



OPEN

# Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2-mRNA-vaccinated individuals

Talia Kustin<sup>1,2,15</sup>, Noam Harel<sup>1,2,15</sup>, Uriah Finkel<sup>3</sup>, Shay Perchik<sup>3</sup>, Sheri Harari<sup>1,2</sup>, Maayan Tahor<sup>1</sup>, Itamar Caspi<sup>1</sup>, Rachel Levy<sup>1</sup>, Michael Leshchinsky<sup>3</sup>, Shifra Ken Dror<sup>4</sup>, Galit Bergerzon<sup>4</sup>, Hala Gadban<sup>4</sup>, Faten Gadban<sup>4</sup>, Eti Eliassian<sup>5</sup>, Orit Shimron<sup>5</sup>, Loulou Saleh<sup>6</sup>, Haim Ben-Zvi<sup>6</sup>, Elena Keren Taraday<sup>7</sup>, Doron Amichay<sup>7,8</sup>, Anat Ben-Dor<sup>7</sup>, Dana Sagas<sup>9</sup>, Merav Strauss<sup>9</sup>, Yonat Shemer Avni<sup>10,11</sup>, Amit Huppert<sup>12,13</sup>, Eldad Kepten<sup>3</sup>, Ran D. Balicer<sup>3</sup>, Doron Netzer<sup>14,15</sup>, Shay Ben-Shachar<sup>3,13,15</sup>✉ and Adi Stern<sup>1,2</sup>✉

**The BNT162b2 mRNA vaccine is highly effective against SARS-CoV-2. However, apprehension exists that variants of concern (VOCs) may evade vaccine protection, due to evidence of reduced neutralization of the VOCs B.1.1.7 and B.1.351 by vaccine sera in laboratory assays. We performed a matched cohort study to examine the distribution of VOCs in infections of BNT162b2 mRNA vaccinees from Clalit Health Services (Israel) using viral genomic sequencing, and hypothesized that if vaccine effectiveness against a VOC is reduced, its proportion among breakthrough cases would be higher than in unvaccinated controls. Analyzing 813 viral genome sequences from nasopharyngeal swabs, we showed that vaccinees who tested positive at least 7 days after the second dose were disproportionately infected with B.1.351, compared with controls. Those who tested positive between 2 weeks after the first dose and 6 days after the second dose were disproportionately infected by B.1.1.7. These findings suggest reduced vaccine effectiveness against both VOCs within particular time windows. Our results emphasize the importance of rigorously tracking viral variants, and of increasing vaccination to prevent the spread of VOCs.**

Mass vaccination against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is currently underway worldwide, providing hope that the coronavirus disease (2019) COVID-19 pandemic may soon be mitigated. In Israel, vaccination commenced on 20 December 2020, primarily with the BNT162b2 mRNA vaccine, and by mid-March 2021, more than 80% of the eligible population (all individuals 16 years old and above) were vaccinated with at least one dose. In clinical trials, the BNT162b2 mRNA vaccine was shown to be 95% efficacious in preventing symptomatic disease; a similarly high protective effectiveness has also been found in real-world settings in Israel<sup>1,2</sup>. However, concerns have emerged regarding the effectiveness of vaccines against various SARS-CoV-2 strains. In particular, three strains have recently been defined as VOCs by the WHO (World Health Organization): the B.1.1.7 strain (first detected in the UK), the B.1.351 strain (first detected in South Africa) and the P.1 strain (first detected in Brazil). Accumulating evidence suggests that the B.1.1.7 strain spreads more rapidly than the original circulating strain and leads to substantially more infections<sup>3,4</sup>.

Concerns have emerged that the B.1.351 and P.1 strains are able to overcome previous immunity to SARS-CoV-2 (refs. <sup>5,6</sup>), yet the evidence has been mixed. Using engineered viruses and/or sequences, laboratory studies have shown that neutralization of B.1.1.7 by BNT162b2-vaccine-elicited sera was either similar to or slightly reduced as compared to neutralization of early circulating isolates<sup>7–12</sup> with or without the globally dominant D614G alteration<sup>13</sup>. Conversely, a significant reduction in neutralization of B.1.351 was observed<sup>7–12</sup>, while other studies suggested neutralization remained relatively high against both B.1.1.7 and B.1.351 (ref. <sup>14</sup>). T cell responses, which are not captured by neutralization studies, were also shown to remain stable against these variants following vaccination<sup>15</sup>. Thus, it remains unknown whether VOCs can mediate BNT162b2 vaccine breakthrough in real-world settings, in which the vaccine elicits persistent antibody and T cell responses. Here, we tested the hypothesis that the B.1.1.7 and B.1.351 strains are able to overcome BNT162b2 mRNA vaccine protection, by comparing their distributions in infected vaccinated individuals and in infected non-vaccinated individuals.

<sup>1</sup>The Shmunis School of Biomedicine and Cancer Research, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel. <sup>2</sup>Edmond J. Safra Center for Bioinformatics, Tel Aviv University, Tel Aviv, Israel. <sup>3</sup>Clalit Research Institute, Innovation Division, Clalit Health Services, Ramat Gan, Israel.

<sup>4</sup>Clalit Health Services, Central Laboratories, Haifa and Western Galilee, Neshet, Israel. <sup>5</sup>Progenin Laboratories, Jerusalem District, Clalit Health Services, Tel Aviv, Israel. <sup>6</sup>Microbiology lab, Rabin Medical Center, Beilinson Hospital, Petah Tiqva, Israel. <sup>7</sup>Central Laboratory, Clalit Health Services, Tel Aviv, Israel.

<sup>8</sup>Department of Clinical Biochemistry and Pharmacology, Faculty of Health Sciences, Ben Gurion University of the Negev, Beersheba, Israel. <sup>9</sup>Microbiology Laboratory, Emek Medical Center, Afula, Israel. <sup>10</sup>Laboratory of Clinical Virology, Soroka University Medical Center, Beersheba, Israel. <sup>11</sup>Faculty of Health Sciences, Ben Gurion University of the Negev, Beersheba, Israel. <sup>12</sup>The Bio-statistical and Bio-mathematical Unit, The Gertner Institute for Epidemiology and Health Policy Research, Chaim Sheba Medical Center, Tel HaShomer, Ramat Gan, Israel. <sup>13</sup>The Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv, Israel. <sup>14</sup>Clalit Health Services, Tel Aviv, Israel. <sup>15</sup>These authors contributed equally: Talia Kustin, Noam Harel, Doron Netzer, Shay Ben-Shachar.

✉e-mail: [shayb@clalit.org.il](mailto:shayb@clalit.org.il); [sternadi@tauex.tau.ac.il](mailto:sternadi@tauex.tau.ac.il)

## Results

**Study population.** We began by identifying the relatively rare vaccinees with documented SARS-CoV-2 infection—symptomatic or asymptomatic—among members of Clalit Health Services (CHS), the largest health care organization in Israel, which insures 4.7 million patients (53% of the population). We divided these individuals into two categories: individuals who had a positive PCR test that was performed between 14 days after the first dose and 6 days after the second dose (denoted as the dose1 group); and individuals who had a positive PCR test that was performed at least 7 days after the second vaccine dose (denoted as the dose2 group). The definitions of these two categories were chosen to match the original BNT162b2 vaccine efficacy study<sup>2</sup>, as well as our ensuing real-world effectiveness study in Israel<sup>1</sup>, both of which revealed very high vaccine protection using these particular criteria. Each vaccinee (case) was matched with an unvaccinated infected individual (control) who tested positive on a similar date ( $\pm 3$  days) and had similar demographic characteristics (age, sex, ethnic sector and geographic location) to reduce bias associated with differential exposure (Methods). Next, we obtained RNA from the nasopharyngeal swabs sampled for PCR and performed complete viral genome sequencing for 813 samples from different individuals, consisting of 149 pairs of dose2–controls, 247 pairs of dose1–controls and additional samples whose match did not undergo successful sequencing (see below; Table 1, Supplementary Table 1 and Supplementary Fig. 1).

**Analysis of variant distributions.** We next used a stringent method of lineage assignment for each viral sequence (Methods). Aside from B.1.1.7 and B.1.351, no other VOCs or variants of interest, as defined by the WHO, were found in our sample (Supplementary Fig. 2). We hence collectively denoted all non-B.1.1.7 and non-B.1.351 lineages found as wild type (WT). All of these WT lineages bore the D614G alteration, in line with the very high frequency of this alteration across the globe<sup>13</sup>. We did not find evidence for the increased presence of any additional alterations that are not lineage-defining alterations of B.1.1.7 or B.1.351.

When examining lineage frequency across time, we noted that B.1.1.7 was the predominant strain of the virus in Israel over the entire sampling period (712/813 sequences), increasing in frequency over time (Fig. 1a). Conversely, B.1.351 was at an overall frequency of 1.6% in our sample of both vaccinated and non-vaccinated individuals (13/813 sequences) (Fig. 1b), similar to previous reports of B.1.351 frequency in Israel from January 2021<sup>16</sup>.

On the basis of previous results from neutralization assays, we hypothesized that B.1.1.7 may be slightly vaccine resistant as compared to WT, whereas B.1.351 may be more vaccine resistant when compared to both B.1.1.7 and WT. Under this hypothesis of ordered resistance, we performed our statistical analyses first on the B.1.1.7 strain, while excluding B.1.351 sequences (to avoid obscuring a potential signal), and then compared the B.1.351 with the B.1.1.7 and WT sequences combined (Fig. 2). We use the McNemar test on our paired vaccinees–controls to examine discordant pairs, defined as pairs where a different variant was found in the vaccinee as compared to its matched control. The McNemar test is particularly useful for comparing paired proportions in retrospective cohorts, where each case is paired with a control, as in the study herein. The null model of this test was that under a hypothesis of equal effectiveness of the vaccine against all variants, the different variants should be evenly distributed across the discordant pairs (see Fig. 2 for a more elaborate explanation).

No statistically significant difference was observed in the discordant rates of B.1.1.7 infection in dose2 cases versus unvaccinated controls (McNemar odds ratio (OR) of 6.4; one-sided exact McNemar test,  $P=0.38$ ), but a significantly higher proportion of B.1.351 was observed in dose2 cases versus unvaccinated controls (McNemar OR of 8.1; one-sided exact McNemar test,  $P=0.02$ ).

Of note, about half of dose2 cases tested positive on days 7–13 after the second dose, and about half tested positive 14 days or more after the second dose (Table 1). However, seven out of eight B.1.351 dose2 cases were isolated 7–13 days after the second dose and the eighth B.1.351 dose2 case was isolated exactly 14 days after the second dose (Fig. 3).

On the other hand, a significantly higher rate of B.1.1.7 was observed in dose1 cases versus unvaccinated controls (McNemar OR of 26.10; one-sided exact McNemar test,  $P=0.006$ ). For B.1.351 in the dose1 category, the sparsity of data (one infection in each category) precluded statistical analysis (Fig. 2). A conditional logistic regression was further performed on the dose1 B.1.1.7 data (as more data were available in this category), supporting the previous analysis: an OR of 2.4 was observed (95% confidence interval of 1.2 to 5.1). Age was included in the regression and was found to be a nonsignificant confounder, suggesting that its possible role in propensity for infection by a specific VOC was corrected through our matching scheme.

**Testing for biases and inclusion of missing data.** To test whether our sampling scheme was biased, we reconstructed a phylogenetic tree of all the sequenced samples together with additional available sequences from Israel, and observed that vaccinated and unvaccinated samples were highly interspersed along the tree (Fig. 4), ruling out strong biases in sampling. Moreover, we focused on the nine dose2 B.1.351 samples (eight cases and one control), and noted that they were from seven different municipalities spread across the geography of Israel. When examining the phylogenetic structure of B.1.351 sequences in Israel in general, we noted that most sequences belonged to one clade whose isolation dates ranged from 28 December 2020 until 9 February 2021, and accordingly most sequences in this clade were quite similar (Supplementary Fig. 5A). Nevertheless, most pairs of dose2 sequences were separated by one to six alterations (Supplementary Fig. 5B). Combined with the data on different municipalities, we conclude that it is unlikely (but not impossible) that these dose2 B.1.351 sequences were part of the same direct transmission chain.

Finally, we noted an additional two B.1.351 sequences, consisting of one dose2 case and one dose1 control, where the sequencing of the matched pair did not undergo successful sequencing, most often due to a high cycle threshold (Ct) value (low viral load (VL)). An additional dose1 control sequence was also ambiguously classified (Methods, Supplementary Fig. 3). Importantly, these sequences would either leave our conclusions regarding B.1.351 unchanged, or would increase the McNemar OR in favor of B.1.351 in the dose2 category (Supplementary Fig. 3), strengthening the results reported above. With regards to B.1.1.7, we found a total of 28 non-paired sequences, once again because a control or case yielded unreliable sequencing. These sequences might change the significance of our results with regards to B.1.1.7 but would not change the trend we found for this variant (Supplementary Fig. 3).

## Discussion

Our results show that there is an increased proportion of VOCs in vaccine breakthrough infections that occurs within two particular windows of time. An increased proportion of B.1.351 was found in individuals fully vaccinated with BNT162b2, 7–14 days after the second dose, as compared to the matched unvaccinated controls. Furthermore, an increased proportion of B.1.1.7 was found in partially vaccinated individuals, 14 days after the first dose until 6 days after the second dose, as compared to the matched unvaccinated control, yet we find no evidence for increased breakthrough rates of B.1.1.7 a week or more after the second dose (Figs. 2 and 3). Not enough data were available to assess vaccine breakthrough of B.1.351 in the dose1 category. These results are generally aligned with those from *in vitro* neutralization assays that have shown a large reduction

**Table 1 | Demographic statistics on paired cases and controls sequenced herein**

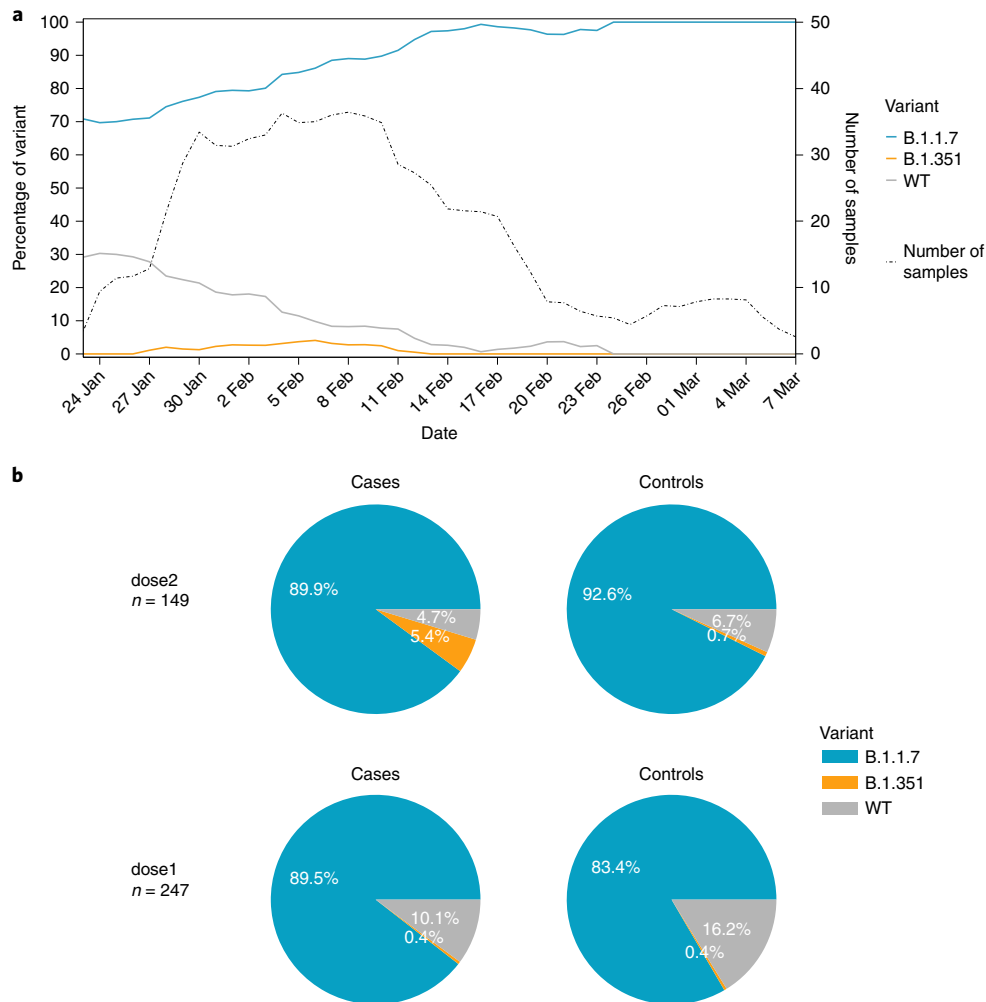
	Control dose2 vaccinee (n = 149)	dose2 vaccinee (n = 149)	Control dose1 vaccinee (n = 247)	dose1 vaccinee (n = 247)
Age group				
0–19	2 (1.3)		4 (1.6)	1 (0.4)
20–29	20 (13.4)	5 (3.4)	31 (12.6)	31 (12.6)
30–39	32 (21.5)	12 (8.1)	59 (23.9)	48 (19.4)
40–49	33 (22.1)	31 (20.8)	59 (23.9)	64 (25.9)
50–59	22 (14.8)	24 (16.1)	55 (22.3)	53 (21.5)
60–69	24 (16.1)	30 (20.1)	25 (10.1)	32 (13.0)
70–79	9 (6.0)	24 (16.1)	10 (4.0)	13 (5.3)
80–89	7 (4.7)	22 (14.8)	4 (1.6)	5 (2.0)
90+		1 (0.7)		
Sex				
Female	87 (58.4)	81 (54.4)	152 (61.5)	152 (61.5)
Male	62 (41.6)	68 (45.6)	95 (38.5)	95 (38.5)
District				
Dan	25 (16.8)	23 (15.4)	22 (8.9)	22 (8.9)
South	3 (2.0)	3 (2.0)	5 (2.0)	5 (2.0)
Haifa	45 (30.2)	45 (30.2)	70 (28.3)	70 (28.3)
Jerusalem	29 (19.5)	29 (19.5)	71 (28.7)	71 (28.7)
Center	23 (15.4)	23 (15.4)	28 (11.3)	28 (11.3)
North	7 (4.7)	7 (4.7)	22 (8.9)	22 (8.9)
Sharon-Shomron	7 (4.7)	9 (6.0)	19 (7.7)	19 (7.7)
Tel Aviv	10 (6.7)	10 (6.7)	10 (4.0)	10 (4.0)
Sector				
General Jewish	117 (78.5)	117 (78.5)	163 (66.0)	163 (66.0)
Jewish Orthodox	11 (7.4)	11 (7.4)	28 (11.3)	28 (11.3)
Non-Jewish	21 (14.1)	21 (14.1)	56 (22.7)	56 (22.7)
Vaccine status				
Non-vaccinated	149 (100.0)		247 (100.0)	
14–20 days from first dose				133 (53.8)
21–28 days from first dose				95 (38.5)
28+ days from first dose				19 (7.7)
7–13 days from second dose		73 (49.0)		
14–20 days from second dose		30 (20.1)		
21+ days from second dose		46 (30.9)		

Absolute counts are shown; relative proportions are in brackets.

in neutralization against B.1.351, and little to no reduction against B.1.1.7 in fully vaccinated individuals<sup>7–11,17</sup>. Overall, our data also suggest that serum-based neutralization studies may provide a good proxy for real-life protection in the case of SARS-CoV-2 (ref.<sup>18</sup>). Although this remains to be tested in a more widespread manner, it suggests that neutralization studies may be valid as a prompt first step before the establishment of real-world studies in the case of the emergence of new SARS-CoV-2 VOCs.

The power of our approach stems from the combination of real-world evaluation with the stringent case–control matching strategy employed, allowing us to rule out that a high proportion of a given variant was due to a confounding effect. For example, an outbreak of B.1.351 in a given city would have led to spurious identification of infected vaccinees there, yet by matching them with unvaccinated individuals, this confounding effect is controlled for. However, it is still possible that other confounding effects were

present and were not controlled for, such as various behavioral effects among vaccinees. Additionally, sequencing limitations prevented us from sequencing very low-VL samples (Methods), and thus the focus of our study was on vaccinees who generated higher VLs. However, it has been shown that cases with a low VL may be a lesser concern from a public health perspective, as they are associated with fewer symptoms and lower risk of transmission<sup>19</sup>. Finally, our dose2 cohort is based on infections documented 7 or more days after the second vaccine dose (Table 1). Some individuals in this cohort may have been infected before the immunity from the boost was fully established, and it is thus possible that enhanced immunity from the boost, which develops over time<sup>20</sup>, may more effectively prevent infection with the B.1.351 variant. Notably, when focusing on the eight B.1.351 cases in the dose2 group, all tested positive during days 7–14 after the second dose, and none tested positive more than 14 days after the second dose. This observation suggests



**Fig. 1 | Variant frequencies of SARS-CoV-2-positive samples. a**, Variant frequencies across the time of the study, including the number of samples collected throughout the study. All values were calculated by averaging over a sliding window of 7 days. All samples sequenced in this study are included herein, including unpaired samples. **b**, Breakdown of variant frequencies based on the four groups of this study. The pie charts display the proportion of each variant (B.1.1.7, B.1.351 and WT) for paired vaccinated cases versus non-vaccinated controls separated by dosage (as defined in the main text), with cases on the left and their associated control on the right. Only paired samples are shown in the figure.

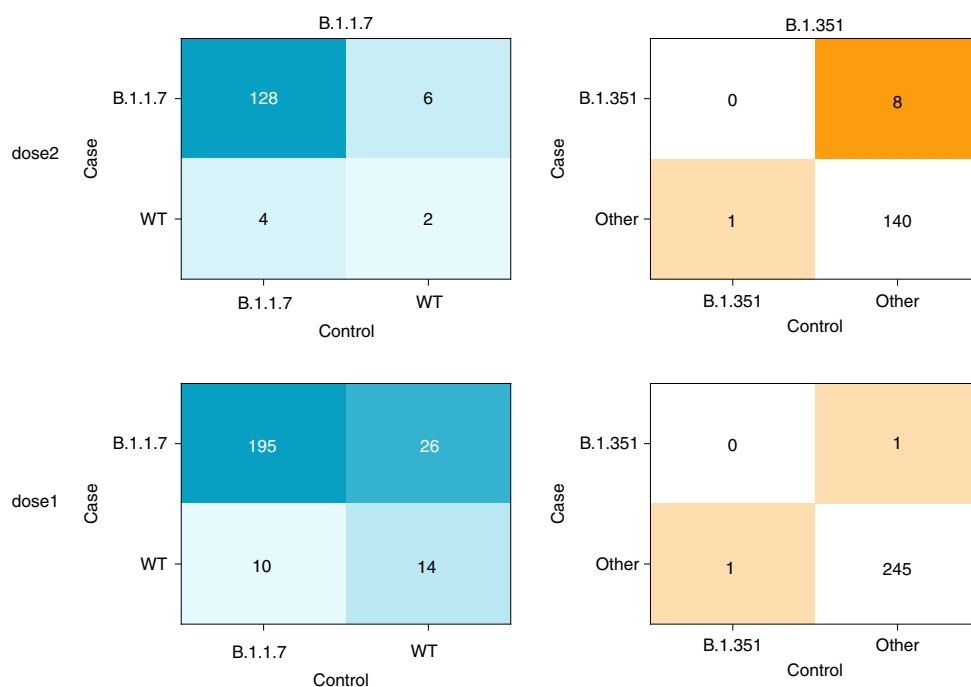
that increased breakthrough of B.1.351 in our cohort occurs mainly in a limited time window post vaccination.

The main caveat of our study was the small sample size of both the WT and B.1.351 variants. These small sample sizes are a product of: the dramatic increase in frequency of the B.1.1.7 variant, first detected in Israel in mid-December 2020, and reaching an overall frequency of ~90% or higher during the period of this study (Fig. 1a); and the low frequency of the B.1.351 variant in Israel at the time of writing<sup>16</sup>. In fact, in our latest samples obtained in late February and early March 2021, we noted fixation of the B.1.1.7 variant, but this interpretation requires caution as our sample size was low (Fig. 1a). Furthermore, caution is required from overinterpreting the McNemar ORs obtained, for two reasons: statistically, they do not necessarily represent the OR of breakthrough; and the absolute numbers we found, in particular for B.1.351 infections, are very small.

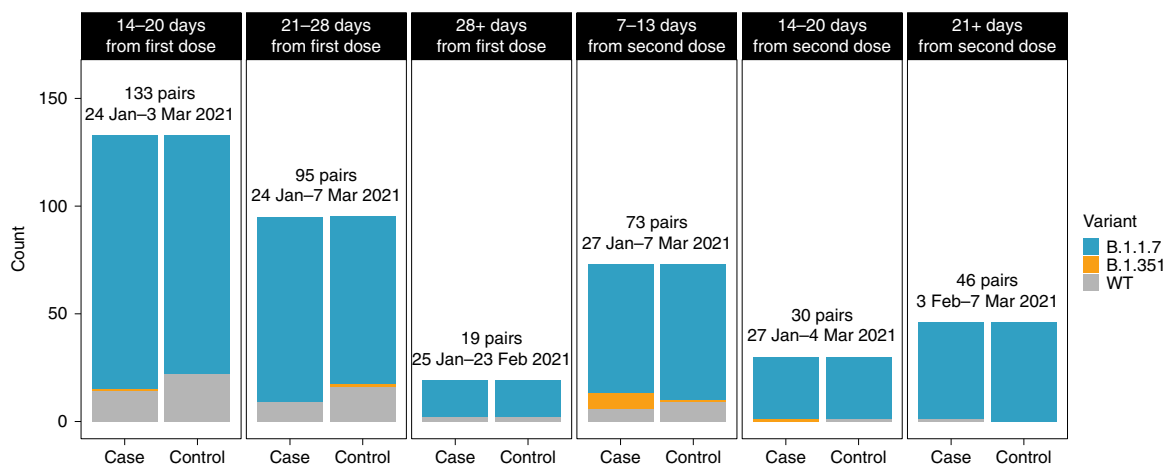
Our study design was not intended to deduce vaccine effectiveness against either variant, as we observe VOCs conditioned on infection. In other words, our focus is only on infected individuals; we ignore non-infected individuals, and do not measure absolute infection rates in the vaccinated or control population. Thus, we can only cautiously speculate on vaccine effectiveness against the B.1.1.7 and B.1.351 strains. Previous real-world work has shown

a very high effectiveness of the BNT162b2 vaccine starting a week after the second dose in a large-scale study performed in Israel among CHS patients<sup>1</sup>. During the period of that study, B.1.1.7 rose to a high frequency in Israel, suggesting that the high vaccine effectiveness observed in the study included high effectiveness against this strain as well. However, our current study may suggest a lower protection against B.1.1.7 in the first weeks after the first vaccine dose. As some countries opt to increase the gap between the first and the second BNT162b2 vaccine from the recommended 3 weeks to a longer period<sup>21</sup>, it is important to carefully assess whether this delay impacts vaccine effectiveness against the B.1.1.7 strain among individuals who received only the first dose. In our data we do not observe increased breakthrough of either the B.1.1.7 or the B.1.351 strain 2 weeks after the second dose, yet we note our data are relatively limited in this period (76 cases, Table 1). Our results are overall aligned with recent results that have shown slightly reduced vaccine effectiveness against the B.1.1.7 and B.1.351 variants as compared to WT, 14 days after the second dose<sup>22</sup>.

We conclude by discussing mechanistic explanations for why we see increased breakthrough rates at very particular and different time windows following vaccination for B.1.1.7 and B.1.351. From a biological point of view, the breakthrough cases observed in this



**Fig. 2 | Results of matched vaccinated cases and non-vaccinated controls separated by effectiveness and VOC.** In each table, a cell reflects the number of pairs concordant (upper left and lower right) or discordant (upper right or lower left) for a given variant. The left panel focuses on the comparison between B.1.1.7 and WT (pairs with B.1.351 were removed), whereas the right panel focuses on comparing B.1.351 and either WT or B.1.1.7 (denoted collectively as ‘other’). Of note, the McNemar test focuses on a comparison of only discordant samples. Under a null hypothesis of equal vaccine effectiveness against all variants, we expect an equal number of discordant pairs in the upper right cell and the lower left cell, in each of the tables.



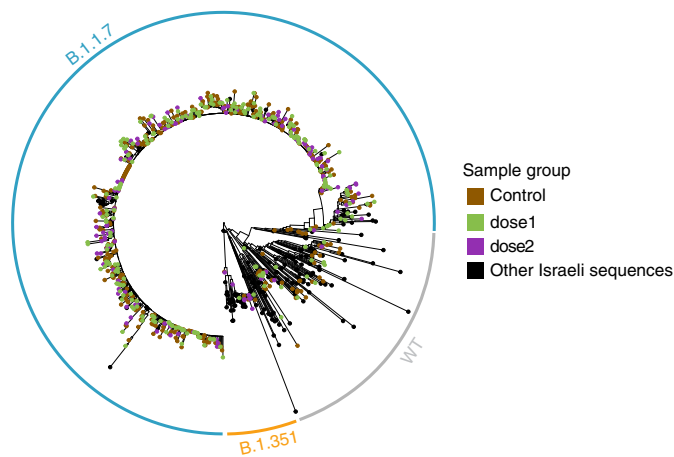
**Fig. 3 | Breakdown of SARS-CoV-2 variant distribution during windows of weeks post vaccination.** The first three panels correspond to the dose1 group and the last three panels correspond to the dose2 group. The number of pairs and the isolation date range of the samples are noted for each panel. The dose2 B.1.351 case that is shown in the 14–20 days category was isolated exactly 14 days after the second dose.

study might be due to immune evasion, mediated by particular alterations present in these strains<sup>9,23–27</sup>. Alternatively, it has been previously reported that B.1.1.7 is associated with lower Ct values, corresponding to higher VLs<sup>3</sup>, which may be sufficient to overcome the less potent immune response elicited by the vaccine before its augmentation by a second dose. We note that in this study we did not observe higher VL in B.1.1.7 infections as compared to other variants (see also ref.<sup>28</sup>), yet we did note higher VL in B.1.351 infections, while noting lower VL in vaccinees as compared to unvaccinated individuals (Supplementary Fig. 6)<sup>29,30</sup>. We stress that these findings are preliminary and may be affected by various behavioral biases. In particular, vaccinees in Israel were exempt from quarantine and

testing following exposure to a positive patient, and this may affect when and how they chose to be tested. Moreover, this created a bias in their symptomatic status—most vaccinees were likely tested only when they were symptomatic. For this reason, we refrain from reporting the rate of symptoms in our cohort.

We were reassured to observe the low frequency of B.1.351 across time (Fig. 1a)<sup>16</sup>. Of note, both B.1.1.7 and B.1.351 were first detected in Israel in late December, at the time vaccination commenced. Our sampling began during a peak of epidemic growth, during increasing rates of vaccination, into a phase of epidemic contraction (Fig. 1 and Supplementary Fig. 4). Owing to these complex dynamics, we can only speculate that selection does not strongly favor the B.1.351





**Fig. 4 | A maximum-likelihood phylogenetic tree of Israeli SARS-CoV-2 samples including those sequenced herein.** Vaccinees are colored in violet or green, non-vaccinees are colored in brown, and black sequences are publicly available sequences from Israel (marked as 'other', Supplementary Table 2). Clades composed of the B.1.1.7, B.1.351 and WT sequences are encircled in blue, orange and gray, respectively.

variant in the particular conditions in Israel, despite the increased rate of vaccination. This may be due to its limited ability to evade vaccine-elicited immunity, mainly during days 7–13 after the second dose. Alternatively, from an evolutionary point of view, it is possible that immune evasion alterations incur a fitness cost in the form of reduced transmissibility, especially as compared to the highly transmissible B.1.1.7 (ref.<sup>3</sup>). More research is required to further understand the evolutionary pressures operating on VOCs, in Israel, and around the world. At the time of revision of this paper, May 2021, we note that case counts in Israel have dramatically dropped to around 35 cases per day in a population of ~9 million (ref.<sup>31</sup>), suggesting that while vaccine breakthrough infections at particular windows of time may be more frequent with the VOCs B.1.1.7 and B.1.351, mass vaccination with two doses controls and contains their spread.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-021-01413-7>.

Received: 30 April 2021; Accepted: 26 May 2021;

Published online: 14 June 2021

### References

- Dagan, N. et al. BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *N. Engl. J. Med.* **384**, 1412–1423 (2021).
- Polack, F. P. et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
- Davies, N. G. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021).
- Davies, N. G. et al. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* **593**, 270–274 (2021).
- Tegally, H. et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. Preprint at *medRxiv* <https://doi.org/10.1101/2020.12.21.20248640> (2020).
- Naveca, F. et al. Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein. *Virological.org* <https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585> (2021).

- Liu, Y. et al. Neutralizing activity of BNT162b2-elicited serum. *N. Engl. J. Med.* **384**, 1466–1468 (2021).
- Xie, X. et al. Neutralization of SARS-CoV-2 spike 69/70 deletion E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* **27**, 620–621 (2021).
- Collier, D. A. et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136–141 (2021).
- Zhou, D. et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184**, 2348–2361.e6 (2021).
- Wang, P. et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021).
- Kuzmina, A. et al. SARS-CoV-2 spike variants exhibit differential infectivity and neutralization resistance to convalescent or post-vaccination sera. *Cell Host Microbe* **29**, 522–528 (2021).
- Korber, B. et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812–827 (2020).
- Liu, Y. et al. Neutralizing activity of BNT162b2-elicited serum—preliminary report. *N. Engl. J. Med.* **384**, 1466–1468 (2021).
- Tarke, A. et al. Negligible impact of SARS-CoV-2 variants on CD4<sup>+</sup> and CD8<sup>+</sup> T cell reactivity in COVID-19 exposed donors and vaccinees. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.02.27.433180> (2021).
- South African variant found in about 1% of all positive tests. *MOH* <https://www.gov.il/en/departments/news/22022021-01> (2021).
- Planas, D. et al. Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 to neutralizing antibodies. *Nat. Med.* **27**, 917–924 (2021).
- Khoury, D. S. et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat. Med.* <https://doi.org/10.1038/s41591-021-01377-8> (2021).
- Marks, M. et al. Transmission of COVID-19 in 282 clusters in Catalonia, Spain: a cohort study. *Lancet Infect. Dis.* **21**, 629–636 (2021).
- Walsh, E. E. et al. Safety and immunogenicity of two RNA-based Covid-19 vaccine candidates. *N. Engl. J. Med.* **383**, 2439–2450 (2020).
- Optimising the COVID-19 vaccination programme for maximum short-term impact. *GOV.UK* <https://www.gov.uk/government/publications/prioritising-the-first-covid-19-vaccine-dose-jvci-statement/optimising-the-covid-19-vaccination-programme-for-maximum-short-term-impact> (2021).
- Abu-Raddad, L. J. et al. Effectiveness of the BNT162b2 Covid-19 vaccine against the B.1.1.7 and B.1.351 variants. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMc2104974> (2021).
- Greaney, A. J. et al. Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.12.31.425021> (2021).
- Weisblum, Y. et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* **9**, e61312 (2020).
- Greaney, A. J. et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **29**, 44–57 (2021).
- Starr, T. N. et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850–854 (2021).
- García-Beltrán, W. F. et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372–2383.e9 (2021).
- Walker, A. S. et al. Increased infections, but not viral burden, with a new SARS-CoV-2 variant. Preprint at *medRxiv* <https://doi.org/10.1101/2021.01.13.21249721> (2021).
- Levine-Tiefenbrun, M. et al. Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine. *Nat. Med.* **27**, 790–792 (2021).
- Petter, E. et al. Initial real world evidence for lower viral load of individuals who have been vaccinated by BNT162b2. Preprint at *medRxiv* <https://doi.org/10.1101/2021.02.08.21251329> (2021).
- COVID-19 Dashboard* (Israeli Ministry of Health); <https://datadashboard.health.gov.il/COVID-19/general> (accessed 26 May 2021)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

## Methods

**Ethics statement.** The study was approved by the CHS institutional review board (no. 0016-21-COM2) and was exempt from the requirement for informed consent. The study was further approved by the Tel Aviv University ethics committee (0002706-1).

**Sample matching.** Data for this study were obtained from CHS's data repositories (Supplementary Fig. 1). The study population consisted of individuals who tested positive for SARS-CoV-2 by PCR with reverse transcription from six major CHS testing laboratories located throughout Israel. Individuals with a positive PCR test were then classified into one of two groups: controls who were not vaccinated before the positive PCR result; and cases that were vaccinated at least 14 days before the PCR result. Cases were further divided into two additional subgroups: individuals who had a positive PCR test that was performed between 14 days after the first dose and 6 days after the second dose were denoted as the dose1 group, and individuals who had a positive PCR test that was performed at least 7 days after the second vaccine dose were denoted as the dose2 group. Next, each case was matched to a control using six parameters: date of sampling for PCR ( $\pm 3$  days), sex, age ( $\pm 10$  years), municipality of residence, geographical district of residence and sector. If two or more controls were available, one was chosen at random. In preliminary analyses, we noted that matching often failed for dose2 samples due to their small sample size, as well as due to increasing proportions of vaccinated individuals in older age categories across time, in line with the vaccine rollout policy in Israel. To increase dose2 matching, we enforced matching on the date of PCR sampling, but allowed for four out of five matches in the remaining parameters, while prioritizing municipality, then sector, then age, and then the additional criteria. We found that the failed parameter match was most often age, sex or municipality. We note that ten control samples served as controls for both a dose1 and a dose2 sample. Table 1 summarizes statistics on the parameters used for matching and other parameters for the various groups of our sample. We also note that some vaccinees (7.7%, Table 1) received the second vaccine dose more than 28 days after the first vaccine, yet were still annotated as described above.

**Obtaining RNA samples and sequencing.** Following matching, RNA from cases and controls was obtained from the main testing laboratories of CHS, with one major limitation: only samples with Ct values of 33 or lower were collected. Ct values were averaged over all genes tested (per laboratory). The dates of the samples ranged from 23 January 2021 to 7 March 2021 (Supplementary Fig. 4). Full-genome sequencing of SARS-CoV-2 was performed based on the ARTIC protocol with a V3 primer set (<https://artic.network/ncov-2019>), with slight modifications detailed below. Briefly, reverse transcription and multiplex PCR was performed in two amplicon pools, and NEBNext Multiplex Oligos for Illumina were ligated to allow for sequencing. All samples were run on an Illumina MiSeq using 250-cycle V2 kits at either the Technion Genome Center (Israel) or at the Genomics Research Unit at Tel Aviv University (Israel). We and others have previously noted amplicon dropout of amplicons 74 and 76 (ref.<sup>32</sup>), both of which cover the spike gene, and in particular some of the lineage-defining alterations of B.1.1.7 and B.1.351 (such as E484K and N501Y). To increase the sequencing yield of these amplicons, we doubled the primer concentrations of both amplicons in our primer pool and lowered the annealing–extension temperature to 63°C.

**Bioinformatic analysis and lineage assignment.** Sequencing reads were trimmed using pTrimmer v1.3.1, a multiplexing primer trimming tool<sup>33</sup>, and then aligned to the reference genome of SARS-CoV-2 (GenBank ID MN908947) using our AccuNGS V1 pipeline<sup>34</sup> that is based on BLAST<sup>35</sup>, using a particular stringent e-value of  $10^{-30}$ . We then set out to determine the consensus sequence of each sample. Typically studies report a majority-rule consensus sequence; that is, the consensus base at each position in the genome is the base that most reads (>50%) support. However, the biological meaning of variable positions where more than one base is observed is complex, especially if such positions are abundant: they may indicate within-host variation, they may indicate multiple genotype infection, they may indicate sample contamination, and they may indicate sequencing errors. To overcome these limitations, we constructed two consensus sequences for each sample, one based on majority rule, and a more strict consensus sequence where we required at least 80% of reads to support a given base. Bases with lower support were assigned an N (ambiguous base). We also noted some regions with fluctuating ambiguity: if a continuous subsequence of length 20 or lower was flanked on both ends by ambiguous bases, we masked out this entire subsequence by assigning it with N, under the assumption it represents unreliable sequencing. In both types of consensus sequencing assignments, we required sequencing coverage of at least ten reads. Finally, we used the Pangolin v2.2.2 software (<https://github.com/cov-lineages/pangolin>) to assign lineages for each consensus sequence using the Pango nomenclature (pangoLEARN 2021-02-06)<sup>36</sup>, which requires that at least 50% of bases sequenced are unambiguous. After verifying the type of lineages we obtained, we labeled all consensus sequences as either B.1.1.7, B.1.351 or WT. Samples for which Pangolin labels of the strict and majority-rule consensus sequences did not coincide were discarded. Thirty-two pairs in which one sample did not undergo successful sequencing were discarded from the paired analyses (but see

Supplementary Fig. 3). The unpaired successful samples were, however, included in the variant frequencies across time analysis (Fig. 1a).

Following classification by Pangolin, we noted that one dose1 control sequence, originally classified as WT (B.1.235), was located within the B.1.351 clade on the phylogenetic tree. Its pair was classified as B.1.1.7, and we included this pair in our extreme scenarios analysis (Supplementary Fig. 3). This is in line with recent concerns regarding misclassifications of Pangolin<sup>37</sup>, and led us to manually verify the phylogenetic location of all sequences in our study.

R v4.0.4, Python v3.7.4, pandas v0.24.2 (ref.<sup>38</sup>), Matplotlib v3.2.1 (ref.<sup>39</sup>), Seaborn v0.10.1 and ggplot2 (ref.<sup>40</sup>) were used during the data analysis and visualization.

**Statistical analysis.** For all primary analyses, a one-sided paired (exact) McNemar's test was used to compare breakthrough of a variant in partially or fully vaccinated individuals. For the analysis of B.1.351, all other variants were defined as the reference group, while for the B.1.1.7 analysis, we excluded any paired observation that included B.1.3.5 (assuming ordinality of breakthrough), while any other variant was defined as the reference. A conditional logistic regression was used as a sensitivity analysis to include age as a possible confounder in case that matching was not sufficient, under the assumption that it was sometimes only partially mediated through matching. The regression was performed only on the dose1 B.1.1.7 analysis, as not enough data were available in other categories. All analyses were conducted with R software version 4.03 and the survival and exact2x2 packages.

**Phylogenetic analysis.** All Israeli sequences available on GISAID (<https://www.gisaid.org/>) from August onwards were downloaded, focusing on high-quality sequences with 10% or less ambiguous sites. Of these sequences, owing to computational limitations, we sampled the most distant 100 WT sequences and 50 B.1.1.7 sequences, and included all available B.1.351 sequences (Supplementary Table 2). The reference genome sequence (MN908947.3) was added on as well, and these sequences were combined with sequences from this study that contained at most 10% ambiguous sites. Alignment was performed using MAFFT v7.300b (ref.<sup>41</sup>) with default parameters. Next, a maximum-likelihood phylogeny was reconstructed using PhyML v3.0\_360-500M (ref.<sup>42</sup>) with default parameters as well, and the tree was rooted using the MN908947.3 sequence from the original outbreak first detected in Wuhan. The ggtree v2.5.1 package was used to visualize the phylogenetic trees.

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

## Data availability

All sequences were uploaded to GISAID and the accession numbers are stated in Supplementary Table 1. The raw sequencing reads were deposited in the National Center for Biotechnology Information Sequence Read Archive database under BioProject accession number PRJNA728463.

## Code availability

Code used to generate consensus sequences as described in the section entitled Bioinformatic analysis and lineage assignment is available at <https://github.com/SternLabTAU/COVID19-VACC>.

## References

- Itokawa, K. et al. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLoS ONE* **15**, e0239403 (2020).
- Zhang, X. et al. pTrimmer: an efficient tool to trim primers of multiplex deep sequencing data. *BMC Bioinformatics* **20**, 236 (2019).
- Gelbart, M. et al. Drivers of within-host genetic diversity in acute infections of viruses. *PLoS Pathog.* **16**, e1009029 (2020).
- Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
- Dudas, G. et al. Travel-driven emergence and spread of SARS-CoV-2 lineage B.1.620 with multiple VOC-like mutations and deletions in Europe. Preprint at *medRxiv* <https://doi.org/10.1101/2021.05.04.21256637> (2021).
- McKinney, W. Data structures for statistical computing in Python. In *Proc. 9th Python in Science Conference* (2010).
- Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
- Villanueva, R. A. M. & Chen, Z. J. *ggplot2: Elegant Graphics for Data Analysis* (Taylor & Francis, 2019).
- Katoh, K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
- Guindon, S. et al. in *Bioinformatics for DNA Sequence Analysis* (ed Posada, D.) 113–137 (Springer, 2009).

## Acknowledgements

We thank R. Milo, N. Kopelman, Y. Woodbridge, D. Burstein and T. Hagai for their advice and discussions, S. Krispin for editing support, and O. Tirosh and M. Meir for extensive technical support. We also thank the Technion Genome Center for rapid sequencing services. This publication is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 852223, RNAVirFitness). This study was further supported by an Israeli Science Foundation grant (3963/19) and by kind donations from the Milner and AppsFlyer foundations. This study was supported in part by fellowships to T.K., N.H. and S.H. from the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Author contributions

T.K. and N.H. performed the bioinformatics analyses. S.P., U.F., E.K. and M.L. performed the clinical data analyses. A.H. provided consultation on study design and statistical analyses. U.F., S.P., I.C., T.K., N.H. and A.S. performed the statistical analyses. T.K., N.H., U.F., S.P., E.K., S.B.-S. and A.S. drafted the paper. T.K., N.H., S.H., M.T. and R.L. coordinated clinical sample collection and sequencing. S.H. and M.T. performed library

preparation for sequencing. S.K.D., G.B., H.G., F.G., E.E., O.S., L.S., H.B.-Z., E.K.T., D.A., A.B.-D., D.S., M.S. and Y.S.A. contributed clinical samples. T.K., N.H., S.P., E.K., D.N., R.D.B., S.B.-S. and A.S. conceived and coordinated the study. A.S., D.N. and S.B.-S. supervised and led the study.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41591-021-01413-7>.

**Correspondence and requests for materials** should be addressed to S.B.-S. or A.S.

**Peer review information** *Nature Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work. Alison Farrell was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



## Reporting Summary

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

### Statistics

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

n/a Confirmed

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

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection None

Data analysis pTrimmer v1.3.1, AccuNGS V1, pangolin 2.2.2, pangoLEARN 2021-02-06, R software version 4.03 and the survival and exact2x2 packages, Python v3.7.4, pandas v0.24.2, Matplotlib v3.2.1, Seaborn v0.10.1, ggplot2 v3.3.3, ggtree v2.5.1, MAFFT v7.300b, PhyML v3.0\_360-500M, BLAST. All software used is cited.

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

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

Data that support the findings of this study have been deposited in the relevant databases: a viral sequence per each patient sample was deposited in the GISAID database (<https://www.gisaid.org>) and accession numbers are stated in the study (Table S1). The raw sequencing reads were deposited in the NCBI Sequence Read Archive (SRA) database under BioProject accession number PRJNA728463 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA728463>). The reference genome of SARS-CoV-2, ID MN908947, was downloaded from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) and the list of all sequence accession numbers used in this study is available in table S2.

## Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

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

Study description	SARS-CoV-2 samples were collected from infected individuals. Samples of vaccinees with the BNT162b2 mRNA vaccine were matched with samples from non-vaccinees. All samples were sequenced and the variant of the virus was determined.
Research sample	RNA extracted from nasopharynx swabs of SARS-CoV2 positive individuals was collected by six major Clalit Health Services testing labs located throughout Israel for the purpose of testing for COVID-19 (not related to this research). This study used leftovers of these samples. Samples were collected across the time interval of January 23, 2021 and March 07, 2021. The samples included both males and females of all ages. Our sample consisted of the vaccinated infected individuals that were sampled in one of the six laboratories of Clalit Health Services and that were successfully matched with an unvaccinated control (See Fig S1)
Sampling strategy	Each sample of a vaccinated individual was matched to a sample of a non-vaccinated individual based on the following parameters: date of sampling for PCR (+/- three days), sex, age (+/- 10 years), municipality of residence, geographical district of residence, and sector. We performed power analysis in order to determine the desired sample size. Since the B.1.351 frequency in Israel prior to our research was estimated at 1%, our analysis resulted in a desired sample size of 800-1000 samples (400-500 matched pairs). We have reached close to 400 pairs and concluded our sampling mainly due to the decline in the number of cases in Israel and more specifically almost no infected vaccinees.
Data collection	Extracted RNA was collected and delivered by Clalit Health Services for sequencing in the Genomic centers at the Technion or at Tel Aviv University. Metadata of samples was provided by Clalit Health Services and included sampling date, age, sex and vaccination status. Data regarding the match of each sample of vaccinee according to the matching scheme was provided by Clalit Health Services.
Timing and spatial scale	Samples were collected starting on January 23, 2021 till March 07, 2021 from six major Clalit Health Services testing labs that are spread across Israel. Our sampling started at the start of the reaserch and our end date was when we reached around 800 samples as described above.
Data exclusions	Samples whose sequencing results were unsuccessful were excluded from the analysis. This led to the exclusion of pairs of sequences: if a control or case was excluded, their paired case or control was excluded, respectively. .
Reproducibility	There is no possibility to reproduce this type of cohort study, especially given the small numbers of infected vaccinated individuals.
Randomization	Allocation to groups was made based on vaccination status. Covariates were controlled to as much as possible using our matching scheme between vaccinees and non-vaccinees based on the following parameters: date of sampling for PCR (+/- three days), sex, age (+/- 10 years), municipality of residence, geographical district of residence, and sector.
Blinding	All samples were anonymized, and were assigned a random identifier by Clalit Health Services
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

## Reporting for specific materials, systems and methods

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

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	SARS-CoV-2 samples were collected from infected individuals. Samples of vaccinees with the BNT162b2 mRNA vaccine were matched with samples from non-vaccinees. Each sample of a vaccinated individual was matched to a sample of a non-vaccinated individual by the following parameters: date of sampling for PCR (+/- three days), sex, age (+/- 10 years), municipality of residence, geographical district of residence, and sector. Aggregated population data is available in Table 1.
Recruitment	No direct recruitment was performed. This was a retrospective study, that relied on leftover samples from routine testing. Biases may include the subjective inclination of individuals to go and get tested and the fact we discarded samples with high CTs (low viral load) and their pairs due to inability to sequence those samples. These biases are described in the main text.
Ethics oversight	The study was approved by the Clalit Health Services institutional review board (IRB #0016-21-COM2) and was exempt from the requirement for informed consent. The study was further approved by the Tel-Aviv University ethics committee (0002706-1).

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