

SARS-CoV-2 evolution during prolonged infection in immunocompromised patients

Andrew D. Marques,¹ Jevon Graham-Wooten,² Ayannah S. Fitzgerald,² Ashley Sobel Leonard,³ Emma J. Cook,¹ John K. Everett,¹ Kyle G. Rodino,¹ Louise H. Moncla,⁴ Brendan J. Kelly,¹ Ronald G. Collman,² Frederic D. Bushman¹

AUTHOR AFFILIATIONS See affiliation list on p. 14.

ABSTRACT Prolonged infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in immunocompromised patients provides an opportunity for viral evolution, potentially leading to the generation of new pathogenic variants. To investigate the pathways of viral evolution, we carried out a study on five patients experiencing prolonged SARS-CoV-2 infection (quantitative polymerase chain reaction-positive for 79–203 days) who were immunocompromised due to treatment for lymphoma or solid organ transplantation. For each timepoint analyzed, we generated at least two independent viral genome sequences to assess the heterogeneity and control for sequencing error. Four of the five patients likely had prolonged infection; the fifth apparently experienced a reinfection. The rates of accumulation of substitutions in the viral genome per day were higher in hospitalized patients with prolonged infection than those estimated for the community background. The *spike* coding region accumulated a significantly greater number of unique mutations than other viral coding regions, and the mutation density was higher. Two patients were treated with monoclonal antibodies (bebtelovimab and sotrovimab); by the next sampled timepoint, each virus population showed substitutions associated with monoclonal antibody resistance as the dominant forms (*spike* K444N and *spike* E340D). All patients received remdesivir, but remdesivir-resistant substitutions were not detected. These data thus help elucidate the trends of emergence, evolution, and selection of mutational variants within long-term infected immunocompromised individuals.

IMPORTANCE SARS-CoV-2 is responsible for a global pandemic, driven in part by the emergence of new viral variants. Where do these new variants come from? One model is that long-term viral persistence in infected individuals allows for viral evolution in response to host pressures, resulting in viruses more likely to replicate efficiently in humans. In this study, we characterize replication in several hospitalized and long-term infected individuals, documenting efficient pathways of viral evolution.

KEYWORDS COVID-19, SARS-CoV-2, coronavirus, long-term infection, prolonged infection

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has been characterized by the regular emergence of new variants. The origin of these new variants is unclear, with some speculating emergence from zoonotic spillover into other vertebrates and spill back into humans. (1–5). An alternative and widely discussed potential source of new variants is infection in immunocompromised patients—in these patients, the virus is resident long term and so can be exposed to a series of selective pressures in a weakened host that may develop sub-optimal immune responses, allowing changes to accumulate in response in the viral genome (4, 6, 7).

Editor Michael S. Diamond, Washington University in St. Louis School of Medicine, St. Louis, Missouri, USA

Address correspondence to Ronald G. Collman, collmanr@penmedicine.upenn.edu, or Frederic D. Bushman, bushman@penmedicine.upenn.edu.

The authors declare no conflict of interest.

See the funding table on p. 14.

Received 23 January 2024

Accepted 25 January 2024

Published 16 February 2024

Copyright © 2024 Marques et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Several studies have examined the viral genetic changes that occur during infection in immunocompromised individuals, providing compelling evidence that they are linked to unusually high rates of within-host mutation accumulation (6–26). These studies have monitored individuals with various immunocompromising conditions, such as organ transplant patients on immunosuppressive drug treatment, cancer patients undergoing chemotherapy or immunotherapy, patients on immunosuppressive drug treatment for autoimmune disorders, and people with HIV infection. Previous studies have characterized at least 43 human patients and identified multiple examples of immune escape and drug-resistant amino acid substitutions (Table S1). In one study, 13.9% of B cell lymphoma patients infected with SARS-CoV-2 had infections lasting 30 days or greater (27). Some individuals with B cell lymphoma have reduced ability to produce SARS-CoV-2-neutralizing antibodies, placing them at higher risk for prolonged infection (28–30) and failure to respond to vaccination (31, 32). In one example, an individual with B cell lymphoma sustained a SARS-CoV-2 infection for 156 days, with the virus accumulating 16 mutations that included four neutralizing antibody escape substitutions (19). In another example, a group of immunocompromised heart transplant patients independently developed the E484K immune evasion substitution in spike within just 14 days (12). The main driver of selection in many of these cases is thought to result from selective pressure for increased fitness of cell–cell transmission within the host (33). These studies suggest that new SARS-CoV-2 variants may accumulate rapidly in immune-compromised individuals, though this has not been extensively quantified.

In this study, we performed genomic analysis of five immunocompromised patients with suspected prolonged SARS-CoV-2 infections to monitor the emergence of new intrahost single-nucleotide variants (iSNVs). To minimize the influence of sequencing error and account for within-host heterogeneity, we generated at least two independent viral whole-genome sequences per timepoint, allowing us to focus on iSNVs that are reproducibly detected. Using these data, we characterize viral evolutionary dynamics within hosts, monitor the development of drug resistance, and compare evolution rates to those seen in the nonimmunocompromised population.

RESULTS

Genomic signatures of prolonged infection

We enrolled 27 immunocompromised patients persistently infected with SARS-CoV-2 (>21 days positive by the nucleic acid amplification test). Of these, five patients had samples spanning multiple days of collection where viral whole-genome sequencing returned high-quality genomes (Fig. 1). We defined high-quality genomes as having >200 × mean coverage with no more than 2% ambiguous calls. A total of 71 replicates representing 20 unique samples were recovered, with 64 of 71 replicates having a mean coverage >1,000 × (Fig. 2A; Table S2). We then selected samples with at least two independent replicates of the viral genome sequence per timepoint to ensure reproducibility. Only patients with at least two replicates from at least two timepoints were included. Replicates at each timepoint were largely consistent in the mutations identified and the proportional occurrence of each of these mutations (Fig. S1). We recovered data from patients 486, 637, 640, 641, and 663, who had prolonged infections, as observed by quantitative polymerase chain reaction (qPCR)-positive tests, lasting 97, 86, 112, 203, and 79 days, respectively (Fig. S2). From these, viral genomic data were recovered spanning 16, 51, 112, 91, and 19 days of infection (Table S3). Patients 486, 637, 641, and 663 were immunocompromised patients with B cell lymphoma; patient 640 was a heart transplant recipient (Table S3 and S4).

Identifying an example of suspected reinfection

The five patients tested positive by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) for viral RNA over time, but it was possible that the initially infecting strain might have been replaced by a different newly infecting strain over the sampling

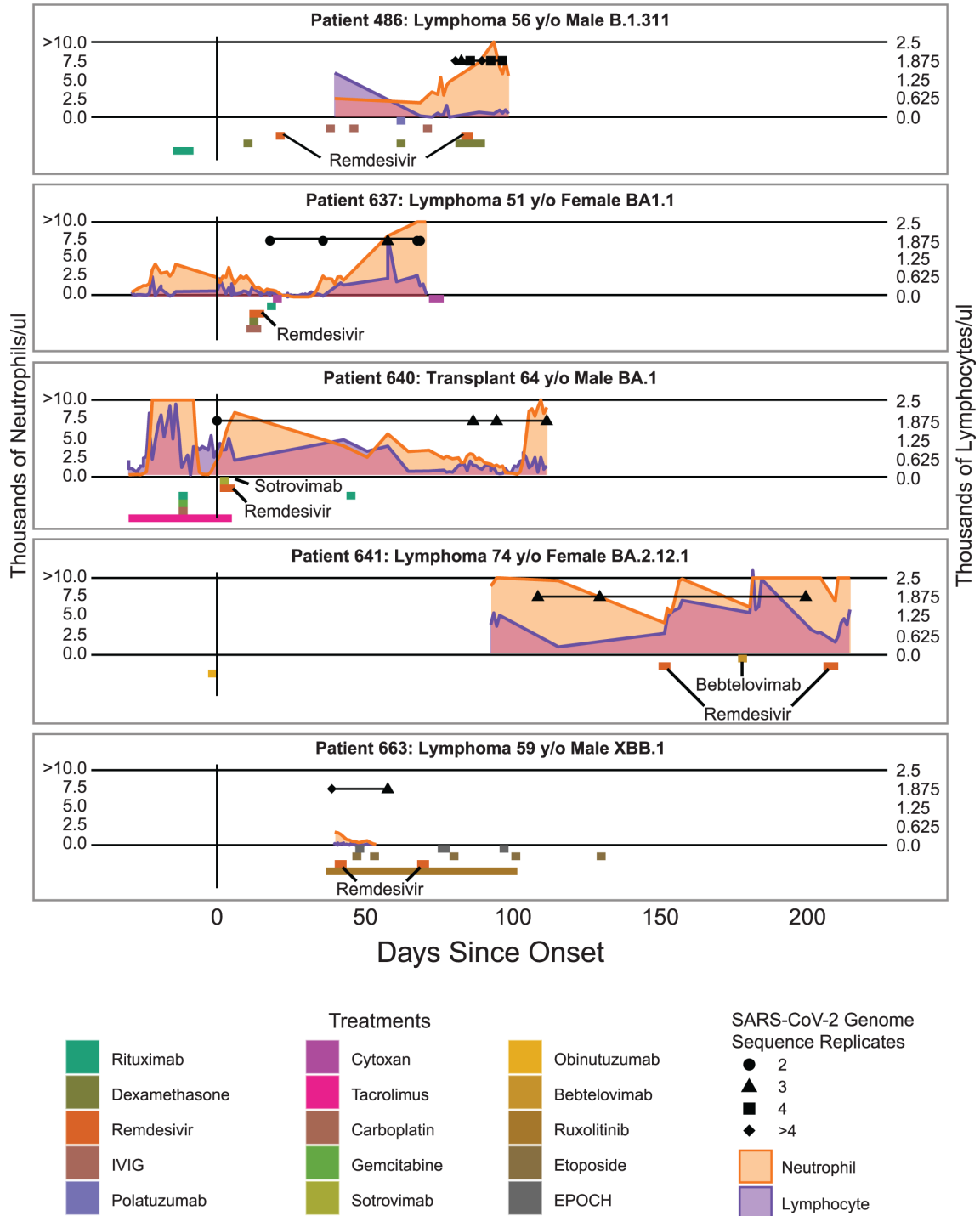


FIG 1 Timeline showing days since symptom onset along the x-axis and patient data on the y-axis. Treatments are colored as boxes below the x-axis, where the length of the box represents the window of treatment administration. Orange and purple line graphs represent absolute neutrophil count (ANC) and absolute lymphocyte count (ALC), respectively, across the left and right y-axis with units of thousands of cells per microliter. Black shapes represent replicates for a given timepoint, and the horizontal black line indicates the period for which sequence data are available. The black vertical line represents the day of onset either by symptom if applicable or by first positive test if initially asymptomatic.

period. We thus generated phylogenetic trees containing genomes for each of the patients together with genomes from the community background using IQ-Tree and Nextclade (34–36). To characterize the community background, we subsampled genomes from the same lineage found within the United States and with 75% of isolates

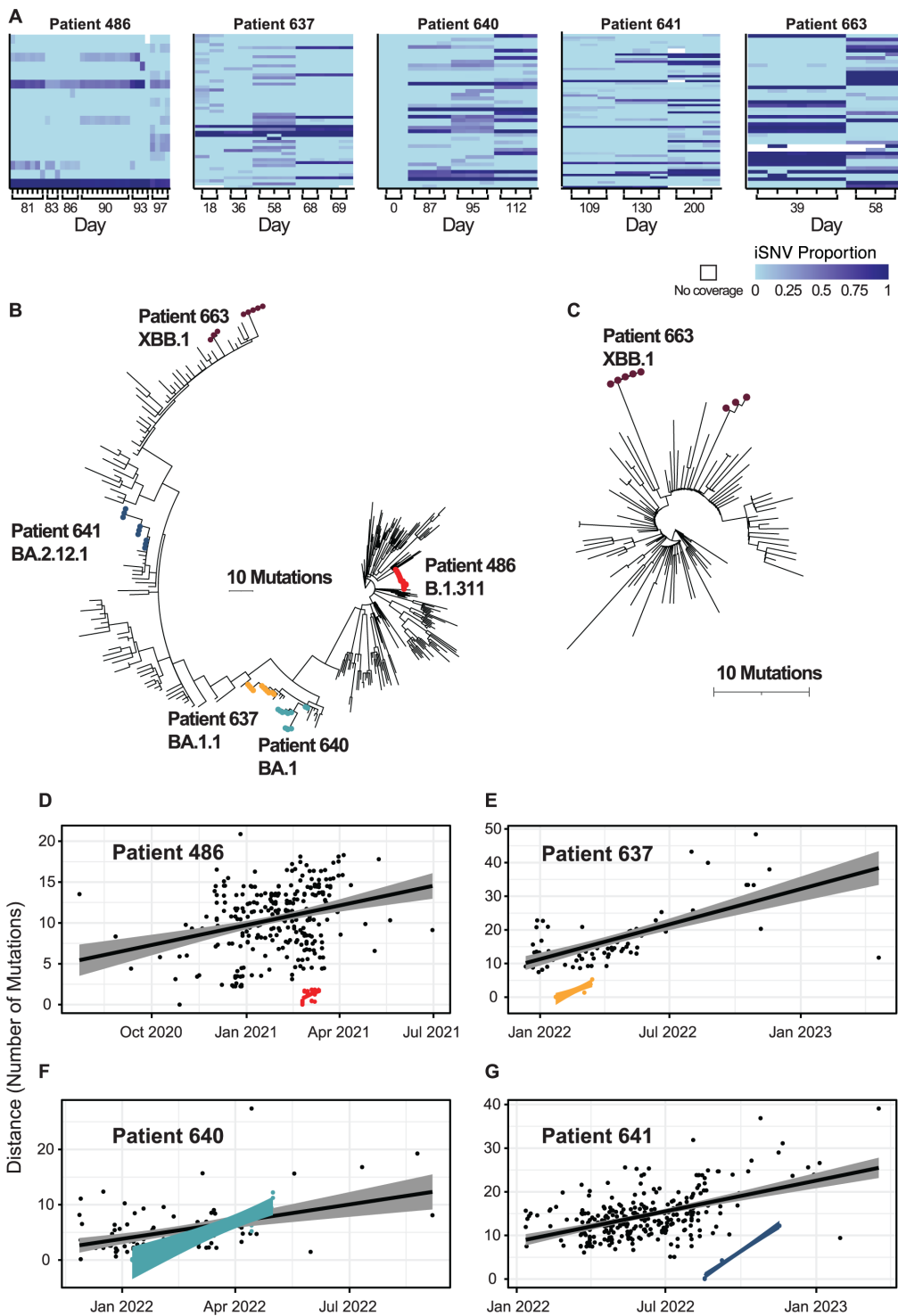


FIG 2 Overview of SARS-CoV-2 evolution in long-term infected individuals. (A) Heatmap showing timepoints and replicates for each patient. Columns represent the genome sequence samples, and rows represent mutations, where light blue indicates a match to the ancestral Wuhan reference strain, and the darker shades indicate increasing prevalence. Only mutations that were found in more than one replicate with >0.03 occurrence and showed changes over the course of infection are included in the heatmap. (B) Phylogenetic tree showing patients in a representative tree covering the course of the pandemic. (C) Phylogenetic tree of patient 663, the case of suspected reinfection, shown in the context of subsampled sequences representing viral lineage XBB.1. (D) Root-to-tip plot for infection in patient 486 (red) and background (black) of the same viral strain, B.1.311. (E) Root-to-tip plot for infection in patient 637 (orange) and background (black) of the same viral strain, BA.1.1. (Continued on next page)

FIG 2 (Continued)

(F) Root-to-tip plot for infection in patient 640 (light blue) and background (black) of the same viral strain, BA.1. (G) Root-to-tip plot for infection in patient 641 (dark blue) and background (black) of the same viral strain, XBB.1.

from our geographical region (Pennsylvania, New Jersey, and Delaware); the remaining 25% of background isolates came from other locations within the United States. Following alignments, IQ-Tree was used to infer maximum-likelihood trees using 1,000 bootstrap replicates (37). These trees contextualize the prolonged infection samples within representative sampling of contemporaneous, genetically similar surveillance isolates. We found that one patient, 663, had two distinct clusters separated by contemporaneous, genetically similar background sequences. Co-infection is not likely, as indicated by a lack of shared iSNVs from the earliest timepoint and later timepoint (Fig. 2A; Table S6); however, super-infection at an intermediate timepoint cannot be ruled out. The discrete phylogenetic clusters of 663 isolates suggest that the two timepoints represented distinct strains and thus a likely reinfection with potential super-infection (Fig. 2B and C; Fig. S3).

Accelerated within-host viral evolution rates

The remaining four patients had all timepoint/replicates clustered together, consistent with prolonged infections caused by a single strain (Fig. S3). Fig. 2B places the prolonged-infection patients in the context of the global population evolution beginning with the emergence of SARS-CoV-2 in late 2019. We observed an accumulation of up to 18 consensus mutations in the patient sampled over the longest period (patient 640, sampled over 112 days).

To estimate and compare evolutionary rates, we calculated the root-to-tip distances from the patient-specific phylogenetic trees with contemporaneous, genetically similar subsampled background isolates (Fig. 2D through G; Fig. S3). Each tree was rooted with the earliest detected isolate of that lineage in the United States with a complete genome submitted to GISAID. The root for the prolonged-infection individual was assigned to the earliest sequence of the highest mean coverage that was recovered; therefore, the pairwise distance to subsequent timepoints represents the genetic distance from the earliest known point. Evolutionary rate is defined in this study as the number of consensus mutations that accumulated from the root to tip per year. This allows for direct comparison of rates of mutation accumulation within a single host versus individuals sampled across the regional population. To estimate the rate of mutation accumulation in each prolonged infection, we modeled the number of consensus mutations that accumulated in the genome as a function of time (in months) with linear regression. For patients with fewer than 2 months of sequence data, the evolution rate of the prolonged-infection patients was higher than background rates but did not achieve significance. The evolution rate for patient 486 was 1.47 consensus mutations per month (95% confidence interval [95% CI], 0–3.12) with the background evolution rate of 0.82 (95% CI, 0.52–1.12), and the evolution rate for patient 637 was 2.82 (95% CI, 1.48–4.16) with the background evolution rate as 1.75 (95% CI, 1.39–2.12). For patients with more than 2 months of sequence data, we observed statistically greater evolution rates than background. The evolution rate for patient 640 was measured as 2.78 consensus mutations per month (95% CI, 1.69–3.87) with the background evolution rate as 0.89 (95% CI, 0.41–1.37), and the evolution rate for patient 641 was 3.83 (95% CI, 3.17–4.5) and the background evolution rate was 1.33 (95% CI, 1.1–1.56). Together, these data suggest that the measured evolution rates of prolonged-infection immunocompromised individuals were 1.6–3.1 times higher than those of contemporaneous samples of a similar genetic background. For sampling windows that were longer than 2 months in duration, we measured a greater evolutionary rate in immunocompromised prolonged infections than expected compared with the community background, which is consistent with the finding of a previous report (38).

Accumulation and persistence of minor variants

To investigate changes in viral sequences in more detail, we examined the emergence and changing proportions of intrahost single-nucleotide variants (iSNV) over the sampling period. Fig. 3A through D displays the 10 most rapidly changing iSNVs for each of the patients. For patient 640, data were available for the earliest stage of infection. This individual was initially asymptomatic when first tested positive, with subsequent symptom onset; therefore, day 0 represents the first positive test. Unique to our data set, this is the only timepoint with no iSNVs present in our data set, suggesting the dominance of a single viral strain (Fig. 3C). The absence of iSNVs at this patient's timepoint is consistent with previous findings of tight bottlenecks at transmission (39). At later timepoints for this patient, and for all other patients' timepoints, iSNVs are detectable across replicates.

We hypothesized that over the course of infection, there would be an increased accumulation of iSNVs. For the remainder of the study, we classify iSNVs as a mutation within the cutoff range detected in two or more replicates at a single timepoint. For example, a cutoff value of 0.01 would account for iSNVs that are between 0.01 and 0.99 of reads for that position. Not surprisingly, we found that at lower thresholds, there was an increasing accumulation of iSNV compared to higher thresholds, which are largely unchanged over the duration of infection (Fig. 4A and B; Fig. S4). At a cutoff of 0.01, we found the accumulation of iSNVs at a rate of 10.9 (95% CI, -4.8 to 26.6), 2.3 (95% CI, -6.6 to 11.2), 4.6 (95% CI, 1.9 to 7.3), and 1.6 (95% CI, -6.7 to 10.0) SNVs per month for patients 486, 637, 640, and 641. At this threshold, an average of 4.9 iSNVs accumulated with each month of infection. Relatively low abundance iSNVs increased in frequency with time, while the frequency of more abundant iSNVs remained more constant.

iSNV mutation density within the SARS-CoV-2 genome

To estimate whether iSNVs accumulate preferentially in specific coding regions compared to others, we aggregated iSNVs by the coding region and normalized based on coding region size. The iSNV counts per coding region are depicted in Fig. 4C. We found that coding regions with the fewest detected iSNVs include *nsp12* (RNA-dependent RNA-polymerase), *nsp14* (exonuclease), and *nsp2* (reprograms host translation machinery) (40; Fig. 4D). These coding regions are also found to be less variable over the course of the global pandemic (41–43). The coding regions with the highest iSNV counts were *spike*, *nucleocapsid*, and *envelope*. *Spike* was the only coding region with a statistically significant greater accumulation of iSNVs compared to other coding regions after using the Benjamini–Hochberg procedure to control for the false discovery rate (adj. $P = 0.014$). Across the four patients, we identified 41 unique *spike* iSNVs (Fig. 4D). Moreover, we found that positions with iSNVs across more than one patient were consistently nonsynonymous substitutions (Fig. 4D). Thus, in immunocompromised hosts, we observe preferential accumulation of substitutions in *spike* relative to the rest of the genome, as has been observed in the general population (44).

Within-host evolution and selection

To assess the evolution and selection within patients, we used linear regression to estimate the rate of iSNV change in proportion over time within each infected patient for all iSNVs that appear in multiple replicates with >3% occurrence (Fig. S5). This threshold was selected because it was effective at removing false-positive iSNVs that were not reproducible in replicate sequencing (39, 45, 46). Slopes provided estimates of rates of change in iSNV frequency per month (Fig. 5A).

We identified several mutations in our patients, which were concordant with published data on long-term infected patients. Helicase L1220 is a position that rarely experiences mutations in random surveillance samples (<0.1% of uploaded sequences, 10,414/16,351,920, GISAID data accessed 12/20/2023) but has previously been identified as having mutations including L1220I in lymphoma patients with depleted B cells; here,

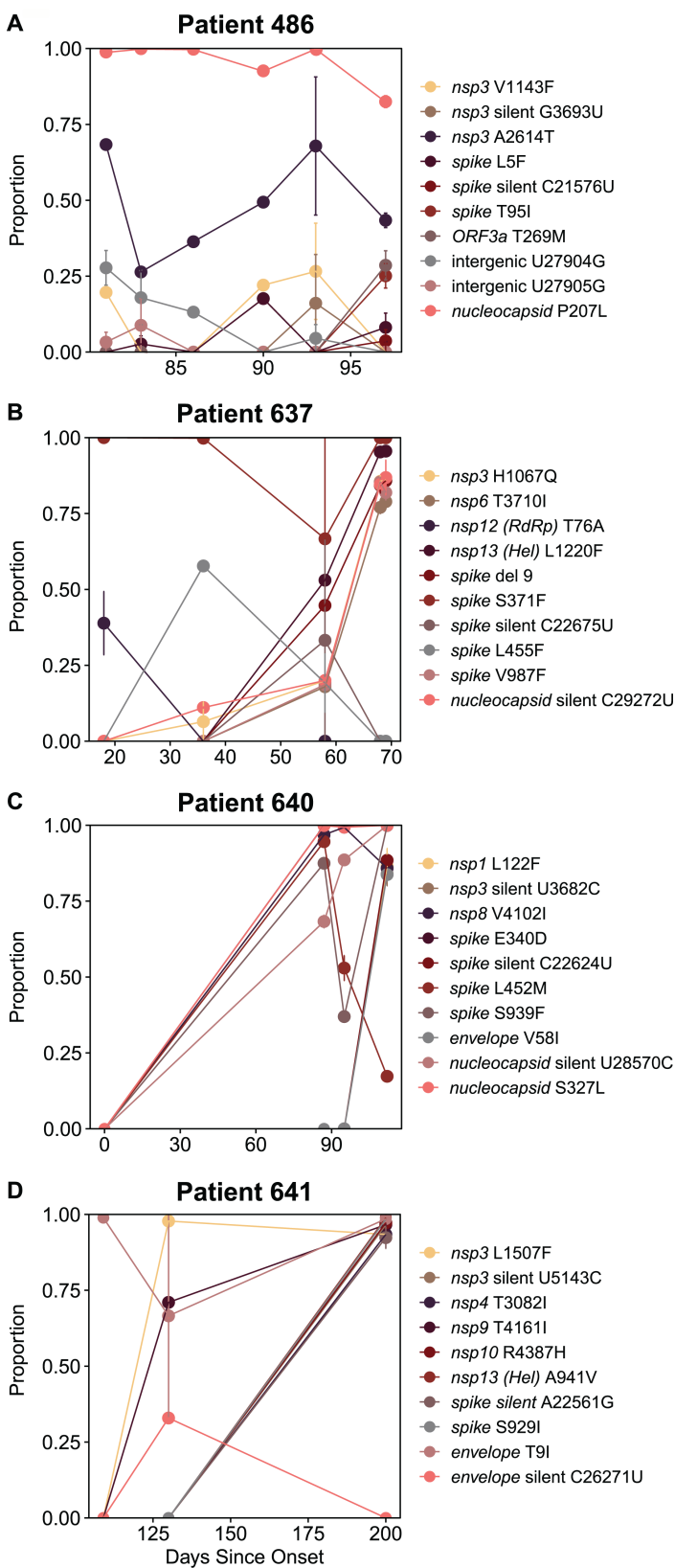


FIG 3 Summary of the top 10 most variable iSNVs within each patient. Different mutations are distinguished by color. Error bars show variability across replicates; lack of error bars indicates less than 0.001 deviation in proportion. (A) Patient 486's top 10 most variable iSNVs. (B) Patient 637's top 10 most variable iSNVs. (C) Patient 640's top 10 most variable iSNVs. (D) Patient 641's top 10 most variable iSNVs.

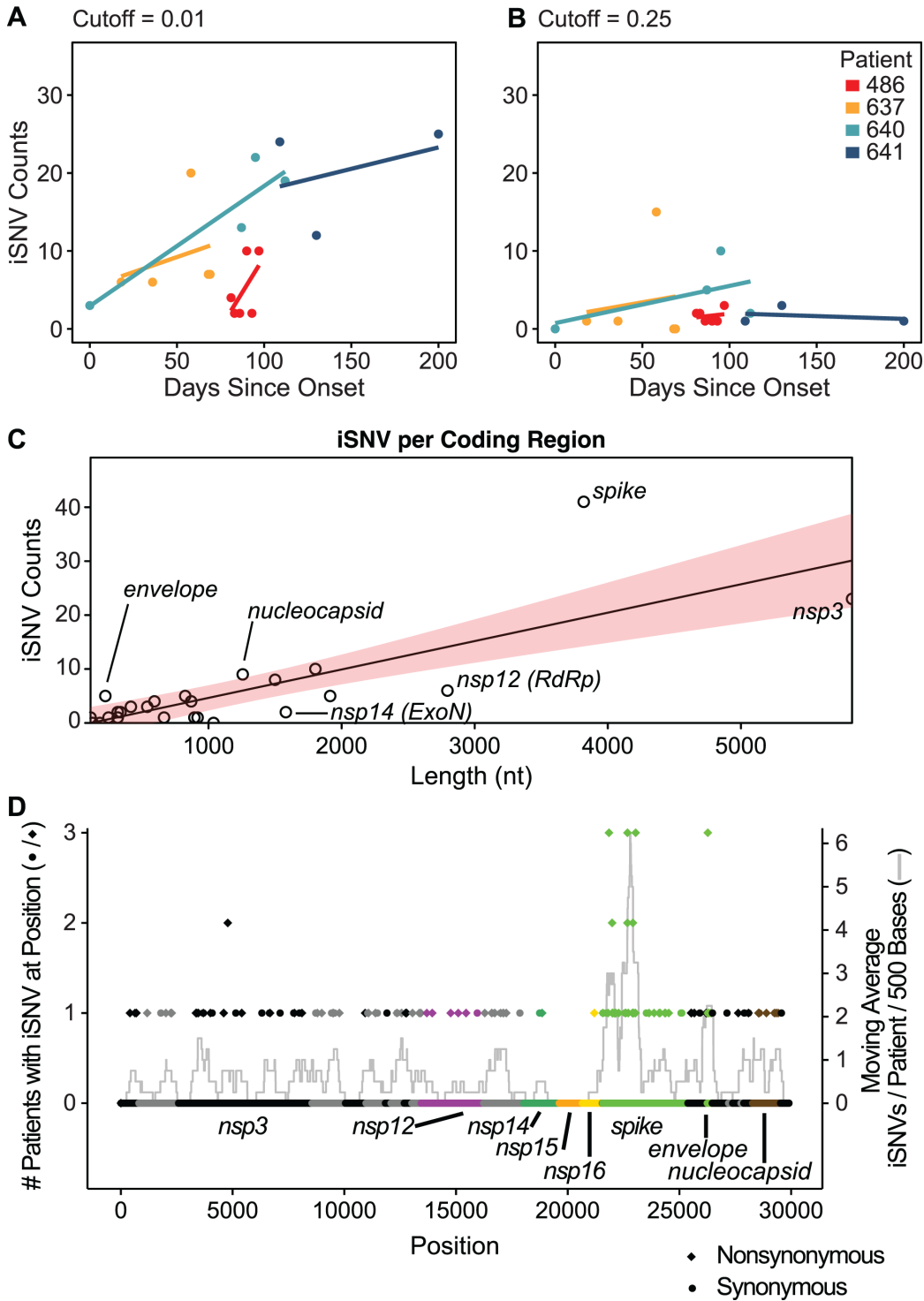


FIG 4 iSNV accumulation over time and by the coding region. (A) iSNV accumulation across patients, shown over the full duration of sampled infection. Linear regression lines were fitted for a cutoff of 0.01 (iSNV found between 1% and 99% prevalence). (B) iSNV accumulation for a cutoff of 0.25 (25%–75% prevalence). (C) Number of iSNVs by the coding region showing length as a nucleotide on the x-axis and iSNV counts on the y-axis. Shaded region indicates the 95% CI for linear regression of coding region length to iSNV counts. (D) Occurrence of iSNVs throughout the genome by position. The left axis indicates the number of patients with any iSNV in a given position, indicated by either circles or diamonds. Circles indicate that the iSNV was synonymous, and diamonds indicate that the iSNV was nonsynonymous. The right axis indicates the moving average centered on each position spanning a window of 500 bases. Units for the moving average are average iSNVs per patient per 500 bases. Coding regions are colored in alternating black or gray, with the regions of interest labeled and colored.

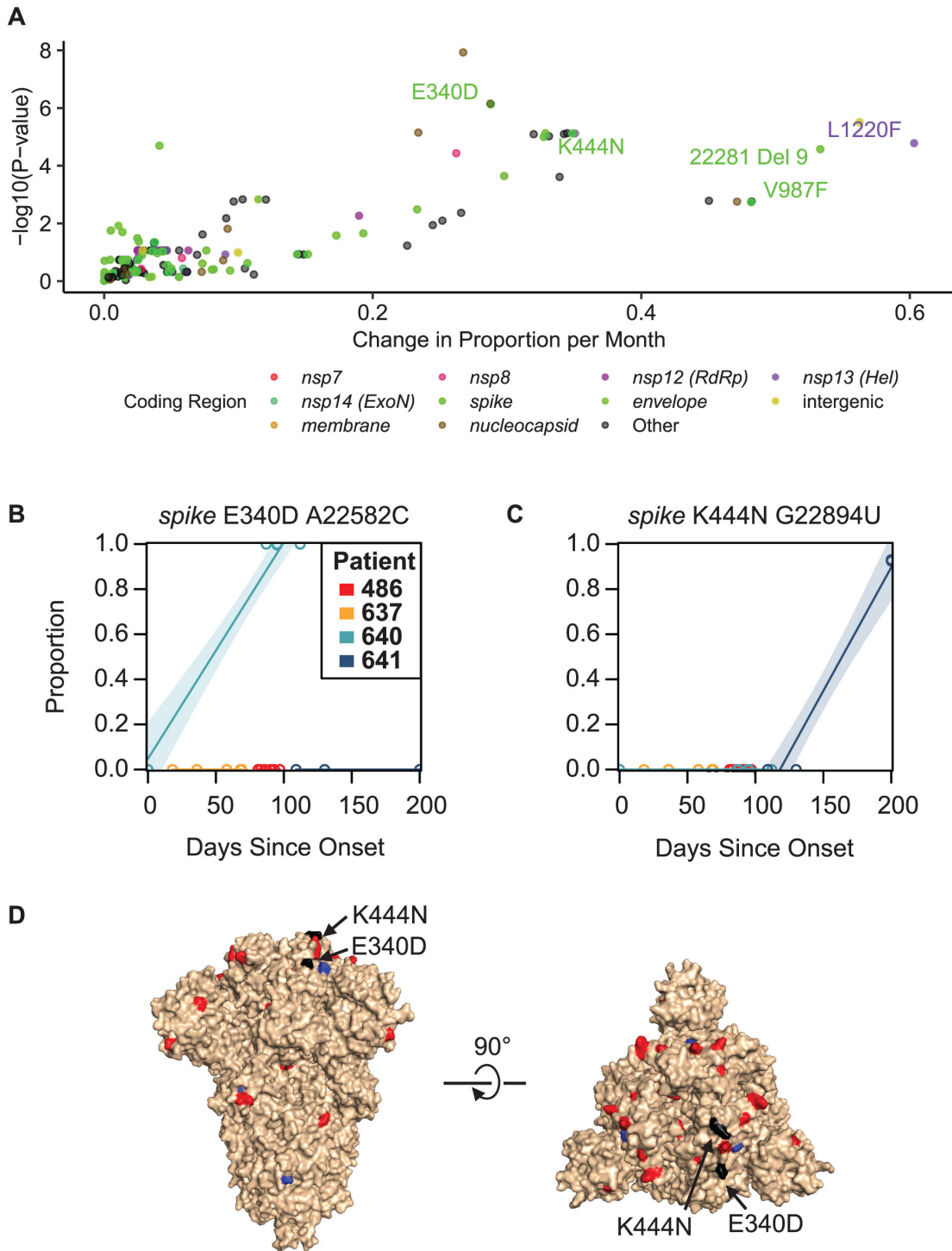


FIG 5 Rate of change of mutations and evasion of monoclonal antibody therapy. (A) Plot showing the change in proportion per month for iSNVs and *P*-values indicating a certainty that there was a change in iSNV proportion, colored by the coding region. (B) Plot showing *spike* E340D, which confers sotrovimab resistance, across all patients. (C) Plot showing *spike* K444N substitution, which confers resistance to bebtelovimab, across all patients. (D) Structure of SARS-CoV-2 *spike* showing nonsynonymous iSNVs as red, synonymous iSNVs as blue, and known drug-resistant mutations that rose to dominate in our cohort marked in black.

we identified L1220F in patient 637 (47–50). *Spike* V987F emerged in patient 637, which was reported previously in a B cell lymphoma patient with prolonged infection but is rarely observed in surveillance samples (<0.01% of uploaded sequences,

1,084/16,351,920, GISAID data accessed 12/20/2023) (48–51). Kemp et al. identified a decrease in *spike* T240I in a prolonged-infection individual; we identified a deletion of nine nucleotides in this region (starting at position 22281), suggesting another common site found in immunocompromised individuals (15). We speculate that the presence of these mutations in prolonged infections could be explained by a replication fitness advantage in immunocompromised subjects, while in immunocompetent subjects, these mutations allow immune recognition, which constrains their emergence.

Antibody and drug escape mutations

We hypothesized that viruses replicating in patients treated with antiviral drugs or monoclonal antibodies (mAbs) would develop mutations conferring drug resistance. Two of the patients received mAb treatments (bebtelovimab or sotrovimab), which target the SARS-CoV-2 spike protein to neutralize the virus (37, 39). In both cases, mAb-resistant mutations replaced the dominant amino acid by the next timepoint following antibody administration (Fig. 5B through D). The bebtelovimab-resistant mutation, *spike* K444N, increased from a prevalence of 0% to 92.7% within 28 days post-treatment. The sotrovimab-resistant mutant *spike* E340D went from 0% to 99.9% prevalence within 87 days post-treatment. These findings are consistent with previous reports where mAbs were administered and resistance was detected (7, 9, 24).

In contrast to mAb treatment, the drug remdesivir was administered to all of the patients studied here, but known remdesivir-resistant mutations were not detected (52, 53).

DISCUSSION

Understanding evolutionary trajectories in SARS-CoV-2 is critical for pandemic preparedness and optimizing therapeutic strategies. Surveillance in immunocompromised individuals with prolonged infections allows for a greater understanding of potential sources of new variants that may emerge and circulate in the wider population. Treatment regimens used in these individuals may result in accumulation of new drug-resistant substitutions that could result in less effective future treatments after transmission to the broader community. In this study, we performed genomic analysis of five potential SARS-CoV-2 prolonged infections, finding four of five to be likely true prolonged infections—the fifth was a likely reinfection. For inclusion, we required recovery of at least two independent viral whole-genome sequences from at least two timepoints, allowing for reproducible detection of novel sequence variants.

There have been several prior case studies identifying the emergence of drug resistance in immunocompromised individuals (7, 14, 21, 38). Table S1 summarizes data for 43 patients over 26 publications. Of these, nine of the 10 who were reported to be treated with mAbs developed substitutions in the spike protein previously documented to confer resistance (7, 9, 12, 21). For our two patients treated with mAbs, *spike* K444N, conferring resistance to bebtelovimab, and *spike* E340D, conferring resistance to sotrovimab, emerged and became dominant by the next timepoint after treatment initiation. Bamlanivimab treatment has also been associated with development of *spike* E484K and E484Q antibody-resistant mutations in immunocompromised patients (12). Mutations in position 484 were commonly observed in studies of immunocompromised patients, but in our study, position 484 remained constant and contained the omicron-specific substitution E484A (11–13, 16, 18). Multiple further substitutions in spike have been reported in immunocompromised subjects but were not seen here (Table S1). Our findings thus are generally consistent with previous findings of the emergence of mAb resistance but emphasize the speed with which such changes can reach fixation (54, 55).

All patients in this study were treated with remdesivir, but we found no mutations annotated as conferring remdesivir resistance. In previous literature, 20 long-term infected patients were reported to be treated with remdesivir, and only one was reported to develop resistance mutations (6, 7, 9, 10, 12–17, 21, 24, 56–61) (Table S1). nsp12 E802D emerged as a remdesivir-resistant mutation in a patient with acquired B cell deficiency,

suggesting that remdesivir drug resistance development is possible but not inevitable (21) and is less frequent than accumulation of resistance mutations to monoclonal antibodies. Possibly remdesivir resistance mutations confer a greater fitness cost to the virus, and therefore do not accumulate despite treatment (62).

We measured within-host evolution rates, quantified as the number of consensus substitutions per month, to be about two times higher than that observed with interhost background surveillance samples in three out of four of our patients. Over all four patients, a statistically greater mutational burden was documented in the *spike* coding region. The median estimated evolution rate across previous studies was 3.1 consensus substitutions per month. While some case studies reported no difference between intrahost and interhost evolution rates (16, 19, 20), other groups reported intrahost rates greater than the published rate for interhost variation (6, 7, 13, 15–17, 63) (Table S1). Accumulation of interhost substitutions in acute infections in otherwise healthy individuals appears to be minimal, although more data would be helpful to compare the evolution rates of acute and prolonged infections (46, 64, 65). These data suggest a high degree of variability within prolonged infection cases, although it is important to note that not all studies used the same approach to calculate evolution rates. A strength of our analysis is that we calculated the intrahost rate and interhost rate simultaneously using a representative sampling of contemporaneous genetically similar background isolates, facilitating direct and statistically robust comparison. These data highlight the potential for new viral evolution in immunocompromised patients.

Our study has several potential limitations. Our samples were collected as part of a retrospective study, as opposed to a prospective study with enrollment before infection. This results in sample collection that depends on the availability of previously banked specimens. In addition, our cohort size limited the statistical power of our analysis. The inability to reconstruct viral haplotypes is a current technical challenge in the field for within-host viral evolution studies, with the current 600-cycle Illumina kits not long enough to span most unique mutations for reconstruction. Additionally, this study represents a conservative estimate for evolutionary rates due to limitations of timepoints available for sampling. Some of the mutations most under selective pressure rise to dominance between sampling timepoints; therefore, it is possible that they dominated in fewer days than we estimated. Consequently, the estimates of fixation for iSNVs without intermediate timepoints collected represent a maximum amount of time. The patient's iSNVs may have changed more rapidly than we could measure, given collection timepoints available. We did not have serum or blood cells available to monitor relevant immune responses. Interhost evolution rates for the BA.1 background appeared faster than other within-lineage Omicron rates (66). A possible explanation for this could be that the background included recombinant sublineages grouped in with BA.1 as labeled through GISAID's lineage calls. Accumulation of iSNVs was analyzed using linear regression due to the number of timepoints available for each patient—with more timepoints collected a more thorough analysis could assess whether linear curves best represent the data. Regressions that place zero-intercept before timepoint 0 suggest that the initial infection may have been seeded by a heterogeneous mixture. Finally, the study would have benefited from having a cohort of acute SARS-CoV-2 infections in otherwise healthy individuals with multiple timepoints to compare within-host evolution to immunocompromised individuals directly.

This study describes the evolutionary changes in immunocompromised, prolonged-infection individuals, including responses to drug treatment pressures. We identified efficient evolutionary escape from monoclonal antibody therapy, though not from remdesivir therapy. The emergence and onward transmission of mAb-resistant mutations could undermine the effectiveness of this element of our current therapeutic arsenal; our study provides an insight into outcomes in individual patients. Given a higher rate of SARS-CoV-2 evolution observed in a majority of our patients, our results underscore the importance of closely monitoring immunocompromised individuals as sources for new and concerning viral variants.

MATERIALS AND METHODS

Human subjects

Patients hospitalized at the Hospital of the University of Pennsylvania were enrolled following informed consent received under institutional review board-approved protocol #823392. Sample types included oropharyngeal and nasopharyngeal swabs or saliva, as previously described (67). Clinical metadata were manually abstracted from EPIC, the electronic medical record system used by the Hospital of the University of Pennsylvania. These abstracted data included the patients' medical comorbidities, laboratory results, and medications administered. The laboratory data comprised lymphocyte and neutrophil counts recorded from the month prior to the date of the initial viral isolate to the date of the final isolate; both automated and manual cell counts were incorporated. Medication details provided the dates of administration spanning from the month prior to the initial viral isolate to the date of the final isolate, focusing on the following medication classes: anti-neoplastic agents, immunomodulators, and SARS-CoV-2-directed therapies. Medication administration was deemed continuous if doses were administered less than 5 days apart.

Sequencing methods

The ARTIC POLAR protocol was used to obtain viral genomic sequences (doi: <https://doi.org/10.1101/2020.04.25.061499>). Sequences were obtained on Illumina NextSeq. Library preparation was performed as follows: a pre-reverse transcription reaction was performed with 5 μ L of viral RNA, 0.5 μ L of random hexamers at 50 μ M (Thermo Fisher, N8080127), 0.5 μ L of a 10 mM deoxynucleoside triphosphate (dNTP) Mix (Thermo Fisher, 18427013), and 1 μ L of nuclease-free water heated for 5 minutes at 65°C and then incubated for 1 minute at 4°C. This followed a reverse transcription reaction using 6.5 μ L of the above mixture combined with 0.5 μ L of SuperScript III Reverse Transcriptase (Thermo Fisher, 18080085), 2 μ L of 5 \times First-Strand Buffer (Thermo Fisher, 18080085), 0.5 μ L of 0.1 M dithiothreitol (DTT) (Thermo Fisher, 18080085), and 0.5 μ L of RNaseOut (Thermo Fisher, 18080051). This was incubated at 42°C for 50 minutes, at 70°C for 10 minutes, and then incubated at 4°C. The resulting amplicons from both primer sets for the sample were combined and then diluted to 0.25 ng/ μ L. For the cDNA amplification, artic-ncov2019 version 4.1 primers from IDT were used. The SARS-CoV-2 PCR was prepared using 2.5 μ L of the previous mixture with 0.25 μ L Q5 Hot Start DNA polymerase (NEB, M0493S), 0.5 μ L of 10 mM dNTP Mix (NEB, N0447S), 5 μ L of 5 \times Q5 reaction buffer (NEB, M0493S), and either 4.0 μ L of primer set 1 or 3.98 μ L of primer set 2 with nuclease-free water to achieve a 25 μ L total volume. The PCR conditions were set to 98°C for 30 s (single cycle), followed by 25 cycles of 98°C for 15 s and 65°C for 5 minutes, and concluded at 4°C. The library was then prepared using the Nextera XT Library Preparation kit (Illumina, FC-131–1096) using IDT for Illumina DNA/RNA UD Indexes Set A, B, C, and D (Illumina, 20027213–20027216). The DNA content for each sample was gauged using the Quant-iT PicoGreen dsDNA quantitation kit (Invitrogen, P7589). After pooling samples in equal concentrations, the combined library's quantity was determined using the Qubit1X dsDNA HS assay kit (Invitrogen, Q33230), and sequencing was performed on the Illumina NextSeq using a P1 2 \times 300 chemistry.

Sequence assembly

Sequence reads were first filtered to remove bases with a quality below Q20. These trimmed sequences were then aligned to the original Wuhan reference sequence ([NC_045512.2](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2)) using the BWA aligner tool (v0.7.17). The Samtools package (v1.10) was used for filtering alignments. Variant positions were identified with the Bcftools package (v1.10.2–34), using PHRED scores of 20 or higher and variant read frequencies that constitute 50% or more of total reads. Variants were categorized using the Pangolin lineage software, specifically Pangolin version 4.1.3 coupled with pangolin-data 1.16.

Point mutations were classified using a previously published bioinformatics pipeline (67). Key reagents are outlined in Table S5.

Phylogenetic trees

Nextstrain's augur tools (CLI v7.1.0) were used to generate representative subsamples of background surveillance data for each of the lineages detected in the five suspected prolonged-infection individuals. The earliest detection of the lineage, in the United States, with a complete, high-coverage genome was used as the root. Trees include all prolonged-infection samples and a 1:3 split of one USA sequence for every three tri-state sequences (Pennsylvania, New Jersey, or Delaware) to allow for a focused sampling in our region and representative samples from across the country. Trees were generated with 150–400 subsampled background sequences in addition to our prolonged infection samples. To generate trees, subsampled sequences were aligned with Nextclade v2.14.0, and maximum-likelihood phylogenetic trees with 1,000 bootstraps were generated using IQ-Tree v1.6.12³⁷. Tree visualization was performed using iTOL v6.

Root-to-tip analysis

Root-to-tip analysis was performed using the custom code to estimate evolution rates. Phylogenetic trees described above were analyzed using Ape (68) to compute the pairwise distance between the pairs of tips from the rooted phylogenetic trees. The root of the background sequences was selected as the earliest high-quality detection of the patient's lineage in the United States, and the root for the prolonged-infection patient was the earliest sequenced timepoint collected. For each patient, the background and patient linear regressions are computed as follows:

$$\text{distance} = \beta_0 + \beta_1 * \text{date} + \varepsilon,$$

where β_0 represents the intercept, β_1 represents the coefficient for *date*, and ε represents the error term. The 95% CI was calculated after estimating the standard error (*SE*) and the t-value using $df = n - 2$. These values were used to calculate the margin of error (*MOE*), where $MOE = t_{\text{value}} * SE$. The 95% CI for each coefficient β is estimated as Lower limit of C.I. = $\beta - MOE$ and Upper limit of C.I. = $\beta + MOE$.

iSNV statistical analysis

Major variants were called using bcftools, and iSNVs were called with bbtools (69, 70). Reads were filtered for a minimum mapping score of 40 from BWA and a quality score of 30. Patients were considered for analysis if they had two or more timepoints with two or more high-quality replicates per timepoint. High-quality replicates are defined as having 98% coverage with $>200 \times$ mean coverage, although 64 of our 71 samples had $>1,000 \times$ mean coverage. Regions with no coverage (containing base calls "Ns") were excluded from any analysis. Thresholds for iSNVs were assessed at cutoffs for 0.01, 0.03, 0.05, 0.1, 0.15, 0.2, and 0.25. A cutoff of 0.03 with a required detection in two or more replicates was used to call subsequent iSNVs. All insertions, deletions, and substitutions (synonymous and nonsynonymous) that met these criteria were included. Coding region locations and lengths were used as described in [NC_045512.2](#) (71). To determine if any coding regions had more or fewer than expected iSNVs, a linear regression was computed as defined by $n = \beta_0 + \beta_1 * \text{length} + \varepsilon$, where n represents the number of predicted unique iSNVs given the coding region length, β_0 represents the y-intercept, β_1 represents the coefficient associated with the coding region length, and ε is the error term capturing the variability in mutations not explained by coding region length. Adjusted *P*-values were obtained after calculating the standardized residuals (*SR*), where $SR = \frac{\text{residuals}}{\text{standard deviation of residuals}}$. Using the t-distribution, the two-tailed test's *P*-value for the *SR* was computed and then corrected for by using the Benjamini–Hochberg procedure

to control for the false discovery rate. To estimate the rates of change in iSNVs over time, linear regression was used as described by $p = \beta_0 + \beta_1 * date + \epsilon$, where β_0 represents the y-intercept, β_1 represents the coefficient associated with the coding region length, ϵ is the error term capturing the variability in mutations not explained by the coding region length, and p represents the proportion of reads containing n number of predicted unique iSNVs given the coding region length.

ACKNOWLEDGMENTS

We are grateful to members of the Bushman and Collman laboratories for help and suggestions, to patients who enrolled in the study, and to the nursing staff for assistance with sampling.

This work was supported in part by the Penn University Research Foundation. Funding was provided by a contract award from the Centers for Disease Control and Prevention (CDC BAA 200–2021-10986 and 75D30121C11102/000HCVL1-2021-55232), the Penn Center for Global Genomics and Health Equity Keystone Pilot Grant (GGHE-KP-2021–001), philanthropic donations to the Penn Center for Research on Coronaviruses and Other Emerging Pathogens, and in part by NIH grant R61/33-HL137063 and AI140442-supplement for SARS-CoV-2. B.J.K. was supported by NIH K23 AI 121485. Additional assistance was provided by the Penn Center for AIDS Research (P30-AI045008).

AUTHOR AFFILIATIONS

¹Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

²Division of Pulmonary, Allergy, and Critical Care, Philadelphia, Pennsylvania, USA

³Division of Infectious Diseases, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, Pennsylvania, USA

⁴Department of Pathobiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

AUTHOR ORCID*s*

Andrew D. Marques  <http://orcid.org/0000-0001-8887-7242>

Brendan J. Kelly  <http://orcid.org/0000-0003-2023-646X>

Ronald G. Collman  <http://orcid.org/0000-0002-0508-3701>

Frederic D. Bushman  <http://orcid.org/0000-0003-4740-4056>

FUNDING

Funder	Grant(s)	Author(s)
HHS Centers for Disease Control and Prevention (CDC)	CDC BAA 200-2021-10986, 75D30121C11102/000HCVL1-2021-55232	John Everett Kyle G. Rodino Louise H. Moncla Brendan J. Kelly Ronald G. Collman Frederic D. Bushman Andrew D. Marques
HHS National Institutes of Health (NIH)	R61/33-HL137063, AI140442-supplement for SARS-CoV-2, P30-AI045008	Ronald G. Collman Frederic D. Bushman
HHS National Institutes of Health (NIH)	K23 AI 121485	Frederic D. Bushman

Funder	Grant(s)	Author(s)
Penn Center for Global Genomics	GGHE-KP-2021-001	Andrew D. Marques Frederic D. Bushman
Penn Center for Research on Coronaviruses and Other Emerging Pathogens		Andrew D. Marques Jevon Graham-Wooten Ayannah S. Fitzgerald Ashley Sobel-Leonard Emma J. Cook John Everett Kyle G. Rodino Louise H. Moncla Brendan J. Kelly Ronald G. Collman Frederic D. Bushman

AUTHOR CONTRIBUTIONS

Andrew D. Marques, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Writing – original draft, Writing – review and editing | Jevon Graham-Wooten, Resources, Writing – review and editing | Ayannah S. Fitzgerald, Resources, Writing – review and editing | Ashley Sobel Leonard, Resources, Writing – review and editing | Emma J. Cook, Resources, Writing – review and editing | John K. Everett, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review and editing | Kyle G. Rodino, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review and editing | Louise H. Moncla, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review and editing | Brendan J. Kelly, Conceptualization, Formal analysis, Funding acquisition, Investigation, Resources, Writing – original draft, Writing – review and editing | Ronald G. Collman, Conceptualization, Formal analysis, Funding acquisition, Investigation, Resources, Writing – original draft, Writing – review and editing | Frederic D. Bushman, Conceptualization, Formal analysis, Funding acquisition, Investigation, Resources, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Consensus sequences can be accessed on GISAID and GenBank with accessions listed in Table S2. Isolates used to generate phylogenetic trees can be accessed through their respective GISAID EPI_SET IDs. Patient 486 background isolates EPI_SET_231026xg (doi: [10.55876/gis8.231026xg](https://doi.org/10.55876/gis8.231026xg)). Patient 637 background isolates EPI_SET_231026tg (doi: [10.55876/gis8.231026tg](https://doi.org/10.55876/gis8.231026tg)). Patient 640 background isolates EPI_SET_231026pw (doi: [10.55876/gis8.231026pw](https://doi.org/10.55876/gis8.231026pw)). Patient 641 background isolates EPI_SET_231026yv (doi: [10.55876/gis8.231026yv](https://doi.org/10.55876/gis8.231026yv)). Patient 663 background isolates EPI_SET_231026um (doi: [10.55876/gis8.231026um](https://doi.org/10.55876/gis8.231026um)). Variant calls can be accessed in Table S6. Raw data can be accessed at BioProject [PRJNA1011815](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1011815). Computer code used can be accessed at <https://github.com/andrewdmarques/Prolonged-Infection-Analysis>

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Figure S1 (mBio00110-24-s0001.tif). Deviation of iSNV proportion by replicate.

Figure S2 (mBio00110-24-s0002.eps). Patient-specific testing results and medical records for symptom onset.

Figure S3 (mBio00110-24-s0003.eps). Patient-specific phylogenetic trees containing lineage-specific background.

Figure S4 (mBio00110-24-s0004.eps). iSNV by day.

Figure S5 (mBio00110-24-s0005.tif). Linear regression for iSNV viral population rates of change.

Tables S1 to S6 (mBio00110-24-s0006.xlsx). Tables S1 to S6.

REFERENCES

- Lu L, Sikkema RS, Velkers FC, Nieuwenhuijse DF, Fischer EAJ, Meijer PA, Bouwmeester-Vincken N, Rietveld A, Wegdam-Blans MCA, Tolsma P, Koppelman M, Smit LAM, Hakze-van der Honing RW, van der Poel WHM, van der Spek AN, Spiereburg MAH, Molenaar RJ, Rond J de, Augustijn M, Woolhouse M, Stegeman JA, Lycett S, Oude Munnink BB, Koopmans MPG. 2021. Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands. *Nat Commun* 12:6802. <https://doi.org/10.1038/s41467-021-27096-9>
- Pickering B, Lung O, Maguire F, Kruczkiewicz P, Kotwa JD, Buchanan T, Gagnier M, Guthrie JL, Jardine CM, Marchand-Austin A, et al. 2022. Divergent SARS-CoV-2 variant emerges in white-tailed deer with deer-to-human transmission. *Nat Microbiol* 7:2011–2024. <https://doi.org/10.1038/s41564-022-01268-9>
- Marques AD, Sherrill-Mix S, Everett JK, Adhikari H, Reddy S, Ellis JC, Zelif H, Greening SS, Cannuscio CC, Strelau KM, Collman RG, Kelly BJ, Rodino KG, Bushman FD, Gagne RB, Anis E, Dandekar S. 2022. Multiple introductions of SARS-CoV-2 alpha and delta variants into white-tailed deer in Pennsylvania. *mBio* 13:e0210122. <https://doi.org/10.1128/mbio.02101-22>
- Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, Katzourakis A. 2023. The evolution of SARS-CoV-2. *Nat Rev Microbiol* 21:361–379. <https://doi.org/10.1038/s41579-023-00878-2>
- Munnink BBO, Nijhuis RHT, Worp N, Boter M, Weller B, Verstrepen BE, GeurtsvanKessel C, Corsten MF, Russcher A, Koopmans M. 2022. Highly divergent SARS-CoV-2 alpha variant in chronically infected immunocompromised person. *Emerg Infect Dis* 28:1920–1923. <https://doi.org/10.3201/eid2809.220875>
- Borges V, Isidro J, Cunha M, Cochicho D, Martins L, Banha L, Figueiredo M, Rebelo L, Trindade MC, Duarte S, Vieira L, Alves MJ, Costa I, Guiomar R, Santos M, Cortê-Real R, Dias A, Póvoas D, Cabo J, Figueiredo C, Manata MJ, Maltez F, Gomes da Silva M, Gomes JP, Paul Duprex W. 2021. Long-term evolution of SARS-CoV-2 in an immunocompromised patient with non-Hodgkin lymphoma. *mSphere* 6:e0024421. <https://doi.org/10.1128/mSphere.00244-21>
- Gonzalez-Reiche AS, Alshammery H, Schaefer S, Patel G, Polanco J, Carreño JM, Amoako AA, Rooker A, Cognigni C, Floda D, et al. 2023. Sequential intrahost evolution and onward transmission of SARS-CoV-2 variants. *Nat Commun* 14:3235. <https://doi.org/10.1038/s41467-023-38867-x>
- Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, Hosmillo M, Rees-Spear C, Mlcochova P, Lumb IU, et al. 2022. Author correction: SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 608:E23. <https://doi.org/10.1038/s41586-022-05104-2>
- Ko KKK, Yingtaweesittikul H, Tan TT, Wijaya L, Cao DY, Goh SS, Abdul Rahman NB, Chan KXL, Tay HM, Sim JHC, Chan KS, Oon LLE, Nagarajan N, Suphavitai C. 2022. Emergence of SARS-CoV-2 spike mutations during prolonged infection in immunocompromised hosts. *Microbiol Spectr* 10:e0079122. <https://doi.org/10.1128/spectrum.00791-22>
- Weigang S, Fuchs J, Zimmer G, Schnepf D, Kern L, Beer J, Luxenburger H, Ankerhold J, Falcone V, Kemming J, Hofmann M, Thimme R, Neumann-Haefelin C, Ulferts S, Grosse R, Hornuss D, Tanriver Y, Rieg S, Wagner D, Huzly D, Schwemmler M, Panning M, Kochs G. 2021. Within-host evolution of SARS-CoV-2 in an immunosuppressed COVID-19 patient as a source of immune escape variants. *Nat Commun* 12:6405. <https://doi.org/10.1038/s41467-021-26602-3>
- Chen L, Zody MC, Di Germanio C, Martinelli R, Mediavilla JR, Cunningham MH, Composto K, Chow KF, Kordalewska M, Corvelo A, Oschwald DM, Fennessey S, Zetkovic M, Dar S, Kramer Y, Mathema B, Germer S, Stone M, Simmons G, Busch MP, Maniatis T, Perlin DS, Kreiswirth BN, Foy PD. 2021. Emergence of multiple SARS-CoV-2 antibody escape variants in an immunocompromised host undergoing convalescent plasma treatment. *mSphere* 6:e0048021. <https://doi.org/10.1128/mSphere.00480-21>
- Jensen B, Luebke N, Feldt T, Keitel V, Brandenburger T, Kindgen-Milles D, Lutterbeck M, Freise NF, Schoeler D, Haas R, Dilthey A, Adams O, Walker A, Timm J, Luedde T. 2021. Emergence of the E484K mutation in SARS-CoV-2-infected immunocompromised patients treated with bamlanivimab in Germany. *Lancet Reg Health Eur* 8:100164. <https://doi.org/10.1016/j.lanepe.2021.100164>
- Choi B, Choudhary MC, Regan J, Sparks JA, Padera RF, Qiu X, Solomon IH, Kuo H-H, Boucau J, Bowman K, et al. 2020. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N Engl J Med* 383:2291–2293. <https://doi.org/10.1056/NEJMc2031364>
- Hensley MK, Bain WG, Jacobs J, Nambulli S, Parikh U, Cillo A, Staines B, Heaps A, Sobolewski MD, Rennick LJ, et al. 2021. Intractable coronavirus disease 2019 (COVID-19) and prolonged severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) replication in a chimeric antigen receptor-modified T-cell therapy recipient: a case study. *Clin Infect Dis* 73:e815–e821. <https://doi.org/10.1093/cid/ciab072>
- Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, Hosmillo M, Rees-Spear C, Mlcochova P, Lumb IU, et al. 2021. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 592:277–282. <https://doi.org/10.1038/s41586-021-03291-y>
- Truong TT, Ryutov A, Pandey U, Yee R, Goldberg L, Bhojwani D, Aguayo-Hiraldo P, Pinsky BA, Pekosz A, Shen L, et al. 2021. Increased viral variants in children and young adults with impaired humoral immunity and persistent SARS-CoV-2 infection: a consecutive case series. *EBioMedicine* 67:103355. <https://doi.org/10.1016/j.ebiom.2021.103355>
- Leung WF, Chorlton S, Tyson J, Al-Rawahi GN, Jassem AN, Prystajek N, Masud S, Deans GD, Chapman MG, Mirzanejad Y, Murray MCM, Wong PHP. 2022. COVID-19 in an immunocompromised host: persistent shedding of viable SARS-CoV-2 and emergence of multiple mutations: a case report. *Int J Infect Dis* 114:178–182. <https://doi.org/10.1016/j.ijid.2021.10.045>
- Harari S, Tahor M, Rutsinsky N, Meijer S, Miller D, Henig O, Halutz O, Levytskyi K, Ben-Ami R, Adler A, Paran Y, Stern A. 2022. Drivers of adaptive evolution during chronic SARS-CoV-2 infections. *Nat Med* 28:1501–1508. <https://doi.org/10.1038/s41591-022-01882-4>
- Khatamzas E, Antwerpen MH, Rehn A, Graf A, Hellmuth JC, Hollaus A, Mohr A-W, Gaitzsch E, Weiglein T, Georgi E, et al. 2022. Accumulation of mutations in antibody and CD8 T cell epitopes in a B cell depleted lymphoma patient with chronic SARS-CoV-2 infection. *Nat Commun* 13:5586. <https://doi.org/10.1038/s41467-022-32772-5>
- Sonnleitner ST, Prelog M, Sonnleitner S, Hinterbichler E, Halbfurter H, Kopecky DBC, Almanzar G, Koblmüller S, Sturmbauer C, Feist L, Horres R, Posch W, Walder G. 2022. Cumulative SARS-CoV-2 mutations and corresponding changes in immunity in an immunocompromised patient indicate viral evolution within the host. *Nat Commun* 13:2560. <https://doi.org/10.1038/s41467-022-30163-4>
- Gandhi S, Klein J, Robertson AJ, Peña-Hernández MA, Lin MJ, Roychoudhury P, Lu P, Fournier J, Ferguson D, Mohamed Bakhsh SAK, Catherine Muenker M, Srivathsan A, Wunder EA, Kerantzias N, Wang W, Lindenbach B, Pyle A, Wilen CB, Ogbuagu O, Greninger AL, Iwasaki A, Schulz WL, Ko AI. 2022. *De novo* emergence of a remdesivir resistance mutation during

- treatment of persistent SARS-CoV-2 infection in an immunocompromised patient: a case report. *Nat Commun* 13:1547. <https://doi.org/10.1038/s41467-022-29104-y>
22. Avanzato VA, Matson MJ, Seifert SN, Pryce R, Williamson BN, Anzick SL, Barbian K, Judson SD, Fischer ER, Martens C, Bowden TA, de Wit E, Riedo FX, Munster VJ. 2020. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. *Cell* 183:1901–1912. <https://doi.org/10.1016/j.cell.2020.10.049>
 23. Tarhini H, Recoing A, Bridier-Nahmias A, Rahi M, Lambert C, Martres P, Lucet J-C, Rioux C, Bouzid D, Lebourgeois S, Descamps D, Yazdanpanah Y, Le Hingrat Q, Lescure F-X, Visseaux B. 2021. Long-term severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infectiousness among three immunocompromised patients: from prolonged viral shedding to SARS-CoV-2 superinfection. *J Infect Dis* 223:1522–1527. <https://doi.org/10.1093/infdis/jiab075>
 24. Monrad I, Sahlertz SR, Nielsen SSF, Pedersen LØ, Petersen MS, Kobel CM, Tarpgaard IH, Storgaard M, Mortensen KL, Schleimann MH, Tolstrup M, Vibholm LK. 2021. Persistent severe acute respiratory syndrome coronavirus 2 infection in immunocompromised host displaying treatment induced viral evolution. *Open Forum Infect Dis* 8:ofab295. <https://doi.org/10.1093/ofid/ofab295>
 25. Cele S, Karim F, Lustig G, San JE, Hermanus T, Tegally H, Snyman J, Moyo-Gwete T, Wilkinson E, Bernstein M, et al. 2022. SARS-CoV-2 prolonged infection during advanced HIV disease evolves extensive immune escape. *Cell Host Microbe* 30:154–162. <https://doi.org/10.1016/j.chom.2022.01.005>
 26. Maponga TG, Jeffries M, Tegally H, Sutherland A, Wilkinson E, Lessells RJ, Msomi N, van Zyl G, de Oliveira T, Preiser W. 2023. Persistent severe acute respiratory syndrome coronavirus 2 infection with accumulation of mutations in a patient with poorly controlled human immunodeficiency virus infection. *Clin Infect Dis* 76:e522–e525. <https://doi.org/10.1093/cid/ciac548>
 27. Lee CY, Shah MK, Hoyos D, Solovyov A, Douglas M, Taur Y, Maslak P, Babady NE, Greenbaum B, Kamboj M, Vardhana SA. 2022. Prolonged SARS-CoV-2 infection in patients with lymphoid malignancies. *Cancer Discov* 12:62–73. <https://doi.org/10.1158/2159-8290.CD-21-1033>
 28. Hueso T, Pouderoux C, Péré H, Beaumont A-L, Raillon L-A, Ader F, Chatenoud L, Eshagh D, Szwebel T-A, Martinot M, et al. 2020. Convalescent plasma therapy for B-cell-depleted patients with protracted COVID-19. *Blood* 136:2290–2295. <https://doi.org/10.1182/blood.202008423>
 29. Malin JJ, Di Cristanziano V, Horn C, Pracht E, Garcia Borrega J, Heger E, Knops E, Kaiser R, Böll B, Lehmann C, Jung N, Borchmann P, Fätkenheuer G, Klein F, Hallek M, Rybniker J. 2022. SARS-CoV-2-neutralizing antibody treatment in patients with COVID-19 and immunodeficiency due to B-cell non-Hodgkin lymphoma. *Blood Adv* 6:1580–1584. <https://doi.org/10.1182/bloodadvances.2021006655>
 30. Gaitzsch E, Passerini V, Khatamzas E, Strobl CD, Muenchhoff M, Scherer C, Osterman A, Heide M, Reischer A, Subklewe M, et al. 2021. COVID-19 in patients receiving CD20-depleting immunochemotherapy for B-cell lymphoma. *Hemasphere* 5:e603. <https://doi.org/10.1097/HS9.0000000000000603>
 31. Terpos E, Gavriatopoulou M, Fotiou D, Giatra C, Asimakopoulos I, Dimou M, Sklirou AD, Ntanasis-Stathopoulos I, Darmani I, Briasoulis A, Kastritis E, Angelopoulos M, Baltadakis I, Panayiotidis P, Trougakos IP, Vassilakopoulos TP, Pagoni M, Dimopoulos MA. 2021. Poor neutralizing antibody responses in 132 patients with CLL, NHL and HL after vaccination against SARS-CoV-2: a prospective study. *Cancers (Basel)* 13:4480. <https://doi.org/10.3390/cancers13174480>
 32. Herishanu Y, Avivi I, Aharon A, Shefer G, Levi S, Bronstein Y, Morales M, Ziv T, Shorer Arbel Y, Scarfò L, Joffe E, Perry C, Ghia P. 2021. Efficacy of the BNT162b2 mRNA COVID-19 vaccine in patients with chronic lymphocytic leukemia. *Blood* 137:3165–3173. <https://doi.org/10.1182/blood.2021011568>
 33. Wilkinson SAJ, Richter A, Casey A, Osman H, Mirza JD, Stockton J, Quick J, Ratcliffe L, Sparks N, Cumley N, Poplawski R, Nicholls SN, Kele B, Harris K, Peacock TP, Loman NJ. 2022. Recurrent SARS-CoV-2 mutations in immunodeficient patients. *Virus Evol* 8:veac050. <https://doi.org/10.1093/ve/veac050>
 34. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34:4121–4123. <https://doi.org/10.1093/bioinformatics/bty407>
 35. Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 30:1188–1195. <https://doi.org/10.1093/molbev/mst024>
 36. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
 37. Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol* 65:997–1008. <https://doi.org/10.1093/sysbio/syw037>
 38. Chaguza C, Hahn AM, Petrone ME, Zhou S, Ferguson D, Breban MI, Pham K, Peña-Hernández MA, Castaldi C, Hill V, Schulz W, Swanstrom RI, Roberts SC, Grubaugh ND, Yale SARS-CoV-2 Genomic Surveillance Initiative. 2023. Accelerated SARS-CoV-2 Intra-host evolution leading to distinct genotypes during chronic infection. *Cell Rep Med* 4:100943. <https://doi.org/10.1016/j.xcrm.2023.100943>
 39. Braun KM, Moreno GK, Wagner C, Accola MA, Rehauer WM, Baker DA, Koelle K, O'Connor DH, Bedford T, Friedrich TC, Moncla LH. 2021. Acute SARS-CoV-2 infections harbor limited within-host diversity and transmit via tight transmission bottlenecks. *PLoS Pathog* 17:e1009849. <https://doi.org/10.1371/journal.ppat.1009849>
 40. Korneeva N, Khalil MI, Ghosh I, Fan R, Arnold T, De Benedetti A. 2023. SARS-CoV-2 viral protein Nsp2 stimulates translation under normal and hypoxic conditions. *Virology* 20:55. <https://doi.org/10.1186/s12985-023-02021-2>
 41. Hsu J-C, Laurent-Rolle M, Pawlak JB, Wilen CB, Cresswell P. 2021. Translational shutdown and evasion of the innate immune response by SARS-CoV-2 Nsp14 protein. *Proc Natl Acad Sci U S A* 118:e2101161118. <https://doi.org/10.1073/pnas.2101161118>
 42. Abbasian MH, Mahmanzar M, Rahimian K, Mahdavi B, Tokhanbigli S, Moradi B, Sisakht MM, Deng Y. 2023. Global landscape of SARS-CoV-2 mutations and conserved regions. *J Transl Med* 21:152. <https://doi.org/10.1186/s12967-023-03996-w>
 43. Hillen HS, Kovic G, Farnung L, Dienemann C, Tegenov D, Cramer P. 2020. Structure of replicating SARS-CoV-2 polymerase. *Nature* 584:154–156. <https://doi.org/10.1038/s41586-020-2368-8>
 44. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, Peacock SJ, Robertson DL, COVID-19 Genomics UK (COG-UK) Consortium. 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 19:409–424. <https://doi.org/10.1038/s41579-021-00573-0>
 45. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, Tan AL, Paul LM, Brackney DE, Grewal S, Gurfield N, Van Rompay KKA, Isern S, Michael SF, Coffey LL, Loman NJ, Andersen KG. 2019. An amplicon-based sequencing framework for accurately measuring Intra-host virus diversity using PrimalSeq and iVar. *Genome Biol* 20:8. <https://doi.org/10.1186/s13059-018-1618-7>
 46. Lythgoe KA, Hall M, Ferretti L, de Cesare M, MacIntyre-Cockett G, Trebes A, Andersson M, Otecko N, Wise EL, Moore N, et al. 2021. SARS-CoV-2 within-host diversity and transmission. *Science* 372:eabg0821. <https://doi.org/10.1126/science.abg0821>
 47. Hossain A, Akter S, Rashid AA, Khair S, Alam A. 2022. Unique mutations in SARS-CoV-2 Omicron subvariants' non-spike proteins: potential impacts on viral pathogenesis and host immune evasion. *Microb Pathog* 170:105699. <https://doi.org/10.1016/j.micpath.2022.105699>
 48. Shu Y, McCauley J. 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 22:30494. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>
 49. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, Akite N, Ho J, Lee RT, Yeo W, Curation Team GC, Maurer-Stroh S. 2021. GISAID's role in pandemic response. *China CDC Wkly* 3:1049–1051. <https://doi.org/10.46234/ccdcw2021.255>
 50. Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 1:33–46. <https://doi.org/10.1002/gch2.1018>
 51. Nakamura K, Sugiyama M, Ishizuka H, Sasajima T, Minakawa Y, Sato H, Miyazawa M, Kitakawa K, Fujita S, Saito N, Kashiwabara N, Kohata H, Hara Y, Kanari Y, Shinka T, Kanemitsu K. 2023. Prolonged infective SARS-CoV-2 Omicron variant shedding in a patient with diffuse large B cell

- lymphoma successfully cleared after three courses of remdesivir. *J Infect Chemother* 29:820–824. <https://doi.org/10.1016/j.jiac.2023.05.003>
52. Szemiel AM, Merits A, Orton RJ, MacLean OA, Pinto RM, Wickenhagen A, Lieber G, Turnbull ML, Wang S, Furnon W, Suarez NM, Mair D, da Silva Filipe A, Willett BJ, Wilson SJ, Patel AH, Thomson EC, Palmarini M, Kohl A, Stewart ME. 2021. *In vitro* selection of remdesivir resistance suggests evolutionary predictability of SARS-CoV-2. *PLoS Pathog* 17:e1009929. <https://doi.org/10.1371/journal.ppat.1009929>
53. Tchesnokov EP, Gordon CJ, Woolner E, Kocinkova D, Perry JK, Feng JY, Porter DP, Götte M. 2020. Template-dependent inhibition of coronavirus RNA-dependent RNA polymerase by remdesivir reveals a second mechanism of action. *J Biol Chem* 295:16156–16165. <https://doi.org/10.1074/jbc.AC120.015720>
54. Ordaya EE, Vergidis P, Razonable RR, Yao JD, Beam E. 2023. Genotypic and predicted phenotypic analysis of SARS-CoV-2 Omicron subvariants in immunocompromised patients with COVID-19 following tixagevimab-cilgavimab prophylaxis. *J Clin Virol* 160:105382. <https://doi.org/10.1016/j.jcv.2023.105382>
55. Focosi D, McConnell S, Sullivan DJ, Casadevall A. 2023. Analysis of SARS-CoV-2 mutations associated with resistance to therapeutic monoclonal antibodies that emerge after treatment. *Drug Resist Updat* 71:100991. <https://doi.org/10.1016/j.drug.2023.100991>
56. Baang JH, Smith C, Mirabelli C, Valesano AL, Manthei DM, Bachman MA, Wobus CE, Adams M, Washer L, Martin ET, Luring AS. 2021. Prolonged severe acute respiratory syndrome coronavirus 2 replication in an immunocompromised patient. *J Infect Dis* 223:23–27. <https://doi.org/10.1093/infdis/jiaa666>
57. Pérez-Lago L, Aldámiz-Echevarría T, García-Martínez R, Pérez-Latorre L, Herranz M, Sola-Campoy PJ, Suárez-González J, Martínez-Laperche C, Comas I, González-Candelas F, Catalán P, Muñoz P, García de Viedma D On Behalf Of Gregorio Marañón Microbiology-Id Covid Study Group2021. Different within-host viral evolution dynamics in severely immunosuppressed cases with persistent SARS-CoV-2. *Biomedicines* 9:808. <https://doi.org/10.3390/biomedicines9070808>
58. Reuken PA, Stallmach A, Pletz MW, Brandt C, Andreas N, Hahnfeld S, Löffler B, Baumgart S, Kamradt T, Bauer M. 2021. Severe clinical relapse in an immunocompromised host with persistent SARS-CoV-2 infection. *Leukemia* 35:920–923. <https://doi.org/10.1038/s41375-021-01175-8>
59. Stanevich OV, Alekseeva EI, Sergeeva M, Fadeev AV, Komissarova KS, Ivanova AA, Simakova TS, Vasilyev KA, Shurygina A-P, Stukova MA, et al. 2023. SARS-CoV-2 escape from cytotoxic T cells during long-term COVID-19. *Nat Commun* 14:149. <https://doi.org/10.1038/s41467-022-34033-x>
60. Riddell AC, Kele B, Harris K, Bible J, Murphy M, Dakshina S, Storey N, Owoyemi D, Pade C, Gibbons JM, Harrington D, Alexander E, McKnight Á, Cutino-Moguel T. 2022. Generation of novel severe acute respiratory syndrome coronavirus 2 variants on the B.1.1.7 lineage in 3 patients with advanced human immunodeficiency virus-1 disease. *Clin Infect Dis* 75:2016–2018. <https://doi.org/10.1093/cid/ciac409>
61. Ciuffreda L, Lorenzo-Salazar JM, Alcoba-Florez J, Rodriguez-Pérez H, Gil-Campesino H, Iñigo-Campos A, García-Martínez de Artola D, Valenzuela-Fernández A, Hayek-Peraza M, Rojo-Alba S, Alvarez-Argüelles ME, Díez-Gil O, González-Montelongo R, Flores C. 2021. Longitudinal study of a SARS-CoV-2 infection in an immunocompromised patient with X-linked agammaglobulinemia. *J Infect* 83:607–635. <https://doi.org/10.1016/j.jinf.2021.07.028>
62. Agostini ML, Andres EL, Sims AC, Graham RL, Sheahan TP, Lu X, Smith EC, Case JB, Feng JY, Jordan R, Ray AS, Cihlar T, Siegel D, Mackman RL, Clarke MO, Baric RS, Denison MR. 2018. Coronavirus susceptibility to the antiviral remdesivir (GS-5734) is mediated by the viral polymerase and the proofreading exoribonuclease. *mBio* 9:e00221-18. <https://doi.org/10.1128/mBio.00221-18>
63. Choudhary MC, Crain CR, Qiu X, Hanage W, Li JZ. 2022. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) sequence characteristics of coronavirus disease 2019 (COVID-19) persistence and reinfection. *Clin Infect Dis* 74:237–245. <https://doi.org/10.1093/cid/ciab380>
64. Valesano AL, Rumpfelt KE, Dimcheff DE, Blair CN, Fitzsimmons WJ, Petrie JG, Martin ET, Luring AS. 2021. Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. *PLoS Pathog* 17:e1009499. <https://doi.org/10.1371/journal.ppat.1009499>
65. Tonkin-Hill G, Martincorena I, Amato R, Lawson ARJ, Gerstung M, Johnston I, Jackson DK, Park N, Lensing SV, Quail MA, et al. 2021. Patterns of within-host genetic diversity in SARS-CoV-2. *Elife* 10:e66857. <https://doi.org/10.7554/eLife.66857>
66. Neher RA. 2022. Contributions of adaptation and purifying selection to SARS-CoV-2 evolution. *Virus Evol* 8:veac113. <https://doi.org/10.1093/ve/veac113>
67. Everett J, Hokama P, Roche AM, Reddy S, Hwang Y, Kessler L, Glascock A, Li Y, Whelan JN, Weiss SR, Sherrill-Mix S, McCormick K, Whiteside SA, Graham-Wooten J, Khatib LA, Fitzgerald AS, Collman RG, Bushman F, Stallings CL. 2021. SARS-Cov-2 genomic variation in space and time in hospitalized patients in Philadelphia. *mBio* 12:e03456-20. <https://doi.org/10.1128/mBio.03456-20>
68. Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528. <https://doi.org/10.1093/bioinformatics/bty633>
69. Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 13:238. <https://doi.org/10.1186/1471-2105-13-238>
70. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
71. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, Yuan M-L, Zhang Y-L, Dai F-H, Liu Y, Wang Q-M, Zheng J-J, Xu L, Holmes EC, Zhang Y-Z. 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579:265–269. <https://doi.org/10.1038/s41586-020-2008-3>