

SARS-CoV-2 escape from cytotoxic T cells during long-term COVID-19

Oksana Stanevich

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia; First Pavlov State Medical University, Saint-Petersburg, Russia

Evgeniia Alekseeva (✉ evg.alekseeva93@gmail.com)

Skolkovo Institute of Science and Technology

Maria Sergeeva

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia

Artem Fadeev

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia <https://orcid.org/0000-0003-3558-3261>

Kseniya Komissarova

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia <https://orcid.org/0000-0002-1465-5548>

Anna Ivanova

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia

Tamara Simakova

Parseq Lab Co. Ltd, Saint Petersburg, Russia

Kirill Vasilyev

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia

Anna-Polina Shurygina

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia

Marina Stukova

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia

Ksenia Safina

Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia

Elena Nabieva

Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia

Sofya Garushyants

A.A. Kharkevich Institute for Information Transmission Problems of the Russian Academy of Sciences, Moscow, Russia

Galya Klink

Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences

Evgeny Bakin

Bioinformatics Institute, Saint Petersburg, Russia

Jullia Zabutova

City Hospital 31, Saint-Petersburg, Russia

Anastasia Kholodnaia

First Pavlov State Medical University, Saint-Petersburg, Russia; City Hospital 31, Saint-Petersburg, Russia

Olga Lukina

First Pavlov State Medical University, Saint-Petersburg, Russia

Irina Skorokhod

City Hospital 31, Saint-Petersburg, Russia

Viktoria Ryabchikova

City Hospital 31, Saint-Petersburg, Russia

Nadezhda Medvedeva

City Hospital 31, Saint-Petersburg, Russia

Dmitry Lioznov

"Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia; First Pavlov State Medical University, Saint-Petersburg, Russia "

Daria Danilenko

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia <https://orcid.org/0000-0001-6174-0836>

Dmitriy Chudakov

Pirogov Russian National Research Medical University <https://orcid.org/0000-0003-0430-790X>

Andrey Komissarov

Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia

Georgii Bazykin



Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute) <https://orcid.org/0000-0003-2334-2751>

Biological Sciences - Article

Keywords: Immunocompromised Hosts, Mode of Selection, Intro-host Evolution, Cellular Immunity, Virus Immunoediting, Cytotoxic CD8 T Cell Clones

Posted Date: July 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-750741/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

SARS-CoV-2 escape from cytotoxic T cells during long-term COVID-19

Oksana V. Stanevich^{1,2†}, Evgeniia I. Alekseeva^{3**}, Maria Sergeeva¹, Artem V. Fadeev¹, Kseniya S. Komissarova¹, Anna A. Ivanova¹, Tamara S. Simakova⁴, Kirill A. Vasilyev¹, Anna-Polina Shurygina¹, Marina A. Stukova¹, Ksenia R. Safina³, Elena R. Nabieva³, Sofya K. Garushyants⁵, Galya V. Klink⁵, Evgeny A. Bakin^{2,6}, Jullia V. Zabutova⁷, Anastasia N. Kholodnaia^{2,7}, Olga V. Lukina², Irina A. Skorokhod⁷, Viktoria V. Ryabchikova⁷, Nadezhda V. Medvedeva⁷, Dmitry A. Lioznov^{1,2}, Daria M. Danilenko¹, Dmitriy M. Chudakov^{3,8}, Andrey B. Komissarov¹, Georgii A. Bazykin^{3,5*}

1. Smorodintsev Research Institute of Influenza, Saint Petersburg, Russia;
2. First Pavlov State Medical University, Saint-Petersburg, Russia;
3. Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia;
4. Parseq Lab Co. Ltd, Saint Petersburg, Russia;
5. A.A. Kharkevich Institute for Information Transmission Problems of the Russian Academy of Sciences, Moscow, Russia;
6. Bioinformatics Institute, Saint Petersburg, Russia;
7. City Hospital 31, Saint-Petersburg, Russia;
8. Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russia.

* e-mail: evg.alekseeva93@gmail.com, g.bazykin@skoltech.ru

[†]Equal contribution

Abstract

Evolution of SARS-CoV-2 in immunocompromised hosts may result in novel variants with changed properties, but the mode of selection underlying this process remains unclear¹. While escape from humoral immunity certainly plays a role in intra-host evolution²⁻⁴, escape from cellular immunity is poorly understood⁵⁻⁷. Here, we report a case of long-term COVID-19 in an immunocompromised patient with non-Hodgkin's lymphoma who received treatment with rituximab and lacked neutralizing antibodies. Over the 318 days of the disease, the SARS-CoV-2 genome gained a total of 40 changes, 34 of which were present by the end of the study period. Among the acquired mutations, 12 reduced or prevented binding of known immunogenic SARS-CoV-2 HLA class I antigens, suggesting that virus immunoediting is largely driven by cytotoxic CD8 T cell clones. The two changes with the strongest effect, nsp3:T504A and nsp3:T504P, were experimentally assessed in a cytotoxic assay of the patient's CD8 T cells. Both these changes were associated with immune escape, with a stronger effect observed for nsp3:T504P, the change which ultimately got fixed. Together, these results suggest that CD8 T cell escape may be an underappreciated contributor to SARS-CoV-2 evolution in humans.

40 Main

41 SARS-CoV-2 evolution in the global human population has involved accumulation of mutations
42 that increase viral transmissibility and cause immune escape¹. A similar set of mutations is also
43 observed in intra-host evolution of SARS-CoV-2 during long-term COVID-19, particularly in
44 immunocompromised patients^{2-4,8-15} treated with monoclonal antibodies and/or convalescent
45 plasma^{13,16}. Both the mutations that spread rapidly in the general population and those that
46 accumulate in intra-host evolution facilitate entry of viral particles into host cells^{17,18} and/or
47 affect binding sites of neutralising antibodies^{2-4,19,20}.

48 In addition to escape from humoral immunity, SARS-CoV-2 evolution can also involve escape
49 from cellular response. Indeed, the landscape of antigen presentation, which is determined by the
50 particular set of patient's HLA alleles, has a significant impact on the course of COVID-19²¹⁻²³.
51 Consistently, in the general population, SARS-CoV-2 acquires changes that reduce binding of
52 viral antigens to HLA I molecules, weakening antigen recognition by corresponding cytotoxic T
53 clones^{5,6}. However, at the population level, accumulation of T cell escape mutations is
54 complicated by the diversity of HLA molecules, and the changes that lie at the origin of
55 SARS-CoV-2 variants of concern are insignificant for CD4 and CD8 T cell reactivity in most
56 patients⁷. Long-term COVID-19 may facilitate T cell escape, as the selection pressure favoring
57 such an escape remains constant throughout the disease. Indeed, intra-host escape from T cells of
58 both types was described for other long-term infections including HIV-1 and hepatitis C²⁴⁻²⁸.
59 Here, we report a case of long-term evolution of SARS-CoV-2 in an immunocompromised
60 patient involving escape from T cell mediated immunity.

61 Case description

62 Patient S, a female previously diagnosed with Non-Hodgkin's diffuse B-cell lymphoma IV stage
63 B, tested positive for SARS-CoV-2 for the first time on April 17, 2020. In the preceding week,
64 she had had close contact with patient A, who later died of COVID-19 pneumonia; paraffin
65 blocks with post-mortem material were subsequently analyzed for SARS-CoV-2 by PCR,
66 followed by RNA extraction and sequencing, as a probable source of infection. Patient S has
67 undergone three periods of positive tests, alternating with two periods of negative tests, between
68 April 17, 2020 and March 1, 2021, spanning a total of 318 days (see Fig. 1a, Extended Data
69 Table 1). She had symptoms of severe COVID-19 between June 6 - September 1, 2020
70 (Extended Data Fig. 1a,b), and again between January 9 - March 1, 2021 (Extended Data Fig.
71 1c), including subfebrile fever and pneumonia with typical COVID-19 patterns. We isolated live
72 virus from swab samples obtained in both of these periods (August 20, 2020 and February 19,
73 2021).

74 Between April 30, 2020 and February 16, 2021, patient S received several cycles of
75 chemotherapy under several different regimens, including monoclonal antibody rituximab. On
76 December 28, 2020, autologous haematopoietic stem cell transplantation (auto-HSCT) was
77 performed. In January 2021, near the end of the study period, patient S received three doses of
78 convalescent plasma. Six nasopharyngeal swab samples suitable for next generation sequencing,
79 together spanning 308 days of the disease, were obtained, alongside two blood samples
80 (Extended Data Table 1).

81 Intrahost evolution of SARS-CoV-2

82 Whole-genome sequencing was performed on six nasopharyngeal swab samples obtained from
83 patient S in August 2020 - February 2021, as well as in an April 2020 sample obtained from
84 patient A (Fig. 1a). Phylogenetic analysis (Supplementary Note 1) indicates that both PCR
85 positive periods of patient S in August 2020 and January-February 2021 constitute a single
86 infection. Indeed, all patient S samples form a single clade within the B.1.1 lineage on the global
87 SARS-CoV-2 phylogeny, with the patient A sample as its ancestor (Fig. 1b). No other Russian
88 samples available in GISAID nest within the patient S clade (Fig. 1b), indicating that the virus
89 evolved in patient S has not seeded observable onward transmission.

90 The two August 2020 samples were characterised respectively by 12 and 18 mutations specific to
91 patient S. In turn, the January-February 2021 samples gained additional 10 to 21 changes.
92 Overall, a total of 40 changes compared to the ancestral state were observed in at least one of the
93 samples, 34 of which were observed by the end of the study period (Supplementary Note 2). This
94 corresponds to the point substitution rate of 15.3×10^{-4} per nucleotide per year, which
95 substantially exceeds the evolutionary rate of SARS-CoV-2 in the general population
96 (permutation test, $p < 10^{-4}$; Fig. 1c). Nearly all accumulated changes were detected in samples
97 obtained before convalescent plasma transfusions (Fig. 1a,d; Extended Data Table 1), indicating
98 that these transfusions could not have affected the observed viral evolution.

99 The accumulated mutations were distributed throughout the viral genome, affecting 18 of the 26
100 viral genes (Fig. 1d). However, there was an excess of nonsynonymous changes in the genes
101 encoding surface proteins: out of the 25 changes, 8 (41%) fell in the spike gene which by length
102 constitutes 13% of the viral genome, while 2 (9%) fell in the envelope gene which constitutes
103 0.8% of the genome (two-sided Binomial test, $p = 0.018$ and 0.016 , respectively). Many of the
104 observed amino acid substitutions were indicative of positive selection in the general population
105 (Supplementary Note 3), and some were previously implicated in antibody escape
106 (Supplementary Note 3). However, virus evolution did not lead to detectable reduction in
107 sensitivity to neutralizing antibodies by the end of the study period compared to a prototype viral
108 strain (Extended Data Fig. 3).

109

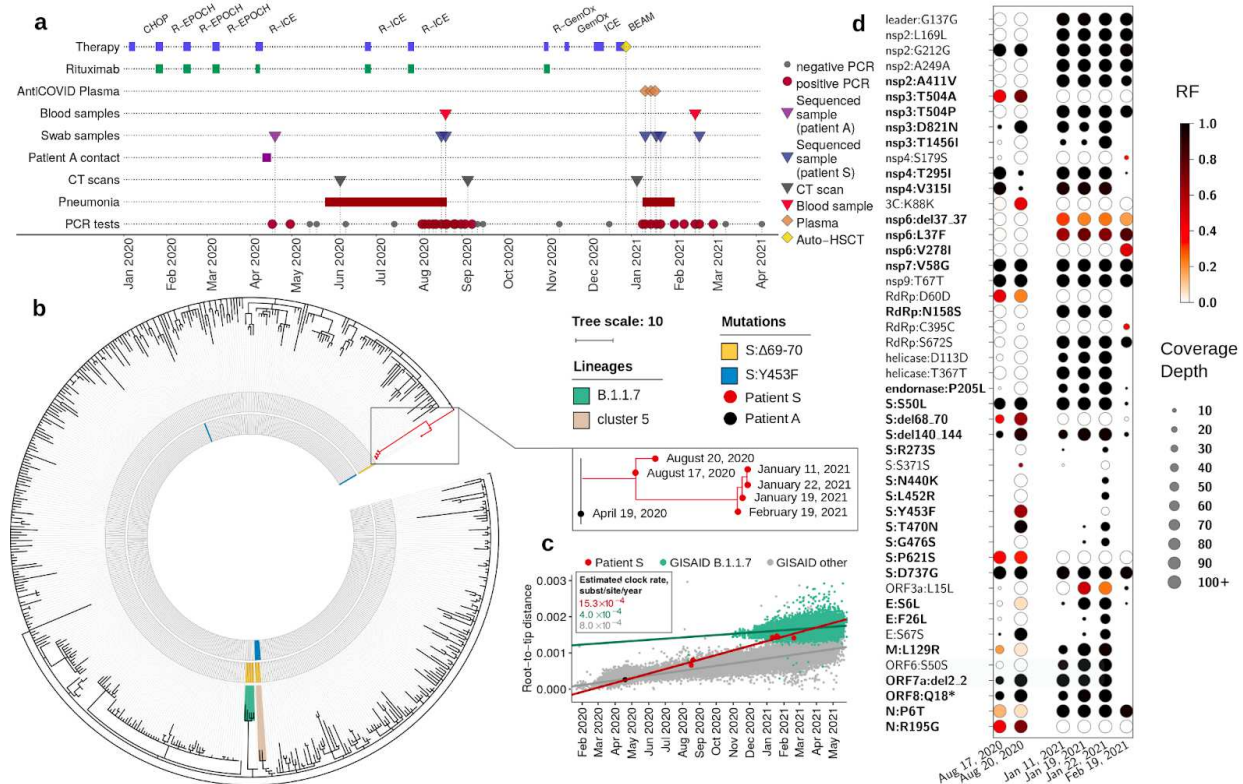


Fig. 1. Intrahost evolution of SARS-CoV-2 in patient S. **a:** The timeline of patient S disease and therapy. **b:** The phylogenetic tree of B.1.1 pruned to contain a random set of 1% of all samples, including the patient A sample (black dot) and the complete clade carrying the patient S samples (red dots). The 2020 samples carried the ΔF combination of mutations (S:Δ69-70HV and S:Y435F; Supplementary Note 3) marked in the two inner circles in yellow and blue. The B.1.1.7 lineage and cluster 5 are shaded. **c:** Regression of root-to-tip genetic distances vs. sampling dates, for patient S samples (together with the ancestral sample of Patient A), B.1.1.7 lineage GISAID samples, and other GISAID samples. Estimated slopes (molecular clock rates) are provided in the inset. In **b** and **c**, the consensus nucleotides (i.e., those supported by more than 50% of the reads, RF>50%) were used to position patient S and A samples. **d:** Variant frequencies in the six patient S samples. All consensus variants (RF>50%, N=40) and nonconsensus variants with 30%<RF<50% (N=7) are shown.

Host immune response

To understand the functional features of immune response in patient S, we analysed her blood samples collected at multiple timepoints spanning the course of the disease. Flow cytometry revealed the absence of B lymphocytes throughout the period of PCR positivity (Extended Data Fig. 4a). Blood serum samples were also analyzed by ELISA for IgG antibodies specific to the SARS-CoV-2 S-antigen; a weak IgG response was registered in one of the samples but no response in the remaining samples. No neutralizing antibodies were detected at any time point by a VN assay using live SARS-CoV-2 strain (Extended Data Table 1).

131 By contrast, a pronounced SARS-CoV-2 specific T-cell response was detected. Indeed, *in vitro*
132 stimulation with an overlapping peptide pool (OPP) caused the expansion of
133 SARS-CoV-2-specific CD4 and CD8 effector memory T-cells (Tem) at both time points
134 (Extended Data Fig. 4b,c).

135 Mutational escape from cytotoxic T cells

136 Given the lack of B-cell but the presence of T-cell immune response in patient S, we
137 hypothesized that the 31 amino acid sequence-altering mutations acquired by SARS-CoV-2 may
138 have led to escape from T cell immunity. First, we asked if these mutations affect presentation of
139 the peptides carrying them by the HLA alleles of patient S (Extended Data Table 3). For this, we
140 adapted an existing pipeline²⁹ to calculate the PHBR (patient harmonic best rank) score (Fig. 2a)
141 for both the ancestral and the derived state at site of each of the 30 mutations (except
142 ORF8:Q18*, Supplementary Note 4). Most sites could be presented in their ancestral state by at
143 least one HLA allele of both classes (27 out of 30 by HLA I, and 24 out of 30 by HLA II). We
144 found that five of the observed mutations substantially (>3-fold) increased the PHBR score for
145 the peptides presented by HLA I, indicating impaired presentation (Fig. 2b). One of these
146 mutations, S:del141-144, also increased the PHBR score for HLA II (Fig. 2c).

147 While an increase in PHBR score can help a peptide escape antigen presentation, this can only
148 affect T cell response if the corresponding peptide is recognised by T cells. To specifically
149 address the effect of mutations on immunogenic peptides, we used IEDB³⁰ to obtain the list of
150 SARS-CoV-2 peptides that were shown to be immunogenic in complexes with the HLA alleles
151 carried by patient S. There were 17 such peptides for HLA I alleles, together overlapping the
152 sites of 11 of the mutations (some of the sites were covered by more than one peptide) (Extended
153 Data Table 4). All these mutations were fixed in the course of intra-host evolution by the end of
154 the study period. No HLA class II immunogenic peptides covering the changed sites were found
155 in IEDB. To focus on the immunogenic peptides, we calculated the imBR (immunogenic best
156 rank) for each of these sites in the ancestral state and compared it to the corresponding value for
157 the derived state. The mutations strongly decreased presentation of immunogenic peptides,
158 indicating that they cause escape from CD8 T cell response (Fig. 2b,d). Together with
159 ORF8:Q18* which prevented presentation of the bulk of the ORF8 protein (Supplementary Note
160 4), this totals to 12 changes with cytotoxic T cell escape effect.

161

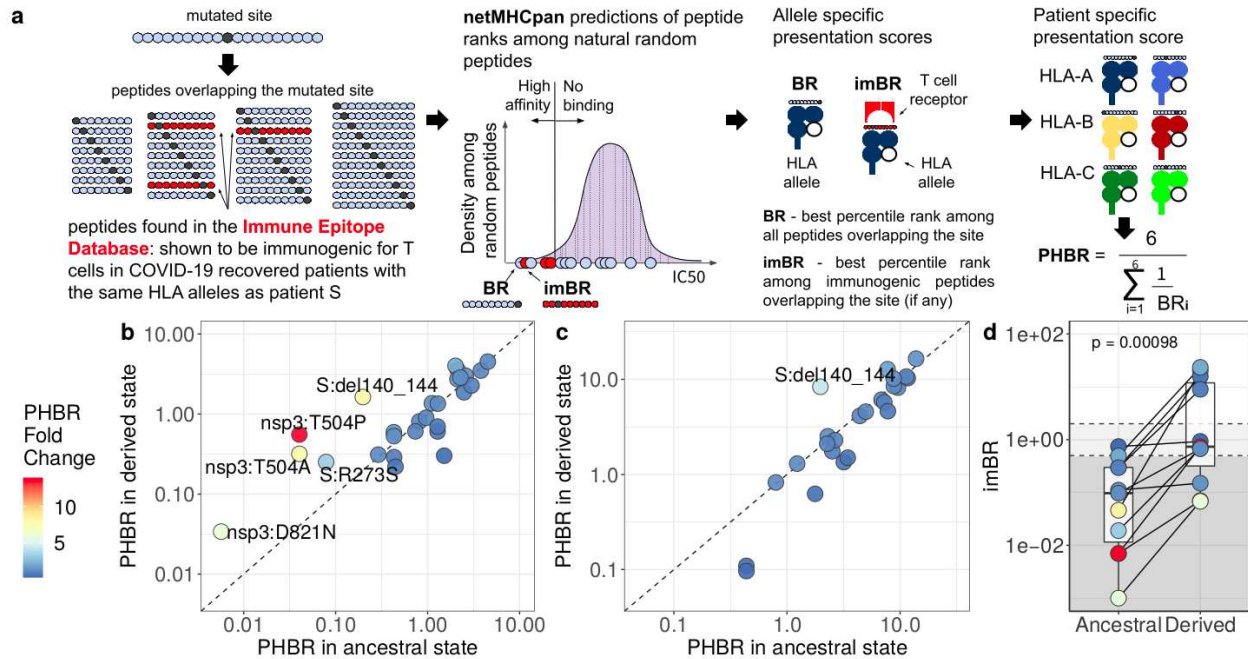


Fig. 2. Mutational escape from cytotoxic T cells. **a:** Calculation of site presentation scores (adapted from Marty et al.²⁹). **b, c:** Change of PHBR scores caused by mutations for HLA I (**b**) and HLA II (**c**) respectively. Dot color corresponds to PHBR fold change; the mutations that substantially (>3-fold) increase PHBR are signed. Sites that did not bind any of the patient's HLA alleles both in ancestral and derived states are not shown. **d:** Comparison of imBR scores for the mutated sites in their ancestral and derived states. The level of significance is calculated by the Wilcoxon sign-rank test.

nsp3 mutations affect T-cell reactivity

Next, we assessed the change in T-cell response caused by the observed mutations. We focused on the two mutations causing the largest PHBR fold change (Fig. 2b). These were the two mutations at position 504 of the nsp3 protein, nsp3:T504A and nsp3:T504P, which were fixed sequentially at the first (T1, August 20, 2020) and the second (T2, February 19, 2021) sampled time points respectively (Fig. 1d). We asked how well the peptides covering these three amino acid variants elicited T-cell response in samples corresponding to these time points. We used the highest ranking peptides covering the mutated site in its ancestral (PTDNYITTY) and derived (PADNYITTY, PPDNYITTY) states; PTDNYITTY was previously shown to be immunogenic in complex with the HLA-A:01*01 allele which is carried by patient S³¹⁻³³.

In the T1 sample, when just nsp3:T504A was detected at intermediate frequencies (Fig. 1d), *in vitro* stimulation of CD8⁺ T cells indicated response to both the ancestral (PTDNYITTY) and the derived (PADNYITTY) peptide changed by nsp3:T504A (Fig. 3). This response was mediated primarily by polyfunctional IFN γ ⁺IL2⁺TNF α ⁺ effector memory T-cells. The response to PADNYITTY was slightly weaker, suggesting a partial escape caused by nsp3:T504A. Stimulation by PPDNYITTY corresponding to the nsp3:T504P allele caused no cytokine response in the T1 sample. In the T2 sample (Fig. 1a), when nsp3:T504P was already fixed, still

no cytokine response to PPDNYITTY was observed, confirming invisibility of this peptide to cellular immune response due to weak binding with HLA. Response to PTDNYITTY and PADNYITTY also vanished at T2; this could indicate that the CD8 T cell clones specific to T and A amino acids became irrelevant with the loss of the corresponding viral variants, and got no antigenic re-stimulation that could drive clonal expansion after auto-HSCT³⁴.

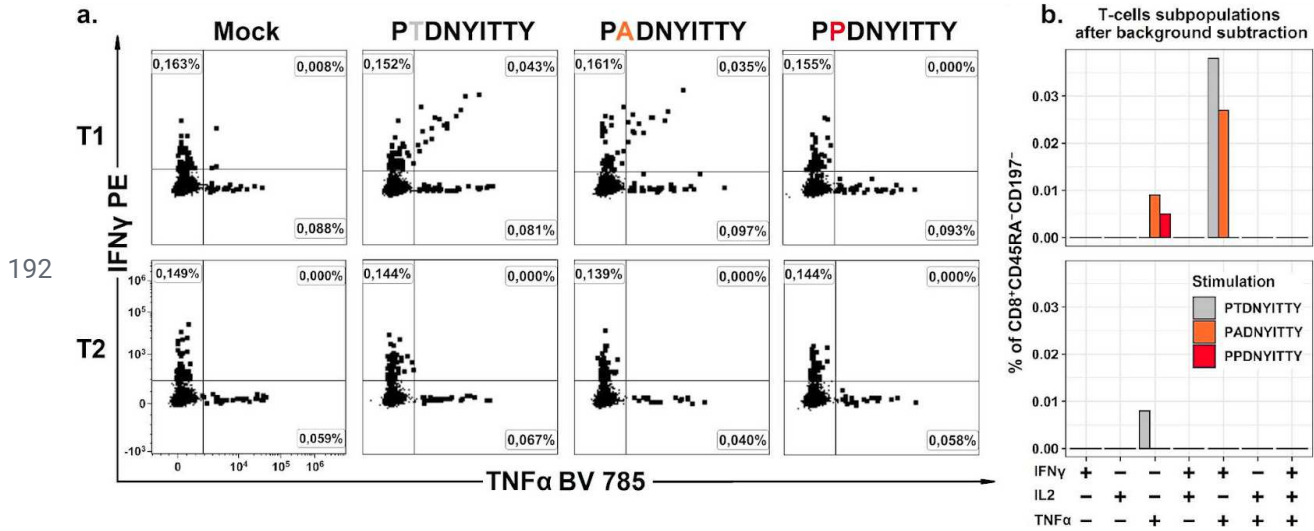


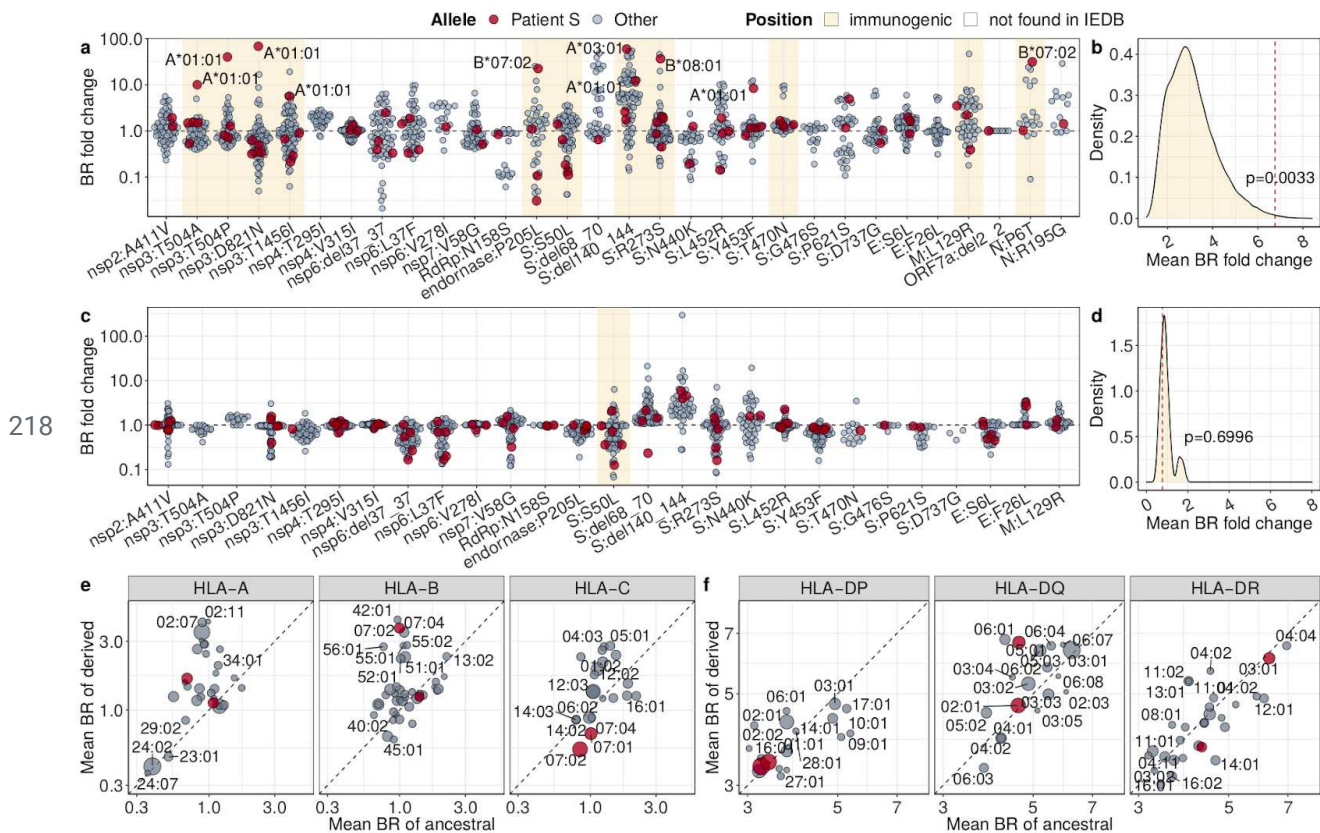
Fig. 3. The CD8 T-cellular immune response to the SARS-CoV-2 epitope carrying the ancestral and the two derived (nsp3:T504A and nsp3:T504P) amino acid variants. a: Representative flow cytometry plots showing the cytokine profiles of SARS-CoV-2-specific CD8 effector memory T cells after stimulation. **b:** Bar plots representing the percentage of different cytokine-producing populations of SARS-CoV-2-specific CD8 T cells after background subtraction (data from the mock-stimulated sample were subtracted from peptide-stimulated samples). T1, August 2020 sample; T2, February 2021 sample.

Possible population-level effects

It has been suggested that escape from humoral immunity in immunosuppressed patients may give rise to SARS-CoV-2 strains with increased fitness in the general population¹. Similarly, escape from cellular immunity in the course of intra-host evolution could affect immune response to descendant SARS-CoV-2 strains outside the host where it evolved. We aimed to estimate the possible effect of the viral evolution in patient S for the human population at large. For this, we compared the BR (Fig. 2a) fold change caused by the mutations observed in patient S for the globally most frequent HLA alleles of each family that together cover 95% of worldwide population frequency^{35,36}. This set of alleles includes all 12 HLA alleles of both classes (I and II) of patient S, which happen to be quite frequent globally (Extended Data Table 3).

As expected, the mutations observed in immunogenic epitopes tended to escape the HLA I alleles of patient S to a larger extent than other frequent HLA I alleles (Fig. 4a,b); no such difference was observed for HLA II alleles (Fig. 4c,d). Nevertheless, these same mutations also reduced binding for other globally frequent HLA I alleles (mean BR fold change = 1.59, Fig. 4e),

215 although not HLA II alleles (mean fold change = 1.02, Fig. 4f). This indicates that the
 216 within-host evolution in patient S indeed could facilitate escape from cytotoxic T cells in the
 217 global population.



219 **Fig. 4 Population-level effect of T cell escape mutations.** **a, c:** The effect of each of the 30
 220 mutations observed in SARS-CoV-2 of patient S on T cell immune escape, for each of the HLA I
 221 **(a)** or HLA II **(c)** alleles carried by patient A (red) and frequent globally (grey). The mutations
 222 that change immunogenic peptides (for HLA I) or are adjacent to such peptides (for HLA II)
 223 according to IEDB are highlighted. Alleles that do not present the corresponding position in both
 224 ancestral and derived state are not shown. For the mutations that correspond to >5-fold increase
 225 in BR, the corresponding HLA alleles are signed. **b, d:** Distribution of mean BPR fold changes
 226 among immunogenic positions for HLA I **(b)** or II **(d)** alleles, based on 10^5 random generations
 227 of individual allele composition; the red dashed line is the percentile corresponding to the allele
 228 composition of patient S. **e, f:** The sum effect of the amino acid changing mutations observed in
 229 SARS-CoV-2 of patient S on antigen presentation by the globally most frequent HLA class I **(e)**
 230 and class II **(f)**. Alleles of patient S are in red.

231 Discussion

232 We have described a case of unprecedentedly long COVID-19 characterized by a large amount
 233 of intrahost evolution. For over 10 months, an evolving SARS-CoV-2 lineage accumulated
 234 changes at a rate which substantially exceeded that in the general population, suggesting
 235 prevalent viral adaptation. Some of the observed changes recapitulated mutations previously

observed in other immunocompromised patients and/or variants of concern (Extended Data Fig. 7, Supplementary Note 3). This is consistent with the hypothesis that immunocompromised patients represent a hotspot of viral adaptation, causing “saltations” in the otherwise clock-like evolutionary rate of SARS-CoV-2¹; notably, such a jump could have happened at the origin of the B.1.1.7 (“alpha”) variant which has attained global dominance in early 2021^{1,37}.

Unlike previously described cases, however, the case described here is characterized by an unusual immune environment. The absence of own B cells, convalescent plasma therapy or monoclonal antibodies therapy during most of the study period indicates that the bulk of viral mutations have accumulated in the absence of humoral immune response. Instead, our data shows that evolution was largely driven by T cell escape. Our computational analysis revealed that many mutations changed the amino acid composition of known immunogenic CD8 T cell antigens and worsened or prevented their presentations on HLA class I alleles of the patient. We experimentally validated the escape effect of two of these mutations.

These results indicate that immunoediting by cytotoxic CD8 clones is an underappreciated factor in intrahost evolution of SARS-CoV-2. Similar to antibody escape, the T cell escape mutations acquired within an individual host may give rise to new epidemiologically important variants if they spill over to the general population. We predict that the changes observed in this study would also substantially affect SARS-CoV-2 antigenicity in the general population in case of onward transmission of the evolved variant. While no such transmission was detected in this case, our results emphasize an additional dimension of SARS-CoV-2 evolution which merits careful surveillance.

References

- 1 Harvey WT, Carabelli AM, Jackson B, *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 2021; **19**: 409–24.
- 2 Kemp S, Collier D, Datir R, *et al.* Neutralising antibodies in Spike mediated SARS-CoV-2 adaptation. *Infectious Diseases (except HIV/AIDS)*, 2020 DOI:10.1101/2020.12.05.20241927.
- 3 Williamson MK, Hamilton F, Hutchings S, *et al.* Chronic SARS-CoV-2 infection and viral evolution in a hypogammaglobulinaemic individual. *Infectious Diseases (except HIV/AIDS)*, 2021 DOI:10.1101/2021.05.31.21257591.
- 4 Khatamzas E, Rehn A, Muenchhoff M, *et al.* Emergence of multiple SARS-CoV-2 mutations in an immunocompromised host. *Infectious Diseases (except HIV/AIDS)*, 2021 DOI:10.1101/2021.01.10.20248871.
- 5 Agerer B, Koblishke M, Gudipati V, *et al.* SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8⁺ T cell responses. *Sci Immunol* 2021; **6**: eabg6461.
- 6 Dolton G, Rius C, Hasan MS, *et al.* Emergence of immune escape at dominant SARS-CoV-2 killer T-cell epitope. *Infectious Diseases (except HIV/AIDS)*, 2021 DOI:10.1101/2021.06.21.21259010.
- 7 Tarke A, Sidney J, Methot N, *et al.* Negligible impact of SARS-CoV-2 variants on CD4⁺ and CD8⁺ T cell reactivity in COVID-19 exposed donors and vaccinees. *Immunology*, 2021 DOI:10.1101/2021.02.27.433180.
- 8 Avanzato VA, Matson MJ, Seifert SN, *et al.* Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell* 2020;

279 **183**: 1901-1912.e9.

280 9 Choi B, Choudhary MC, Regan J, *et al.* Persistence and Evolution of SARS-CoV-2 in an
 281 Immunocompromised Host. *N Engl J Med* 2020; **383**: 2291–3.

282 10 Sepulcri C, Dentone C, Mikulska M, *et al.* The longest persistence of viable SARS-CoV-2
 283 with recurrence of viremia and relapsing symptomatic COVID-19 in an immunocompromised
 284 patient – a case study. *Infectious Diseases (except HIV/AIDS)*, 2021
 285 DOI:10.1101/2021.01.23.21249554.

286 11 Nakajima Y, Ogai A, Furukawa K, *et al.* Prolonged viral shedding of SARS-CoV-2 in an
 287 immunocompromised patient. *J Infect Chemother* 2021; **27**: 387–9.

288 12 Moore JL, Ganapathiraju PV, Kurtz CP, Wainscoat B. A 63-Year-Old Woman with a History
 289 of Non-Hodgkin Lymphoma with Persistent SARS-CoV-2 Infection Who Was Seronegative
 290 and Treated with Convalescent Plasma. *Am J Case Rep* 2020; **21**.
 291 DOI:10.12659/AJCR.927812.

292 13 Hueso T, Pouderoux C, Péré H, *et al.* Convalescent plasma therapy for B-cell-depleted
 293 patients with protracted COVID-19. *Blood* 2020; **136**: 2290–5.

294 14 Wei L, Liu B, Zhao Y, Chen Z. Prolonged shedding of SARS-CoV-2 in an elderly liver
 295 transplant patient infected by COVID-19: a case report. *Ann Palliat Med* 2020; **9**: 8–8.

296 15 Karataş A, İnkaya AÇ, Demiroğlu H, *et al.* Prolonged viral shedding in a lymphoma patient
 297 with COVID-19 infection receiving convalescent plasma. *Transfus Apher Sci* 2020; **59**:
 298 102871.

299 16 Betraíns A, Godinas L, Woei-A-Jin FJSH, *et al.* Convalescent plasma treatment of persistent
 300 severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infection in patients with
 301 lymphoma with impaired humoral immunity and lack of neutralising antibodies. *Br J*
 302 *Haematol* 2021; **192**: 1100–5.

303 17 Chen J, Wang R, Wang M, Wei G-W. Mutations Strengthened SARS-CoV-2 Infectivity. *J Mol*
 304 *Biol* 2020; **432**: 5212–26.

305 18 Peacock TP, Goldhill DH, Zhou J, *et al.* The furin cleavage site of SARS-CoV-2 spike protein
 306 is a key determinant for transmission due to enhanced replication in airway cells.
 307 *Microbiology*, 2020 DOI:10.1101/2020.09.30.318311.

308 19 Starr TN, Greaney AJ, Addetia A, *et al.* Prospective mapping of viral mutations that escape
 309 antibodies used to treat COVID-19. *Science* 2021; **371**: 850–4.

310 20 Garrett ME, Galloway J, Chu HY, *et al.* High resolution profiling of pathways of escape for
 311 SARS-CoV-2 spike-binding antibodies. *Cell* 2021; : S0092867421005808.

312 21 Shkurnikov M, Nersisyan S, Jankevic T, *et al.* Association of HLA Class I Genotypes With
 313 Severity of Coronavirus Disease-19. *Front Immunol* 2021; **12**: 641900.

314 22 Pretti MAM, Galvani RG, Vieira GF, Bonomo A, Bonamino MH, Boroni M. Class I HLA
 315 Allele Predicted Restricted Antigenic Coverages for Spike and Nucleocapsid Proteins Are
 316 Associated With Deaths Related to COVID-19. *Front Immunol* 2020; **11**: 565730.

317 23 La Porta CAM, Zapperi S. Estimating the Binding of Sars-CoV-2 Peptides to HLA Class I in
 318 Human Subpopulations Using Artificial Neural Networks. *Cell Syst* 2020; **11**: 412-417.e2.

319 24 Bronke C, Almeida C-AM, McKinnon E, *et al.* HIV escape mutations occur preferentially at
 320 HLA-binding sites of CD8 T-cell epitopes. *AIDS* 2013; **27**: 899–905.

321 25 Troyer RM, McNevin J, Liu Y, *et al.* Variable Fitness Impact of HIV-1 Escape Mutations to
 322 Cytotoxic T Lymphocyte (CTL) Response. *PLoS Pathog* 2009; **5**: e1000365.

323 26 Goulder PJR, Brander C, Tang Y, *et al.* Evolution and transmission of stable CTL escape
 324 mutations in HIV infection. *Nature* 2001; **412**: 334–8.

325 27 Erdmann N, Du VY, Carlson J, *et al.* HLA Class-II Associated HIV Polymorphisms Predict
 326 Escape from CD4+ T Cell Responses. *PLOS Pathog* 2015; **11**: e1005111.
 327 28 Erickson AL, Kimura Y, Igarashi S, *et al.* The Outcome of Hepatitis C Virus Infection Is
 328 Predicted by Escape Mutations in Epitopes Targeted by Cytotoxic T Lymphocytes. *Immunity*
 329 2001; **15**: 883–95.
 330 29 Marty R, Kaabinejadian S, Rossell D, *et al.* MHC-I Genotype Restricts the Oncogenic
 331 Mutational Landscape. *Cell* 2017; **171**: 1272-1283.e15.
 332 30 Vita R, Mahajan S, Overton JA, *et al.* The Immune Epitope Database (IEDB): 2018 update.
 333 *Nucleic Acids Res* 2019; **47**: D339–43.
 334 31 Ferretti AP, Kula T, Wang Y, *et al.* Unbiased Screens Show CD8+ T Cells of COVID-19
 335 Patients Recognize Shared Epitopes in SARS-CoV-2 that Largely Reside outside the Spike
 336 Protein. *Immunity* 2020; **53**: 1095-1107.e3.
 337 32 Gangaev A, Ketelaars SLC, Isaeva OI, *et al.* Identification and characterization of a
 338 SARS-CoV-2 specific CD8+ T cell response with immunodominant features. *Nat Commun*
 339 2021; **12**: 2593.
 340 33 Saini SK, Hersby DS, Tamhane T, *et al.* SARS-CoV-2 genome-wide T cell epitope mapping
 341 reveals immunodominance and substantial CD8⁺ T cell activation in COVID-19 patients. *Sci*
 342 *Immunol* 2021; **6**: eabf7550.
 343 34 Mamedov IZ, Britanova OV, Bolotin DA, *et al.* Quantitative tracking of T cell clones after
 344 haematopoietic stem cell transplantation. *EMBO Mol Med* 2011; **3**: 201–7.
 345 35 Solberg OD, Mack SJ, Lancaster AK, *et al.* Balancing selection and heterogeneity across the
 346 classical human leukocyte antigen loci: A meta-analytic review of 497 population studies.
 347 *Hum Immunol* 2008; **69**: 443–64.
 348 36 Sarkizova S, Klaeger S, Le PM, *et al.* A large peptidome dataset improves HLA class I
 349 epitope prediction across most of the human population. *Nat Biotechnol* 2020; **38**: 199–209.
 350 37 Peacock TP, Penrice-Randal R, Hiscox JA, Barclay WS. SARS-CoV-2 one year on: evidence
 351 for ongoing viral adaptation. *J Gen Virol* 2021; **102**. DOI:10.1099/jgv.0.001584.
 352

353

354

355 Methods

356 Sample collection and sequencing

357 Special informed consent was obtained from the patient before the specimen for additional tests
358 were taken. RT-PCR of swabs and sequencing of viral RNA was performed in the Smorodintsev
359 Influenza Research Institute. All specimens were obtained and transported according to standard
360 sampling protocol. RNA from nasopharyngeal swabs was extracted using QIAamp Viral RNA
361 Mini Kit (QIAGEN). RNA from patient A post-mortem FFPE specimens was extracted using
362 RNeasy FFPE Kit (QIAGEN). Samples were tested for SARS-CoV-2 viral RNA by real-time
363 RT-PCR on thermal cycler CFX96 (BioRad) using "Intifica SARS-CoV-2" Kit (Alkor Bio).
364 Whole-genome amplification of SARS-CoV-2 virus genome for samples from August 2020 and
365 from January 2021 was performed using BioMaster RT-PCR Premium kit (Biolabmix) and
366 primers from ARTIC Network protocol version 3³⁸ and ARTIC Network protocol version 1³⁹
367 with modifications, respectively. Nextera XT (Illumina) kit was used for library preparation in
368 August 2020 and DNA Prep (Illumina) kit was used for library preparation in January 2021, and
369 the libraries were sequenced using the MiSeq platform (Illumina) with version 3 600-cycle
370 chemistry.

371 The DNA of patient S was extracted from peripheral blood using QIAamp Blood DNA Mini kit.
372 DNA sample was prepared and captured with the SureSelect Human All Exon kit v7 (Agilent),
373 and whole exome was sequenced using MGISEQ-2000 at Pirogov Russian National Research
374 Medical University (Moscow, Russia).

375 Flow cytometry assays

376 Flow cytometry assays were performed using cryopreserved PBMCs. Cells were isolated from
377 patients' heparinized blood by gradient centrifugation with lymphocyte separation medium
378 Lymphosep (BioWest), frozen in freezing medium containing 10% DMSO (AppliChem) in FBS
379 (Gibco) and stored in liquid nitrogen until usage.

380 For B-cells analysis presented in Extended Data Fig. 4a, PBMCs samples were towed in a 37°C
381 water bath and stained with fluorescently-labeled antibodies to surface markers CD19-APC/Fire
382 750 (Clone: SJ25C1, Biolegend), BV421-CD20 (Clone: 2H7, Biolegend), CD3-BV605 (Clone:
383 OKT3, Biolegend). PBMCs from a healthy volunteer were used as a control. B-cells were
384 identified as a live CD3-/CD19+/CD20+ population.

385 The T-cell response was assessed by intracellular cytokine staining. For further analysis, cells
386 were towed in a 37°C water bath and stimulated for 6 hours with 5 µg/ml of the overlapping
387 peptide pool mixture containing the entire Nucleocapsid (N) (GenScript) and RBD region of
388 SARS-CoV2 Spike (S) protein (Miltenyi Biotec) (for Extended Data Fig. 4b,c) or one of the
389 peptides PTDNYITTY, PADNYITTY or PPDNYITTY (for Fig. 3) in the RPMI medium
390 (Gibco), containing 10% of FBS (Gibco), 1% of penicillin-streptomycin solution (Gibco),
391 Brefeldin A (BD) and costimulatory CD28/CD49 reagent (BD). Negative control samples were
392 stimulated with the complete medium; for positive control, PMA/ionomycin (Sigma)

combination was used. Surface markers were stained with fluorescent antibody panel containing CD3-APC/Fire (Clone: SK7, Biolegend), CD4-AF647 (Clone: SK3, Biolegend), CD8a-AF600 (Clone: HIT8a, Biolegend), CD45RA-PE/Dazzle (Clone: HI100, Biolegend), CD197-BV421 (Clone: G043H7, Biolegend). Intracellular cytokines were stained using IL-2-FITC (Clone: MQ1-17H12, Biolegend), IFN γ -PE (Clone: 45.15, Beckman Coulter), TNF α -BV785 (Clone: MAb11, Biolegend) antibodies. Cells were permeabilized with BD Cytotfix/Cytoperm™ Fixation/Permeabilization Solution Kit (BD) according to the manufacturer's instructions. Data were collected on a CytoFlex flow cytometer (Beckman Coulter). The results were analyzed using the Kaluza Analysis v2.1 program (Beckman Coulter). Interleukin (IL) 2, interferon γ (IFN γ) and tumor necrosis factor (TNF α) response was measured in effector memory T cells (Tem). To identify Tem, lymphocytes were gated based on their size and granularity. Live CD3⁺T cells were identified and subdivided into CD4⁺ and CD8⁺ T cells. These populations were further subdivided based on the expression of CD45RA and CD197(CCR7). CD3⁺CD4⁺ or CD3⁺CD8⁺ lymphocytes with the CD45RA⁻/CCR7⁻ phenotype were considered Tem cells. Cut-off values for the definition of cytokine-producing T cell responses stimulated with SARS-CoV-2 peptides were ≥ 5 events and a ≥ 2 -fold difference in the magnitude of TNF⁺, IFN γ ⁺ or IL-2⁺ Tem cells compared to the non stimulated control.

Virus isolation and antigenicity

Live viruses (samples 30579V and 30769V from August 20, 2020 and 22748V and 23680V from February 19, 2020) were isolated from patient S swab samples in Vero E6 cells (IZSLER #BSCL87). Culture was inoculated for 2 hours with swab material diluted 1/10 in DMEM (Biolot) supplemented with 2% HI-FBS (Gibco), 1% anti-anti (Gibco) and then incubated for 3 days until first CPE signs. Samples were subsequently passaged one time in Vero cells (ATCC #CCL81).

A total of 16 serum samples were obtained during the first wave of the COVID-19 pandemic in spring-summer 2020 from recovered volunteers with PCR-confirmed SARS-CoV-2 infection and tested in a microneutralization assay.

Microneutralization was performed with hCoV-19/St_Petersburg-3524V/2020 virus (GISAID EPI_ISL_415710, with the ΔF combination of mutations absent, designated as Reference), and 30769V and 23680V viruses isolated from the patient S (designated patient S August 2020 and patient S January 2021, respectively). Serum was heat inactivated (56°C, 60 min), serially diluted 2-fold starting from 1/10, mixed with 25 TCID₅₀ of virus, incubated for 1h at 37°C and inoculated into Vero cells in triplicates in 96-well plate. 5 days after inoculation, neutralizing antibody titer was calculated as the reciprocal of the highest serum dilution preventing CPE.

Serum samples obtained from patient S were tested for virus specific antibodies in ELISA and in microneutralization assay with either Reference or patient S viruses. ELISA was performed with "SARS-CoV-2-IgG-IFA-BEST" (VEKTOR BEST #D-5501) according to manufacturer's instructions.

431 HLA genotyping

432 HLA typing was performed using a commercial kit according to the manufacturer's instructions
433 (PARAllele™ HLA solution v3, Parseq Lab). HLA-A, -B, -C, -DRB1 and -DQB1 loci were
434 genotyped with 3-field resolution. Simultaneously, HLA calling was performed from WES data
435 using HLA-HD version 1.3.0⁴⁰ with IPD-IMGT/HLA database Release Version 3.43. The
436 inferred alleles are listed in Extended Data Table 3.

437 Using HLA-2-Haplo software tool⁴¹ this set of alleles was split into two haplotypes presented in
438 Extended Data Table 3. A European population database was used in this procedure. An
439 a-posteriori probability of found combination was 97.6%. As one can see, the found haplotypes
440 are among the most common variants in the European population.

441 Consensus calling

442 Raw reads were trimmed with Trimmomatic version 0.39⁴² to remove adapter sequences and
443 low-quality ends. Trimmed reads were mapped onto the Wuhan-Hu-1 (MN908947.3) reference
444 genome with BWA MEM version 0.7.17⁴³. The following reads were then removed from the
445 mapping: reads with abnormal insert length to read ratio (greater than 10 or less than 0.8), reads
446 with insert length greater than 1100, reads with more than 50% soft-clipped bases. Soft-clipped
447 ends were trimmed from the remaining reads, 10 nucleotides were cropped from read ends using
448 custom scripts, and primer sequences were removed with ivar version 1.3⁴⁴. Only reads with at
449 least 30 nucleotides remaining after the procedure were kept. SNV and short indel calling was
450 done with LoFreq version 2.1.5⁴⁵, with SNVs considered consensus if they were covered by at
451 least 4 reads and supported by more than 50% of those reads; indels were considered consensus
452 if they were covered by at least 20 reads with at least 50% of those supporting the variant.
453 Regions that were covered by fewer than 2 reads or that were covered by 2 or 3 reads and called
454 non-reference were masked as NC. Consensus was created by bcftools version 1.9^{46,47} consensus.

455 Phylogenetic analysis

456 255,389 genomes of SARS-CoV-2 were downloaded from GISAID on December 12, 2020,
457 (Supplementary Data ACKN) and aligned with MAFFT v7.453⁴⁸ against the reference genome
458 Wuhan-Hu-1/2019 (NCBI ID: MN908947.3⁴⁹ with --addfragments --keeplength options. 100
459 nucleotides from the beginning and from the end of the alignment were trimmed. After that, we
460 excluded sequences (1) shorter than 29,000 bp, (2) with more than 300 positions of missing data
461 (Ns) and gaps, (3) excluded by Nextstrain, (4) from animals other than minks, or (5)
462 corresponding to resequencing of the same patients, leaving us with 201,948 sequences. Identical
463 sequences were then collapsed within the country and host and annotated by the Pangolin
464 package version 2.1.0⁵⁰. To this dataset, we added the two patient S samples obtained in August,
465 2020 as well as the patient A sample. As sequences of patient S belonged to the B.1.1 lineage,
466 we further only kept sequences annotated as B.1.1, excluding a large clade defined by mutation
467 G25563T (GH clade in GISAID⁵¹ nomenclature). We additionally masked the highly homoplasic
468 site 11083. The final set of 49,083 sequences was used to construct the phylogenetic tree with
469 IQ-Tree v2.1.1⁵² under the GTR substitution model and '-fast' option. Ancestral sequences at the
470 internal tree nodes were reconstructed with TreeTime v. 0.8.0⁵³. Having ensured that the two

471 patient S samples form a clade rooted at the patient A sample and not carrying any samples other
472 than those of patient S, we then separately reconstructed the phylogeny of all six samples of
473 patient S, rooted it with patient A, and manually added the resulting clade to the downsampled
474 B.1.1 tree. For visualisation purposes, the tree was downsampled to contain 1% of samples,
475 including the patient A sample and the complete clade containing all patient S samples.

476

477 To estimate the molecular clock rate of the patient S lineage (Fig. 1c), we downloaded all
478 sequences available in GISAID on May 31, 2021, filtered them as described above, and
479 subsampled the filtered dataset to 50,000 samples preserving all Russian sequences. To this
480 dataset, we added the six patient S samples and the ancestral patient A sample. We then aligned
481 the obtained 50,007 sequences against the reference sequence and reconstructed the phylogeny
482 with Fasttree version 2.1.11⁵⁴. Finally, we computed root-to-tip distances and calculated the slope
483 of the root-to-tip distance vs. sampling dates regression line for the three separate datasets: (1)
484 patient S samples, (2) B.1.1.7 samples, and (3) the remaining samples from the subsampled
485 GISAID dataset. To validate the difference between the estimated clock rates for patient S
486 samples and samples belonging to dataset (3), we subsampled this dataset, picking six random
487 samples collected on the same dates as the patient S samples, and computed the linear regression
488 slope, in each of the 10,000 trials. (For dataset (2), this procedure was impossible because there
489 were no B.1.1.7 samples in August 2020). None of the 10,000 samples resulted in the estimated
490 clock rate above 15.3×10^{-4} , implying the p-value of < 0.0001 .

491 Effect of SARS-CoV-2 mutations on antigen presentation in 492 patient S

493 To study the effect of mutations in SARS-CoV-2 proteins on their antigen presentation, we
494 adapted a pipeline from Marty and et.²⁹ (Fig. 2a). For each mutated site in both its ancestral and
495 derived states, we inferred all possible peptides of certain lengths overlapping it, and calculated
496 their percentile ranks (Rank_El) relative to a set of random natural peptides by netMHCpan
497 version 4.1 and netMHCIIpan version 4.0⁵⁵ for HLA I and HLA II respectively. We used peptide
498 lengths between 8 and 12 amino acids for HLA I alleles, and between 12 and 18 amino acids for
499 HLA II. If the mutated site was not presented by any of the HLA alleles either in the ancestral or
500 derived states, we excluded it from analysis. To exclude non-presenting peptides, we used the
501 percentile rank $< 2\%$ threshold for HLA I, and $< 10\%$ threshold for HLA II, as recommended by
502 the netMHCpan manual. For derived states of deletions, we extended the peptide in the
503 C-direction as necessary to preserve its length. We paired the predicted A and B chains of HLA
504 class II alleles as suggested in the tool allele list: HLA-DQA10101-DQB10501,
505 HLA-DQA10501-DQB10201, HLA-DPA10103-DPB10402, HLA-DPA10103-DPB10401,
506 DRB1_0301, DRB1_0101. We excluded the stop-codon producing mutation ORF8:Q18* from
507 comparisons of ancestral and derived states, since the corresponding values for the derived state
508 were undefined.

509 As in Marty et al.²⁹, we used the best percentile rank (BR) among all possible peptides
510 overlapping the mutated site as the presentation score of this site for the particular HLA allele.
511 To estimate the overall presentation of the site in the patient, we calculated the patient harmonic
512 best rank (PHBR), i.e., the harmonic mean of BRs of HLA alleles of the same class. To compare
513 the effect of a mutation on site presentation, we calculated the fold change of PHBR score as the

ratio of the derived PHBR to the ancestral PHBR (so that fold change > 1 indicates weakening of presentation).

To focus on the peptides shown to be immunogenic to T cells in other SARS-CoV-2 infected patients carrying the same HLA alleles as patient S, we used IEDB³⁰ (Immune Epitope Database and Analysis Resource, accessed on June 1, 2021) with the “positive assay only” filter. For those sites inferred to be contained in immunogenic peptide, we calculated the best percentile rank of immunogenic peptide overlapping the site of mutation (imBR).

Population-level effects of mutations on antigen presentation

To check the effect of detected SARS-CoV-2 mutations on presentation by the HLA alleles other than those of patient S, we calculated the BR scores as explained above for the most frequent classical HLA alleles of each family that together represented 95% of the HLA alleles in the world population. The list and frequencies of such alleles were taken from Sarkisova et al. and Solberg et al.^{35,36}.

For most mutations detected in immunogenic epitopes, at least one of the HLA I alleles of patient S demonstrated extreme values of BR fold change in comparison with other alleles (Fig. 4a). To check the probability of such an observation happening by chance, we performed a permutation test, calculating the probability that a randomly chosen set of alleles has the same or a more extreme value of mean BR fold change across all mutations overlapped by immunogenic peptides as that of alleles of patient S. This was true for 33 out of 100000 permutations, corresponding to $p = 0.0033$ (Fig. 4b). None of the HLA II immunogenic epitopes overlapped any of the mutated sites; the only mutated site adjacent to such an epitope (S:S50L) did not stand out in the permutation test ($p = 0.6996$; Fig. 4c,d).

To compare the effects of mutations between different HLA alleles in Fig. 4e,f, we calculated the mean BR across all changed sites. This analysis again excluded ORF8:Q18*, which nevertheless prevented production of high-affinity epitopes for most alleles.

Data analysis and visualisation

Data analysis was performed in R version 4.0.0⁵⁶, and figures were visualised with ggplot2 package version 3.3.2⁵⁷. SARS-CoV-2 phylogenetic tree was visualized with ITOL version 6⁵⁸.

Data availability

Sequence data is available from the Sequence Read Archive: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA749008/> (SRA: PRJNA749008, Supplementary Data 1). Consensus sequences are available from the GISAID with identifiers, presented in Supplementary Data 1. Code is available at https://github.com/EvgeniiaAleksseeva/patient_S.

547 Ethics declaration

548 The study was approved by the Local Ethics Review Board of the Smorodintsev Research
549 Institute of Influenza and by the Biomedical Ethics Committee of the I.P. Pavlov First Saint
550 Petersburg State Medical University. All necessary patient/participant consent has been obtained
551 and the appropriate institutional forms have been archived.

552 Acknowledgements

553 We are thankful to patient S for participation in this study. This study was supported by RFBR
554 project 20-04-60556 to G.A.B. E.I.A. was supported by RFBR project 20-34-90153. D.M.C. was
555 supported by a grant from the Ministry of Science and Higher Education of the Russian
556 Federation (075-15-2019-1789). D.M.D, A.B.K. and D.A.L. were supported by the
557 Governmental contract “Genetic Analysis of SARS-CoV-2 in Russia in the first wave of
558 COVID-19 pandemic”. We are thankful to Mikhail Shugay, Aleksandr Tashkeev and Vadim
559 Karnauhov for helpful discussions. We thank all of the authors who have contributed genome
560 data on GISAID (see Supplementary Data 1 for the list).

562 Author contributions

563 O.V.S. provided the detailed clinical picture and participated in study design; E.I.A. conceived
564 and performed analysis of mutation escape effects; M.S. cultured the virus and performed
565 neutralization assay; A.V.F., K.S.K. and A.A.I. produced sequencing data; T.S.S. performed HLA
566 genotyping; K.A.V., A.-P.S. and M.A.S. performed T-cells assays; K.R.S., E.R.N., S.K.G.,
567 G.V.K. and G.A.B. performed genome analysis; K.R.S. analysed the evolutionary rate; E.A.B.
568 performed HLA-calling and HLA-haplotyping; J.V.Z. collected samples and communicated with
569 the hospital; A.N.K. described the clinical picture and formalised the patient agreement; O.V.L.
570 described the CT scans; I.A.S. is the patient’s attending doctor; V.V.R. provided detailed
571 diagnostic and treatment information about lymphoma; N.V.M. coordinated screening for
572 SARS-CoV-2 between the oncohematology department and the reference laboratory; D.A.L.
573 provided coordination, supervision and funding acquisition; D.M.D. participated in study design
574 and logistics; A.B.K., D.M.C. and G.A.B. planned the study; O.V.S., E.I.A., K.R.S., E.R.N.,
575 S.K.G., A.B.K., D.M.C. and G.A.B. drafted the manuscript; E.I.A. and G.A.B. wrote the
576 manuscript, with contributions from all authors.

577 Competing Interest Declaration

578 Authors declare no competing interests.

579

580 Supplementary Notes

581 Supplementary Note 1

582 Between 19 and 47 genetic changes distinguish the patient S samples from the
583 Wuhan-Hu-1/2019 reference strain⁴⁹. Seven of these changes, including the three SNPs at
584 adjacent positions 21881-21883, were contained in each of the patient S samples, placing them in
585 the B.1.1 lineage. The lineage of patient S carries the remaining 12 to 40 genetic changes. The
586 patient A sample carries the seven mutations characteristic of B.1.1 but no other mutations,
587 confirming patient A as the likely source of infection for patient S, and indicating that the
588 remaining changes are specific to patient S.

589

Supplementary Note 2

Among the 12 mutations specific to patient S and observed in both August 2020 samples, 10 were single-nucleotide mutations (6 nonsynonymous, 3 synonymous and 1 creating a premature stop codon), and the remaining 2 were in-frame deletions. In the second August 2020 sample, 6 additional changes (4 nonsynonymous, 1 synonymous and 1 in-frame deletion) reached consensus frequencies.

Six of the mutations that reached consensus frequencies in the August 2020 samples reversed back to the ancestral state by January 2021, including the ΔF combination (see Supplementary Note 3). Additionally, the January-February 2021 samples gained 21 new mutations compared to the August 17, 2020 sample. 10 of these mutations (6 nonsynonymous and 4 synonymous) were detected in all winter samples. The other 11 mutations (8 nonsynonymous and 3 synonymous) were each called in a subset of the winter samples; in the remaining samples, the corresponding sites were usually poorly covered. Overall, 34 changes were observed in the January 22, 2021 sample, which is the highest-quality sample among the winter 2021 samples (Fig. 1d, Extended Data Table 2). Together with the six reverted changes, this totals to 40 observed changes.

In addition to changes in the consensus sequence, we observed a number of variants at intermediate frequencies (above 30% in at least one of the samples, but below 50% in all samples and therefore not included in the consensus sequence; Fig. 1d, Extended Data Table 2), indicating within-host polymorphism. Three such variants (1 nonsynonymous and 2 synonymous) were observed in the August samples (all of them were lost in the January-February samples), and three (2 synonymous and 1 frame-disrupting deletion) were observed in the January-February samples (all absent in the August samples).

614 Supplementary Note 3

615 Among the positions that acquired amino acid mutations, ten (nsp2:A411V, nsp3:T1456I,
616 nsp4:V315I, endonase:P205L, S:del68_70, S:del140_144, S:N440K, S:L452R, S:G476S,
617 N:R195G) experienced pervasive positive selection according to the FEL (fixed effects
618 likelihood) model⁵⁹ and/or their frequencies grew in the global viral population according to
619 Jonckheere's trend test (Extended Data Table 2), as reported in ref.⁵¹ (accessed on 31th March
620 2020).

621 Many of the detected mutations are known from other studies. Notably, these include the ΔF
622 combination (S:Y453F + S: Δ 69-70HV), which is observed in the consensus of the August 20,
623 2020 sample; S:Y453F is also found at high read frequency in the other 2020 sample, indicating
624 that ΔF was probably also present at this time point (S:69-70 is too poorly covered at this time
625 point to be called). The ΔF combination was previously described as associated with
626 mink-related clusters. It has arisen in parallel in multiple mink populations; among humans, it
627 was mainly found in cases traceable to minks ("Cluster 5", or B.1.1.298), indicating reverse
628 transmission⁶⁰. Despite the presence of the ΔF combination, patient S cannot be placed into the
629 cluster 5 clade because cluster 5 is separated from B.1.1 by two additional mutations (those at
630 positions 15656 and 25936) which are absent in patient S (Extended Data Fig. 2). Furthermore,
631 the ΔF combination was not fixed in the August 2020 samples of patient S but segregated at an
632 intermediate frequency (Extended Data Table 2). Together, this indicates that the ΔF combination
633 was acquired by patient S independently. It was not observed in any of the 2021 patient S
634 samples, indicating that it had been reversed by that time.

635 The ΔF combination confers the ability to rapidly replicate to high titers and to evade recognition
636 by neutralizing antibodies⁶¹, raising concerns that these mutations may affect vaccine efficiency.
637 Y453F affects the receptor-binding domain (RBD), possibly increasing hACE2 binding^{17,19}. It
638 allows immune escape from monoclonal antibodies and polyclonal sera; in particular, it has led
639 to 57% escape from the REGN10933 monoclonal antibody, a component of FDA-approved
640 Regeneron's REGN-COV2 cocktail for treatment of COVID-19 patients, although it did not
641 allow escape from the full cocktail of two antibodies (REGN10933+REGN10987)¹⁹. It was also
642 shown to escape cellular immunity in HLA-A24-restricted patients⁶².

643 The second mutation of the ΔF combination, S: Δ 69-70HV, was recently shown to occur in a
644 virus from another immunocompromised patient with COVID-19² (Extended Data Fig. 5). In
645 that study, S: Δ 69-70HV has been fixed during convalescent plasma therapy, suggesting antibody
646 selection pressure, which is consistent with decreased virus sensitivity to neutralisation with sera
647 from recovered patients. However, patient S was not treated with convalescent plasma in 2020
648 and had no detectable neutralizing antibody response, suggesting that S: Δ 69-70HV could have
649 been favored by some other factor of selection. In patient S, both the S:Y453F and the
650 S: Δ 69-70HV mutations were polymorphic in 2020 and were lost by 2021, suggesting that this
651 other factor may have been transient. The presence of the ΔF combination in the August 19,
652 2020 sample may underlie reduced sensitivity to neutralizing antibodies for this time point
653 (Extended Data Fig. 3). Reacquired sensitivity to neutralizing antibodies by February 19, 2021 is
654 also consistent with the loss of the ΔF combination by this time (Extended Data Fig. 3).

655 Besides S:Δ69-70HV, patient S has acquired six additional mutations that were also observed in
 656 other immunocompromised patients: S:S50L⁶³, S:N440K⁶⁴, S:Δ141-144^{4,8,9,63,64}, nsp3:T504I⁴,
 657 nsp3:T295I⁶³ and nsp6:L37F⁶⁴ (Extended Data Fig. 5). The most recurrent of these mutations,
 658 S:Δ141-144, was shown to lead to an escape from neutralising antibodies⁶⁵. It falls into the
 659 recurrent deletion region⁶⁵ where frequent deletions are observed, including S:Y144del, the
 660 lineage defining deletion of B.1.1.7⁶⁶ which is speculated to have been founded by a chronically
 661 infected individual¹). Another recurrent mutation, nsp6:L37F, is associated with asymptomatic
 662 course of infection⁶⁷; plausibly, it could have contributed to the ultimate recovery of the two
 663 immunocompromised patients in whom it has been observed (patient S and patient 3 from
 664 Truong et al., 2021).

665 Two of the mutations that emerged in patient S also spread in the general population as part of
 666 variants of concern (VOCs). The first is S:Δ69-70HV, which is a lineage-defining mutation of
 667 B.1.1.7. The second is S:L452R, which is found in several VOCs, including AY.1, AY.2 and
 668 B.1.617.2, as well as in multiple variants of interest. S:L452R was shown to have a pleiotropic
 669 effect, causing an escape both from T cell immunity and from neutralising antibodies^{62,68}. Finally,
 670 ORF8:Q18*, a stop-inducing mutation in ORF8, is reminiscent of the stop-inducing mutation in
 671 a different codon of the ORF8 protein (27th, as opposed to the 18th in this study) which is another
 672 of the lineage-defining mutations of B.1.1.7. The functions of ORF8 and its role in immune
 673 response and disease progression are extensively debated⁶⁹⁻⁷¹.

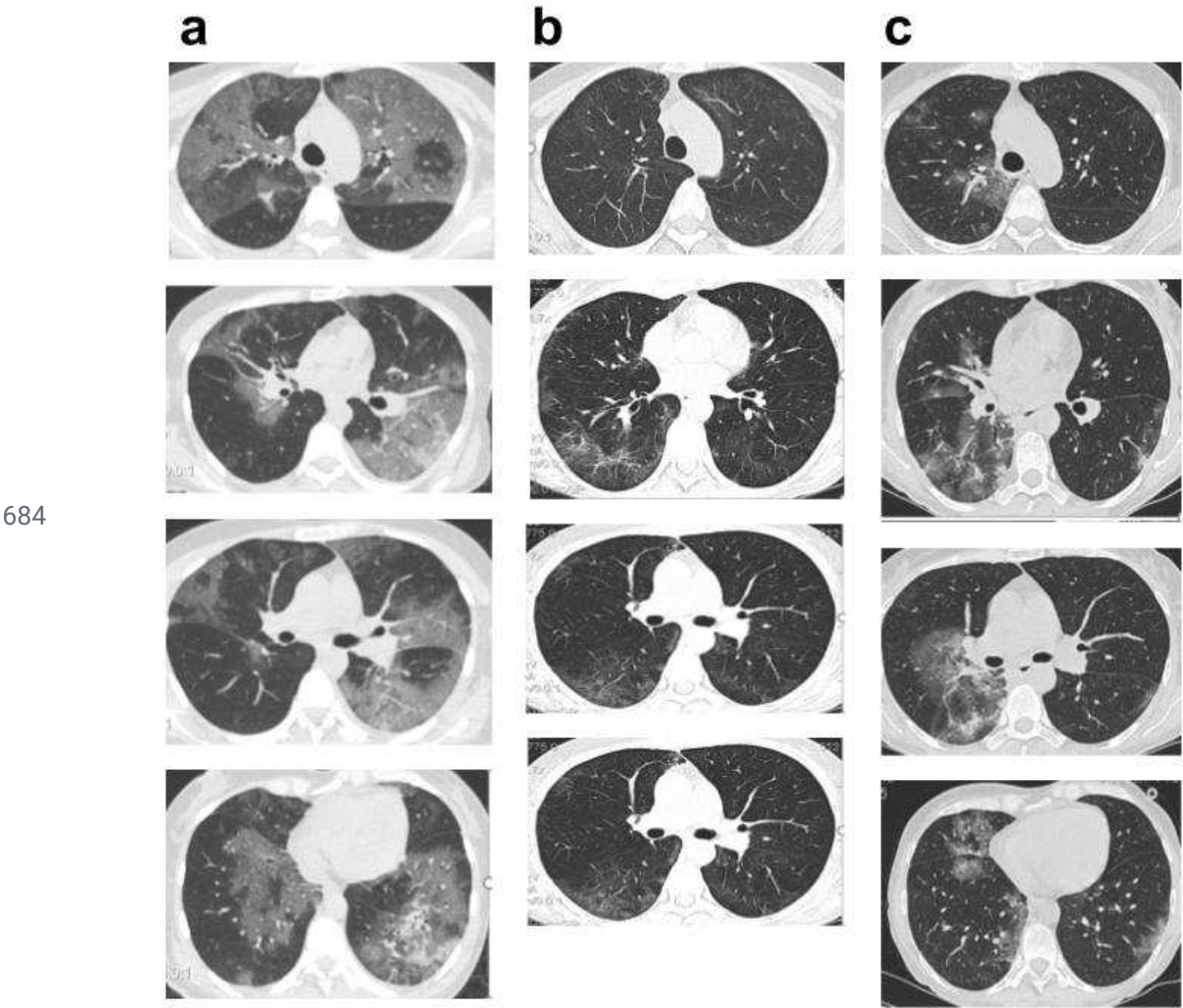
674

675 Supplementary Note 4

676 We excluded ORF8:Q18* from the analysis presented in Fig. 2 because it is impossible to
677 calculate a presentation score (BR, PHBR or imBR) for a site lost from an amino acid sequence
678 due to a stop gain. However, among the peptides absent from the ORF8 sequence after the
679 ORF8:Q18* change, three were listed in IEDB as immunogenic on HLA I alleles carried by
680 patient S, and one, as immunogenic on an HLA II allele carried by patient S. All of them are
681 strong binders (Extended Table 4). Therefore, ORF8:Q18* also causes T cell escape.

682

683 Extended Data



685 **Extended Data Fig. 1** CT scans of patient S lungs obtained by Optima 660 (GE) CT-scanner with
686 standard in-house protocols without contrast enhancement in June 2020 **(a)**, August 2020 **(b)** and January
687 2021 **(c)**. In both lungs in **(a)**, **(b)**, **(c)** there were bilateral, multifocal and diffuse ground-glass
688 opacifications with small regions of subpleural consolidations, but without predominant distribution in
689 **(a)**, mild reticulation and regions of architectural distortion with the formation of subpleural bands in **(b)**,
690 small regions of mild reticulation, vascular dilatation, and regions of linear consolidation with the
691 formation of bands in **(c)**. CT-patterns can be determined as typical for COVID-19 pulmonary disease
692 according to the Radiological Society of North America expert consensus⁷².

693

694

695 **Extended Data Table 1:** The timeline of patient S survey and therapy.

696

Event or procedure	Dates and results	Comment
Contact with patient A	April 10, 2020 - April 16, 2020	Shared a ward in a hospital
SARS-CoV-2 PCR tests	In 2020: April 17 (+), April 30 (+), May 14 (-), May 19 (-), June 9 (-), July 14 (-), August 3 (+), August 5 (+), August 8 (+), August 11 (+), August 13 (+), August 17 (+, Ct 22), August 20 (+, Ct 19), August 21 (+), August 26 (+), August 27 (+), August 31 (+), September 2 (+, Ct 32), September 3 (+), September 8 (+), September 12 (-), September 16 (-), November 10 (-), December 16 (-); In 2021: January 9 (+), January 11 (+), January 15 (+), January 19 (+, Ct = 21), January 22 (+, Ct = 24), January 22 (+, Ct = 31), February 1 (+, Ct = 24), February 8 (+, Ct = 30), February 16 (+, Ct = 19), February 19 (+, Ct = 29), March 1 (+, Ct = 32), March 10 (-), April 5 (-)	(+) - positive PCR; (-) - negative PCR; Ct - real time PCR cycle threshold when known..
Periods of pneumonia	In 2020: June 6 - September 1; In 2021: January 9 - February 1.	
ELISA for anti-S-SARS-CoV-2 IgG	In 2020: August 17, 2020: negative (Cut-off-Index 0.54); November 12, 2020: positive (Cut-off-Index 3.75); December 15: ambiguous (Cut-off-Index 1.03).	Cut-off-index for ELISA: <0,8 – negative; 0,8 – 1,1 – ambiguous; >1 - positive.
VN assay	In 2020: August 17: neutralizing antibodies not detected (titer <10); November 12: neutralizing antibodies not detected (titer <10); December 15: neutralizing antibodies not detected (titer <10).	
Sequenced samples	Patient A: April 19, 2020; Patient S: In 2020: August 17, August 20; In 2021: January 11, January 19, January 22, February 19.	The January 11, 2021 swab was obtained prior to convalescent plasma transfusion on the same date.
Blood samples	August 20, 2020; February 16, 2021	
Chemotherapy	In 2020: CHOP: January 5 - January 9; R-EPOCH: January 24 - January 29; R-EPOCH: February 13 - February 18; R-EPOCH: March 5 - March 10; R-ICE: April 5 - April 8; R-ICE: June 23 - June 27; R-ICE: July 24 - July 28; R-GemOx: October 30; GemOx: October 30; GemOx: November 14; ICE: December 5 - December 12; BEAM: December 21 - December 27.	Abbreviations of chemotherapy regimens are standardized for DBCL ⁷³ ..
Autological transplantation of hematopoietic stem cells (Auto-HSCT)	December 28, 2020	
Convalescent plasma	In 2021: January 11, January 15, January 18.	

697

698

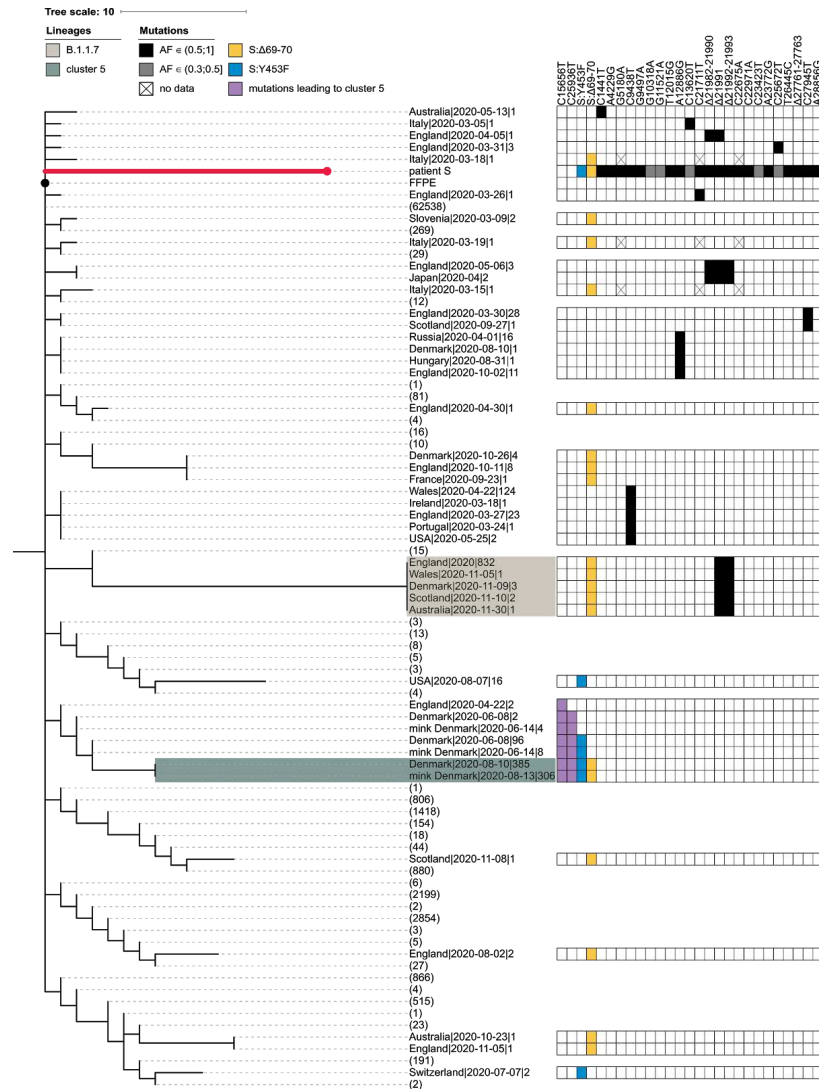
699

700

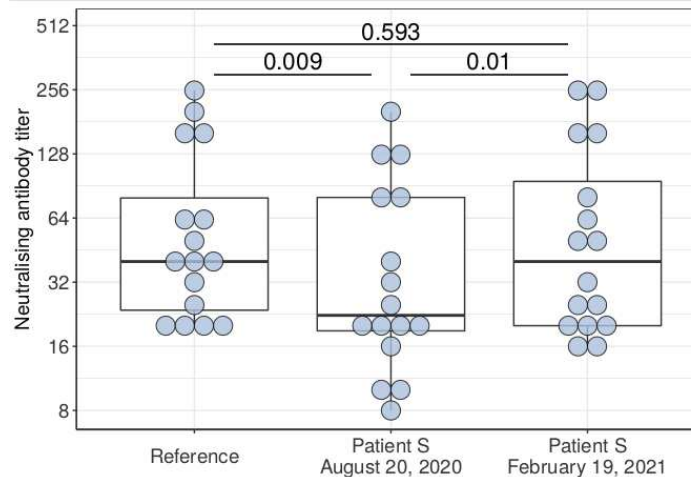
Extended Data Table 2: The list of mutations and their frequencies in sequencing reads obtained from patient S swab samples. Only variants reaching 30% frequency at least in one of the samples are shown. The consensus variants are highlighted in blue, nonsynonymous nucleotide substitutions are in bold. NC (no coverage) indicates coverage depth less than 4 reads. “Selection” and “Trend Z” columns mark positions that experience positive selection and increase of corresponding changes in frequency, according to observablehq.com⁵¹ accessed on 31th March 2020.

Gene	Nucleotide change	AA change	Aug 17 2020	Aug 20 2020	Jan 11 2021	Jan 19 2021	Jan 22 2021	Feb 19 2021	Selection	TrendZ
leader	C:676:T	leader:G137G	0.000	0.000	0.992	0.990	0.998	1.000		
nsp2	G:1312:A C:1441:T T:1552:C C:2037:T	nsp2:L169L nsp2:G212G nsp2:A249A nsp2:A411V	0.000 0.995 0.000 0.000	0.000 0.995 0.000 0.000	0.998 0.996 0.999 0.996	0.998 0.996 0.996 0.996	0.995 0.995 0.998 0.999	0.999 0.992 0.996 1.000	+	
nsp3	A:4229:G A:4229:C G:5180:A C:7086:T	nsp3:T504A nsp3:T504P nsp3:D821N nsp3:T1456I	0.434 0.000 1.000 0.000	0.730 0.000 1.000 0.000	0.000 0.994 1.000 1.000	0.000 1.000 1.000 1.000	0.000 1.000 1.000 1.000	0.000 1.000 NC NC	+	+
nsp4	T:9091:C C:9438:T G:9497:A	nsp4:S179S nsp4:T295I nsp4:V315I	0.000 0.995 0.974	0.000 1.000 1.000	0.000 0.997 0.937	0.000 1.000 0.958	0.000 0.999 0.916	0.333 1.000 NC		+
3C	G:10318:A	3C:K88K	0.014	0.460	0.000	0.000	0.000	0.000		
nsp6	TG:11082:T G:11083:T G:11804:A	nsp6:del37 nsp6:L37F nsp6:V278I	0.000 0.006 0.000	0.005 0.007 0.000	0.302 0.621 0.000	0.221 0.743 0.000	0.238 0.685 0.000	0.175 0.810 0.480		
nsp7	T:12015:G	nsp7:V58G	0.997	0.998	1.000	0.999	0.999	0.999		
nsp9	A:12886:G	nsp9:T67T	1.000	0.993	0.997	0.999	0.998	1.000		
RdRp	C:13620:T A:13913:G C:14625:T A:15456:G	RdRp:D60D RdRp:N158S RdRp:C395C RdRp:S672S	0.453 0.000 0.000 0.000	0.222 0.000 0.000 0.000	0.000 1.000 0.000 1.000	0.000 0.998 0.000 0.997	0.000 0.999 0.000 1.000	NC NC 0.391 1.000		
helicase	C:16575:T A:17337:G	helicase:D113D helicase:T367T	0.000 0.000	0.000 0.000	1.000 1.000	1.000 0.999	1.000 0.996	NC NC		
endornase	C:20234:T	endornase:P205L	0.000	0.000	1.000	1.000	1.000	1.000	+	+
S	C:21711:T ATACATG:21764:A TTTTGGGTGTTTA:21981:T G:22381:T C:22675:A T:22882:G T:22917:G A:22920:T C:22971:A G:22988:A C:23423:T A:23772:G	S:S50L S:del68_70 S:del140_144 S:R273S S:S371S S:N440K S:L452R S:Y453F S:T470N S:G476S S:P621S S:D737G	1.000 0.354 1.000 NC NC NC NC NC NC NC 0.349 0.999	0.998 0.626 0.922 0.000 0.636 0.000 0.000 0.625 0.992 0.000 0.322 1.000	1.000 0.000 0.962 1.000 0.000 NC NC NC NC NC 0.000 0.991	1.000 0.000 0.965 NC NC NC NC 1.000 1.000 1.000 0.000 0.999	1.000 0.000 0.967 1.000 NC NC 1.000 1.000 1.000 NC 1.000	0.000 0.000 1.000 NC NC NC NC NC NC NC 0.000 0.989	+	+
ORF3a	T:25435:C	ORF3a:L15L	0.000	0.000	0.000	0.530	0.239	1.000		
E	C:26261:T T:26320:C T:26445:C	E:S6L E:F26L E:S67S	0.000 0.000 1.000	0.060 0.000 1.000	1.000 NC NC	1.000 1.000 1.000	1.000 1.000 1.000	1.000 NC NC		
M	T:26908:G	M:L129R	0.186	0.047	1.000	1.000	0.967	NC		
ORF6	T:27351:C	ORF6:S50S	0.000	0.000	0.968	1.000	1.000	NC		
ORF7a	GATT:27758:G	ORF7a:del2	1.000	1.000	1.000	1.000	1.000	NC		
ORF8	C:27945:T	ORF8:Q18*	1.000	0.995	1.000	0.987	1.000	NC		
N	C:28289:A A:28856:G	N:P6T N:R195G	0.140 0.428	0.061 0.702	0.993 0.000	0.993 0.000	0.997 0.000	0.960 0.000	+	

707

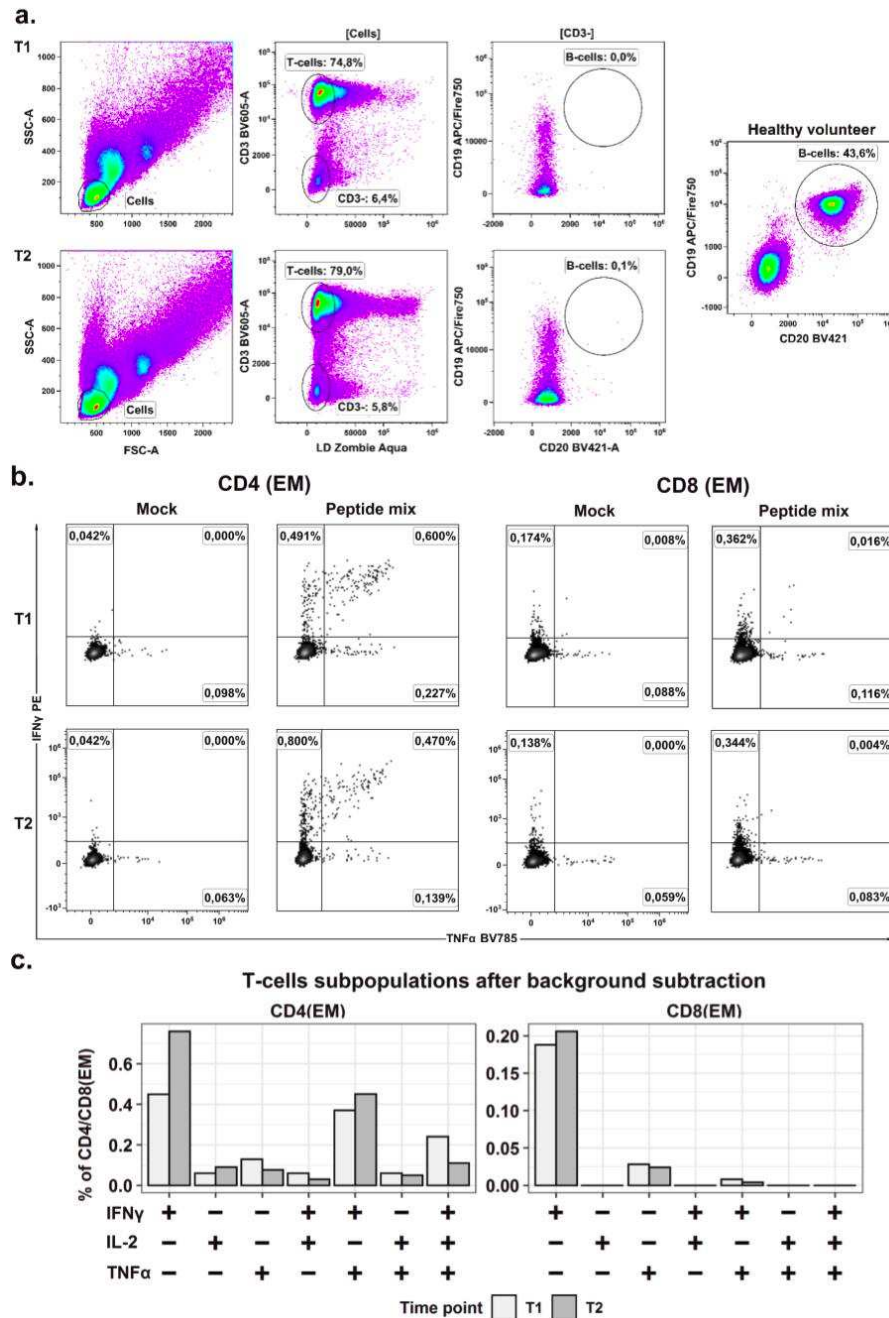


Extended Data Fig. 2. Patient S is robustly placed outside the cluster 5 clade. The abridged phylogeny of the B.1.1 lineage phylogeny is shown. Only those samples are shown which met either of the following conditions: (i) carried any of the differences found between the B.1.1 root and the patient S. sample (black cells), and these mutations had occurred in the branch immediately descendant from the B.1.1 root; or (ii) carried either the S:Δ69-70HV (blue cells) or the S:Y435F (yellow cells) mutation, independent of the timing of their origin. Additionally, we retained the samples from the branches that separate the cluster 5 clade from the rest of the phylogeny (two additional mutations, purple cells). Samples that didn't meet these criteria were collapsed, with the number of such samples shown in parentheses. The retained samples were then grouped by country, with names formatted as 'country|date of the earliest sample|number of samples'. B.1.1.7 and cluster 5 samples are shaded as in Fig. 1b. The presence of the above-mentioned mutations is indicated by the matrix at the right. Two mutations distinguishing cluster 5 from the B.1.1 root (purple) reject uniting patient S and cluster 5 in the same clade. For patient S, mutations with allele frequency below 50% in all three samples are shown in grey. Missing data ('N's in sequences) are shown as crosses. FFPE (black dot), patient A sample (the presumed source of infection for patient S).



Extended Data Fig. 3 Effect of viral evolution in patient S on neutralization by antibodies.

Neutralizing activity of serum obtained from 16 convalescent donors against patient S virus samples obtained on August 20, 2020 and February 19, 2021, as well as a reference viral strain of the B.1 lineage isolated from a swab sample in the beginning of the pandemic in Russia in March 2020. The August 20, 2020 isolate demonstrated reduced sensitivity to neutralizing antibodies compared to the reference strain, with the geometric mean fold decrease of 1.6 (CI 1.2-2.0, range 0.8-4.0). The February 19, 2021 isolate carried no signature of reduced sensitivity, and was indistinguishable in its sensitivity to neutralizing antibodies from the reference strain. Each sample was tested in triplicate and GMTs are plotted. Mann-Whitney-Wilcoxon test with Holm adjustments was used for pairwise comparisons.



738

739 **Extended Data Fig. 4 Features of immune response in patient S.** **a:** Flow cytometry plots showing the
 740 absence of B cells); **b:** CD4 and CD8 T-cell responses to SARS-CoV-2 overlapping peptide pools
 741 (N+RBD). Representative flow cytometry plots showing the cytokine profiles of SARS-CoV-2-specific
 742 CD4 and CD8 effector memory T cells after the stimulation; **c:** Bar-plots representing the percentage of
 743 different cytokine-producing populations of SARS-CoV-2-specific CD4 and CD8 T cells after
 744 background subtraction (data from the mock-stimulated sample were subtracted from peptide-stimulated
 745 samples). Time points T1 and T2 correspond to August 20, 2020 and February 16, 2021 respectively.
 746 Stimulation with Nucleocapsid (N) and RBD region of SARS-CoV2 Spike (S) protein OPP provoked
 747 expansion of both SARS-CoV-2-specific CD4 and CD8 T cells; the CD4 T-cell response predominated
 748 over CD8, as usual for COVID-19 patients^{71,74}.

749 **Extended Data Table 3.** Results of HLA calling from WES data using HLA-HD. Classical HLA genes
750 A, B, C, DRB1 and DQB1 were additionally confirmed with conventional genotyping. These alleles were
751 further combined into haplotypes via HLA-2-Haplo software. Thus, haplotype 1 corresponds to A - 01:01,
752 B - 08:01, C - 07:01, DRB1 - 03:01, DQB1 - 02:01 (population frequency 5.98e-2); and haplotype 2
753 corresponds to A - 03:01, B - 07:02, C - 07:02, DRB1 - 01:01, DQB1 - 05:01 (population frequency
754 3.37e-3). Worldwide allele frequencies are presented according to ref ^{35,36}.

755

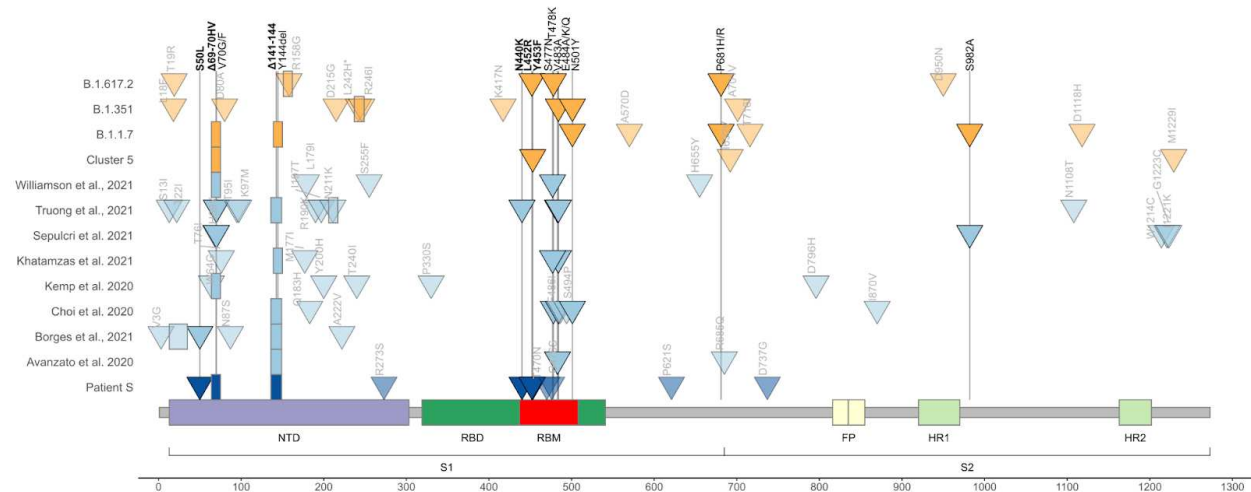
HLA	Allele 1	Frequency (Allele 1)	Allele 2	Frequency (Allele 2)
A	HLA-A*03:01:01	0.04272	HLA-A*01:01:01	0.04843
B	HLA-B*08:01:01	0.02960	HLA-B*07:02:01	0.04104
C	HLA-C*07:02:01	0.13101	HLA-C*07:01:01	0.06887
DRB1	HLA-DRB1*03:01:01	0.0676	HLA-DRB1*01:01:01	0.04123
DQA1	HLA-DQA1*01:01:01	-	HLA-DQA1*05:01:01	-
DQB1	HLA-DQB1*05:01:01	0.09307	HLA-DQB1*02:01:08	0.15003
DPA1	HLA-DPA1*01:03:01	-	-	-
DPB1	HLA-DPB1*04:02:01	0.18989	HLA-DPB1*04:01:01	0.23267

756

757

758 **Extended Data Table 4.** List of peptides experimentally validated in previous studies and included in the
759 IEDB database that overlapped the mutations observed in this study.

Mutation	HLA class	Peptide before mutation	Peptide after mutation	HLA and change of percentile rank
S:S50L	HLA I	FRSSVLH S T	FRSSVLH L T	HLA-C*07:01: Weak (0.74) -> Weak (0.93)
S:S50L	HLA I	S TQDLFLPF	L TQDLFLPF	HLA-A*01:01: Weak (1.1) -> Weak (1.6)
S:S50L	HLA I	S TQDLFLPFF	L TQDLFLPFF	HLA-A*01:01: Weak (1.2) -> Weak (1.7)
S:del140_144	HLA I	CNDP F LG V Y	CNDPYHKNN	HLA-A*01:01: Strong (0.39) -> No binding
S:del140_144	HLA I	G V Y YHKNNK	KNNKSWMES	HLA-A*03:01: Strong (0.046) -> No binding
S:del140_144	HLA I	FCNDP F LG V Y Y	FCNDPYHKNNK	HLA-A*01:01: Weak (0.59) -> No binding
S:R273S	HLA I	YLQP R T F LL	YLQP S T F LL	HLA-B*08:01: Strong (0.019) -> Weak (0.69)
S:T470N	HLA I	KPFERDIS T E I	KPFERDIS N E I	HLA-B*07:02: Strong (0.11) -> Strong (0.15)
nsp3:T504A	HLA I	T DN I YITTY	A DN I YITTY	HLA-A*01:01: Strong (0.14) -> Weak (0.72)
nsp3:T504A	HLA I	P T DN I YITTY	P A DN I YITTY	HLA-A*01:01: Strong (0.007) -> Strong (0.07)
nsp3:T504P	HLA I	T DN I YITTY	P DN I YITTY	HLA-A*01:01: Strong (0.14) -> No binding
nsp3:T504P	HLA I	P T DN I YITTY	P P DN I YITTY	HLA-A*01:01: Strong (0.007) -> Weak (0.73)
nsp3:D821N	HLA I	TT D PSFLGRY	TT N PSFLGRY	HLA-A*01:01: Strong (0.001) -> Strong (0.068)
nsp3:D821N	HLA I	HTT D PSFLGRY	HTT N PSFLGRY	HLA-A*01:01: Strong (0.04) -> Strong (0.2)
nsp3:D821N	HLA I	TT D PSFLGRY M	TT N PSFLGRY M	HLA-A*01:01: Strong (0.089) -> Weak (1.2)
nsp3:T1456I	HLA I	S T N V T IATY	S I N V T IATY	HLA-A*01:01: Strong (0.097) -> Weak (0.67)
endornase:P205L	HLA I	K P RSQMEIDF	K L RSQMEIDF	HLA-B*07:02: Strong (0.3) -> No binding
ORF8:Q18*	HLA I	Q S C T Q H Q P Y	-	HLA-A*01:01: Strong (0.48) -> Lost
ORF8:Q18*	HLA I	E P K L G S L V V	-	HLA-B*07:02: Strong (0.49) -> Lost
ORF8:Q18*	HLA I	V D D P C P I H F Y	-	HLA-A*01:01: Strong (0.16) -> Lost
N:P6T	HLA I	G P Q N Q R NAPRITF	G T Q N Q R NAPRITF	HLA-B*07:02: Strong (0.49) -> No binding
M:L129R	HLA I	VPLHG T I L	VPLHG T I R	HLA-B*07:02: Strong (0.29) -> No binding
ORF8:Q18*	HLA II	P C P I H F Y S K W I R V G	-	HLA-DRB1*01:01: Strong (0.54) -> Lost



Extended Data Fig. 5. Concordant origin of spike mutations in notable COVID-19 variants and reported cases of persistent COVID-19. Shown are the locations of mutations in the amino acid sequence encoded by the spike gene. Rows, from top to bottom: VOCs Delta (B.1.617.2), Beta (B.1.351), Alpha (B.1.1.7); Cluster 5 variant; immunosuppressed individual with persistent infection for 290 days (Williamson et al., 2021); three patients with acute lymphoblastic leukemia who were persistently positive for SARS-CoV-2 (Truong et al., 2021); immunosuppressed individual treated with immunoglobulin (Sepulcri et al. 2021); immunosuppressed individual treated with convalescent plasma (Khatamzas et al. 2021); immunosuppressed individual treated with convalescent plasma (Kemp et al. 2020); immunosuppressed individual treated with Regeneron monoclonal antibody cocktail (Choi et al. 2020; only those mutations present at the final timepoint (T3, day 152) are shown); immunocompromised patient without convalescent plasma treatment (Borges et al., 2021); immunocompromised individual treated with convalescent plasma (Avanzato et al. 2020); immunosuppressed individual not treated with convalescent plasma or antibodies (patient S, this study). Triangles, point mutations; rectangles, deletions. Bright colors represent mutations observed in at least two studies. Mutations labelled on top in black were observed in multiple lineages/experiments, among those, mutations that are present in Patient S are highlighted with bold font.

785 Online Reference List

- 786 38 Tyson JR, James P, Stoddart D, *et al.* Improvements to the ARTIC multiplex PCR method for
787 SARS-CoV-2 genome sequencing using nanopore. *Genomics*, 2020
788 DOI:10.1101/2020.09.04.283077.
- 789 39 Itokawa K, Sekizuka T, Hashino M, Tanaka R, Kuroda M. Disentangling primer interactions
790 improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLOS ONE* 2020; **15**:
791 e0239403.
- 792 40 Kawaguchi S, Matsuda F. High-Definition Genomic Analysis of HLA Genes Via
793 Comprehensive HLA Allele Genotyping. In: Tomar N, ed. *Immunoinformatics*. New York,
794 NY: Springer US, 2020: 31–8.
- 795 41 Geffard E, Limou S, Walencik A, *et al.* Easy-HLA: a validated web application suite to reveal
796 the full details of HLA typing. *Bioinformatics* 2020; **36**: 2157–64.
- 797 42 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
798 *Bioinformatics* 2014; **30**: 2114–20.
- 799 43 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
800 *ArXiv13033997 Q-Bio* 2013; published online May 26. <http://arxiv.org/abs/1303.3997>
801 (accessed May 13, 2021).
- 802 44 Grubaugh ND, Gangavarapu K, Quick J, *et al.* An amplicon-based sequencing framework for
803 accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* 2019;
804 **20**: 8.
- 805 45 Wilm A, Aw PPK, Bertrand D, *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant
806 caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets.
807 *Nucleic Acids Res* 2012; **40**: 11189–201.
- 808 46 Li H. A statistical framework for SNP calling, mutation discovery, association mapping and
809 population genetical parameter estimation from sequencing data. *Bioinformatics* 2011; **27**:
810 2987–93.
- 811 47 Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools.
812 *Bioinforma Oxf Engl* 2009; **25**: 2078–9.
- 813 48 Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
814 Improvements in Performance and Usability. *Mol Biol Evol* 2013; **30**: 772–80.
- 815 49 Wu F, Zhao S, Yu B, *et al.* A new coronavirus associated with human respiratory disease in
816 China. *Nature* 2020; **579**: 265–9.
- 817 50 Rambaut A, Holmes EC, O’Toole Á, *et al.* A dynamic nomenclature proposal for
818 SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020; **5**: 1403–7.
- 819 51 Sergei Pond. Evolutionary annotation of global SARS-CoV-2/COVID-19 genomes enabled by
820 data from GISAID. *observablehq.com* 2020; published online Dec 24.
- 821 52 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective
822 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* 2015;
823 **32**: 268–74.
- 824 53 Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis.
825 *Virus Evol* 2018; **4**. DOI:10.1093/ve/vex042.
- 826 54 Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for
827 Large Alignments. *PLoS ONE* 2010; **5**: e9490.
- 828 55 Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and
829 NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif

deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res* 2020; **48**: W449–54.

56 R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing,. 2018. <https://www.R-project.org/>.

57 Ginestet C. ggplot2: Elegant Graphics for Data Analysis: Book Reviews. *J R Stat Soc Ser A Stat Soc* 2011; **174**: 245–6.

58 Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinforma Oxf Engl* 2007; **23**: 127–8.

59 Kosakovsky Pond SL, Frost SDW. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol Biol Evol* 2005; **22**: 1208–22.

60 Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, *et al.* Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* 2021; **371**: 172–7.

61 Ria Lassaunière, Jannik Fona, Morten Rasmussen, Anders Frische, Charlotta Polacek Strandh, Thomas Bruun Rasmussen, Anette Bøtner, Anders Fomsgaard ger. Working paper on SARS-CoV-2 spike mutations arising in Danish mink, their spread to humans and neutralization data. 2020. https://files.ssi.dk/Mink-cluster-5-short-report_AFO2.

62 Motozono C, Toyoda M, Zahradnik J, *et al.* SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host Microbe* 2021; : S1931312821002845.

63 Vítor Borges, Joana Isidro, Mário Cunha, Daniela Cochicho, Luis Martins, Luis Banha, Margarida Figueiredo, Leonor Rebelo, Maria Céu Trindade, Sílvia Duarte, Luís Vieira, Maria João Alves, Inês Costa, Raquel Guiomar, Madalena Santos, Rita Cortê-Real, André Dias, Diana Póvoas, João Cabo, Carlos Figueiredo, Maria José Manata, Fernando Maltez, Maria Gomes da Silva, João Paulo Gomes. Long-term evolution of SARS- CoV-2 in an immunocompromised patient with non-Hodgkin lymphoma. *virological.org* 2020; published online Feb 18.

64 Truong TT, Ryutov A, Pandey U, *et al.* Increased Viral Variants in Children and Young Adults with Impaired Humoral Immunity and Persistent Sars-Cov-2 Infection: A Consecutive Case Series. *EBioMedicine* 2021; **67**: 103355.

65 McCarthy KR, Rennick LJ, Nambulli S, *et al.* Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* 2021; **371**: 1139–42.

66 Andrew Rambaut, Nick Loman, Oliver Pybus, Wendy Barclay, Jeff Barrett, Alesandro Carabelli, Tom Connor, Tom Peacock, David L Robertson, Erik Volz, on behalf of COVID-19 Genomics Consortium UK (CoG-UK). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Virological* 2020; published online Dec 18. <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/563/1>.

67 Wang R, Chen J, Hozumi Y, Yin C, Wei G-W. Decoding Asymptomatic COVID-19 Infection and Transmission. *J Phys Chem Lett* 2020; **11**: 10007–15.

68 Focosi D, Maggi F. Neutralising antibody escape of SARS-CoV-2 spike protein: Risk assessment for antibody-based Covid-19 therapeutics and vaccines. *Rev Med Virol* 2021; : rmv.2231.

69 Pereira F. Evolutionary dynamics of the SARS-CoV-2 ORF8 accessory gene. *Infect Genet Evol* 2020; **85**: 104525.

70 Zinzula L. Lost in deletion: The enigmatic ORF8 protein of SARS-CoV-2. *Biochem Biophys Res Commun* 2021; **538**: 116–24.

876 71 Zhang Y, Chen Y, Li Y, *et al.* The ORF8 protein of SARS-CoV-2 mediates immune evasion
877 through down-regulating MHC-I. *Proc Natl Acad Sci* 2021; **118**: e2024202118.
878 72 Simpson S, Kay FU, Abbara S, *et al.* Radiological Society of North America Expert
879 Consensus Document on Reporting Chest CT Findings Related to COVID-19: Endorsed by
880 the Society of Thoracic Radiology, the American College of Radiology, and RSNA. *Radiol*
881 *Cardiothorac Imaging* 2020; **2**: e200152.
882 73 Cancer Therapy Adviser.
883 [https://www.cancertherapyadvisor.com/home/cancer-topics/hematologic-cancers/hematologic-](https://www.cancertherapyadvisor.com/home/cancer-topics/hematologic-cancers/hematologic-cancers-treatment-regimens/non-hodgkin-lymphoma-nhl-treatment-regimens-diffuse-large-b-cell-lymphoma/)
884 [cancers-treatment-regimens/non-hodgkin-lymphoma-nhl-treatment-regimens-diffuse-large-b-c](https://www.cancertherapyadvisor.com/home/cancer-topics/hematologic-cancers/hematologic-cancers-treatment-regimens/non-hodgkin-lymphoma-nhl-treatment-regimens-diffuse-large-b-cell-lymphoma/)
885 [ell-lymphoma/](https://www.cancertherapyadvisor.com/home/cancer-topics/hematologic-cancers/hematologic-cancers-treatment-regimens/non-hodgkin-lymphoma-nhl-treatment-regimens-diffuse-large-b-cell-lymphoma/).
886 74 Habel JR, Nguyen THO, van de Sandt CE, *et al.* Suboptimal SARS-CoV-2-specific CD8⁺
887 T-cell response associated with the prominent HLA-A*02:01 phenotype. *Infectious Diseases*
888 (except HIV/AIDS), 2020 DOI:10.1101/2020.08.17.20176370.
889

890

891

892

893

894

895

896

897

898

899

900

901

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [S11.xlsx](#)