Intrahost evolution and forward transmission of a novel SARS-CoV-2 Omicron BA.1 subvariant

Ana S. Gonzalez-Reiche[1], Hala Alshammary[2], Sarah Schaefer[3], Gopi Patel[3], Jose Polanco[2,4], Angela A. Amoako[2,4], Aria Rooker[2,4], Christian Cognigni[2,4], Daniel Floda[1], Adriana van de Guchte[1], Zain Khalil[1], Keith Farrugia[1], Nima Assad[5], Jian Zhang[1], Bremy Alburquerque[1,5], Levy Sominsky[2,4], Komal Srivastava[2,4], Robert Sebra[1,8,9,10], Juan David Ramirez[6,11], Radhika Banu[6], Paras Shrestha[6], Alberto Paniz-Mondolfi[6], Emilia Mia Sordillo[6 †], Viviana Simon[2,3,4,6,7, †], Harm van Bakel[1,6,8,†]

1. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
2. Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
3. Division of Infectious Diseases, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
4. Center for Vaccine Research and Pandemic Preparedness (C-VARPP), Icahn School of Medicine at Mount Sinai, New York, NY, USA
5. Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
6. Department of Pathology, Molecular, and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
7. The Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
8. Icahn Genomics Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
9. Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA
10. Sema4, a Mount Sinai venture, Stamford CT, 06902
11. Centro de Investigaciones en Microbiología y Biotecnología-UR (CIMBIUR), Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia.

Corresponding authors/senior authors:

† Dr. Emilia Mia Sordillo (Emilia.Sordillo@mountsinai.org), Dr. Viviana Simon (viviana.simon@mssm.edu) and Dr. Harm van Bakel (harm.vanbakel@mssm.edu)

ABSTRACT

Persistent SARS-CoV-2 infections have been reported in immune-compromised individuals and people undergoing immune-modulatory treatments. It has been speculated that the emergence of antigenically diverse SARS-CoV-2 variants such as the Omicron variant may be the result of intra-host viral evolution driven by suboptimal immune responses, which must be followed by forward transmission. However, while intrahost evolution has been documented, to our knowledge no direct evidence of subsequent forward transmission is available to date.

Here we describe the emergence of an Omicron BA.1 sub-lineage with 8 additional amino acid substitutions within the spike (E96D, L167T, R346T, L455W, K458M, A484V, H681R, A688V) in an immune-compromised host along with evidence of 5 forward transmission cases. Our findings show that the Omicron BA.1 lineage can further diverge from its exceptionally mutated genome during prolonged SARS-CoV-2 infection; highlighting an urgent need to employ therapeutic strategies to limit duration of infection and spread in vulnerable patients.

52 **INTRODUCTION**

53 Emergence of antigenically diverse SARS-CoV-2 variants of concern (VOC) poses a
54 challenge to SARS-CoV-2 vaccine-induced immunity. Prolonged evolution in
55 immunocompromised hosts is thought to have played a role in the emergence of
56 several SARS-CoV-2 VOC such as Alpha and Omicron[1,2]. However, although persistent
57 SARS-CoV-2 replication with stepwise acquisition of mutations in spike within
58 immunocompromised COVID-19 patients has been documented[3-7], clear evidence for
59 forward transmission has been lacking.

60 The Mount Sinai Pathogen Surveillance Program (MS-PSP) has profiled the evolving
61 landscape of SARS-CoV-2 in New York City (NYC) since the beginning of the
62 pandemic[8,9]. We routinely perform complete viral genome sequencing from at least 10%
63 of randomly selected contemporaneous SARS-CoV-2 positive specimens collected from
64 patients seeking care at our health system.

65 We describe a persistent SARS-CoV-2 Omicron BA.1 infection in an immuno-
66 compromised individual during a 12-week period and document the accumulation of
67 eight additional amino acid substitutions in the already antigenically distinct Omicron
68 BA.1 spike protein. In addition, we identified five subsequent cases harboring the same
69 genotype; three of which were identified in the same health system. The presence of the
70 unique combination of mutations in all six cases is indicative of forward transmission of
71 this novel Omicron BA.1 sub-lineage. Although the majority of amino acid changes
72 occur at positions known to confer either immune escape[10,11] or altered viral
73 fusogenicity[12-14], some of the mutations have rarely been seen in other lineages, and
74 their overall constellation is unique.

75 **MATERIALS AND METHODS**

76 *Molecular SARS-CoV-2 diagnostics*

77 SARS-CoV-2 molecular diagnostic testing was performed in the Molecular Microbiology
78 Laboratories of the MSHS Clinical Laboratory by nucleic acid amplification tests (NAAT)
79 that have been validated for nasopharyngeal, anterior nares swabs and saliva
80 specimens. All but one positive samples included in this study were tested using the
81 PerkinElmer® New Coronavirus Nucleic Acid Detection Kit which provides qualitative
82 detection of nucleic acid from SARS-CoV-2. It includes two SARS-CoV-2 targets
83 (ORF1ab, N) and one internal positive control (IC; bacteriophage MS2). Cycle threshold
84 (Ct) values are generated for all three targets and provide a quantitative estimate of the
85 viral load. The only other positive sample that was run on another testing platform was
86 the first specimen of the index case (P1). It was tested as a point of care test (POC)
87 using the LIAT SARS-CoV-2 assay, and the biospecimen was discarded prior to
88 transfer to the MS-PSP. In addition, two other negative specimens registered for patient
89 P3 during the same time window were run in the LumiraDx assay, which does not
90 provide Ct values.

91

92 *The Mount Sinai Pathogen Surveillance Program (MS-PSP)*

93 Residual nasopharyngeal and anterior nares (AN) swab specimens were collected after
94 completion of the diagnostic process, as part of the Mount Sinai Pathogen Surveillance
95 Program (IRB approved HS#13-00981).

96

97 *Complete SARS-CoV-2 genome sequencing*

98 RNA was extracted using the Chemagic™ Viral DNA/RNA 300 Kit H96 (PerkinElmer,
99 CMG-1033-S) on the automated Chemagic™ 360 instrument (PerkinElmer, 2024-0020)
100 from 300uL of viral transport media per the manufacturer's protocol. cDNA synthesis
101 followed by whole genome amplification with two custom primer panel sets targeting 1.5
102 and 2kb regions across the SARS-CoV-2 genome was performed as previously
103 described[8] with addition of new oligonucleotides to minimize amplicon dropout for

104 Omicron lineages derived from PANGO lineage B.1.1.529 (available on request).
105 Paired-end (2x150bp) Nextera XT (Illumina, cat. FC-131-1096) libraries prepared from
106 amplicons were sequenced on a MiSeq instrument. SARS-CoV-2 genomes were
107 assembled using a custom reference-based pipeline as previously described[8]. The final
108 sequencing depth per genome ranged between ~135k to ~415k reads.

109

110 *Phylogenetic analysis*

111 Global background SARS-CoV-2 sequences were downloaded from the GISAID
112 EpiCOV database (as of 2022-05-13). The GISAID database was queried for the novel
113 mutations observed in P1-P4 sequences in order to identify their closest matches.
114 Sequence hits were confirmed by their Mash distance[15]. For this, a genome sketch was
115 generated from the sanitized alignment of the global sequences filtered for lineages
116 BA.1.* as produced by Nextstrain[16,17] v11 for SARS-CoV-2 with default parameters
117 (https://github.com/nextstrain/ncov). This allowed pairwise comparisons of our data with
118 high-quality global sequences with available metadata. Phylogenetic inferences of the
119 MS-PSP SARS-CoV-2 genomes, including the closest sequence matches and all other
120 BA.* lineage sequences from the MS-PSP surveillance program were done using an
121 Omicron BA.* and New York State-focused background with proximity subsampling
122 scheme with Nextstrain. Clade and lineage assignments were done with Nextclade CLI
123 v3.2 (2021-11-04)[18] and with PANGO-v1.8 (pangolin v3.1.17, pangoLEARN v.2022-04-
124 22) [19,20].

125

126 *Viral cultures*

127 Replication-competent SARS-CoV-2 was obtained as previously described[21]. Briefly,
128 Vero-E6 cells expressing TMPRSS2 were cultured in Dulbecco's Modified Eagle
129 Medium containing 10% heat-inactivated fetal bovine serum and 1% Minimum Essential
130 Medium (MEM) Amino Acids Solution, supplemented with 100 U/ml penicillin, 100 µg/ml
131 streptomycin, 100 µg/ml normocin, and 3 µg/ml puromycin. 200ul of viral transport
132 media from the nasopharyngeal or anterior nares swab specimen was added to Vero-

133 E6-TMPRSS2 cells in culture media supplemented with 0.5 µg/ml Amphotericin B.
134 Cultures were maintained for 10 days. Culture supernatants were collected and clarified
135 by centrifugation (4000g, 5min) upon the appearance of cytopathic effect (CPE). The
136 viral stocks were sequence-verified. Viral cultures were successful for the initial
137 specimens obtained from P2, P3 and P4, but failed for all P1 specimens.

138

139 **RESULTS AND DISCUSSION**

140 In this study we report the genomic analysis of serially collected nasopharyngeal (NP)
141 and anterior nares (AN) samples from an immunocompromised patient with persistent
142 SARS-CoV-2 Omicron BA1 replication between December 2021 and March 2022. Over
143 a 12-week period, we documented the accumulation of eight amino acid substitutions in
144 the spike N-terminal domain (NTD), the receptor binding domain (RBD), and cleavage
145 site region (CSR) (**Figure 1**) within the same patient. The first three mutations R346T,
146 K458M and E484V were detected simultaneously 40 days after initial SARS-CoV-2
147 diagnosis and were fixed. Two weeks later (day 64), L167T and the furin cleavage site
148 mutation P681Y were detected in addition to three mutations listed. During the following
149 weeks (day 72 and day 81), two different viral populations emerged each carrying
150 shared (L455W) as well as additional distinct signature mutations (E96D at day 72,
151 S477D at day 81). The majority of the amino acid substitutions that accumulated in
152 spike have been associated with neutralization escape[10,11] or improved viral
153 fusogenicity[12]. All but one of the mutations observed in the consensus sequences
154 outside the spike gene were non-synonymous, further supporting positive intrahost
155 selection of spike changes with competitive advantage.

156 Background health system-wide SARS-CoV-2 genomic surveillance by MS-PSP
157 conducted during the same time period identified three other patients with COVID-19
158 harboring SARS-CoV-2 Omicron variants that were similar to that of the index case and
159 consistent with multiple forward transmissions based on the timeline of infection (**Figure**
160 **1**, and see **Table 1** and **Figure 2** for details and the CT values for the individual
161 specimens). These three forward transmissions were detected in immunocompromised
162 individuals. In one case, infection persisted for more than three weeks resulting in

163 acquisition of an additional mutation in the spike RBD (V445A) (**Figure 1**). In some
164 specimens we detected minority alleles at variant positions suggesting ongoing intra-
165 host selection (**Figure 1** and **Figure 3**). Although viral isolation failed for specimens
166 available for the index case, we successfully cultured SARS-CoV-2 from the initial
167 positive specimens of the three forward transmission cases, confirming the transmission
168 of replication-competent virus.

169 Lastly, two additional related SARS-CoV-2 genomes from the NYC area were identified
170 from the GISAID database (**Figure 1**). Based on the metadata provided, these two
171 cases differed by age and gender from our cases, consistent with limited community
172 spread of this new Omicron BA1 sub-variant. All transmission cases contained one
173 synonymous change in Orf1a (T6001C) previously only observed at day 72 in the index
174 case (**Figure 4**), narrowing the time window of the first transmission to a few weeks
175 following day 64. As of mid-May 2022, the most recent case was detected in mid-April
176 2022 and the last SARS-CoV-2 positive specimen from a forward transmission was
177 collected in early May 2022.

178 In conclusion, our data demonstrate that intra-host evolution of SARS-CoV-2 occurring
179 during persistent infection in immunocompromised individuals can drive the emergence
180 and spread of novel (sub)variants with extensive mutations in key antigenic regions,
181 even in the context of the already highly divergent Omicron lineage. Our data suggest
182 that genomic monitoring of persistent long-term infections is warranted and further
183 underscore the need to limit the duration of viral infection. Improved early detection of
184 novel SARS-CoV-2 variants, a better understanding of the selection processes driving
185 SARS-CoV-2 evolution, and therapies that limit virus persistence and shedding are
186 essential to reduce emergence of highly mutated viral variants in the future.

187

188 **ACKNOWLEDGMENTS**

189 We thank the Mount Sinai Hospital and School Leadership, in particular Dr. David Reich
190 and Dr. Dennis Charney, for their ongoing support of the MS-PSP throughout the
191 pandemic. We thank the laboratory technicians in the Molecular Microbiology
192 Laboratories and the Rapid Response Laboratories of the Mount Sinai Health System

since without their assistance and help, none of this surveillance work would be possible.

We gratefully acknowledge the authors and originating and submitting laboratories of sequences from GISAID's EpiCoV (www.gisaid.org) that were used as background for our phylogenetic inferences.

FUNDING SOURCES

The work reported was in part supported by a contract from the *National Institute of Allergy and Infectious Diseases* (75N93021C00014, Option 12A) awarded to the Center for Research on Influenza Pathogenesis and Transmission – (a Center of Excellence for Influenza Research and Response) as part of the *SARS-CoV-2 Assessment of Viral Evolution (SAVE) Program*, a contract from the *National Institute of Allergy and Infectious Diseases* (HHSN272201400008C) awarded to the Center for Research on Influenza Pathogenesis – (a Center of Excellence for Influenza Research and Surveillance), philanthropic donations from the *JPB Foundation*, a research grant (2020-215611 [5384]) from the *Open Philanthropy Project*, and awards (S10OD026880 and S10OD030463) from the *NIH Office of Research Infrastructure Programs*.

COMPETING INTERESTS

The Icahn School of Medicine at Mount Sinai has filed patent applications relating to SARS-CoV-2 serological assays listing Viviana Simon as co-inventor. Mount Sinai has spun out a company, Kantaro, to market serological tests for SARS-CoV-2. Robert Sebra is VP of Technology Development and a stockholder at Sema4, a Mount Sinai Venture. This work, however, was conducted solely at Icahn School of Medicine at Mount Sinai.

References

1. Martin DP, Lytras S, Lucaci AG, et al. Selection analysis identifies unusual clustered mutational changes in Omicron lineage BA.1 that likely impact Spike function. bioRxiv 2022:2022.01.14.476382. DOI: 10.1101/2022.01.14.476382.
2. (CoG-UK) C-GCU. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. (https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563).
3. Corey L, Beyrer C, Cohen MS, Michael NL, Bedford T, Rolland M. SARS-CoV-2 Variants in Patients with Immunosuppression. N Engl J Med 2021;385(6):562-566. DOI: 10.1056/NEJMsb2104756.
4. 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;183(7):1901-1912 e9. DOI: 10.1016/j.cell.2020.10.049.
5. Choi B, Choudhary MC, Regan J, et al. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. N Engl J Med 2020;383(23):2291-2293. DOI: 10.1056/NEJMc2031364.
6. Kemp SA, Collier DA, Datir RP, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature 2021;592(7853):277-282. DOI: 10.1038/s41586-021-03291-y.
7. 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. DOI: 10.1016/j.ebiom.2021.103355.

243 8. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, et al. Introductions and early
244 spread of SARS-CoV-2 in the New York City area. Science 2020. DOI:
245 10.1126/science.abc1917.
246 9. Hernandez MM, Gonzalez-Reiche AS, Alshammary H, et al. Molecular evidence
247 of SARS-CoV-2 in New York before the first pandemic wave. Nat Commun
248 2021;12(1):3463. DOI: 10.1038/s41467-021-23688-7.
249 10. Greaney AJ, Starr TN, Barnes CO, et al. Mapping mutations to the SARS-CoV-2
250 RBD that escape binding by different classes of antibodies. Nat Commun
251 2021;12(1):4196. DOI: 10.1038/s41467-021-24435-8.
252 11. Starr TN, Greaney AJ, Hilton SK, et al. Deep Mutational Scanning of SARS-CoV-
253 2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding.
254 Cell 2020;182(5):1295-1310 e20. DOI: 10.1016/j.cell.2020.08.012.
255 12. Escalera A, Gonzalez-Reiche AS, Aslam S, et al. Mutations in SARS-CoV-2
256 variants of concern link to increased spike cleavage and virus transmission. Cell
257 Host Microbe 2022. DOI: 10.1016/j.chom.2022.01.006.
258 13. Saito A, Irie T, Suzuki R, et al. Enhanced fusogenicity and pathogenicity of
259 SARS-CoV-2 Delta P681R mutation. Nature 2022;602(7896):300-306. DOI:
260 10.1038/s41586-021-04266-9.
261 14. Meng B, Abdullahi A, Ferreira I, et al. Altered TMPRSS2 usage by SARS-CoV-2
262 Omicron impacts infectivity and fusogenicity. Nature 2022;603(7902):706-714.
263 DOI: 10.1038/s41586-022-04474-x.

15. Ondov BD, Treangen TJ, Melsted P, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 2016;17(1):132. DOI: 10.1186/s13059-016-0997-x.

16. Bedford T, Neher R, Hadfield J, et al. Nextstrain build for novel coronavirus SARS-CoV-2. 2020.

17. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 2018;34(23):4121-4123. DOI: 10.1093/bioinformatics/bty407.

18. Aksamentov IR, Cornelius; Hodcroft, Emma B.; Neher, Richard A. Nextclade: clade assignment, mutation calling and quality control for viral genomes. Journal of Open Source Software 2021;6(67):3773. DOI: 10.21105/joss.03773.

19. O'Toole Á, McCrone JT. Phylogenetic assignment of named global outbreak lineages. 2020-05-19 ed2020.

20. O'Toole A, Hill V, Pybus OG, et al. Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 with grinch. Wellcome Open Res 2021;6:121. DOI: 10.12688/wellcomeopenres.16661.2.

21. Carreno JM, Alshammary H, Tcheou J, et al. Activity of convalescent and vaccine serum against SARS-CoV-2 Omicron. Nature 2022;602(7898):682-688. DOI: 10.1038/s41586-022-04399-5.

284 **TABLES**

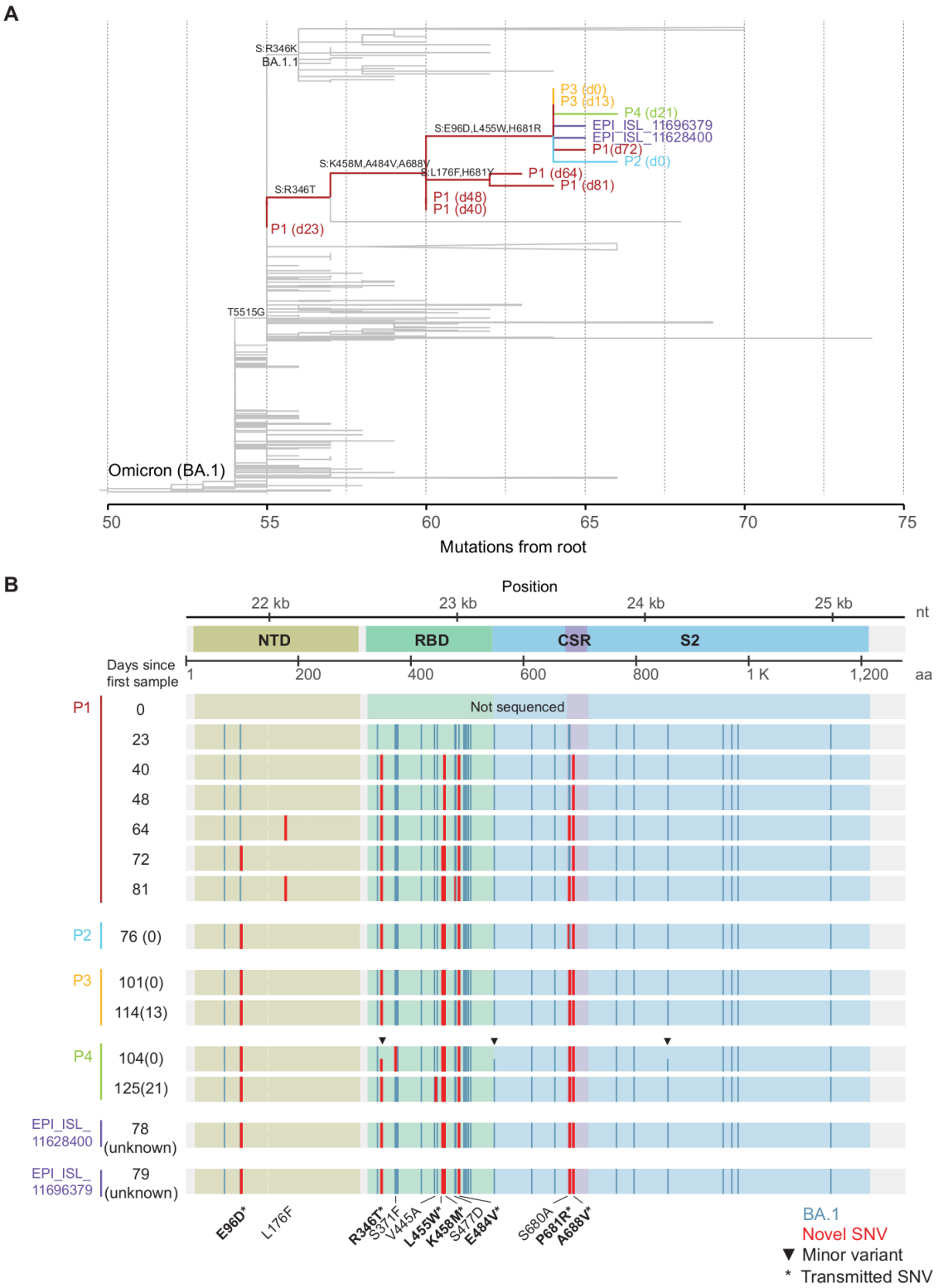
285 **Table 1: Spike mutations detected in the sequenced cases. Mutations are shown relative to BA.1, the lineage of the**

286 **earliest sample. Mutations in bold correspond to changes at positions already mutated in BA.1.**

Patient ID	Sex	Age	Study Day	GISAID Accession	Source	Mutations in Spike relative to earliest sample (BA.1)											
			0	Not sequenced	This study												
			23	EPI_ISL_11897513	This study	BA.1: A67V, T95I, Y145D, L212I, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F											
P1	M	61-65	40	EPI_ISL_11897516	This study			R346T					K458M		**A484V**		
			48	EPI_ISL_11178997	This study			R346T					K458M		**A484V**		
			64	EPI_ISL_11179080	This study		L176F	R346T					K458M		**A484V**		**H681Y**
			72	EPI_ISL_11685451	This study	E96D		R346T				L455W	K458M		**A484V**		
			81	EPI_ISL_11897525	This study		L176F	R346T				L455W	K458M	S477D	**A484V**		**H681Y**
P2	F	66-70	76	EPI_ISL_11685480	This study	E96D		R346T				L455W	K458M		**A484V**	S680A	
P3	M	36-40	101	EPI_ISL_12422465	This study	E96D		R346T				L455W	K458M		**A484V**		**H681R**
			114	EPI_ISL_12711034	This study	E96D		R346T				L455W	K458M		**A484V**		**H681R**
P4	F	61-65	104	Pending	This study	E96D		R346R/T	S371F			L455W	K458M		**A484V**	T547T/K	**H681R**
			125	Pending	This study	E96D		R346T		V445A		L455W	K458M		**A484V**		**H681R**
	F	66-70	79	EPI_ISL_11696379	Howard et al. 2022, CDC. (www.gisaid.org/)	E96D		R346T				L455W	K458M		**A484V**		**H681R**
	M	66-70	78	EPI_ISL_11628400		E96D		R346T				L455W	K458M		**A484V**		**H681R**

287

288 **FIGURES**



289

290 **Figure 1. Intrahost emergence and forward transmission of an Omicron BA.1**
291 **subvariant**

292 **A)** Maximum likelihood phylogenetic subtree with SARS-CoV-2 (BA.1) sequences from
293 the prolonged-infection case and forward transmissions, in a global background of
294 sequences available in GISAID. Patients are depicted in different colors and the
295 number of days after the first SARS-CoV-2 positive specimen is indicated in
296 brackets. Sibling clusters of BA.1 and other sublineages are collapsed for easier
297 visualization. The x-axis shows the number of nucleotide substitutions relative to the
298 root of the phylogenetic tree.

299 **B)** Nucleotide and amino acid substitutions in the spike protein observed in sequential
300 specimens obtained from patient 1 (P1) with prolonged infection with BA.1 and
301 transmission cases (P2, P3 and P4). P1 shows progressive accumulation of
302 mutations in the N-terminal domain (NTD), receptor binding domain (RBD), and
303 cleavage site region (CSR); The same constellation of mutations was subsequently
304 detected in three documented transmission cases (P2, P3 and P4) and in two
305 sequences from GISAID (EPI_ISL_11628400 and EPI_ISL_11696379). The number
306 of days since the first positive test in P1 is shown on the left, with the number of
307 days after the first positive test for each patient between brackets. The BA.1
308 substitutions are shown in blue and novel substitutions are shown in red, relative to
309 the reference genome sequence NC_045512.2. Positions with minor single
310 nucleotide variants at the relevant positions are indicated.

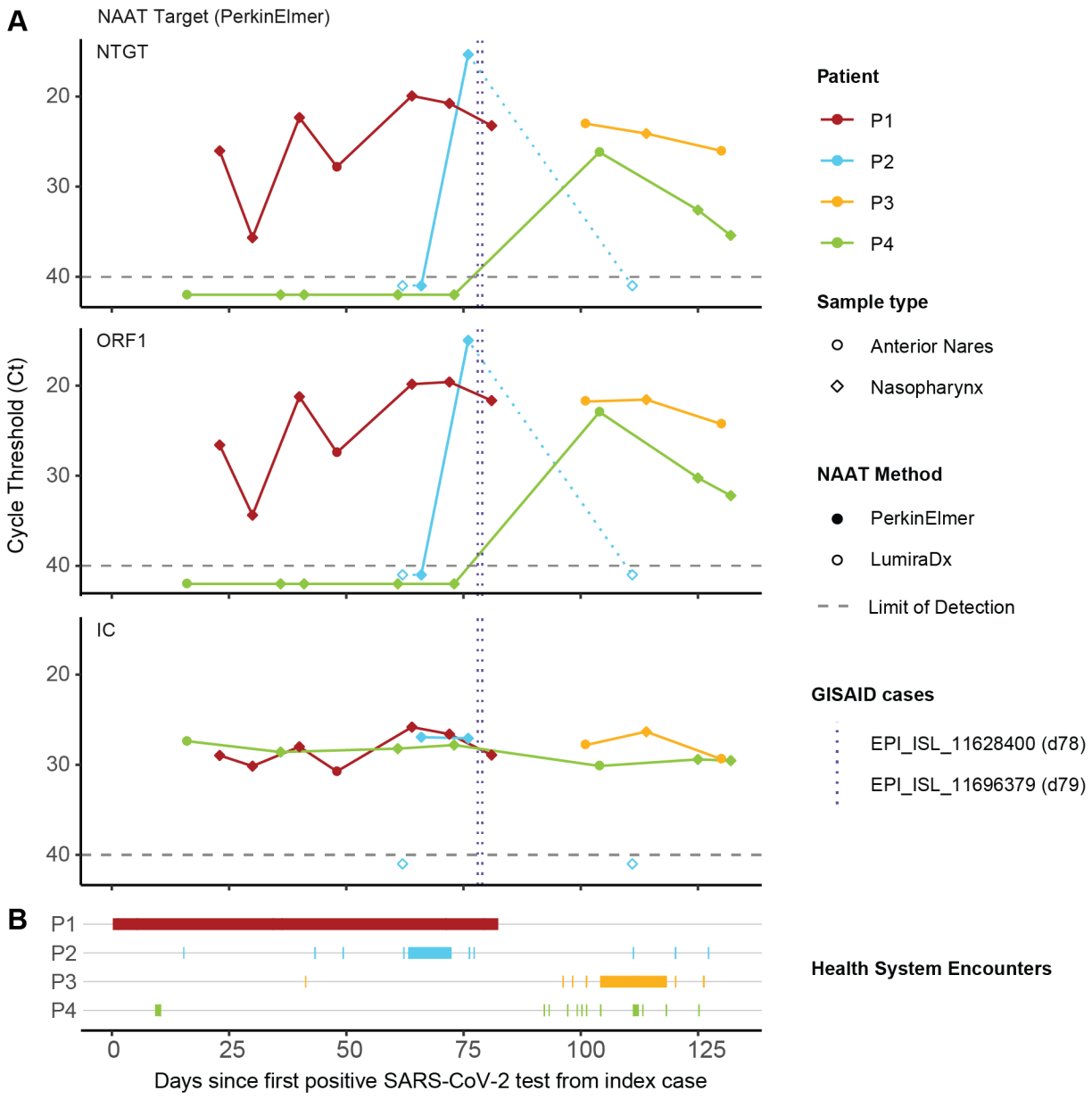
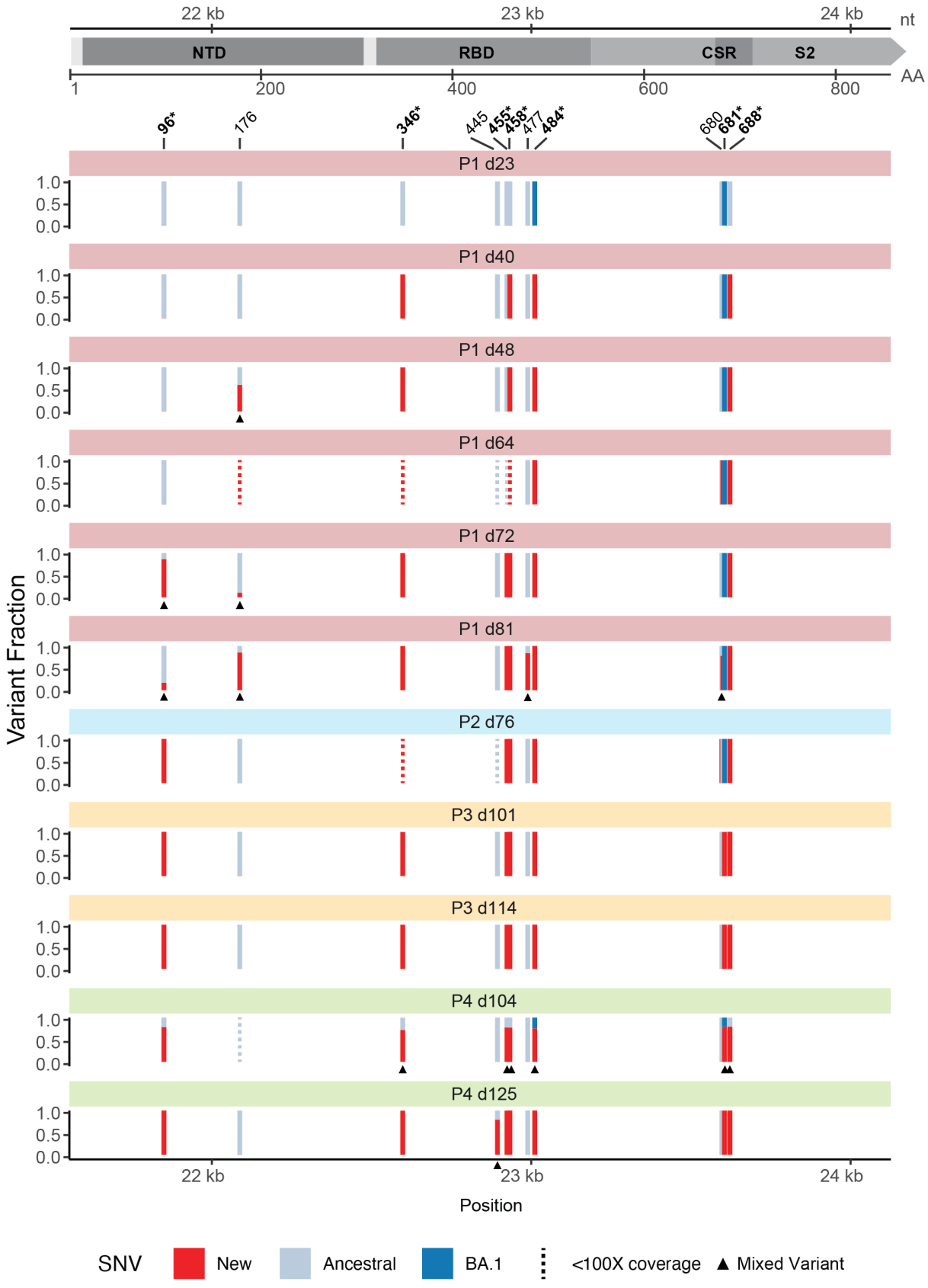


Figure 2: Overview of the longitudinal SARS-CoV-2 viral load estimates and the clinical encounters for the four cases captured by the MS-PSP.

A) PerkinElmer® Real-time RT-PCR SARS-CoV-2 cycle threshold (Ct) values for nasopharyngeal and anterior nares specimens. The horizontal dashed line represents the cutoff value for a positive result. Vertical lines represent the sampling days of the cases recorded in GISAID relative to P1's first positive test. For P2, the

negative tests by alternative diagnostic method (LumiraDx™) with no Ct value available, are indicated with empty markers. NTGT: N gene, ORF1: ORF1ab gene, IC: Internal Control.

B) Timeline of the encounters of P1, P2, P3 and P4 with the health system. Each tick represents an individual hospital encounter while the horizontal bars denote hospital admissions (duration is indicated by the length of the bar).


22 kb
23 kb
24 kb
nt
NTD
RBD
CSR
S2
1
200
400
600
800
AA
96*
176
346*
445
455*
458*
477
484*
680
681*
688*
P1 d23
P1 d40
P1 d48
P1 d64
P1 d72
P1 d81
P2 d76
P3 d101
P3 d114
P4 d104
P4 d125
Variant Fraction
1.0
0.5
0.0
Position
SNV
New
Ancestral
BA.1
<100X coverage
Mixed Variant

325

326 **Figure 3: Summary of the SARS-CoV-2 minority viral variant distribution within**
327 **diagnostic nasal biospecimen.**

328 The distribution of nucleotide variants below the consensus level for positions that
329 changed over the course of the infection of the index and transmitted infections are
330 shown. Single nucleotide variants (SNVs) are colored by strain. The dotted bars indicate
331 positions with coverage below the threshold of 100X coverage for calling minority
332 variants. The positions with minority variants present at frequencies > 0.1% are
333 indicated with triangles for their respective samples.

334

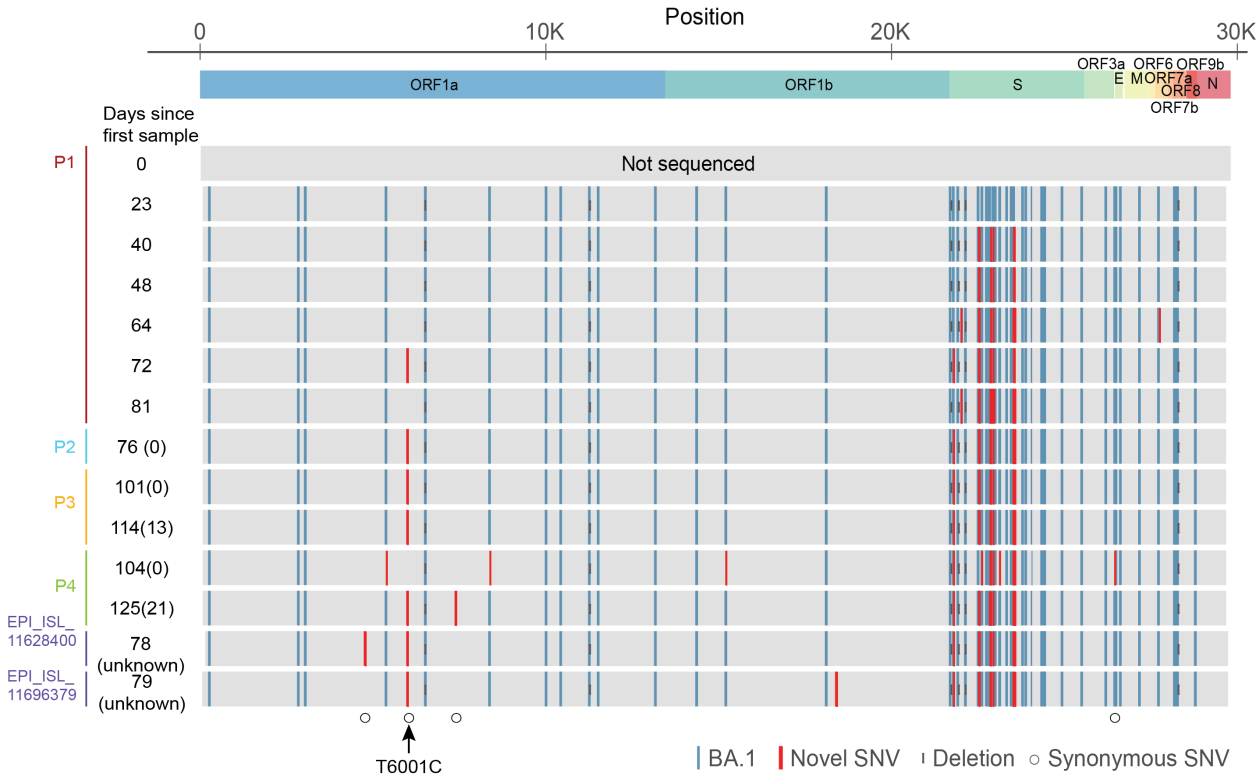


Figure 4: Overview of genome-wide SARS-CoV-2 mutations found in the diagnostic nasal biospecimen.

The mutations present across the complete SARS-CoV-2 genome are shown. Nucleotide substitutions observed in sequential specimens obtained from patient 1 (P1) with prolonged infection with BA.1 and forward transmission cases (P2, P3 and P4). There is accumulation and fixation of new SNVs in the spike region in P1.

The same constellation of mutations was subsequently detected in three documented transmission cases (P2, P3 and P4) and in two sequences from GISAID (EPI_ISL_11628400 and EPI_ISL_11696379). The only shared synonymous SNV outside of the spike (ORF1a:T6001C) is indicated with an arrow.

The number of days since the first positive test in P1 is shown on the left, with the number of days after the first positive test for each patient between brackets. The BA.1 substitutions are shown in blue and novel substitutions are shown in red, relative to the reference genome sequence NC_045512.2.