Article

SARS-CoV-2 evolution during treatment of chronic infection

https://doi.org/10.1038/s41586-021-03291-y

Received: 1 December 2020

Accepted: 26 January 2021

Published online: 5 February 2021

Check for updates

Steven A. Kemp[1,148], Dami A. Collier[1,2,3,148], Rawlings P. Datir[2,3,148], Isabella A. T. M. 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[4], Elizabeth Blane[3], Yorgo Modis[2,3,8], Kendra E. Leigh[2,3,8], John A. G. Briggs[8], Marit J. van Gils[9], Kenneth G. C. Smith[2,3], John R. Bradley[3,10], Chris Smith[11], Rainer Doffinger[12], Lourdes Ceron-Gutierrez[12], Gabriela Barcenas-Morales[12,13], David D. Pollock[14], Richard A. Goldstein[1], Anna Smielewska[5,11], Jordan P. Skittrall[4,15,16], Theodore Gouliouris[4], Ian G. Goodfellow[5], Effrossyni Gkrania-Klotsas[4], Christopher J. R. Illingworth[15,17], Laura E. McCoy[1] & Ravindra K. Gupta[2,3,18] ✉

The spike protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is critical for virus infection through the engagement of the human ACE2 protein[1] and is a major antibody target. Here we show that chronic infection with SARS-CoV-2 leads to viral evolution and reduced sensitivity to neutralizing antibodies in an immunosuppressed individual treated with convalescent plasma, by generating whole-genome ultra-deep sequences for 23 time points that span 101 days and using in vitro techniques to characterize the mutations revealed by sequencing. There was little change in the overall structure of the viral population after two courses of remdesivir during the first 57 days. However, after convalescent plasma therapy, we observed large, dynamic shifts in the viral population, with the emergence of a dominant viral strain that contained a substitution (D796H) in the S2 subunit and a deletion (ΔH69/ΔV70) in the S1 N-terminal domain of the spike protein. As passively transferred serum antibodies diminished, viruses with the escape genotype were reduced in frequency, before returning during a final, unsuccessful course of convalescent plasma treatment. In vitro, the spike double mutant bearing both ΔH69/ΔV70 and D796H conferred modestly decreased sensitivity to convalescent plasma, while maintaining infectivity levels that were similar to the wild-type virus.The spike substitution mutant D796H appeared to be the main contributor to the decreased susceptibility to neutralizing antibodies, but this mutation resulted in an infectivity defect. The spike deletion mutant ΔH69/ΔV70 had a twofold higher level of infectivity than wild-type SARS-CoV-2, possibly compensating for the reduced infectivity of the D796H mutation. These data reveal strong selection on SARS-CoV-2 during convalescent plasma therapy, which is associated with the emergence of viral variants that show evidence of reduced susceptibility to neutralizing antibodies in immunosuppressed individuals.

A septuagenarian male was admitted to a tertiary hospital in the summer of 2020 and had tested positive for SARS-CoV-2 using reverse-transcription quantitative PCR (RT–qPCR) 35 days previously in a nasopharyngeal swab (day 1) at a local hospital (Extended Data Figs. 1, 2). His past medical history included 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 combined immunodeficiency of B and T cells (Extended Data Figs. 2, 3 and Supplementary Table 1). Computed tomography of the chest showed widespread abnormalities consistent with pneumonia associated with coronavirus disease 2019 (COVID-19) (Supplementary Fig. 1). Treatment included two 10-day courses of remdesivir with a 5-day gap in between (Extended Data Fig. 1). Two units of convalescent plasma were administered on days 63 and 65 (Extended Data Fig. 3). After clinical deterioration, remdesivir and a unit of convalescent plasma were administered on day 95, but the individual died on day 102 (Supplementary Note).

Comparative analysis of sequential virus samples

Most samples were respiratory samples from the nose and throat or endotracheal aspirates during the period of intubation

A list of affiliations appears at the end of the paper.

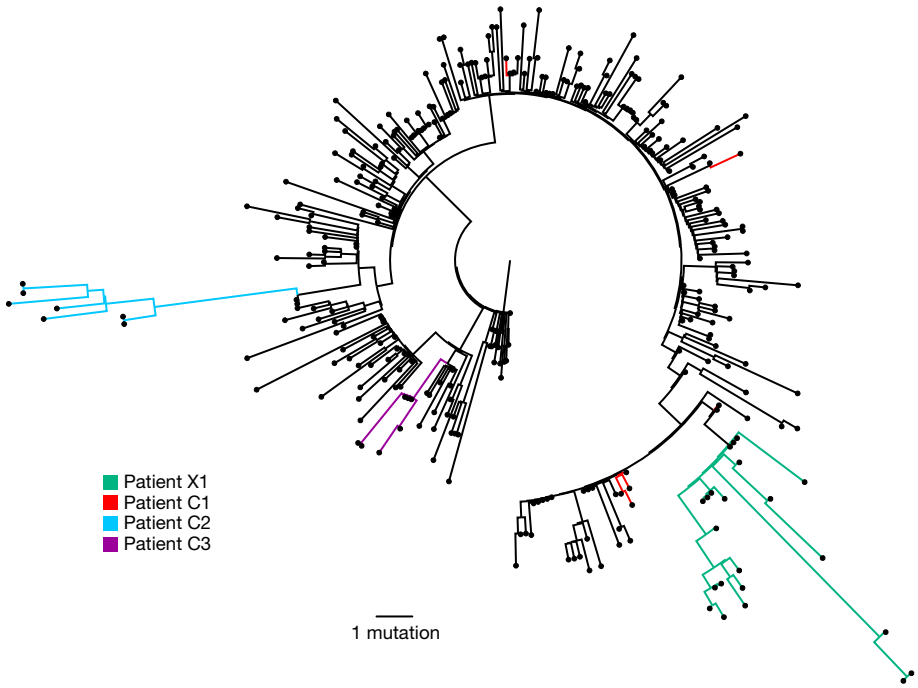


Fig. 1 | Analysis of 23 patient-derived whole-genome sequences of SARS-CoV-2 in the context of local sequences and other patients with chronic SARS-CoV-2 shedding. The circularized maximum-likelihood phylogenetic tree is rooted using the Wuhan-Hu-1 reference sequence, showing a subset of 250 local SARS-CoV-2 genomes from GISAID. This diagram highlights the considerable diversity of patient X1 (green) compared with three other local patients with prolonged shedding (patients C1–C3 are shown as red, blue and purple sequences, respectively). All 'United Kingdom/English' SARS-CoV-2 genomes were downloaded from the GISAID database and a random subset of 250 sequences were selected as background (shown in black).

(Supplementary Table 3). Cycle threshold (C_t) values ranged from 16 to 34 and all 23 respiratory samples were successfully sequenced by a standard single-molecule sequencing approach as per the ARTIC protocol implemented by The COVID-19 Genomics UK (COG-UK; https://www.cogconsortium.uk/) Consortium; of these samples, 20 additionally underwent short-read deep sequencing using the Illumina platform (Supplementary Table 4). There was general agreement between the two methods (Extended Data Fig. 4). However, owing 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 the observed mutations (Extended Data Fig. 4). Finally, we detected no evidence of recombination, based on two independent methods (data not shown).

Maximum-likelihood analysis of patient-derived whole-genome consensus sequences demonstrated clustering with other local sequences from the same region (Fig. 1). The infecting strain was assigned to lineage 20B bearing the spike(D614G) variant. Environmental sampling showed evidence of virus on surfaces such as the telephone and call bell. Sequencing of these surface viruses showed clustering with those derived from the respiratory tract (Extended Data Fig. 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 recently reported long-term immunosuppressed individuals with SARS-CoV-2[4,5] (Extended Data Fig. 2 and Supplementary Table 2). Whereas the sequences from the three local patients as well as one of the previous studies[5] showed little divergence with no amino acid changes in the spike protein over time, the case patient (hereafter, patient X1) showed considerable diversification. The other, previously reported patient[4] showed a similar degree of diversification as patient X1. Further investigation of the sequence data suggested the existence of an underlying structure to the viral population in patient X1, with samples collected at days 93 and 95 being rooted within, but divergent from, the original population (Extended Data Figs. 5, 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 for SARS-CoV-2 by RT–qPCR and there was no sustained change in C_t values throughout the 101 days after 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 Fig. 3). Notably, we were not able to culture virus from stored swab samples. Consensus sequences from short-read deep-sequencing Illumina data revealed dynamic population changes after day 65, as shown in a highlighter plot (Extended Data Fig. 6). In addition, we were also able to follow the dynamics of virus populations down to low frequencies during the entire period (Fig. 2 and Supplementary Table 4). After treatment with remdesivir on day 41, the low frequency variant analysis showed transient amino acid changes in populations at below 50% abundance in open reading frame (ORF) 1b, 3a and spike, with a C27509T (causing a T39I substitution) mutation in *orf7a* reaching 79% on day 45 (Fig. 2, and Supplementary Information). We found that an I513T substitution (encoded by T2343C) in NSP2 and a V157L substitution (encoded by G13936T) in RdRp had emerged from undetectable at day 54 to almost 100% frequency on day 66 (Fig. 2), with the mutation in the polymerase being the more plausible candidate for driving this sweep. Notably, spike(N501Y)—which can increase the affinity for the ACE2 receptor[6] and which is present in the UK B1.1.7 lineage[7]—was observed on day 55 at a frequency of 33%, but was eliminated by the sweep of the NSP2 and RdRp variant.

In contrast to the early period of infection, between days 66 and 82, after the first two administrations of convalescent sera, a shift in the virus population was observed, with a variant bearing a D796H substitution in S2 and a ΔH69/ΔV70 deletion in the S1 N-terminal domain (NTD) of spike becoming the dominant population at day 82. This was identified in a nose and throat swab sample with high viral load as indicated by a C_t value of 23 (Fig. 3a). The deletion was detected transiently at baseline on the basis of to short-read deep sequencing. The ΔH69/ΔV70 deletion was due to an out-of-frame six-nucleotide deletion, which results in the sequence of codon 68 changing from ATA to ATC.

On days 86 and 89, viruses obtained from upper respiratory tract samples were characterized by the spike(Y200H, T240I) double mutant, with the mutation–deletion pair (spike(D796H, ΔH69/ΔV70)) observed on day 82 having decreased to frequencies of 10% or less (Figs. 2, 3).

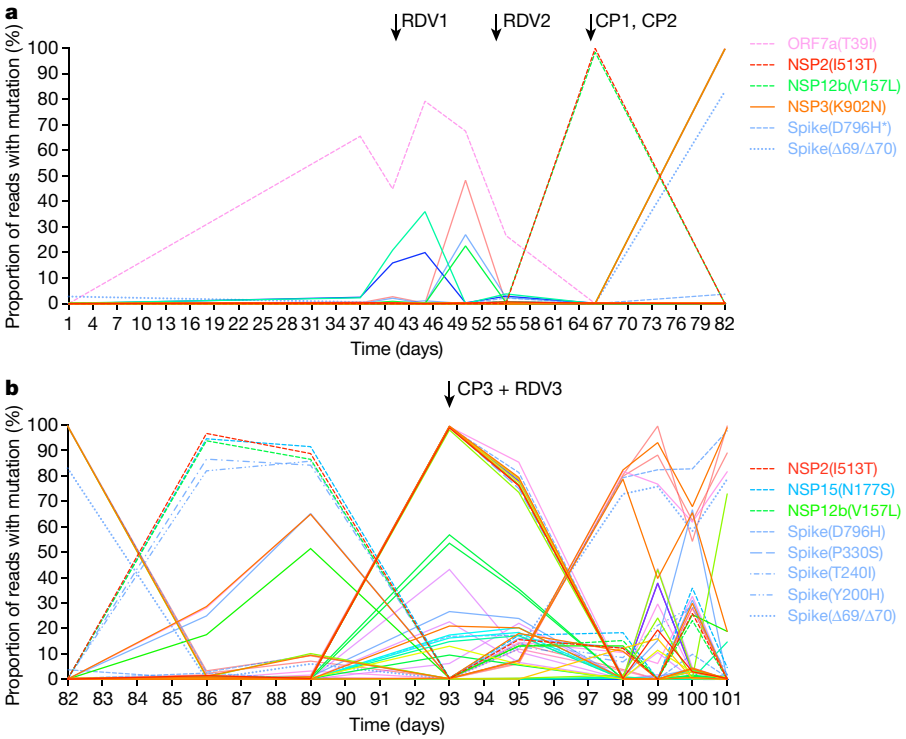


Fig. 2 | Whole-genome variant trajectories showing amino acid changes and the relationship to treatments. Data are based on Illumina short-read ultra-deep sequencing values at 1,000× coverage. Variants shown reached a frequency of at least 10% in at least two samples. CP, convalescent plasma; RDV, remdesivir. Amino acid substitutions described in the text are indicated by labels using the same colouring as the position in the genome (associated proteins are included for each name). Labelled variants are represented by dashed lines. **a**, Variants detected in patient X1 on days 1–82 after the first positive RT–qPCR test for SARS-CoV-2. Spike(D796H*) (light blue) has the same frequency as NSP3(K902N) (orange) and is therefore hidden beneath the orange line. **b**, Variants detected in patient X1 on days 82–101.

Spike(Y200H, T240I) were accompanied at high frequency by three other nonsynonymous variants with similar allele frequencies, encoding I513T in NSP2, V157L in RdRp and N177S in NSP15 (Fig. 2a). Both of these were also previously observed at a frequency of more than 98% in the sample on day 66 (Fig. 2a), suggesting that this new lineage emerged out of a population that existed previously.

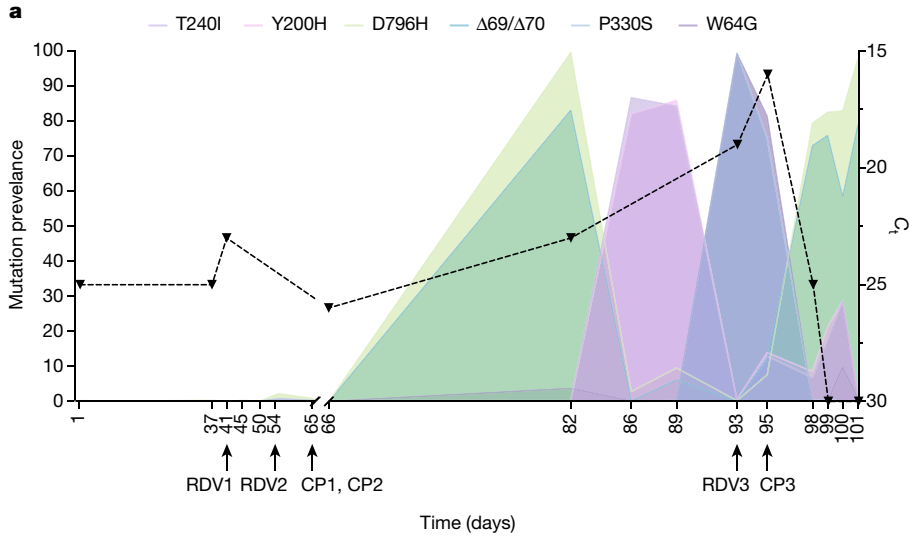
Sequencing of a nose and throat swab sample on day 93 identified viruses characterized by spike(P330S) at the edge of the receptor-binding domain (RBD) and spike(W64G) in the S1 NTD at close to 100% abundance, whereas spike(D796H, ΔH69/ΔV70) showed an abundance of less than 1% and the variants spike(Y200H) and spike(T240I) had frequencies of less than 2%. Viruses with the spike(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 (Figs. 2, 3b and Extended Data Figs. 5, 6) suggests the possibility that they represent a compartmentalized subpopulation.

Patterns in the variant frequencies suggest competition between virus populations carrying different mutations; viruses with the mutation–deletion pair spike(D796H, ΔH69/ΔV70) rose to high frequency during convalescent plasma therapy, but were then 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, which was dominant on day 66 but was outcompeted during therapy by the mutation–deletion variant. With the lapse in therapy, the original strain—which had acquired NSP15(N1773S) and the spike(Y200H, T240I)—regained dominance, followed by the emergence of a separate population with the spike(W64G, P330S) variant.

In a final attempt to reduce the viral load, a third course of remdesivir (day 93) and a third dose of convalescent plasma (day 95) were administered. We observed the re-emergence of the spike(D796H, ΔH69/ΔV70) viral population (Figs. 2, 3). The inferred linkage of spike(D796H) and spike(ΔH69/ΔV70) was maintained as evidenced by the highly similar frequencies of the two variants, suggesting that the third unit of convalescent plasma led to the re-emergence of this population under renewed positive selection. In further support of our proposed idea of competition, frequencies of these two variants appeared to mirror changes in the NSP2(I513T) variant (Fig. 2), suggesting that these variants are markers of opposing clades in the viral population. C_t values remained low throughout this period with hyperinflammation, which eventually led to multi-organ failure and death on day 102. The repeated increase in the frequency of the viral population with convalescent plasma therapy strongly supports the hypothesis that the combination of the deletion and mutation in the spike protein conferred a selective advantage.

Spike mutations impair neutralizing antibody potency

Using lentiviral pseudotyping we generated wild-type spike, spike(D796H, ΔH69/ΔV70) and single mutant spike proteins in enveloped virions to measure the neutralization activity of convalescent plasma against these viruses (Fig. 4). This system has been shown to give generally similar results to replication-competent viruses[8,9]. Spike protein from each mutant was detected in pelleted virions (Fig. 4a). We also used an HIV-1(p24) antibody to monitor the levels of lentiviral particle production (Fig. 4a and Supplementary Fig. 2). We then measured the infectivity of the pseudoviruses, correcting for virus input using measurements of the reverse transcriptase activity, and found that spike(ΔH69/ΔV70) appeared to have twofold higher infectivity over a single round of infection compared to wild-type spike (Fig. 4b and Extended Data Fig. 7). By contrast, the spike(D796H) single mutant had significantly lower infectivity compared to the wild-type spike protein



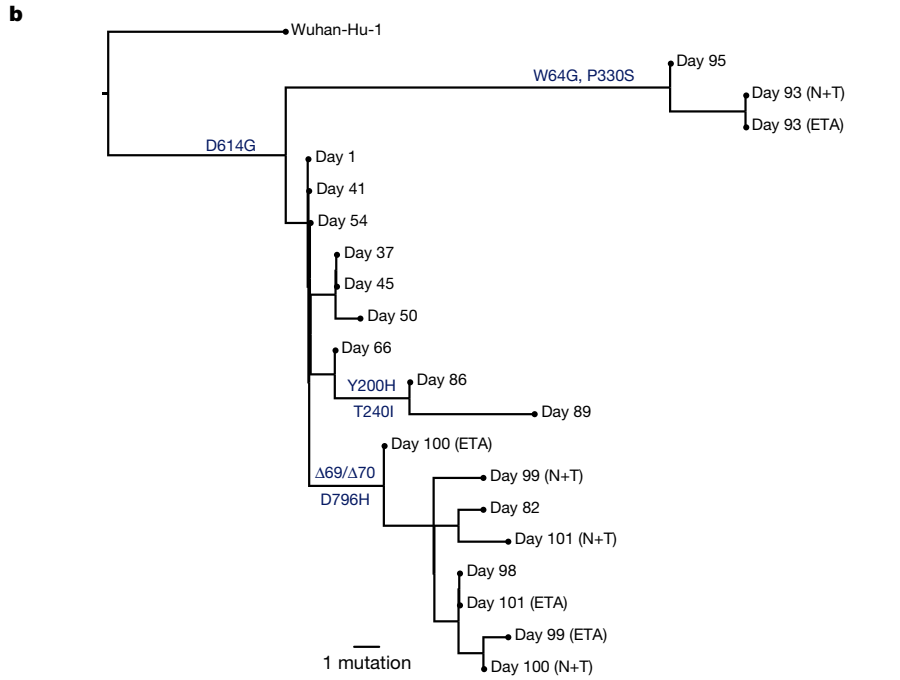


Fig. 3 | Frequencies of longitudinal variants and phylogenetic relationships for virus populations bearing six mutations in the spike protein. a, At baseline, all six spike variants (Illumina sequencing) except for spike(ΔH69/ΔV70) were absent (less than 1% and fewer than 20 reads). Approximately two weeks after receiving two units of convalescent plasma, viral populations carrying spike(ΔH69/ΔV70) and spike(D796H) increased to frequencies of more than 80% but decreased significantly 4 days later. This population was replaced by a population bearing spike(Y200H) and spike(T240I), detected in two samples over a period of 6 days. These viral populations were then replaced by virus carrying spike(W64G) and spike(P330S), which both dominated at day 93. Following a third course of remdesivir and an additional unit of convalescent plasma, the spike(ΔH69/ΔV70) and spike(D796H) virus population re-emerged to become the dominant viral strain reaching variant frequencies of more than 75%. Pairs of mutations arose and disappeared simultaneously, indicating linkage on the same viral haplotype. C_t values from respiratory samples are indicated on the right *y* axis (black dashed line and triangles). In cases in which there were duplicate readings on the same day, to remain consistent, samples from nose and throat swabs were plotted. **b**, Maximum-likelihood phylogenetic tree of patient X1 with the day of sampling indicated. Spike mutations defining each of the clades are shown ancestrally on the branches on which they arose. On dates for which multiple samples were collected, these are indicated as endotracheal aspirate (ETA) and nose and throat swabs (N+T).

and the double mutant (spike(D796H, ΔH69/ΔV70)) had a similar infectivity to wild-type spike (Fig. 4b and Extended Data Fig. 7).

We found that spike(D796H) alone and the spike(D796H, ΔH69/ΔV70) double mutant were less sensitive to neutralization by convalescent plasma samples (Fig. 4c–e and Extended Data Fig. 7). By contrast, the spike(ΔH69/ΔV70) single mutant did not reduce neutralization sensitivity. In addition, patient-derived serum from days 64 and 66 (one day before and after the infusion with the second convalescent plasma therapy) similarly showed lower potency against the spike(D796H, ΔH69/ΔV70) double mutant (Fig. 4f, g).

A panel of nineteen monoclonal antibodies isolated from three donors was previously identified to neutralize SARS-CoV-2. To establish whether the mutations that occurred in vivo (spike(D796H) and spike(ΔH69/ΔV70)) resulted in a global change in neutralization sensitivity, we tested neutralizing monoclonal antibodies targeting the seven major epitope clusters that have previously been described (excluding the non-neutralizing clusters II and V and the small ($n \leq 2$) neutralizing clusters IV and X). The eight RBD-specific monoclonal antibodies (Extended Data Fig. 8) exhibited no major change in neutralization potency and the non-RBD-specific antibody COVA1-21 showed a 3–5-fold reduction in potency against spike(D796H, ΔH69/ΔV70) and spike(ΔH69/ΔV70), but not against spike(D796H) alone[9] (Extended Data Fig. 8). We observed no differences in neutralization between single or double mutants and the wild-type spike protein, suggesting that the mechanism of escape was probably outside of these epitopes in the RBD. These data confirm the specificity of the findings from convalescent plasma and suggest that the observed mutations are related to antibodies that target regions outside of the RBD. Notably, spike(ΔH69/ΔV70)-containing viruses showed reduced neutralization sensitivity to the monoclonal antibody COVA1-21, targeting an as-yet undefined epitope outside of the RBD[10].

To understand how spike(ΔH69/ΔV70) and spike(D796H) might confer antibody resistance, we assessed how these mutations might affect the spike structure (Extended Data Fig. 9). We based this analysis primarily on a structure that lacked stabilizing modifications (PDB 6XR8)[11], but also referred to stabilized structures determined at different pH values[12]. ΔH69/ΔV70 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 Fig. 9). As this loop is flexible and highly accessible, ΔH69/ΔV70 could in principle affect antibody binding in this region. D796 is located near the base of

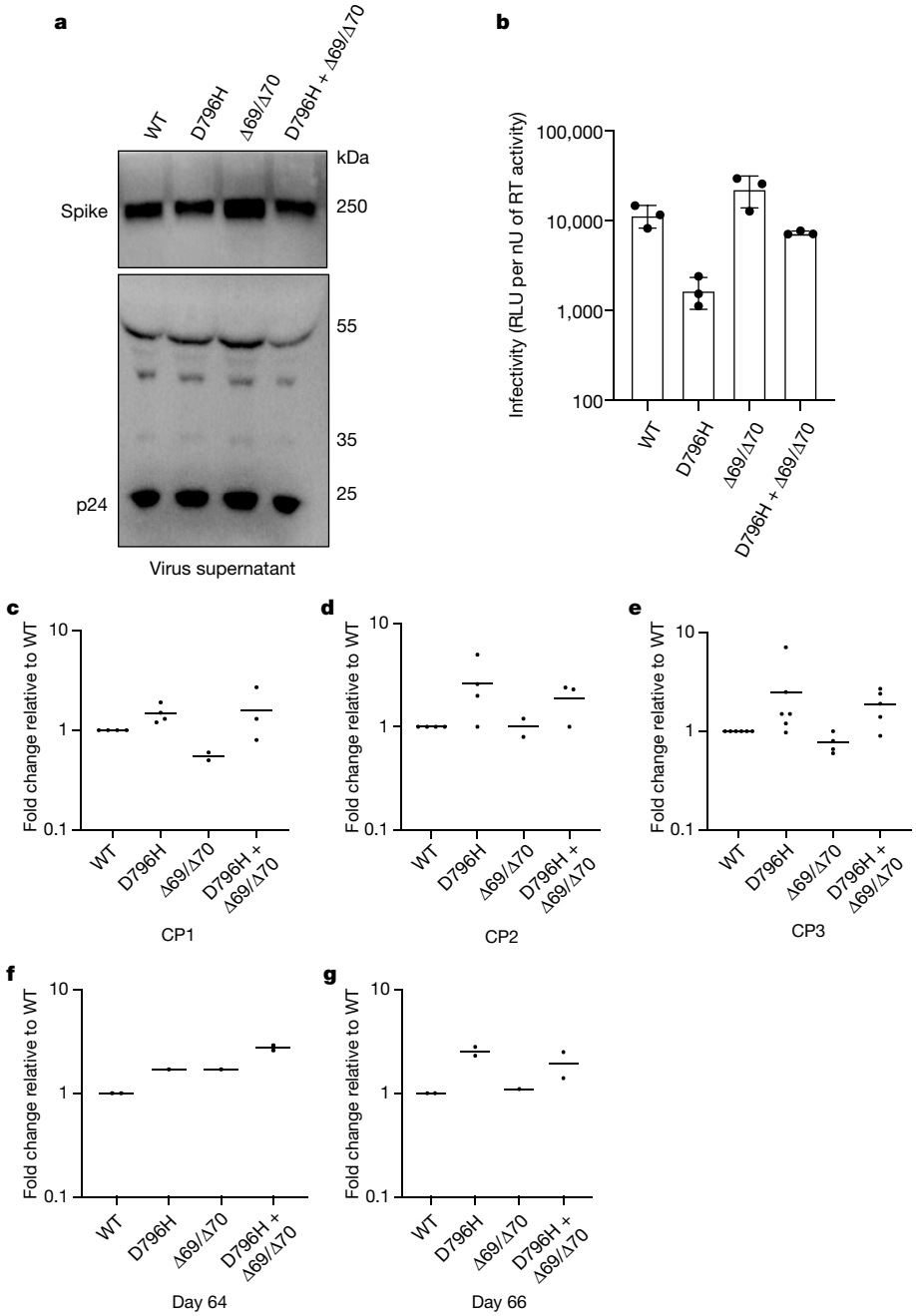


Fig. 4 | Infectivity of the spike(D796H, ΔH69/ΔV70) variant and sensitivity to convalescent plasma treatment. a, Western blot of virus pellets after centrifugation of supernatants from cells transfected with lentiviral pseudotyping plasmids that included the spike protein. Blots are representative of two independent transfections. **b**, Single-round infectivity of luciferase-expressing lentiviruses pseudotyped with the spike protein (wild-type (WT) or mutant) of SARS-CoV-2 in HEK293T cells co-transfected with ACE2 and TMPRSS2 plasmids. Infectivity is corrected for reverse transcriptase activity in the virus supernatant as measured by qPCR. Data points represent technical replicates ($n = 3$). Data are mean ± s.e.m. of two independent experiments. RLU, relative light units; nU, nanounit. **c**–**e**, Neutralization potency of convalescent plasma units (CP1–CP3) against pseudotyped viruses bearing spike(D796H), spike(ΔH69/ΔV70) and spike(D796H, ΔH69/ΔV70). **f**, **g**, Neutralizing potency of the serum of patient X1 against pseudotyped virus bearing spike(D796H), spike(ΔH69/ΔV70) and spike(D796H, ΔH69/ΔV70). Patient serum was taken on the indicated days. The serum dilution required to inhibit 50% of virus infection (ID_{50}) is shown, expressed as a fold change relative to the wild-type virus. Data points represent means of technical replicates (horizontal bars) obtained an independent experiments ($n = 2$–6).

the spike protein, in a surface loop that is structurally disordered in the prefusion conformation and becomes part of a large disordered region in the post-fusion S2 trimer[11] (Extended Data Fig. 9). The loop containing residue 796 is proposed to be targeted by antibodies[15], despite mutations at position 796 being relatively uncommon (Extended Data Fig. 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 Fig. 9).

Discussion

Here we 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 the twofold 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 noted in our analysis signs of compartmentalized 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 characterized by spatially distinct viral populations[16,17]. In the analysis of data, it is important to distinguish genetic changes that occur in the primary viral population from apparent changes that arise from the stochastic observation of spatially distinct subpopulations in the host. Although the samples that we obtained on days 93 and 95 of infection are genetically distinct from the other samples, the remaining samples are consistent with arising from a consistent viral population. We note that detection of viral RNA in post-mortem tissue was observed previously 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 the

movement of secretions from lower to upper respiratory tract—could lead to fluctuations in viral populations at particular sampling sites.

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

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 patient.

In addition to documenting the emergence of SARS-CoV-2 spike(ΔH69/ΔV70) in vivo, we show that this variant modestly increases infectivity of the spike protein in a pseudotyping assay. The deletion was observed contemporaneously with the rare S2 variant spike(D796H) after two separate courses of convalescent plasma, with other viral populations emerging. D796H, but not ΔH69/ΔV70, conferred a reduction in susceptibility to polyclonal antibodies in the units of convalescent plasma administered, although we cannot speculate as to their individual effect on sera from other individuals. It is notable that the spike(D796H, ΔH69/ΔV70) double mutant decreased in prevalence in between convalescent plasma courses, suggesting that there were other selective forces at play in the intervening period, which are possibly driven by the inflammation observed in the individual. This includes the possibility that the haplotype comprising spike (D796H, ΔH69/ΔV70) may have carried mutations in other regions that were deleterious during that intervening period. Although ΔH69/ΔV70 is expanding at a high rate[18], D796 mutations are also increasing. D796H has been documented in 0.02% of global sequences and D796Y appears in 0.05% of global sequences (Extended Data Fig. 9).

The effects of convalescent plasma on virus evolution found here are unlikely to apply in immunocompetent hosts in whom viral diversity is likely to be lower owing to better immune control. Our data highlight that infection control measures may need to be tailored to the needs of immunocompromised patients and also caution in the interpretation of guidelines from the US Centres of Disease Control that recommend 20 days as the upper limit of infection-prevention precautions in immunocompromised patients who are afebrile[19]. Owing to the difficulty of culturing clinical isolates, use of surrogate experiments are warranted[20]. However, in cases in which the detection of ongoing viral evolution is possible, this serves as a clear proxy for the existence of infectious virus. In our case, we detected environmental contamination while in a single-occupancy room and the patient was moved to a negative-pressure high air-change infectious-disease isolation room.

The clinical efficacy of convalescent plasma in patients with severe COVID-19 has not been demonstrated[21], and its use in different stages of infection and disease remains experimental; as such, we suggest that it should be reserved for use within clinical trials, and with rigorous monitoring of clinical and virological parameters. The data from this single patient indicate caution should be used for convalescent plasma therapy in patients with immunosuppression of both T cell and B cell arms; in these patients, the administered antibodies have little support from cytotoxic T cells, thereby reducing the chances of clearance and theoretically raising the potential for the evolution of escape mutations in SARS-CoV-2. Although we await further data, in cases in which clinical trial enrolment is not possible, convalescent plasma administered for clinical need in immunosuppressed individuals should ideally only be considered as part of observational studies, undertaken preferably in single-occupancy rooms with enhanced infection-control precautions, including environmental sampling of SARS-CoV-2 and real-time sequencing. Understanding the viral dynamics and characterizing the viral evolution in response to different selection pressures in an immunocompromised individual are necessary not only for improved patient management but also for public health benefit.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03291-y.

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 (2020).
2. Kim, K. W. et al. Respiratory viral co-infections among SARS-CoV-2 cases confirmed by virome capture sequencing. Preprint at https://doi.org/10.21203/rs.3.rs-105996/v1 (2020).
3. Bull, R. A. et al. Analytical validity of Nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat. Commun.* **11**, 6272 (2020).
4. Choi, B. et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N. Engl. J. Med.* **383**, 2291–2293 (2020).
5. Avanzato, V. A. et al. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised cancer patient. *Cell* **183**, 1901–1912 (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 (2020).
7. Rambaut, A. et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Preprint at 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. *J. Exp. Med.* **217**, e20201181 (2020).
9. Brouwer, P. J. M. et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643–650 (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 (2012).
11. Cai, Y. et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586–1592 (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 (2020).
13. Robbiani, D. F. et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (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 (2020).
15. Shrock, E. et al. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* **370**, eabd4250 (2020).
16. Sobel Leonard, A. et al. The effective rate of influenza reassortment is limited during human infection. *PLoS Pathog.* **13**, e1006203 (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**, e02063-17 (2018).
18. Kemp, S. A. et al. Recurrent emergence and transmission of a SARS-CoV-2 spike deletion H69/V70. Preprint at https://doi.org/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. Preprint at https://doi.org/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.* https://doi.org/10.1056/NEJMoa2031304 (2020).

[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, UK. [6]NHS Blood and Transplant, Oxford and BRC Haematology Theme, University of Oxford, Oxford, UK. [7]Viral Pseudotype Unit, Medway School of Pharmacy, University of Kent, Canterbury, UK. [8]Medical Research Council Laboratory of Molecular Biology, Cambridge, UK. [9]Department of Medical Microbiology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. [10]NIHR Cambridge Bioresource, Cambridge, UK. [11]Department of Virology, Cambridge University NHS Hospitals Foundation Trust, Cambridge, UK. [12]Department of Clinical Biochemistry and Immunology, Addenbrooke's Hospital, Cambridge, UK. [13]FES-Cuautitlán, UNAM, Cuautitlán Izcalli, Mexico. [14]Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO, USA. [15]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK. [16]Clinical Microbiology and Public Health Laboratory, Addenbrooke's Hospital, Cambridge, UK. [17]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK. [18]Africa Health Research Institute, Durban, South Africa. [148]These authors contributed equally: Steven A. Kemp, Dami A. Collier, Rawlings P. Datir. *Lists of authors and their affiliations appear online. ✉e-mail: rkg20@cam.ac.uk

The CITIID-NIHR BioResource COVID-19 Collaboration

Principal investigators
Stephen Baker[2,3], Gordon Dougan[2,3], Christoph Hess[2,3,19,20], Nathalie Kingston[15,21], Paul J. Lehner[2,3], Paul A. Lyons[2,3], Nicholas J. Matheson[2,3], Willem H. Owehand[21], Caroline Saunders[22], Charlotte Summers[3,22,23,24], James E. D. Thaventhiran[2,3,25], Mark Toshner[3,23,24] & Michael P. Weekes[2]

CRF & volunteer research nurses
Ashlea Bucke[22], Jo Calder[22], Laura Canna[22], Jason Domingo[22], Anne Elmer[22], Stewart Fuller[22], Julie Harris[26], Sarah Hewitt[22], Jane Kennet[22], Sherly Jose[22], Jenny Kourampa[22], Anne Meadows[22], Criona O'Brien[26], Jane Price[22], Cherry Publico[22], Rebecca Rastall[22], Carla Ribeiro[22], Jane Rowlands[22], Valentina Ruffolo[22] & Hugo Tordesillas[22]

Sample logistics
Ben Bullman[2], Benjamin J. Dunmore[3], Stuart Fawke[27], Stefan Gräf[3,15,21], Josh Hodgson[3], Christopher Huang[3], Kelvin Hunter[2,3], Emma Jones[28], 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[29]

Sample processing & 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[30], Nick Gleadall[21], Richard Grenfell[31], Andrew Hinch[2,3], Oisin Huhn[30], Sarah Jackson[3], Isobel Jarvis[3], Daniel Lewis[3], Joe Marsden[3], Francesca Nice[32], 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[21], Jonathan Stephens[21], Mateusz Strezlecki[31] & Lori Turner[2,3]

Clinical data collection
Eckart M. D. D. De Bie[3], Katherine Bunclark[3], Masa Josipovic[33], Michael Mackay[3], Sabrina Rossi[34], Mayurun Selvan[3], Sarah Spencer[14] & Cissy Yong[34]

Royal Papworth Hospital ICU
Ali Ansaripour[24], Alice Michael[24], Lucy Mwaura[24], Caroline Patterson[24] & Gary Polwarth[24]

Addenbrooke's Hospital ICU
Petra Polgarova[22] & Giovanni di Stefano[22]

Cambridge & Peterborough Foundation Trust
Codie Fahey[35] & Rachel Michel[35]

ANPC & Centre for Molecular Medicine & Innovative Therapeutics
Sze-How Bong[36], Jerome D. Coudert[37] & Elaine Holmes[38]

NIHR BioResource
John Allison[15,21], Helen Butcher[15,39], Daniela Caputo[15,39], Debbie Clapham-Riley[15,39], Eleanor Dewhurst[15,39], Anita Furlong[15,39], Barbara Graves[15,39], Jennifer Gray[15,39], Tasmin Ivers[15,39], Mary Kasanicki[15,22], Emma Le Gresley[15,39], Rachel Linger[15,39], Sarah Meloy[15,39], Francesca Muldoon[15,39], Nigel Ovington[15,21], Sofia Papadia[15,39], Isabel Phelan[15,39], Hannah Stark[15,39], Kathleen E. Stirrups[15,21], Paul Townsend[15,21], Neil Walker[15,21] & Jennifer Webster[15,39]

[19]Department of Biomedicine, University and University Hospital Basel, Basel, Switzerland. [20]Botnar Research Centre for Child Health (BRCCH) University Basel & ETH Zurich, Basel, Switzerland. [21]Department of Haematology, University of Cambridge, Cambridge, UK. [22]Cambridge Clinical Research Centre, NIHR Clinical Research Facility, Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Cambridge, UK. [23]Heart and Lung Research Institute, Cambridge, UK. [24]Royal Papworth Hospital NHS Foundation Trust, Cambridge, UK. [25]MRC Toxicology Unit, School of Biological Sciences, University of Cambridge, Cambridge, UK. [26]Department of Paediatrics, University of Cambridge, Cambridge, UK. [27]Cambridge Institute for Medical Research, Cambridge, UK. [28]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK. [29]Department of Biochemistry, University of Cambridge, Cambridge, UK. [30]Department of Obstetrics & Gynaecology, The Rosie Maternity Hospital, Cambridge, UK. [31]Cancer Research UK, Cambridge Institute, University of Cambridge, Cambridge, UK. [32]Cancer Molecular Diagnostics Laboratory, Department of Oncology, University of Cambridge, Cambridge, UK. [33]Metabolic Research Laboratories, Wellcome Trust-Medical Research Council Institute of Metabolic Science, University of Cambridge, Cambridge, UK. [34]Department of Surgery, Addenbrooke's Hospital, Cambridge, UK. [35]Cambridge and Peterborough Foundation Trust, Fulbourn Hospital, Fulbourn, Cambridge, UK. [36]Australian National Phenome Centre, Murdoch University, Murdoch, Western Australia, Australia. [37]Centre for Molecular Medicine and Innovative Therapeutics, Health Futures Institute, Murdoch University, Perth, Western Australia, Australia. [38]Centre of Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth, Western Australia, Australia. [39]Department of Public Health and Primary Care, School of Clinical Medicine, University of Cambridge, Cambridge, UK.

The COVID-19 Genomics UK (COG-UK) Consortium

Funding acquisition, leadership & supervision, metadata curation, project administration, samples & logistics, sequencing & analysis, software & analysis tools and visualization
Samuel C. Robson[40]

Funding acquisition, leadership & supervision, metadata curation, project administration, samples & logistics, sequencing & analysis and software & analysis tools
Nicholas J. Loman[41] & Thomas R. Connor[42,43]

Leadership & supervision, metadata curation, project administration, samples & logistics, sequencing & analysis, software & analysis tools and visualization
Tanya Golubchik[44]

Funding acquisition, metadata curation, samples & logistics, sequencing & analysis, software & analysis tools and visualization
Rocio T. Martinez Nunez[45]

Funding acquisition, leadership & supervision, metadata curation, project administration and samples & logistics
Catherine Ludden[46]

Funding acquisition, leadership & supervision, metadata curation, samples & logistics and sequencing & analysis
Sally Corden[43]

Funding acquisition, leadership & supervision, project administration, samples & logistics and sequencing & analysis
Ian Johnston[47] & David Bonsall[44]

Funding acquisition, leadership & supervision, sequencing & analysis, software & analysis tools and visualization
Colin P. Smith[48] & Ali R. Awan[49]

Funding acquisition, samples & logistics, sequencing & analysis, software & analysis tools and visualization
Giselda Bucca[48]

Leadership & supervision, metadata curation, project administration, samples & logistics and sequencing & analysis
M. Estee Torok[3,50,51]

Leadership & supervision, metadata curation, project administration, samples & logistics and visualization
Kordo Saeed[52,53] & Jacqui A. Prieto[52,54]

Leadership & supervision, metadata curation, project administration, sequencing & analysis and software & analysis tools
David K. Jackson[47]

Metadata curation, project administration, samples & logistics, sequencing & analysis and software & analysis tools
William L. Hamilton[50,51]

Metadata curation, project administration, samples & logistics, sequencing & analysis and visualization
Luke B. Snell[55]

Funding acquisition, leadership & supervision, metadata curation and samples & logistics
Catherine Moore[43]

Funding acquisition, leadership & supervision, project administration and samples & logistics
Ewan M. Harrison[46,47]

Leadership & supervision, metadata curation, project administration and samples & logistics
Sonia Goncalves[47]

Leadership & supervision, metadata curation, samples & logistics and sequencing & analysis
Derek J. Fairley[56,57], Matthew W. Loose[58] & Joanne Watkins[43]

Leadership & supervision, metadata curation, samples & logistics and software & analysis tools
Rich Livett[47]

Leadership & supervision, metadata curation, samples & logistics and visualization
Samuel Moses[59,60]

Leadership & supervision, metadata curation, sequencing & analysis and software & analysis tools
Roberto Amato[47], Sam Nicholls[41] & Matthew Bull[43]

Leadership & supervision, project administration, samples & logistics and sequencing & analysis
Darren L. Smith[61,62,63]

Leadership & supervision, sequencing & analysis, software & analysis tools and visualization
Jeff Barrett[47] & David M. Aanensen[64]

Metadata curation, project administration, samples & logistics and sequencing & analysis
Martin D. Curran[65], Surendra Parmar[65], Dinesh Aggarwal[46,47,66] & James G. Shepherd[67]

Metadata curation, project administration, sequencing & analysis and software & analysis tools
Matthew D. Parker[68]

Metadata curation, samples & logistics, sequencing & analysis and visualization
Sharon Glaysher[69]

Metadata curation, sequencing & analysis, software & analysis tools and visualization
Matthew Bashton[61,62], Anthony P. Underwood[64], Nicole Pacchiarini[43] & Katie F. Loveson[70]

Project administration, sequencing & analysis, software & analysis tools and visualization
Alessandro M. Carabelli[46]

Funding acquisition, leadership & supervision and metadata curatio
Kate E. Templeton[71,72]

Funding acquisition, leadership & supervision and project administration
Cordelia F. Langford[47], John Sillitoe[47], Thushan I. de Silva[68] & Dennis Wang[68]

Funding acquisition, leadership & supervision and sequencing & analysis
Dominic Kwiatkowski[47,73], Andrew Rambaut[72], Justin O'Grady[74,75] & Simon Cottrell[43]

Leadership & supervision, metadata curation and sequencing & analysis
Matthew T. G. Holden[76] & Emma C. Thomson[67]

Leadership & supervision, project administration and samples & logistics
Husam Osman[66,77], Monique Andersson[78], Anoop J. Chauhan[69] & Mohammed O. Hassan-Ibrahim[79]

Leadership & supervision, project administration and sequencing & analysis
Mara Lawniczak[47]

Leadership & supervision, samples & logistics and sequencing & analysis
Alex Alderton[47], Meera Chand[66], Chrystala Constantinidou[80], Meera Unnikrishnan[80], Alistair C. Darby[81], Julian A. Hiscox[81] & Steve Paterson[81]

Leadership & supervision, sequencing & analysis and software & analysis tools
Inigo Martincorena[47], David L. Robertson[67], Erik M. Volz[82], Andrew J. Page[74] & Oliver G. Pybus[83]

Leadership & supervision, sequencing & analysis and visualization
Andrew R. Bassett[47]

Metadata curation, project administration and samples & logistics
Cristina V. Ariani[47], Michael H. Spencer Chapman[46,47], Kathy K. Li[67], Rajiv N. Shah[67], Natasha G. Jesudason[67] & Yusri Taha[84]

Metadata curation, project administration and sequencing & analysis
Martin P. McHugh[71] & Rebecca Dewar[71]

Metadata curation, samples & logistics and sequencing & analysis
Aminu S. Jahun[85], Claire McMurray[41], Sarojini Pandey[86], James P. McKenna[56], Andrew Nelson[62,63], Gregory R. Young[61,62], Clare M. McCann[62,63] & Scott Elliott[69]

Metadata curation, samples & logistics and visualization
Hannah Lowe[59]

Metadata curation, sequencing & analysis and software & analysis tools
Ben Temperton[87], Sunando Roy[88], Anna Price[42], Sara Rey[43] & Matthew Wyles[68]

Metadata curation, sequencing & analysis and visualization
Stefan Rooke[72] & Sharif Shaaban[76]

Project administration, samples & logistics and sequencing & analysis
Mariateresa de Cesare[89]

Project administration, samples & logistics and software & analysis tools
Laura Letchford[47]

Project administration, samples & logistics and visualization
Siona Silveira[52], Emanuela Pelosi[52] & Eleri Wilson-Davies[52]

Samples & logistics, sequencing & analysis and software & analysis tools
Myra Hosmillo[85]

Sequencing & analysis, software & analysis tools and visualization
Áine O'Toole[72], Andrew R. Hesketh[48], Richard Stark[80], Louis du Plessis[83], Chris Ruis[46], Helen Adams[90] & Yann Bourgeois[91]

Funding acquisition and leadership & supervision
Stephen L. Michell[87], Dimitris Grammatopoulos[86,92], Jonathan Edgeworth[93], Judith Breuer[88,94], John A. Todd[89] & Christophe Fraser[44]

Funding acquisition and project administration
David Buck[89] & Michaela John[95]

Leadership & supervision and metadata curation
Gemma L. Kay[74]

Leadership & supervision and project administration
Steve Palmer[47], Sharon J. Peacock[46,66] & David Heyburn[43]

Leadership & supervision and samples & logistics
Danni Weldon[47], Esther Robinson[66,77], Alan McNally[41,96], Peter Muir[66], Ian B. Vipond[66], John Boyes[97], Venkat Sivaprakasam[98], Tranprit Salluja[99], Samir Dervisevic[100] & Emma J. Meader[100]

Leadership & supervision and sequencing & analysis
Naomi R. Park[47], Karen Oliver[47], Aaron R. Jeffries[87], Sascha Ott[80], Ana da Silva Filipe[67], David A. Simpson[57] & Chris Williams[43]

Leadership & supervision and visualization
Jane A. H. Masoli[87,101]

Metadata curation and samples & logistics
Bridget A. Knight[87,101], Christopher R. Jones[87,101], Cherian Koshy[102], Amy Ash[102], Anna Casey[103], Andrew Bosworth[66,77], Liz Ratcliffe[103], Li Xu-McCrae[77], Hannah M. Pymont[66], Stephanie Hutchings[66], Lisa Berry[86], Katie Jones[86], Fenella Halstead[98], Thomas Davis[104], Christopher Holmes[105], Miren Iturriza-Gomara[81], Anita O. Lucaci[81], Paul Anthony Randell[106,107], Alison Cox[106,107], Pinglawathee Madona[106,107], Kathryn Ann Harris[94], Julianne Rose Brown[94], Tabitha W. Mahungu[108], Dianne Irish-Tavares[108], Tanzina Haque[108], Jennifer Hart[108], Eric Witele[108], Melisa Louise Fenton[99], Steven Liggett[109], Clive Graham[110], Emma Swindells[111], Jennifer Collins[84], Gary Eltringham[84], Sharon Campbell[112], Patrick C. McClure[113], Gemma Clark[114], Tim J. Sloan[115], Carl Jones[114] & Jessica Lynch[116,117]

Metadata curation and sequencing & analysis
Ben Warne[118], Steven Leonard[47], Jillian Durham[47], Thomas Williams[72], Sam T. Haldenby[81], Nathaniel Storey[94], Nabil-Fareed Alikhan[74], Nadine Holmes[58], Christopher Moore[58], Matthew Carlile[58], Malorie Perry[43], Noel Craine[43], Ronan A. Lyons[119], Angela H. Beckett[40], Salman Goudarzi[70], Christopher Fearn[70], Kate Cook[70], Hannah Dent[70] & Hannah Paul[70]

Metadata curation and software & analysis tools
Robert Davies[47]

Project administration and samples & logistics
Beth Blane[46], Sophia T. Girgis[46], Mathew A. Beale[47], Katherine L. Bellis[46,47], Matthew J. Dorman[47], Eleanor Drury[47], Leanne Kane[47], Sally Kay[47], Samantha McGuigan[47], Rachel Nelson[47], Liam Prestwood[47], Shavanthi Rajatileka[47], Rahul Batra[93], Rachel J. Williams[88], Mark Kristiansen[88], Angie Green[89], Anita Justice[78], Adhyana I. K. Mahanama[52,120] & Buddhini Samaraweera[52,120]

Project administration and sequencing & analysis
Nazreen F. Hadjirin[46] & Joshua Quick[41]

Project administration and software & analysis tools
Radoslaw Poplawski[41]

Samples & logistics and Sequencing & analysis
Leanne M. Kermack[46], Nicola Reynolds[121], Grant Hall[85], Yasmin Chaudhry[85], Malte L. Pinckert[85], Iliana Georgana[85], Robin J. Moll[47], Alicia Thornton[66], Richard Myers[66], Joanne Stockton[41], Charlotte A. Williams[88], Wen C. Yew[62], Alexander J. Trotter[74], Amy Trebes[89], George MacIntyre-Cockett[89], Alec Birchley[43], Alexander Adams[43], Amy Plimmer[43], Bree Gatica-Wilcox[43], Caoimhe McKerr[43], Ember Hilvers[43], Hannah Jones[43], Hibo Asad[43], Jason Coombes[43], Johnathan M. Evans[43], Laia Fina[43], Lauren Gilbert[43], Lee Graham[43], Michelle Cronin[43], Sara Kumziene-Summerhayes[43], Sarah Taylor[43], Sophie Jones[43], Danielle C. Groves[68], Peijun Zhang[68], Marta Gallis[68] & Stavroula F. Louka[68]

Samples & logistics and software & analysis tools
Igor Starinskij[67]

Sequencing & analysis and software & analysis tools
Chris Jackson[17], Marina Gourtovaia[47], Gerry Tonkin-Hill[47], Kevin Lewis[47], Jaime M. Tovar-Corona[47], Keith James[47], Laura Baxter[80], Mohammad T. Alam[80], Richard J. Orton[67], Joseph Hughes[67], Sreenu Vattipally[67], Manon Ragonnet-Cronin[82], Fabricia F. Nascimento[82], David Jorgensen[82], Olivia Boyd[82], Lily Geidelberg[82], Alex E. Zarebski[83], Jayna Raghwani[83], Moritz U. G. Kraemer[83], Joel Southgate[42,43], Benjamin B. Lindsey[68] & Timothy M. Freeman[68]

Software & analysis tools and visualization
Jon-Paul Keatley[47], Joshua B. Singer[67], Leonardo de Oliveira Martins[74], Corin A. Yeats[64], Khalil Abudahab[64], Ben E. W. Taylor[64] & Mirko Menegazzo[64]

Leadership & supervision
John Danesh[47], Wendy Hogsden[98], Sahar Eldirdiri[104], Anita Kenyon[104], Jenifer Mason[122], Trevor I. Robinson[122], Alison Holmes[106,123], James Price[106,123], John A. Hartley[88], Tanya Curran[56], Alison E. Mather[74], Giri Shankar[43], Rachel Jones[43], Robin Howe[43] & Sian Morgan[95]

Metadata curation
Elizabeth Wastenge[71], Michael R. Chapman[46,47,124], Siddharth Mookerjee[106,123], Rachael Stanley[100], Wendy Smith[114], Timothy Peto[78], David Eyre[78], Derrick Crook[78], Gabrielle Vernet[125], Christine Kitchen[42], Huw Gulliver[42], Ian Merrick[42], Martyn Guest[42], Robert Munn[42], Declan T. Bradley[57,126] & Tim Wyatt[126]

Project administration
Charlotte Beaver[47], Luke Foulser[47], Sophie Palmer[46], Carol M. Churcher[46], Ellena Brooks[46], Kim S. Smith[46], Katerina Galai[46], Georgina M. McManus[46], Frances Bolt[106,123], Francesc Coll[127], Lizzie Meadows[74], Stephen W. Attwood[83], Alisha Davies[43], Elen De Lacy[43], Fatima Downing[43], Sue Edwards[43], Garry P. Scarlett[91], Sarah Jeremiah[52] & Nikki Smith[68]

Samples & logistics
Danielle Leek[46], Sushmita Sridhar[46,47], Sally Forrest[46], Claire Cormie[46], Harmeet K. Gill[46], Joana Dias[46], Ellen E. Higginson[46], Mailis Maes[46], Jamie Young[46], Michelle Wantoch[121], Dorota Jamrozy[47], Stephanie Lo[47], Minal Patel[47], Verity Hill[72], Claire M. Bewshea[87], Sian Ellard[87,101], Cressida Auckland[101], Ian Harrison[66], Chloe Bishop[66], Vicki Chalker[66], Alex Richter[128], Andrew Beggs[128], Angus Best[96], Benita Percival[96], Jeremy Mirza[96], Oliver Megram[96], Megan Mayhew[96], Liam Crawford[96], Fiona Ashcroft[96], Emma Moles-Garcia[96], Nicola Cumley[96], Richard Hopes[66], Patawee Asamaphan[67], Marc O. Niebel[67], Rory N. Gunson[129], Amanda Bradley[130], Alasdair Maclean[130], Guy Mollett[130], Rachel Blacow[130], Paul Bird[105], Thomas Helmer[105], Karlie Fallon[105], Julian Tang[105], Antony D. Hale[131], Louissa R. Macfarlane-Smith[131], Katherine L. Harper[131], Holli Carden[131], Nicholas W. Machin[66,132], Kathryn A. Jackson[81], Shazaad S. Y. Ahmad[66,132], Ryan P. George[132], Lance Turtle[81], Elaine O'Toole[122], Joanne Watts[122], Cassie Breen[122], Angela Cowell[122], Adela Alcolea-Medina[133,134], Themoula Charalampous[45,93], Amita Patel[55], Lisa J. Levett[135], Judith Heaney[135], Aileen Rowan[82], Graham P. Taylor[82], Divya Shah[94], Laura Atkinson[94], Jack C. D. Lee[94], Adam P. Westhorpe[88], Riaz Jannoo[88], Helen L. Lowe[88], Angeliki Karamani[88], Leah Ensell[88], Wendy Chatterton[135], Monika Pusok[135], Ashok Dadrah[99], Amanda Symmonds[99], Graciela Sluga[136], Zoltan Molnar[57], Paul Baker[109], Stephen Bonner[109], Sarah Essex[109], Edward Barton[110], Debra Padgett[110], Garren Scott[110], Jane Greenaway[111], Brendan A. I. Payne[84], Shirelle Burton-Fanning[84], Sheila Waugh[84], Veena Raviprakash[112], Nicola Sheriff[112], Victoria Blakey[112], Lesley-Anne Williams[112], Jonathan Moore[137], Susanne Stonehouse[137], Louise Smith[138], Rose K. Davidson[75], Luke Bedford[139], Lindsay Coupland[100], Victoria Wright[58], Joseph G. Chappell[113], Theocharis Tsoleridis[113], Jonathan Ball[113], Manjinder Khakh[114], Vicki M. Fleming[114], Michelle M. Lister[114], Hannah C. Howson-Wells[114], Louise Berry[114], Tim Boswell[114], Amelia Joseph[114], Iona Willingham[114], Nichola Duckworth[115], Sarah Walsh[115], Emma Wise[116,117], Nathan Moore[116,117], Matilde Mori[116,117,140], Nick Cortes[116,117], Stephen Kidd[116,117], Rebecca Williams[125], Laura Gifford[43], Kelly Bicknell[69], Sarah Wyllie[69], Allyson Lloyd[69], Robert Impey[69], Cassandra S. Malone[79], Benjamin J. Cogger[79], Nick Levene[141], Lynn Monaghan[141], Alexander J. Keeley[68], David G. Partridge[68,142], Mohammad Raza[68,142], Cariad Evans[68,142], Kate Johnson[68,142], Irina Abnizova[147], Louise Aigrain[147], Alex Alderton[147], Mozam Ali[147], Laura Allen[147], Roberto Amato[147], Ralph Anderson[147], Cristina Ariani[147], Siobhan Austin-Guest[147], Sendu Bala[147], Jeffrey Barrett[147], Andrew Bassett[147], Kristina Battleday[147], James Beal[147], Mathew Beale[147], Charlotte Beaver[147], Sam Bellany[147], Tristram Bellerby[147], Katie Bellis[147], Duncan Berger[147], Matt Berriman[147], Emma Betteridge[147], Paul Bevan[147], Simon Binley[147], Jason Bishop[147], Kirsty Blackburn[147], James Bonfield[147], Nick Boughton[147], Sam Bowker[147], Timothy Brendler-Spaeth[147], Iraad Bronner[147], Tanya Brooklyn[147], Sarah Kay Buddenborg[147], Robert Bush[147], Catarina Caetano[147], Alex Cagan[147], Nicola Carter[147], Joanna Cartwright[147], Tiago Carvalho Monteiro[147], Liz Chapman[147], Tracey-Jane Chillingworth[147], Peter Clapham[147], Richard Clark[147], Adrian Clarke[147], Catriona Clarke[147], Daryl Cole[147], Elizabeth Cook[147], Maria Coppola[147], Linda Cornell[147], Clare Cornwell[147], Craig Corton[147], Abby Crackett[147], Alison Cranage[147], Harriet Craven[147], Sarah Craw[147], Mark Crawford[147], Tim Cutts[147], Monika Dabrowska[147], Matt Davies[147], Robert Davies[147], Joseph Dawson[147], Callum Day[147], Aiden Densem[147], Thomas Dibling[147], Cat Dockree[147], David Dodd[147], Sunil Dogga[147], Matthew Dorman[147], Gordon Dougan[147], Martin Dougherty[147], Alexander Dove[147], Lucy Drummond[147], Eleanor Drury[147], Monika Dudek[147], Jillian Durham[147], Laura Durrant[147], Elizabeth Easthope[147], Sabine Eckert[147], Pete Ellis[147], Ben Farr[147], Michael Fenton[147], Marcella Ferrero[147], Neil Flack[147], Howerd Fordham[147], Grace Forsythe[147], Luke Foulser[147], Matt Francis[147], Audrey Fraser[147], Adam Freeman[147], Anastasia Galvin[147], Maria Garcia-Casado[147], Alex Gedny[147], Sophia Girgis[147], James Glover[147], Sonia Goncalves[147], Scott Goodwin[147], Oliver Gould[147], Marina Gourtovaia[147], Andy Gray[147], Emma Gray[147], Coline Griffiths[147], Yong Gu[147], Florence Guerin[147], Will Hamilton[147], Hannah Hanks[147], Ewan Harrison[147], Alexandria Harrott[147], Edward Harry[147], Julia Harvison[147], Paul Heath[147], Anastasia Hernandez-Koutoucheva[147], Rhiannon Hobbs[147], Dave Holland[147], Sarah Holmes[147], Gary Hornett[147], Nicholas Hough[147], Liz Huckle[147], Lena Hughes-Hallet[147], Adam Hunter[147], Stephen Inglis[147], Sameena Iqbal[147], Adam Jackson[147], David Jackson[147], Keith James[147], Dorota Jamrozy[147], Carlos Jimenez Verdejo[147], Ian Johnston[147], Matthew Jones[147], Kalyan Kallepally[147], Leanne Kane[147], Keely Kay[147], Sally Kay[147], Jon Keatley[147], Alan Keith[147], Alison King[147], Lucy Kitchin[147], Matt Kleanthous[147], Martina Klimekova[147], Petra Korlevic[147], Ksenia Krasheninnkova[147], Dominic Kwiatkowski[147], Greg Lane[147], Cordelia Langford[147], Adam Laverack[147], Katharine Law[147], Mara Lawniczak[147], Stefanie Lensing[147], Steven Leonard[147], Laura Letchford[147], Kevin Lewis[147], Amanah Lewis-Wade[147], Jennifer Liddle[147], Quan Lin[147], Sarah Lindsay[147], Sally Linsdell[147], Rich Livett[147], Stephanie Lo[147], Rhona Long[147], Jamie Lovell[147], Jon Lovell[147], Catherine Ludden[147], James Mack[147], Mark Maddison[147], Aleksei Makunin[147], Irfan Mamun[147], Jenny Mansfield[147], Neil Marriott[147], Matt Martin[147], Inigo Martincorena[147], Matthew Mayho[147], Shane McCarthy[147], Jo McClintock[147], Samantha McGuigan[147], Sandra McHugh[147], Liz McMinn[147], Carl Meadows[147], Emily Mobley[147], Robin Moll[147], Maria Morra[147], Leanne Morrow[147], Kathryn Murie[147], Sian Nash[147], Claire Nathwani[147], Plamena Naydenova[147], Alexandra Neaverson[147], Rachel Nelson[147], Ed Nerou[147], Jon Nicholson[147], Tabea Nimz[147], Guillaume G. Noell[147], Sarah O'Meara[147], Valeriu Ohan[147], Karen Oliver[147], Charles Olney[147], Doug Ormond[147], Agnes Oszlanczi[147], Steve Palmer[147], Yoke Fei Pang[147], Barbora Pardubska[147], Naomi Park[147], Aaron Parmar[147], Gaurang Patel[147], Minal Patel[147], Maggie Payne[147], Sharon Peacock[147], Arabella Petersen[147], Deborah Plowman[147], Tom Preston[147], Liam Prestwood[147], Christoph Puethe[147], Michael Quail[147], Diana Rajan[147], Shavanthi Rajatileka[147], Richard Rance[147], Suzannah Rawlings[147], Nicholas Redshaw[147], Joe Reynolds[147], Mark Reynolds[147], Simon Rice[147], Matt Richardson[147], Connor Roberts[147], Katrina Robinson[147], Melanie Robinson[147], David Robinson[147], Hazel Rogers[147], Eduardo Martin Rojo[147], Daljit Roopra[147], Mark Rose[147], Luke Rudd[147], Ramin Sadri[147], Nicholas Salmon[147], David Saul[147], Frank Schwach[147], Carol Scott[147], Phil Seekings[147], Lesley Shirley[147], John Sillitoe[147], Alison Simms[147], Matt Sinnott[147], Shanthi Sivadasan[147], Bart Siwek[147], Dale Sizer[147], Kenneth Skeldon[147], Jason Skelton[147], Joanna Slater-Tunstill[147], Lisa Sloper[147], Nathalie Smerdon[147], Chris Smith[147], Christen Smith[147], James Smith[147], Katie Smith[147], Michelle Smith[147], Sean Smith[147], Tina Smith[147], Leighton Sneade[147], Carmen Diaz Soria[147], Catarina Sousa[147], Emily Souster[147], Andrew Sparkes[147], Michael Spencer-Chapman[147], Janet Squares[147], Robert Stanley[147], Claire Steed[147], Tim Stickland[147], Ian Still[147], Mike Stratton[147], Michelle Strickland[147], Allen Swann[147], Agnieszka Swiatkowska[147], Neil Sycamore[147], Emma Swift[147], Edward Symons[147], Suzanne Szluha[147], Emma Taluy[147], Nunu Tao[147], Katy Taylor[147], Sam Taylor[147], Stacey Thompson[147], Mark Thompson[147], Mark Thomson[147], Nicholas Thomson[147], Scott Thurston[147], Gerry Tonkin-Hill[147], Dee Toombs[147], Benjamin Topping[147], Jaime Tovar-Corona[147], Daniel Ungureanu[147], James Uphill[147], Jana Urbanova[147], Philip Jansen Van[147], Valerie Vancollie[147], Paul Voak[147], Danielle Walker[147], Matthew Walker[147], Matt Waller[147], Gary Ward[147], Charlie Weatherhogg[147], Niki Webb[147], Danni Weldon[147], Alan Wells[147], Eloise Wells[147], Luke Westwood[147], Theo Whipp[147], Thomas Whiteley[147], Georgia Whitton[147], Andrew Whitwham[147], Sara Widaa[147], Mia Williams[147], Mark Wilson[147] & Sean Wright[147]

Sequencing & analysis
Emma Betteridge[47], Ben W. Farr[47], Scott Goodwin[47], Michael A. Quail[47], Carol Scott[47], Lesley Shirley[47], Scott A. J. Thurston[47], Diana Rajan[47], Iraad F. Bronner[47], Louise Aigrain[47], Nicholas M. Redshaw[47], Stefanie V. Lensing[47], Shane McCarthy[47], Alex Makunin[47], Carlos E. Balcazar[72], Michael D. Gallagher[72], Kathleen A. Williamson[72], Thomas D. Stanton[72], Michelle L. Michelsen[87], Joanna Warwick-Dugdale[87], Robin Manley[87], Audrey Farbos[87], James W. Harrison[87], Christine M. Sambles[87], David J. Studholme[87], Angie Lackenby[66], Tamyo Mbisa[66], Steven Platt[66], Shahjahan Miah[66], David Bibby[66], Carmen Manso[66], Jonathan Hubb[66], Gavin Dabrera[66], Mary Ramsay[66], Daniel Bradshaw[66], Ulf Schaefer[66], Natalie Groves[66], Eileen Gallagher[66], David Lee[66], David Williams[66], Nicholas Ellaby[66], Hassan Hartman[66], Nikos Manesis[66], Vineet Patel[66], Juan Ledesma[66], Katherine A. Twohig[66], Elias Allara[46,66], Clare Pearson[46,66], Jeffrey K. J. Cheng[80], Hannah E. Bridgewater[80], Lucy R. Frost[80], Grace Taylor-Joyce[80], Paul E. Brown[80], Lily Tong[67], Alice Broos[67], Daniel Mair[67], Jenna Nichols[67], Stephen N. Carmichael[67], Katherine L. Smollett[143], Kyriaki Nomikou[67], Elihu Aranday-Cortes[67], Natasha Johnson[67], Seema Nickbakhsh[67,76], Edith E. Vamos[81], Margaret Hughes[81], Lucille Rainbow[81], Richard Eccles[81], Charlotte Nelson[81], Mark Whitehead[81], Richard Gregory[81], Matthew Gemmell[81], Claudia Wierzbicki[81], Hermione J. Webster[81], Chloe L. Fisher[49], Adrian W. Signell[144], Gilberto Betancor[144], Harry D. Wilson[144], Gaia Nebbia[93], Flavia Flaviani[145], Alberto C. Cerda[134], Tammy V. Merrill[134], Rebekah E. Wilson[134], Marius Cotic[88], Nadua Bayzid[88], Thomas Thompson[57], Erwan Acheson[57], Steven Rushton[146], Sarah O'Brien[146], David J. Baker[74], Steven Rudder[74], Alp Aydin[74], Fei Sang[58], Johnny Debebe[58], Sarah Francois[83], Tetyana I. Vasylyeva[83], Marina Escalera Zamudio[83], Bernardo Gutierrez[83], Angela Marchbank[42], Joshua Maksimovic[95], Karla Spellman[95], Kathryn McCluggage[95], Mari Morgan[43], Robert Beer[95], Safiah Afifi[95], Trudy Workman[42], William Fuller[42], Catherine Bresner[42], Adrienn Angyal[68], Luke R. Green[68], Paul J. Parsons[68], Rachel M. Tucker[68], Rebecca Brown[68] & Max Whiteley[68]

Software & analysis tools
James Bonfield[47], Christoph Puethe[47], Andrew Whitwham[47], Jennifier Liddle[47], Will Rowe[41], Igor Siveroni[82], Thanh Le-Viet[74] & Amy Gaskin[43]

Visualization
Rob Johnson[82]

[40]Centre for Enzyme Innovation, University of Portsmouth (PORT), Portsmouth, UK. [41]Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK. [42]Department of Medicine, Cardiff University, Cardiff, UK. [43]Public Health Wales NHS Trust, Cardiff, UK. [44]Nuffield Department of Medicine, University of Oxford, Oxford, UK. [45]Department of Virology, King's College London, London, UK. [46]Department of Medicine, University of Cambridge, Cambridge, UK. [47]Department of Pathogen Genomics, Wellcome Sanger Institute, London, UK. [48]Department of Medicine, University of Brighton, Brighton, UK. [49]Genomics Innovation Department, Guy's and St Thomas' NHS Foundation Trust, London, UK. [50]Department of Infectious Diseases, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [51]Department of Microbiology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [52]Department of Medicine, University Hospital Southampton NHS Foundation Trust, Southampton, UK. [53]Department of Medicine, School of Medicine, University of Southampton, Southampton, UK. [54]Department of Medicine, School of Health Sciences, University of Southampton, Southampton, UK. [55]Department of Clinical Infection & Diagnostics Research, St Thomas' Hospital and Kings College London, London, UK. [56]Belfast Health & Social Care Trust, Belfast, UK. [57]Department of Medicine, Queen's University Belfast, Belfast, UK. [58]Deep Seq Department, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK. [59]East Kent Hospitals University NHS Foundation Trust, Canterbury, UK. [60]Department of Medicine, University of Kent, Canterbury, UK. [61]Hub for Biotechnology in the Built Environment, Northumbria University, Newcastle upon Tyne, UK. [62]Department of Medicine, Northumbria University, Newcastle upon Tyne, UK. [63]NU-OMICS, Northumbria University, Newcastle upon Tyne, UK. [64]Centre for Genomic Pathogen Surveillance, University of Oxford, Oxford, UK. [65]Clinical Microbiology and Public Health Laboratory, Public Health England, Cambridge, UK. [66]Public Health England, London, UK. [67]MRC-University of Glasgow Centre for Virus Research, Glasgow, UK. [68]University of Sheffield, Sheffield, UK.

[69]Portsmouth Hospitals University NHS Trust, Portsmouth, UK. [70]School of Pharmacy and Biomedical Sciences, University of Portsmouth (PORT), Portsmouth, UK. [71]NHS Lothian, Edinburgh, UK. [72]University of Edinburgh, Edinburgh, UK. [73]Department of Medicine, University of Oxford, Oxford, UK. [74]Quadram Institute Bioscience, Norwich, UK. [75]Department of Medicine, University of East Anglia, Norwich, UK. [76]Public Health Scotland, Glasgow, UK. [77]Heartlands Hospital, Birmingham, UK. [78]Department of Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [79]Department of Medicine, Brighton and Sussex University Hospitals NHS Trust, Brighton, UK. [80]Department of Medicine, University of Warwick, Warwick, UK. [81]Department of Medicine, University of Liverpool, Liverpool, UK. [82]Department of Medicine, Imperial College London, London, UK. [83]Department of Zoology, University of Oxford, Oxford, UK. [84]Newcastle Hospitals NHS Foundation Trust, Newcastle, UK. [85]Division of Virology, Department of Pathology, University of Cambridge, Cambridge, UK. [86]University Hospitals Coventry and Warwickshire, Coventry, UK. [87]University of Exeter, Exeter, UK. [88]University College London, London, UK. [89]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [90]Betsi Cadwaladr University Health Board, Betsi Cadwaladr, UK. [91]School of Biological Sciences, University of Portsmouth (PORT), Portsmouth, UK. [92]Warwick Medical School and Institute of Precision Diagnostics, Pathology, UHCW NHS Trust, Warwick, UK. [93]Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, London, UK. [94]Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK. [95]Cardiff and Vale University Health Board, Cardiff, UK. [96]Turnkey Laboratory, University of Birmingham, Birmingham, UK. [97]Gloucestershire Hospitals NHS Foundation Trust, Gloucester, UK. [98]Microbiology Department, Wye Valley NHS Trust, Hereford, UK. [99]Sandwell and West Birmingham NHS Trust, Birmingham, UK. [100]Norfolk and Norwich University Hospital, Norwich, UK. [101]Royal Devon and Exeter NHS Foundation Trust, Exeter, UK. [102]Barking, Havering and Redbridge University Hospitals NHS Trust, Barking, UK. [103]Queen Elizabeth Hospital, Birmingham, UK. [104]Department of Microbiology, Kettering General Hospital, Kettering, UK. [105]Clinical Microbiology, University Hospitals of Leicester NHS Trust, Leicester, UK. [106]Imperial College Hospitals NHS Trust, London, UK. [107]North West London Pathology, London, UK. [108]Royal Free NHS Trust, London, UK. [109]South Tees Hospitals NHS Foundation Trust, Newcastle, UK. [110]North Cumbria Integrated Care NHS Foundation Trust, Carlisle, UK. [111]North Tees and Hartlepool NHS Foundation Trust, Stockton-on-Tees, UK. [112]County Durham and Darlington NHS Foundation Trust, Durham, UK. [113]Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK. [114]Clinical Microbiology Department, Queens Medical Centre, Nottingham, UK. [115]PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, Scunthorpe, UK. [116]Basingstoke Hospital, Basingstoke, UK. [117]University of Surrey, Guildford, UK. [118]Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [119]Swansea University, Swansea, UK. [120]Ministry of Health, Colombo, Sri Lanka. [121]Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. [122]Liverpool Clinical Laboratories, Liverpool, UK. [123]NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, London, UK. [124]Health Data Research UK Cambridge, Cambridge, UK. [125]Hampshire Hospitals NHS Foundation Trust, Winchester, UK. [126]Public Health Agency, London, UK. [127]Department of Infection Biology, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK. [128]University of Birmingham, Birmingham, UK. [129]West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, Glasgow, UK. [130]NHS Greater Glasgow and Clyde, Glasgow, UK. [131]National Infection Service, PHE and Leeds Teaching Hospitals Trust, Leeds, UK. [132]Manchester University NHS Foundation Trust, Manchester, UK. [133]Guy's and St Thomas' Hospitals, London, UK. [134]Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's College Hospital NHS Foundation Trust, London, UK. [135]Health Services Laboratories, London, UK. [136]Maidstone and Tunbridge Wells NHS Trust, Maidstone, UK. [137]Gateshead Health NHS Foundation Trust, Gateshead, UK. [138]Norfolk County Council, Norwich, UK. [139]East Suffolk and North Essex NHS Foundation Trust, Ipswich, UK. [140]Department of Medicine, University of Southampton, Southampton, UK. [141]Microbiology Department, Princess Alexandra Hospital, Harlow, UK. [142]Sheffield Teaching Hospitals, Sheffield, UK. [143]Institute of Biodiversity, Animal Health & Comparative Medicine, Glasgow, UK. [144]Department of Infectious Diseases, King's College London, London, UK. [145]Guy's and St Thomas' BRC, London, UK. [146]Newcastle University, Newcastle, UK. [147]Wellcome Sanger Centre, Hinxton, UK.

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 500 µl of sample with a dilution of MS2 bacteriophage to act as an internal control, using the easyMAG platform (Biomerieux) according to the manufacturers' instructions. All samples were tested for presence of SARS-CoV-2 with a validated one-step RT–qPCR assay developed in conjunction with the Public Health England Clinical Microbiology[22]. Amplification reactions were all performed on a Rotorgene PCR instrument. Samples with a C_t value of ≤36 were considered to be positive.

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

SARS-CoV-2 serology by multiplex particle-based flow cytometry

Recombinant SARS-CoV-2 nucleocapsid, spike and RBD proteins were covalently coupled to distinct carboxylated bead sets (Luminex) to form a triplex and analysed as previously described[24]. Specific binding was reported as mean fluorescence intensities.

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 phytohaemagglutinin (10 µg ml^{-1}; Sigma-Aldrich), lipopolysaccharide (1 µg ml^{-1}, List Biochemicals) or by costimulation with anti-CD3 (MEM57, Abcam, 200 ng ml^{-1}, 1:1,000) and IL-2 (Immunotools, 1,430 U ml^{-1}, 1:1,000) antibodies. Supernatants were taken after 24 h. Levels (pg ml^{-1}) are shown for IFNγ, IL-17, IL-2, TNF, IL-6, IL-1b and IL-10. Cytokines were measured by multiplexed particle-based flow cytometry on a Luminex analyser (Bio-Plex, Bio-Rad) using an R&D Systems custom kit (R&D Systems).

For viral genomic sequencing, total RNA was extracted from samples as described previously[25]. Samples were sequenced using MinION flow cells v.9.4.1 (Oxford Nanopore Technologies) following the ARTICnetwork v.3 protocol[26] and BAM files assembled using the ARTICnetwork assembly pipeline[27]. A representative set of 10 sequences was selected and also sequenced using the Illumina MiSeq platform. Amplicons were diluted to 2 ng µl^{-1} 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 the manufacturer's instructions. In brief, amplicons were end-repaired and had A-overhang added; these were then ligated with 15 mM of NEXTflex DNA Barcodes (Bio Scientific). Post-ligation products were cleaned using AMPure beads and eluted in 25 µl. Then, 20 µl was used for library amplification by 5 cycles of PCR. For the negative controls, 1 ng was used for ligation-based library preparation. All libraries were assayed using TapeStation (Agilent Technologies) to assess fragment size and quantified by qPCR. All libraries were then pooled in equimolar ratios accordingly. Libraries were loaded at 15 nM and spiked in 5% PhiX (Illumina) and sequenced on for MiSeq 500 cycle using a Miseq Nano v.2 with 2 × 250 bp paired-end sequencing. A minimum of ten reads was required for a variant call.

Bioinformatics processes

For long-read sequencing, genomes were assembled with reference-based assembly and a curated bioinformatics pipeline with 20× minimum coverage across the whole genome[27]. For short-read sequencing, FASTQ files were downloaded, poor-quality reads were identified and removed, and both Illumina and PHiX adapters were removed using TrimGalore v.0.6.6[28]. 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[29]. BAM files were then sorted and indexed with Samtools v.1.11 and PCR optical duplicates were removed using Picard (http://broadinstitute.github.io/picard). 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 (https://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 2 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 (Thermo Fisher Scientific) and gene-specific primers were used for reverse transcription. Template RNA was degraded with RNase H (Thermo Fisher Scientific). All primers used were 'in-house' primers designed using the multiple sequence alignment of the patient's consensus sequences obtained by next-generation sequencing. A partial gene sequence of *spike* (encoding amino acids 21–800) was amplified as one continuous length of DNA (the *spike* gene is approximately 1.8 kb) by nested PCR. Terminally diluted cDNA was amplified by PCR using Platinum Taq DNA Polymerase High Fidelity (Invitrogen) so that 30% of reactions were positive[30]. By Poisson statistics, sequences were deemed ≥80% likely to be derived from HIV-1 single genomes. We obtained between 20 and 60 single genomes at each sample time point to achieve 90% confidence of detecting variants present at ≥8% of the viral population in vivo[31,32]. Partial *spike* amplicons obtained from terminal-dilution PCR amplification were Sanger-sequenced to form a contiguous sequence using another set of eight in-house primers. Sanger sequencing was provided by Genewiz and manual sequence editing was performed using DNA Dynamo software (Blue Tractor Software).

Phylogenetic analysis

All available full-genome SARS-CoV-2 sequences were downloaded from the GISAID database (http://gisaid.org/)[33] on 16 December 2020. Duplicate and low-quality sequences (>5% nucleocapsid regions) were removed, leaving a dataset of 212,297 sequences with a length of more than 29,000 bp. All sequences were sorted by name and only sequences sequenced with United Kingdom/England identifiers were retained. From this dataset, sequences were deduplicated and, in figures in which background sequences were required, randomly subsampled using seqtk (https://github.com/lh3/seqtk). All sequences were aligned to the SARS-CoV-2 reference strain MN908947.3, using MAFFT v.7.475 with automatic flavour selection[34]. Major SARS-CoV-2 clade memberships were assigned to all sequences using both the Nextclade server v.0.9 (https://clades.nextstrain.org/) and Phylogenetic Assignment Of Named Global Outbreak Lineages (PANGOLIN)[35].

Maximum-likelihood phylogenetic trees were produced using the above curated dataset using IQ-TREE v.2.1.2[36]. Evolutionary model selection for trees were inferred using ModelFinder[37] and trees were estimated using the GTR + F + I model with 1,000 ultrafast bootstrap replicates[38]. All trees were visualized 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 less than 50 were collapsed using an in-house script.

In-depth allele frequency variant calling

The SAMFIRE package v.1.06[39] was used to call allele frequency trajectories from BAM file data. Reads were included in this analysis if they had a median PHRED score of at least 30, trimming the ends of reads to achieve this if necessary. Nucleotides were then filtered to have a PHRED score of at least 30; reads with fewer than 30 such reads were discarded. Distances between sequences, accounting for low-frequency variant information, were also obtained using SAMFIRE. The sequence distance metric, described in a previous paper[40], combines allele frequencies across the whole genome. Where L is the length of the genome, we define $\mathbf{q}(t)$ as a $4 \times L$ 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 generalization of the Hamming distance

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

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

$$D(\mathbf{q}(t_1), \mathbf{q}(t_2)) = \sum_i d\left(q_i(t_1), q_i(t_2)\right)$$

The Mathematica software package was used 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 that the samples collected on days 93 and 95 arose 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 (https://github.com/PollockLaboratory/AnCovMulti).

Western blot analysis

Cells were transfected with the indicated plasmid preparations and 48 h after transfection the culture supernatant was collected 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 4–12% 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 using a ChemiDoc MP imaging system (Biorad) using anti-spike S2 (Invitrogen; 1:1,000 dilution) and anti-p24 Gag (NIH AIDS Reagents; 1:1,000 dilution) antibodies.

Recombination detection

All sequences were tested for potential recombination, as this would affect 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[41]. To corroborate any findings, ClonalFrameML v.1.12[42] was also used to infer recombination breakpoints. Neither of the programs found evidence of recombination in our data.

Structural viewing

The PyMOL Molecular Graphics System v.2.4.0 (https://github.com/schrodinger/pymol-open-source/releases) was used to map the location of the four spike mutations of interested onto a previously published SARS-CoV-2 spike structure (PDB: 6ZGE)[43].

Testing of convalescent plasma for antibody titres

The anti-SARS-CoV-2 ELISA (IgG) assay used to test convalescent plasma for antibody titres was Euroimmun Medizinische Labordiagnostika. This indirect ELISA-based assay uses a recombinant structural spike 1 (S1) protein of SARS-CoV-2 expressed in the human cell line HEK293 for the detection of SARS-CoV-2 IgG.

Generation of spike mutants

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

Pseudotype virus preparation

Viral vectors were prepared by transfection of HEK293T cells using the Fugene HD transfection reagent (Promega). HEK293T cells were transfected with a mixture of 11 µl of Fugene HD, 1 µg of pCDNAΔ19Spike–HA, 1 µg of p8.91 HIV-1 gag–pol expression vector[45,46] and 1.5 µg of pCSFLW (expressing the firefly luciferase reporter gene with the HIV-1 packaging signal). Viral supernatant was collected at 48 and 72 h after transfection, filtered through 0.45-µm filter and stored at −80 °C. The 50% tissue-culture infectious dose ($TCID_{50}$) of SARS-CoV-2 pseudovirus was determined using the Steady-Glo Luciferase assay system (Promega).

Standardization of virus input by SYBR-Green-based product-enhanced PCR assay

The reverse transcriptase activity of virus preparations was determined by qPCR using a SYBR-Green-based product-enhanced PCR assay as previously described[47]. In brief, tenfold dilutions of virus supernatant were lysed in a 1:1 ratio in a 2× lysis solution (made up of 40% glycerol (v/v), 0.25% Triton X-100 (v/v), 100 mM KCl, RNase inhibitor 0.8 U ml^{-1}, Tris HCL 100 mM, buffered to pH 7.4) for 10 min at room temperature.

Then, 12 µl of each sample lysate was added to 13 µl of a SYBR Green master mix (containing 0.5 µM of MS2 RNA forward and reverse primers, 3.5 pmol ml^{-1} of MS2 RNA and 0.125 U $µl^{-1}$ 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 reverse transcriptase activity calculated by comparing the relative amounts of reverse transcription to a reverse transcriptase standard of known activity.

Serum/plasma pseudotype neutralization assay

Spike pseudotype assays have been shown to have similar characteristics as neutralization testing using fully infectious wild-type SARS-CoV-2[8]. Virus neutralization assays were performed on HEK293T cells transiently transfected with ACE2 and TMPRSS2 using SARS-CoV-2 spike pseudotyped virus expressing luciferase[48]. Pseudotyped virus was incubated with serial dilutions of heat-inactivated human serum samples or convalescent plasma in duplicate for 1 h at 37 °C. Virus- and cell-only controls were also included. Then, freshly trypsinized HEK293T ACE2- and TMPRSS2-expressing cells were added to each well. After 48 h incubation in a 5% CO_2 environment at 37 °C, luminescence was measured using Steady-Glo Luciferase assay system (Promega).

Monoclonal antibody pseudotype neutralization assay

Virus neutralization assays were performed using HeLa cells stably expressing ACE2 and using SARS-CoV-2 spike pseudotyped virus expressing luciferase as previously described[49]. Pseudotyped virus was incubated with serial dilutions of purified monoclonal antibodies[9] in duplicate for 1 h at 37 °C. Then, freshly trypsinized HeLa ACE2-expressing cells were added to each well. After 48 h incubation in a 5% CO_2 environment at 37 °C, luminescence was measured using a

Bright-Glo Luciferase assay system (Promega) and neutralization was calculated relative to virus-only controls. IC_{50} values were calculated in GraphPad Prism.

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 his family. Additional control patients with COVID-19 were enrolled to the NIHR BioResource Centre Cambridge under ethics review board approval (17/EE/0025).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Long-read sequencing data that support the findings of this study have been deposited in the NCBI SRA database with accession codes SAMN16976824–SAMN16976846 under 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. Source data are provided with this paper.

Code availability

The SAMFIRE package v.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.

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. *Lancet Infect. Dis.* **20**, 1263–1271 (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.* **1**, 100062 (2020).
24. Xiong, X. et al. A thermostable, closed SARS-CoV-2 spike protein trimer. *Nat. Struct. Mol. Biol.* **27**, 934–941 (2020).
25. Meredith, L. W. et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.* **20**, 1263–1271 (2020).
26. Quick, J. nCoV-2019 sequencing protocol. *Protocols.io* https://doi.org/10.17504/protocols.io.bbmuik6w (2020).
27. Loman, N., Rowe, W. & Rambaut, A. nCoV-2019 novel coronavirus bioinformatics protocol. Protocol version 1 https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html (Artic Network, 2020).
28. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
29. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
30. Jordan, M. R. et al. Comparison of standard PCR/cloning to single genome sequencing for analysis of HIV-1 populations. *J. Virol. Methods* **168**, 114–120 (2010).
31. Palmer, S. et al. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J. Clin. Microbiol.* **43**, 406–413 (2005).
32. Keele, B. F. et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl Acad. Sci. USA* **105**, 7552–7557 (2008).
33. Shu, Y. & McCauley, J. GISAID: global initiative on sharing all influenza data — from vision to reality. *Euro Surveill.* **22**, 30494 (2017).
34. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
35. Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
36. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
37. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
38. Minh, B. Q., Nguyen, M. A. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
39. Illingworth, C. J. SAMFIRE: multi-locus variant calling for time-resolved sequence data. *Bioinformatics* **32**, 2208–2209 (2016).
40. Lumby, C. K., Zhao, L., Breuer, J. & Illingworth, C. J. A large effective population size for established within-host influenza virus infection. *eLife* **9**, e56915 (2020).
41. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).
42. Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* **11**, e1004041 (2015).
43. Wrobel, A. G. et al. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* **27**, 763–767 (2020).
44. Gregson, J. et al. Human immunodeficiency virus-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. *J. Infect. Dis.* **222**, 1108–1116 (2020).
45. Naldini, L., Blömer, 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. USA* **93**, 11382–11388 (1996).
46. Gupta, R. K. et al. Full-length HIV-1 Gag determines protease inhibitor susceptibility within in vitro assays. *AIDS* **24**, 1651–1655 (2010).
47. Vermeire, J. et al. Quantification of reverse transcriptase activity by real-time PCR as a fast and accurate method for titration of HIV, lenti- and retroviral vectors. *PLoS ONE* **7**, e50859 (2012).
48. Mlcochova, P. et al. Combined point-of-care nucleic acid and antibody testing for SARS-CoV-2 following emergence of D614G spike variant. *Cell Rep. Med.* **1**, 100099 (2020).
49. Seow, J. et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nat. Microbiol.* **5**, 1598–1607 (2020).
50. Brouwer, P. J. M. et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643–650 (2020).

Acknowledgements We are grateful to the patient and his family. We thank the staff at CUH and the NIHR Cambridge Clinical Research Facility; R. Kugathasan and W. Barclay for discussions; M. Curran, W. Hamilton and D. Sparkes, A. Floto and F. Gallagher; J. Voss for the gift of HeLa cells stably expressing ACE2; and J. Nathan for the RBD protein and L. James for the nucleocapsid 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. R.K.G. is supported by a Wellcome Trust Senior Fellowship in Clinical Science (WT108082AIA). L.E.M. is supported by a Medical Research Council Career Development Award (MR/R008698/1). S.A.K. is supported by the Bill and Melinda Gates Foundation via PANGEA grant (OPP1175094). D.A.C. is supported by a Wellcome Trust Clinical PhD Research Fellowship. C.J.R.I. acknowledges MRC funding (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 (to L.E.M.). J.A.G.B. is supported by the Medical Research Council (MC_UP_1201/16). I.G.G. is a Wellcome Senior Fellow and supported by the Wellcome Trust (207498/Z/17/Z). D.D.P. is supported by NIH GM083127.

Author contributions R.K.G., S.A.K., D.A.C., A.S., T.G. and E.G.-K. conceived the study. R.K.G., S.A.K., D.A.C., L.E.M., J.A.G.B., E.G.-K., N.T., A. Chandra, C.S., R.D., R.A.G., D.D.P. and Y.M. designed experiments. S.A.K., D.A.C., L.E.M., R.D., C.R.-S., A.J., I.A.T.M.F., K.S., T.G., C.J.R.I., J.R.B., J.P.S., M.J.v.G., L.G.-C., G.B.-M. and K.L. performed experiments. R.K.G., S.A.K., D.A.C., P.M., L.E.M., J.A.G.B., S.G., K.S., T.G., J.B., K.G.C.S., I.G.G., C.J.R.I., I.U.L., D.J.R., J.P.S., J.R.B., R.A.G., D.D.P., R.D., L.C.-G. and G.B.-M. interpreted data. R.K.G., D.A.C. and S.A.K wrote the first draft of the paper. All authors contributed to revising the paper.

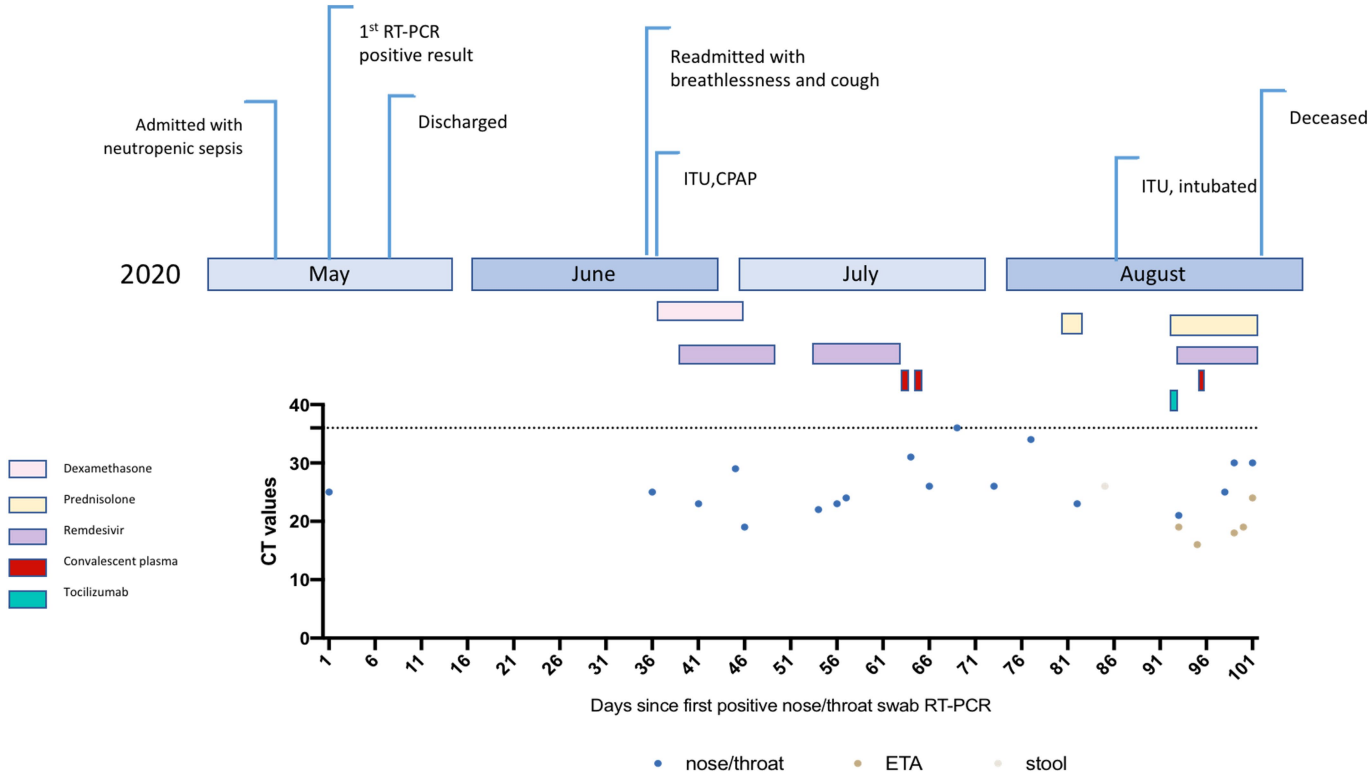
Competing interests The authors declare no competing interests.

Additional information

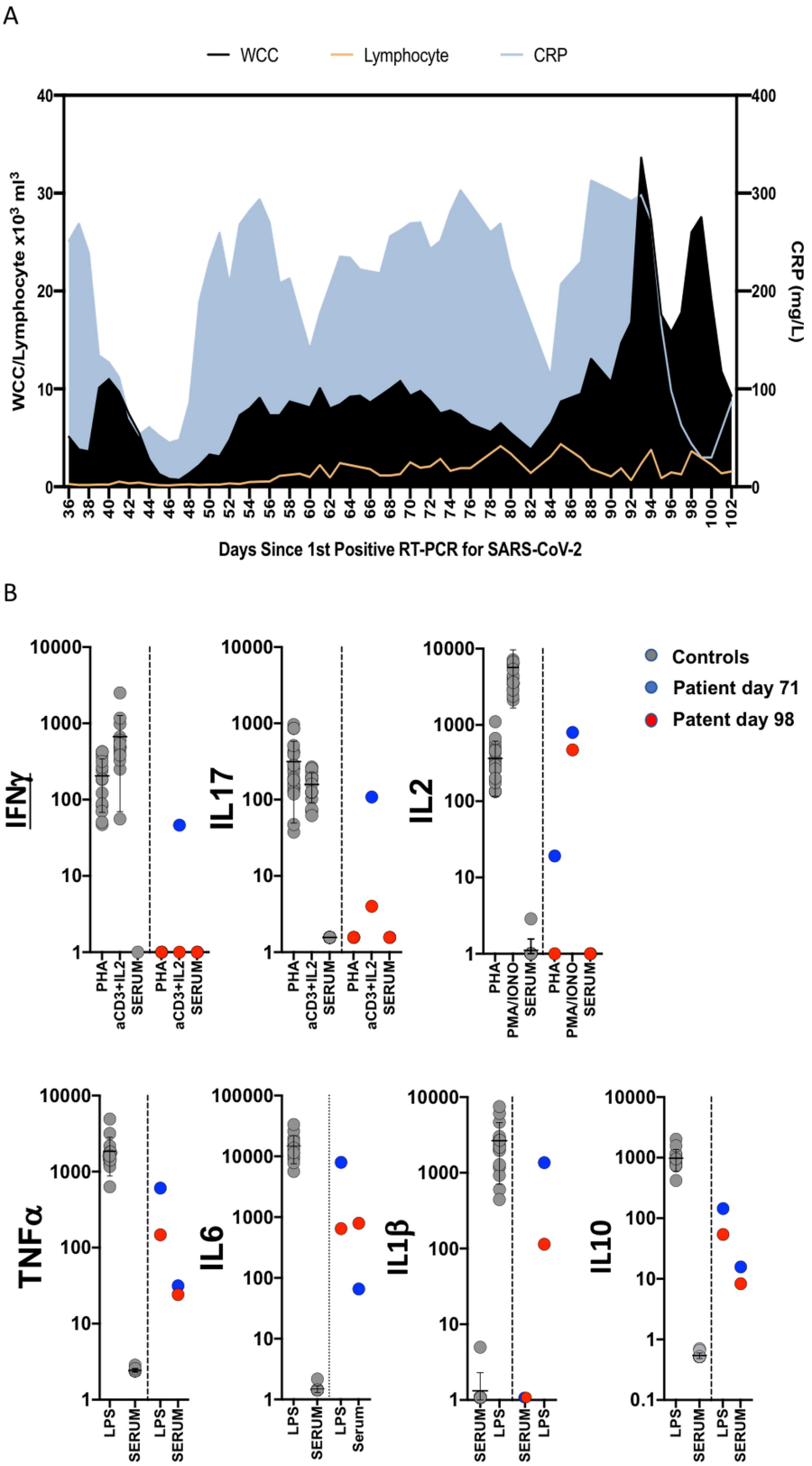
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03291-y.

Correspondence and requests for materials should be addressed to R.K.G.

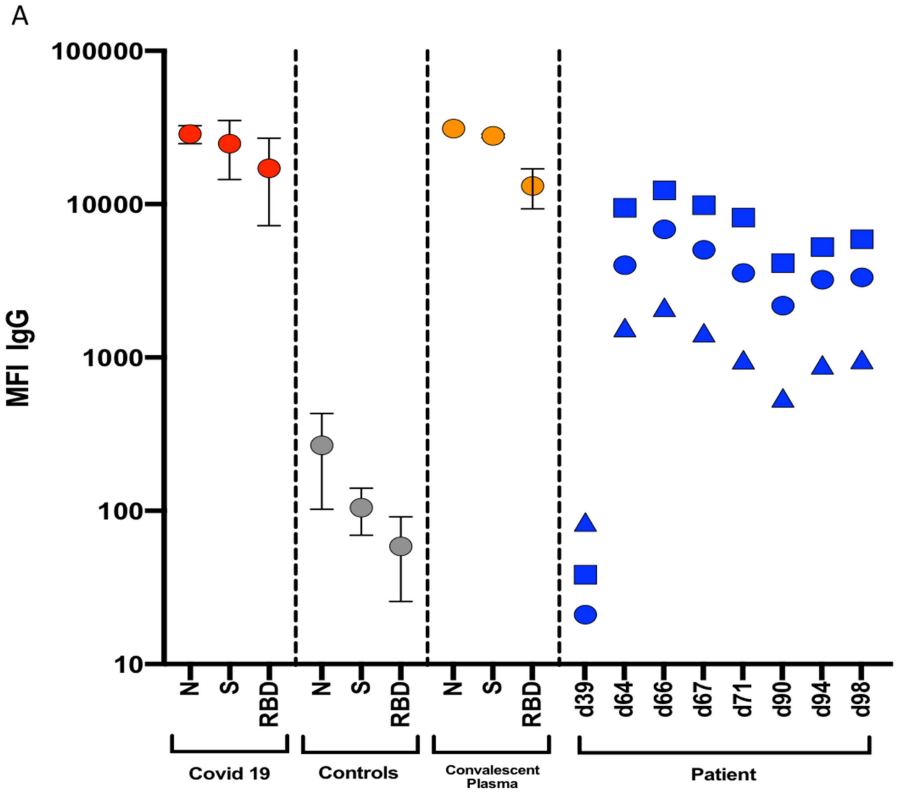
Peer review information *Nature* thanks Richard Neher and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

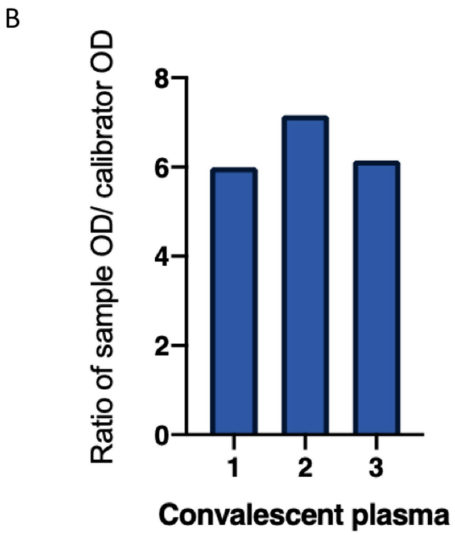


Extended Data Fig. 1 | Clinical time line of events with longitudinal respiratory sample C_t values. CPAP, continuous positive airway pressure; CT, cycle threshold; ITU, intensive therapy unit.



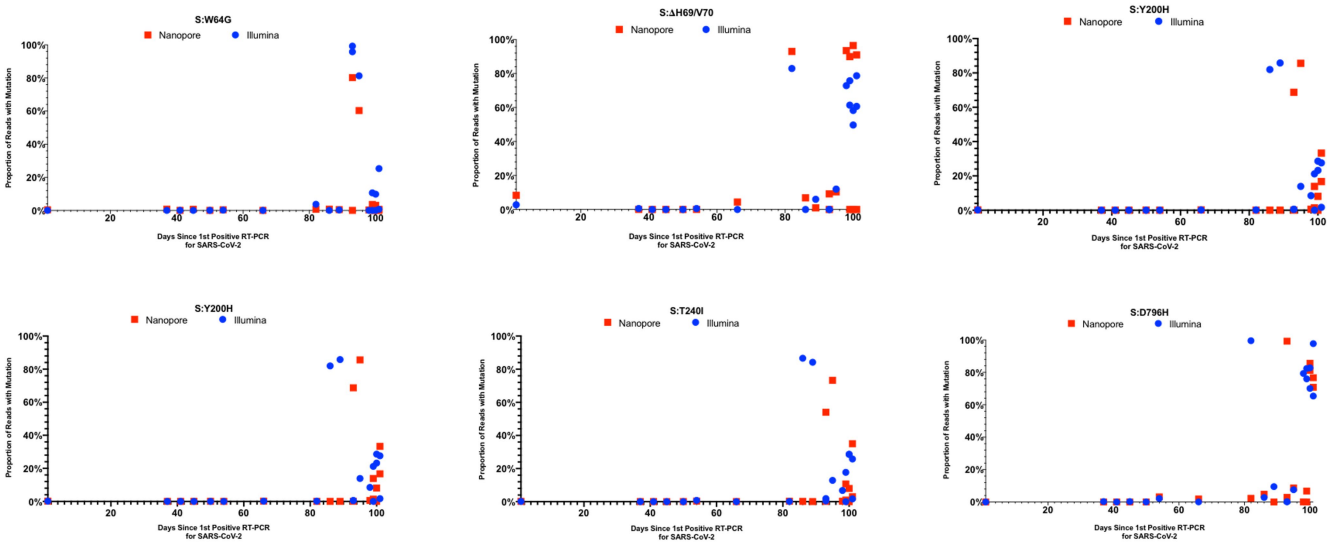
Extended Data Fig. 2 | Blood parameters over time in patient X1. a, White cell (WCC) and lymphocyte counts are expressed as 10^3 cells per mm^3. CRP, C-reactive protein. **b**, Assessment of T cell and innate function. Cytokines were measured in whole blood after stimulation for 24 h either after T cell stimulation with phytohaemagglutinin (PHA) or anti-CD3/IL2 antibodies, or innate stimulation with lipopolysaccharide (LPS). Data from healthy control individuals are shown as grey circles ($n = 15$), data from patient X1 at days 71 and 98 are shown as blue and red circles, respectively. Cytokine levels are shown as pg ml^{-1}. Data are mean ± s.d.





Extended Data Fig. 3 | Serum antibody levels against SARS-CoV-2. **a**, Anti-SARS-CoV-2 IgG antibodies in patient X1, before and after convalescent plasma treatment (days 66 and 95) compared with RNA$^+$ patients with COVID-19 and pre-pandemic healthy control individuals. Red, grey and gold, IgG antibodies against the nucleocapsid (N) protein, trimeric spike (S) protein and RBD of SARS-CoV-2 were measured by multiplexed particle-based flow cytometry (Luminex) in RNA$^+$ patients with COVID-19 ($n = 20$, red dots), pre-pandemic healthy control individuals ($n = 20$, grey dots) and convalescent donor plasma (orange dots). Patient sera over time are shown in blue: anti-SARS-CoV-2 IgG to nucleocapsid (blue squares), spike (blue circles) and RBD (blue triangles). Data are mean fluorescent intensity (MFI) ± s.d. The timing of convalescent plasma units (days after first positive test) is also shown. **b**, SARS-CoV-2 antibody titres in convalescent plasma. Measurement of SARS-CoV-2-specific IgG antibody titres in three units of convalescent plasma by Euroimmun assay. OD, optical density.

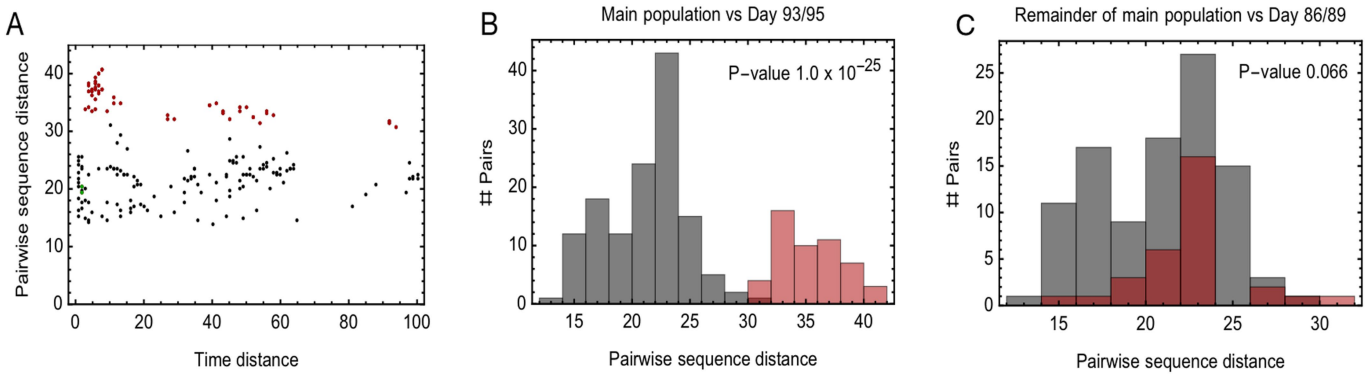
A



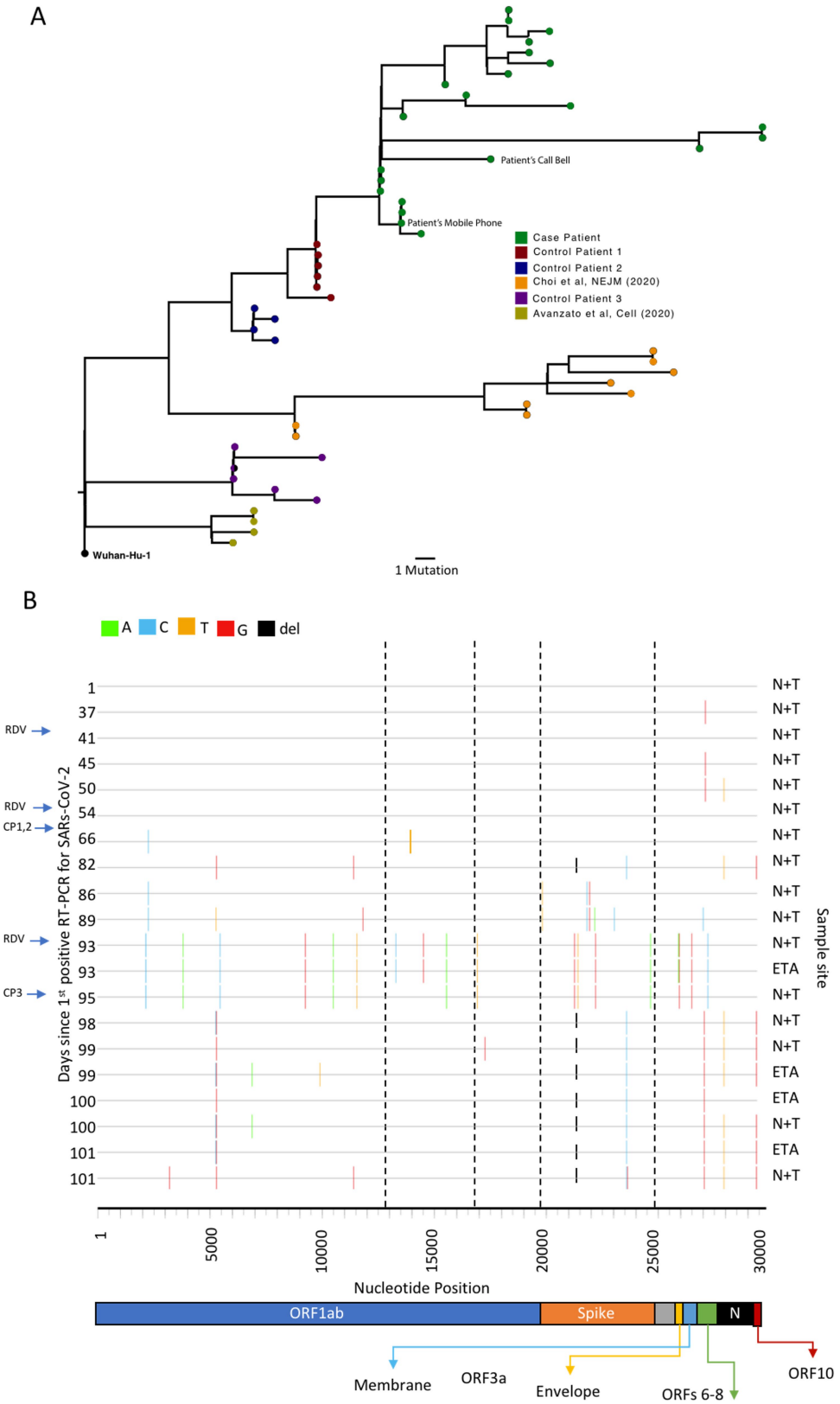
B

	W64G	P330S	ΔH69/V70	D796H	T200I	Y240H
Day 1 (n=7)	0	0	0	0	0	0
Day 37 (n=38)	0	0	0	0	0	0
Day 98 (n=21)	1 (4.8)	1 (4.8)	17 (81.0)	13*(68.4)	3 (14.3)	3 (14.3)

Extended Data Fig. 4 | Cross-comparison of sequencing approaches. **a**, Comparison between short-read (Illumina) and long-read single-molecule (Oxford Nanopore) sequencing methods for the six observed mutations in the spike protein. Concordance was generally good between the majority of time points; however, owing to large discrepancies in a number of time points, we suggest that due to the high base-calling error rate, the Nanopore method is not yet suitable for calling minority variants. As such, all figures in the main paper were produced using only the Illumina data. **b**, Single-genome sequencing data from respiratory samples at the indicated days. The number of single genomes obtained at each time point with the mutations of interest (identified by deep sequencing) are shown. *The denominator is 19, as the primer reads for two samples were poor quality at amino acid 796 on day 98. Amino acid variant and corresponding nucleotide position: spike(W64G), 21752; spike(Δ69), 21765–21770; spike(Y200H), 22160; spike(T240I), 22281; spike(P330S), 22550; spike(D795H), 23948.

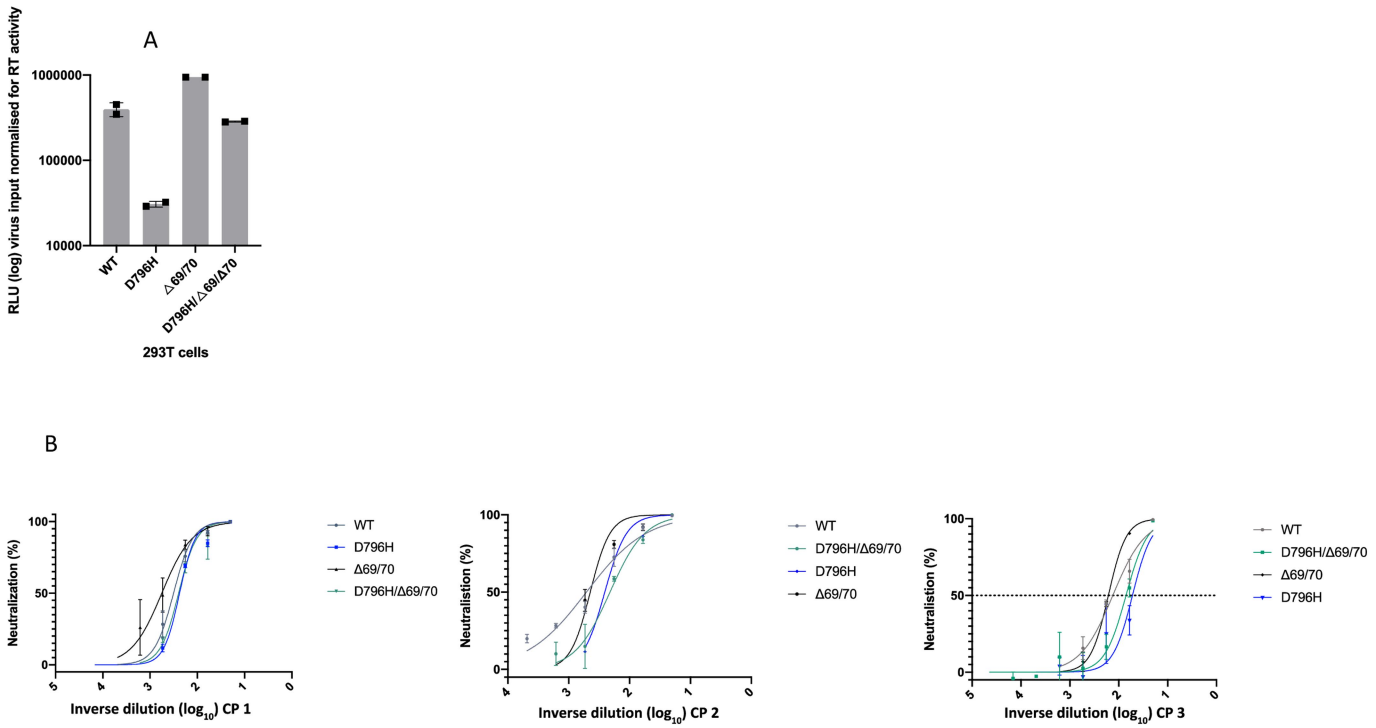


Extended Data Fig. 5 | Evidence of a within-host cladal structure. a, Pairwise distances between samples measured using the all-locus distance metric plotted against pairwise distances in time (measured in days) between samples being collected. Internal distances between samples in the proposed main 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 *U*-test. **c**, Pairwise distances between samples in the main clade, once the data collected on days 86, 89, 93 and 95 were 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 *U*-test.



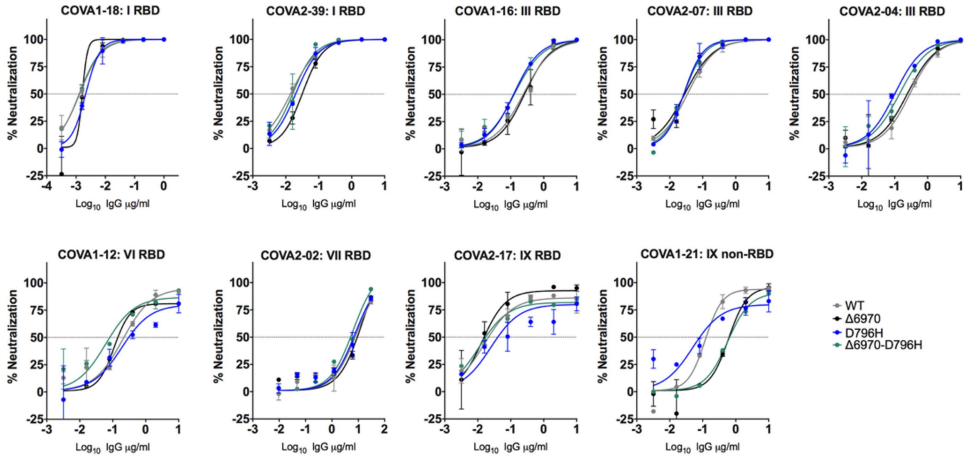
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Sequence analyses and comparisons with other patients with SARS-CoV-2. a, Magnified view of the maximum-likelihood phylogenetic tree showing the diversity of patient X1 and three other patients who showed long-term shedding of SARS-CoV-2 from the local area (red, blue and purple), compared to recently published sequences[4,5] (orange and gold). Control patients with SARS-CoV-2 generally showed limited diversity temporally, although the sequences from one of the previous studies[4] were highly divergent. Environmental samples (the call bell and mobile phone of the patient) are indicated. Tree branches have been collapsed where bootstrap support was <60. **b**, Highlighter plot showing the nucleotide changes at consensus level in sequential respiratory samples compared to the consensus sequence at first diagnosis of COVID-19. Each row indicates the time point the sample was collected (number of days after the first positive RT–qPCR result for SARS-CoV-2). Black dashed lines indicate the RNA-dependent RNA polymerase (RdRp) and spike regions of the genome. There were few nucleotide substitutions between days 1 and 54, despite the patient receiving two courses of remdesivir. The first major changes in the spike genome occurred on day 82, after convalescent plasma was administered on days 63 and 65. The amino acid deletion in S1 (ΔH69/ΔV70) is indicated by the black lines. Sample sites included endotracheal aspirate (ETA) and nose and throat swabs (N+T).



Extended Data Fig. 7 | In vitro infectivity and neutralization sensitivity of spike-pseudotyped lentiviruses. a, Infection of target HEK293T cells expressing TMPRSS2 and ACE2 receptors using equal amounts of virus as determined by reverse transcriptase activity. Data points are mean ± s.e.m . of technical replicates ($n = 2$); data are representative of $n = 2$ independent experiments. **b**, Representative inverse dilution plots for spike variants against convalescent plasma units 1–3. Data are mean ±s.e.m. of the neutralization of technical replicates ($n = 2$). Data are representative of two independent experiments ($n = 2$).

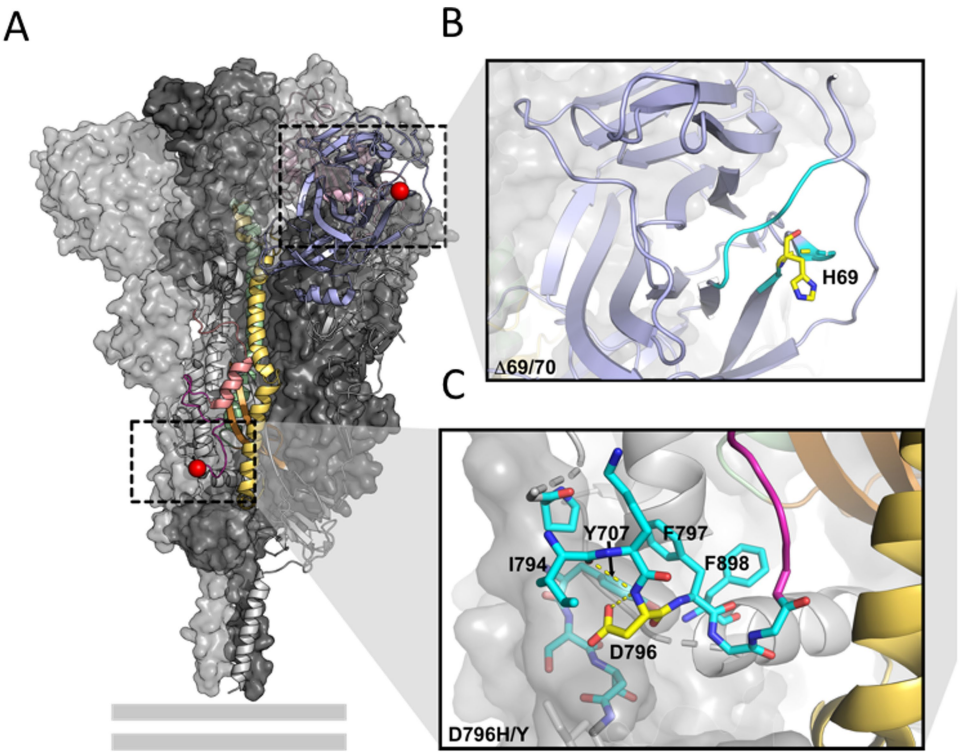
A



B

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

Extended Data Fig. 8 | Neutralizing activity of a panel of monoclonal antibodies against SARS-CoV-2 spike-pseudotyped viruses. a, Lentiviruses pseudotyped with the spike protein (wild type (WT) (D614G background), D796H, ΔH69/ΔV70, D796H + ΔH69/ΔV70) of SARS-CoV-2 were produced in HEK293T cells and used to infect target Hela cells stably expressing ACE2 in the presence of serial dilutions of the indicated monoclonal antibodies (mAb). Data are mean ± s.d. of $n = 2$ technical replicates. Data are representative of at least two independent experiments. **b**, Classes of RBD-binding antibodies and fold changes for spike mutations D796H or ΔH69/ΔV70 are indicated on a previously published study[50]. Clusters II and V contain only non-neutralizing monoclonal antibodies; the smaller neutralizing monoclonal antibody clusters IV ($n = 2$) and X ($n = 1$) were not tested. Red indicates significant fold changes.



D

Mutation	Number of Sequences	Global Prevalence (%)
W64G	0	0.00
ΔH69/V70	12883	4.32
Y200H	7	<0.01
T240I	77	0.02
P330S	167	0.06
D796H	65	0.02
D796Y	141	0.05

Extended Data Fig. 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**, Magnification of the region defined by the box around H69 in **a**. H69 is highlighted in yellow. Residues containing atoms that are within 6 Å of H69 are highlighted in cyan. **c**, Magnification of the region defined by the box around D796 in **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 described in this paper. All high-coverage sequences were downloaded from the GISAID database on 6 January 2021 and aligned using MAFFT; as of this date there were 298,254 sequences available. The global prevalence of each of the six spike variants (W64G, ΔH69/ΔV70, 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.

nature research

Corresponding author(s): Ravindra K Gupta (rkg20@cam.ac.uk)

Last updated by author(s): Jan 4, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Graphad Prism v8 and v9 were used to produce figures. Mafft v7.475 was used for multiple sequence alignments. IQTREE and ModelFinder v2.1.2 was used to infer maximum-likelihood phylogenies. Figtree v1.4.4 was used to annotate and manipulate phylogeny trees. NextClade server v0.9 and Pangolin v2.12 were used to assign lineages to sequences.
Data analysis	Software versions and parameters used for sequence conservation analysis are reported in methods. For in-depth variant analyses, the SAMFIRE package v1.06 was used (https://github.com/cjri/samfire/). To validate frequency variants, custom code was used as part of the package AnCovMutlti (github.com/PollockLaboratory/AnCovMulti).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequences fro SARS-CoV-2 were obtained from GISAID database (https://gisaid.org/) using the filters and search parameters defined in the methods section.

Structural models were obtained from the Protein Data Bank (PDB) https://www.rcsb.org/.
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-read and consensus fasta files have been deposited and are available to view and download on GitHub (https://github.com/Steven-Kemp/sequence_files). Raw data used to create figures are available to view and download on GitHub (https://github.com/Steven-Kemp/sequence_files/tree/main/figure_data).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No formal sample size calculation was undertaken as this case study report focused only on a single patient. By definition n=1
Data exclusions	No data were excluded.
Replication	All experimental data were reproducible and we have shown data representative of at least two independent experiments
Randomization	not applicable as this is a descriptive study of one patient
Blinding	not applicable as this is a descriptive study of one patient and blinding was not appropriate

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Panel of non-commercially available antibodies fully described in Science: DOI: 10.1126/science.abc5902. were provided by Brouwer et al at request. All antibiodies were diluted to a starting concentration of 10 ug/ml and then titrated 5-fold, except for COVA1-18 which was at a starting concentration of 1 ug/ml and COVA2-02 at a starting assay concentration of 30 ug/ml due to their different neutralization potency. Spike S2 antibody (Invitrogen,, Cat no: Pa1-41165), p24 antibody (NIH AIDS Reagents cat no: ARP465). Anti-CD3 antibody (MEM57, Abcam, 200 ng/ml, 1:1000, Cat no Ab8090)
Validation	Validated by providing lab: see supplementary: https://science.sciencemag.org/content/sci/suppl/2020/06/15/science.abc5902.DC1/abc5902-Brouwer-SM.pdf

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HeLa cells were donated by kind request from James Voss as noted in the acknowledgments section. HEK 293T cells from ATCC were used for transfection work.
Authentication	None of the cell lines used were authenticated.

Mycoplasma contamination	All cell lines used were tested a(by PCR) and were mycoplasma free.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.

Human research participants

Policy information about studies involving human research participants

Population characteristics	A single septuagenarian male was treated at a local Cambridge hospital for COVID-19 symptoms.
Recruitment	As part of routine testing, nose + throat samples and endotracheal aspirates were collected from the patient at 23 time-points.
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.