SARS-CoV-2 variants in the making: Sequential intrahost evolution and forward transmissions in the context of persistent infections

Ana S. Gonzalez-Reiche[1], Hala Alshammary[2,4], Sarah Schaefer[3], Gopi Patel[3], Jose Polanco[2,4], Juan Manuel Carreño [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], PARIS/PSP study group[2,4], Levy Sominsky[2,4], Charles Gleason[2,4] Komal Srivastava[2,4], Robert Sebra[1,8,9,10], Juan David Ramirez[6,11], Radhika Banu[6], Paras Shrestha[6], Florian Krammer[2,4,6], Alberto Paniz-Mondolfi[6], Emilia Mia Sordillo[6 †], Viviana Simon[2,3,4,6,7, †], Harm van Bakel[1,8,†]

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)

PARIS/PSP study group:

Giulio Kleiner, Dalles Andre, Katherine F Beach, Maria C Bermúdez-González, Gianna Cai, Neko Lyttle, Lubbertus C F Mulder, Annika Oostenink, Ashley Beathrese T Salimbangon, Gagandeep Singh, Morgan van Kesteren, Brian Monahan, Jake Mauldin, Mahmoud Awawda

ABSTRACT – 193 WORDS

Persistent severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections have been reported in immune-compromised individuals and people undergoing immune-modulatory treatments. Although intrahost evolution has been documented, to our knowledge, no direct evidence of subsequent transmission and stepwise adaptation is available.

Here we describe sequential persistent SARS-CoV-2 infections in three individuals that led to the emergence, forward transmission, and continued evolution of a new Omicron sublineage, BA.1.23, over an eight-month period. The initially transmitted BA.1.23 variant encoded seven additional amino acid substitutions within the spike protein (E96D, R346T, L455W, K458M, A484V, H681R, A688V), and displayed substantial resistance to neutralization by sera from boosted and/or Omicron BA.1-infected study participants. Subsequent continued BA.1.23 replication resulted in additional substitutions in the spike protein (S254F, N448S, F456L, M458K, F981L, S982L) as well as in five other virus proteins.

Our findings demonstrate that the Omicron BA.1 lineage can diverge further from its already exceptionally mutated genome during persistent infection in more than one host, and also document ongoing transmission of these novel variants. There is an urgent need to implement strategies to prevent prolonged SARS-CoV-2 replication and to limit the spread of newly emerging, neutralization-resistant variants in vulnerable patients.

59 **INTRODUCTION**

60 The genomic landscape of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-
61 2) has been shaped by the emergence of antigenically diverse variants of concern (VOC)[1]. These
62 viral variants carry mutations that render them more transmissible, more fusogenic, and/or permit
63 escape not only from infection but also vaccine-induced immunity[2-4]. Some variants such as the
64 Omicron VOCs, which swept the globe starting November 2021, also display partial or complete
65 resistance to therapeutic or prophylactic monoclonal treatments[5]. Over the past ten months
66 Omicron sublineages with increasing numbers of mutations in spike have continued to emerge
67 (e.g., BA.1, BA.2, BA.4/5), posing great challenges to the containment of SARS-CoV-2 spread
68 and providing the rationale for formulation of bivalent SARS-CoV-2 booster vaccines that include
69 an Omicron BA.1 or BA.5 spike in addition to the ancestral spike.

70 Most patients with coronavirus disease 2019 (COVID-19) clear the virus upon resolution
71 of the acute infection. However, ongoing SARS-CoV-2 replication has been documented in a
72 subset of immunocompromised individuals. In these chronically infected cases, recovery of
73 infectious virus over several months and stepwise acquisition of mutations in spike has been
74 observed, pointing to positive selection[6-13]. It has been speculated that prolonged intrahost viral
75 evolution played a role in the emergence of several SARS-CoV-2 VOCs such as Alpha and
76 Omicron[14,15], but clear evidence for forward transmissions of mutated variants from chronic
77 infection cases has been lacking.

78 We describe herein a primary case of persistent SARS-CoV-2 Omicron BA.1 infection
79 marked by intrahost evolution of a variant (Omicron BA.1.23) encoding seven additional amino
80 acid substitutions in the already antigenically distinct Omicron BA.1 spike protein, and that
81 resulted in at least five further cases of Omicron BA.1.23 infection. Persistent infections
82 documented in two of these cases (one lasting four weeks, the other more than four months) were
83 associated with continued BA.1.23 evolution and acquisition of additional mutations within and
84 outside of spike. Although the variants that evolved in the persistently infected cases during the
85 cumulative eight-month period reveal unique constellations of mutations in each, they also
86 strongly point to convergent viral evolution with other co-circulating SARS-CoV-2 lineages.
87 Notably, most amino acid changes occurred at positions known to confer either immune escape[16,17]
88 or altered viral fusogenicity[18-20]. Some of these escape mutations in BA.1.23 are now signature

89 substitutions in emerging Omicron lineages such as BA.2.75.2, indicating that persistent viral
90 replication in the context of suboptimal immune responses is an important driver of SARS-CoV-
91 2 diversification.

92 **RESULTS**

93 **Emergence of a novel BA.1 sublineage through intrahost evolution.**

94 We performed genomic analysis of serially collected nasopharyngeal (NP) and anterior nares (AN)
95 samples from an immunocompromised patient (P1) with diffuse B-cell lymphoma and persistent
96 SARS-CoV-2 Omicron BA.1 replication between December 2021 and March 2022. Over a 12-
97 week period, we documented the accumulation of nine amino acid substitutions in the spike protein
98 N-terminal domain (NTD), the receptor binding domain (RBD), and in the S1/S2 furin cleavage
99 site (FCS) (**Fig. 1**) within the same patient. The first four mutations R346T, K458M, E484V, and
100 A688V were detected simultaneously 40 days after the initial SARS-CoV-2 diagnosis and were
101 fixed. Two weeks later (day 64), L167T and the FCS mutation P681Y were detected in addition
102 to the four initial mutations. During the following weeks (day 72 and day 81), two different viral
103 populations emerged each carrying shared (L455W) as well as additional distinct signature
104 mutations (E96D on day 72, S477D on day 81). All but one of the mutations observed in the
105 consensus sequences outside the spike gene were non-synonymous, providing further support for
106 positive intrahost selection of spike protein changes with competitive replication advantages.

107

108 **Forward spread confirms transmission potential of the novel BA.1.23 sublineage.**
109 Background health system-wide SARS-CoV-2 genomic surveillance conducted by the Mount
110 Sinai Pathogen Surveillance Program (MS-PSP) during the same time period identified three other
111 patients harboring SARS-CoV-2 Omicron variants that shared seven of the amino acid
112 substitutions found in the index case P1 (E96D, R346T, L455W, K458M, A484V, H681R,
113 A688V), and were consistent with multiple forward transmissions based on the timeline of
114 infection (**Fig. 2**, **Table 1**). Although viral isolation failed for specimens available from the index
115 case, we successfully cultured SARS-CoV-2 from the initial positive specimens of the subsequent
116 cases, confirming the transmission of replication-competent virus. All three forward transmissions
117 were detected in patients with underlying hematologic malignancies.

118 Lastly, two additional related SARS-CoV-2 genomes from the NYC area were identified from the
119 GISAID (Global Initiative on Sharing Avian Influenza Data) database (**Fig. 2**). Based on the
120 metadata provided, these two cases differed by age and gender from our cases, consistent with
121 independent but limited community spread of this new Omicron BA.1 sub-variant. All

122 transmission cases contained one synonymous change in Orf1a (T6001C) previously only
123 observed at day 72 in the index case (**Extended Data Fig. 1**), narrowing the time window of the
124 first transmission to a few weeks following day 64. As of August 2022, the most recent case was
125 detected in mid-April 2022 and the last SARS-CoV-2 positive specimen from a forward
126 transmission case (P3) was collected in September 2022. Taken together, the presence of a unique
127 combination of mutations in all six cases is indicative of forward transmission of this novel
128 Omicron subvariant, which received the BA.1.23 Pango lineage designation.

129

130 **Sequential persistent infections drive further intrahost evolution of BA.1.23**

131 While patient P2 cleared the BA.1.23 infection, patients P3 and P4 both developed persistent
132 infections (**Fig. 2 and 3A**). The MS-PSP continued monitoring these patients during follow-up
133 visits and hospitalizations. Patient P4 remained positive by nucleic acid amplification tests
134 (NAAT) for four weeks, with acquisition of an additional mutation in the spike RBD (S:V445A)
135 within one week of the initial positive test (**Fig. 2 and 3A**). Patient P3 developed a much longer
136 persistent infection lasting for more than four months. Genomic analysis of all 10 serially collected
137 specimens from P3 identified 14 additional amino acid substitutions throughout the viral genome
138 (**Fig. 2B and Extended Data Fig. 1**). The most recent specimen collected 131 days after P3's
139 COVID-19 onset contained 11 additional amino acid substitutions compared to the originally
140 transmitted BA.1.23 variant. These included six substitutions in the spike NTD (S254F), RBD
141 (N448S, F456L, reversion of 458M to K), and S2 (reversion of 981L to F, S982L); as well as five
142 substitutions in other SARS-CoV-2 proteins (ORF1a nsp3: T1001/183I, K1795/977Q,
143 ORF1ab_nsp11 S4398/6L; ORF1ab_nsp12: G5063/671S; and ORF7a: T39I/T24I). Notably, two
144 spike amino acid reversions to the WA1/USA (S:M458K) or BA.1 (S:F981L) state were
145 accompanied by changes at neighboring positions (S:F456L and S:S982L) (**Fig. 2B and Extended**
146 **Data Fig. 1**).

147 We further identified multiple positions with intra-host single nucleotide variants that were present
148 in only a minority of the SARS-CoV-2 viral populations, hereafter referred to as minor single
149 nucleotide variants (mSNVs), within each of the three persistent infection cases (P1, P3, and P4).
150 In some cases, we detected mSNVs within or outside the spike gene in earlier specimens, prior to
151 their fixation in subsequent specimens from the same patient (**Figure 2B and Extended Data**

152 **Figure 1,2**). We also observed numerous positions with mSNVs that persisted across multiple
153 sequential specimens without becoming dominant. This was most notable for patient P3: between
154 the first specimen collected on day 101 and the ninth specimen collected on day 218, the number
155 of positions with mSNVs at a level >10% in two or more specimens increased gradually from 0 to
156 19 – eventually outnumbering the consensus sequence changes by a factor of three (**Fig. 2B and**
157 **Extended Data Fig. 3**). Approximately half of these mutations occurred within the spike gene,
158 where nonsynonymous mSNVs were clustered in the S1 (NTD:S254F, RBD:K356T, S371F,
159 P384L, F456L, K458M, and FCS:N679K) and S2 (D936Y,V963F, N978D, L981F, S982L,
160 E1195G) domains (**Fig. B2 and Extended Data Fig. 4**). Mutations at these positions are rare at
161 the consensus level, with a maximum prevalence of 0.1% in the GISAID reference database [as of
162 2022-09-10]. The final specimen collected two weeks later, on day 232, showed eleven additional
163 substitutions (including five at positions with prior mSNVs) and two positions with mSNVs (**Fig.**
164 **2B, Extended Data Fig. 3 and Supplementary Table S1**). The progressive accumulation of
165 mutations as well as the low prevalence of the specific substitutions in co-circulating SARS-CoV-
166 2 lineages in NYC at the time, makes it unlikely that the mutated genotypes are the result of
167 recombination with other viral variants. The large shift in mutational profile may reflect a fitness
168 advantage of the dominant variant in this specimen. Thus, intra-host adaptation occurs with
169 different dynamics pointing to bottlenecks and competing selection pressures resulting in the
170 appearance and disappearance of specific mutations.

171

172 **Impact of treatment and host immunity on BA.1.23 evolution**

173 To determine the potential impact of SARS-CoV-2 antiviral treatments on intrahost evolution we
174 examined the medication histories of patients P1 (index), P2, P3 and P4 (**Fig. 3**). We also assessed
175 SARS-CoV-2 spike binding antibody levels using available serological data from clinical tests.
176 The index patient P1 was vaccinated at the time of admission with two doses of an unspecified
177 vaccine administered six months and one week before hospitalization. A third vaccine dose was
178 administered two days after admission. Additional SARS-CoV-2 antiviral treatments included a
179 course of remdesivir on days 1-6 and a dose of Gamunex IgG on day 9 (**Fig. 3D**). Moderate titers
180 of SARS-CoV-2 antibodies were detected on day 38, around the time the first intra-host mutations
181 were found (**Fig. 2C**). These titers could be due to residual antibodies from the Gamunex IgG

182 treatment, an immune response to vaccination and/or infection, or a combination of treatment and
183 host response. Notably, the detection of the first three spike mutations followed an increase in
184 NAAT cycle thresholds (Ct) on day 30, suggesting a potential selection bottleneck around this
185 time (**Fig. 2A**).

186 Patient P2 was fully vaccinated and had high titers of SARS-CoV-2 antibodies when assayed three
187 months before and one month after their single positive SARS-CoV-2 NAAT (**Fig. 3B**). This
188 patient also received a one-month course of nirmatrelvir/ritonavir (Paxlovid™) (**Fig. 3D**). The
189 combination of high levels of host antibodies and antiviral treatment likely explains why this
190 patient did not develop a persistent infection.

191 The long-term case P3 was fully vaccinated at the time of their first positive test and received a 3-
192 week course of nirmatrelvir/ritonavir (Paxlovid™). However, no SARS-CoV-2 antibodies were
193 detected in serological assays performed during the first 4 months of their persistent infection (**Fig.**
194 **3B**). Notably, during this time we also did not detect new substitutions in the BA.1.23 lineage,
195 despite indications of active viral replication based on consistently low NAAT Ct values (less than
196 25) in NP and AN samples. Active replication was confirmed by isolation of replication-competent
197 SARS-CoV-2 from five specimens collected between study days 101–147. Notably, the
198 appearance of new intra-host substitutions during the last month of persistent infection occurred
199 around the time of detection of SARS-CoV-2 antibodies at levels close to the lower limit of the
200 serological assay (**Fig. 3**). P3 did not receive antibody therapy during their hospitalization,
201 indicating that the seroconversion was the result of a delayed and weak host immune response.

202 Patient P4 was unvaccinated and received bebtelovimab mAb after their first positive SARS-CoV-
203 2 PCR test, followed by a course of nirmatrelvir/ritonavir. P4 had low SARS-CoV-2 antibody titers
204 at the time of the first positive PCR prior to the mAb treatment (**Fig. 3B**). The S:V445A mutation
205 that emerged in this patient may be associated with the mAb treatment, as it was shown to result
206 in a 111-fold reduction in susceptibility to bebtelovimab for viral variants carrying this mutation
207 in pseudotyped virus-like particle (VLP) neutralization assays[21]. We did not find SARS-CoV-2
208 mutations associated with remdesivir[22-27] or nirmatrelvir/ritonavir resistance[28] in any of the
209 patients treated with these drugs.

210

211 **Serum neutralization profile of Omicron BA.1.23**

212 Neutralization of Omicron BA.1 isolates by sera from convalescent or twice boosted individuals
213 is strongly reduced compared to the levels obtained for neutralization of the ancestral strains[29-31].
214 We therefore assessed the degree of neutralizing activity of BA.1.23 (BA.1 + S:E96D, R346T,
215 L455W, K548M, E484V, P681R, A688V) compared to Omicron BA1 (B.1.1.529), and the
216 ancestral SARS-CoV-2 (USA-WA1/2020, WA). We used a multi-cycle replication assay in which
217 serum is present continuously to best mimic physiological *in vivo* conditions. We selected sera
218 from two subsets of PARIS study participants representing distinct levels of immunity. The first
219 panel of sera tested was collected before and after booster vaccination while the second set of
220 samples was obtained before and after Omicron BA.1 breakthrough infection.

221 Prior to booster vaccination with monovalent SARS-CoV-2 RNA-based vaccines, neutralization
222 of BA.1, BA.1.23 and the WA1/USA were comparable (geometric mean titer; GMT WA1: 31;
223 GMT BA.1: 12; GMT BA.1.23: 9; **Fig. 4A**). However, 4/9 and 6/9 samples failed to display any
224 neutralization activity against BA.1 and BA.1.23 respectively while this was true for only 2/9
225 samples when looking at activity against the ancestral virus. After booster vaccination, the loss in
226 neutralization for BA.1.23 was greater than 15-fold compared to only a three-fold reduction for
227 BA.1 relative to WA1/USA (geometric mean titer; GMT WA1: 674; BA.1:234; GMT BA.1.23:
228 45; **Fig. 4A**). Of note, the loss of neutralization activity for BA.1 measured here is less than what
229 has previously been reported which could be due to assay variation in combination with the limited
230 number of samples tested[29].

231 We next looked how sera collected before and after breakthrough infection would neutralize
232 BA.1.23. We noted a significant difference in neutralization activity for BA.1.23 but not BA.1
233 compared to WA1 in the samples collected before and after BA.1 breakthrough infection (**Fig.
234 4B**). Prior to BA.1 infection, the difference was >25-fold with 3/11 samples having undetectable
235 neutralization activity for BA.1.23 and 5-fold difference for BA.1 (GMT WA1: 406; GMT BA.1:
236 83; GMT BA.1.23: 16; **Fig. 4B**). After infection, neutralization titers increased for all three viruses
237 reducing the difference to two-fold for BA.1 and 8-fold for BA.1.23 (GMT WA1: 2134; GMT
238 BA.1: 1026; GMT BA.1.23: 253; **Fig. 4B**).

239 Lastly, we analyzed sera collected from patient P2 at different time points (e.g., before booster
240 vaccination, after booster vaccination but before BA.1.23 infection, and at two time points after
241 infection with BA.1.23). P2 mounted neutralizing antibodies after booster vaccination with the

242 ancestral variant being neutralized >25 times more efficiently than BA.1 or BA.1.23 (**Fig. 4C**).
243 One month after breakthrough infection with BA.1.23, neutralization of all three viruses increased
244 relative to the titers measured after booster vaccination but to very different degrees (e.g. WA1: 6-
245 fold, BA.1:45 fold, BA.1.23: 25-fold). Of note, two months after infection, the BA.1 titers
246 remained stable, while neutralization of the ancestral virus as well as BA.1.23 declined equally.

247 Collectively these results indicate that BA.1.23 is significantly more resistant to neutralization than
248 the parental BA.1 variant.

249

250 **DISCUSSION**

251 It has been speculated that the emergence of antigenically diverse SARS-CoV-2 variants such as
252 Omicron result from intrahost viral evolution driven initially by suboptimal immune responses and
253 then spread by forward transmissions. Our data demonstrate that intrahost evolution of SARS-
254 CoV-2 during persistent infection in immunocompromised individuals can drive the emergence
255 and spread of novel (sub)variants with extensive mutations in key antigenic regions, even in the
256 context of the already highly mutated Omicron lineage. Further sequential evolution in
257 transmission cases shows that rapid divergence can occur stepwise and persist in small numbers
258 of individuals, complicating early detection of novel variants.

259 The most extensive changes in the evolution of the BA.1.23 lineage were seen during its
260 emergence in patient P1 and its subsequent divergence in patient P3 over a total period of 8 months.
261 Based on available clinical testing data, both patients were initially negative for SARS-CoV-2
262 antibodies but seroconverted to moderate to low levels around the time spike mutations were
263 accumulating. These conditions are likely optimal for the selection of immune escape mutations
264 and provide an explanation for the clustering of BA.1.23-specific changes in the spike protein. It
265 is important to note that the antiviral treatments administered over limited time periods as
266 monotherapies (e.g., monoclonal antibodies, remdesivir, and nirmatrelvir/ritonavir (Paxlovid™)
267 failed to eliminate the persistent infection, highlighting the need for improved or potential
268 combination therapy tailored for these special situations.

269 During the initial emergence of the BA.1.23 lineage in index patient P1 mutations were
270 concentrated in the spike protein. The amino acid substitutions or positions at which changes
271 accumulated have been associated with neutralization escape (e.g., R346T, E484V, S477)[16,17],
272 increased angiotensin converting enzyme 2 (ACE2) binding avidity (e.g., R346T, L455W)[32],
273 improved viral fusogenicity (e.g., P681R/Y)[18], or a predicted increase in fitness (e.g., positions
274 346, 484 and 681)[33]. Subsequent stepwise changes in patient P3 occurred more broadly throughout
275 the SARS-CoV-2 genome. The RBD mutations N448S and F456L present in the most recent
276 specimen from P3 are within the ACE2 binding region with potential for changes in binding
277 affinity and antibody escape[32]. Mutations outside the spike mapped to nonstructural proteins, many
278 of which have previously been detected in other VOCs such as Nsp3_T183I (Alpha),

279 Nsp3_K977Q (Gamma), Nsp12_G671S (Delta AY.* and Omicron BA.2.75). Although the effects
280 of these mutations have not been characterized, they occurred in the virus polymerase subunit
281 (Nsp12) and in proteins with antagonistic activity towards the host's innate immunity (Nsp3,
282 ORF6, ORF7a, ORF9b)[34].

283 Consistent with the extensive changes in the spike protein, resistance of the transmitted BA.1.23
284 variant to neutralization was substantially increased compared to the ancestral SARS-CoV-2 strain
285 as well as to BA.1 Omicron. This was true both for sera collected before and after monovalent
286 booster vaccination as well as after BA.1 breakthrough infection (**Figure 4**). Escape from
287 neutralization was likely mediated by the combination of R346T, L455W, K458M, and E/A484V
288 within the receptor binding region of spike. Of note, substitutions at position R346 are found in
289 several Omicron sublineages (e.g., BA.4, BA.5) and subvariants (e.g., BA.2.72.2, BA.4.6/BF.7
290 (R346T), BA.4.7 (R346S), and BA.5.9 (R346I))[35]. These variants were not circulating at the time
291 of BA.1.23 emergence pointing to convergent evolution in the context of suboptimal immune
292 responses.

293 We further found that changes in minority viral populations can greatly increase the spectrum of
294 mutations during persistent infection beyond what is observed at the consensus level. Viral 'quasi-
295 species' have been described for pandemic coronaviruses such as SARS-CoV-1, Middle Eastern
296 Respiratory Syndrome Coronavirus (MERS-CoV) and SARS-CoV-2[36-42], but their role remains to
297 be fully elucidated[43]. Interestingly, mSNVs predominantly occurred in conserved regions and
298 remained present in multiple sequential specimens. Only a small subset of these mutations became
299 dominant in later specimens. We speculate that these viral populations enable sub-optimal
300 mutations to persist within the host, increasing the likelihood for the emergence of compensatory
301 mutations that offset their fitness defects.

302 Despite careful monitoring, we did not encounter new BA.1.23 cases in our routine SARS-CoV-2
303 surveillance or in the GISAID database between June-September 2022, suggesting that the lineage
304 has not spread further beyond the initial outbreak. This may be due to increased caution by
305 immunocompromised patients to avoid contacts that could increase their risk of infection. We also
306 note that patients P1, P3 and P4 were hospitalized for long stretches of time during their persistent
307 infections, further limiting community exposure. Finally, the limited transmission could reflect a
308 reduced fitness of BA.1.23 compared to contemporary lineages. Notable in this respect is that the

309 emergence of the BA.1.23 lineage coincided with the general displacement of BA.1 by BA.2
310 sublineages in the NYC area, in particular BA.2.12.1, which dominated SARS-CoV-2 cases during
311 the spring and early summer of 2022. A longer observation period with continued background
312 surveillance will be needed to rapidly identify any potential re-emergence of BA.1.23 in patients
313 other than the ones included in this study.

314 Our findings add to the accumulating knowledge regarding SARS-CoV-2 persistent infections,
315 and document subsequent forward transmissions. Furthermore, we show that persistent infections
316 can drive emergence of viral variants with the potential to spread to other individuals, some of
317 whom may themselves develop prolonged infections, and thereby establish a pathway for
318 continued virus evolution. This warrants genomic monitoring of persistent infections in particular,
319 and further underscores the need to limit the duration of viral infection. Improved early detection
320 of novel SARS-CoV-2 variants and forward transmission tracking, a better understanding of the
321 selection processes driving SARS-CoV-2 evolution and the role of mSNVs, as well as therapies
322 that limit virus persistence and shedding, are essential to reduce the emergence of highly mutated
323 viral variants in the future.

324

325 **MATERIALS AND METHODS**

326 *Molecular SARS-CoV-2 diagnostics*

327 SARS-CoV-2 molecular diagnostic testing was performed in the Molecular Microbiology
328 Laboratories of the MSHS Clinical Laboratory by nucleic acid amplification tests (NAAT) that
329 have been validated for nasopharyngeal (NP), anterior nares (AN) swabs, and saliva specimens.
330 All but one positive sample included in this study were tested using the PerkinElmer® New
331 Coronavirus Nucleic Acid Detection Kit which provides qualitative detection of nucleic acid from
332 SARS-CoV-2. It includes two SARS-CoV-2 targets (ORF1ab, N) and one internal positive control
333 (IC; bacteriophage MS2). Cycle threshold (Ct) values are generated for all three targets and
334 provide a quantitative estimate of the viral load. The only other positive sample that was run on
335 another testing platform was the first specimen of the index case (P1). It was tested as a point of
336 care test (POC) using the cobas® Liat® System (cobas® SARS-CoV-2 & Influenza A/B) assay, and
337 the biospecimen was discarded prior to transfer to the MS-PSP.

338

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

340 The Mount Sinai Pathogen Surveillance Program (MS-PSP) has profiled the evolving landscape
341 of SARS-CoV-2 in New York City (NYC) since the beginning of the pandemic[44,45]. We routinely
342 perform complete viral genome sequencing of randomly selected contemporaneous SARS-CoV-2
343 positive specimens collected from patients seeking care at our health system. Residual respiratory
344 specimens (NP, AN swabs) were collected after completion of the diagnostic process, as part of
345 the Mount Sinai Pathogen Surveillance Program. The program was reviewed and approved by the
346 Mount Sinai Hospital (MSH) Institutional Review Board (IRB-13-00981). The detailed
347 investigation into persistent SARS-CoV-2 infections in patients receiving care at the Mount Sinai
348 Health System was separately reviewed and approved by the MSH IRB (IRB-22-00760).

349

350 *Human Serum Samples*

351 We used a panel of sera collected as part of the longitudinal observational PARIS (Protection
352 Associated with Rapid Immunity to SARS-CoV-2) cohort[46]. This study follows health care
353 workers longitudinally since April 2020. Samples were selected based on the study participants

354 different levels of immunity (before and after booster vaccination, as well as before and after
355 Omicron BA.1 infection). The study was reviewed and approved by the MSH IRB (IRB-20-
356 03374). All participants signed written consent forms prior to sample and data collection and
357 provided permission for sample banking and sharing. Sera collected before and after SARS-CoV-
358 2 booster vaccination (N=9 participants, 18 sera) as well as before and after Omicron BA.1 break-
359 through infections (N=11 participants, 22 sera) were selected for testing in this study (see
360 **Supplementary Table S3** for infection, vaccine and demographics information).

361 In addition, patient P2 participated in another of our observational study (IRB-16-01215). They
362 provided written consent prior to biospecimen and data collection. Several serum samples from
363 this patient before and after BA.1.23 infection were available for neutralization assays.

364 The 44 sera (40 PARIS, 4 from P2) from our observational cohorts are unique to this study and
365 are not publicly available.

366

367 *RNA extraction and SARS-CoV-2 genome sequencing*

368 RNA was extracted using the Chemagic ™ Viral DNA/RNA 300 Kit H96 (PerkinElmer, CMG-
369 1033-S) on the automated Chemagic ™ 360 instrument (PerkinElmer, 2024-0020) from 300uL of
370 viral transport media per the manufacturer's protocol. cDNA synthesis followed by whole genome
371 amplification with two custom primer panel sets targeting 1.5 and 2kb regions across the SARS-
372 CoV-2 genome was performed as previously described[44] with the addition of new oligonucleotides
373 to minimize amplicon dropout for Omicron lineages derived from PANGO lineage B.1.1.529
374 (**Supplementary Table S4**). Paired-end (2x150bp) Nextera XT (Illumina, cat. FC-131-1096)
375 libraries prepared from amplicons were sequenced on a MiSeq instrument. SARS-CoV-2 genomes
376 were assembled using a custom reference-based pipeline as previously described[44]. The final
377 average sequencing depth per genome ranged between ~135k to ~415k reads.

378

379 *Analysis of minor nucleotide variants*

380 Minority intra-host single nucleotide variants (mSNVs) were annotated when present in both reads
381 of paired-end reads of a single sample and at a minimum frequency of 0.1 (10%). A second quality
382 control filter was applied by only including positions for which mSNVs were present in more than

383 one sample from the study (from different time pointes) or for positions that changed at the
384 consensus level at any point of the investigation.

385

386 *Phylogenetic analysis*

387 Global background SARS-CoV-2 sequences were downloaded from the GISAID EpiCOV
388 database (as of 2022-08-30). The GISAID database was queried for the novel mutations observed
389 in P1-P4 sequences to identify their closest matches. Sequence hits were confirmed by their Mash
390 distance[47]. For this, a genome sketch was generated with Mash v.2.3 from the sanitized alignment
391 of the global sequences filtered for lineages BA.1.* as produced by Nextstrain[1,48] v11 for SARS-
392 CoV-2 with default parameters (https://github.com/nextstrain/ncov). This allowed pairwise
393 comparisons of our data with high-quality global sequences with available metadata. Maximum
394 likelihood phylogenetic inferences of the MS-PSP SARS-CoV-2 genomes, including the closest
395 sequence matches and all other BA.* lineage sequences from the MS-PSP surveillance program,
396 were done using an Omicron BA.* and New York State-focused background with proximity
397 subsampling scheme with Nextstrain under the GTR+G model of nucleotide substitution. Branch
398 support was assessed with the ultrafast bootstrap method with 1000 replicates in IQ-TREE
399 multicore version 2.1.2 [PMID: 29077904, 25371430]. Clade and lineage assignments were done
400 with Nextclade CLI v3.2 (2021-11-04)[49] and with PANGO-v1.8 (pangolin v3.1.17, pangoLEARN
401 v.2022-04-22) [50,51].

402

403 *SARS-CoV-2 viral cultures*

404 Replication-competent SARS-CoV-2 was obtained as previously described[29]. Briefly, Vero-E6
405 cells expressing transmembrane serine protease 2 (TMPRSS2) (BPS Biosciences, catalog (cat.)
406 no. 78081) were cultured in Dulbecco's modified Eagle medium containing 10% heat-inactivated
407 fetal bovine serum and 1% minimum essential medium (MEM) Amino Acids Solution,
408 supplemented with 100 U/ml penicillin, 100 µg/ml streptomycin, 100 µg/ml normocin, and 3 µg/ml
409 puromycin. 200ul of viral transport media from the nasopharyngeal or anterior nares swab
410 specimen was added to Vero-E6-TMPRSS2 cells in culture media supplemented with 2% heat-
411 inactivated fetal bovine serum, 100g/ml normocin, and 0.5 µg/ml amphotericin B. Cultures were

412 maintained for a maximum of 10 days. Culture supernatants were collected and clarified by
413 centrifugation (3,739g for 5 min) upon the appearance of cytopathic effects. Viral cultures were
414 successful for the initial specimens obtained from P2, P3, and P4, but failed for all P1 specimens.

415

416 *Titration and concentration of viral isolates*

417 Viral stocks were sequence-verified and then titered using the 50% tissue culture infectious dose
418 ($TCID_{50}$) method on Vero-E6-TMPRSS2 cells. Based on sequencing verification, viral isolate
419 from P3 (PV56567) was selected for further experiments. Since the initial viral titer of the
420 expanded BA.1. 23 PV56567 viral stock was too low (4x10^2 $TCID_{50}$/ml) to perform
421 microneutralization assays, we concentrated the viral isolate 4x using Pierce™ Protein
422 Concentrator PES, 100K MWCO (Thermo Scientific, Catalog number: 88537) columns as
423 described by the manufacturer. Briefly, culture supernatant was collected upon appearance of a
424 cytopathic effect (CPE), clarified by centrifugation (3,739g, 5min) and then concentrated. Titers
425 after concentration were 1.27x10^5 $TCID_{50}$/ml.

426

427 *Micro-neutralization assays with replication-competent SARS-CoV-2 isolates*

428 Sera collected from three different groups of study participants were used to assess the
429 neutralization of wild type (WA1), BA.1, and BA.1.23 SARS-CoV-2 isolates. The first panel of
430 samples includes sera collected before and after mRNA SARS-CoV-2 booster vaccination (N: 9,
431 PARIS cohort), the second panel includes sera collected before and after BA.1 Omicron break-
432 through infection (N:11, PARIS cohort). The last series includes longitudinal samples from P2,
433 the first of the forward transmission case, collected prior as well as after the infection with
434 BA.1.23.

435 Sera from study participants were serially diluted (three-fold) from a starting dilution of 1:10 using
436 infection media consisting of minimum essential media (MEM, Gibco) supplemented with 2 mM
437 L-glutamine, 0.1% sodium bicarbonate (w/v, Gibco), 10 mM 4-(2-hydroxyethyl)-1-
438 piperazineethanesulfonic acid (HEPES, Gibco), 100 U/ml penicillin, 100 µg/ml streptomycin
439 (Gibco) and 0.2% bovine serum albumin (MP Biomedicals). Diluted sera were incubated for one
440 hour with 10,000 $TCID_{50}$ of the three different viruses at room temperature. After the incubation,

441 120 µl of the serum-virus mix were transferred to Vero-E6-TMPRSS2 (plated in 96 well plates the
442 prior day) and incubated at 37°C, 5% CO_2 for one hour. The serum-virus mix was removed and
443 100 µl/well of the diluted sera were added in addition to 100 µl/well of MEM 2% FBS. The plates
444 were incubated for 48 hours at 37°C incubation. The cells were fixed (200 µl/well, 10%
445 formaldehyde solution) overnight at 4°C prior to staining an anti-SARS nucleoprotein antibody
446 produced in house (mAb 1C7C7)[52]. All procedures above were performed in the Biosafety Level
447 3 (BSL-3) facility at the Icahn School of Medicine at Mount Sinai following approved standard
448 safety guidelines. The nucleoprotein staining was performed as previously described[29]. The
449 analyses were performed using Prism 9 software (GraphPad).

450

451 *Statistics*

452 Statistical analyses were performed using Prism 9 software (GraphPad). A one-way ANOVA with
453 Tukey's multiple comparisons test was used to compare the neutralization titers for the three
454 viruses.

ACKNOWLEDGMENTS

We thank the Mount Sinai Hospital and School Leadership, in particular Dr. David Reich, Dr. Dennis Charney, and Dr. Carlos Cordon-Cardo, for their ongoing support of the MS-PSP throughout the pandemic. We thank the laboratory technologists and staff in the Molecular Microbiology 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* and by an Option to the Collaborative Influenza Vaccine Innovation Centers (CIVIC) contract 75N93019C00051 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, Option 20) awarded to the *Center for Research on Influenza Pathogenesis*, and awards (S10OD026880 and S10OD030463) from the *NIH Office of Research Infrastructure Programs*. The Mount Sinai Pathogen Surveillance Program is supported in part by Institutional funds from the Icahn School of Medicine as well as the Mount Sinai Hospital.

AUTHOR CONTRIBUTION STATEMENT

Conceptualization: A.S.G.-R., H.A., E.M.S., V.S., H.v.B. Sample acquisition: S.S, G.P, J.P., A.A., A.R., C.C., D.F., L.S., C.G., K.S., J.D.R., R.B., P.S, A.P-M, PSP-PARIS Study Group. Sequencing and genome assembly: A.vd.G, K.F., Z.K., J.Z., B.A., R.S. Methodology: A.S.G.-R., H.A., N.A., J.M.C, F.K. Investigation: A.S.G.-R., H.A., J.M.C., E.M.S., V.S., H.v.B. Visualization: A.S.G.-R., H.A., V.S., H.v.B. Funding Acquisition: E.M.S., V.S., H.v.B. Project administration: K.S., E.M.S., V.S., H.v.B. Supervision: E.M.S., V.S., H.v.B. Writing – First draft: A.S.G.-R., H.A., H.v.B. Writing – Review and editing: A.S.G.-R., H.A., J.M.C, F.K. E.M.S., V.S., H.v.B.

484

485 ## COMPETING INTERESTS

486 The Icahn School of Medicine at Mount Sinai has filed patent applications relating to SARS-CoV-
487 2 serological assays (U.S. Provisional Application Numbers: 62/994,252, 63/018,457, 63/020,503
488 and 63/024,436) and NDV-based SARS-CoV-2 vaccines (U.S. Provisional Application Number:
489 63/251,020) which list Florian Krammer as co-inventor. Viviana Simon is also listed on the
490 serological assay patent application as co-inventor. Patent applications were submitted by the
491 Icahn School of Medicine at Mount Sinai. Mount Sinai has spun out a company, Kantaro, to market
492 serological tests for SARS-CoV-2. Florian Krammer has consulted for Merck and Pfizer (before
493 2020), and is currently consulting for Pfizer, Third Rock Ventures, Seqirus and Avimex. The
494 Krammer laboratory is also collaborating with Pfizer on animal models of SARS-CoV-2. Robert
495 Sebra is VP of Technology Development and a stockholder at Sema4, a Mount Sinai Venture. This
496 work, however, was conducted solely at Icahn School of Medicine at Mount Sinai.

497

498 ## ADDITIONAL INFORMATION STATEMENT

499 Correspondence and requests for materials should be addressed to Drs. Sordillo, Simon and/or van
500 Bakel.

501

502 ## DATA AND CODE AVAILABILITY STATEMENT

503 Custom code used in this study is available at https://github.com/mjsull/COVID_pipe. Complete
504 genome sequences for the viral isolates cultured from nasal swabs (BA.1 and BA.1.23) are
505 available in GISAID.

506 **REFERENCES**

507 1 Nextstrain build for novel coronavirus SARS-CoV-2 (2020).

508 2 DeGrace, M. M. *et al.* Defining the risk of SARS-CoV-2 variants on immune protection. *Nature*
509 **605**, 640-652, doi:10.1038/s41586-022-04690-5 (2022).

510 3 Flemming, A. Omicron, the great escape artist. *Nat Rev Immunol* **22**, 75, doi:10.1038/s41577-022-
511 00676-6 (2022).

512 4 Plante, J. A. *et al.* The variant gambit: COVID-19's next move. *Cell Host Microbe* **29**, 508-515,
513 doi:10.1016/j.chom.2021.02.020 (2021).

514 5 VanBlargan, L. A. *et al.* An infectious SARS-CoV-2 B.1.1.529 Omicron virus escapes
515 neutralization by therapeutic monoclonal antibodies. *Nat Med* **28**, 490-495, doi:10.1038/s41591-
516 021-01678-y (2022).

517 6 Corey, L. *et al.* SARS-CoV-2 Variants in Patients with Immunosuppression. *N Engl J Med* **385**,
518 562-566, doi:10.1056/NEJMsb2104756 (2021).

519 7 Avanzato, V. A. *et al.* Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an
520 Asymptomatic Immunocompromised Individual with Cancer. *Cell* **183**, 1901-1912 e1909,
521 doi:10.1016/j.cell.2020.10.049 (2020).

522 8 Choi, B. *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N Engl*
523 *J Med* **383**, 2291-2293, doi:10.1056/NEJMc2031364 (2020).

524 9 Kemp, S. A. *et al.* SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **592**, 277-
525 282, doi:10.1038/s41586-021-03291-y (2021).

526 10 Truong, T. T. *et al.* Increased viral variants in children and young adults with impaired humoral
527 immunity and persistent SARS-CoV-2 infection: A consecutive case series. *EBioMedicine* **67**,
528 103355, doi:10.1016/j.ebiom.2021.103355 (2021).

529 11 Scherer, E. M. *et al.* SARS-CoV-2 Evolution and Immune Escape in Immunocompromised
530 Patients. *N Engl J Med* **386**, 2436-2438, doi:10.1056/NEJMc2202861 (2022).

531 12 Cele, S. *et al.* SARS-CoV-2 prolonged infection during advanced HIV disease evolves extensive
532 immune escape. *Cell Host Microbe* **30**, 154-162 e155, doi:10.1016/j.chom.2022.01.005 (2022).

533 13 Hoffman, S. A. *et al.* SARS-CoV-2 Neutralization Resistance Mutations in Patient with HIV/AIDS,
534 California, USA. *Emerg Infect Dis* **27**, 2720-2723, doi:10.3201/eid2710.211461 (2021).

535 14 Martin, D. P. *et al.* Selection analysis identifies unusual clustered mutational changes in Omicron
536 lineage BA.1 that likely impact Spike function. *bioRxiv*, 2022.2001.2014.476382,
537 doi:10.1101/2022.01.14.476382 (2022).

538 15 (CoG-UK), C.-G. C. U. *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage*
539 *in the UK defined by a novel set of spike mutations*, <https://virological.org/t/preliminary-genomic-

540 characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-
541 mutations/563> (2020).
542 16 Greaney, A. J. *et al.* Mapping mutations to the SARS-CoV-2 RBD that escape binding by different
543 classes of antibodies. *Nat Commun* **12**, 4196, doi:10.1038/s41467-021-24435-8 (2021).
544 17 Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals
545 Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310 e1220,
546 doi:10.1016/j.cell.2020.08.012 (2020).
547 18 Escalera, A. *et al.* Mutations in SARS-CoV-2 variants of concern link to increased spike cleavage
548 and virus transmission. *Cell Host Microbe*, doi:10.1016/j.chom.2022.01.006 (2022).
549 19 Saito, A. *et al.* Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation.
550 *Nature* **602**, 300-306, doi:10.1038/s41586-021-04266-9 (2022).
551 20 Meng, B. *et al.* Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and
552 fusogenicity. *Nature* **603**, 706-714, doi:10.1038/s41586-022-04474-x (2022).
553 21 *Fact sheet for healtcare providers: Emergency use authorization for bebtelovimab.*, <
554 https://pi.lilly.com/eua/bebtelovimab-eua-factsheet-hcp.pdf#page=12> (2022).
555 22 Szemiel, A. M. *et al.* In vitro selection of Remdesivir resistance suggests evolutionary predictability
556 of SARS-CoV-2. *PLoS Pathog* **17**, e1009929, doi:10.1371/journal.ppat.1009929 (2021).
557 23 Gandhi, S. *et al.* De novo emergence of a remdesivir resistance mutation during treatment of
558 persistent SARS-CoV-2 infection in an immunocompromised patient: a case report. *Nat Commun*
559 **13**, 1547, doi:10.1038/s41467-022-29104-y (2022).
560 24 Checkmahomed, L. *et al.* In Vitro Selection of Remdesivir-Resistant SARS-CoV-2 Demonstrates
561 High Barrier to Resistance. *Antimicrob Agents Chemother* **66**, e0019822, doi:10.1128/aac.00198-
562 22 (2022).
563 25 Agostini, M. L. *et al.* Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is
564 Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. *mBio* **9**,
565 doi:10.1128/mBio.00221-18 (2018).
566 26 Martinot, M. *et al.* Emerging RNA-Dependent RNA Polymerase Mutation in a Remdesivir-Treated
567 B-cell Immunodeficient Patient With Protracted Coronavirus Disease 2019. *Clin Infect Dis* **73**,
568 e1762-e1765, doi:10.1093/cid/ciaa1474 (2021).
569 27 Tchesnokov, E. P. *et al.* Template-dependent inhibition of coronavirus RNA-dependent RNA
570 polymerase by remdesivir reveals a second mechanism of action. *J Biol Chem* **295**, 16156-16165,
571 doi:10.1074/jbc.AC120.015720 (2020).
572 28 Zhou, Y. *et al.* Nirmatrelvir Resistant SARS-CoV-2 Variants with High Fitness in Vitro. *bioRxiv*,
573 2022.2006.2006.494921, doi:10.1101/2022.06.06.494921 (2022).

29 Carreno, J. M. *et al.* Activity of convalescent and vaccine serum against SARS-CoV-2 Omicron. *Nature* **602**, 682-688, doi:10.1038/s41586-022-04399-5 (2022).

30 Cele, S. *et al.* Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature* **602**, 654-656, doi:10.1038/s41586-021-04387-1 (2022).

31 Dejnirattisai, W. *et al.* SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. *Cell* **185**, 467-484 e415, doi:10.1016/j.cell.2021.12.046 (2022).

32 *Omicron Receptor Binding Domain Deep Mutational Scanning Data*, <https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS_Omicron/> (

33 Obermeyer, F. *et al.* Analysis of 6.4 million SARS-CoV-2 genomes identifies mutations associated with fitness. *Science* **376**, 1327-1332, doi:10.1126/science.abm1208 (2022).

34 Thorne, L. G. *et al.* Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature* **602**, 487-495, doi:10.1038/s41586-021-04352-y (2022).

35 Jian, F. *et al.* Further humoral immunity evasion of emerging SARS-CoV-2 BA.4 and BA.5 subvariants. *bioRxiv*, 2022.2008.2009.503384, doi:10.1101/2022.08.09.503384 (2022).

36 Jary, A. *et al.* Evolution of viral quasispecies during SARS-CoV-2 infection. *Clin Microbiol Infect* **26**, 1560 e1561-1560 e1564, doi:10.1016/j.cmi.2020.07.032 (2020).

37 Xu, D., Zhang, Z. & Wang, F. S. SARS-associated coronavirus quasispecies in individual patients. *N Engl J Med* **350**, 1366-1367, doi:10.1056/NEJMc032421 (2004).

38 Karamitros, T. *et al.* SARS-CoV-2 exhibits intra-host genomic plasticity and low-frequency polymorphic quasispecies. *J Clin Virol* **131**, 104585, doi:10.1016/j.jcv.2020.104585 (2020).

39 Briese, T. *et al.* Middle East respiratory syndrome coronavirus quasispecies that include homologues of human isolates revealed through whole-genome analysis and virus cultured from dromedary camels in Saudi Arabia. *mBio* **5**, e01146-01114, doi:10.1128/mBio.01146-14 (2014).

40 Tonkin-Hill, G. *et al.* Patterns of within-host genetic diversity in SARS-CoV-2. *Elife* **10**, doi:10.7554/eLife.66857 (2021).

41 Valesano, A. L. *et al.* Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. *PLoS Pathog* **17**, e1009499, doi:10.1371/journal.ppat.1009499 (2021).

42 Wang, Y. *et al.* Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients. *Genome Med* **13**, 30, doi:10.1186/s13073-021-00847-5 (2021).

43 Sun, F. *et al.* SARS-CoV-2 Quasispecies Provides an Advantage Mutation Pool for the Epidemic Variants. *Microbiol Spectr* **9**, e0026121, doi:10.1128/Spectrum.00261-21 (2021).

44 Gonzalez-Reiche, A. S. *et al.* Introductions and early spread of SARS-CoV-2 in the New York City area. *Science*, doi:10.1126/science.abc1917 (2020).

607 45 Hernandez, M. M. *et al.* Molecular evidence of SARS-CoV-2 in New York before the first
608 pandemic wave. *Nat Commun* **12**, 3463, doi:10.1038/s41467-021-23688-7 (2021).
609 46 Simon, V. *et al.* PARIS and SPARTA: Finding the Achilles' Heel of SARS-CoV-2. *mSphere* **7**,
610 e0017922, doi:10.1128/msphere.00179-22 (2022).
611 47 Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash.
612 *Genome Biol* **17**, 132, doi:10.1186/s13059-016-0997-x (2016).
613 48 Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121-
614 4123, doi:10.1093/bioinformatics/bty407 (2018).
615 49 Aksamentov, I. R., Cornelius; Hodcroft, Emma B.; Neher, Richard A. Nextclade: clade assignment,
616 mutation calling and quality control for viral genomes. *Journal of Open Source Software* **6**, 3773,
617 doi:10.21105/joss.03773 (2021).
618 50 Phylogenetic assignment of named global outbreak lineages v. 2020-05-19 (2020).
619 51 O'Toole, A. *et al.* Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and
620 B.1.351/501Y-V2 with grinch. *Wellcome Open Res* **6**, 121,
621 doi:10.12688/wellcomeopenres.16661.2 (2021).
622 52 Amanat, F. *et al.* An In Vitro Microneutralization Assay for SARS-CoV-2 Serology and Drug
623 Screening. *Curr Protoc Microbiol* **58**, e108, doi:10.1002/cpmc.108 (2020).
624

625 **TABLES**

626 **Table 1: Spike mutations detected in the sequenced cases.**

Pat	Sex	Age	Days from P1 first positive specimen	GISAID Accession number (EPI_ISL)	Source	Lineage	Mutations in Spike relative to earliest sample (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) #																	
							E96	L176	S254	R346	**S371**	V445	N448	L455	F456	K458	S477	**E484**	T547	S680	**P681**	**A688**	**L981**	S982
			0	N/A	This study																			
			23	11897513	This study	BA.1					L							A			H	V	F	
			40	11897516	This study	BA.1				T	L					M		**V**			H	V	F	
P1	M	61-65	48	11178997	This study	BA.1.23				T	L					M		**V**			H	V	F	
			64	11179080	This study	BA.1.23		F		T	L					M		**V**			**Y**	V	F	
			72	11685451	This study	BA.1.23	D			T	L			W		M		**V**			H	V	F	
			81	11897525	This study	BA.1.23+		F		T	L			W		M	D	**V**			**Y**	V	F	
P2	F	66-70	76	11685480	This study	BA.1.23	D			T	L			W		M		**V**		A	H	V	F	
			101	12422465	This study	BA.1.23	D			T	L			W		M		**V**			**R**	V	F	
			111	13350580	This study	BA.1.23	D			T	L			W		M		**V**			**R**	V	F	
			114	12711034	This study	BA.1.23	D			T	L			W		M		**V**			**R**	V	F	
P3	M	36-40	130	13350591	This study	BA.1.23+	D			T	**F**			W		M		**V**			**R**	V	F	
			147	n.s.*	This study	BA.1.23	D			T	L			W		M		**V**			**R**	V	F	
			192	n.s.*	This study	BA.1.23+	D			T	**F**			W		M		**V**			**R**	V	F	
			209	n.s.*	This study	BA.1.23+	D			T	**F**			W		M		**V**			**R**	V	F	

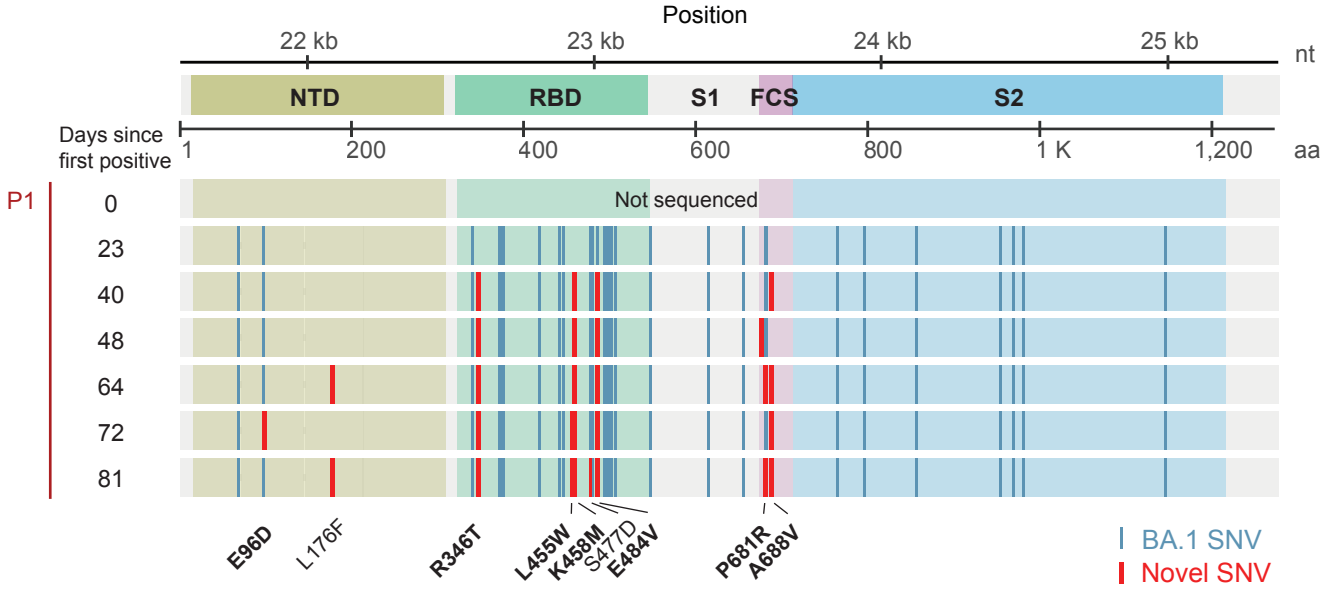
			213	n.s.*	This study	BA.1.23	D			T	L			W		M	**V**			**R**	V	F	
			218	n.s.*	This study	BA.1.23+	D			T	**F**			W		M	**V**			**R**	V	F	
			232	n.s.*	This study	BA.1.23+	D		F	T	**F**		S	W	L		**V**			**R**	V		**L**
			104	n.s.*	This study	BA.1.23+	D			T	**F**			W		M	**V**	K		**R**	V	F	
P4	F	61-65	111	12961673	This study	BA.1.23	D			T	L			W		M	**V**			**R**	V	F	
			125	12896972	This study	BA.1.23+	D			T	L	A		W		M	**V**			**R**	V	F	
			132	n.s.*	This study	BA.1.23+	D			T	L			W	L	M	**V**			**R**	V	F	
S1	F	81-85	78	11628400	Howard et al. 2022. GISAID	BA.1.23	D			T	L			W		M	**V**			**R**	V		
S2	F	26-30	79	11696379	Howard et al. 2022. GISAID	BA.1.23	D			T	L			W		M	**V**			**R**	V		

* n.s.: Consensus not submitted due to presence of >10 intrahost single nucleotide variants. N/A: not sequenced .

Mutations in bold correspond to changes at positions already mutated in BA.1.

FIGURES

Figure 1

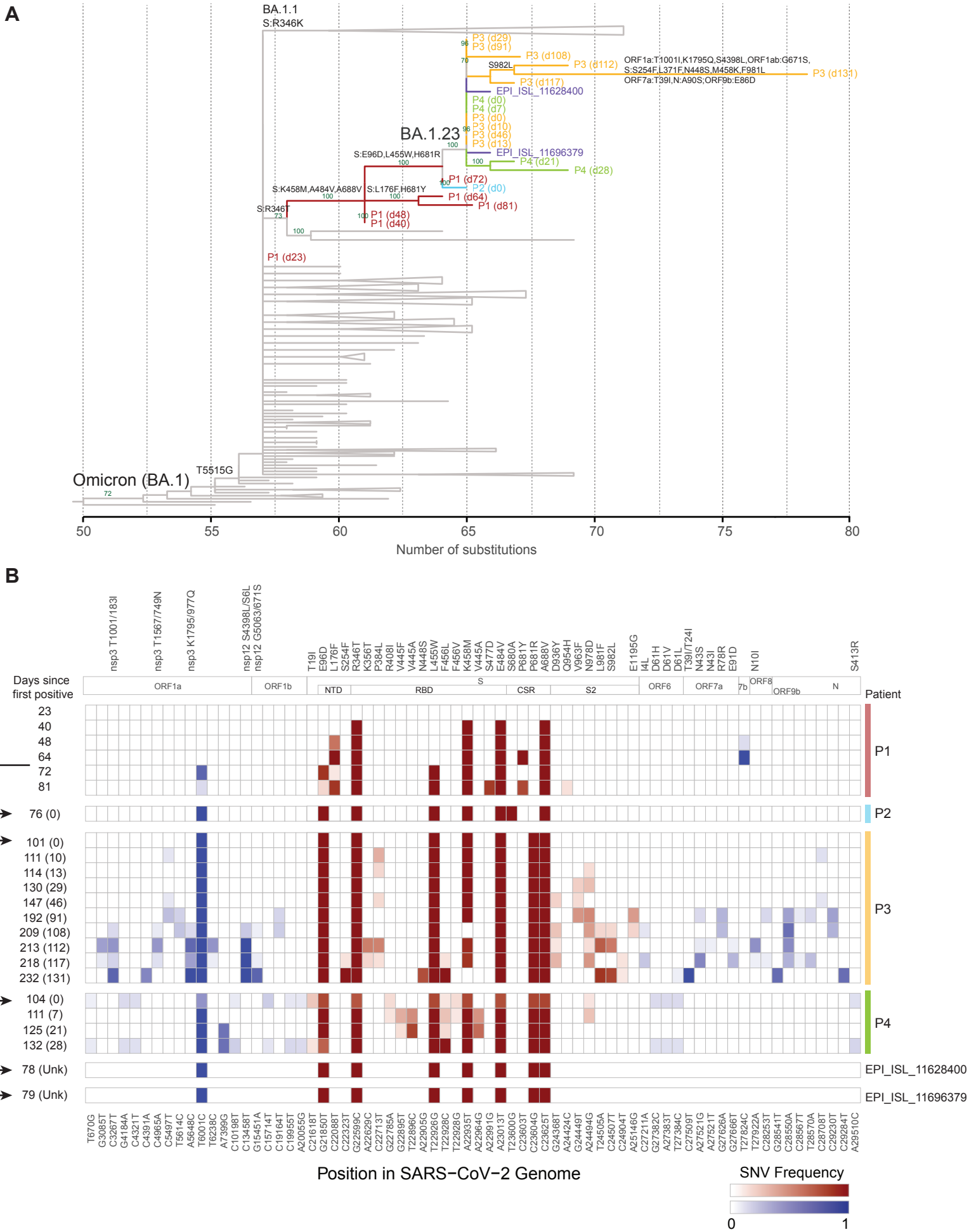

Position
22 kb
23 kb
24 kb
25 kb
nt
NTD
RBD
S1
FCS
S2
Days since first positive
1
200
400
600
800
1 K
1,200
aa
P1
0
23
40
48
64
72
81
Not sequenced
E96D
L176F
R346T
L455W
K458M
S477D
E484V
P681R
A688V
BA.1 SNV
Novel SNV

632 **Figure 1. Intrahost emergence of an Omicron BA.1.23 subvariant**

633 Multiple sequence alignment of the SARS-CoV-2 spike protein showing the appearance of
634 nucleotide and amino acid substitutions in the spike protein in sequential specimens obtained from
635 the index case (Patient 1, P1).

636

Figure 2

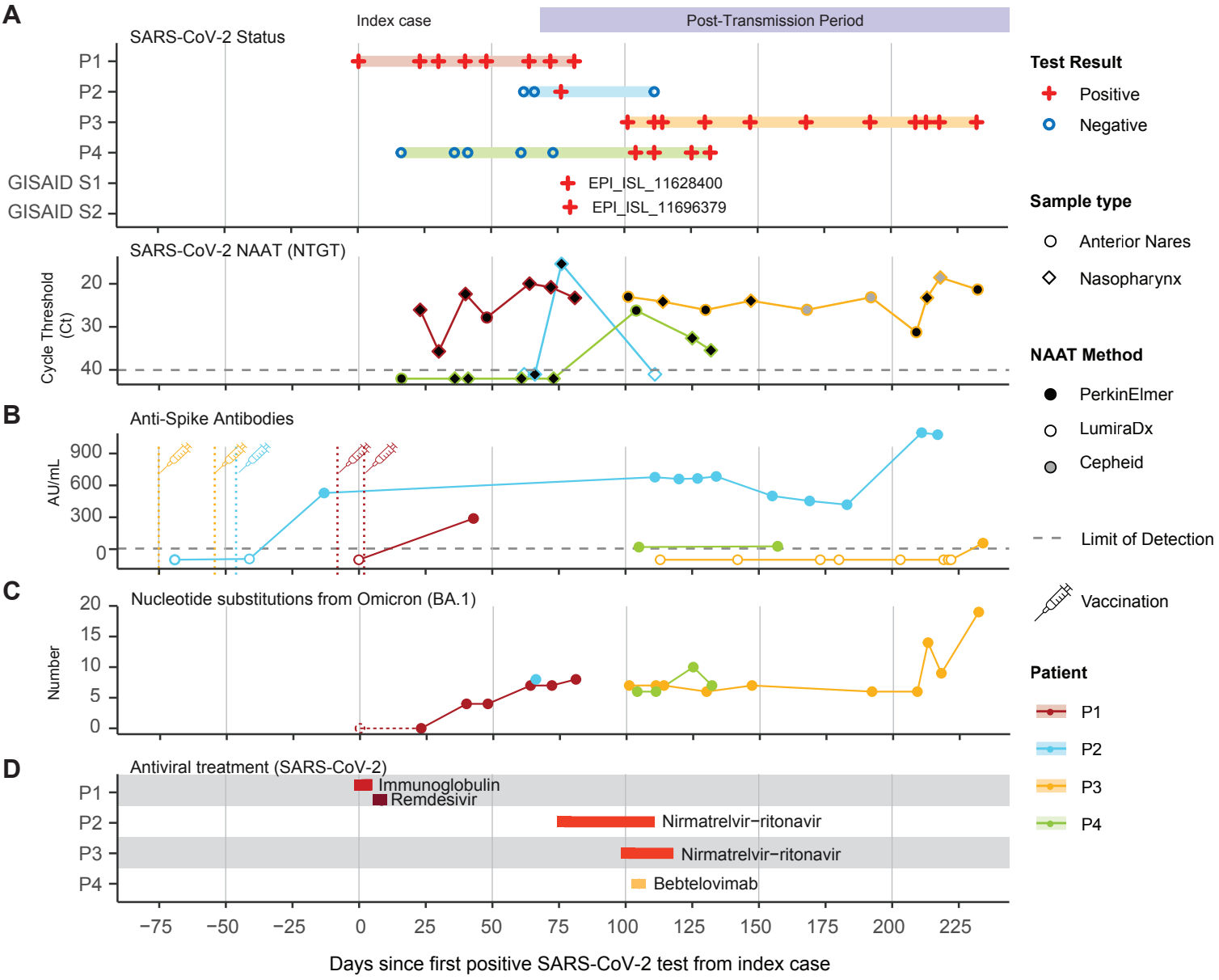

A
BA.1.1
S:R346K
P3 (d29)
P3 (d91)
P3 (d108)
S982L
P3 (d112)
ORF1a:T1001I,K1795Q,S4398L,ORF1ab:G671S,
S:S254F,L371F,N448S,M458K,F981L
ORF7a:T39I,N:A90S;ORF9b:E86D
P3 (d131)
P3 (d117)
EPI_ISL_11628400
P4 (d0)
P4 (d7)
P3 (d0)
P3 (d10)
P3 (d46)
P3 (d13)
BA.1.23
EPI_ISL_11696379
P4 (d21)
P4 (d28)
S:E96D,L455W,H681R
P1 (d72)
P2 (d0)
S:K458M,A484V,A688V
S:L176F,H681Y
P1 (d64)
P1 (d81)
S:R346T
P1 (d48)
P1 (d40)
P1 (d23)
T5515G
Omicron (BA.1)
50
55
60
65
70
75
80
Number of substitutions
B
nsp3 T1001/183I
nsp3 T1567/749N
nsp3 K1795/977Q
nsp12 S4398L/S6L
nsp12 G5063/671S
T19I
E96D
L176F
S254F
R346T
K356T
P384L
R408I
V445F
V445A
N448S
L455W
F456L
F456V
K458M
V445A
S477D
E484V
S680A
P681Y
P681R
A688V
D936Y
Q954H
V963F
N978D
L981F
S982L
E1195G
I4L
D61H
D61V
D61L
T39I/T24I
N43S
N43I
R78R
E91D
N10I
S413R
Days since first positive
ORF1a
ORF1b
S
NTD
RBD
CSR
S2
ORF6
ORF7a
7b
ORF8
ORF9b
N
Patient
23
40
48
64
72
81
P1
76 (0)
P2
101 (0)
111 (10)
114 (13)
130 (29)
147 (46)
192 (91)
209 (108)
213 (112)
218 (117)
232 (131)
P3
104 (0)
111 (7)
125 (21)
132 (28)
P4
78 (Unk)
EPI_ISL_11628400
79 (Unk)
EPI_ISL_11696379
T670G
G3085T
C3267T
G4184A
C4321T
C4391A
C4965A
C5497T
T5614C
A5648C
T6001C
T6238C
A7399G
C10198T
C13458T
G15451A
C15714T
C19164T
C19955T
A20055G
C21618T
G21850T
C22088T
C22323T
G22599C
A22629C
C22713T
G22785A
G22895T
T22896C
A22905G
T22926G
T22928C
T22928G
A22935T
A22964G
A22991G
A23013T
T23600G
C23603T
C23604G
C23625T
G24368T
A24424C
G24449T
A24494G
T24505A
C24507T
C24904T
A25146G
C27211A
G27382C
A27383T
T27384C
C27509T
A27521G
A27521T
G27626A
G27666T
T27824C
T27922A
C28253T
G28541T
C28550A
C28567T
T28570A
C28708T
C29230T
C29284T
A29510C
Position in SARS−CoV−2 Genome
SNV Frequency
0
1

638 **Figure 2. Forward transmission of the Omicron BA.1.23 subvariant**

639 **A)** Maximum likelihood phylogenetic subtree with SARS-CoV-2 (BA.1) sequences from the
640 persistent infection case P1 and the onward transmissions (P2, P3, P4), in a global background
641 of sequences available in GISAID. Branches are colored to identify the patient. The number of
642 days after the first SARS-CoV-2 positive specimen of P1 is indicated in brackets. Sibling
643 clusters of BA.1 and other sublineages are collapsed for easier visualization. The x-axis shows
644 the number of nucleotide substitutions relative to the root of the phylogenetic tree. Bootstrap
645 support values above 70% are shown for the un-collapsed branches. The distinct BA.1
646 subvariant that was transmitted was designated as PANGO lineage BA.1.23.

647 **B)** Frequency of single nucleotide variants (SNV), and amino acid substitutions for SARS-CoV-
648 2 genomic positions with consensus changes from ancestral Wuhan-1 and Omicron BA.1.
649 Positions with mixed nucleotides below consensus are also shown for intrahost SNVs
650 (mSNVs) seen in more than one time point. The sequenced specimens are shown sequentially
651 for patient 1 (P1) with prolonged BA.1 infection and transmission cases of BA.1.23 (P2, P3
652 and P4). The viral genomes from P1 show progressive accumulation of mutations in the N-
653 terminal domain (NTD), receptor binding domain (RBD), and S1/S2 furin cleavage site (FCS);
654 The same constellation of mutations was subsequently detected in three documented
655 transmission cases (P2, P3 and P4). The number of days since the first positive test in P1 is
656 shown on the left, with the number of days after the first positive test for each patient between
657 brackets. The mutation frequencies within the Spike protein are shown in red, and for non-
658 Spike genome regions in blue. Positions are numbered according to the reference genome
659 sequence NC_045512.2. Arrows on the left indicate the likely window of transmission from
660 the index patient between days 64-72.

Figure 3


A
SARS-CoV-2 Status
Index case
Post-Transmission Period
P1
P2
P3
P4
GISAID S1
GISAID S2
EPI_ISL_11628400
EPI_ISL_11696379
SARS-CoV-2 NAAT (NTGT)
Cycle Threshold (Ct)
20
30
40
B
Anti-Spike Antibodies
AU/mL
900
600
300
0
C
Nucleotide substitutions from Omicron (BA.1)
Number
20
15
10
5
0
D
Antiviral treatment (SARS-CoV-2)
Immunoglobulin
Remdesivir
Nirmatrelvir−ritonavir
Nirmatrelvir−ritonavir
Bebtelovimab
−75
−50
−25
0
25
50
75
100
125
150
175
200
225
Days since first positive SARS-CoV-2 test from index case
Test Result
Positive
Negative
Sample type
Anterior Nares
Nasopharynx
NAAT Method
PerkinElmer
LumiraDx
Cepheid
Limit of Detection
Vaccination
Patient
P1
P2
P3
P4

662 **Figure 3. Timeline of BA.1.23 evolution compared to antibody levels and treatments**

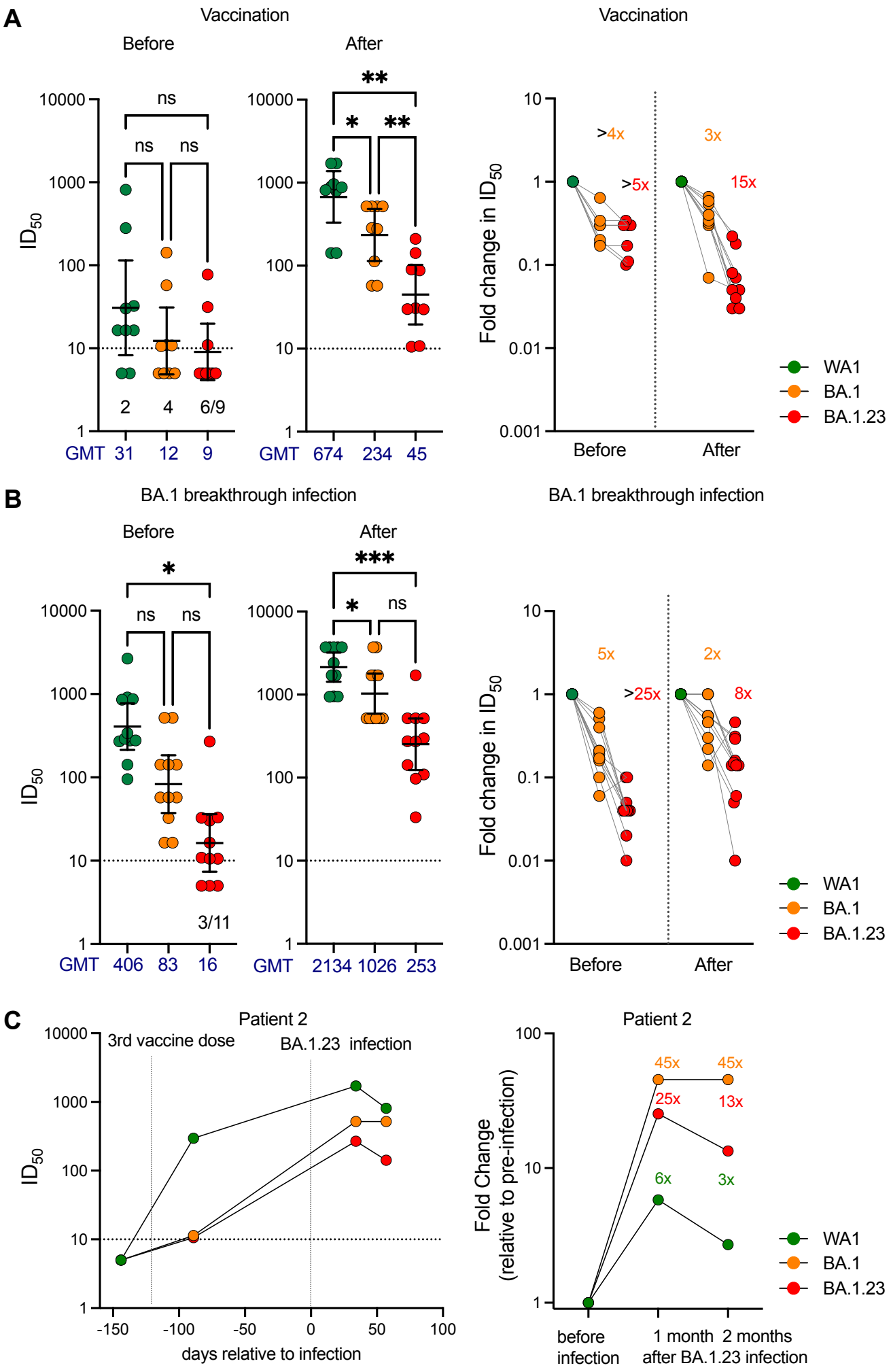
663 **A)** The timeline of BA.1.23 infected patients (top) is shown together with nucleic acid
664 amplification test (NAAT) SARS-CoV-2 (N gene) cycle threshold (Ct) values for respiratory
665 specimens, for different diagnostic methods (bottom). Empty markers indicate negative tests
666 by alternative diagnostic tests with no Ct value readout (LumiraDxTM).

667 **B)** Number of nucleotide substitutions in the consensus sequence relative to Omicron BA.1 in the
668 sequenced specimens.

669 **C)** SARS-CoV-2 spike binding IgG antibody levels for P1- P4. Antibody levels are shown in
670 arbitrary units per mL (AU/mL). Available vaccine information is indicated.

671 **D)** Timeline of SARS-CoV-2 antiviral treatments received by P1-P4. Treatment duration is
672 indicated by the length of the bar.

673


A
Vaccination
Before
After
ns
ns
ns
**
*
**
ID50
10000
1000
100
10
1
2
4
6/9
GMT 31 12 9
GMT 674 234 45
Vaccination
Fold change in ID50
>4x
>5x
3x
15x
0.001
0.01
0.1
Before
After
WA1
BA.1
BA.1.23
B
BA.1 breakthrough infection
Before
After
*
ns
ns
***
*
ns
3/11
GMT 406 83 16
GMT 2134 1026 253
BA.1 breakthrough infection
5x
>25x
2x
8x
C
Patient 2
3rd vaccine dose
BA.1.23 infection
days relative to infection
-150
-100
-50
0
50
100
Patient 2
Fold Change
(relative to pre-infection)
45x
45x
25x
13x
6x
3x
before infection
1 month after BA.1.23 infection
2 months

675 **Figure 4. The Omicron BA.1.23 subvariant displays strongly reduced neutralizing activity**
676 **compared to the parental Omicron variant as well as ancestral SARS-CoV-2 variant.**

677 **A)** Absolute neutralization titers (left) and fold reduction (right) for WA-1, BA.1 and BA.1.23
678 variants by paired sera from study participants collected before and after booster vaccination.
679 The two samples with levels below detection for WA-1 were excluded from the fold-change
680 calculation shown in the right panel.
681 **B)** Absolute neutralization titers (left) and fold-reduction (right) for WA-1, BA.1 and BA.1.23
682 isolates by sera from study participants who experienced breakthrough infection with BA.1.
683 Data for paired sera from 11 participants collected before and after BA.1 infection are shown.
684 **C)** Absolute neutralization titers (left) and fold-changes in neutralization titers relative to the post
685 booster timepoint (right) for each of the isolates (WA-1, BA.1 and BA.1.23) by sera collected
686 from patient P2 before and after BA1.23 infection. The first vertical dotted line (left) represents
687 the third vaccine dose. The second vertical dotted line (right) indicates the infection with the
688 BA.1.23 variant.

689 GMT denotes mean geometric mean of the inhibitory dilution 50% (ID50) values. The
690 horizontal dotted line represents the limit of detection (10). Samples with neutralization titers
691 below the level of detection were assigned the neutralization value of 5 (equaling to half of the
692 limit of detection). Fold changes in neutralization are A one-way ANOVA with Tukey's
693 multiple comparison test was used to compare the neutralization titers before and after booster
694 vaccination or BA.1 breakthrough infection. Significant p values (<0.05) are indicated in the
695 figure. Source data is provided in **Supplementary Table S2**.

696

EXTENDED DATA FIGURES

Extended Data Figure 1: Overview of genome-wide SARS-CoV-2 mutations found in 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.

709 **Extended Data Figure 2: Summary of SARS-CoV-2 mutations within the spike gene in**
710 **minority viral populations.**

711 The distribution of minority intrahost single nucleotide variants (mSNV) for positions that were
712 fixed over the course of the infection of the index and transmitted infections are shown indicated
713 with asterisks. Single nucleotide variants are colored by strain. The dotted bars indicate positions
714 with coverage below the threshold of 100X coverage for calling mSNVs. The positions with
715 minority variants present at frequencies > 0.1 are indicated with triangles for their respective
716 samples.

717

718 **Extended Data Figure 3: Summary of genome-wide SARS-CoV-2 mutations in minority**
719 **viral populations for patient P3**

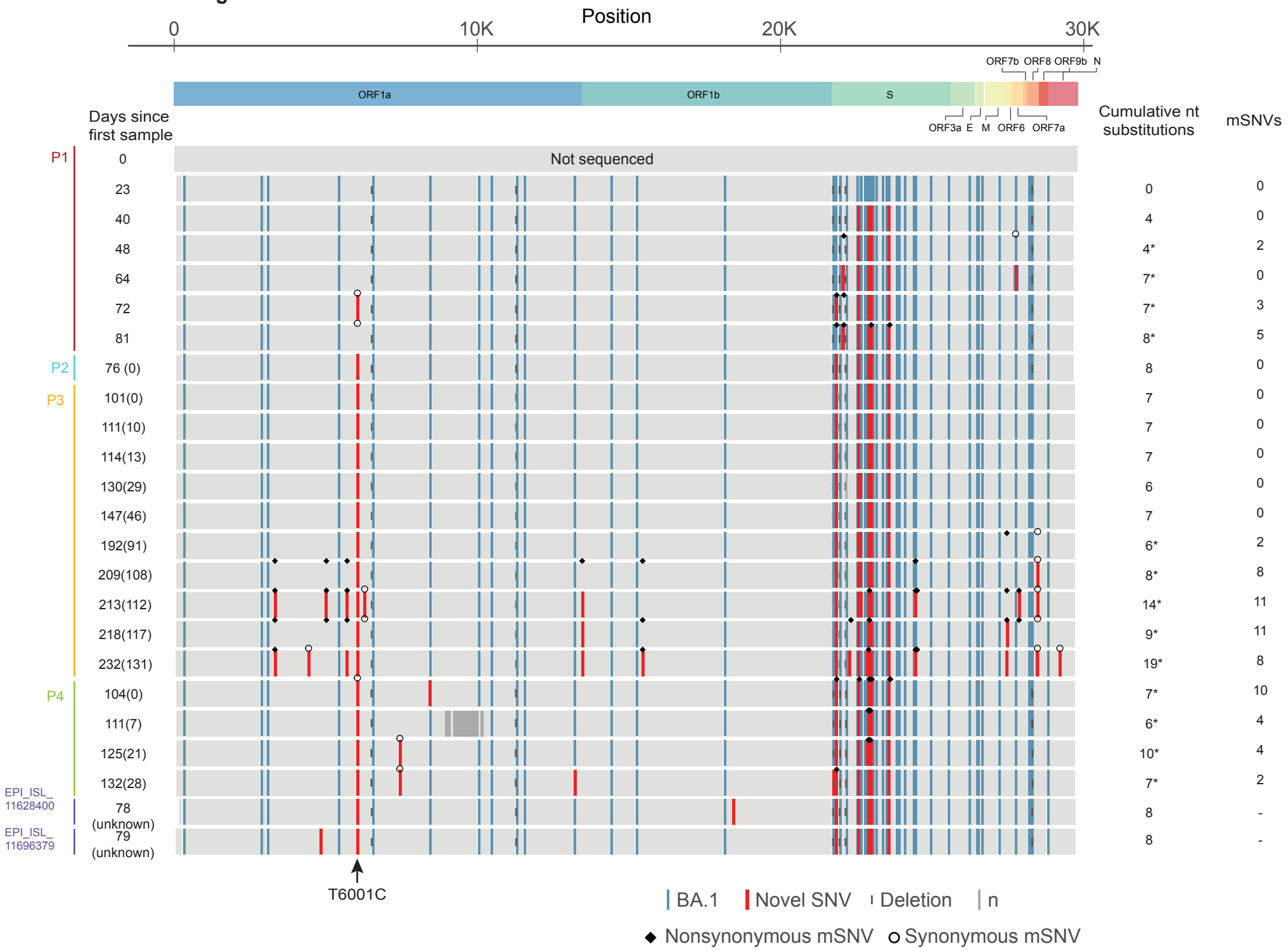
720 Similar to Extended Data Fig. 2 but showing the distribution of nucleotide variants below the
721 consensus level observed in one or more specimens from P3. Single nucleotide variants (SNVs)
722 are colored by strain. The dotted bars indicate positions with coverage below the threshold of 100X
723 coverage for calling mSNVs. The positions with minority variants present at frequencies > 0.1%
724 are indicated with triangles (nonsynonymous) or circles (synonymous) for their respective
725 samples.

726 **Extended Data Figure 4: Summary of SARS-CoV-2 mutations within the Spike gene in**
727 **minority viral populations for patient P3**

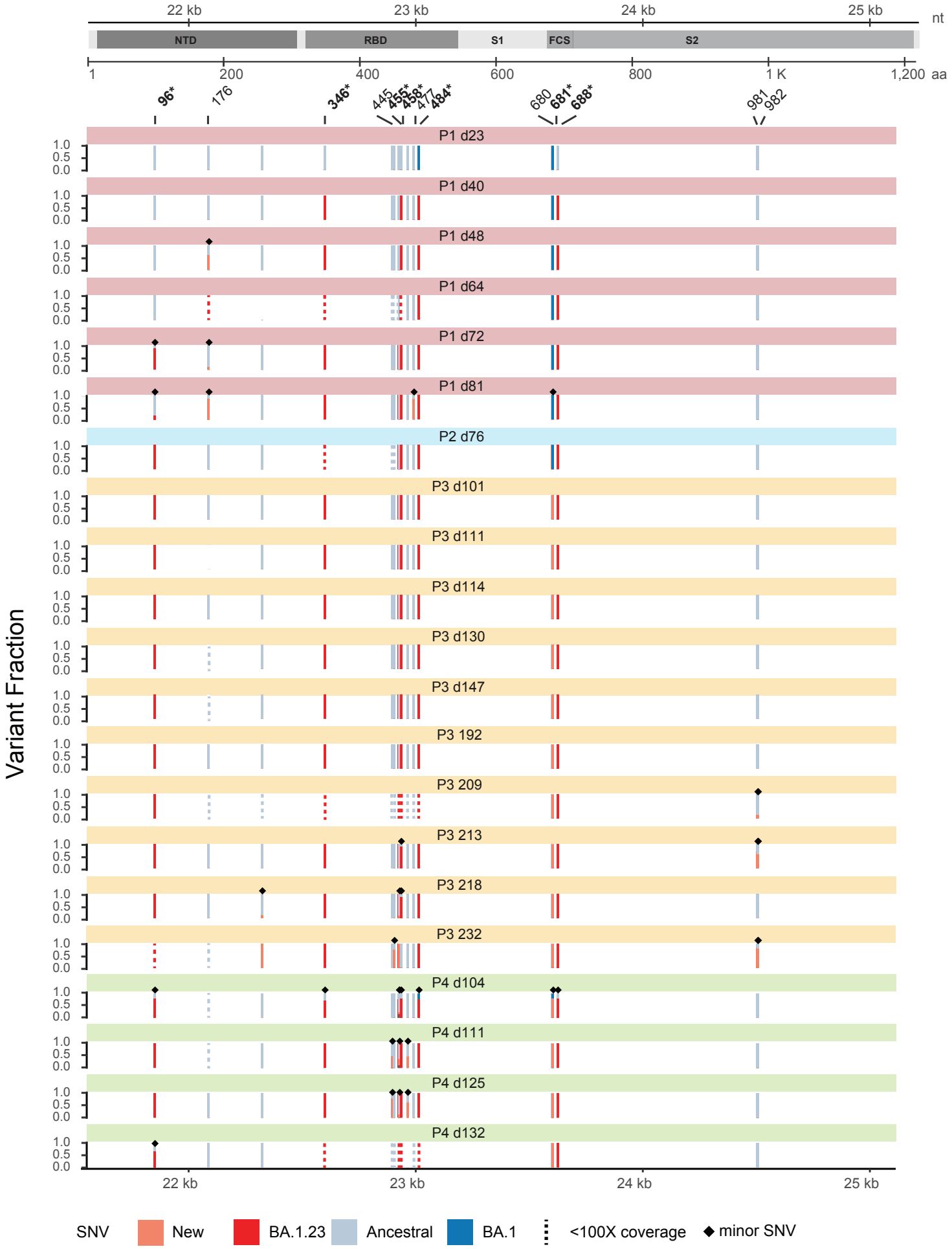
728 Similar to Extended Data Fig. 3 but showing minority mutations in the Spike gene region. Single
729 nucleotide variants (SNVs) are colored by strain. The dotted bars indicate positions with coverage
730 below the threshold of 100X coverage for calling minor SNVs (mSNVs). The positions with
731 minority variants present at frequencies > 0.1 are indicated with triangles for their respective
732 samples.

733

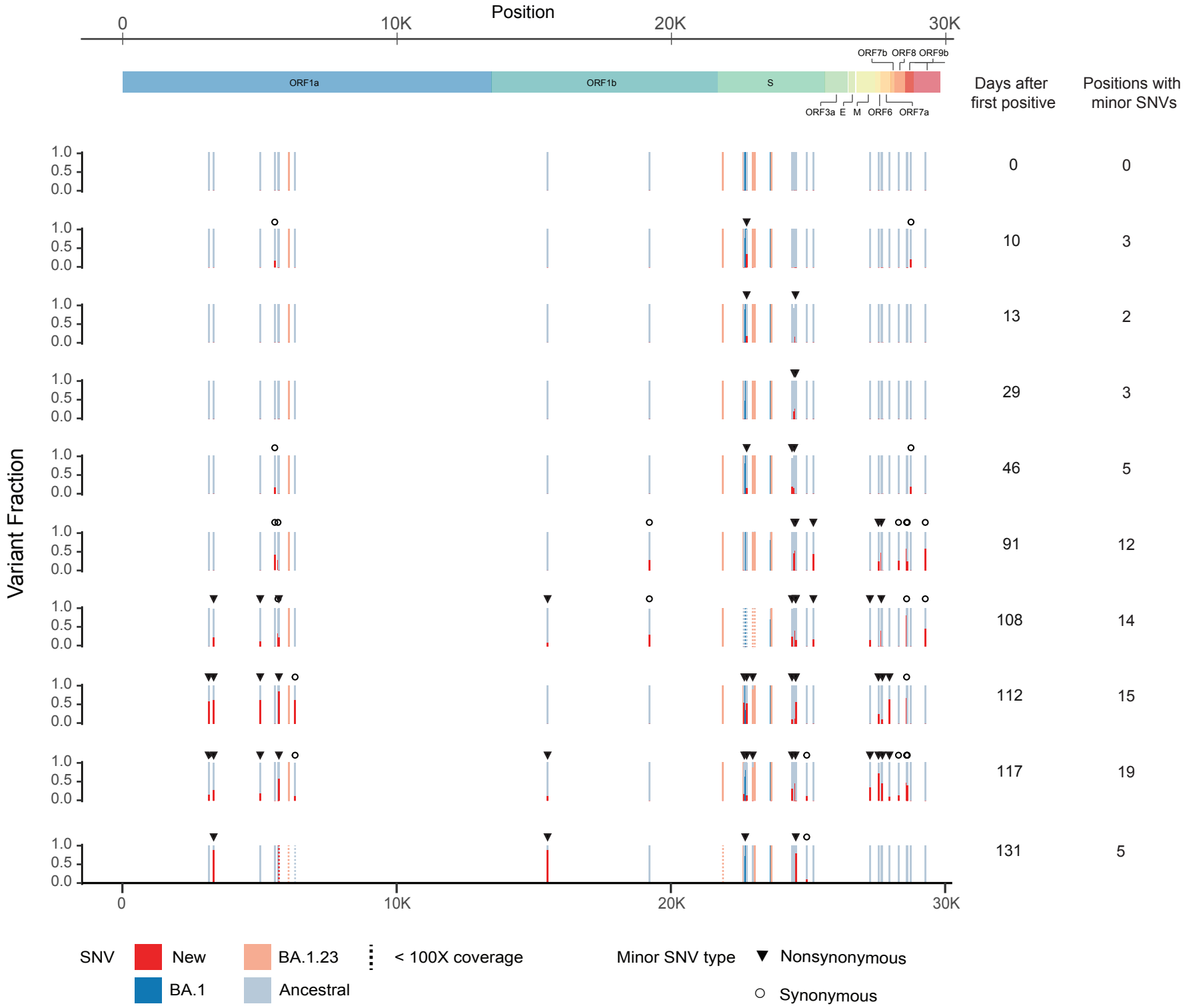
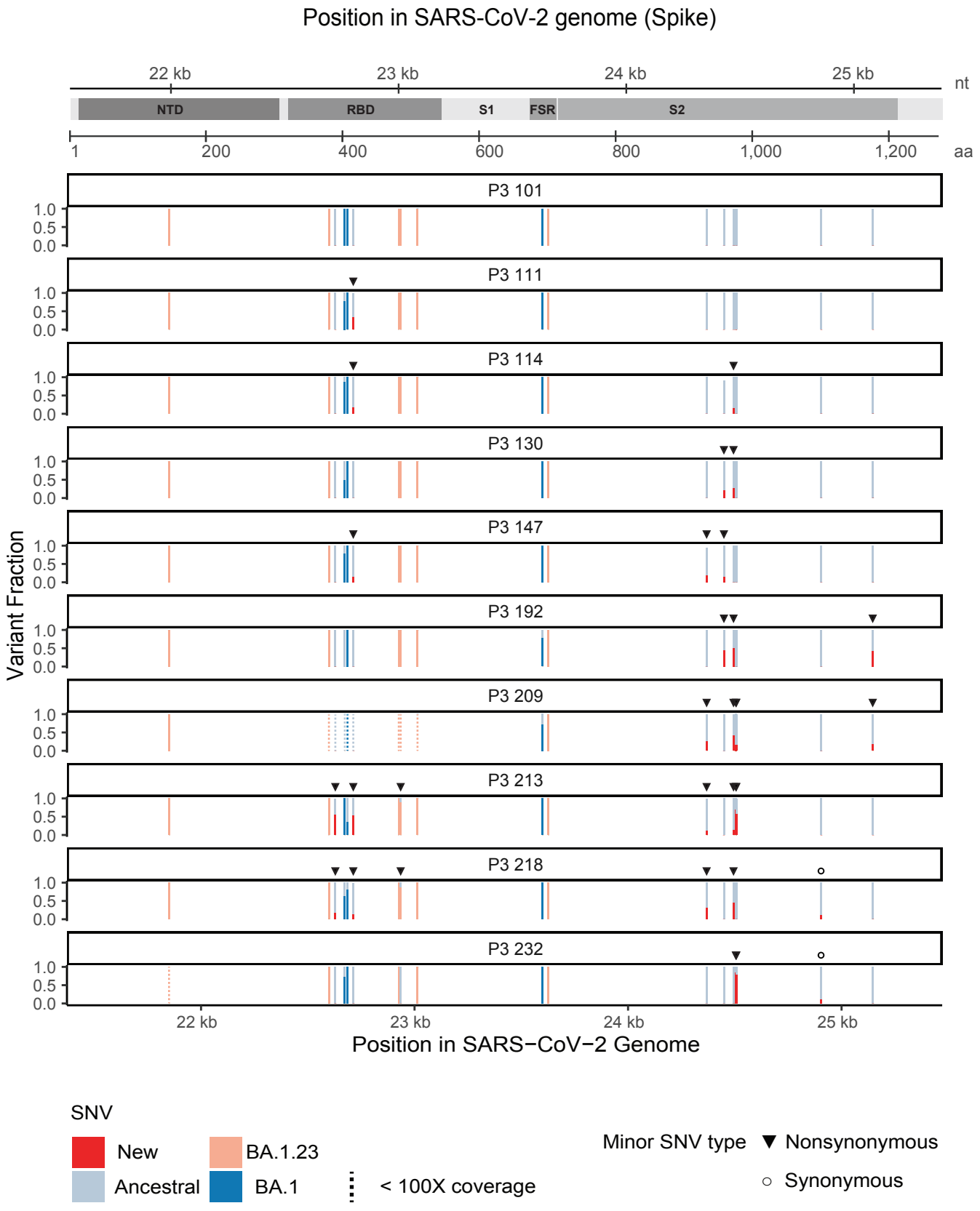
Extended Data Figure 1



Extended Data Figure 2



Extended Data Figure 3



Extended Data Figure 4