherine Bresner [51], Adrienn Angyal [134], Luke R Green [134], Paul J Parsons [134], Rachel M Tucker [134], Rebecca Brown [134] and Max Whiteley [134].

Software and analysis tools:

James Bonfield [140], Christoph Puethe [140], Andrew Whitwham [140], Jennifier Liddle [140], Will Rowe [82], Igor Siveroni [80], Thanh Le-Viet [111], and Amy Gaskin [110].

Visualisation:

Rob Johnson [80].

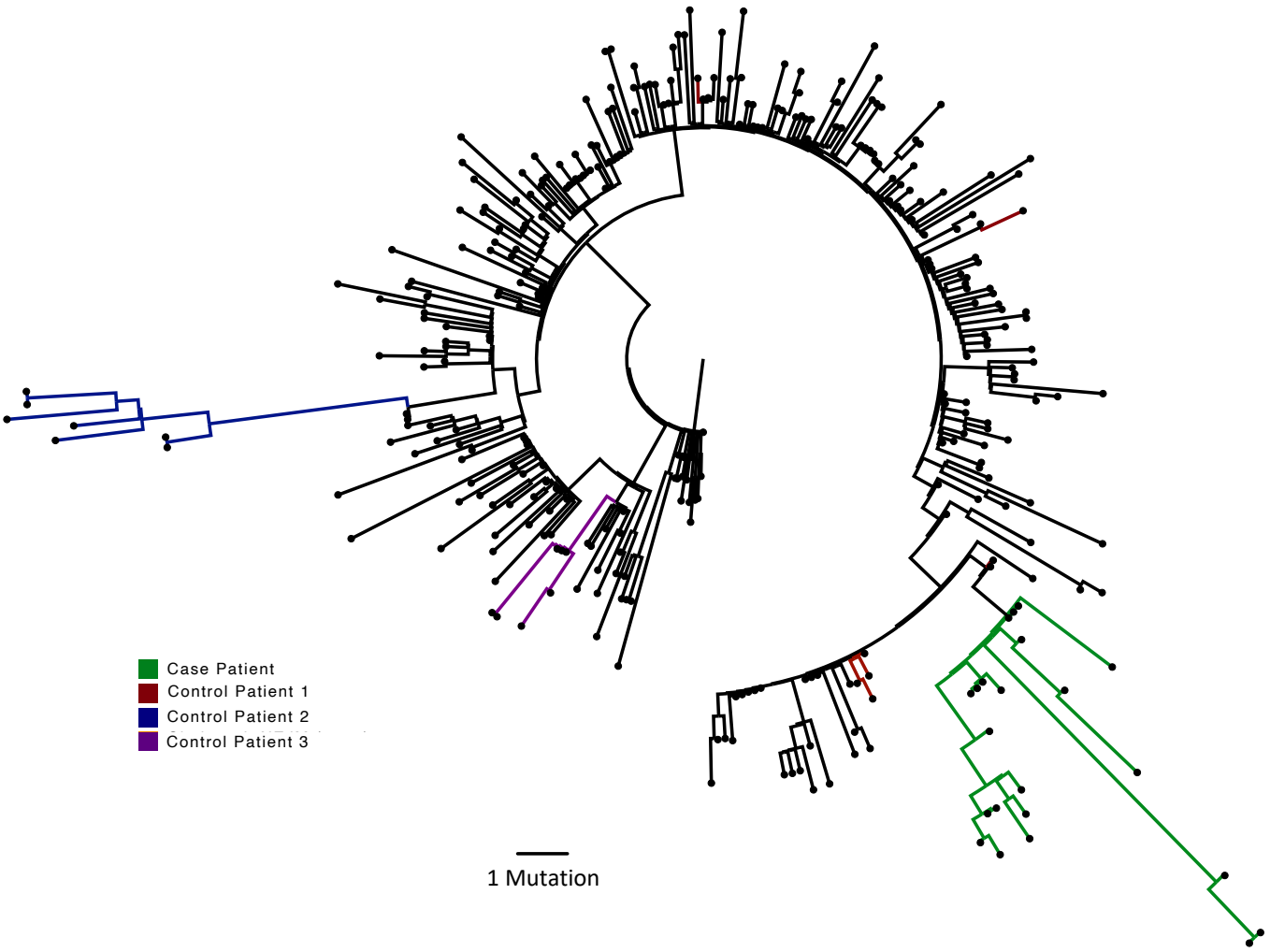
42. Barking, Havering and Redbridge University Hospitals NHS Trust, Barking, United Kingdom
43. Basingstoke Hospital, Basingstoke, United Kingdom
44. Belfast Health & Social Care Trust, Belfast, United Kingdom
45. Betsi Cadwaladr University Health Board, Betsi Cadwaladr, United Kingdom
46. Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
47. Brighton and Sussex University Hospitals NHS Trust, Brighton & Sussex, United Kingdom
48. Cambridge Stem Cell Institute, University of Cambridge, Cambridge, United Kingdom
49. Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom
50. Cardiff and Vale University Health Board, Cardiff, United Kingdom
51. Cardiff University, Cardiff, United Kingdom
52. Centre for Clinical Infection & Diagnostics Research, St. Thomas' Hospital and Kings College London, London, United Kingdom
53. Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

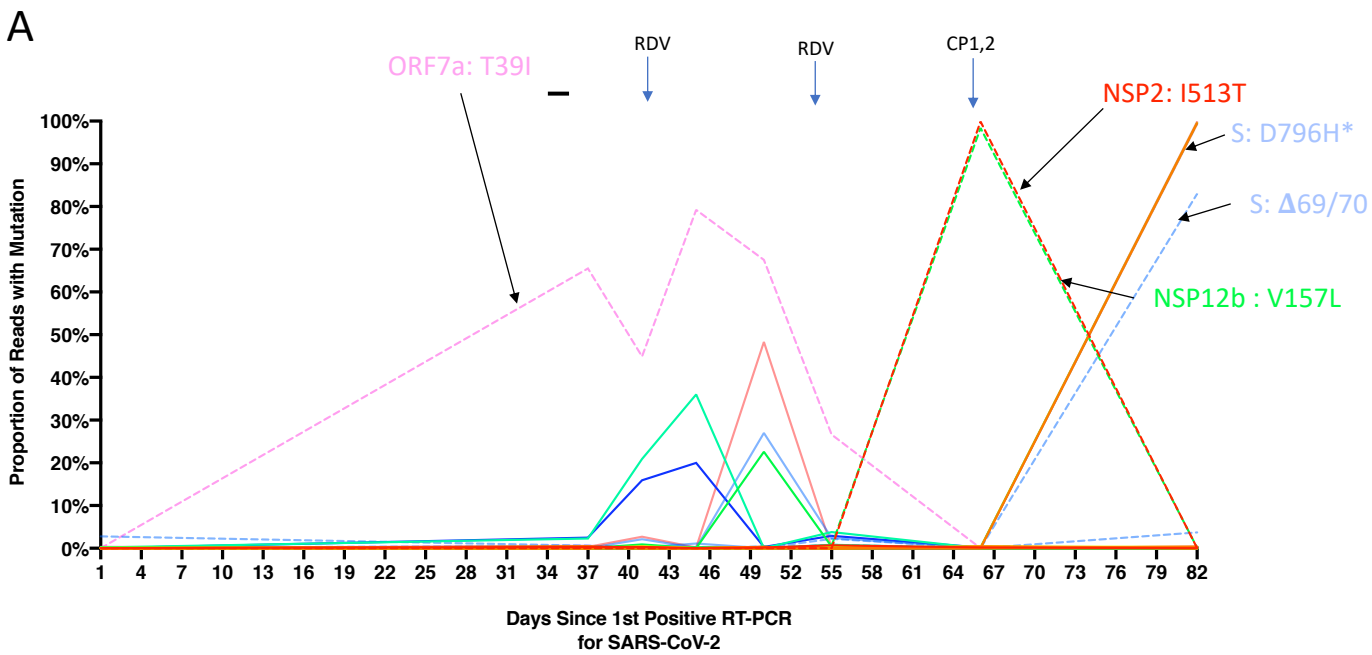
54. Centre for Enzyme Innovation, University of Portsmouth (PORT), Portsmouth, United Kingdom
55. Centre for Genomic Pathogen Surveillance, University of Oxford, Oxford, United Kingdom
56. Clinical Microbiology Department, Queens Medical Centre, Nottingham, United Kingdom
57. Clinical Microbiology, University Hospitals of Leicester NHS Trust, Leicester, United Kingdom
58. County Durham and Darlington NHS Foundation Trust, Durham, United Kingdom
59. Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, United Kingdom
60. Department of Infection Biology, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom
61. Department of Infectious Diseases, King's College London, London, United Kingdom
62. Department of Microbiology, Kettering General Hospital, Kettering, United Kingdom
63. Departments of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust; Cambridge, UK, Cambridge, United Kingdom
64. Department of Zoology, University of Oxford, Oxford, United Kingdom
65. Division of Virology, Department of Pathology, University of Cambridge, Cambridge, United Kingdom
66. East Kent Hospitals University NHS Foundation Trust, Kent, United Kingdom
67. East Suffolk and North Essex NHS Foundation Trust, Suffolk, United Kingdom
68. Gateshead Health NHS Foundation Trust, Gateshead, United Kingdom
69. Genomics Innovation Unit, Guy's and St. Thomas' NHS Foundation Trust, London, United Kingdom
70. Gloucestershire Hospitals NHS Foundation Trust, Gloucester, United Kingdom
71. Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom
72. Guy's and St. Thomas' BRC, London, United Kingdom
73. Guy's and St. Thomas' Hospitals, London, United Kingdom
74. Hampshire Hospitals NHS Foundation Trust, Hampshire, United Kingdom
75. Health Data Research UK Cambridge, Cambridge, United Kingdom
76. Health Services Laboratories, London, United Kingdom
77. Heartlands Hospital, Birmingham, Birmingham, United Kingdom
78. Hub for Biotechnology in the Built Environment, Northumbria University, Northumbria, United Kingdom
79. Imperial College Hospitals NHS Trust, London, United Kingdom
80. Imperial College London, London, United Kingdom
81. Institute of Biodiversity, Animal Health & Comparative Medicine, Glasgow, United Kingdom
82. Institute of Microbiology and Infection, University of Birmingham, Birmingham, United Kingdom
83. King's College London, London, United Kingdom
84. Liverpool Clinical Laboratories, Liverpool, United Kingdom
85. Maidstone and Tunbridge Wells NHS Trust, Maidstone, United Kingdom
86. Manchester University NHS Foundation Trust, Manchester, United Kingdom
87. Microbiology Department, Wye Valley NHS Trust, Hereford, United Kingdom
88. MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom

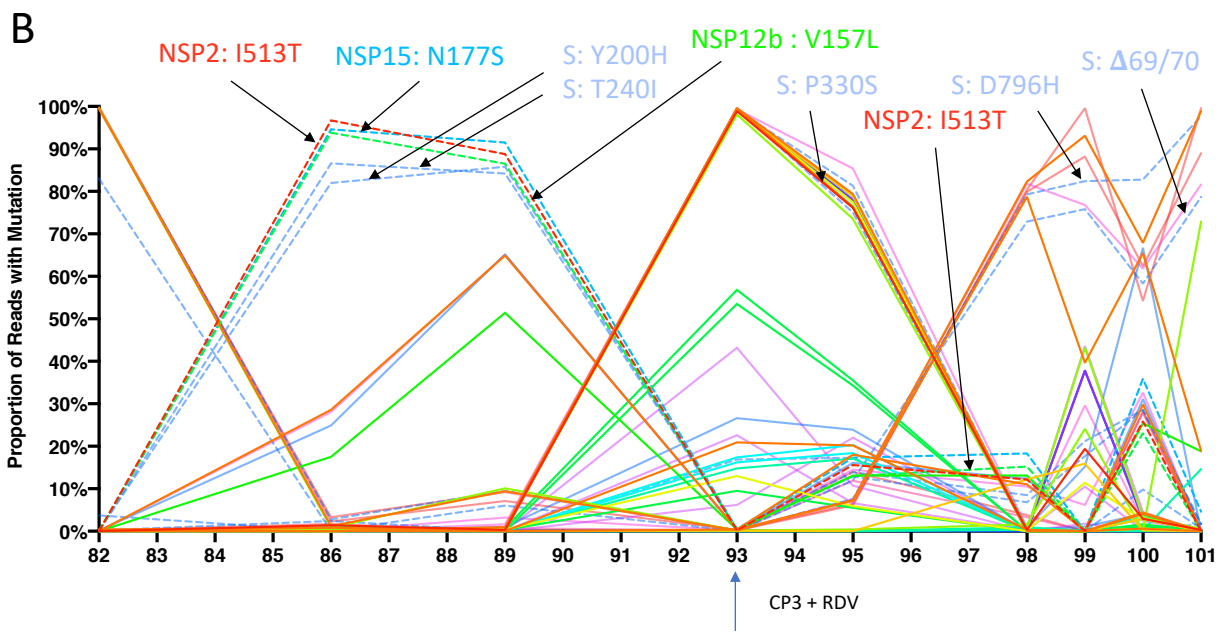
89. MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom
90. National Infection Service, PHE and Leeds Teaching Hospitals Trust, Leeds, United Kingdom
91. Newcastle Hospitals NHS Foundation Trust, Newcastle, United Kingdom
92. Newcastle University, Newcastle, United Kingdom
93. NHS Greater Glasgow and Clyde, Glasgow, United Kingdom
94. NHS Lothian, Edinburgh, United Kingdom
95. Norfolk and Norwich University Hospital, Norfolk, United Kingdom
96. Norfolk County Council, Norfolk, United Kingdom
97. North Cumbria Integrated Care NHS Foundation Trust, Carlisle, United Kingdom
98. North Tees and Hartlepool NHS Foundation Trust, Stockton-on-Tees, United Kingdom
99. Northumbria University, Northumbria, United Kingdom
100. Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
101. PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, Lincolnshire, United Kingdom
102. Portsmouth Hospitals University NHS Trust, Portsmouth, United Kingdom
103. Princess Alexandra Hospital Microbiology Dept., Harlow, United Kingdom
104. Public Health Agency, London, United Kingdom
105. Public Health England, London, United Kingdom
106. Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, United Kingdom
107. Public Health England, Colindale, London, United Kingdom
108. Public Health England, Colindale, London, United Kingdom
109. Public Health Scotland, Glasgow, United Kingdom
110. Public Health Wales NHS Trust, Cardiff, United Kingdom
111. Quadram Institute Bioscience, Norwich, United Kingdom
112. Queen Elizabeth Hospital, Birmingham, United Kingdom
113. Queen's University Belfast, Belfast, United Kingdom
114. Royal Devon and Exeter NHS Foundation Trust, Devon, United Kingdom
115. Royal Free NHS Trust, London, United Kingdom
116. Sandwell and West Birmingham NHS Trust, Sandwell, United Kingdom
117. School of Biological Sciences, University of Portsmouth (PORT), Portsmouth, United Kingdom
118. School of Pharmacy and Biomedical Sciences, University of Portsmouth (PORT), Portsmouth, United Kingdom
119. Sheffield Teaching Hospitals, Sheffield, United Kingdom
120. South Tees Hospitals NHS Foundation Trust, Newcastle, United Kingdom
121. Swansea University, Swansea, United Kingdom
122. University Hospitals Southampton NHS Foundation Trust, Southampton, United Kingdom
123. University College London, London, United Kingdom
124. University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom
125. University Hospitals Coventry and Warwickshire, Coventry, United Kingdom
126. University of Birmingham, Birmingham, United Kingdom
127. University of Birmingham Turnkey Laboratory, Birmingham, United Kingdom

128. University of Brighton, Brighton, United Kingdom
129. University of Cambridge, Cambridge, United Kingdom
130. University of East Anglia, East Anglia, United Kingdom
131. University of Edinburgh, Edinburgh, United Kingdom
132. University of Exeter, Exeter, United Kingdom
133. University of Liverpool, Liverpool, United Kingdom
134. University of Sheffield, Sheffield, United Kingdom
135. University of Warwick, Warwick, United Kingdom
136. University of Cambridge, Cambridge, United Kingdom
137. Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's College Hospital NHS Foundation Trust, London, United Kingdom
138. Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, United Kingdom
139. Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
140. Wellcome Sanger Institute, London, United Kingdom
141. West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, Glasgow, United Kingdom
142. Department of Medicine, University of Cambridge, Cambridge, United Kingdom
143. Ministry of Health, Colombo, Sri Lanka
144. NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, London, United Kingdom
145. North West London Pathology, London, United Kingdom
146. NU-OMICS, Northumbria University, Northumbria, United Kingdom
147. University of Kent, Kent, United Kingdom
148. University of Oxford, Oxford, United Kingdom
149. University of Southampton, Southampton, United Kingdom
150. University of Southampton School of Health Sciences, Southampton, United Kingdom
151. University of Southampton School of Medicine, Southampton, United Kingdom
152. University of Surrey, Guildford, United Kingdom
153. Warwick Medical School and Institute of Precision Diagnostics, Pathology, UHCW NHS Trust, Warwick, United Kingdom

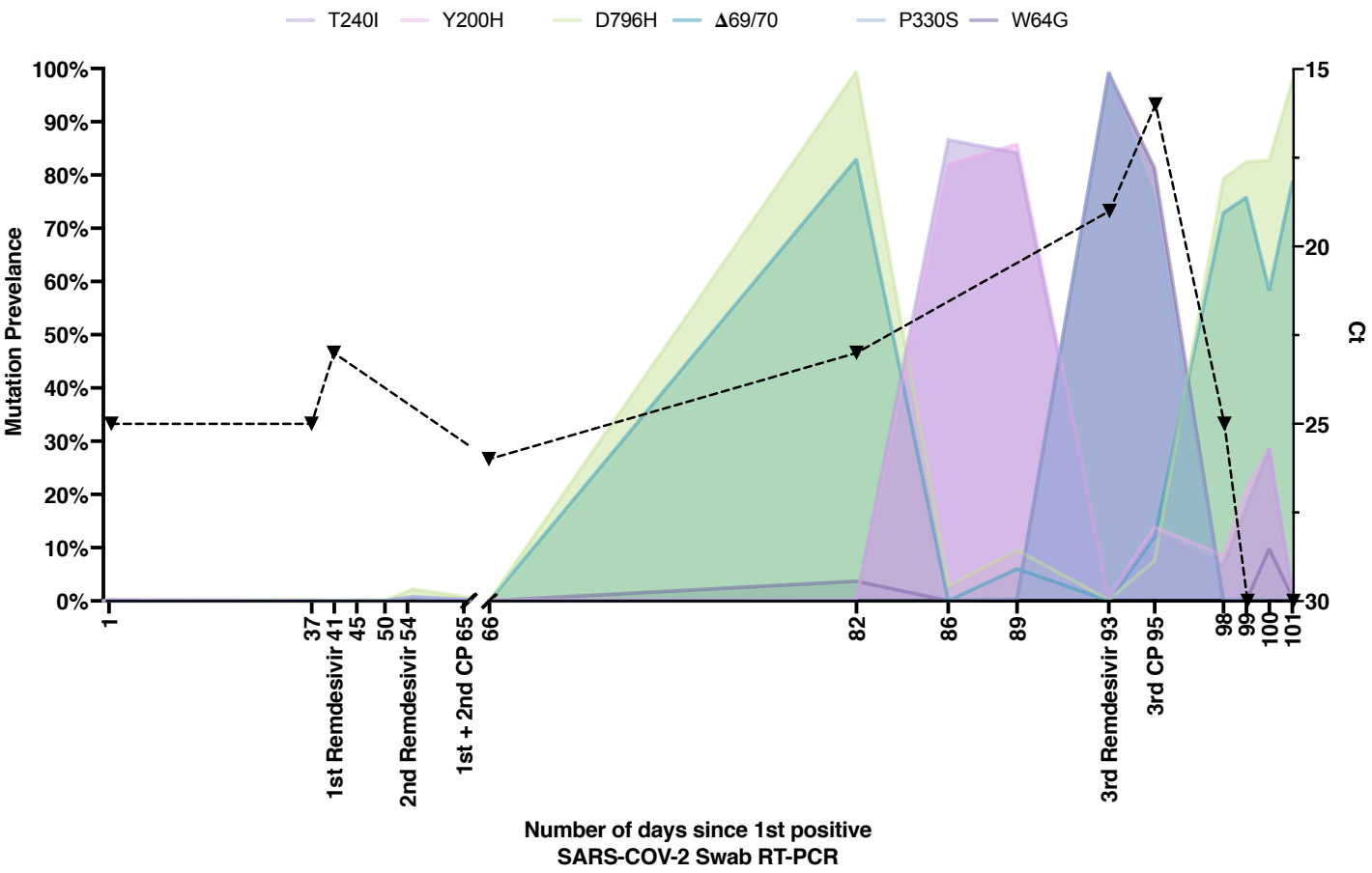
1252


Case Patient
Control Patient 1
Control Patient 2
Control Patient 3
1 Mutation

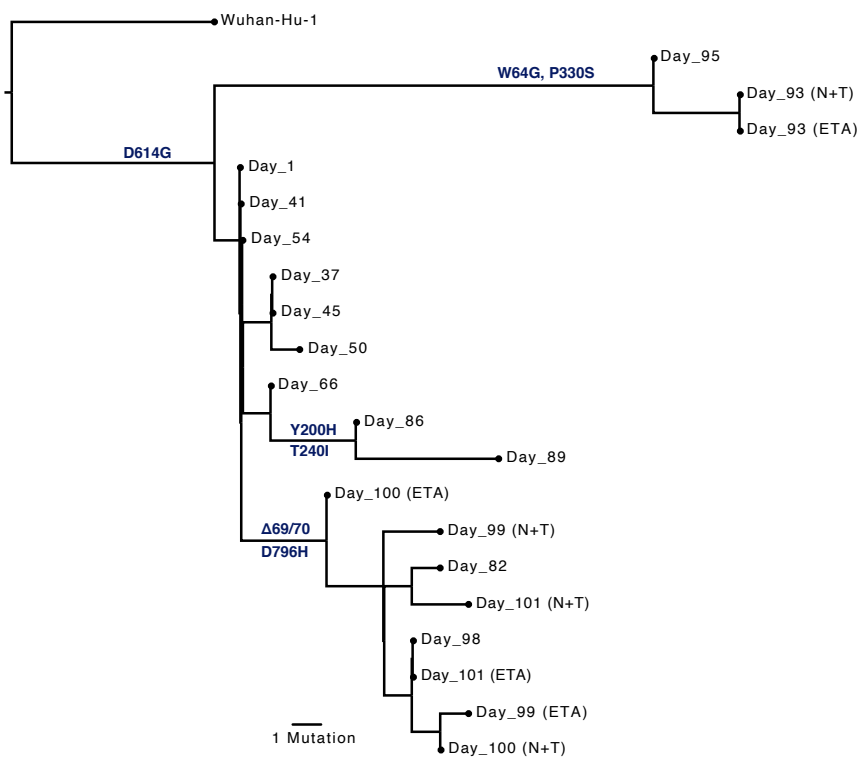

A
ORF7a: T39I
RDV
RDV
CP1,2
NSP2: I513T
S: D796H*
S: Δ69/70
NSP12b : V157L
Proportion of Reads with Mutation
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79 82
Days Since 1st Positive RT-PCR
for SARS-CoV-2

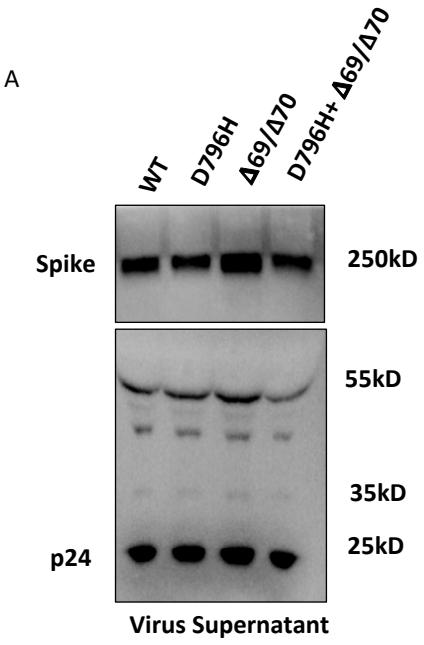

B
NSP2: I513T
NSP15: N177S
S: Y200H
S: T240I
NSP12b : V157L
S: P330S
NSP2: I513T
S: D796H
S: Δ69/70
Proportion of Reads with Mutation
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101
CP3 + RDV

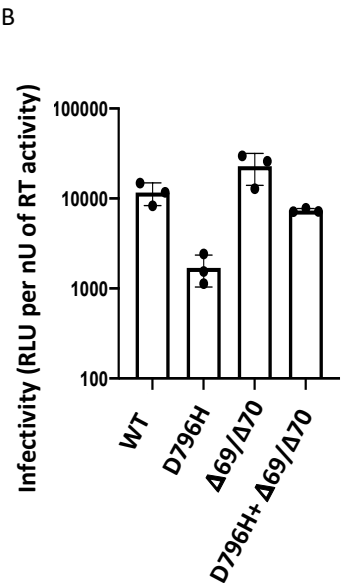
A

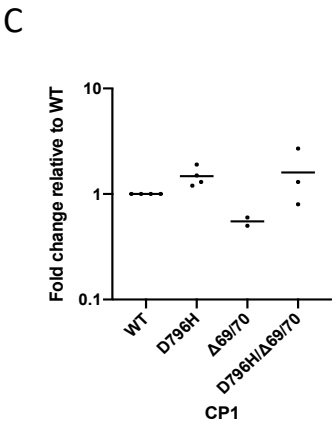

T240I
Y200H
D796H
Δ69/70
P330S
W64G
Mutation Prevelance
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Ct
15
20
25
30
1
37
41
45
50
54
65
66
82
86
89
93
95
98
99
100
101
1st Remdesivir
2nd Remdesivir
1st + 2nd CP
3rd Remdesivir
3rd CP
Number of days since 1st positive
SARS-COV-2 Swab RT-PCR

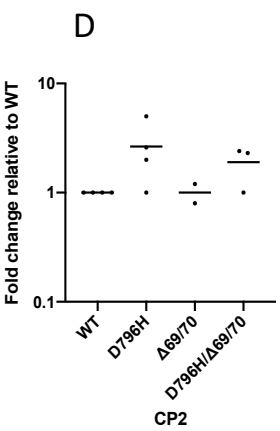
B

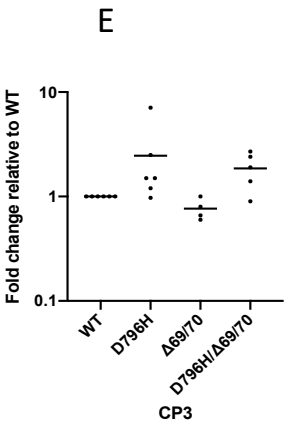

Wuhan-Hu-1
W64G, P330S
Day_95
Day_93 (N+T)
Day_93 (ETA)
D614G
Day_1
Day_41
Day_54
Day_37
Day_45
Day_50
Day_66
Y200H
T240I
Day_86
Day_89
Day_100 (ETA)
Δ69/70
D796H
Day_99 (N+T)
Day_82
Day_101 (N+T)
Day_98
Day_101 (ETA)
Day_99 (ETA)
Day_100 (N+T)
1 Mutation

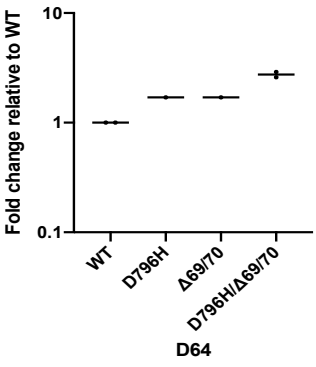

A
WT
D796H
Δ69/Δ70
D796H+ Δ69/Δ70
Spike
250kD
55kD
35kD
p24
25kD
Virus Supernatant

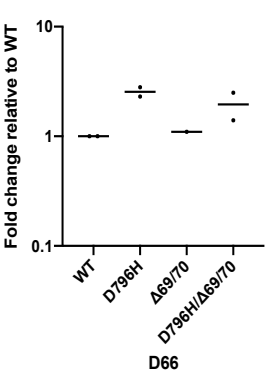

B
Infectivity (RLU per nU of RT activity)
100000
10000
1000
100
WT
D796H
Δ69/Δ70
D796H+ Δ69/Δ70


C
Fold change relative to WT
10
1
0.1
WT
D796H
Δ69/70
D796H/Δ69/70
CP1


D
Fold change relative to WT
10
1
0.1
WT
D796H
Δ69/70
D796H/Δ69/70
CP2


E
Fold change relative to WT
10
1
0.1
WT
D796H
Δ69/70
D796H/Δ69/70
CP3


F
Fold change relative to WT
10
1
0.1
WT
D796H
Δ69/70
D796H/Δ69/70
D64


G
Fold change relative to WT
10
1
0.1
WT
D796H
Δ69/70
D796H/Δ69/70
D66