

## RESEARCH ARTICLE

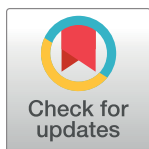
# Risk of rapid evolutionary escape from biomedical interventions targeting SARS-CoV-2 spike protein

Debra Van Egeren<sup>1,2,3</sup>, Alexander Novokhodko<sup>4</sup>, Madison Stoddard<sup>5</sup>, Uyen Tran<sup>5</sup>, Bruce Zetter<sup>6</sup>, Michael Rogers<sup>6</sup>, Bradley L. Pentelute<sup>7</sup>, Jonathan M. Carlson<sup>8</sup>, Mark Hixon<sup>9</sup>, Diane Joseph-McCarthy<sup>10</sup>, Arijit Chakravarty<sup>5\*</sup>

**1** Department of Systems Biology, Harvard Medical School, Boston, MA, United States of America, **2** Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, United States of America, **3** Stem Cell Program, Boston Children's Hospital, Boston, MA, United States of America, **4** Department of Mechanical Engineering, University of Washington, Seattle, WA, United States of America, **5** Fractal Therapeutics, Cambridge, MA, United States of America, **6** Vascular Biology Program, Boston Children's Hospital, Boston, MA, United States of America, **7** Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, United States of America, **8** Microsoft Research, Redmond, WA, United States of America, **9** Mark S. Hixon Consulting, LLC, San Diego, CA, United States of America, **10** Boston University, Boston, MA, United States of America

☞ These authors contributed equally to this work.

\* [arijit@fractal.tx](mailto:arijit@fractal.tx)



## OPEN ACCESS

**Citation:** Van Egeren D, Novokhodko A, Stoddard M, Tran U, Zetter B, Rogers M, et al. (2021) Risk of rapid evolutionary escape from biomedical interventions targeting SARS-CoV-2 spike protein. PLoS ONE 16(4): e0250780. <https://doi.org/10.1371/journal.pone.0250780>

**Editor:** Mahmoud Naguib, Uppsala University, SWEDEN

**Received:** February 9, 2021

**Accepted:** April 14, 2021

**Published:** April 28, 2021

**Copyright:** © 2021 Van Egeren et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting Information](#) files.

**Funding:** D.V.E. acknowledges funding from the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard, award number #1764269, and the Harvard Quantitative Biology Initiative. A. N. acknowledges funding from the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1762114. Any opinions, findings, and

## Abstract

The spike protein receptor-binding domain (RBD) of SARS-CoV-2 is the molecular target for many vaccines and antibody-based prophylactics aimed at bringing COVID-19 under control. Such a narrow molecular focus raises the specter of viral immune evasion as a potential failure mode for these biomedical interventions. With the emergence of new strains of SARS-CoV-2 with altered transmissibility and immune evasion potential, a critical question is this: how easily can the virus escape neutralizing antibodies (nAbs) targeting the spike RBD? To answer this question, we combined an analysis of the RBD structure-function with an evolutionary modeling framework. Our structure-function analysis revealed that epitopes for RBD-targeting nAbs overlap one another substantially and can be evaded by escape mutants with ACE2 affinities comparable to the wild type, that are observed in sequence surveillance data and infect cells *in vitro*. This suggests that the fitness cost of nAb-evading mutations is low. We then used evolutionary modeling to predict the frequency of immune escape before and after the widespread presence of nAbs due to vaccines, passive immunization or natural immunity. Our modeling suggests that SARS-CoV-2 mutants with one or two mildly deleterious mutations are expected to exist in high numbers due to neutral genetic variation, and consequently resistance to vaccines or other prophylactics that rely on one or two antibodies for protection can develop quickly -and repeatedly- under positive selection. Predicted resistance timelines are comparable to those of the decay kinetics of nAbs raised against vaccinal or natural antigens, raising a second potential mechanism for loss of immunity in the population. Strategies for viral elimination should therefore be diversified across molecular targets and therapeutic modalities.

conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Fractal Therapeutics provided support in the form of salaries for authors A.C., M.S., and U.T., but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: A.C., M.S., and U.T. are employees and shareholders of Fractal Therapeutics. D.V.E., A.N., B.Z., and D.J.-M. are shareholders of Fractal Therapeutics. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

## Introduction

The deployment of vaccines against SARS-CoV-2 brings the question of mutational escape from antibody prophylaxis to the forefront. Rapid evolutionary evasion of neutralizing antibodies (nAbs) poses a number of threats to biomedical interventions aimed at bringing the virus under control, namely the risk of reduced vaccinal efficacy over time as resistant variants continue to emerge (which may or may not be rectifiable with annual vaccine updates), the risk of waning effectiveness of natural immunity as a result of evasion of common nAbs, and the risk of antibody-dependent enhancement (ADE).

SARS-CoV-2 is commonly considered to acquire mutations more slowly than other RNA viruses [1,2]. However, the SARS-CoV-2 mutation burden and evolutionary rate ( $1 \times 10^{-3}$  substitutions per base per year [2]) have only been estimated under conditions of neutral genetic drift (distinct from antigenic drift) [3], in the absence of strong positive selection pressure provided by population-level immunity or other interventions that select for resistance mutations. In immunologically naïve COVID-19 patients, viral load and transmission [4] peak near the time of symptom onset, while the host antibody response peaks approximately 10 days later [5]. Thus, transmission in immunologically naïve individuals occurs well in advance of the appearance of a robust humoral response. These kinetics suggest the immune response in naïve individuals exerts limited selection pressure on the virus, consistent with direct genetic evidence from deep sequencing showing little to no positive selection [6]. Hence, the evolutionary rate prior to the widespread deployment of vaccines or development of natural immunity (based primarily on neutral genetic drift) may underestimate the evolutionary potential of the virus to evade nAbs deployed as active immunity (vaccines) or passive immunity (nAb prophylactics). When nAbs are broadly present in the population, population-level selection for antibody-evading, infection-competent viral mutants may result in a rapid resurgence of SARS-CoV-2 infections.

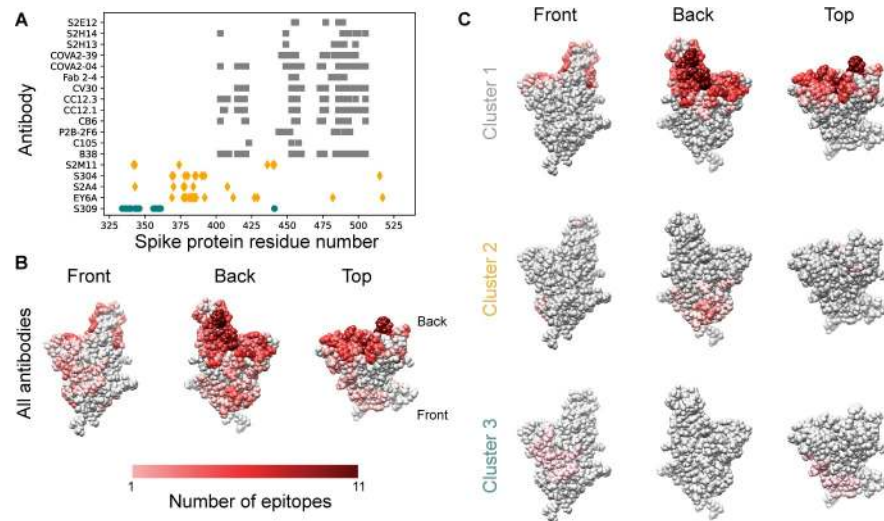
Mutation rates alone offer a limited picture of the ability of viruses to generate successful escape mutations. While some vaccine-preventable viruses have very low mutation rates (such as smallpox,  $\sim 1 \times 10^{-6}$  sub/nuc/yr [7]), others have high mutation rates (such as poliovirus,  $1 \times 10^{-2}$  sub/nuc/yr) (S1 Table). There is a sharp contrast between the high antigenic evolvability of viruses such as influenza [8], notable for their evolutionary capacity for immune evasion, and the low antigenic evolvability of viruses like poliovirus, which have proven highly tractable to antibody-mediated prophylaxis via vaccines [9] despite a high evolutionary rate (S1 Table). Studies of other infectious diseases support the concept that natural selection promotes antigenic evolvability [10].

To better understand the potential for immune evasion mediated by SARS-CoV-2 RBD mutations in the presence of nAbs, singly or in combination, we focused on three questions. First, what is the evolutionary cost of harboring nAb-evading RBD mutations? Second, given this evolutionary cost, how likely is it that SARS-CoV-2 patients will harbor viruses with pre-existing nAb-evading RBD mutations as their dominant viral sequence? Third, how rapidly will such nAb-evading RBD mutants become fixed in the population once nAb vaccines and therapies are deployed widely?

## Results

### There is a low evolutionary cost to developing resistance to RBD-targeting nAbs

To explore the diversity of the B-cell response against the RBD, we catalogued the reported spike RBD epitopes recognized by the natural human immune response. Consistent with prior



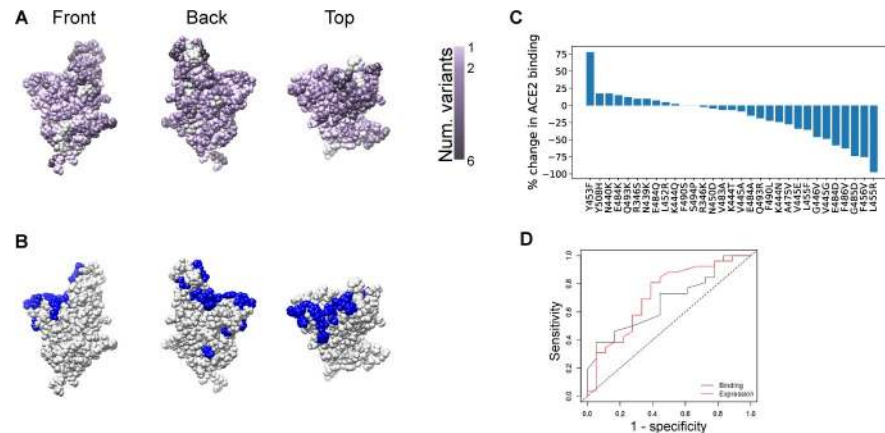
**Fig 1. Epitopes for antibodies targeting the spike protein RBD overlap substantially.** A. Contact residues for spike protein RBD antibody epitopes. Colors and symbols denote antibody clusters: Grey squares: Cluster 1, yellow diamonds: Cluster 2, green circles: Cluster 3. B. RBD structure with each residue colored by the number of antibody epitopes including it, compiled from PDB data. C. RBD structure, colored by the number of antibody epitopes that each residue is part of, by epitope cluster.

<https://doi.org/10.1371/journal.pone.0250780.g001>

work [11], we found that the reported epitopes show substantial overlap (Fig 1A and 1B). Clustering revealed three clusters representing distinct immunogenic sites on the RBD, the largest of which overlaps substantially with the ACE2 binding interface (Fig 1A and 1C). These clusters resemble those reported by other groups [12]. There was limited evidence for glycosylation in these epitope clusters (S1 Fig). The observed overlap in residues included in epitopes from independently-generated natural human antibodies shows that parts of the RBD surface are repeatedly targeted by the human B-cell response in different individuals. Spontaneous mutations at these key epitope residues could render many nAbs ineffective.

Genomic sequencing of SARS-CoV-2 from infected individuals has revealed several point mutations in the RBD, some of which have been shown experimentally to confer resistance to nAbs. As of 8/18/20, multiple amino acid changes have been reported in the GISAID sequence database [13] in RBD residues within antibody epitopes (Fig 2A), showing that SARS-CoV-2 antibody binding region variants are capable of causing human infection. Some of these naturally-occurring variants confer *in vitro* resistance to SARS-CoV-2 nAbs (Fig 2B) [14–16]. Additionally, 23 of the escape mutations experimentally identified by Starr et al [17] have been reported in the GISAID database, many of which do not compromise spike-ACE2 binding when examined *in vitro* (Fig 2C). This suggests that escape mutants that evade nAb binding have a low evolutionary cost. In fact, many antibodies have escape mutants that have increased ACE2 binding affinities (S2 Table).

To further understand the evolutionary cost of escape mutations, we evaluated the link between ACE2 binding affinity and/or RBD expression and viral infectivity (Methods). We determined how well changes in RBD binding to ACE2 or RBD expression caused by a mutation [17] predicted a 10% loss of infectivity as measured by luciferase reporter pseudoviral assay [15]. The low-to-moderate level of sensitivity and specificity of ACE2 binding affinity and RBD expression as predictors of pseudoviral infectivity suggest that changes in ACE2 binding affinity and RBD expression are well-tolerated in many immune-evading mutants. This further provides the virus with a range of possibilities for generating mutants that can evade nAbs without compromising infectivity (Fig 2D).



**Fig 2. The spike protein RBD tolerates mutations that confer resistance to one or more nAbs.** **A.** Spike protein RBD structure, with each residue colored by the number of distinct amino acid changes present in the GISAID sequencing database. **B.** RBD structure with residues at which mutations have been shown to confer escape from antibody neutralization marked in blue. **C.** Experimentally-measured effects of immune escape mutations on ACE2 binding, as taken from [17]. **D.** ROC curve showing the low predictive value of ACE2 binding measurements (grey) and expression (red) for *in vitro* infectivity of SARS-CoV-2 mutants. Area under the curve (AUC) is 0.67 for ACE2 binding as a predictor of infectivity and 0.72 for RBD expression as a predictor of infectivity.

<https://doi.org/10.1371/journal.pone.0250780.g002>

Taken together, the narrow focus of the immune response on a specific region of the RBD (Fig 1), the immunodominance of the spike protein [12], and the ability of RBD nAb escape mutations to yield viable and infectious viral particles capable of ACE2 binding (Fig 2) suggest a low evolutionary cost for the virus in generating escape mutants for nAbs.

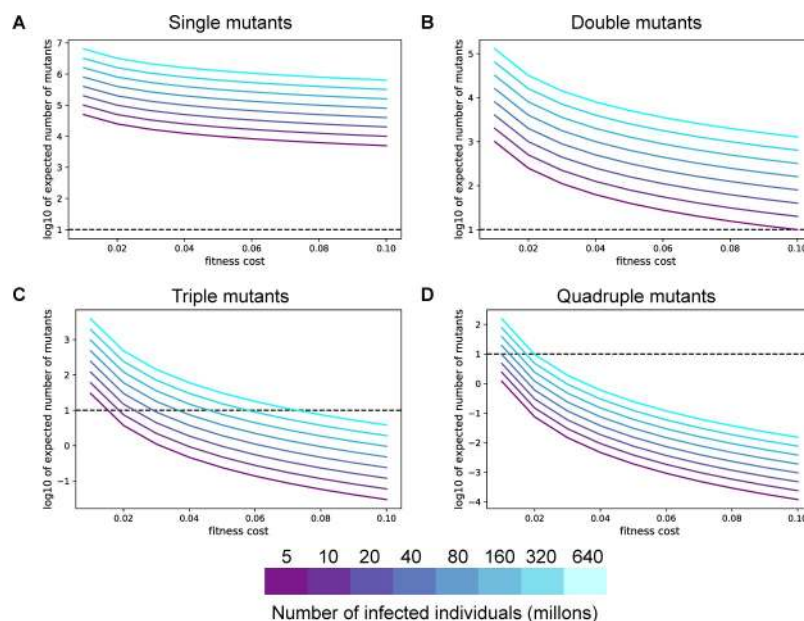
### Mutant frequency under neutral drift is likely to lead to escape from single and double antibody combinations

Based on this assessment, we used evolutionary theory to predict the frequency of immune escape mutants in the population both before and after the widespread presence of nAbs due to vaccines, passive immunization or natural immunity. Before immunity or antibody prophylaxis is widely established in the population, there is no transmission advantage for viruses with immune escape mutations since most people are equally susceptible to infection from wild-type and mutant SARS-CoV-2. Instead, these mutations may have a small evolutionary fitness cost due to negative effects on ACE2 binding affinity or other factors, similar to the observed fitness cost of drug resistance mutations in HIV [18] and consistent with results suggesting that much of the SARS-CoV-2 genome is under weak purifying selection [19]. Indeed, many point mutations modestly reduce the ability of SARS-CoV-2 to infect cells *in vitro*, which could lead to reduced host-host transmission [16]. Although these mutants are at a fitness disadvantage compared to the wild-type virus before nAbs are broadly present in the population, they are constantly generated through *de novo* mutation which allows them to exist at nonzero frequencies. However, once nAbs are common in the population, these mutants will have a selective advantage. If they already exist at sufficient frequency in the population, the escape mutants will expand deterministically and lead to widespread SARS-CoV-2 resistance to nAbs.

Using mathematical modeling methods developed to study intrahost evolutionary dynamics during HIV infection, we calculated the expected number of infected individuals whose dominant viral sequence harbors one or more mildly deleterious immune-evading mutations under drift conditions (referred to as “mutants”, see Methods for details) [20]. Most reported

nAbs are susceptible to at least one single-nucleotide change resulting in evasion [17], suggesting that a single point mutation may correspond to the evasion of one antibody in a combination. This model predicts the frequency of such mutants using the mutation rate of the virus, the typical fitness cost to the virus from an immune escape mutation, and the total number of infected individuals. We estimated the per base per transmission mutation rate of SARS-CoV-2 from population phylodynamic studies to be between  $1 \times 10^{-5}$  to  $1 \times 10^{-4}$  (Methods) [2]. Many nAbs are evaded by multiple distinct point mutations, so the per-transmission rate of generating a new mutant that evades a particular neutralizing antibody can be more than an order of magnitude higher than the per base mutation rate [21]. We investigated a range of infected population sizes (from 5 million to 640 million) and a range of transmission fitness costs for each mutation before the widespread presence of nAbs.

The expected number of SARS-CoV-2-infected individuals whose dominant viral sequence harbors one or two immune escape point mutations is high enough to eventually lead to widespread resistance to nAbs. Over a range of fitness costs and assuming a population of at least 5 million active infections (which we note is vastly lower than estimates as of 12/28/2020), we predicted that over 10,000 SARS-CoV-2-infected individuals worldwide would harbor a dominant viral sequence capable of evading one antibody (Fig 3A). This number far exceeds the threshold number of individuals required for natural selection and not neutral genetic drift to drive evolution (dashed lines in Fig 3). Assuming an immune escape mutant will eventually have a fitness advantage of 0.1, corresponding to approximately 14% of the population receiving an effective prophylactic impacted by this mutation (S2 Fig), positive selection will allow an escape mutant to expand and eventually outcompete the wild-type virus if 10 or more individuals are infected with this variant [22]. More than one nucleotide change may be required to confer resistance to an antibody combination if it contains more than one antibody with distinct escape mutation profiles. If a specific two-mutation combination is required for



**Fig 3. SARS-CoV-2 mutants with one or two mildly deleterious mutations are expected to exist at high numbers.** A-D. The expected number of individuals infected with a specific single (A), double (B), triple (C), or quadruple (D) SARS-CoV-2 mutant viruses at different values of the fitness cost. For all panels, the colors denote the total number of individuals with active SARS-CoV-2 infection globally. The horizontal dashed line is the drift boundary calculated at a fitness benefit of 0.1 for the mutation combination.

<https://doi.org/10.1371/journal.pone.0250780.g003>

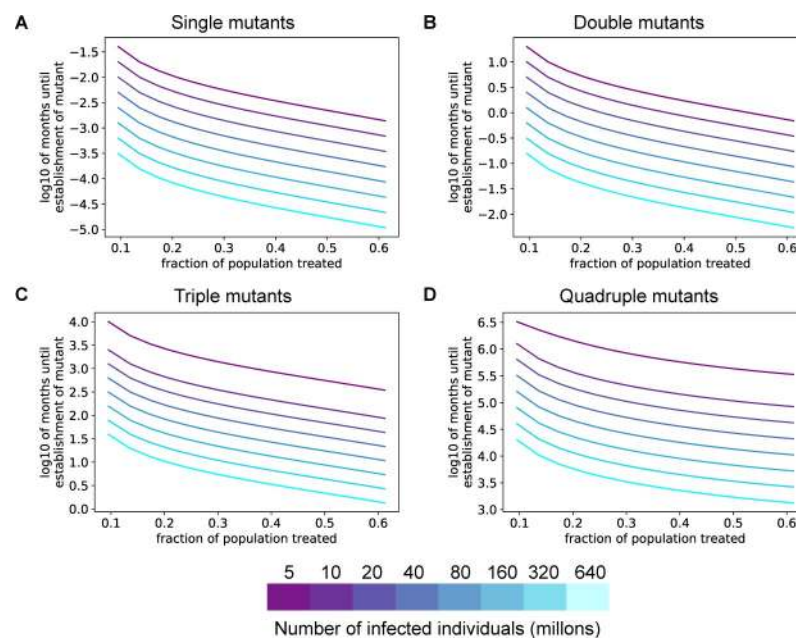


resistance, the expected number of infected individuals harboring a dominant viral sequence capable of evading the antibody combination can be orders of magnitude lower (Fig 3B). However, substantial double mutant populations (~hundreds of individuals) are expected if there are more than 50 million active infections worldwide (a plausible count as of 12/28/2020, see Methods for details).

If more than two mutations are required for a virus to escape nAbs, it is much less likely that population-level resistance will arise immediately. Each triple mutant is expected to be at appreciable frequencies only when the fitness cost of immune-evading mutations is lower than 0.04 (Fig 3C). Specific quadruple mutants are not expected to exist at significant frequencies in the population due to standing genetic variation alone for all but the lowest fitness costs (Fig 3D).

### Single and double resistance mutants are expected to establish quickly under selection

Even if a specific combination of mutations that confers resistance to an antibody combination is not present before the intervention is released, spontaneous mutation and positive selection will eventually lead to expansion of an escape mutant. To estimate how quickly population-level resistance to SARS-CoV-2 antibodies will emerge under natural selection, we modeled the acquisition of multiple mutations over time as a fitness valley-crossing problem (Methods). To acquire a specific combination of mutations that confers therapeutic resistance, the wild-type virus must transit through a valley of intermediate lower-fitness genotypes that have some, but not all, of the mutations required for immune escape. Previous theoretical expressions describing the time required to cross a fitness valley [23] were used to estimate the time needed for SARS-CoV-2 to acquire a given combination of one to four mutations (Fig 4). The



**Fig 4. Resistance to single or double antibody combinations will develop quickly under positive selection pressure.** A-D. Expected time to establishment of a successful single (A), double (B), triple (C), or quadruple (D) immune escape mutant assuming a per-site per-transmission mutation rate of  $1 \times 10^{-4}$ . The advantageous antibody resistant phenotype is acquired only after a specific combination of 1–4 mutations is present in the same virus. For all panels, the colors denote the total number of individuals with active SARS-CoV-2 infection. The fitness cost for each intermediate mutant is 0.05.

<https://doi.org/10.1371/journal.pone.0250780.g004>

time required for establishment of population-level resistance depends on how beneficial resistance is for virus transmission, and this benefit increases as more individuals in the population harbor nAbs against a given antigen. For antibodies or antibody combinations capable of being defeated by a single mutation, our modeling predicts the pre-existence of a resistant fraction before deployment of the intervention (Fig 4A). When examining double mutants, for infectious population sizes of 40 million or more, resistance to a widely deployed nAb combination will occur within months (Fig 4B). However, triple and quadruple mutation combinations will take much longer to establish in the population, even if nAbs are used widely and exert a strong selection pressure for these mutants (Fig 4C and 4D). These results hold under a range of intermediate fitness costs for viral mutants that harbor only a subset of the mutations required for escape (S3 Fig).

### Population-level resistance occurs more quickly at higher viral mutation rates

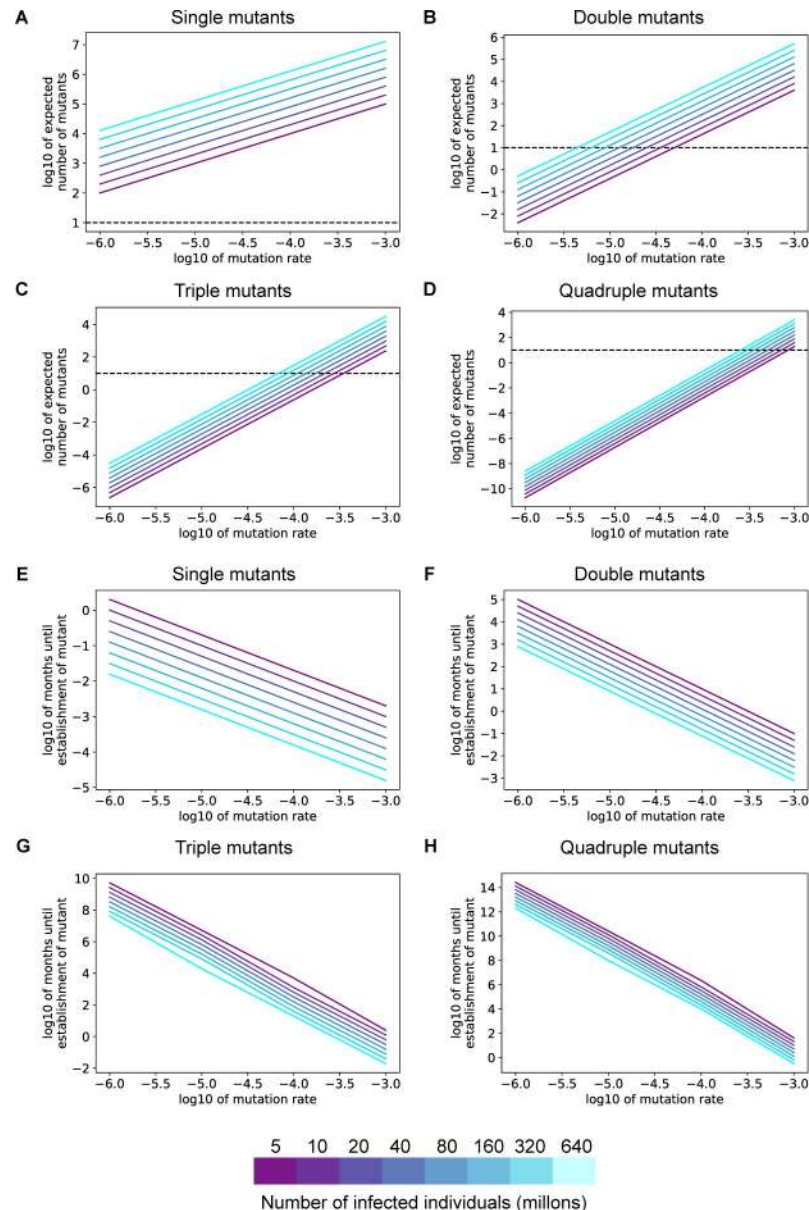
The SARS-CoV-2 mutation rate is a key parameter that determines how quickly the virus will acquire resistance to antibody interventions. While we estimated the per transmission rate of generating an antibody escape mutant at  $1 \times 10^{-4}$  (Methods), differences between antibody epitope sizes or changes in the mutation rate of the virus population over time [24] could influence this effective mutation rate. Our analysis revealed that many individuals would be infected with single or double SARS-CoV-2 mutants at a range of mutation rates greater than  $1 \times 10^{-5}$  (Fig 5A and 5B), while at higher mutation rates even triple and quadruple mutants will occur at sufficient frequencies to quickly establish in the population (Fig 5C and 5D). Similarly, we found that resistance to antibody combinations requiring two or fewer mutations for resistance would establish quickly after widespread presence of nAbs (Fig 5E and 5F). With a higher mutation rate ( $1 \times 10^{-3}$  per transmission), resistance could emerge against even combinations of nAbs that require the acquisition of 4 mutations (Fig 5H).

The juxtaposition of a relatively constrained immune response against the high degree of evolutionary plasticity of the spike RBD (visible even under neutral drift conditions) suggests that SARS-CoV-2 has extensive capacity to evolve to evade nAbs targeting a small number of antigenic regions. This capacity will negatively impact SARS-CoV-2 immunity in humans, whether active (vaccinal or natural) or passive (nAb prophylactics).

## Discussion

The work described in this paper points to the mutation tolerance of SARS-CoV-2 spike protein, placing this property in context with the mutation rates and pandemic sizes (number of active infections) to estimate the ease with which the virus will mutate to defeat combinations of neutralizing antibodies. Numerous COVID-19 antibody prophylactics and vaccines target the spike protein [25,26], and the immunodominance of the spike RBD in the natural immune response [12] implies that even vaccines that use live-attenuated or inactivated SARS-CoV-2 will rely to some extent on nAbs that target the RBD [27]. Thus, anticipating the viral population's response to widespread spike RBD-targeting nAbs has significant implications for SARS-CoV-2 prophylaxis.

The evolvability of SARS-CoV-2 spike protein RBD in the presence of nAbs depends on both the mutation rate in the presence of selection pressure and the mutational tolerance of the spike protein. The mutation rate of SARS-CoV-2 is in line with that of other single-strand RNA viruses, [28], and is relatively high when compared against some members of this group (such as Hepatitis C) for which evolution has practical clinical consequences (S1 Table) [29,30]. Mutation rates themselves are evolvable and may increase over time due to natural



**Fig 5. Resistance to single or double antibody combinations will develop quickly across a range of SARS-CoV-2 mutation rates.** A-D. The expected number of individuals infected with a specific single (A), double (B), triple (C), or quadruple (D) SARS-CoV-2 mutant viruses at different values of the per transmission mutation rate. E-H. Expected time to establishment of a successful single (E), double (F), triple (G), or quadruple (H) immune escape. The fitness benefit of resistance is 0.1. For all panels, the colors denote the total number of individuals with active SARS-CoV-2 infection. The fitness cost for each intermediate mutant is 0.05.

<https://doi.org/10.1371/journal.pone.0250780.g005>

selection [31]. A SARS-CoV-2 RNA dependent RNA polymerase (RdRp) variant that increases the mutation rate by two to five times has already been identified in some clinical isolates [24]. At the same time, our analyses point to a relatively high tolerance of the spike protein RBD for immune-evading mutations. Experimentally determined immune escape mutations bind host ACE2, in many cases with little to no loss of affinity relative to the wild-type (Fig 2C), and one prevalent RBD immune-evading mutation (N439K) has been shown experimentally to



enhance ACE2 binding and have similar *in vitro* replication fitness to wild-type virus [32]. Compromising spike RBD function (either through loss of ACE2-binding or expression levels) has a weak impact on *in vitro* infectivity (Fig 2D).

Our work suggests that it is likely that standing genetic variation alone has already produced a substantial population of viruses with single and double nucleotide changes that confer nAb resistance. These variants will establish quickly in the population under selection pressure. In fact, there is already a precedent for this behavior, as one such selective sweep occurred early on in the SARS-CoV-2 pandemic when the D614G mutation rose to nearly 80% frequency in under 6 months [33]. This mutation confers increased infectivity on the virus [34] and was readily generated in sufficient numbers to ensure its expansion. Additional selective sweeps have occurred in Europe (20A.EU1 variant) [35] and in the UK (VOC 202012/01 and B.1.1.7 variants) [36,37]. Of note, the B.1.1.7 variant has been shown to evade the 4A8 neutralizing antibody [36]. Additionally, the recent outbreak in minks of a variant with a combination of mutations that reduce antibody binding suggests that even variants with multiple mutations can be generated and are viable [38]. Currently, most of the SARS-CoV-2 genome is not under positive selection [19], but if nAbs are widely present in the population, mutations that confer resistance via immune evasion will expand rapidly under positive selection pressure. Evidence from multiple experimental studies showing that single RBD point mutations can lead to resistance [36] to neutralizing convalescent plasma from multiple donors [16,39,40] suggests that specific single mutants may be able to evade spike-targeting vaccinal immunity in many individuals and rapidly lead to spread of vaccine-resistant SARS-CoV-2. One variant that can escape convalescent plasma neutralization is already circulating in South Africa [41] and could experience greater positive selection pressure once vaccines are deployed widely.

This has implications for SARS-CoV-2 disease control strategies, as one possible solution to the problem of immune evasion by SARS-CoV-2 that has been proposed is to develop a new vaccine update every year, similar to influenza [42]. In practice, such a solution will only work in the face of a moderate pace of evolution of SARS-CoV-2 and a low degree of clonal diversity among various clades of SARS-CoV-2 as they evolve to evade the current crop of vaccines. Further, if within-host evolution of SARS-CoV-2 contributes to population-level immune evasion, the valley-crossing mechanism described in this paper could accelerate the emergence of vaccine-resistant strains in the months following vaccine deployment. To the extent that new strains of SARS-CoV-2 are antigenically distinct, this may also lead to increased risk of antibody-dependent enhancement (ADE), as one mechanism for ADE involves antibodies that bind to the pathogen but fail to neutralize it [43]. Finally, our work suggests that immune evasion requiring one to two mutations occurs within months, raising the prospect that this phenomenon will further shorten the duration of natural immunity, which is already limited by the relatively short duration of the humoral [44,45] and cellular [46] responses to SARS-CoV-2 infection. Further studies are required to understand the risk immune evasion poses to a strategy of annually updated vaccines.

Going forward, our work suggests strategies for designing SARS-CoV-2 prophylactics that are more resistant to viral evolution. First, nAbs should be used in combinations, preferably targeting more than two non-overlapping epitopes. Strategies for the design of prophylactic antibodies and vaccines should involve combining nAbs that bind to non-overlapping escape mutant regions, including those from smaller, distinct clusters outside the RBD. Alternatively, if antibodies from the same cluster are used, escape mutants must be carefully characterized to ensure they do not overlap [21]. Similarly, vaccines should be evaluated based on the number of SARS-CoV-2 point mutations required to disarm the neutralizing antibodies they generate.

Second, the evolutionary pressure on the virus will determine the speed at which resistance to nAbs emerges. The more widely a given epitope is targeted by biomedical intervention, and the more effective it is, the more rapidly it will generate resistance (Fig 4). This is a potential weakness of focusing on only a handful of vaccines (or epitopes) for global deployment. The effectiveness of nAb-based interventions for disease control will depend on how many different interventions are deployed, how many mutations are required to evade each intervention, and the extent to which their escape mutations overlap.

Finally, the overall size of the pandemic in terms of number of active infections will play a significant role in whether the virus can be brought under control with nAb prophylactics or vaccines. The speed at which nAb resistance develops in the population increases substantially as the number of infected individuals increases, suggesting that complementary strategies to prevent SARS-CoV-2 transmission that exert specific pressure on other proteins (e.g., antiviral prophylactics) or that do not exert a specific selective pressure on the virus (e.g., high-efficiency air filtration, masking, ultraviolet air purification) are key to reducing the risk of immune escape. In this context, vaccines that do not provide sterilizing immunity (and therefore continue to permit transmission) will lead to the buildup of large standing populations of virus [47], greatly increasing the risk of immune escape.

The evolvability of SARS-CoV-2 in response to selection pressure will determine the ultimate tractability of our efforts at disease control. Our work suggests that the capacity of SARS-CoV-2 to evade the immune system may be greater than originally anticipated and raises the specter of a process of ongoing and continuous evolution in response to antibody-based prophylaxis, occurring on a timescale that may not be convenient or tractable for the design of novel biomedical interventions. Thus, our findings speak to the need for both public health and biomedical intervention strategies targeting SARS-CoV-2 to be designed to account for the risk of rapid evolutionary response to biomedical interventions.

## Materials and methods

### Compilation of published neutralizing antibody epitopes

The authors performed a comprehensive search of all entries in the Protein Data Bank (PDB) [48] as of September 1<sup>st</sup>, 2020 which matched the criteria “Source Organism Taxonomy Name equals SARS-2 AND Source Organism Taxonomy Name equals Homo sapiens”. Structures were included if there were patient-derived nAbs present and the authors reported the binding residues. Structures were excluded if they were not patient-derived, if they were not nAbs, or if the authors did not report the binding epitope because the resolution was too low to identify it precisely. Additionally, other epitopes that met the inclusion criteria but were not found in the PDB were included on an ad hoc basis. The structures used are listed in S3 Table, along with the methods by which they were acquired.

After the search was completed, it was determined that there were too few epitopes reported outside of the RBD to attempt clustering in those residues. Thus, the clustering analysis was limited to the RBD. The RBD was defined as in [17].

### Clustering of antibody epitopes

The antibodies were clustered using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm [49]. The distance function used to compare two epitopes was the median distance between residues in the epitopes. The maximum distance between two epitopes in one cluster was set at 30 residues. The minimum number of epitopes per cluster was set at 2. Epitopes classified as noise were assigned to their own clusters. These parameters were

set based on a subset of 9 antibodies and used for all subsequent clustering. Clustering was done using Python version 3.7 with the scikit-learn package [50].

### ACE2 binding affinity and RBD predictivity for pseudoviral infectivity

Tolerability of SARS-CoV-2 to changes in ACE2 binding affinity or RBD expression were examined based on a receiver operating characteristic (ROC) analysis. Based on publications measuring the impact of 44 spike RBD mutations on the *in vitro* infectivity of SARS-CoV-2 pseudovirus, we examined the predictivity of changes in RBD binding to ACE2 or expression according to [17] for greater than 10% loss of infectivity as measured by luciferase reporter pseudoviral assay [15]. These mutants were either observed circulating in the population in GISAID [16] or experimentally identified to confer escape from nAbs [17,51]. Predictivity of ACE2 binding affinity or RBD expression for *in vitro* infectivity was evaluated based on the ability to predict a 90% or greater infectivity relative to the wildtype (WT) strain.

### Expected number of mutants without positive selection

To investigate the emergence of SARS-CoV-2 resistance to nAbs, we modeled virus transmission dynamics using a modified deterministic susceptible-infectious-recovered (SIR) model with mutation. The simplest version of this model includes two viral genotypes, WT and mutant viruses with a specific single nucleotide change. At the population level there are four different compartments: susceptible individuals ( $S$ ), individuals infected with WT virus ( $I_0$ ), infected with mutant virus ( $I_1$ ), and recovered or resistant to infection from WT virus ( $R$ ). The number of individuals in each compartment obey the following differential equations:

$$\frac{dS}{dt} = -\beta S(I_0 + (1-s)I_1)$$

$$\frac{dI_0}{dt} = \beta S[(1-\mu)I_0 + \mu(1-s)I_1] - \delta I_0$$

$$\frac{dI_1}{dt} = \beta S[\mu I_0 + (1-\mu)(1-s)I_1] - \delta I_1$$

$$\frac{dR}{dt} = \delta(I_0 + I_1)$$

where  $\beta$  is the transmission rate of WT virus,  $\mu$  is the per transmission probability of mutation at a specific site,  $s$  is the fitness cost to transmission of the mutation, and  $\delta$  is the recovery rate from the infection. The infected compartments of this SIR model have the same mathematical description as the virus-infected cells in the intrahost virus dynamics model presented by Ribeiro and co-workers [20], assuming the birth and death rates of uninfected cells in the intrahost model is negligible. We also assume that the frequency of recovered individuals in the population is small enough so  $S$  and  $R$  can be treated as constant. Therefore, for the frequency of virus mutants present in the population at steady state before establishment of widespread immunity or vaccination, we used the frequency of individuals infected with a virus with a single mutation  $f_1 = \mu/s$ , which agrees with the expected frequency of single mutants under mutation-selection equilibrium [23]. As given in [20], the frequency of viral mutants with  $k$  mutations is

$$f_k = k! \left(\frac{\mu}{s}\right)^k.$$

## Expected time to development of population-level resistance under positive selection

If a substantial fraction of the population is immune to the WT virus (either due to lasting immunity after recovering from the infection, vaccination, or administration of therapeutic antibodies), viruses with antibody escape mutations will have a fitness advantage over the WT virus. This advantage comes from the ability of the mutant virus to infect individuals who are immune to the WT virus and depends on the fraction of the population who are immune. If a single mutant has the ability to infect recovered/resistant individuals, the SIR model equations for infected individuals change to

$$\frac{dI_0}{dt} = \beta[S(1 - \mu)I_0 + (S + R)\mu(1 - s)I_1] - \delta I_0$$

$$\frac{dI_1}{dt} = \beta[S\mu I_0 + (S + R)(1 - \mu)(1 - s)I_1] - \delta I_1$$

and the effective transmission rate for the mutant virus is given by  $\beta \frac{(S+R)}{S} (1 - s)$ . The selective advantage  $w$  of the mutant virus is therefore  $\frac{(S+R)}{S} (1 - s) - 1$ .

In order for a virus with one or more mutations that confers immune escape to expand deterministically due to positive selection and establish in the population, the variant must first be created through mutation of a single virion. Then, the mutant virus must infect enough individuals so that it is unlikely to go extinct due to stochastic drift. Assuming the total number of infected individuals  $N$  is constant, if a single mutation is sufficient to lead to immune escape, the time needed to establish an immune escape mutant is exponentially distributed with expected time  $1/N\mu w$  generations [23].

To estimate the time needed for establishment of double-, triple-, or higher-order mutants that confer immune escape, we adapted previously reported work on the dynamics of asexual populations crossing fitness valleys [23]. If  $k$  mutations are required for immune escape and all intermediates with less than  $k$  mutations have fitness cost  $s$ , the time  $\tau_k$  to establishment of the  $k$ -mutant was approximated as

$$\tau_k = \frac{1}{N\mu p_1} + \sum_{i=1}^{k-1} t_i + \frac{\gamma}{w},$$

where  $\gamma$  is Euler's constant,  $t_i \approx \ln\left(\frac{2}{\sqrt{\mu p_{i+1}}}\right)$ , and the probability  $p_i$  of an  $i$ -mutant to be successful was approximated as

$$p_i \approx \begin{cases} -s + \sqrt{s^2 + 4\mu p_{i+1}} & \text{for } i < k \\ w & \text{for } i = k \end{cases}.$$

This approximation holds for intermediate population sizes ( $N\mu < 1$ ). Following the argument for large populations in [23], when  $N\mu > 1$  we treated mutants with few mutations deterministically. We again used the results in [20] to estimate the frequency of  $i$ -mutants at steady state under mutation-selection equilibrium  $f_i$  for  $i \leq k$ . The minimum value of  $i$  for which the estimated  $i$ -mutant population size  $Nf_i$  was  $< 1/\mu$  was taken as the mutant with the most mutations that could be treated deterministically as a constant-sized population. Denoting this number of mutations as  $j$ , the modified expression for the expected time to escape for large

population sizes is

$$\tau_k = \frac{1}{Nf_j\mu p_1} + \sum_{i=1}^{k-j-1} t_i + \frac{\gamma}{w}.$$

### Evolutionary model parameter value selection

The SARS-CoV-2 infection length was set to 2 weeks, based on published estimates of infectious period length [52].

The effective rate of acquiring nucleotide substitutions that escape a given nAb was estimated to be  $1 \times 10^{-4}$  per transmission. The overall single nucleotide substitution rate has been estimated at approximately  $1 \times 10^{-3}$  per site per year from multiple phylogenetic analyses of global SARS-CoV-2 genomic sequences [2], which is a  $3.8 \times 10^{-5}$  per site per transmission mutation rate assuming a two-week infection generation time. However, since multiple different single nucleotide mutations have been shown to confer resistance to many nAbs [21], we estimated the effective mutation rate (defined as the per transmission rate of producing a mutation that generates resistance to a particular nAb) to be 2–3x higher than the rate of producing a particular nucleotide substitution. The effect of changing the mutation rate on the mutant frequency and escape time estimates is shown in Fig 5.

We assumed that, without selective pressure imposed by deployment of an intervention, mutant virus is less fit than wild-type virus and is not transmitted as effectively. A similar fitness cost is assumed to apply to mutant intermediates that only harbor a subset of mutations required for escape from an antibody combination. This fitness cost to viral transmission is difficult to directly measure, so we used a range of fitness costs from 0.01 to 0.1, corresponding to a 1–10% reduction in transmission rate for mutant viruses. These fitness costs are of similar magnitude to those measured for HIV drug resistance mutations in treatment-naïve patients [18] and are broadly justified by the limited impact of spike RBD mutations on ACE2 binding and the limited ability of ACE2 binding and expression to predict infectivity (Fig 2).

We estimated the total number of individuals infected with SARS-CoV-2 using the number of diagnosed cases. As of 11/8/20, the number of active diagnosed cases worldwide is 14 million [53], and the number of infections is expected to be 5–10 times the number of diagnosed cases, as determined by modeling [54] and seroprevalence studies [55].

### Supporting information

**S1 Fig. Glycosylation in SARS-CoV-2 spike protein RBD.** Glycosylated residues are marked in blue, while the remainder of residues are colored by the number of epitopes that contain the residue (red color bar).

(PDF)

**S2 Fig. Relationship between the fraction of the population that receive a prophylactic that is completely effective in preventing infection from wild-type virus and the strength of selection for an escape mutant.**

(PDF)

**S3 Fig.** Time required to establish a resistant viral single (A), double (B), triple (C), or quadruple (D) mutant with different fitness costs for intermediate mutants. In our model, viral variants with some, but not all, mutations required for resistance to an antibody intervention have a fitness cost (ranging from 1–9% less infectious). Increasing the fitness cost of these intermediates prolongs the time required for a resistant variant with a specific combination of 2–4 mutations (B–D) to establish in the population.

(PDF)



**S1 Table. Evolutionary rates of pathogenic RNA viruses.**

(PDF)

**S2 Table. Antibody escape mutants with highest ACE2 binding affinities.**

(PDF)

**S3 Table. Antibody structures used.**

(PDF)

**Author Contributions****Conceptualization:** Arijit Chakravarty.**Data curation:** Alexander Novokhodko, Madison Stoddard, Uyen Tran.**Formal analysis:** Debra Van Egeren, Alexander Novokhodko, Madison Stoddard.**Investigation:** Debra Van Egeren, Alexander Novokhodko, Madison Stoddard, Uyen Tran.**Methodology:** Debra Van Egeren, Alexander Novokhodko, Madison Stoddard, Arijit Chakravarty.**Software:** Debra Van Egeren.**Supervision:** Bruce Zetter, Michael Rogers, Bradley L. Pentelute, Jonathan M. Carlson, Mark Hixon, Diane Joseph-McCarthy, Arijit Chakravarty.**Visualization:** Debra Van Egeren, Alexander Novokhodko, Madison Stoddard, Uyen Tran.**Writing – original draft:** Debra Van Egeren, Alexander Novokhodko, Madison Stoddard, Arijit Chakravarty.**Writing – review & editing:** Debra Van Egeren, Alexander Novokhodko, Madison Stoddard, Uyen Tran, Bruce Zetter, Michael Rogers, Bradley L. Pentelute, Jonathan M. Carlson, Mark Hixon, Diane Joseph-McCarthy, Arijit Chakravarty.**References**

1. Zhao Z, Li H, Wu X, Zhong Y, Zhang K, Zhang Y-P, et al. Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol Biol.* 2004; 4. <https://doi.org/10.1186/1471-2148-4-21> PMID: [15222897](https://pubmed.ncbi.nlm.nih.gov/15222897/)
2. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution.* 2020; 83: 104351. <https://doi.org/10.1016/j.meegid.2020.104351> PMID: [32387564](https://pubmed.ncbi.nlm.nih.gov/32387564/)
3. CDC. How Flu Viruses Can Change. In: Centers for Disease Control and Prevention [Internet]. 15 Oct 2019 [cited 8 Nov 2020]. Available: <https://www.cdc.gov/flu/about/viruses/change.htm>.
4. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med.* 2020; 26: 672–675. <https://doi.org/10.1038/s41591-020-0869-5> PMID: [32296168](https://pubmed.ncbi.nlm.nih.gov/32296168/)
5. To KK-W, Tsang OT-Y, Leung W-S, Tam AR, Wu T-C, Lung DC, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *The Lancet Infectious Diseases.* 2020; 20: 565–574. [https://doi.org/10.1016/S1473-3099\(20\)30196-1](https://doi.org/10.1016/S1473-3099(20)30196-1) PMID: [32213337](https://pubmed.ncbi.nlm.nih.gov/32213337/)
6. Berrio A, Gartner V, Wray GA. Positive selection within the genomes of SARS-CoV-2 and other Coronaviruses independent of impact on protein function. *PeerJ.* 2020; 8: e10234. <https://doi.org/10.7717/peerj.10234> PMID: [33088633](https://pubmed.ncbi.nlm.nih.gov/33088633/)
7. Babkin IV, Babkina IN. The Origin of the Variola Virus. *Viruses.* 2015; 7: 1100–1112. <https://doi.org/10.3390/v7031100> PMID: [25763864](https://pubmed.ncbi.nlm.nih.gov/25763864/)

8. Thyagarajan B, Bloom JD. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *Elife*. 2014; 3. <https://doi.org/10.7554/eLife.03300> PMID: 25006036
9. Quadeer AA, Barton JP, Chakraborty AK, McKay MR. Deconvolving mutational patterns of poliovirus outbreaks reveals its intrinsic fitness landscape. *Nature Communications*. 2020; 11: 377. <https://doi.org/10.1038/s41467-019-14174-2> PMID: 31953427
10. Graves CJ, Ros VID, Stevenson B, Sniegowski PD, Brisson D. Natural selection promotes antigenic evolvability. *PLoS Pathog*. 2013; 9: e1003766. <https://doi.org/10.1371/journal.ppat.1003766> PMID: 24244173
11. Yuan M, Liu H, Wu NC, Lee C-CD, Zhu X, Zhao F, et al. Structural basis of a shared antibody response to SARS-CoV-2. *Science*. 2020; 369: 1119–1123. <https://doi.org/10.1126/science.abd2321> PMID: 32661058
12. Piccoli L, Park Y-J, Tortorici MA, Czudnochowski N, Walls AC, Beltramello M, et al. Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell*. 2020; 183: 1024–1042.e21. <https://doi.org/10.1016/j.cell.2020.09.037> PMID: 32991844
13. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges*. 2017; 1: 33–46. <https://doi.org/10.1002/gch2.1018> PMID: 31565258
14. Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science*. 2020; 369: 1014–1018. <https://doi.org/10.1126/science.abd0831> PMID: 32540904
15. Schmidt F, Weisblum Y, Muecksch F, Hoffmann H-H, Michailidis E, Lorenzi JCC, et al. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses. *J Exp Med*. 2020; 217. <https://doi.org/10.1084/jem.20201181> PMID: 32692348
16. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell*. 2020; 182: 1284–1294.e9. <https://doi.org/10.1016/j.cell.2020.07.012> PMID: 32730807
17. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Diggins AS, et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*. 2020; 182: 1295–1310. <https://doi.org/10.1016/j.cell.2020.08.012> PMID: 32841599
18. Kühnert D, Kouyos R, Shirreff G, Pečerska J, Scherrer AU, Böni J, et al. Quantifying the fitness cost of HIV-1 drug resistance mutations through phylodynamics. *PLoS Pathog*. 2018; 14. <https://doi.org/10.1371/journal.ppat.1006895> PMID: 29462208
19. Cagliani R, Forni D, Clerici M, Sironi M. Computational Inference of Selection Underlying the Evolution of the Novel Coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2. *Journal of Virology*. 2020; 94. <https://doi.org/10.1128/JVI.00411-20> PMID: 32238584
20. Ribeiro RM, Bonhoeffer S, Nowak MA. The frequency of resistant mutant virus before antiviral therapy. *AIDS*. 1998; 12: 461–465. <https://doi.org/10.1097/00002030-199805000-00006> PMID: 9543443
21. Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, et al. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host & Microbe*. 2021; 29: 44–57.e9. <https://doi.org/10.1016/j.chom.2020.11.007> PMID: 33259788
22. Rouzine IM, Rodrigo A, Coffin JM. Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology. *Microbiol Mol Biol Rev*. 2001; 65: 151–185. <https://doi.org/10.1128/MMBR.65.1.151-185.2001> PMID: 11238990
23. Weissman DB, Desai MM, Fisher DS, Feldman MW. The Rate at Which Asexual Populations Cross Fitness Valleys. *Theor Popul Biol*. 2009; 75: 286–300. <https://doi.org/10.1016/j.tpb.2009.02.006> PMID: 19285994
24. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine*. 2020; 18: 179. <https://doi.org/10.1186/s12967-020-02344-6> PMID: 32321524
25. Amanat F, Krammer F. SARS-CoV-2 Vaccines: Status Report. *Immunity*. 2020; 52: 583–589. <https://doi.org/10.1016/j.immuni.2020.03.007> PMID: 32259480
26. Baum A, Ajithdoss D, Copin R, Zhou A, Lanza K, Negron N, et al. REGN-COV2 antibodies prevent and treat SARS-CoV-2 infection in rhesus macaques and hamsters. *Science*. 2020 [cited 1 Nov 2020]. <https://doi.org/10.1126/science.abe2402> PMID: 33037066
27. Gao Q, Bao L, Mao H, Wang L, Xu K, Yang M, et al. Development of an inactivated vaccine candidate for SARS-CoV-2. *Science*. 2020; 369: 77–81. <https://doi.org/10.1126/science.abc1932> PMID: 32376603
28. Peck KM, Lauring AS. Complexities of Viral Mutation Rates. *Journal of Virology*. 2018; 92. <https://doi.org/10.1128/JVI.01031-17> PMID: 29720522

29. Di Lello FA, Culasso ACA, Campos RH. Inter and inpatient evolution of hepatitis C virus. *Ann Hepatol*. 2015; 14: 442–449. PMID: [26019029](#)
30. Khera T, Todt D, Vercauteren K, McClure CP, Verhoye L, Farhoudi A, et al. Tracking HCV protease population diversity during transmission and susceptibility of founder populations to antiviral therapy. *Antiviral Res*. 2017; 139: 129–137. <https://doi.org/10.1016/j.antiviral.2017.01.001> PMID: [28062191](#)
31. Duffy S. Why are RNA virus mutation rates so damn high? *PLoS Biol*. 2018; 16. <https://doi.org/10.1371/journal.pbio.3000003> PMID: [30102691](#)
32. Thomson EC, Rosen LE, Shepherd JG, Spreafico R, Filipe A da S, Wojcechowskyj JA, et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell*. 2021; 184: 1171–1187.e20. <https://doi.org/10.1016/j.cell.2021.01.037> PMID: [33621484](#)
33. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*. 2020; 182: 812–827.e19. <https://doi.org/10.1016/j.cell.2020.06.043> PMID: [32697968](#)
34. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*. 2020; 1–9. <https://doi.org/10.1038/s41586-020-2895-3> PMID: [33106671](#)
35. Hodcroft EB, Zuber M, Nadeau S, Comas I, Candelas FG, Consortium S-S, et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv*. 2020; 2020.10.25.20219063. <https://doi.org/10.1101/2020.10.25.20219063> PMID: [33269368](#)
36. Kemp SA, Harvey WT, Datir RP, Collier DA, Ferreira I, Carabelli AM, et al. Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion  $\Delta$ H69/V70. *bioRxiv*. 2020; 2020.12.14.422555. <https://doi.org/10.1101/2020.12.14.422555>
37. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. 2021 [cited 6 Apr 2021]. <https://doi.org/10.1126/science.abg3055> PMID: [33658326](#)
38. WHO | SARS-CoV-2 mink-associated variant strain—Denmark. In: WHO [Internet]. World Health Organization; [cited 8 Nov 2020]. Available: <http://www.who.int/csr/don/06-november-2020-mink-associated-sars-cov2-denmark/en/>.
39. Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature*. 2021; 1–10. <https://doi.org/10.1038/s41586-021-03291-y> PMID: [33545711](#)
40. Andreano E, Piccini G, Licastro D, Casalino L, Johnson NV, Paciello I, et al. SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. *bioRxiv*. 2020; 2020.12.28.424451. <https://doi.org/10.1101/2020.12.28.424451> PMID: [33398278](#)
41. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nature Medicine*. 2021; 1–4. <https://doi.org/10.1038/s41591-020-01213-5> PMID: [33442018](#)
42. Regalado A. Moderna believes it could update its coronavirus vaccine without a big new trial. In: MIT Technology Review [Internet]. 13 Jan 2021 [cited 24 Jan 2021]. Available: <https://www.technologyreview.com/2021/01/13/1016098/moderna-variant-coronavirus-vaccine-update/>.
43. Lee WS, Wheatley AK, Kent SJ, DeKosky BJ. Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies. *Nature Microbiology*. 2020; 5: 1185–1191. <https://doi.org/10.1038/s41564-020-00789-5> PMID: [32908214](#)
44. Seow J, Graham C, Merrick B, Acors S, Pickering S, Steel KJA, et al. Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans. *Nature Microbiology*. 2020; 5: 1598–1607. <https://doi.org/10.1038/s41564-020-00813-8> PMID: [33106674](#)
45. Choe PG, Kang CK, Suh HJ, Jung J, Song K-H, Bang JH, et al. Waning Antibody Responses in Asymptomatic and Symptomatic SARS-CoV-2 Infection—Volume 27, Number 1—January 2021—Emerging Infectious Diseases journal—CDC. [cited 27 Jan 2021]. <https://doi.org/10.3201/eid2701.203515> PMID: [33050983](#)
46. Dan JM, Mateus J, Kato Y, Hastie KM, Yu ED, Faliti CE, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science*. 2021; 371. <https://doi.org/10.1126/science.abf4063> PMID: [33408181](#)
47. Stoddard M, Sarkar S, Nolan RP, White DE, White L, Hochberg NS, et al. Beyond the new normal: assessing the feasibility of vaccine-based elimination of SARS-CoV-2. *medRxiv*. 2021; 2021.01.27.20240309. <https://doi.org/10.1101/2021.01.27.20240309>
48. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28: 235–242. <https://doi.org/10.1093/nar/28.1.235> PMID: [10592235](#)

49. Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Portland, OR: AAAI Press; 1996. pp. 226–231.
50. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011; 12: 2825–2830.
51. Hu J, He C-L, Gao Q-Z, Zhang G-J, Cao X-X, Long Q-X, et al. D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity. *bioRxiv*. 2020; 2020.06.20.161323. <https://doi.org/10.1101/2020.06.20.161323>
52. Byrne AW, McEvoy D, Collins AB, Hunt K, Casey M, Barber A, et al. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ Open*. 2020; 10: e039856. <https://doi.org/10.1136/bmjopen-2020-039856> PMID: [32759252](https://pubmed.ncbi.nlm.nih.gov/32759252/)
53. Worldometers.info. Coronavirus Update (Live). [cited 8 Nov 2020]. Available: <https://www.worldometers.info/coronavirus/>.
54. Johnson KE, Stoddard M, Nolan RP, White DE, Hochberg N, Chakravarty A. This time is different: model-based evaluation of the implications of SARS-CoV-2 infection kinetics for disease control. *medRxiv*. 2020; 2020.08.19.20177550. <https://doi.org/10.1101/2020.08.19.20177550>
55. Anand S, Montez-Rath M, Han J, Bozeman J, Kerschmann R, Beyer P, et al. Prevalence of SARS-CoV-2 antibodies in a large nationwide sample of patients on dialysis in the USA: a cross-sectional study. *The Lancet*. 2020; 396: 1335–1344. [https://doi.org/10.1016/S0140-6736\(20\)32009-2](https://doi.org/10.1016/S0140-6736(20)32009-2) PMID: [32987007](https://pubmed.ncbi.nlm.nih.gov/32987007/)