medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Clinical and genomic signatures of rising SARS-CoV-2 Delta breakthrough infections in New York

Author Affiliations:

Ralf Duerr[1*], Dacia Dimartino[2], Christian Marier[2], Paul Zappile[2], Samuel Levine[3], Fritz François[4], Eduardo Iturrate[4], Guiqing Wang[5], Meike Dittmann[1], Jennifer Lighter[6], Brian Elbel[7, 8], Andrea B. Troxel[7], Keith S. Goldfeld[7], and Adriana Heguy[2,5*]

1. Department of Microbiology, NYU Grossman School of Medicine

2. Genome Technology Center, Office of Science and Research, NYU Langone Health

3. Hospital Operations, NYU Langone Health

4. Department of Medicine, NYU Grossman School of Medicine

5. Department of Pathology, NYU Grossman School of Medicine

6. Department of Pediatric Infectious Diseases, NYU Grossman School of Medicine

7. Department of Population Health, NYU Grossman School of Medicine

8. NYU Wagner Graduate School of Public Service

* Co-corresponding authors

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Corresponding authors:

Ralf Duerr, MD, PhD

Department of Microbiology, NYU Grossman School of Medicine

Alexandria Center for Life Science (ACLS), West Tower

430 East, 29th Street, Room 323

New York, NY 10016

Email: ralf.duerr@nyulangone.org

Tel: 212-263-4159

Adriana Heguy, PhD

Department of Pathology, NYU Grossman School of Medicine

Genome Technology Center, NYU Langone Health

550 First Avenue, MSB 294A

New York, NY 10018

Email: adriana.heguy@nyulangone.org

Tel: 212-263-8048

Conflict of Interest Statement: The authors declare that no conflict of interest exists

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

In 2021, Delta has become the predominant SARS-CoV-2 variant worldwide. While vaccines effectively prevent COVID-19 hospitalization and death, vaccine breakthrough infections increasingly occur. The precise role of clinical and genomic determinants in Delta infections is not known, and whether they contribute to increased rates of breakthrough infections compared to unvaccinated controls. Here, we show a steep and near complete replacement of circulating variants with Delta between May and August 2021 in metropolitan New York. We observed an increase of the Delta sublineage AY.25, its spike mutation S112L, and nsp12 mutation F192V in breakthroughs. Delta infections were associated with younger age and lower hospitalization rates than Alpha. Delta breakthroughs increased significantly with time since vaccination, and, after adjusting for confounders, they rose at similar rates as in unvaccinated individuals. Our data indicate a limited impact of vaccine escape in favor of Delta's increased epidemic growth in times of waning vaccine protection.

Introduction

The SARS-CoV-2 pandemic has been a showcase for observing viral evolution in real time. New variants emerged in various parts of the globe and caused waves of infection that reached different countries in rapid succession. Variants being monitored (VBM), particularly Alpha, Beta, and Gamma, and variant of concern (VOC) Delta emerged and subsequently increased despite worldwide vaccination efforts[1-3]. In the United States, vaccinations started in late December 2020, with three vaccines currently employed: two mRNA-based vaccines, BNT162b2 (Pfizer/BioNTech), now FDA-approved, mRNA-1273 (Moderna), and the adenovirus-based

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Janssen COVID-19 vaccine, JNJ-78436735[4-6]. All three vaccines utilize the spike sequence of SARS-CoV-2 Wuhan-Hu-1 isolated in January 2020[7,8], and thus their epitopes are not perfectly matched to those of currently circulating variants. Despite this mismatch, the vaccines are highly effective at preventing symptomatic disease, hospitalization, death, and forward transmission[9-11]. However, post-vaccination infections (vaccine breakthroughs) do occur. The epidemiological data on whether these breakthroughs are driven by properties inherent to specific variants are controversial. Two studies earlier in 2021 indicated increased breakthrough rates of the Beta or Gamma variants following two doses of mRNA vaccines[12,13]. In contrast, other studies, including our own, found that breakthrough cases had a similar variant distribution as in unvaccinated individuals in the respective cohort at that specific time[14-17].

In 2021, Delta became the predominant variant globally. Data from the CDC indicate that coincident with the rise of Delta in the US, vaccine effectiveness decreased from 91% to 66%, which is consistent with waning immunity observed in Israel and Qatar[18-20]. Delta has been associated with higher replication and transmission[21,22], and *in vitro* studies report slightly decreased neutralizing efficacy of monoclonal antibodies, convalescent sera or sera from vaccinated individuals against Delta, suggesting that some mutations present in Delta may cause partial immune escape[23-25]. Clinical and full SARS-CoV-2 genome data from post-vaccination infections in the era of Delta are still scarce[26] but will be key to revealing features critical for vaccine effectiveness/escape. Here, we determined the SARS-CoV-2 genetic makeup of 132 post-vaccination infections and their associated clinical characteristics in fully vaccinated individuals within NYU Langone Health (NYULH), a large metropolitan New York healthcare

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

system with hospitals across the region. We assessed the probability of Delta to result in breakthrough infection relative to other variants.

Results

Demographic parameters in our cohort of vaccinated and unvaccinated SARS-CoV-2-infected participants

Between May 1st and August 3rd, 2021, a total of n=1613 SARS-CoV-2 infections were recorded in our multicenter healthcare system. The majority of SARS-CoV-2-positive tests (82%) were from unvaccinated individuals (n=1297), whereas 18% were from vaccinated (n=297) among a total of 168,127 fully vaccinated individuals within our system (**Table 1**). We sequenced all breakthrough cases with available specimens and randomly selected as many specimens from unvaccinated individuals as allowed by our internal weekly processing capacity (up to 96), irrespective of demographic or clinical parameters. We obtained high quality SARS-CoV-2 sequences from 132 vaccinated and 283 unvaccinated individuals. The median age (37 versus 42 years) and sex distribution were similar in both groups. The majority of recorded breakthrough infections (63%) occurred at >120 days after full vaccination with a median of 136 days after vaccination (**Table 1**).

Delta infections were associated with a distinct network of clinical and demographic predictors.

More than 45% of the breakthrough infections in our cohort were asymptomatic (n=60); only 6% (n=8) were hospitalized, all of whom were > 50 years (5/8 >70 years) and had co-

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

morbidities. There was one death from metastatic cancer (**Table S1, Figs S1, S2**). The hospitalization rate was significantly higher in the unvaccinated (29%, n=83, *P*<0.0001); for 66 of those, COVID-19 was the main reason for hospitalization, 10 required ICU admission, and seven died from COVID-19 (**Table S2**). To identify differential features of the circulating variants, we compared their clinical and demographic characteristics (**Figs. 1a, b, S1-S3**). Among the SARS-CoV-2-positive cases studied, hospitalizations due to Delta were significantly lower compared with Alpha, Gamma, Iota, and the other variants combined, and significance was maintained against Alpha when assessed separately in vaccinated and unvaccinated individuals (**Fig. S1**). Among Delta infections, we recorded only one COVID-19 related death (0.7%), whereas Alpha had the highest rates with at least four COVID-19-related deaths (5.9%) (**Tables S1, S2**). Delta affected younger individuals, particularly when compared to Alpha. Hospitalizations were associated with older age, specifically in Alpha, Delta, and Iota infections. There was a trend to lower cycle threshold (Ct) values in Delta infections, which translated into a weak but significant inverse correlation between Delta and Ct values in the overall data set and among breakthrough infections (**Figs. 1b, S2a**). The Ct values did not differ in hospitalized versus non-hospitalized. In all, Delta infections exhibited a characteristic clinical profile regarding age (younger), time of infection (later in the year) and hospitalization rate (lower), all of which significantly correlated with each other and might be mutual drivers (**Fig. 1b**).

Delta rapidly replaced Alpha and Iota variants in both unvaccinated and vaccine breakthrough cases recorded at NYULH.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

At the beginning of our study period (May 2021), several variants circulated in the New York metropolitan area, mainly Alpha, Iota and Gamma (**Fig. 1c, d**). In June, we observed a nadir of case numbers (<10 per week), followed by a rise of Delta infections starting end of June that caused a dramatic increase in total cases. This rapid shift in variants towards Delta was detected in both vaccinated and unvaccinated individuals. By the end of our study, we saw a near complete replacement of the locally-originating Iota and the imported Alpha and Gamma variants by Delta.

Delta infections including breakthroughs exhibit a characteristic subvariant makeup.

To determine the diversity and relatedness of SARS-CoV-2 genomes, we performed phylogenetic analyses including our 132 breakthrough and 283 unvaccinated cases, using US and global reference sequences (**Figs. 2a, S4** and **Table S3**). Our SARS-CoV-2 sequences covered a broad set of variants with predominance of Delta and its numerous subvariants, Alpha, Gamma, Iota, Mu, and additional B.1 lineages (**Fig. 2b)**. Delta sequences appeared more densely packed and clustered, indicative of transmission chains (**Figs. 2A, S4**). One of these clusters consisted of the Delta subvariant AY.25 that emerged recently and included sequences with spike mutation S112L. There was no evidence of transmission clusters solely composed of breakthrough infections. Alpha and Iota were overall less frequently found among breakthrough infections (odds ratios and 95% confidence intervals (95%CI) of 0.35 (0.19-0.66) and 0.25 (0.09-0.69), respectively), whereas Delta counts were more frequent over the study period, which was primarily due to B.1.617.2 and AY.25 (odds ratios and 95%CI of 2.12 (1.34-3.36) and 1.97 (1.01-3.73), respectively (**Fig. 2b**). The S112L-positive AY.25 sub-cluster

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

contained SARS-CoV-2 sequences from one unvaccinated and five vaccinated cases (**Figs. 3, S5, S6**), with four cases occurring in New York County.

Potential sites of adaptive evolution of the Delta variant under vaccine pressure.

To screen for genomic signatures of vaccine breakthrough and possible adaptive selection, we assessed all base pair mutations over the full genome and also all amino acid mutations across spike, enriched in vaccinated or unvaccinated participants (**Figs. 3, 4, S5**, **S6 Tables S4-S6**). The numbers of full genome and spike mutations were comparable in viruses obtained from vaccinated and unvaccinated individuals. Breakthrough sequences had in average 35.8 base pair mutations over the full genome (range: 24-44) and 9.7 amino acid mutations per spike (range: 3-14), compared to 35.6 base pair mutations/full genome (range: 19-56) and 9.6 amino acid mutations/spike (range: 3-15) in the unvaccinated. In spike, we found a total of 168 amino acid changes compared to Wuhan-Hu-1: 115 sites were more frequently mutated in unvaccinated controls, compared to 52 sites in breakthroughs (**Fig. 3a**). To identify spike amino acid mutations that are enriched in breakthrough sequences, we performed a two-pronged approach: I) to investigate the enrichment of mutations across variants, we statistically analyzed sequences from all variants and combined it with a posthoc variant assessment to assign the predominant variant(s) for each mutation and to correct for potential bias introduced by unequal variant distribution between groups (**Fig. 3b**); II) a Delta-specific mutation analysis was done to identify mutations enriched among Delta sequences and to confirm the results using all variants (**Fig. 3c**). Among the 52 mutations preferentially found in breakthroughs, 10 significantly differed between breakthrough and unvaccinated cases (**Fig. 3b,**

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table S4). The posthoc analysis revealed that nine of these 10 sites were Delta-defining mutations, whilst one, S112L, is not Delta-defining but contributes to the array of Delta subvariant AY.25 mutations. The comparative spike mutation analysis exclusive for the subset of Delta sequences confirmed S112L as the only substantially enriched mutation among Delta breakthroughs ($p<0.05$, **Fig. S5** and **Table S5**). The S112L mutation locates at the surface-exposed top part of NTD, one of the major antigenic regions of the spike protein (**Fig. 3d**)[27].

To identify sites of potential adaptive evolution in Delta beyond spike, we extended our statistical analysis to the full genome (**Fig. 4**). Among 240 Delta sequences (101 breakthroughs and 139 unvaccinated controls), we observed 791 mutation sites, of which 448 were more frequently mutated in controls, 328 more frequent in breakthroughs, and 15 equally mutated in both groups. In total, we found nine sites with significantly different mutation frequencies between the groups (**Fig. 4a, b, Table S6**). Two mutations were enriched in breakthrough sequences compared to unvaccinated controls, which, besides spike mutation S112L included mutation F192V in nsp12, the RNA-dependent RNA polymerase (RdRp) gene. In addition, seven mutations were less frequent in breakthrough sequences, which included four non-ambiguous mutations in nsp2, nsp14 (2x), and ORF7a, and three ambiguous mutations in nsp2, nsp9, and nsp15. A genomic region-specific mutation analysis further revealed that most coding regions had higher numbers of mutations in unvaccinated control sequences, however, ORF3a (56%) and ORF8 (67%) had higher numbers of sites with enriched mutations in breakthrough sequences. These results indicate that beyond the spike S112L and nsp12 F192V mutations, ORF3 and ORF8 might be under selective immune pressure in Delta.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Delta cases increased at comparable rates in vaccinated and unvaccinated individuals and Delta breakthrough is associated with a distinct set of clinical and genomic factors.

We expanded our current analysis of recorded breakthrough infections in our cohort, by adding data from our previous study[14] that covered February to April 2021, for a total of 208 breakthrough and 1329 control sequences (**Fig. 5**). The pattern of total COVID-19 case numbers in New York State over time is comparable to our SARS-CoV-2 sequences, supporting our attempt to sequence a representative set of viruses (**Fig. 5a, b**). While increasing vaccination rates were associated with an overall decline in COVID-19 cases and SARS-CoV-2 sequences obtained, emerging Delta infections partially reverted this trend (**Fig. 5a-c**). Iota breakthrough significantly decreased and Alpha breakthrough exhibited no significant change or slightly decreased over time since vaccination, whereas Delta breakthrough infections rose significantly in near linear fashion within the first six months post full vaccination, both in absolute and relative counts (**Fig. 5d, S7-S9**).

To assess whether Delta infections are statistically enriched among breakthrough infections compared to other variants, we performed logistic regression analyses and matched the 208 vaccinated breakthrough patients 1:1 to the 1329 unvaccinated patients, correcting for the confounding variables clinical collection date, sex, and age. While unmatched analyses suggested elevated numbers of Delta infections in vaccinated compared to unvaccinated individuals, the adjusted analysis in matched participants of the study cohort revealed no significant difference between both groups (**Fig. S8, Tables S7, S8**). Recorded Delta cases increased significantly compared to all other variants starting in May 2021 (**Table S8**). Notably, the rise of Delta occurred at comparable rates in vaccinated in unvaccinated individuals (**Fig.**

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

5c), indicating no major advantage of Delta *per se* in conferring breakthrough after adjustment for confounding variables.

In summary, the study of Delta infections revealed that, compared to unvaccinated controls, breakthrough infections were associated with asymptomatic disease, lower rates of hospitalization, and weakly also with spike mutation S112L, nsp12 mutation F192V, and the absence of seven mutations across different regions of the SARS-CoV-2 full genome (**Fig. 5e**).

Discussion

Since emergency-use authorization of the first COVID-19 vaccine in December 2020, the spotlight has been on viral sequences from vaccinated individuals to closely monitor the potential emergence of vaccine escape mutations. While the first half of 2021 was characterized by a wealth of variants, some more prevalent than others regionally or globally, the Delta variant has rapidly replaced all other variants worldwide[28]. Our cohort in metropolitan New York is a prime example of Delta's takeover, both in unvaccinated and vaccinated individuals with a near linear increase of Delta breakthrough infections according to elapsed time post vaccination (**Figs. 1, 5**). Breakthrough numbers in our cohort approach the absolute numbers in the unvaccinated, which is in line with other reports[29,30]. The reasons for this change are multifactorial, including epidemiological and virological factors. First, the probability of a positive case being a breakthrough versus an unvaccinated case has risen over time due to increased vaccination rates in our area and elsewhere (**Fig. 5**)[31]. Second, Delta has been detected in increased rates in younger hosts, who tend to be vaccinated at lower rates than older individuals, or were ineligible for vaccination at the time of our study. The younger

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

population is also associated with lower hospitalization rates (**Figs. 1, S1**)[32]. Third, Delta carries the P681R substitution in the spike furin cleavage site, which triggers enhanced S1/S2 cleavage and might explain Delta's increased replication rate *in vitro* [21,33,34]. A substitution at the same site (P681H) is found in Alpha, which evolved independently of Delta and dominated the global infection landscape between January and June 2021[35-39]. P681H/R mutations were both detected more frequently among breakthrough infections in our previous study of breakthrough infections in New York when Alpha and Iota variants were dominant regionally[14], and in this study during the current Delta wave (**Fig. 3**). Although Delta shows compromised sensitivity to some RBD and NTD neutralizing antibodies with up to 8-fold reduced sensitivity *in vitro* to vaccine-induced antibodies compared to D614G viruses (including infectious virus assays)[21,23,40], neutralization escape is substantially lower in magnitude as compared to Beta, Gamma, and Mu[41,42]. Efficient spike cleavage (P681H/R) and replicative competence appear to be central checkpoints for SARS-CoV-2's epidemiological success, with Delta being improved compared to Alpha[21,33,43-45].

In line with data from other laboratories, the pre-Delta breakthrough infections that we studied in early 2021 displayed a variant distribution similar to infections in unvaccinated individuals[14-17], though with signs of a starting sieve effect of neutralizing antibody escape mutations, e.g., E484K[12-14,46]. The variant pattern in both breakthrough and unvaccinated cases has dramatically changed towards the second half of 2021 with a clear dominance of Delta-associated mutations (**Fig. 3**). In the second half of 2021, breakthrough appears to be primarily shaped by virological factors that increased transmissibility, facilitated by immunological permeability in times of waning vaccine efficacy[47]. Delta's global rise in the second half of 2021 coincided when large

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

parts of high-risk populations were past six months post full vaccination. The lift of US mask recommendations in May 2021, right before the ignition of the Delta wave, may have also added to its spread. Mask mandates were reinstated halfway through the Delta wave, i.e., on July 28th 2021, at NYU Langone Health, whereas New York State only issued a recommendation to wear masks in early August[48].

Delta is a primary source for the evolution of next-generation variants, evidenced by the rapid diversification into more than 120 Delta subvariants. We observed an uptick of the subvariant AY.25 and the S112L mutation in NTD, which, while still at low numbers, preferentially spread among the vaccinated individuals in our cohort. In our phylogenetic analysis, we observed multiple Delta clusters that suggest efficient spread of distinct sublineages in the population. As shown for AY.25, they involve unvaccinated and vaccinated individuals, and their presence in clusters implies transmission chains.

High vaccine efficacy was reported against Delta infections, particularly at the beginning of the Delta wave; however, there is growing evidence that vaccine efficacy against Delta decreases over time[20,22,49-51]. A central question remains whether reduced vaccine efficiency is due to Delta *per se* or whether it is a matter of waning immunity and dependent on elapsed time since full vaccination. Current knowledge indicates that it may be both. Statistical analyses of retrospective cohort studies indicated that waning immunity with time, coinciding with a period of easing societal public health restrictions, played a larger role than Delta's vaccine escape[50,51]. However, an observational study on SARS-Cov-2-infected index cases and contacts indicated that vaccination reduced transmission more effectively in Alpha compared to Delta[22]. Our data suggest that selective adaptation processes might be in process that may eventually lead to the

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

accumulation of mutations (spike S112L, nsp12 F192V) or new (sub)variants (AY.25) under vaccine immune pressure (**Figs. 4** and **5e**). It remains elusive at this time whether spike S112L, located at the apical part of spike NTD, nsp12 F192V, located distant from RdRp's active site, or the other sites with differential mutation rates in breakthroughs play a role in immune evasion, enhanced transmission, or as bystander mutations.

As with all observational genomic surveillance studies, limitations and confounding factors exist. These include demographic, temporal, and behavioral factors, diagnostic testing practices, sequencing sensitivity thresholds, and sampling, e.g., undersampling of mild or asymptomatic cases. Some but not all of these factors can be adjusted or minimized by unbiased sample collection, consistent testing procedures and guidelines, and matched data analyses. Our adjusted model indicates that the probability of Delta infection has increased at similar rates in vaccinated and unvaccinated individuals. However, Delta breakthrough cases appear to rise continuously with extended time post vaccination (**Fig. 5**), which justifies the use of booster doses [52,53]. Monitoring Delta, its evolving subvariants, and newly emerging variants including their early signs of adaptive evolution will be critical to adequately address upcoming SARS-CoV-2 outbreaks and the increasing numbers of vaccine breakthrough infections.

Methods

Study design and sample collection. This study was approved by the NYULH Institutional Review Board, protocol numbers i21-00493 and i21-00561. SARS-CoV-2 infections were studied in vaccinated and unvaccinated individuals from May 1st until August 3rd in the NYULH system. Cases were identified using DataCore, the system's clinical data management and extraction

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

resource. Vaccine breakthroughs were defined by positive real-time (RT)-PCR test for SARS-CoV-2 RNA regardless of Ct at least 14 days after the second dose of BNT162b2 (Pfizer/BioNTech) or mRNA-1273 (Moderna) vaccines, or the single-dose COVID-19 Janssen vaccine. The unvaccinated control group consisted of all SARS-CoV-2 positive cases in our healthcare system who had not received a dose of any vaccine at the time of RT-PCR positivity, collected and sequenced in the same period as the breakthrough infections. Nasopharyngeal swabs were sampled from individuals with exposure to SARS-CoV2, suspected to have an infection with SARS-CoV-2, or as part of clinical diagnostics or hospital admission.

RNA extraction, library preparation and sequencing

Detailed methods were described recently[14] (**Supplemental Methods**). SARS-CoV-2 sequences that were <23,000 bp or <4000x genome coverage were considered inadequate and not included in the analyses.

Phylogenetic and mutation analyses

Methods were described recently[14] (**Supplemental Methods**). Briefly, maximum likelihood IQ trees were produced on Mafft-aligned SARS-CoV-2 full genome sequences (1000 bootstrap replicates). Spike amino acid counts and site-specific mutation frequencies were analyzed using Fisher exact tests with multiplicity corrections (Benjamini-Hochberg)[54].

Statistical analysis

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The comparison of 132 breakthrough and 283 unvaccinated control samples achieved 96% power in detecting a 15% difference (10% versus 25%) in mutation rates, clinical or demographic variables in a two-tailed chi squared/Fisher Exact test with a type I error of 5% (G*Power v.3.1.9.4)[55]. We evaluated the variant distribution (Alpha, Delta, Gamma, Iota, and all other variants) in fully vaccinated compared to unvaccinated individuals and addressed confounding variables arising from the use of observational data via matching and adjustment. We matched the vaccinated breakthrough cases 1:1 to the unvaccinated cases, correcting for the confounding variables: clinical collection date, sex, and age[56]. Quality controls of matching included the analysis of propensity score distributions and empirical quantile-quantile (eQQ) plots of distribution balance of co-variates using the R v.4.1.0 MatchIt package. To compare the probability of a Delta variant for the vaccinated and unvaccinated groups over time, we performed logistic regression analyses on the full data adjusting for sex, age (centered and standardized), and month of test. The relationship between variant distribution and time since vaccination was studied using linear regression analyses in Prism v.8.4.3 using two-sided Pearson tests. We compared clinical and demographic variables between groups using non-parametric Mann-Whitney tests or Kruskal-Wallis tests with Dunn's multiplicity correction in Prism. Correlation analyses were done using two-sided Spearman rank tests in Prism. For all analyses unless otherwise stated, a two-sided Type I error rate of 0.05 was applied. Details are described in **Supplemental Methods**.

Data availability

All sequences are publicly available in SRA (BioProject PRJNA769411) and GISAID (**Tables S1-S3**).

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Author contributions:

Ralf Duerr, MD, PhD designed the study, analyzed data, created figures, and wrote the manuscript; Dacia Dimartino, PhD and Paul Zappile made libraries and sequenced the samples; Christian Marier analyzed genomic data; Samuel Levine and Fritz Francois, MD, generated total numbers of vaccinations for the entire healthcare system; Guiqing Wang, MD, PhD collected samples and performed clinical SARS-CoV-2 detection; Eduardo Iturrate, MD, and Brian Elbel, PhD monitored epidemiological data and generated lists of breakthrough infections; Meike Dittmann, PhD wrote the manuscript; Jennifer Lighter, MD reviewed clinical information; Andrea Troxel, ScD and Keith S. Goldfeld, DrPH performed statistical analyses; and Adriana Heguy, PhD designed the study, generated genomic data and wrote the manuscript. All authors reviewed and edited the manuscript.

Acknowledgements. We thank NYU Langone Health DataCore for support extracting the data for this study from clinical databases and the clinical laboratory technicians for assistance in testing, saving and retrieving specimens, especially Joanna Fung. We also thank Dr. Joan Cangiarella for her continuous support of genomic surveillance for SARS-CoV-2 at NYULH, including providing institutional funding for this study. The authors wish to acknowledge the software support of New York University's Data Services, Bobst Library, and, in particular, the expertise shared by Christopher Schwarz and Senior Academic Technology Specialist Denis Rubin. We are grateful to the submitting laboratories who deposited data in GISAID, in

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

particular to those whose sequences we used to create the phylogenetic tree (**Table S3**). The NYULH Genome Technology Center is partially supported by the Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center.

References

1. World Health Organization (WHO). WHO Coronavirus (COVID-19) Dashboard. Vol. 2021 (2021).
2. Centers for Disease Control and Prevention (CDC). SARS-CoV-2 Variant Classifications and Definitions. Vol. 2021 (2021).
3. GISAID. GISAID database. Vol. 2021 (2021).
4. Baden, L.R., *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med* **384**, 403-416 (2021).
5. Polack, F.P., *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* **383**, 2603-2615 (2020).
6. Sadoff, J., *et al.* Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. *N Engl J Med* (2021).
7. Walls, A.C., *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292 e286 (2020).
8. Wrapp, D., *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).
9. Dagan, N., *et al.* BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. *N Engl J Med* (2021).
10. Haas, E.J., *et al.* Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet* **397**, 1819-1829 (2021).
11. Harris, R.J., *et al.* Effect of Vaccination on Household Transmission of SARS-CoV-2 in England. *N Engl J Med* **385**, 759-760 (2021).
12. Kustin, T., *et al.* Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2-mRNA-vaccinated individuals. *Nat Med* (2021).
13. McEwen, A.E., *et al.* Variants of Concern Are Overrepresented Among Postvaccination Breakthrough Infections of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Washington State. *Clinical Infectious Diseases* (2021).
14. Duerr, R., *et al.* Dominance of Alpha and Iota variants in SARS-CoV-2 vaccine breakthrough infections in New York City. *J Clin Invest* **131**(2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

15. Pollett, S.D., *et al.* The SARS-CoV-2 mRNA vaccine breakthrough infection phenotype includes significant symptoms, live virus shedding, and viral genetic diversity. *Clin Infect Dis* (2021).
16. CDC Covid- Vaccine Breakthrough Case Investigations Team. COVID-19 Vaccine Breakthrough Infections Reported to CDC - United States, January 1-April 30, 2021. *MMWR Morb Mortal Wkly Rep* **70**, 792-793 (2021).
17. Bergwerk, M., *et al.* Covid-19 Breakthrough Infections in Vaccinated Health Care Workers. *N Engl J Med* (2021).
18. Chemaitelly, H., *et al.* Waning of BNT162b2 Vaccine Protection against SARS-CoV-2 Infection in Qatar. *N Engl J Med* (2021).
19. Levin, E.G., *et al.* Waning Immune Humoral Response to BNT162b2 Covid-19 Vaccine over 6 Months. *N Engl J Med* (2021).
20. Fowlkes, A., *et al.* Effectiveness of COVID-19 Vaccines in Preventing SARS-CoV-2 Infection Among Frontline Workers Before and During B.1.617.2 (Delta) Variant Predominance - Eight U.S. Locations, December 2020-August 2021. *MMWR Morb Mortal Wkly Rep* **70**, 1167-1169 (2021).
21. Mlcochova, P., *et al.* SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* (2021).
22. Eyre, D.W., *et al.* The impact of SARS-CoV-2 vaccination on Alpha & Delta variant transmission. *medRxiv*, 2021.2009.2028.21264260 (2021).
23. Planas, D., *et al.* Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276-280 (2021).
24. Liu, C., *et al.* Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* **184**, 4220-4236 e4213 (2021).
25. Lucas, C., *et al.* Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity. *Nature* (2021).
26. Lopez Bernal, J., *et al.* Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *New England Journal of Medicine* **385**, 585-594 (2021).
27. Harvey, W.T., *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology* **19**, 409-424 (2021).
28. Nextstrain team. Genomic epidemiology of novel coronavirus - Global subsampling. Vol. 2021 (2021).
29. Brown, C.M., *et al.* Outbreak of SARS-CoV-2 Infections, Including COVID-19 Vaccine Breakthrough Infections, Associated with Large Public Gatherings - Barnstable County, Massachusetts, July 2021. *MMWR Morb Mortal Wkly Rep* **70**, 1059-1062 (2021).
30. Keehner, J., *et al.* Resurgence of SARS-CoV-2 Infection in a Highly Vaccinated Health System Workforce. *N Engl J Med* **385**, 1330-1332 (2021).
31. Choi, A., Velasquez, J. & Welch, W. Coronavirus in New York City - Tracking the spread of the pandemic. Vol. 2021 (The City, 2021).
32. Lam-Hine, T., *et al.* Outbreak Associated with SARS-CoV-2 B.1.617.2 (Delta) Variant in an Elementary School - Marin County, California, May-June 2021. *MMWR Morb Mortal Wkly Rep* **70**, 1214-1219 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

33. Lubinski, B., *et al.* Spike protein cleavage-activation mediated by the SARS-CoV-2 P681R mutation: a case-study from its first appearance in variant of interest (VOI) A.23.1 identified in Uganda. *bioRxiv* (2021).
34. Peacock, T.P., *et al.* The SARS-CoV-2 variants associated with infections in India, B.1.617, show enhanced spike cleavage by furin. *bioRxiv*, 2021.2005.2028.446163 (2021).
35. Martin, D.P., *et al.* The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages. *Cell* **184**, 5189-5200.e5187 (2021).
36. Collier, D.A., *et al.* Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* (2021).
37. Starr, T.N., *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310 e1220 (2020).
38. Gu, H., *et al.* Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **369**, 1603-1607 (2020).
39. Meng, B., *et al.* Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the variant of concern lineage B.1.1.7. *Cell Reports*.
40. Chia, P.Y., *et al.* Virological and serological kinetics of SARS-CoV-2 Delta variant vaccine-breakthrough infections: a multi-center cohort study. *medRxiv*, 2021.2007.2028.21261295 (2021).
41. Miyakawa, K., Jeremiah, S.S., Kato, H. & Ryo, A. Neutralizing efficacy of vaccines against the SARS-CoV-2 Mu variant. *medRxiv*, 2021.2009.2023.21264014 (2021).
42. Uriu, K., *et al.* Ineffective neutralization of the SARS-CoV-2 Mu variant by convalescent and vaccine sera. *bioRxiv*, 2021.2009.2006.459005 (2021).
43. Liu, Y., *et al.* Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. *bioRxiv*, 2021.2008.2012.456173 (2021).
44. Saito, A., *et al.* SARS-CoV-2 spike P681R mutation, a hallmark of the Delta variant, enhances viral fusogenicity and pathogenicity. *bioRxiv*, 2021.2006.2017.448820 (2021).
45. Zhang, J., *et al.* Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant. *bioRxiv* (2021).
46. Feder, K.A., *et al.* Association of E484K spike protein mutation with SARS-CoV-2 infection in vaccinated persons---Maryland, January - May 2021. *Clin Infect Dis* (2021).
47. Thomas, S.J., *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine through 6 Months. *New England Journal of Medicine* (2021).
48. Matthews, K. & Hill, M. Masks encouraged in New York, but no city or state mandate. Vol. 2021 (AP News, 2021).
49. Lopez Bernal, J., *et al.* Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *N Engl J Med* **385**, 585-594 (2021).
50. Rosenberg, E.S., *et al.* New COVID-19 Cases and Hospitalizations Among Adults, by Vaccination Status - New York, May 3-July 25, 2021. *MMWR Morb Mortal Wkly Rep* **70**, 1150-1155 (2021).
51. Tartof, S.Y., *et al.* Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study. *The Lancet*.
52. Bar-On, Y.M., *et al.* Protection of BNT162b2 Vaccine Booster against Covid-19 in Israel. *New England Journal of Medicine* (2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

53. World Health Organization (WHO). Interim statement on booster doses for COVID-19 vaccination. Vol. 2021 (2021).
54. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
55. Faul, F., Erdfelder, E., Lang, A.G. & Buchner, A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* **39**, 175-191 (2007).
56. D'Agostino, R.B., Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* **17**, 2265-2281 (1998).

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Tables.

Table 1. Demographic and clinical data and SARS-CoV-2 lineages of fully vaccinated vs non-vaccinated COVID-19 cases with full SARS-CoV-2 genomes

	Fully Vaccinated	Unvaccinated
Age		
Median	37	42
Range	22-90	<1-96
Age Groups*		
<20	0	39 (13.8)
20-29	31 (23.5)	45 (15.9)
30-39	42 (31.8)	49 (17.3)
40-49	13 (9.8)	31 (10.9)
50-59	19 (14.4)	45 (15.9)
60-69	13 (9.8)	31 (10.9)
70-79	8 (6.1)	18 (6.4)
80-89	5 (3.8)	21 (7.4)
>90	1 (0.7)	4 (1.4)
Sex		
Male	67	133
Female	65	155
Vaccine Type		
Pfizer	115	Not applicable
Moderna	14	Not applicable
Janssen	3	Not applicable
Days after vaccination		
Median	136	Not applicable
Range	21-221	
<120	49	Not applicable
>120	83	Not applicable
Hospitalized	8 (6)	83 (29)
Symptomatic	72 (54.5)	207 (84.1)[#]
SARS-CoV-2 variant		
Alpha	13 (9.8)	68 (24.0)
Iota	7 (5.3)	35 (12.4)
Gamma	4 (3.0)	19 (6.7)
Delta	101 (76.5)	139 (49.1)
Other	7 (5.3)	22 (7.8)
Total Cases	132	283

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

All figures are absolute numbers except numbers in parenthesis are % from total cases.

[#] information available for 246/283 unvaccinated individuals.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure legends.

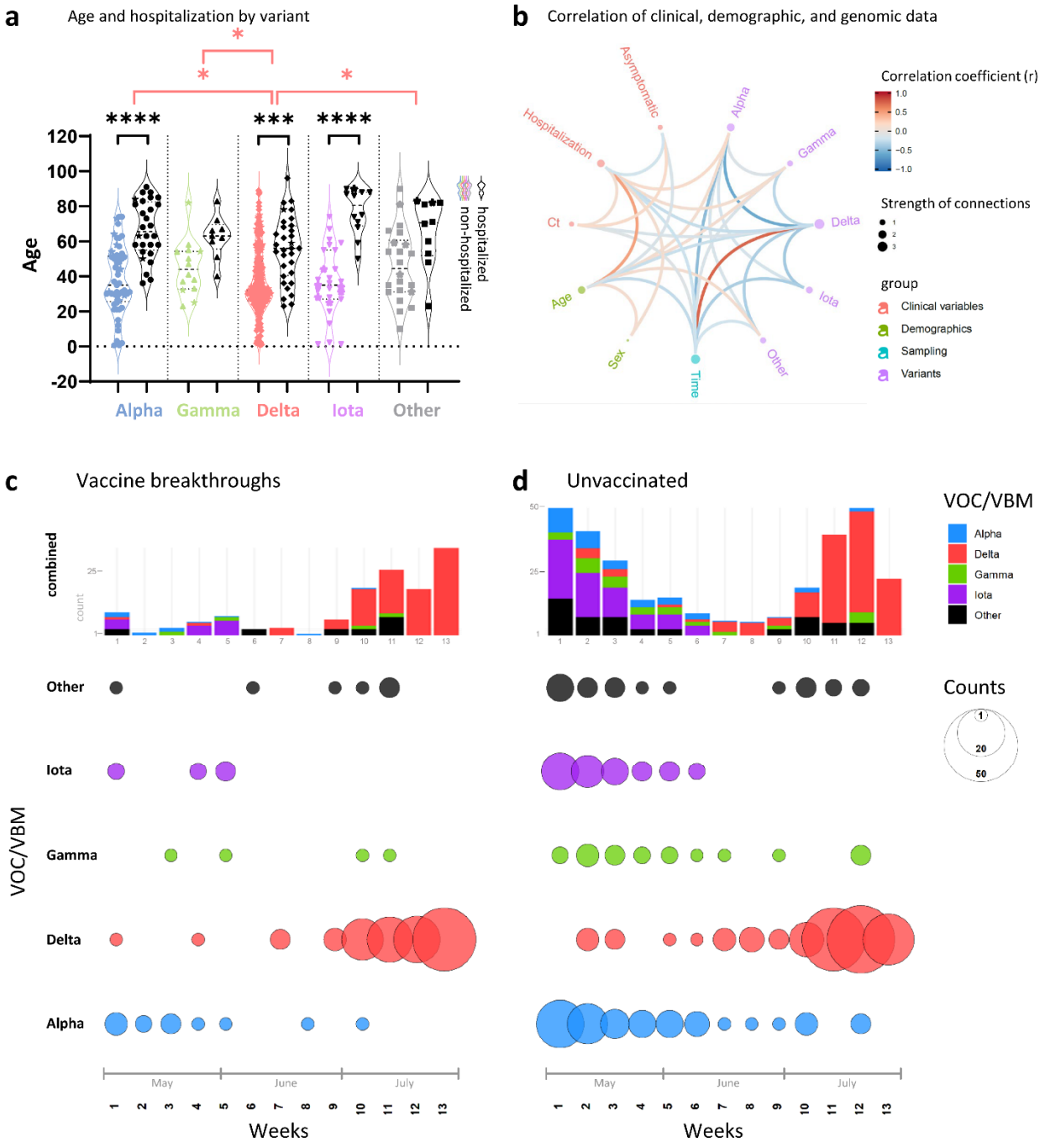


Figure 1. Relationship of clinical, demographic, and genomic data. a, Violin plot summarizing the age distribution of the combined data set of 132 vaccinated and 283 unvaccinated SARS-CoV-2-positive individuals by variant. The pairs of colored and black violins show non-

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

hospitalized versus hospitalized cases per variant. Horizontal lines indicate the median and interquartile ranges of values. Breakthrough cases are shown as stars. Statistical comparisons of age were made using Kruskal-Wallis tests between variants (red brackets) and, for each variant, between non-hospitalized and hospitalized cases (black brackets). All statistically significant results are shown: * P<0.05, ** P<0.01, *** P<0.001, **** P<0.0001. **b**, Correlation analysis of clinical, demographic, and genomic data of SARS-CoV-2 infected individuals (as in A). Red and blue edges represent positive and negative correlations between connected variables, respectively, according to the scale of r values to the right. Only significant correlations (P<0.05, Spearman rank test) are displayed. Nodes are color-coded based on the grouping of variables. Node size corresponds to the strength of correlations. Ct: Cycle threshold in RT-PCR; **s**ex: male sex; time: date of sampling. **c, d**, Distribution of variants (top) and absolute variant counts (bottom), colored and separated into to the most frequently detected variants of concern (VOC) and being monitored (VBM) and all other variants combined (Other). Vaccine breakthrough infections (**c**) are shown side-by-side with unvaccinated controls (**d**) on a weekly basis starting May 1st 2021 (all full weeks shown).

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

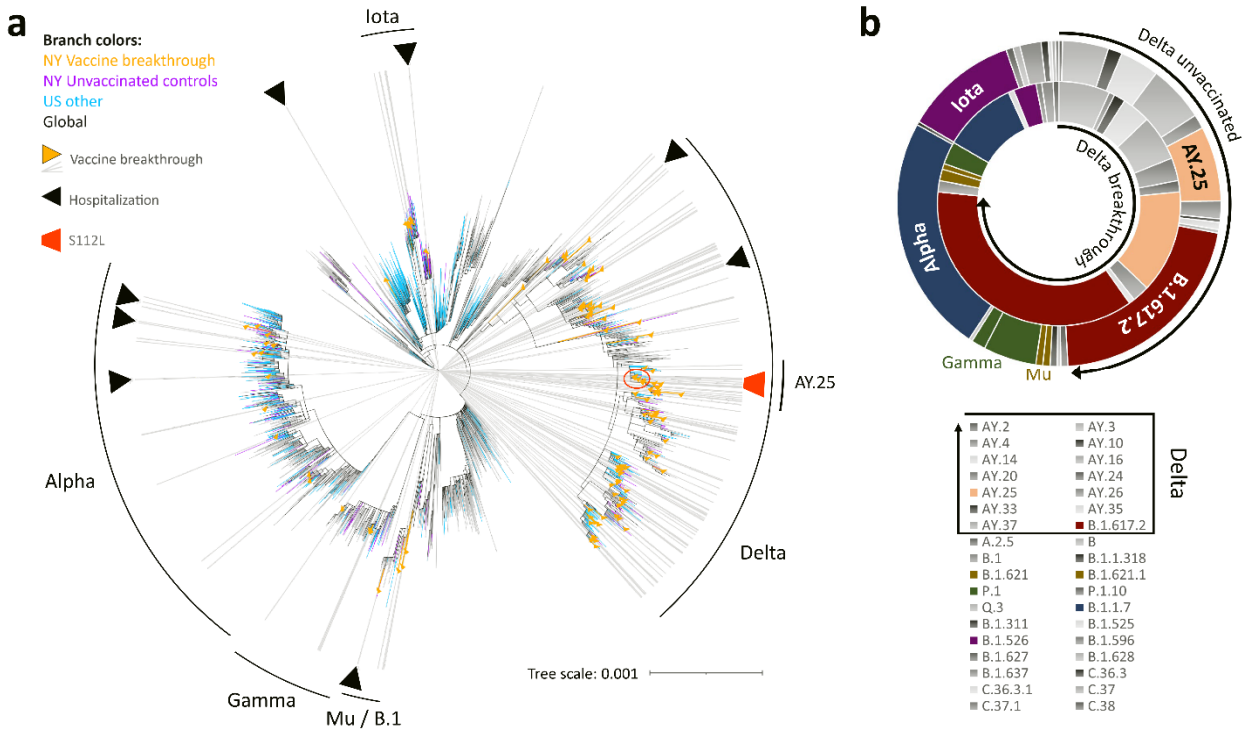


Figure 2. Phylogenetic analysis and variant distribution of SARS-CoV-2 vaccine breakthrough and unvaccinated control sequences.

a, Maximum likelihood (IQ) tree of 3511 SARS-CoV-2 full genome sequences (base pairs 202-29,666 according to Wuhan-Hu-1 as reference), including 132 vaccine breakthrough (orange) and 283 unvaccinated control SARS-CoV-2 sequences from the NYU Langone Health cohort (greater NYC area) (purple) together with 920 other US (non-NYU; cyan) and 2176 global (non-US; black) reference sequences. The substitution scale of the tree, generated with 1000 bootstrap replicates and Wuhan/WH01/2019-12-26 as root, is indicated at the bottom right. Vaccine breakthrough sequences are highlighted by orange triangles (as branch symbols) and gray rays radiating from the root to the outer rim of the tree. Hospitalizations among vaccine breakthrough infections are indicated by black triangles. The variants responsible for most vaccine breakthrough infections are labeled. The Delta plus S112L sub-lineage is encircled and

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

highlighted with a red trapezoid symbol. **b**, Double-donut plot to compare the variant distribution of breakthrough (inner ring) and unvaccinated control sequences (outer ring). The most abundant variants and Delta subvariants (highlighted by black arrows) are shown in color and labelled in the plot (outer ring only). The color code for all detected variants (Pango lineages) is provided below.

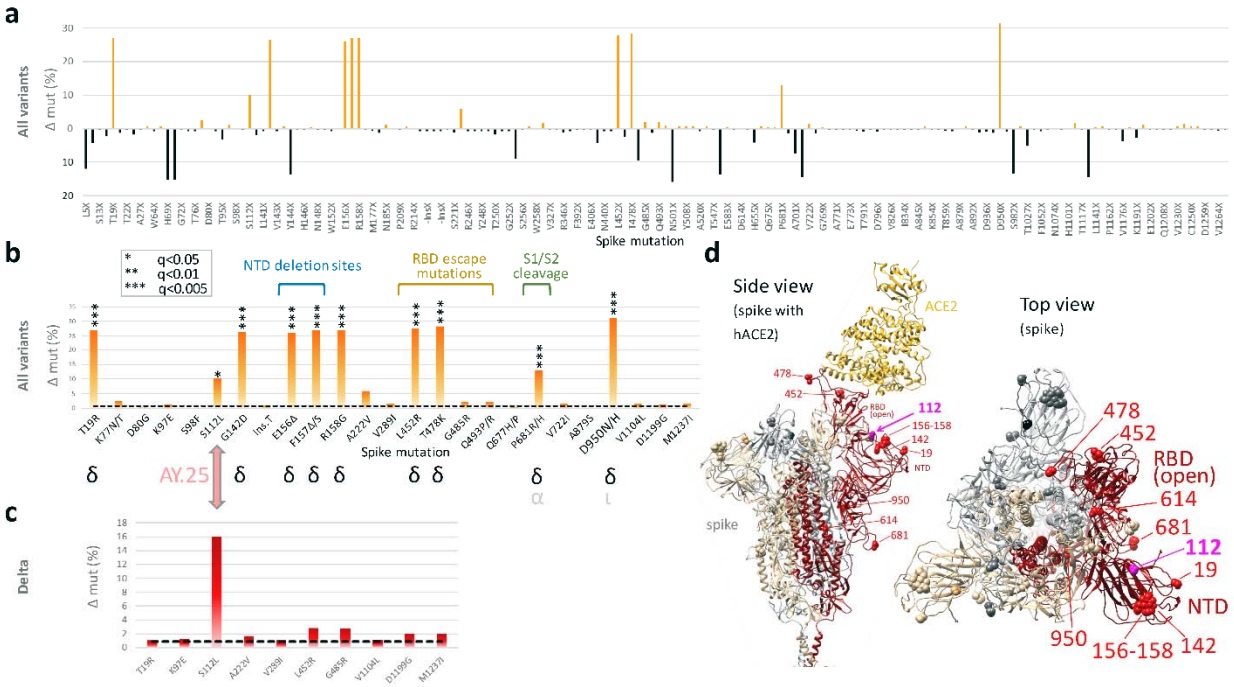


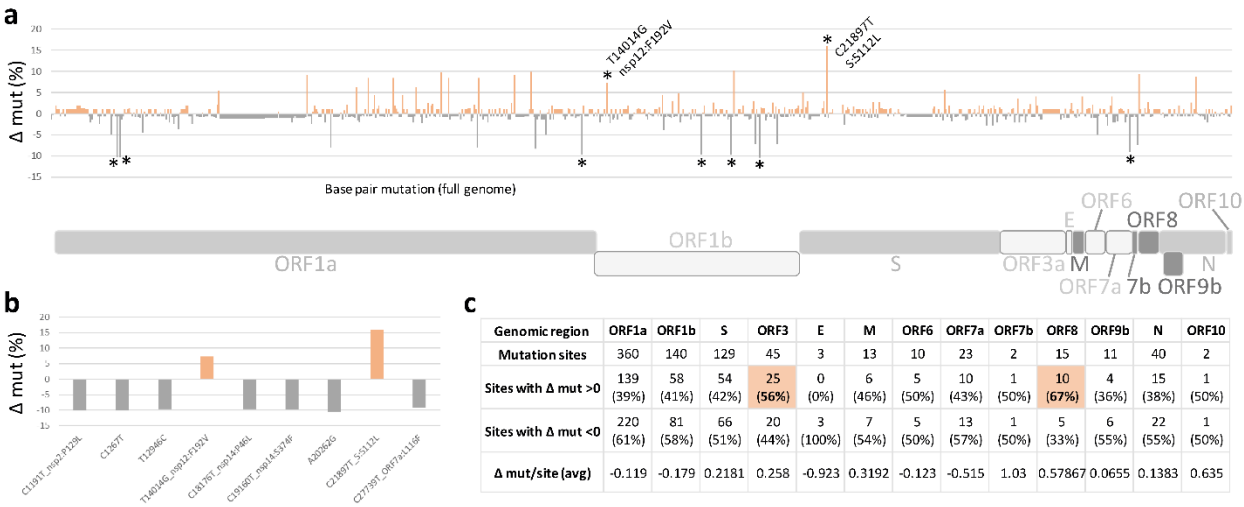
Figure 3. Site-specific spike mutation analysis in SARS-CoV-2 vaccine breakthrough sequences compared to unvaccinated controls.

a, Site-specific amino acid mutation (mut) frequencies in spike in 132 vaccine breakthrough sequences compared to 283 unvaccinated controls from the same cohort. The Wuhan-Hu-1 sequence served as reference. The mirror plot displays differences of mutation frequencies per spike residue between vaccinated and unvaccinated groups, shown along the x-axis (n=168);

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

orange (facing up) and black bars (facing down) refer to elevated mutation frequencies in vaccinated or unvaccinated individuals, respectively. **b**, Enriched spike mutations in vaccine breakthrough sequences compared to unvaccinated controls. Unique occurrences of mutations in breakthrough cases were disregarded. The dashed black line indicates the average mutation frequency across all spike residues in the unvaccinated control data set (n=283) compared to Wuhan-Hu-1. Significantly enriched mutations in Fisher exact tests are indicated by asterisks (* P<0.05, *** P<0.005) and the variants in which these mutations were found are shown below (black: main source, gray: secondary source). Mutations in the spike N-terminal domain (NTD), receptor binding domain (RBD), and near the S1/S2 interface associated with neutralization escape and/or affecting important biological functions are labeled. **c**, The same analysis as in (**b**) but focusing on Delta sequences exclusively. 101 Delta vaccine breakthrough sequences were compared to 139 Delta unvaccinated controls. The dashed black line indicates the average mutation frequency across all spike residues in the Delta unvaccinated control data set compared to Wuhan-Hu-1 as reference (n=139). **d**, Structural analysis of mutation sites on a spike trimer bound to human ACE2 (hACE2) (pdb S_ACE2). Each protomer is colored differently. The hACE2-bound protomer with the RBD in the “up” position is shown in red. Statistically enriched mutation sites in vaccine breakthroughs (according to **b**) are shown as spheres, labeled in one protomer in red (Delta) or pink (AY.25).

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



Genomic region	ORF1a	ORF1b	S	ORF3	E	M	ORF6	ORF7a	ORF7b	ORF8	ORF9b	N	ORF10
Mutation sites	360	140	129	45	3	13	10	23	2	15	11	40	2
Sites with Δ mut >0	139 (39%)	58 (41%)	54 (42%)	**25 (56%)**	0 (0%)	6 (46%)	5 (50%)	10 (43%)	1 (50%)	**10 (67%)**	4 (36%)	15 (38%)	1 (50%)
Sites with Δ mut <0	220 (61%)	81 (58%)	66 (51%)	20 (44%)	3 (100%)	7 (54%)	5 (50%)	13 (57%)	1 (50%)	5 (33%)	6 (55%)	22 (55%)	1 (50%)
Δ mut/site (avg)	-0.119	-0.179	0.2181	0.258	-0.923	0.3192	-0.123	-0.515	1.03	0.57867	0.0655	0.1383	0.635

Figure 4. Full genome mutation analysis in SARS-CoV-2 Delta vaccine breakthrough sequences compared to Delta unvaccinated controls.

a, Site-specific base pair mutation frequencies in full genomes (bp 202-29666) of 101 Delta vaccine breakthrough sequences compared to 139 Delta unvaccinated controls from the same cohort. The Wuhan-Hu-1 sequence served as reference. The mirror plot displays differences of mutation frequencies per site between vaccinated and unvaccinated groups, shown along the x-axis (n=791); orange (facing up) and gray bars (facing down) refer to elevated mutation rates in vaccinated or unvaccinated individuals, respectively. Significantly enriched mutations in Fisher exact tests are indicated by asterisks (* P<0.05), and mutations that were enriched in the vaccinated group are also labeled. SARS-CoV-2 coding genomic regions are shown below the plot. **b**, Bar plot summarizing all enriched mutations from (**a**). **c**, Differential full genome mutation analysis of Delta infections in vaccinated and unvaccinated individuals by SARS-CoV-2 genomic region. Regions in which mutation sites were more frequently mutated in vaccinated compared to unvaccinated individuals are highlighted in orange and percentages shown in bold.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Bp: base pairs; E: envelope; mut: mutation; nsp: non-structural protein; N: nucleocapsid; ORF: open reading frame; S: spike.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

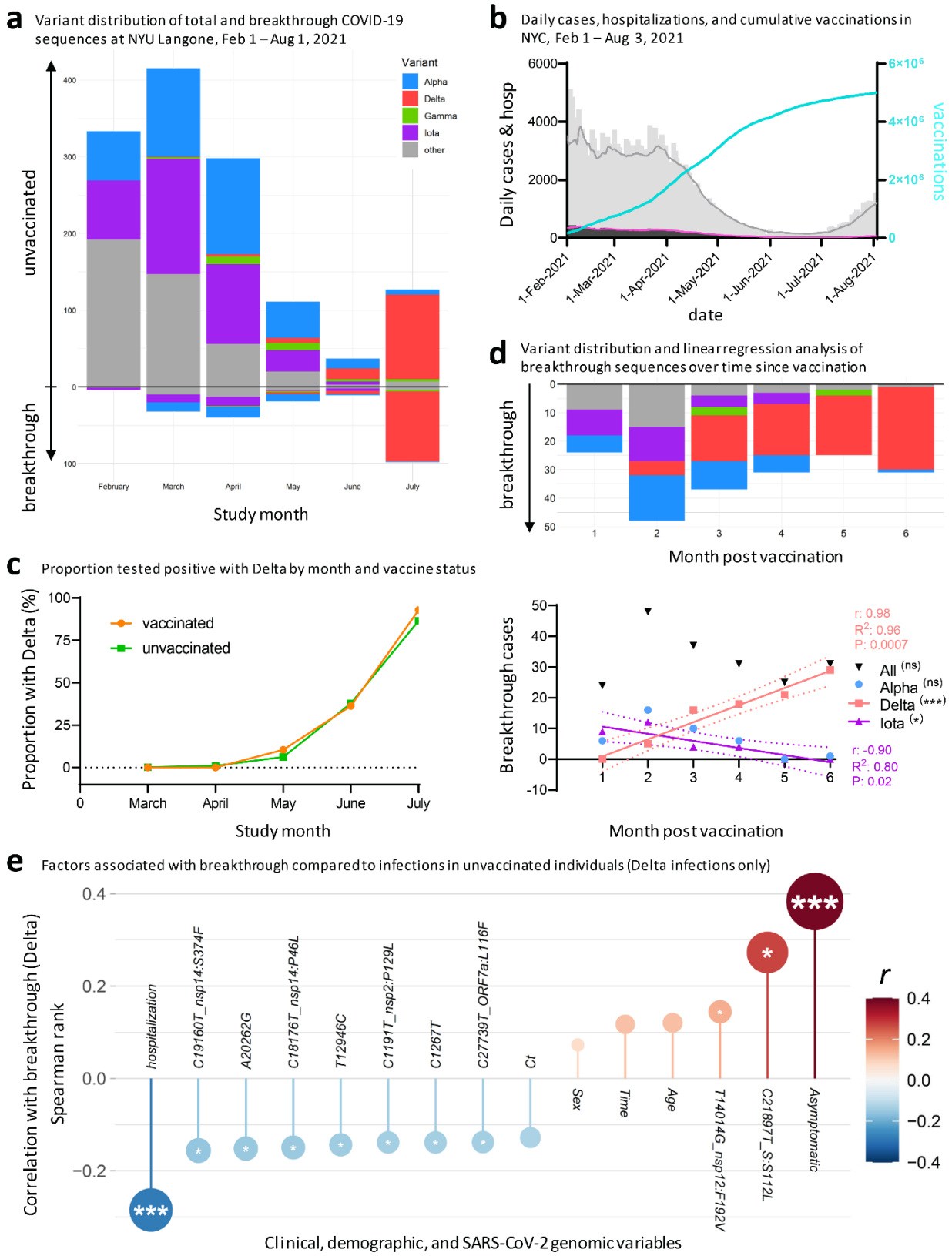


Figure 5. Increasing rates of SARS-CoV-2 breakthrough infections with Delta over six months

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

post vaccination, and clinical and genomic factors associated with Delta breakthrough.

a, Variant distribution of SARS-CoV-2 sequences from unvaccinated (facing up) and vaccinated individuals (facing down), obtained at NYU Langone Health between February and July 2021. **b**, Daily COVID-19 cases (gray bars) and 7-day averages (gray line), daily COVID-19 deaths (black bars) and 7-day averages (pink line), and cumulative vaccination numbers (turquoise line) in New York City between February 1st and August 3rd. Source of data: NYC Open Data and NYC Health, Citywide Immunization Registry (CIR). **c**, Probability of positive test with Delta by month in vaccinated and unvaccinated individuals, adjusted for month of test, sex, and age of participants. **d**, Variant distribution among all NYULH breakthrough sequences, displayed according to months post full vaccination (starting at day 14 after the last dose for full vaccination). The chart below shows a linear regression analysis of breakthrough infections per variant against time post vaccination. Significant results are highlighted by asterisks, labeled with the correlation coefficient (r), goodness of fit (R^2), and P value, and the fitted line with 95% confidence intervals shown. * $P<0.05$, ** $P<0.01$, *** $P<0.005$. **e**, Correlation of clinical, demographic, and SARS-CoV-2 genomic factors with breakthrough by comparing Delta infections in vaccinated (n=101) and unvaccinated individuals (n=139). Spearman rank correlation is displayed on the y axis and color-coded. Multiple comparison-corrected P values (q, Benjamini-Hochberg) are indicated by asterisks within the circles (* $q<0.05$, ** $q<0.01$, *** $q<0.005$). Ct: Cycle threshold in RT-PCR; sex: male sex; time: date of sampling.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Please see Supplementary Information for:

Supplemental Methods.

RNA extraction, cDNA synthesis, library preparation and sequencing

Phylogenetic and mutation analyses

Visualization

Statistical analysis

External data sources

Supplemental Figures.

Figure S1. Hospitalizations and demographic data by variant in breakthrough cases and unvaccinated controls.

Figure S2. Asymptomatic infections and demographic data by variant in breakthrough cases and unvaccinated controls.

Figure S3. Relationship of clinical, demographic, and genomic data in breakthrough cases and unvaccinated controls.

Figure S4. Phylogenetic analysis of SARS-CoV-2 Delta sequences.

Figure S5. Spike mutation patterns in vaccine breakthrough sequences.

Figure S6. Site-specific spike mutation analysis in SARS-CoV-2 Delta vaccine breakthrough sequences compared to Delta infections in unvaccinated controls.

Figure S7. Ratios of breakthrough SARS-CoV-2 infections by variant and time post vaccination

Figure S8. Breakthrough SARS-CoV-2 infections by variant and time post vaccination.

medRxiv preprint doi: https://doi.org/10.1101/2021.12.07.21267431; this version posted December 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure S9. SARS-CoV-2 variant distribution in vaccinate breakthrough and unvaccinated cases and quality control of matching.

Supplemental Tables.

Table S1. Full specifications of 132 SARS-CoV-2 breakthrough infections.

Table S2. Full specifications of 283 SARS-CoV-2 infections in unvaccinated controls.

Table S3. Metadata of GISAID reference sequences used in phylogenetic analyses.

Table S4. Spike mutation statistics of sites with enriched mutation rates in vaccine breakthrough infections compared to unvaccinated controls.

Table S5. Delta variant spike mutation statistics of sites with enriched mutation rates in Delta vaccine breakthrough infections compared to Delta infections in unvaccinated controls.

Table S6. Delta variant full genome mutation statistics of all sites with significantly different mutation rates in Delta vaccine breakthrough infections compared to Delta infections in unvaccinated controls.

Table S7. Contingency tables and chi-square tests of unmatched and 1:1 matched data from vaccinated and unvaccinated SARS-CoV-2^{+} participants.

Table S8. Estimated model comparing the probability of a positive Delta test for the vaccinated/unvaccinated groups, adjusted for sex, age (centered and standardized), and month of test.