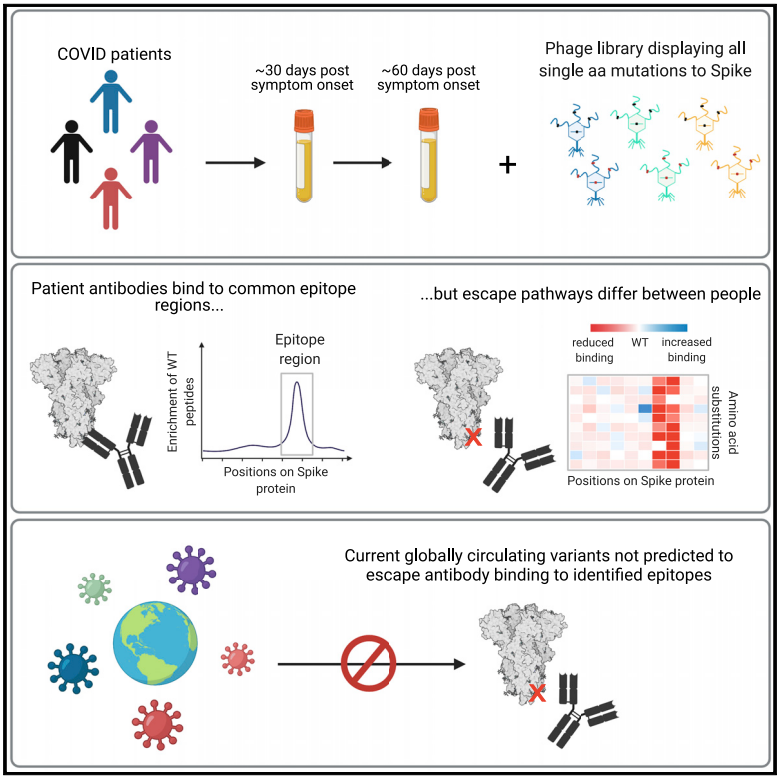
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Cell

High-resolution profiling of pathways of escape for SARS-CoV-2 spike-binding antibodies

Graphical abstract



Highlights

- Profiled mutations that allow antibody escape from COVID-19 plasma using Phage-DMS
- Common epitopes in spike were identified, as were more unique epitopes
- Escape varied among individuals, suggesting different susceptibility to variants
- Little evidence for variation in identified epitopes among circulating viruses

Authors

Meghan E. Garrett, Jared Galloway, Helen Y. Chu, ..., Haidyn Weight, Frederick A. Matsen IV, Julie Overbaugh

Correspondence

joverbau@fredhutch.org

In brief

The humoral response to COVID-19 is mapped to identify antibody escape pathways as well as variation to antibody-mediated virus evolution at the individual level.

Garrett et al., 2021, Cell *184*, 2927–2938
May 27, 2021
https://doi.org/10.1016/j.cell.2021.04.045

CellPress

Cell

CellPress

Article

High-resolution profiling of pathways of escape for SARS-CoV-2 spike-binding antibodies

Meghan E. Garrett,[1,2] Jared Galloway,[3] Helen Y. Chu,[4] Hannah L. Itell,[1,2] Caitlin I. Stoddard,[1] Caitlin R. Wolf,[4] Jennifer K. Logue,[4] Dylan McDonald,[4] Haidyn Weight,[1] Frederick A. Matsen IV,[3,5] and Julie Overbaugh[1,3,6,*]

[1]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98102, USA

[2]Molecular and Cellular Biology Graduate Program, University of Washington and Fred Hutchinson Cancer Research Center, Seattle, WA 98195, USA

[3]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98102, USA

[4]Department of Medicine, University of Washington, Seattle, WA 98195, USA

[5]Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA 98102, USA

[6]Lead contact

*Correspondence: joverbau@fredhutch.org

https://doi.org/10.1016/j.cell.2021.04.045

SUMMARY

Defining long-term protective immunity to SARS-CoV-2 is one of the most pressing questions of our time and will require a detailed understanding of potential ways this virus can evolve to escape immune protection. Immune protection will most likely be mediated by antibodies that bind to the viral entry protein, spike (S). Here, we used Phage-DMS, an approach that comprehensively interrogates the effect of all possible mutations on binding to a protein of interest, to define the profile of antibody escape to the SARS-CoV-2 S protein using coronavirus disease 2019 (COVID-19) convalescent plasma. Antibody binding was common in two regions, the fusion peptide and the linker region upstream of the heptad repeat region 2. However, escape mutations were variable within these immunodominant regions. There was also individual variation in less commonly targeted epitopes. This study provides a granular view of potential antibody escape pathways and suggests there will be individual variation in antibody-mediated virus evolution.

INTRODUCTION

The global outbreak of a novel coronavirus, SARS-CoV-2, has claimed over one million lives within just 1 year after the first detected case (https://coronavirus.jhu.edu/map.html), with countless others experiencing long-term health problems after recovering from infection. Vaccines to prevent SARS-CoV-2 spread and treatments to reduce disease severity are currently under rapid development, with many strategies relying on antibody-mediated immunity. The main viral target of interest for vaccines and antibody therapies against SARS-CoV-2 is the coronavirus spike (S) protein, which decorates the surface of the virion and mediates attachment and entry into host cells (Walls et al., 2020). The S protein is composed of a trimeric assembly of two subunits, S1 and S2, which are proteolytically cleaved at the S1/S2 boundary. The S1 subunit contains an N-terminal domain (NTD) and a receptor binding domain (RBD) within the C-terminal domain (CTD). The S2 subunit contains the fusion peptide (FP) along with two heptad repeat regions (HR1 and HR2), separated by a linker region, responsible for driving viral and host membrane fusion (Walls et al., 2017; Xia et al., 2020). Binding of the S1 protein via the RBD to the human ACE2 receptor is followed by proteolytic cleavage at the S2′ site, which exposes the FP and activates a series of conformational changes resulting in membrane fusion (Belouzard et al., 2009; Fan et al., 2020; Shang et al., 2020).

Neutralizing antibodies targeting the SARS-CoV-2 RBD have been the main focus of vaccine strategies and antibody therapies, as they block virus entry in cell culture (Pinto et al., 2020; Wan et al., 2020; Wang et al., 2020; Wec et al., 2020) and prevent infection or disease in some animal models (Hassan et al., 2020; Rogers et al., 2020; Zost et al., 2020). However, the study of other coronaviruses has illustrated that antibodies elicited by infection can target epitope regions outside of the RBD. For example, a number of neutralizing antibodies directed to regions other than the RBD have been isolated from individuals infected with the closely related viruses SARS-CoV and MERS-CoV (Shanmugaraj et al., 2020). Recent SARS-CoV-2 studies of serum antibodies from coronavirus disease 2019 (COVID-19) patients have led to the identification of neutralization activity directed at linear epitopes found just downstream of the RBD, overlapping the FP, and just upstream of HR2 in the linker region (Li et al., 2020b; Poh et al., 2020). Thus, there may be multiple regions within the SARS-CoV-2 S protein that shape the viral immune response.

While antibodies to the RBD are a logical initial focus for studies of protective antibodies against SARS-CoV-2, it is not yet known whether they are a correlate of protection for

SARS-CoV-2 in humans. A limited understanding of protective immunity is to be expected at the early stage in this new disease, and a broad view is therefore prudent. In this regard, it is important to note that studies from other viruses such as HIV and Ebola have shown that non-neutralizing antibodies are an immune correlate of protection in humans (Gunn et al., 2018; Haynes et al., 2012; Milligan et al., 2015; Saphire et al., 2018). For SARS-CoV-2, antibodies against the S protein likely perform functions other than neutralization given that there is not a direct correlation between the levels of binding to S protein and neutralization titers (Robbiani et al., 2020). Given the incomplete picture we have regarding immunity against SARS-CoV-2, it is important to study the antibody response irrespective of function and/or epitope in order to fill these knowledge gaps.

Given the rapid spread and amplification of SARS-CoV-2 in the population and the high mutation rate of RNA viruses, variants that can evade the immune response are likely to arise. There is strong evidence that immune selection drove the emergence of escape mutants in the S protein of SARS-CoV (Sui et al., 2008) and MERS-CoV (Kleine-Weber et al., 2019), raising the possibility that the same could occur with SARS-CoV-2. Immune selection may be further enhanced by a vaccine if it is not fully protective, making understanding the potential escape pathways of the virus critically important.

There are several recent studies that have examined the effect of select mutations on serum antibody binding to SARS-CoV-2. One study that assessed 82 S protein variants present in circulating SARS-CoV-2 detected a few mutations that resulted in decreased neutralization sensitivity (Li et al., 2020a). However, only a small fraction of the potential mutations that could arise on the S protein were tested. Another study harnessed the power of deep mutational scanning (DMS) to capture a complete picture of the functional consequences of single mutations within the RBD on protein expression, ACE2 binding, and monoclonal antibody binding (Starr et al., 2020). However, no study has yet examined the effect of the polyclonal antibody response on immune escape across the S protein, which is the target in current prophylactic and therapeutic strategies to combat COVID-19.

Previously we developed a comprehensive method of mapping escape mutations for HIV monoclonal antibodies, referred to as Phage-DMS (Garrett et al., 2020). In Phage-DMS, a library of all possible mutations to a protein is generated in peptide fragments, which are expressed by phage. Complexes of antibodies that bind the phage library are immunoprecipitated and sequenced to determine the antibody-binding region(s) and the mutations within that epitope region that disrupt binding. Phage-DMS has several advantages: it is high-throughput, allowing comparison among a large number of samples in parallel; it can determine pathways of escape for both neutralizing and non-neutralizing antibodies that bind to linear epitopes; and results using this method also correlate well with mutational effects measured in other assays (Garrett et al., 2020). Here, we used Phage-DMS to understand the spectrum of single mutations on the S protein that could reduce antibody binding and thus mediate escape from plasma antibodies found in COVID-19 patients. Using a Phage-DMS library displaying both wild-type and mutant peptides tiling across the S protein, we identified a spectrum of single mutants that were capable of reducing antibody binding and found person-to-person variability in the effect of mutations within immunodominant epitopes. In the arms race between the humoral immune response and SARS-CoV-2, these results allow us to predict pathways of escape and forecast the appearance of escape mutants.

RESULTS

Generation of the Spike Phage-DMS library

In order to explore all amino acids (aa) that define the epitope of antibodies directed toward the S protein of SARS-CoV-2, we generated a Phage-DMS library designed to tile across the ectodomain of the S protein from the Wuhan Hu-1 strain (Figure 1) (Wu et al., 2020). We also included DMS peptides generated in the context of the D614G mutation, as clinical and *in vitro* evidence suggests that this variant may have increased infectivity as compared to the original Wuhan Hu-1 strain (Korber et al., 2020). We computationally designed sequences coding for peptides 31 aa long with the variable aa in the central position. To achieve single-aa resolution of epitope boundaries, peptides were designed to overlap by 30 aa. 24,820 unique peptides were designed in total; the peptide library included wild-type peptides that could be used to define the antibody epitope and peptides with all possible mutations to determine those within the defined epitope that disrupt or enhance antibody binding. Two biological replicate libraries of these peptide sequences were cloned as we have done previously (Garrett et al., 2020). Deep sequencing of the final duplicate libraries (library 1 and library 2) indicated that each contained a high percentage of all unique sequences (96.0% and 95.9%, respectively) (Figure S1).

Enrichment of immunodominant linear epitopes by antibodies from COVID patients

We used the Spike Phage-DMS library to determine the pattern of antibody binding in plasma from a cohort of 18 COVID-19 patients from the Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) study in the Seattle area collected between March and May 2020 (Table S1). Most patients had mild symptoms not requiring hospitalization, with the exception of one patient (6) who had moderate symptoms and required supplemental oxygen. Convalescent plasma was collected twice, at approximately day 30 and day 60 post-symptom onset (p.s.o.; Table S1).

To define the epitope region targeted by antibodies, we examined the enrichment of wild-type peptides from the library. We first calculated Pearson's correlation coefficient for the peptide enrichment values from each biological replicate experiment. In general, samples with poor correlation between replicate experiments were those that lacked reproducibly strong antibody binding. The correlation ranged from 0.96 to 0.27, with three patients having no sample from either time point above a correlation of 0.5 (Figure S2A). The latter three cases (7, 16, and 17) were excluded from further analyses. Paired samples taken from the same individual at 30 and 60 days p.s.o. had peptide enrichment values that were well correlated, with a median correlation of 0.87, and were significantly better correlated than randomly paired samples ($p = 1.2e-09$, Wilcoxon rank sum test), which had a median

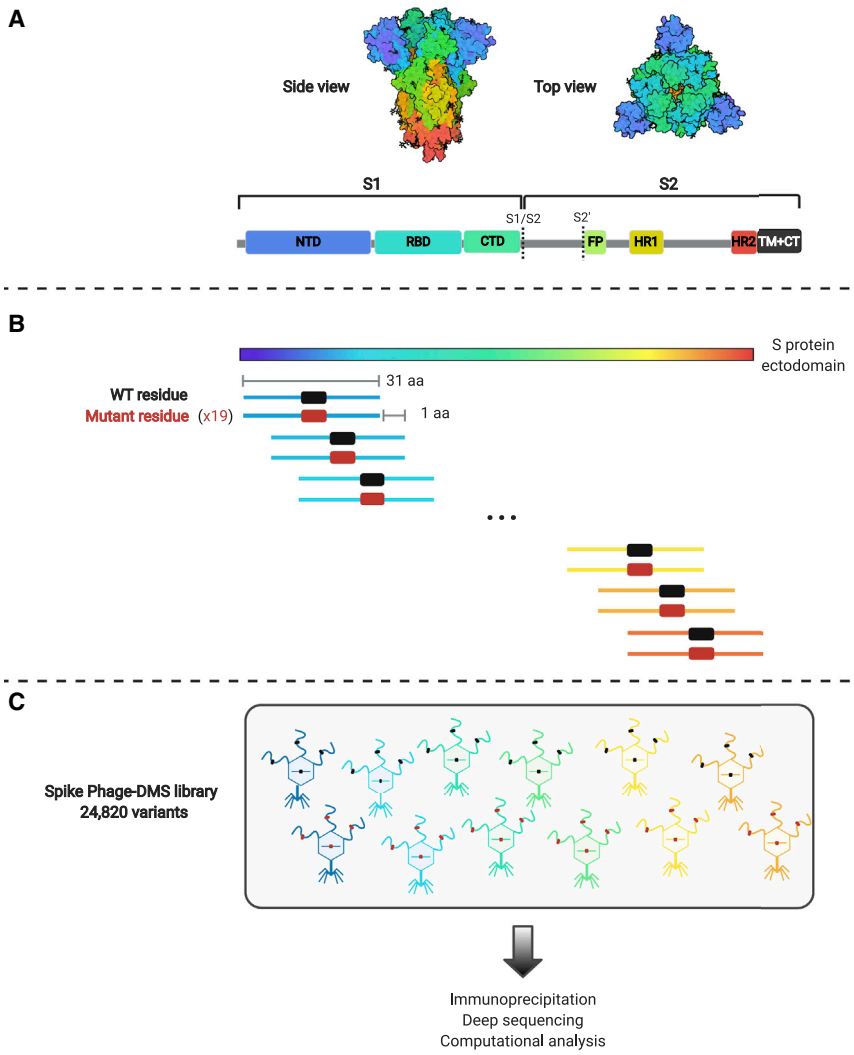


Figure 1. Schematic of the design of the Spike Phage-DMS library

(A) Structure of the S protein and location of important protein domains. Structure was made in BioRender.com (PDB: 6VXX).

(B) Sequences were computationally designed to code for peptides 31aa long and tile stepwise across the Wuhan Hu-1 SARS-CoV-2 S protein ectodomain by 1 aa. There are 20 peptides representing all 20 possible aa at the central position containing either the wild-type residue (shown in black) or a mutant residue (shown in red). Within the 31 aa region surrounding the D614G mutation, peptides were also generated with G614 in addition to the 20 aa variants at the central position.

(C) The designed sequences were cloned into a T7 phage display vector and amplified to create the final S protein Phage-DMS library. This library was then used in downstream immunoprecipitation and deep sequencing experiments with human plasma. See Figure S1 for distribution of sequences in the final library.

correlation of 0.29 (Figure S2B). The correlation between paired samples also tended to be stronger when the correlation between biological replicates, and therefore antibody binding, was stronger (Figure S2C).

Wild-type peptides from two immunodominant regions, which include the FP and upstream sequences spanning aa 809–834 as well as the linker region and N-terminal portion of the adjacent HR2 domain spanning aa 1,140–1,168, were the most enriched peptides (Figure 2). Peptide enrichment data are also available in interactive form using the dms-view online tool at https://github.com/meghangarrett/Spike-Phage-DMS/tree/master/analysis-and-plotting/dms-view (Hilton et al., 2020). Interestingly, while most patients showed enrichment primarily within the linker region upstream of HR2, patient 5 showed strong enrichment of peptides from an epitope within the HR2 domain itself. The enriched peptides included aa 1,167–1,191 and this epitope was unique compared to the other 14 patients. We also observed cases where patient plasma enriched peptides from a less common epitope, albeit weakly, at both 30 and 60 days p.s.o. Plasma from patient 12 enriched peptides within a region of the NTD (aa 255–280), patient 15 enriched peptides within both the RBD (aa 485–500) and the region just downstream of the RBD (aa 540–573), and patient 3 enriched peptides in the region just upstream of the S1/S2 cleavage site (aa 620–644) (Figure 2). Because most patients did not have antibodies that bound linear peptides within the RBD, despite it being a known target of antibodies, we confirmed by ELISA that the patients studied here indeed have plasma antibodies that target RBD (Figure S3A). This indicates that antibodies targeting RBD bind to mainly conformational and/or glycosylated epitopes, which are missed by Phage-DMS.

To assess the potential neutralizing capability of these non-RBD-targeting antibodies, we depleted the 30-day patient plasma samples of RBD-binding antibodies and then performed a neutralization assay with pseudotyped lentivirus expressing the S protein (Crawford et al., 2020). We performed an ELISA on the mock-depleted and RBD-depleted samples and found very low (12.5% of mock) to undetectable (0% of mock) residual RBD-binding activity (Figure S3A). We also performed an ELISA against the S protein and found that RBD antibody depletion did not affect binding to the S protein (Figure S3B), which is consistent with other studies that have found that the majority of S-protein-binding antibodies target non-RBD epitopes (Greaney et al., 2021; Piccoli et al., 2020). We then compared the neutralization capacity of mock-depleted and RBD-depleted patient plasma and found that a wide range (0% to 59%) of the neutralization activity present in patient plasma is directed at non-RBD epitopes (Figure S3D). While we did not find that enrichment of peptides

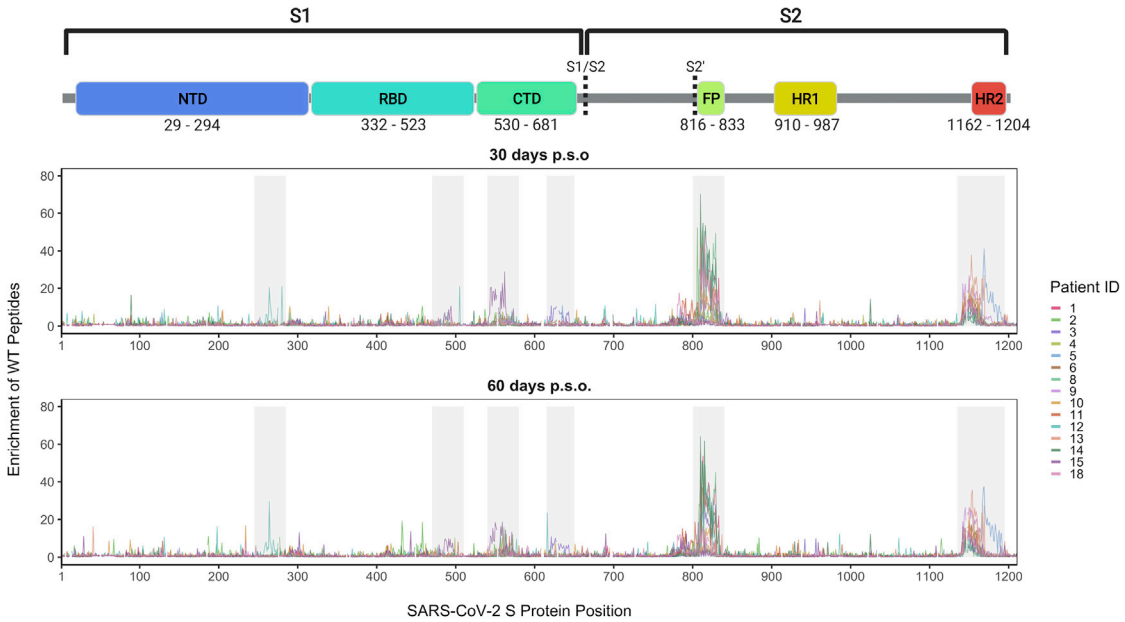


Figure 2. Linear epitopes bound by COVID-19 patient plasma

Lines represent the enrichment of wild-type peptides from the Spike Phage-DMS library from individual plasma samples. Samples from convalescent COVID-19 patient plasma taken at approximately day 30 p.s.o. (top panel) or day 60 p.s.o. (bottom panel) are shown. Lines are colored by patient, with the key to the patient IDs on the right (see Figure S2 for patient inclusion criteria). Grey boxes highlight immunogenic regions where enrichment was detected in at least one individual across time points. Peptides that were included in the design, but absent from the phage library (see Figure S1), are shown as breaks in the line plots. A schematic of S protein domains is shown above, with locations defined based on numbering used in: https://cov.lanl.gov/components/sequence/COV/annt/annt.comp. See also Table S1 and Figures S3 and S4.

within the FP or HR2 region correlated with residual neutralization after depletion (Figure S3E), we did find that binding to the S2 subunit did correlate with NT50 after depletion of RBD antibodies (Figures S3C and S3F). We additionally confirmed that residual antibody binding to RBD after depletion was not associated with the neutralization activity remaining after RBD depletion (Figure S3G). This suggests that antibodies targeting the S2 subunit are at least partially responsible for the residual neutralization activity in these cases.

Patient-to-patient variability in mutations that lead to loss of antibody binding

To determine the effect of mutations on antibody binding to S protein epitopes, we compared the relative enrichment of wild-type peptides and mutant peptides within the epitopes defined above. To quantify the effect of each aa on binding, we calculated the differential selection of mutant peptides versus wild-type peptides and scaled this value by the strength of binding to the wild-type peptide, as we have previously done for Phage-DMS experiments; this measure is highly correlated with the relative binding of individual mutant peptides by ELISA (Garrett et al., 2020). Plotting the scaled differential selection values for all mutants at each site in a heatmap allows for the visualization of sites where mutations led to a detectable loss (in red) or gain (in blue) of binding. The scaled differential selection data for all patients is also available to view in logo plot form at https://github.com/meghangarrett/Spike-Phage-DMS/tree/master/analysis-and-plotting/dms-view.

We generated heatmaps for four representative samples from COVID-19 patients with strong enrichment. We did this for the two immunodominant regions, FP and linker region/HR2, and included paired day 30 and 60 p.s.o. samples (Figure 3 and 4). We also examined the escape profiles within less common epitopes (Figure S4). While the effect of mutations seemed to be consistent between paired patient samples, between individuals, there was more heterogeneity in the specific escape profiles for each region, as follows.

FP region

Within the FP region, sites that led to reduced binding clustered primarily within the 5′ region of the FP itself, which spans from sites S816 to F833 (Figure 3A). Some sites were similarly sensitive between patients. For example, mutations at site 819 and 820 disrupted antibody binding to some extent in all of the patients shown and were the dominant escape positions for patient 8. In contrast, mutations in the adjacent site 818 appeared to contribute to the epitope of some patient antibodies (1, 3, and 12), but not others (8). One interesting site that variably contributed to antibody binding is R815, which is upstream of the FP cleavage site and thus would not be predicted to be present on the post-cleavage form of S2. Antibodies from patient 3 and 12 both showed reduced binding to peptides with mutations at this site; indeed, this was the dominant escape mutation for patient 3. Substitutions at this same aa position had a more modest impact on binding and only a few aa led to disruption for patient 8. In the case of patient 1, several mutations at site 815 had a

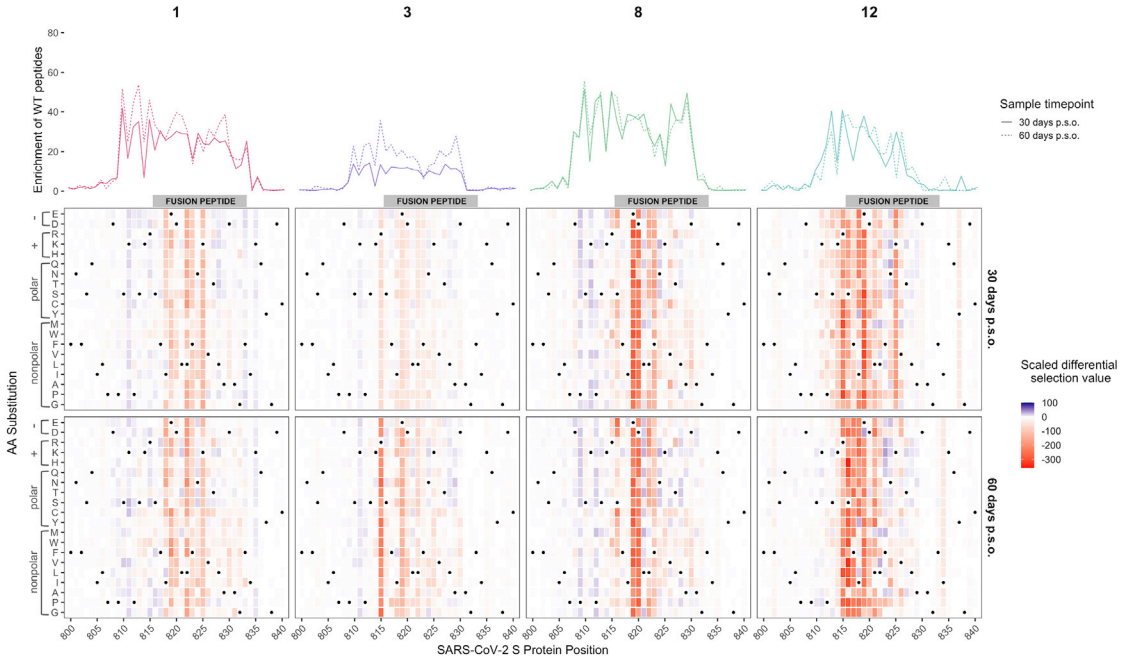


Figure 3. Effect of mutations on binding by COVID-19 patient plasma within the FP region

Heatmaps depicting the effect of all mutations, as measured by scaled differential selection, at each site within the FP epitope for representative COVID-19 patients (numbered at top). Mutations enriched above the wild-type residue are colored blue, and mutations depleted as compared to the wild-type residue are colored red. The intensity of the colors reflects the amount of differential selection as indicated to the right. The wild-type residue is indicated with a black dot. Line plots showing the enrichment of wild-type peptides for each patient are shown above, with a solid line for the day 30 p.s.o. patient samples and a dashed line for the day 60 p.s.o. patient samples. See Figure S5 for alignment of coronavirus FP sequences.

positive differential selection value, suggesting they enhance binding.

Not all mutations at each site were equally disruptive to epitope binding, suggesting specific mutations rather than removal of the wild-type aa per se may be more important for escape. For example, addition of negatively charged aa at site S816 led to loss of binding for patient 8, whereas addition of small aa such as alanine or glycine at the same site had little to no effect. We noted some cases where a mutation had a positive differential selection value and thus presumably bound more strongly than the wild-type SARS-CoV-2 sequence at that position. We aligned the FP sequences for SARS-CoV-2 and human endemic coronaviruses and found that some mutations selected above wild-type were residues present in FP sequences from other coronaviruses (Figure S5). For patient 12, a few mutants at site 818, including I818L, had positive differential selection values, and two coronaviruses (HKU1 and NL63) have a leucine at position 818. In another instance, the N824S mutation had a positive scaled differential selection value for all four samples taken 60 days p.s.o., and we saw that NL63 and 229E both have a serine at position 824.

Linker/HR2 region

Within the region surrounding HR2, we observed that most patients targeted an epitope spanning amino acids 1,123–1,162 within the linker region just upstream of the HR2 domain, with contribution of residues within the N terminus of the HR2 domain in some cases, while one patient targeted an epitope within the HR2 domain itself (Figure 4). In addition to this individual variation in the epitope boundaries, the mutations at different sites exhibited marked variability between patients, as seen for the FP epitope. An example of this variability can be seen among patient antibodies targeting distinct epitopes within the linker region. In this epitope, patient 13 exhibited sensitivity to mutations spanning 1,162 to 1,167, while this was not seen in the other patients. There were also some cases where there were similarities in escape profiles for patient antibodies targeting the same region. For example, there was evidence that aa changes at sites 1,151 and 1,152 disrupted antibody binding in several patients. However, while mutations at site 1,151 disrupted antibody binding for patients 9 and 18, they did not for the antibodies in patient 13.

In patient 5, where there was a distinct epitope within the HR2 domain, we observed unique sensitivity to mutations in sites 1,176 through 1,182 not found in any other patient. Like many other plasma epitope profiles, mutations to the same site had a variable effect. For example, at site 1,180, some mutations improved binding (e.g., Q1180G), some reduced binding (e.g., Q1180L), and some had little to no effect (e.g., Q1180A). There was no overlap in the sites that showed reduced binding in patient 5 and those that were negatively selected in patients 9, 13, or 18, again suggesting that the linker region and HR2 domain epitopes are distinct epitopes that have distinct pathways of escape.

Other regions

There was relatively less differential selection at regions outside of the immunodominant epitopes, but the profiles nonetheless

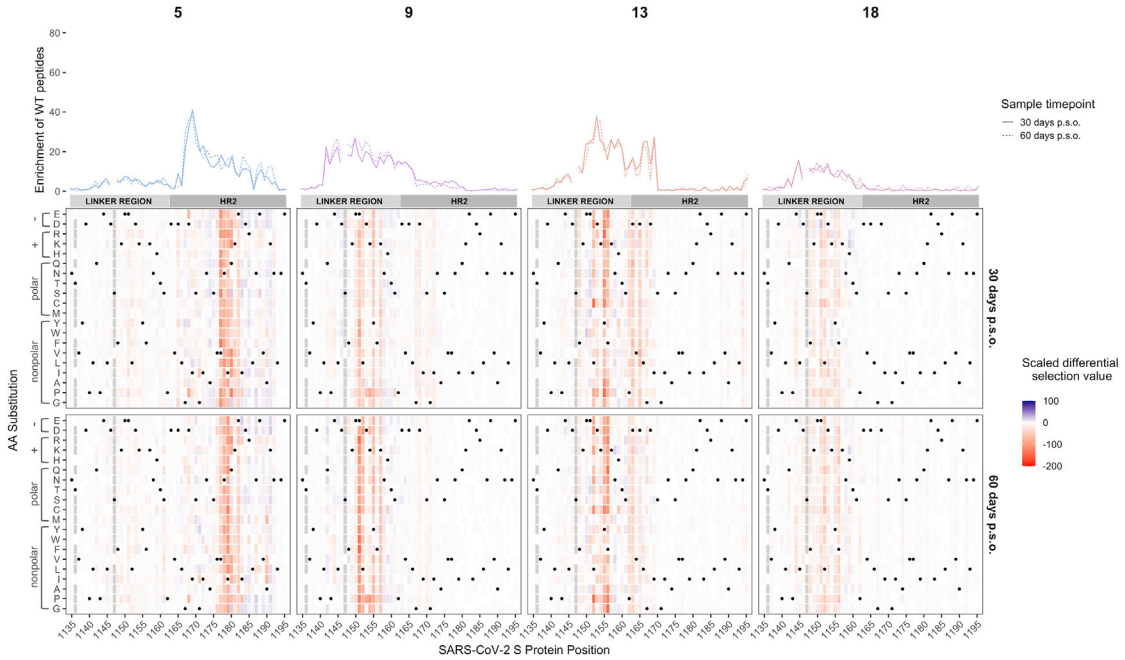


Figure 4. Effect of mutations on binding by COVID-19 patient plasma within the linker/HR2 region

Heatmaps depicting the effect of all mutations, as measured by scaled differential selection, at each site within the linker region/HR2 epitope for representative COVID-19 patients (numbered at top). Mutations enriched above the wild-type residue are colored blue and mutations depleted as compared to the wild-type residue are colored red. The intensity of the colors reflects the amount of differential selection as indicated to the right. The wild-type residue is indicated with a black dot. Line plots showing the enrichment of wild-type peptides for each patient are shown above, with a solid line for the day 30 p.s.o. patient samples and a dashed line for the day 60 p.s.o. patient samples. Peptides missing from the library are shown as gray boxes in the heatmaps and as breaks in the line plots.

provided evidence for some specific mutations that disrupt binding in epitopes in the NTD, RBD, and CTD just downstream of the RBD (Figure S5). For example, within the NTD, positions spanning 264–269 in patient 12 disrupted binding, most notably mutations from A to select aa (W, V, and I) at site 264. The escape profiles for patient 15 suggests that changes at positions 491–495 disrupt binding in the RBD epitope, while mutations at sites 550–553 and 558 disrupted binding within the CTD epitope for this patient. For patient 3, the major effect on binding within the CTD was at positions 628–634, most notably at aa 628 and 633. Although negative scaled differential selection values can be seen at various other aa positions, in many of these cases, the effects were weak and/or inconsistent across time points. Identifying cases of stronger enrichment in these regions will be necessary in order to better define the full spectrum of escape mutations at these regions.

Most common circulating SARS-CoV-2 variants, including D614G, are not predicted to escape antibody binding

One question that arises from the escape profiles described above is whether the mutations that disrupt antibody binding identified here correlate with the global evolution of the SARS-CoV-2 virus to-date. Using variant frequencies present in sequences from the global initiative on sharing avian influenza data (GISAID) and reported at the http://cov.lanl.govcontent/index website, we compared the naturally occurring diversity of every site to the effect on antibody binding as found by Phage-DMS. For each patient and at each site, we averaged the scaled differential selection value for all mutants and then plotted this against the mutational entropy, which is a measure of aa diversity (Figure 5A). Larger mutational entropy values indicate more global diversity at that site, and sites with a mutational entropy value of above 0.02 are flagged as sites of interest by the Los Alamos National Laboratory (LANL) database. We found that the majority of sites that led to loss of antibody binding when mutated were not present at a high frequency in nature. Sites within the immunodominant epitopes did not generally have high mutational entropy, although for patients 13 and 15, there were a few sites within the HR2 region that were both present at high frequency in nature (mutational entropy > 0.02) and led to loss of antibody binding when mutated. Averaging the scaled differential selection values across all patients did not reveal any sites that commonly lead to loss of antibody binding and are present at high frequency (Figure S6).

Focusing on circulating variants that are rising in abundance in the S protein ectodomain, we next examined the effect of mutations at these sites as determined by Phage-DMS. We selected the 20 most common mutations found in nature and saw that these were located across the S1 and S2 proteins and mostly were not found within the major immunodominant regions identified here, with the exception of two mutations that appear within the LR/HR2 region (Figure 5B). We examined the scaled differential selection value of each natural variant for all 60-day patient samples. In general, the variants found in the global population did not show reduced binding to patient

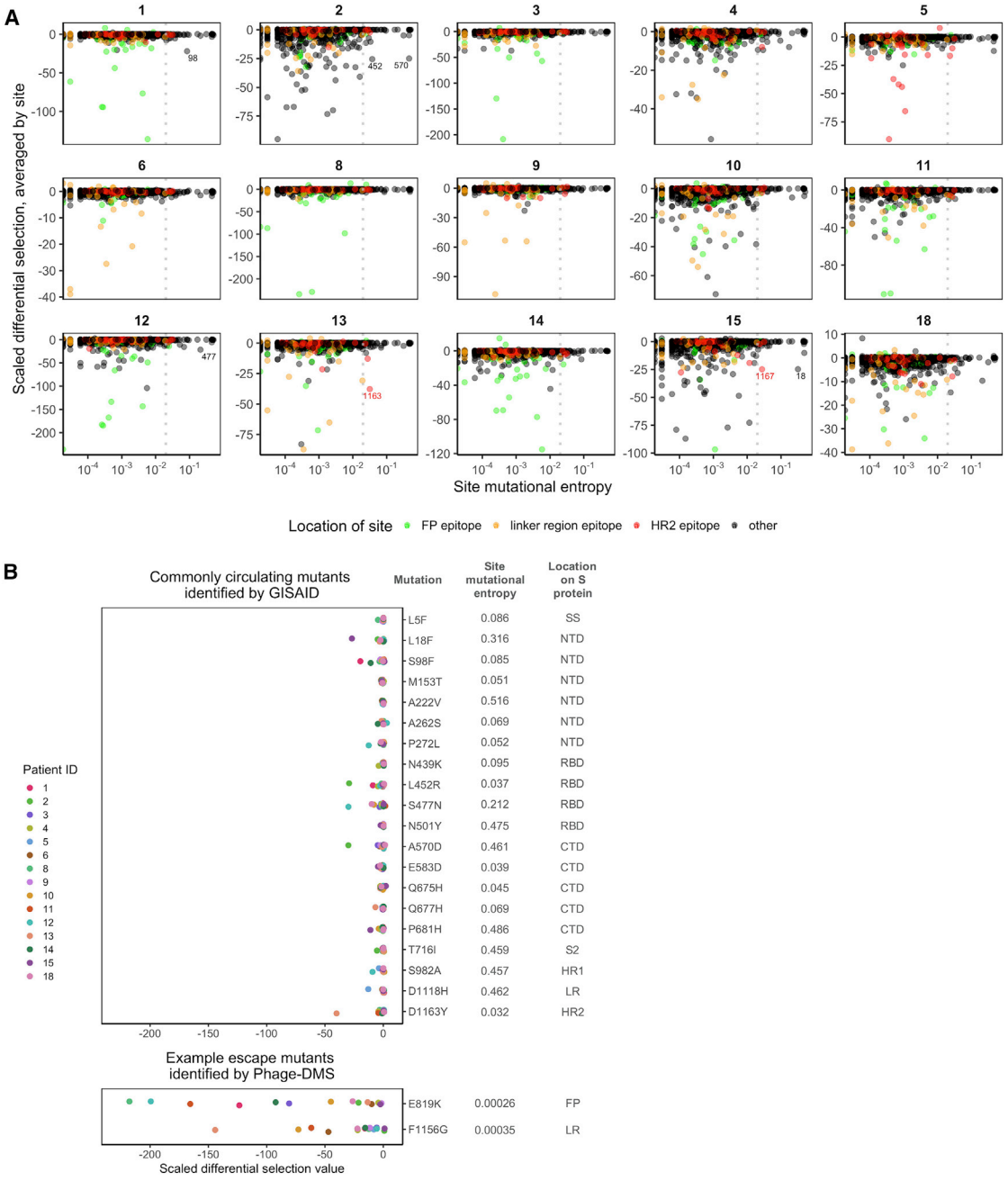


Figure 5. Predicted effects of commonly circulating S protein variants on antibody escape

(A) Scatterplot comparing the effect of mutations on patient plasma antibody binding and the frequency of all circulating S protein variants. The mutational entropy of every circulating protein variant, as reported at the http://cov.lanl.govcontent/index website and based on GISAID global sequencing, is plotted on the x axis. The average of the scaled differential selection values for all mutants at each site is plotted on the y axis. Patient IDs are indicated on the top. Each site is colored by its location, as indicated on the bottom. The dotted line denotes the cutoff (0.02) of mutational entropy that GISAID uses to determine variants of interest. Sites with relatively high mutational entropy and scaled differential selection values are labeled. See Figure S6 for the average effect across all patients.

(B) Effect of mutant peptides representing commonly circulating S protein variants on binding to COVID-19 patient plasma. We selected sites with a mutational entropy of greater than 0.02, as this is the cutoff used by LANL to determine sites of interest. On top are the 20 sites with the highest mutational entropy values, and on the bottom are two selected sites that were noted as sites of antibody escape within immunodominant epitopes by Phage-DMS. On the right are the mutations examined, named according to the wild-type amino acids, followed by the site number, followed by the mutant amino acids of interest. Mutations chosen at sites of high mutational entropy represent the most common variant found in nature. The scaled differential values found by Phage-DMS for each mutant peptide are shown as dots and are colored by patient as indicated to the left. Data are from samples taken day 60 p.s.o. SS, signal sequence; S2, N-terminal region of S2; LR, linker region.

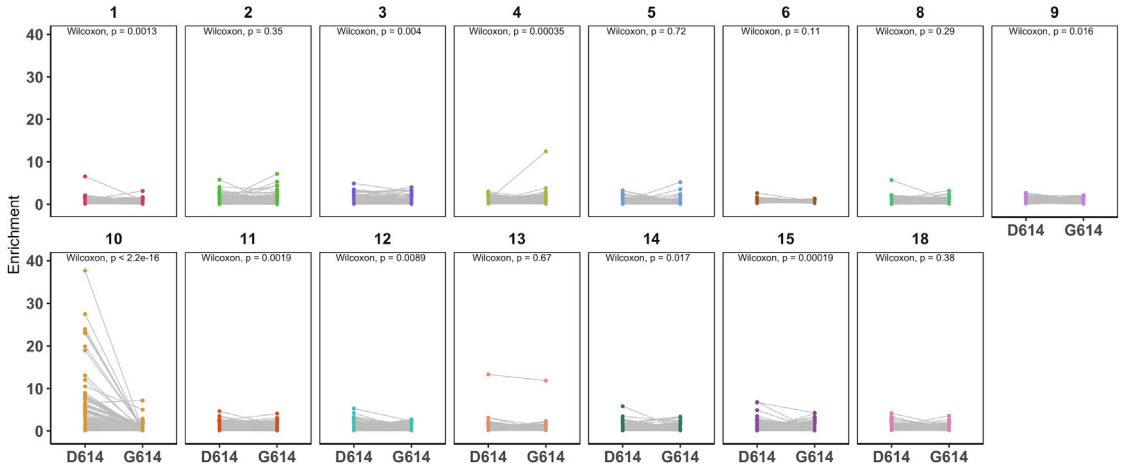


Figure 6. Epistatic effects of D614G mutation on antibody binding

Enrichment values for paired mutant peptides made in either the wild-type Wuhan Hu-1 strain (on the left, D614) or D614G background (on the right, G614) for each patient (numbered at top). All mutant peptides that contained site 614 were included in this analysis (spanning amino acids 599–629). Data are from samples taken day 60 p.s.o. Wilcoxon paired signed-rank test was performed (n = 380 paired mutant peptides). The effect size for all patient samples was small (Wilcoxon $r < 0.3$), except for patient 10, whose antibodies exhibited a moderate effect (Wilcoxon $r = 0.46$).

plasma antibodies, with perhaps a modest effect for a few mutations with individual patient samples. This contrasted with the strong negative selection observed in the Phage-DMS screen for mutations within the two immunodominant epitopes at position 819 and 1,156, two positions where there are not variants rising in abundance in nature. However, the possibility remains that sites of high mutational entropy could exist within conformational epitopes, which are not generally displayed in phage libraries.

Evidence for patient-specific epistatic effects of D614G

We were particularly interested in the S protein mutant D614G, a variant that has emerged as the dominant circulating strain of SARS-CoV-2. Because this library also contains peptides made in the background of the D614G strain, we were able to test whether this mutation could, through epistasis, interact with mutations introduced at other sites on the same peptide and together modify the ability of antibodies to bind. For all mutant peptides that tiled across site 614, the enrichment of each peptide made in the original Wuhan Hu-1 strain background (D614) was compared against the enrichment of the corresponding peptide made in the background of the G614 mutation. We found that the presence of the D614G mutation significantly reduced the ability of antibodies from patient 10 to bind to mutant peptides in this region, with a moderate effect size (Figure 6; Wilcoxon paired signed-rank test). In 8 out of the other 14 patients, there was a statistically significant difference in the binding between mutant peptides with and without the D614G mutation, but the effect sizes were all small. Occasionally, for example in the case of patient 4, a mutation at one position (A609W) showed stronger enrichment in the context of G614 as compared to D614. Conversely, in the case of patient 1, one mutation (A609D) was more enriched when accompanied by D614 as compared to G614. In other cases, such as with patient 2, there was some selective enrichment in both contexts.

DISCUSSION

In this study, we tested all possible mutations on the S protein to provide a map of escape pathways within immunodominant linear epitopes targeted by the plasma of convalescent COVID-19 patients. The responses within an individual were consistent over time, but there were many unique pathways of escape that differed between individuals, even within the same epitope region. These findings suggest that the pattern of virus evolution within the growing pandemic is not likely to be driven by a single antibody escape mutation, which may explain the lack of emergence of circulating strains with mutations that disrupt antibody binding identified here. Thus, the responses to a SARS-CoV-2 S protein vaccine immunogen are not likely to be uniform, nor will the pathways of escape.

While others have defined the linear epitopes bound by antibodies within COVID-19 patient serum (Jiang et al., 2020; Li et al., 2020b; Poh et al., 2020; Shrock et al., 2020; Yi et al., 2020; Zamecnik et al., 2020), there have been limited studies to determine what mutations within these regions could abrogate antibody binding. One study that tested 82 circulating SARS-CoV-2 variants showed that the few single mutations that did result in reduced neutralization sensitivity were unique to each serum sample (Li et al., 2020a), and our data support and extend the idea that antibody-binding sites and escape mutants vary greatly from person to person. The lack of a singular escape signature within each immunodominant epitope also implies that a diverse repertoire of antibodies targets these regions. This provides a functional context for results of studies of SARS-CoV-2 S protein antibodies sequences from COVID-19 patients, which showed that no single clone dominates the antibody response; rather, a diverse collection of variable heavy- and light-chain genes are used in different individuals (Brouwer et al., 2020; Seydoux et al., 2020).

Our results also demonstrate the power of interrogating the role of every possible aa at every site on the S protein.

Mapping COVID-19 patient serum epitopes by alanine scanning has helped identify sites of antibody binding by removing important side-chain interactions (Shrock et al., 2020), but studies with other viruses have shown that escape can be mediated by mutations at sites not directly in contact with the antibody via introduction of nearby charged or bulky aa (Dingens et al., 2019; Doud et al., 2017; Patel et al., 2019). This concept was evident in this study, for example within the FP epitope for patient 8, where addition of negatively charged aa led to escape at site S816, but addition of an alanine did not. We were also able to examine the effects of mutations in the context of the common circulating 614 variant, which we found did not itself drive escape from antibody binding. However, our data suggest that mutation D614G does potentiate escape mutations at other positions in at least one patient, highlighting the power of examining combinations of mutations to better understand the increasing global dominance of the D614G variant.

The FP and linker region/HR2 immunodominant epitopes profiled in this study offer intriguing alternative targets for vaccine design or immunotherapy that could complement efforts focused on the RBD. S2 in general and the FP in particular are highly conserved among coronaviruses, indicating the strong purifying evolution acting in this region due to functional constraints. As with other RNA viruses, this restricted mutational space likely limits the virus' ability to escape antibody binding (Domingo and Holland, 1997; Keck et al., 2009; Kosik et al., 2018; Louie et al., 2018). One study has found that the FP is a target of neutralizing antibodies in SARS-CoV-2 patients (Poh et al., 2020), and these antibodies could act by preventing protease-mediated cleavage at the S2′ site. Interestingly, we saw evidence of antibody binding to sites upstream of the S2′ site in some patients (3 and 12), but not others (1 and 8), potentially indicating that antibodies to the FP could bind during different stages of cell entry before or after S2′ cleavage. HR2 and the upstream linker region are both targets of neutralizing antibodies in SARS-CoV-2 patients (Li et al., 2020b), and a neutralizing antibody raised against the linker region from murine SARS-CoV infection has been shown to cross-react with SARS-CoV-2 (Lip et al., 2006; Zheng et al., 2020). Antibodies targeting the heptad repeat regions, which undergo large conformational transformations in order to facilitate membrane fusion, could neutralize by binding to the fusion intermediate state. Interestingly, our data suggest that non-RBD-directed neutralization correlates with S2 protein binding, although it did not correlate specifically with either the FP or linker region/HR2 enrichment measured by Phage-DMS. These more conserved targets in S2 may be important for designing optimal and durable vaccines, given that RBD has higher mutational entropy, increasing the potential for immune escape from vaccine-induced antibodies. The mutational flexibility of the RBD is supported by *in vitro* studies with RBD-targeting neutralizing antibodies, which found that escape mutants in the RBD of SARS-CoV-2 were rapidly selected (Baum et al., 2020).

There are several important caveats to the results we obtained with Phage-DMS libraries. In general, phage libraries display linear peptides and therefore miss antibodies that bind to complex conformational epitopes or at best provide a partial view of a linear portion of such epitopes. The strong plasma binding response to RBD by ELISA but lack of enriched RBD peptides demonstrates this limitation and indicates that we are unable to make predictions about mutants in conformational epitopes on the RBD. Additionally, the peptide libraries are amplified in bacteria and therefore lack glycans or other post-translational modifications. While these Phage-DMS-derived escape maps define mutations that could lead to loss of antibody binding, it is unknown whether the virus would tolerate mutations at those sites. In fact, a study examining the effect of mutations within the SARS-CoV-2 RBD demonstrated that many led to poor protein expression or loss of function (Starr et al., 2020). Thus, for the antibody-targeting regions described here, which are in more conserved functional domains, the escape mutations should be further examined in the context of their mutational tolerance.

These studies have defined common and variable escape mutations across 18 COVID-19 patients that will be useful for viral surveillance, particularly as SARS-CoV-2 S-protein-based vaccines are introduced into the population. In addition, the Spike Phage-DMS library developed here could be useful for examining larger cohorts, potentially including those with variable clinical outcomes and individuals of variable ages, to define whether mutations that disrupt antibody binding vary in a systematic way across populations and whether this is correlated with clinical outcome or risk of reinfection.

Limitations of the study

This study focuses on defining the epitopes, including linear sequences, targeted by convalescent plasma in mild and moderate cases of COVID-19. Specifically, because the Spike Phage-DMS library displays only short linear peptides, this method is generally not optimal for identifying the full epitope of antibodies that rely on conformational presentation of the antigen. The peptides displayed in a Phage-DMS library do not undergo post-translational modification, and therefore, antibodies that rely on binding to glycans and other similar moieties cannot be defined using this method. The plasma samples used in this study were taken relatively soon after infection and were mostly taken from patients with mild infection not requiring hospitalization. Thus, we may not have detected the full range of possible responses to COVID-19.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects
- METHOD DETAILS
  - Design and generation of the Spike Phage-DMS library
  - Immunoprecipitation of human plasma with Phage-DMS library

- Illumina library preparation and deep sequencing
- Depletion of RBD-binding antibodies from patient plasma
- ELISA with RBD, Spike, and S2 proteins
- Generation and titration of Spike pseudotyped lentivirus
- Neutralization assays
- Multiple sequence alignment
- Graphical illustrations

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Demultiplexing and alignment of Illumina reads
  - Calculating enrichment and scaled differential selection of peptides
  - Data curation

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.04.045.

ACKNOWLEDGMENTS

We gratefully acknowledge Kevin Sung for helpful discussion and Katherine H.D. Crawford for the Python code used to generate the oligonucleotide library sequences. We thank Sarah K. Hilton, John Huddleston, and Jesse D. Bloom for developing and demonstrating the dms-view online tool. We thank the Fred Hutch Genomics core facility, in particular Cassie Sather, for assistance with sequencing. This work was funded by NIH grants AI138709 (principal investigator [PI] J.O.), R01 AI146028, and U19 AI128914 (PI F.A.M.). J.O. received support as the Endowed Chair for Graduate Education (FHCRC). The research of F.A.M. was supported in part by a Faculty Scholar grant from the Howard Hughes Medical Institute and the Simons Foundation.

AUTHOR CONTRIBUTIONS

J.O. conceived the project; M.E.G. and J.O. led the design of the study; H.Y.C. led the HAARVI study, with C.R.W., J.K.L., and D.M. involved in sample collection. M.E.G., H.L.I., and H.W. performed experiments, with C.I.S. performing sample processing and organization; M.E.G. and J.G. performed computational and data analyses, with F.A.M. advising. M.E.G. and J.O. wrote the paper, with input from all authors.

DECLARATION OF INTERESTS

M.E.G. and J.O. are inventors on a patent application on Phage-DMS. H.Y.C. is a consultant for Merck, Pfizer, Ellume, and the Bill and Melinda Gates Foundation and has received support from Cepheid and Sanofi-Pasteur.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. We worked to ensure that the study questionnaires were prepared in an inclusive way. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. One or more of the authors of this paper self-identifies as living with a disability. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: November 13, 2020
Revised: February 26, 2021
Accepted: April 23, 2021
Published: May 4, 2021

REFERENCES

Baum, A., Fulton, B.O., Wloga, E., Copin, R., Pascal, K.E., Russo, V., Giordano, S., Lanza, K., Negron, N., Ni, M., et al. (2020). Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science *369*, 1014–1018.

Belouzard, S., Chu, V.C., and Whittaker, G.R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc. Natl. Acad. Sci. USA *106*, 5871–5876.

Brouwer, P.J.M., Caniels, T.G., van der Straten, K., Snitselaar, J.L., Aldon, Y., Bangaru, S., Torres, J.L., Okba, N.M.A., Claireaux, M., Kerster, G., et al. (2020). Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. Science *369*, 643–650.

Crawford, K.H.D., Eguia, R., Dingens, A.S., Loes, A.N., Malone, K.D., Wolf, C.R., Chu, H.Y., Tortorici, M.A., Veesler, D., Murphy, M., et al. (2020). Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. Viruses *12*, 12.

Dingens, A.S., Arenz, D., Weight, H., Overbaugh, J., and Bloom, J.D. (2019). An Antigenic Atlas of HIV-1 Escape from Broadly Neutralizing Antibodies Distinguishes Functional and Structural Epitopes. Immunity *50*, 520–532.e3.

Dingens, A.S., Crawford, K.H.D., Adler, A., Steele, S.L., Lacombe, K., Eguia, R., Amanat, F., Walls, A.C., Wolf, C.R., Murphy, M., et al. (2020). Serological identification of SARS-CoV-2 infections among children visiting a hospital during the initial Seattle outbreak. Nat. Commun. *11*, 4378.

Domingo, E., and Holland, J.J. (1997). RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. *51*, 151–178.

Doud, M.B., Hensley, S.E., and Bloom, J.D. (2017). Complete mapping of viral escape from neutralizing antibodies. PLoS Pathog. *13*, e1006271.

Fan, X., Cao, D., Kong, L., and Zhang, X. (2020). Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. Nat. Commun. *11*, 3618.

Garrett, M.E., Itell, H.L., Crawford, K.H.D., Basom, R., Bloom, J.D., and Overbaugh, J. (2020). Phage-DMS: A Comprehensive Method for Fine Mapping of Antibody Epitopes. iScience *23*, 101622.

Greaney, A.J., Loes, A.N., Crawford, K.H.D., Starr, T.N., Malone, K.D., Chu, H.Y., and Bloom, J.D. (2021). Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe *29*, 463–476.e6.

Gunn, B.M., Yu, W.H., Karim, M.M., Brannan, J.M., Herbert, A.S., Wec, A.Z., Halfmann, P.J., Fusco, M.L., Schendel, S.L., Gangavarapu, K., et al. (2018). A Role for Fc Function in Therapeutic Monoclonal Antibody-Mediated Protection against Ebola Virus. Cell Host Microbe *24*, 221–233.e5.

Hassan, A.O., Case, J.B., Winkler, E.S., Thackray, L.B., Kafai, N.M., Bailey, A.L., McCune, B.T., Fox, J.M., Chen, R.E., Alsoussi, W.B., et al. (2020). A SARS-CoV-2 Infection Model in Mice Demonstrates Protection by Neutralizing Antibodies. Cell *182*, 744–753.e4.

Haynes, B.F., Gilbert, P.B., McElrath, M.J., Zolla-Pazner, S., Tomaras, G.D., Alam, S.M., Evans, D.T., Montefiori, D.C., Karnasuta, C., Sutthent, R., et al. (2012). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. N. Engl. J. Med. *366*, 1275–1286.

Hilton, S.K., Huddleston, J., Black, A., North, K., Dingens, A.S., Bedford, T., and Bloom, J.D. (2020). dms-view: Interactive visualization tool for deep mutational scanning data. J. Open Source Softw. *5*, 2353.

Jiang, H.W., Li, Y., Zhang, H.N., Wang, W., Yang, X., Qi, H., Li, H., Men, D., Zhou, J., and Tao, S.C. (2020). SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses. Nat. Commun. *11*, 3581.

Keck, Z.Y., Li, S.H., Xia, J., von Hahn, T., Balfe, P., McKeating, J.A., Witteveldt, J., Patel, A.H., Alter, H., Rice, C.M., and Foung, S.K. (2009). Mutations in hepatitis C virus E2 located outside the CD81 binding sites lead to escape from broadly neutralizing antibodies but compromise virus infectivity. J. Virol. *83*, 6149–6160.

Kleine-Weber, H., Elzayat, M.T., Wang, L., Graham, B.S., Müller, M.A., Drosten, C., Pöhlmann, S., and Hoffmann, M. (2019). Mutations in the Spike Protein of Middle East Respiratory Syndrome Coronavirus Transmitted in Korea Increase Resistance to Antibody-Mediated Neutralization. J. Virol. *93*, 93.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., et al.; Sheffield COVID-19 Genomics Group (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell *182*, 812–827.e19.

Kosik, I., Ince, W.L., Gentles, L.E., Oler, A.J., Kosikova, M., Angel, M., Magadán, J.G., Xie, H., Brooke, C.B., and Yewdell, J.W. (2018). Influenza A virus hemagglutinin glycosylation compensates for antibody escape fitness costs. PLoS Pathog. *14*, e1006796.

Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., et al. (2020a). The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. Cell *182*, 1284–1294.e9.

Li, Y., Lai, D.Y., Zhang, H.N., Jiang, H.W., Tian, X., Ma, M.L., Qi, H., Meng, Q.F., Guo, S.J., Wu, Y., et al. (2020b). Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. Cell. Mol. Immunol. *17*, 1095–1097.

Lip, K.M., Shen, S., Yang, X., Keng, C.T., Zhang, A., Oh, H.L., Li, Z.H., Hwang, L.A., Chou, C.F., Fielding, B.C., et al. (2006). Monoclonal antibodies targeting the HR2 domain and the region immediately upstream of the HR2 of the S protein neutralize in vitro infection of severe acute respiratory syndrome coronavirus. J. Virol. *80*, 941–950.

Louie, R.H.Y., Kaczorowski, K.J., Barton, J.P., Chakraborty, A.K., and McKay, M.R. (2018). Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies. Proc. Natl. Acad. Sci. USA *115*, E564–E573.

Mabuka, J., Nduati, R., Odem-Davis, K., Peterson, D., and Overbaugh, J. (2012). HIV-specific antibodies capable of ADCC are common in breastmilk and are associated with reduced risk of transmission in women with high viral loads. PLoS Pathog. *8*, e1002739.

Milligan, C., Richardson, B.A., John-Stewart, G., Nduati, R., and Overbaugh, J. (2015). Passively acquired antibody-dependent cellular cytotoxicity (ADCC) activity in HIV-infected infants is associated with reduced mortality. Cell Host Microbe *17*, 500–506.

Mohan, D., Wansley, D.L., Sie, B.M., Noon, M.S., Baer, A.N., Laserson, U., and Larman, H.B. (2018). PhIP-Seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. Nat. Protoc. *13*, 1958–1978.

Patel, J.S., Quates, C.J., Johnson, E.L., and Ytreberg, F.M. (2019). Expanding the watch list for potential Ebola virus antibody escape mutations. PLoS ONE *14*, e0211093.

Piccoli, L., Park, Y.J., Tortorici, M.A., Czudnochowski, N., Walls, A.C., Beltramello, M., Silacci-Fregni, C., Pinto, D., Rosen, L.E., Bowen, J.E., et al. (2020). Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. Cell *183*, 1024–1042.e21.

Pinto, D., Park, Y.J., Beltramello, M., Walls, A.C., Tortorici, M.A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., De Marco, A., et al. (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. Nature *583*, 290–295.

Poh, C.M., Carissimo, G., Wang, B., Amrun, S.N., Lee, C.Y., Chee, R.S., Fong, S.W., Yeo, N.K., Lee, W.H., Torres-Ruesta, A., et al. (2020). Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. Nat. Commun. *11*, 2806.

Robbiani, D.F., Gaebler, C., Muecksch, F., Lorenzi, J.C.C., Wang, Z., Cho, A., Agudelo, M., Barnes, C.O., Gazumyan, A., Finkin, S., et al. (2020). Convergent antibody responses to SARS-CoV-2 in convalescent individuals. Nature *584*, 437–442.

Rogers, T.F., Zhao, F., Huang, D., Beutler, N., Burns, A., He, W.T., Limbo, O., Smith, C., Song, G., Woehl, J., et al. (2020). Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. Science *369*, 956–963.

Saphire, E.O., Schendel, S.L., Gunn, B.M., Milligan, J.C., and Alter, G. (2018). Antibody-mediated protection against Ebola virus. Nat. Immunol. *19*, 1169–1178.

Seydoux, E., Homad, L.J., MacCamy, A.J., Parks, K.R., Hurlburt, N.K., Jennewein, M.F., Akins, N.R., Stuart, A.B., Wan, Y.H., Feng, J., et al. (2020). Analysis of a SARS-CoV-2-Infected Individual Reveals Development of Potent Neutralizing Antibodies with Limited Somatic Mutation. Immunity *53*, 98–105.e5.

Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., and Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. Proc. Natl. Acad. Sci. USA *117*, 11727–11734.

Shanmugaraj, B., Siriwattananon, K., Wangkanont, K., and Phoolcharoen, W. (2020). Perspectives on monoclonal antibody therapy as potential therapeutic intervention for Coronavirus disease-19 (COVID-19). Asian Pac. J. Allergy Immunol. *38*, 10–18.

Shrock, E., Fujimura, E., Kula, T., Timms, R.T., Lee, I.H., Leng, Y., Robinson, M.L., Sie, B.M., Li, M.Z., Chen, Y., et al.; MGH COVID-19 Collection & Processing Team (2020). Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. Science *370*, eabd4250.

Stadlbauer, D., Amanat, F., Chromikova, V., Jiang, K., Strohmeier, S., Arunkumar, G.A., Tan, J., Bhavsar, D., Capuano, C., Kirkpatrick, E., et al. (2020). SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup. Curr. Protoc. Microbiol. *57*, e100.

Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K.H.D., Dingens, A.S., Navarro, M.J., Bowen, J.E., Tortorici, M.A., Walls, A.C., et al. (2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell *182*, 1295–1310.e20.

Sui, J., Aird, D.R., Tamin, A., Murakami, A., Yan, M., Yammanuru, A., Jing, H., Kan, B., Liu, X., Zhu, Q., et al. (2008). Broadening of neutralization activity to directly block a dominant antibody-driven SARS-coronavirus evolution pathway. PLoS Pathog. *4*, e1000197.

Walls, A.C., Tortorici, M.A., Snijder, J., Xiong, X., Bosch, B.J., Rey, F.A., and Veesler, D. (2017). Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. Proc. Natl. Acad. Sci. USA *114*, 11157–11162.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell *181*, 281–292.e6.

Wan, J., Xing, S., Ding, L., Wang, Y., Gu, C., Wu, Y., Rong, B., Li, C., Wang, S., Chen, K., et al. (2020). Human-IgG-Neutralizing Monoclonal Antibodies Block the SARS-CoV-2 Infection. Cell Rep. *32*, 107918.

Wang, C., Li, W., Drabek, D., Okba, N.M.A., van Haperen, R., Osterhaus, A.D.M.E., van Kuppeveld, F.J.M., Haagmans, B.L., Grosveld, F., and Bosch, B.J. (2020). A human monoclonal antibody blocking SARS-CoV-2 infection. Nat. Commun. *11*, 2251.

Wec, A.Z., Wrapp, D., Herbert, A.S., Maurer, D.P., Haslwanter, D., Sakharkar, M., Jangra, R.K., Dieterle, M.E., Lilov, A., Huang, D., et al. (2020). Broad neutralization of SARS-related viruses by human monoclonal antibodies. Science *369*, 731–736.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., et al. (2020). A new coronavirus associated with human respiratory disease in China. Nature *579*, 265–269. https://doi.org/10.1038/s41586-020-2008-3.

Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., Qi, F., Bao, L., Du, L., Liu, S., et al. (2020). Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. Cell Res. *30*, 343–355.

Yi, Z., Ling, Y., Zhang, X., Chen, J., Hu, K., Wang, Y., Song, W., Ying, T., Zhang, R., Lu, H., and Yuan, Z. (2020). Functional mapping of B-cell linear epitopes of SARS-CoV-2 in COVID-19 convalescent population. Emerg. Microbes Infect. *9*, 1988–1996.

Zamecnik, C.R., Rajan, J.V., Yamauchi, K.A., Mann, S.A., Loudermilk, R.P., Sowa, G.M., Zorn, K.C., Alvarenga, B.D., Gaebler, C., Caskey, M., et al. (2020). ReScan, a Multiplex Diagnostic Pipeline, Pans Human Sera for SARS-CoV-2 Antigens. Cell Rep. Med. *1*, 100123.

Zheng, Z., Monteil, V.M., Maurer-Stroh, S., Yew, C.W., Leong, C., Mohd-Ismail, N.K., Cheyyatraivendran Arularasu, S., Chow, V.T.K., Lin, R.T.P., Mirazimi, A., et al. (2020). Monoclonal antibodies for the S2 subunit of spike of SARS-CoV-1 cross-react with the newly-emerged SARS-CoV-2. Euro Surveill. *25*, 2000291.

Zost, S.J., Gilchuk, P., Case, J.B., Binshtein, E., Chen, R.E., Nkolola, J.P., Schäfer, A., Reidy, J.X., Trivette, A., Nargi, R.S., et al. (2020). Potently neutralizing and protective human antibodies against SARS-CoV-2. Nature *584*, 443–449.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CR3022 monoclonal antibody	BEI Resources	Cat. # NR-52392
Goat anti-human IgG-Fc horseradish peroxidase (HRP)	Bethyl Labs	Cat. # A80-104P, RRID:AB_10630432
Bacterial and virus strains		
T7 Select 10-3b bacteriophage	EMD Millipore Sigma	Cat. # 70014-3
Escherichia coli, strain BLT5403	EMD Millipore Sigma	Cat. # 70548-3
Biological samples		
COVID-19 patient convalescent plasma	This paper	See Table S1
Normal Human Serum, collected 03/2006 – 02/2009 and 05/2014 – 07/2015	Gemini Bioscience	Cat. # 100-110, Lot # H87W00K
Chemicals, peptides, and recombinant proteins		
Q5 High-Fidelity 2X Master Mix	NEB	Cat. # M0492S
SARS-CoV-2 RBD protein	Roland Strong lab	N/A
SARS-CoV-2 Spike protein	Sino Biological	Cat. # 40589-V08B1
SARS-CoV-2 S2 protein	The Native Antigen Company	Cat. # REC31807-100
SARS-CoV-2 (COVID-19) Spike protein RBD-coupled magnetic beads	AcroBiosystems	Cat. # MBS-K002
1-Step Ultra TMB-ELISA Substrate Solution	ThermoFisher	Cat. # 34029
Sulfuric Acid Solution, 1N	Fisher Scientific	Cat. # SA212-1
Fugene-6	Promega	Cat. # E2692
Critical commercial assays		
QIAquick Gel Extraction Kit	QIAGEN	Cat. # 28704
KAPA Library Quantification Kit	KAPA Biosystems	Cat. # KK4828
AMPure XP	Beckman Coulter	Cat. # A63881
Quant-iT PicoGreen dsDNA Assay Kit	ThermoFisher	Cat. # P7589
Bright-Glo Luciferase Assay System	Promega	Cat. # E2620
Deposited data		
Demultiplexed Illumina sequencing reads	This paper	SRA: PRJNA715823
SARS-CoV-2 Wuhan Hu-1 complete genome	Wu et al., 2020	NCBI GenBank: MN908947
Sarbecovirus Spike protein sequences	NCBI	NCBI GenBank: YP_009724390.1, YP_009555241.1, YP_173238.1, YP_003767.1, NP_073551.1
Experimental models: cell lines		
Human: Embryonic Kidney (HEK293T)	ATCC	Cat. # CRL-11268
Human: Embryonic Kidney cells expressing human ACE2 (HEK293T-hACE2)	BEI Resources	Cat. # NR-52511
Recombinant DNA		
Plasmid: HDM-SARS2-Spike-delta21	Addgene	Cat. # 155130
Plasmid: Luciferase-IRES-ZsGreen	BEI Resources	Cat. # NR-52516
Plasmid: HDM-Hgpm2 gag/pol lentiviral helper plasmid	BEI Resources	Cat. # NR-52517
Plasmid: pRC-CMV-Rev1b	BEI Resources	Cat. # NR-52519
Plasmid: HDM-tat1b	BEI Resources	Cat. # NR-52518

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
Library pool primer, Fwd: 5′ AATGATACGGCAGGAATTCTACGCTGAGT 3′	Garrett et al., 2020	N/A
Library pool primer, Rvs: 5′ CGATCAGCAGAGGCAA GCTTGCTATCA 3′	Garrett et al., 2020	N/A
Illumina library prep primer, Round 1, Fwd: 5′ TCGTCGGCAGCGTCTCCAGTCAGGTG TGATGCTC 3′	Garrett et al., 2020	N/A
Illumina library prep primer, Round 1, Rvs: 5′ GTGGGCTCGGAGATGTG TATAAGAGACAGCAAGAC CCGTTTAGAGGCCC	Garrett et al., 2020	N/A
Illumina library prep primer, Round 2, Fwd: 5′ AATGATACGGCGACCACC GAGATCTACACNNNNNNN NTCGTCGGCAGCGTCTCCAGTC 3′	Garrett et al., 2020	N/A
Illumina library prep primer, Round 2, Rvs: 5′ CAAGCAGAAGACGGCATA CGAGATNNNNNNNNGTCTC GTGGGCTCGGAGATGTG TATAAGAGACAG 3′	Garrett et al., 2020	N/A
Custom Illumina sequencing primer: 5′ GCTCGGGGATCCGAATTCTACGCTGAGT 3′	Garrett et al., 2020	N/A
Software and algorithms		
Python		https://www.python.org/
DNAChisel (version 3.2.2)		https://github.com/Edinburgh-Genome-Foundry/DnaChisel
phip-flow	Matsen Lab	https://github.com/matsengrp/phip-flow
Nextflow		https://www.nextflow.io/
Bowtie		https://quay.io/biocontainers/bowtie:1.2.2%5fpy36h2d50403_1
phippery	Matsen Lab	https://github.com/matsengrp/phippery
xarray		http://xarray.pydata.org/en/stable/
SAMtools		https://quay.io/biocontainers/samtools:1.3%5fh0592bc0_3
R (version 4.0.2)		https://www.R-project.org/
tidyverse		https://www.tidyverse.org/
ggpubr		https://github.com/kassambara/ggpubr
corrr		https://github.com/tidymodels/corrr
cowplot		https://github.com/wilkelab/cowplot/
scales		https://github.com/r-lib/scales
rstatix		https://github.com/kassambara/rstatix
coin		http://coin.r-forge.r-project.org/
Custom code	This paper	https://github.com/meghangarrett/Spike-Phage-DMS
Clustal Omega		https://www.ebi.ac.uk/Tools/msa/clustalo/
GraphPad Prism version 9	GraphPad Software	N/A
Other		
Protein A Dynabeads	Invitrogen	Cat. # 10002D

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Protein G Dynabeads	Invitrogen	Cat. # 10004D
1.1 mL 96-deep-well polypropylene U-bottom plate	BrandTech	Cat. # 701350
Clear Flat-Bottom Immuno Nonsterile 384-Well Plates	Thermo Scientific	Cat. # 464718
Epoch plate reader	Biotek	N/A
Steriflip-GP Sterile Centrifuge Tube Top Filter Unit	Millipore Sigma	Cat. # SCGP00525
LUMIstar Omega plate reader	BMG Labtech	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and request for reagents may be directed to the corresponding author Julie Overbaugh (joverbau@fredhutch.org).

Materials availability

The data presented in this manuscript and research materials used in this study are available from the lead contact upon request.

Data and code availability

Sequencing data has been deposited to NCBI and are accessible under BioProject # PRJNA715823. The Nextflow pipeline, used to align all sample reads to the reference library, is available at https://github.com/matsengrp/phip-flow. A custom python package used for all downstream sample curation and analysis is available at https://github.com/matsengrp/phippery. All custom code and input files used to (1) generate oligonucleotide sequences for the S protein Phage-DMS library, (2) run the alignment pipeline, (3) analyze sequencing data for the experiments in this paper, and (4) to generate figures have been deposited at https://github.com/meghangarrett/Spike-Phage-DMS. Data is available to explore using the dms-view website, hosted here: https://github.com/meghangarrett/Spike-Phage-DMS/tree/master/analysis-and-plotting/dms-view.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

Plasma samples were taken from non-hospitalized COVID-19 patients enrolled in the Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) study at the University of Washington. Prior to study initiation, the following institutional human subjects review committee approved the protocol: University of Washington IRB (Seattle, Washington, USA) and concurrent approvals were obtained from the Fred Hutchinson Cancer Research Center for the current study. All plasma samples were heat-inactivated at 56°C for 1 hour before storage and use. Plasma samples were spun in a centrifuge for 10 min at 1,000 x g in order to clarify the supernatant before use.

METHOD DETAILS

Design and generation of the Spike Phage-DMS library

To create a Phage-DMS library for the S protein of SARS-CoV-2 we used the sequence from the Wuhan Hu-1 strain (GenBank: MN908947). Only the ectodomain of the S protein was included (aa 1-1211), excluding the transmembrane and cytoplasmic domains. Additionally, in the region of the D614G variant of the Wuhan Hu-1 strain, we designed DMS peptides spanning positions 599 through 619 that also include the D614G mutation. Sequences were optimized for uniform GC content (to reduce later biases during PCR amplification) and codon usage for expression in *E. coli*. GC and codon optimization was done using the Python package DNAChisel (version 3.2.2), aiming for GC content of between 0.4 and 0.6 within a window of 100 nucleotides. After optimizing the sequences, the two subunits of the S protein (S1 and S2) were then treated as separate proteins. Sequence coding for a glycine-serine linker ($[G_4S]_3$) was added to the beginning and end of the sequence of each protein in order to ensure that the first amino acid of the protein was located in the central position of the peptide. We then generated sequences coding for peptides 31 amino acids long, tiling by 30 amino acids, and containing a single variable residue at the central position of the peptide. This resulted in 20 peptide sequences containing all possible mutations at each position along the proteins, with only one amino acid shift between sequential peptides. Each sequence additionally had 5′ and 3′ adaptor sequences added to facilitate amplification and cloning (5′: AGGAATTCTACGCTGAGT, 3′: TGATAGCAAGCTTGCC). After removing duplicate sequences, 24,820 unique sequences were

synthesized by Twist Bioscience as an oligonucleotide pool. Two biological duplicate libraries were generated by independently cloning the sequences into a T7 phage vector and then amplifying the phage, as we have done previously (Garrett et al., 2020). Peptides are numbered by the corresponding S protein location of the amino acid in the central position of the peptide. Code used to optimize and generate peptide sequences for the S protein Phage-DMS library can be found at https://github.com/meghangarrett/Spike-Phage-DMS.

Immunoprecipitation of human plasma with Phage-DMS library

Immunoprecipitation of phage-antibody complexes was performed as previously described (Garrett et al., 2020; Mohan et al., 2018). Briefly, deep 96-well plates were blocked with 3% BSA in Tris-buffered saline with 0.01% Tween (TBST) overnight at 4°C. The phage library was diluted to a concentration representing 200,000 pfu/mL per unique peptide and 1 mL of the diluted phage was added to each well. We assume that plasma IgG concentrations are about 10 ug/uL (Mabuka et al., 2012) and added 10 ug of each sample to the appropriate wells. For every experiment, samples are run in technical duplicate on the same plate. Plates are sealed and rocked at 4°C for 18-20 hours. To immunoprecipitate the phage-antibody complexes we added 40uL of a 1:1 mixture of Protein A and Protein G Dynabeads to each well and incubated the samples for 4 hours at 4°C while rocking. Dynabeads were magnetically separated and then beads were washed 3x with 400 uL wash buffer (150 mM NaCl, 50 mM Tris-HCl, 0.1% [vol/vol] NP-40, pH 7.5). We resuspended beads in 40 uL of water and then lysed bound phage at 95°C for 10 minutes. Additionally, we lysed 10-20 million phage from the diluted input library to determine the distribution of phage in the starting library. Lysed samples were stored at −20°C before preparing for Illumina sequencing.

Illumina library preparation and deep sequencing

Lysed phage DNA from each sample was amplified and readied for Illumina deep sequencing by performing two rounds of PCR, as previously described (Garrett et al., 2020). Each PCR reaction was performed using Q5 High-Fidelity 2X Master Mix. For the first round of PCR, 10uL of lysed phage was used as the template in a 25 uL reaction. For the second round of PCR, 2 uL of the round 1 PCR product was then used as the template in a 50 uL reaction, with primers that add dual indexing sequences on either side of the insert. PCR products were then cleaned using AMPure XP beads and eluted in 50 uL water. DNA concentrations were quantified via Quant-iT PicoGreen dsDNA Assay Kit. Equimolar amounts of DNA from the samples, along with 10X the amount of the input library samples, was pooled, gel purified, and the final library was quantified using the KAPA Library Quantification Kit. Pools were sequenced on an Illumina MiSeq with 1x125 bp single end reads using a custom sequencing primer.

Depletion of RBD-binding antibodies from patient plasma

Depletion was performed as previously described (Greaney et al., 2021). In brief, magnetic beads conjugated with RBD protein were washed in PBS with 0.05% BSA and resuspended at 1mg/mL. Beads and plasma were incubated together at a ratio of 1:3 plasma: beads, and plasma was allowed to bind to beads overnight on a rotating rack at 4°C. Beads were then magnetically separated and the plasma supernatant was used in downstream assays (accounting for the 1:4 dilution factor that occurred during depletion). A mock depletion control for each plasma sample was also prepared in parallel, adding PBS with 0.05% BSA at a ratio of 1:3 plasma:buffer.

ELISA with RBD, Spike, and S2 proteins

ELISAs to test for plasma IgG binding to RBD, Spike, and S2 proteins were performed as previously described (Dingens et al., 2020; Stadlbauer et al., 2020), with some modifications. 384 well ELISA plates were coated with 25 uL of recombinant protein at a concentration of 2 ug/mL in phosphate buffered saline (PBS), with protein allowed to coat the wells overnight at 4°C. Plates were then washed 4x with 100 uL of wash buffer (PBS with 0.01% Tween-20 [PBST]) and blocked with 100 uL of 3% nonfat dry milk for 2 hours at room temperature. Block was then thrown off, and 25 uL of diluted plasma or antibody control was added to each well. Patient plasma was diluted 1:100 (accounting for prior dilution made during RBD-antibody depletion) in wash buffer with 1% nonfat dry milk and then four 3-fold serial dilutions were made, for a total of 5 dilutions. CR3022 was used as a positive control for RBD and Spike and was diluted to 1 ug/mL followed by nine 4-fold serial dilutions, for a total of 10 dilutions. Pooled plasma from the 18 patients in this study was used as the positive control for S2 and was diluted 1:50 followed by nine 4-fold serial dilutions, for a total of 10 dilutions. Pooled plasma collected pre-pandemic was used as a negative control, and was treated the same as patient plasma. Plates were incubated for 2 hours at 37°C and then samples were washed 4x with wash buffer. Goat anti-human IgG-Fc horseradish peroxidase (HRP)-conjugated antibody was diluted 1:3000 in wash buffer containing 1% nonfat dry milk, and 25 uL was added to the plate. After 1 hour at room temperature, plates were washed 4x with wash buffer and 25 uL of TMB substrate was added. After 15 minutes at room temperature, the reaction was stopped with 25 uL of 1N sulfuric acid and the OD450 was read on a BioTek Epoch plate reader. After subtracting background signal from buffer-only wells, the area under the titration curve was calculated using log-transformed dilution values in GraphPad Prism (v9).

Generation and titration of Spike pseudotyped lentivirus

We generated and determined the titers of pseudotyped lentivirus as previously described (Crawford et al., 2020). We used the codon-optimized Spike sequence from the Wuhan-Hu-1 strain with a 21 amino acid deletion in the cytoplasmic tail (also known

as HDM_Spikedelta21). HEK293T cells were first seeded at a density of 5 × 10^5 cells per well into 6-well plates with DMEM (supplemented with 10% fetal bovine serum (FBS), 2 mM l-glutamine, and penicillin/streptomycin/fungizone). After 16 to 24 hours, cells were transfected using FuGENE-6 with the Luciferase_IRES_ZsGreen backbone, the Gag/Pol-, Rev-, and Tat-containing lentiviral helper plasmids, as well as the plasmid containing Spike. 24 hours after transfection, we removed the media and replaced with fresh DMEM complete. At 50-60 hours post transfection we collected viral supernatants, filtered through a 0.22 um Steriflip filter, and stored at −80°C. To titer the viral supernatant, we plated 1.25 × 10^4 HEK293T-ACE2 cells per well in 50 uL in a 96-well black-walled plate and incubated 16 to 24 hours before adding 100 uL of viral supernatant dilutions to each well. Viral supernatants were diluted 1:10 in DMEM complete followed by seven 2-fold serial dilutions, and each dilution series was run in duplicate. 60 hours post-infection we removed 100 uL of media from each well and added 30 uL of the Bright-Glo reagent, and then read the relative luciferase units (RLU) on the LUMIstar Omega plate reader.

Neutralization assays

We performed neutralization assays as previously described (Crawford et al., 2020). We plated 1.25 × 10^4 HEK293T-ACE2 cells per well in 50 uL in a 96-well black-walled plate and incubated for 12 hours before infecting. Virus was diluted to achieve a final concentration of ~200,000 to 500,000 RLU/well, which was determined by titering as described above. Starting at 1:20 (accounting for final virus/plasma volume), six 3-fold serial dilutions of patient plasma, for a total of 7 dilutions, were made and then added to virus. The virus/plasma mixture was incubated at 37°C for 1 hour and then 100 uL was added to the cells. At 60 hours post infection we measured the RLU per well using the Bright-Glo reagent on the LUMIstar Omega plate reader. We calculated the fraction infectivity of the antibody-containing wells by normalizing to a media-only (no plasma) well infected with the same diluted virus. We calculated the neutralization titer 50% (NT50) using GraphPad Prism software by fitting a four-parameter nonlinear regression curve with the bottom fixed at 0 and the top fixed at 1.

Multiple sequence alignment

Alignment of FP amino acid sequences from SARS-CoV-2, OC43, HKU1, NL63, and 229E was done using Clustal Omega. GenBank sequences used are as follows: YP_009724390.1, YP_009555241.1, YP_173238.1, YP_003767.1, and NP_073551.1.

Graphical illustrations

Graphical abstract, Figure 1, and the diagram of S protein domains in Figure 2 were made in BioRender.com

QUANTIFICATION AND STATISTICAL ANALYSIS

Demultiplexing and alignment of Illumina reads

Demulitplexing and fastq file generation were performed by the Fred Hutch Genomics Core using Illumina MiSeq Reporter software. Demultiplexed sample reads were aligned to the reference library in parallel using a Nextflow data processing pipeline. The pipeline builds a Bowtie index from the peptide metadata by converting the metadata to fasta format and feeds it into the bowtie-build command. The low-quality end of the reads is trimmed to 93bp in order to match the reference lengths before performing end-to-end alignment and allowing for 0 mismatches. For each sample, we quantified the abundance of each peptide by using samtools-idxstats to count the number of reads mapped to each specific peptide in the reference library. The peptide counts were merged into an enrichment matrix organized by unique identifiers for each peptide and sample. The metadata tables were tied with the enrichment matrix into an xarray dataset using shared coordinate dimensions of the unique sample and peptide identifiers. We used this dataset organization as the starting point for all downstream sample curation and analysis.

Calculating enrichment and scaled differential selection of peptides

Enrichment and scaled differential selection were calculated as described previously (Garrett et al., 2020). To calculate enrichment, each peptide's pseudocount frequency was divided by a respective library pseudocount. Each pseudo count was defined as the raw count plus the ratio of the sum of each sample and library count, with a minimum value of p = 1. Differential selection of mutant amino acids was then calculated as the log-fold change between each mutant peptide and the wild-type peptide at a locus. By definition, the differential selection of a wild-type amino acid is always 0. The scaled differential selection of a mutant amino acid was then calculated by multiplying the differential selection of a mutant amino acid by the enrichment of the wild-type peptide at that position.

Data curation

All samples were screened at least once with each Spike Phage-DMS Library 1 and Library 2, but occasionally were screened multiple times with each library. To ensure we used the same amount of data with each sample, we examined the Pearson's correlation value of enrichment values for all peptides between all biological replicate experiments and chose the data from the best two correlated experiments for inclusion in the analyses presented here. If there was not at least one time point where the sample had a correlation of at least 0.5, then all samples from that patient were excluded from enrichment and scaled differential selection analyses. Peptides that were never sequenced in any experiment were removed from the analyses, and all enrichment and scaled differential selection values shown are the average of two biological replicate experiments.

Supplemental figures

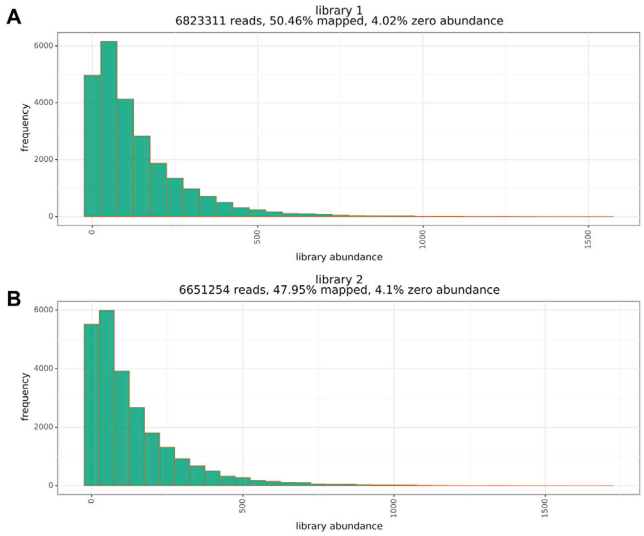


Figure S1. Distribution of sequenced peptides within biological replicate Spike Phage-DMS libraries, related to Figures 1 and 2

(A and B) Histogram showing the distribution of all sequenced peptides from a representative deep sequencing experiment for Spike Phage-DMS Library 1 (A) and Library 2 (B). Reads were stringently aligned to the reference library, allowing for 0 mismatches, and the proportion of unmapped reads is shown at the top. Additionally, the proportion of all non-sequenced peptides for each library is shown at the top.

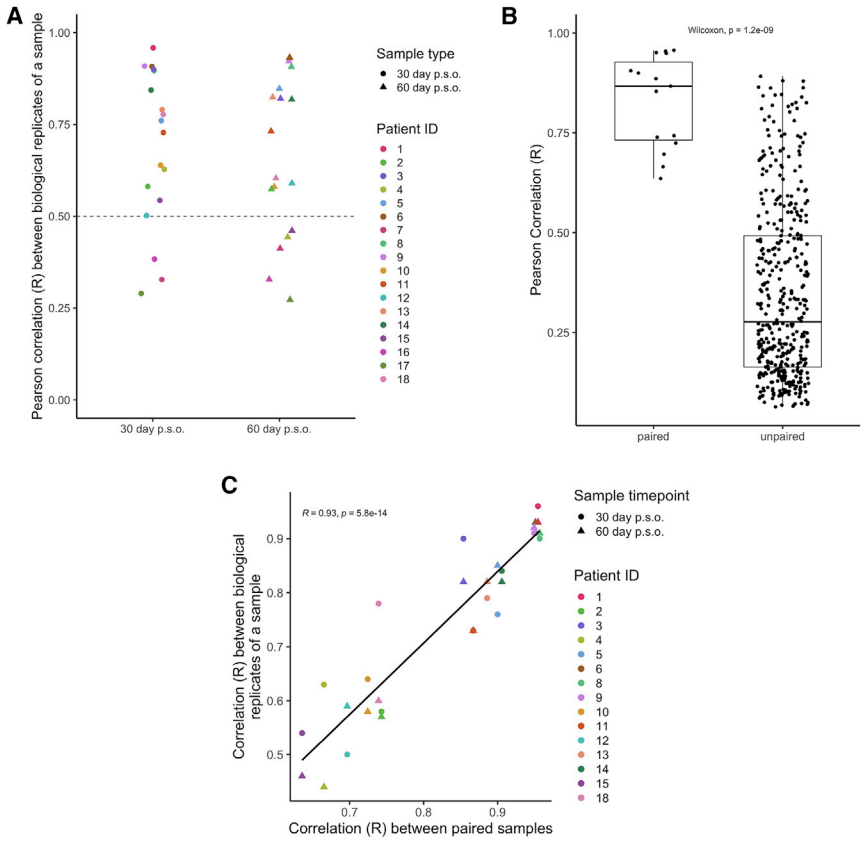


Figure S2. Reproducibility of peptide enrichment by plasma from COVID-19 patients, related to Figure 2

(A) Distribution of correlation values between peptide enrichment values for replicate experiments with samples from COVID-19 patients (Pearson's correlation coefficient, R). Each color corresponds to a unique patient or volunteer, and the shape of each dot represents the type of sample. A dotted line at y = 0.5 represents the cutoff used to determine whether samples were kept in the analysis. (B) Boxplots showing the distribution of correlation values between patient samples that were paired between the day 30 and 60 p.s.o. time points (on the left) or samples that were randomly paired and compared (on the right). (C) Relationship between the biological replicate correlation for a sample and its correlation with its paired time point. Each color corresponds to a unique patient, and the shape of each dot represents the type of sample. Pearson's correlation coefficient shown.

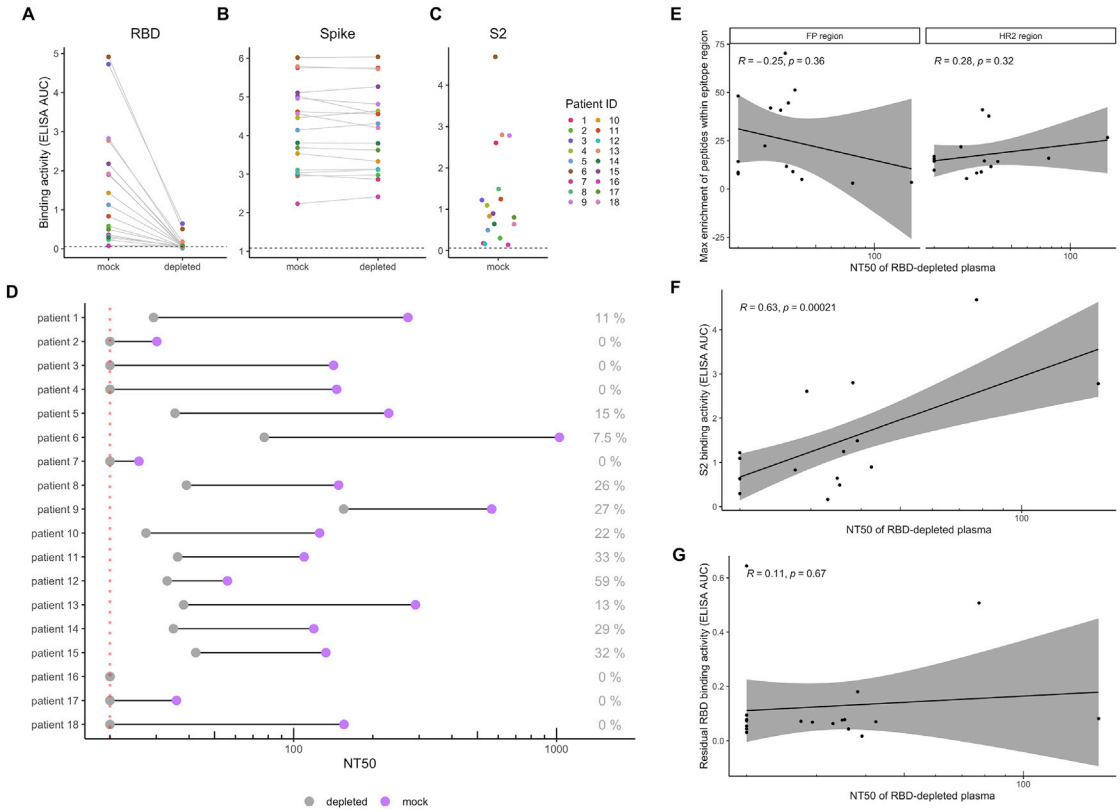


Figure S3. Plasma binding and neutralization with RBD-depleted plasma, related to Figure 2

(A and B) Change in plasma binding to RBD (A) and Spike (B) before and after depletion of RBD-binding antibodies, as measured by ELISA area under the curve (AUC). (C) Plasma binding to S2 subunit protein, as measured by ELISA AUC. Dotted lines indicate the lower limit of detection, as determined by pooled pre-pandemic serum. Each point is colored by patient as indicated to the right. (D) Comparison of neutralization titer 50% (NT50) before and after depletion of RBD binding antibodies. The percent residual neutralization for each patient is shown on the right. The dotted line indicates the lower limit of detection (NT50 of 20). Because the lowest dilution of plasma tested is 1:20, we cannot determine NT50 titers smaller than this. (E) Correlation between the most enriched peptide within the FP and HR2 epitope regions and the residual NT50 values for each patient plasma sample. (F) Correlation between plasma binding to S2 and residual NT50 values for each patient. All plasma tested in these assays were from the 30d time point.

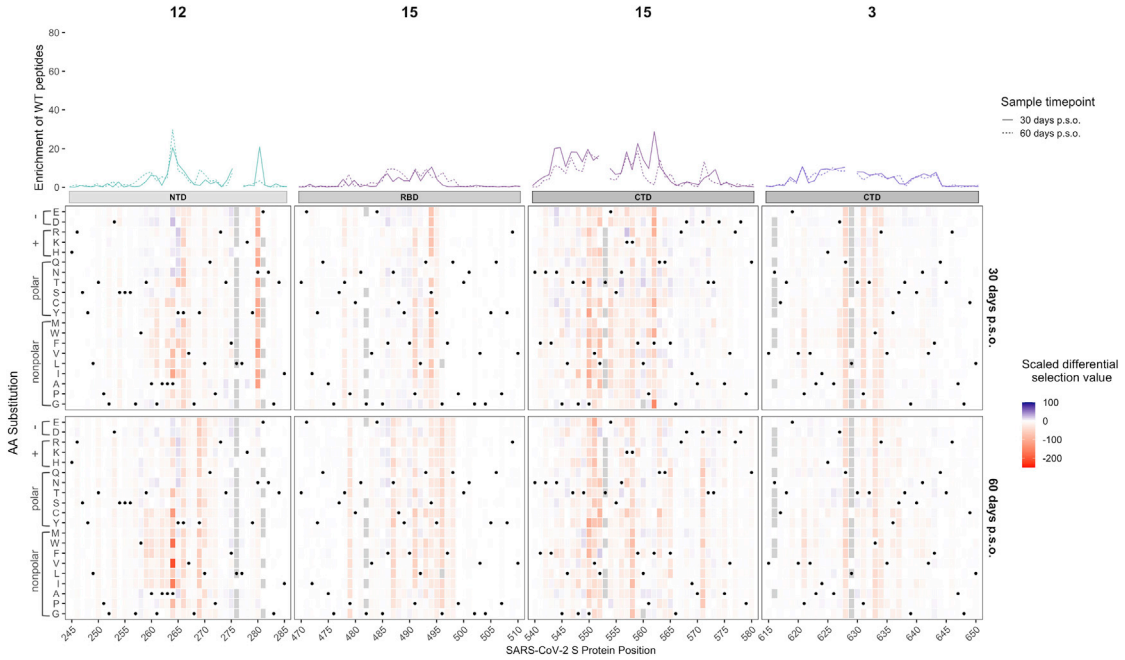


Figure S4. Effect of mutations on binding by COVID-19 patient plasma within various regions, related to Figure 2

Heatmaps depicting the effect of all mutations, as measured by scaled differential selection, at each site within the NTD, RBD, and CTD regions. Mutations enriched above the wild-type residue are colored blue and mutations depleted as compared to the wild-type residue are colored red. The wild-type residue is indicated with a black dot. Line plots showing the enrichment of wild-type peptides for each patient are shown above, with a solid line for patient samples taken at day 30 p.s.o. and a dashed line for patient samples taken at day 60 p.s.o. Peptides missing from the library are shown as gray boxes in the heatmaps and as breaks in the line plots.

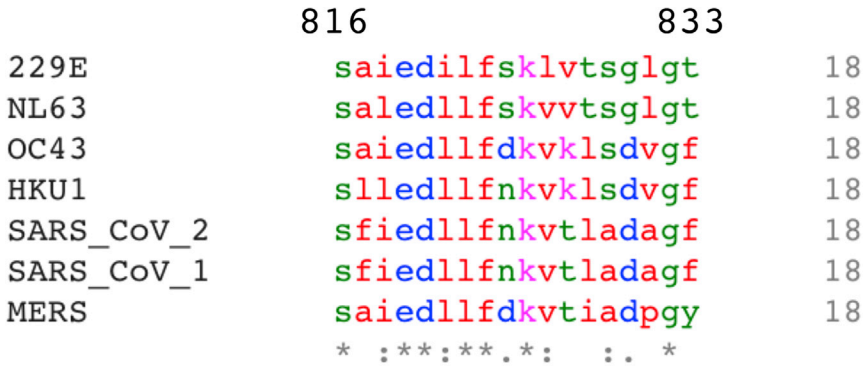


Figure S5. Multiple sequence alignment of the FP for SARS-CoV-2 and human endemic coronaviruses (OC43, HKU1, NL63, and 229E), related to Figure 3

Alignment was performed using Clustal Omega, and aa are colored according to physiochemical properties. Red indicates small and hydrophobic molecules (excluding Y), blue indicates acidic molecules, magenta indicates basic molecules (excluding H), and green indicates glycine, hydroxyl, sulfhydryl, and amine molecules. GenBank accession numbers: YP_009724390.1, YP_009555241.1, YP_173238.1, YP_003767.1, and NP_073551.1, respectively.

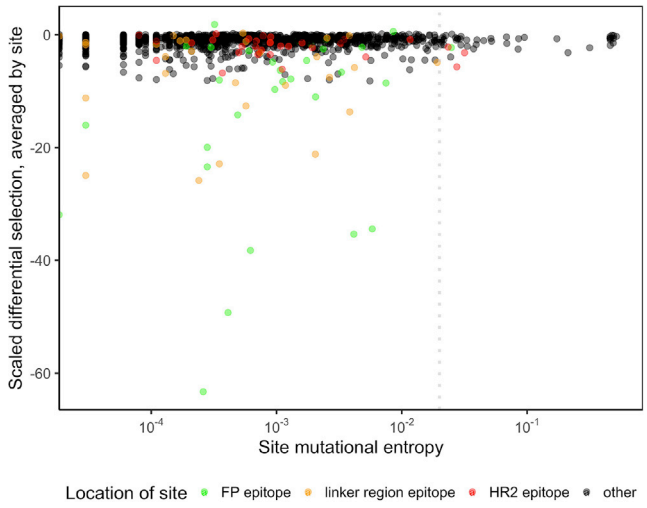


Figure S6. Effect of commonly circulating S protein variants on antibody escape for all patients, related to Figure 5

Scatterplot comparing the effect of mutations on patient plasma antibody binding and the frequency of all circulating S protein variants. The mutational entropy of every circulating protein variant, as reported at the http://cov.lanl.govcontent/index website and based on GISAID global sequencing, is plotted on the x axis. The average scaled differential selection values for all mutants, averaged across all patients, at each site is plotted on the y axis. Each site is colored by its location, as indicated on the bottom. The dotted line denotes the cutoff (0.02) of mutational entropy which GISAID uses to determine variants of interest.