Understanding SARS-CoV-2 antigenic evolution using complete genotype-to-phenotype maps of the effects of mutations on antibody binding

Allison J. Greaney

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Jesse D. Bloom, Chair

Neil P. King

Julie Overbaugh

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Understanding SARS-CoV-2 antigenic evolution using complete genotype-to-phenotype maps of the effects of mutations on antibody binding

Allison J. Greaney

Chair of the Supervisory Committee:
Professor Jesse D. Bloom
Fred Hutchinson Cancer Research Center

Since the COVID-19 pandemic began, global sequencing efforts have allowed scientists to follow the evolution of the SARS-CoV-2 virus in real time. The consequences of that evolution, however, are less readily apparent. Traditional methods to test the functional or antigenic effects of a new mutation can take weeks, resulting in a lag between surveillance and the corresponding experimental data. Thus, we developed a prospective approach to comprehensively measure the effects of mutations and generate complete genotype-to-phenotype maps for the SARS-CoV-2 spike receptor-binding domain (RBD).

The RBD binds to the angiotensin converting enzyme 2 (ACE2) receptor on host cells, mediating viral entry. We developed a yeast-display deep mutational scanning system to measure the effects of all possible single amino-acid mutations to the RBD on expression (a correlate of protein folding and stability) and ACE2 binding. We identified parts of the RBD that are mutationally constrained, as well as mutations that enhance expression or ACE2 binding. We created an interactive visualization to quickly identify the effects of any mutation to the RBD. This "lookup table" has been useful in interpreting the evolution of SARS-CoV-2 as new SARS-CoV-2 variants have emerged.

The RBD is also a major target of neutralizing antibodies that can block the virus' ability

to enter cells. We extended the deep mutational scanning system to comprehensively measure the effects of mutations to the RBD on antibody binding. We found that the resulting antibody-escape maps predict which mutations will be selected during viral growth in the presence of antibody. Additionally, we identified mutations that can escape binding and neutralization of antibody therapeutics used to treat COVID-19. As new SARS-CoV-2 variants emerged, these antibody-escape maps could predict which variants would be resistant to antibody treatment.

As greater fractions of the population gain immunity to SARS-CoV-2 through infection and/or vaccination, immune selection is likely to become a major driving force of the virus' evolution. I applied our deep mutational scanning system to measure the effects of RBD mutations on the binding of polyclonal antibodies from convalescent or vaccine-elicited plasmas and sera. We found that mutations to site 484 within one epitope (the "class 2" epitope) had some of the largest effects on antibody binding. For convalescent plasmas, mutations to site 484 often reduced neutralization to the same degree as removing all RBD-binding antibodies–indicating that the infection-elicited neutralizing response was highly focused on this one site. In late 2020, multiple SARS-CoV-2 variant lineages with mutations to site 484 began to emerge, suggesting that immune selection may already be shaping SARS-CoV-2 evolution.

In addition to evading preexisting immunity, new SARS-CoV-2 variants may also alter the specificity of the antibody response. We compared the antibody response elicited by infection with early 2020 viruses to that elicited by the B.1.351 (Beta) variant. We found that the B.1.351 variant induces an antibody response with a shifted immunodominance hierarchy that is more focused on a different epitope (the "class 3" epitope) spanning sites 443 to 452 in the RBD. While this epitope is conserved between early 2020 and B.1.351 viruses, it is mutated in the Delta lineage, which contains an L452R mutation, and has risen

to near-fixation in many countries. Thus, as SARS-CoV-2 continues to evolve, it will be necessary to reevaluate which mutations might evade immunity elicited by new variants. Overall, my thesis work has contributed to our understanding of how mutations affect the function and antigenicity of the SARS-CoV-2 RBD, and improve our ability to anticipate the consequences of the virus' evolution.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First, I would like to thank my thesis advisor, Jesse Bloom, for your mentorship and support. Thank you for all the scientific guidance, big and small, over the years. Thank you for teaching me how to code when I was a rotation student, for working closely with me on many aspects of my projects in the beginning of my graduate studies, and for empowering me to take more ownership over my projects as I progressed. Thank you for allowing me the opportunities to present my work at meetings and conferences, and to interface directly with collaborators. Thank you for your investment in my career and for ensuring that I grow as a scientist and future physician-scientist. Thank you also for your encouragement and frequent Slack messages of, "amazing progress on all this!". Thank you for supporting me and for kindness and understanding through the various challenges, personal and scientific throughout my time in the lab. I could not have chosen a better lab and thesis advisor.

I would also like to thank the other members of my thesis committee, Neil King, Julie Overbaugh, Trevor Bedford, and Maitreya Dunham. While we had zero pre-pandemic meetings together, I am grateful for the feedback, advice, and collaborations, both formal and informal.

One of the most rewarding parts of graduate school has been the opportunity to collaborate with scientists all over the world. I have learned so much from these many collaborators, and am grateful for their advice, assistance, and for expanding my scientific knowledge. Thank you to Seth Zost, James Crowe, Christopher Barnes, Pamela Bjorkman, Khadija Khan, Alex Sigal, and Cate Speake. I would especially like to thank Helen Chu for being such a wonderful collaborator and physician-scientist mentor. Thank you also to Caitlin

Wolf and Nicholas Franko for coordinating the HAARVI study.

Thank you to the administrative staff who have supported me and helped me get through the bureaucracies and challenges of grad school. Thank you to Cathy Lin, Alex Moreno, Brian Giebel, Marcie Buckner, Sara Carlson, Kim Kwon, and Carolina Chambers. Thank you to Luna Yu for IT support and for always bringing a sense of positivity and optimism to the lab.

Thank you to the amazing scientific core staff at the Fred Hutch, without whom I would have never been able to complete my dissertation research, especially Andy Marty, Dolores Covarrubias, Brian Raden, and Amira Davis. Despite the pandemic, you all went out of your way to ensure that our research was able to progress.

I would also like to thank my previous mentors for their encouragement and ongoing support of my scientific career. Thank you to Tera Levin, who I have looked up to as a scientist and mentor for more than 10 years. Your infectious passion for science first inspired me to attend graduate school. Berkeley could have been a very overwhelming and isolating experience, but you were the first "senior" person who I knew supported me personally and professionally. I certainly would have never gone to graduate school without your encouragement. I can't overstate how much I appreciate your continued support over the subsequent years! Thank you also to Mahtab Moayeri, for being such a supportive mentor and friend. I would never have had the courage to get an MD/PhD or go to graduate school if not for your mentorship, encouragement, and support. I learned so much about how to "do science" working with you. You taught me how to read a paper, attend a seminar, and talk to other scientists. Thank you for the skills and confidence that empowered me to be successful as a graduate student. I am grateful that your support and subtle encouragements continue even now.

I would also like to thank all members of the Bloom Lab with whom I have overlapped

and had the incredible privilege to work with over the years.

First, I would like to thank Tyler Starr, who has been a mentor to me for half of my PhD. Before you joined the Bloom Lab, you were one of my science "heroes," and I read all of your graduate school papers religiously. I am grateful that I got to work so closely with you for the past two years. I was fortunate to have the opportunity to learn from you and the way you do science. Thank you for taking the time to teach me so much and for answering my panicked Slack messages about last-minute experiment doubts or FACS sorting disasters at any hour of the day or night. Thank you for empowering me to take ownership over my own projects and helping me to gain confidence, knowledge, and mastery over my work. I had so much fun working with you, and I am really going to miss talking science and doing experiments with you when I leave the lab. As you move on to open your own lab, I hope that I can be counted as the first unofficial member of the Starr Lab.

Thank you to Rachel Eguia and Andrea Loes, with whom I was fortunate to collaborate. Thank you for all your help and contributions and for teaching me how to run lentivirus experiments and neutralization assays. Thank you Keara for having FACS parties with me in the basement during the darkest days of the pandemic. Thanks to all three of you for also keeping the lab running smoothly despite the pandemic. Thank you to Shirleen Soh, Juhye Lee, and Katherine Xue for emotional and scientific support early in my PhD. Thank you to Shirleen for being a mentor to me in the Bloom Lab, and for helping me to get through the challenges of the first years of grad school. Thank you to Sarah Hilton for being my computational mentor, and to Kate for being a senior MSTP student I could look up to, and someone I could always ask for help or advice, even when you were a busy 3rd year med student. I was so lucky to get to work with you during your last year in grad school.

I am eternally grateful to my parents, Jessie and Michael Greaney. Thank you for supporting me through all the ups and downs of graduate school. You always listened to me

complain about low points, and were quick to congratulate me on milestones along the way.

Finally, I must thank my twin sister and best friend, Samantha Greaney. Thank you for being my strongest supporter, and for always telling me that you knew that I could overcome every challenge, big or small. I am so happy that you moved to Seattle as a resident. I have enjoyed looking up to you as you have gained confidence as a physician, and as anyone who knows me knows, I am so proud of you.

Chapter 1

INTRODUCTION

The COVID-19 pandemic began in late 2019 with the emergence of the SARS-CoV-2 virus in the human population [353, 334, 355]. Thus, SARS-CoV-2 became the seventh coronavirus known to infect humans, after SARS-CoV-1 [60] and MERS-CoV [348], and the seasonal coronaviruses HKU1, NL63, OC43, and 229E. In the following 22 months, there have been more than 242 million confirmed COVID-19 cases and 4.9 million deaths globally [223], and massive disruptions to the global economy and many facets of daily life.

During this time, the global scientific community has focused tremendous efforts on understanding many aspects of SARS-CoV-2 biology, immunity, and host interactions. At the time of its emergence, however, little was known about SARS-CoV-2 or the immune response it would elicit. Moreover, it was unknown whether the SARS-CoV-2 virus would evolve increased fitness, ability to transmit among humans, or evasion of the immune response. My thesis work has aimed to address some of these central questions about SARS-CoV-2 and how to interpret its evolution.

In this introduction, I will provide background on the SARS-CoV-2 spike protein and its receptor-binding domain (RBD), and the humoral immune response to SARS-CoV-2 infection and vaccination. I will explain the need to understand which mutations can affect antibody binding, as well as previously established methods to identify such mutations. Finally, I will introduce the high-throughput technologies of deep mutational scanning, yeast display, and mutational antigenic profiling, which I have used in my thesis work to compre-

hensively identify mutations to the SARS-CoV-2 spike RBD that evade or reduce binding by monoclonal antibodies and infection- or vaccination-elicited polyclonal sera and plasmas.

1.1 The SARS-CoV-2 spike protein and receptor-binding domain

Cell entry of coronaviruses is mediated by the homotrimeric transmembrane spike protein and involves receptor binding, proteolytic processing of the spike protein, and membrane fusion [294]. The spike protein is comprised of the S1 (receptor-binding) and S2 (fusion machinery) subunits. For many coronaviruses, the spike protein is cleaved between S1 and S2, and remains non-covalently bound in the prefusion conformation [32, 208, 157, 308, 307]. The spike protein is subsequently cleaved by host proteases at the S2' site upstream of the fusion peptide [197, 208]. This activates the spike protein for fusion with the cell membrane [24, 308].

The host entry receptor for SARS-CoV-2, SARS-CoV-1, and some related coronaviruses is the cell-surface protein angiotensin converting enzyme 2 (ACE2) [183, 306, 332, 179]. Other coronaviruses, however, have different receptors. OC43 and HKU1 use the S1 domain A (also known as the N-terminal domain (NTD)) to recognize 5-N-acetyl-9-O-acetyl-sialic acid on cell-surface glycoproteins [295, 301]. The MERS-CoV spike NTD binds to sialic acids [185, 227], which is followed by binding of domain B (also known as the RBD) to the entry receptor, dipeptidyl-peptidase 4 (DPP-4) [243, 192].

SARS-CoV-2 spike's RBD binds to the ACE2 receptor with high affinity [139, 179, 306, 332, 268]. Because of its role in viral entry, the RBD is a major determinant of cross-species transmission and evolution for SARS-related coronaviruses (sarbecoviruses) more generally [184, 269, 179, 249]. Despite its important function, the RBD is highly variable among sarbecoviruses [141], reflecting the complex selective pressures shaping its evolution [196, 76, 103].

Prior work on SARS-CoV-1 RBD binding to ACE2 was informative early in the pandemic.

For instance, while the SARS-CoV-1 and SARS-CoV-2 RBDs have high amino-acid sequence similarity (73% to 76%), their receptor-binding motifs (RBMs) are only 50% to 53% identical [311]. The SARS-CoV-2 RBM, which has a higher affinity for ACE2 than SARS-CoV-1, contains amino-acid changes at key ACE2 contact residues including sites 455, 486, 493, 494, and 501 (SARS-CoV-2 numbering) [311], which contribute to this increased affinity.

Early work examining the interactions of human ACE2 with the RBDs of SARS-CoV-2 and related sarbecoviruses [306, 332, 268, 179] suggested that a better understanding of the effects of mutations on the RBD's affinity for ACE2 would inform on the potential for the SARS-CoV-2 RBD to undergo further evolution as it spread among humans. This knowledge would help to interpret the consequences of viral evolution. We address this question in **Chapter 2**.

1.2 The antibody response to the SARS-CoV-2 spike protein

In the early days of the pandemic, little was known about the antibody response to SARS-CoV-2. In fact, when we began our work, antibodies elicited against SARS-CoV-2 were difficult to access. The first antibodies that were studied for their ability to bind the SARS-CoV-2 RBD were isolated from humans previously infected with SARS-CoV-1 [230, 286, 290]. Today, there are over 4,000 monoclonal antibodies that have been described to bind SARS-CoV-2 [246], many of which now have published antibody gene sequences or high-resolution structures. Several antibody therapeutics to prevent or treat SARS-CoV-2 infection are currently in development or clinical use [168, 64].

Not only were there few characterized SARS-CoV-2 monoclonal antibodies, but in early 2020, little was known about the immune response that elicited by SARS-CoV-2 infection. Initial studies in animal models suggested that neutralizing antibodies against the spike protein were protective against SARS-CoV-2 infection and disease [7, 204, 306]. Later, observational studies in humans suggested that the presence of neutralizing antibodies against

SARS-CoV-2 spike elicited by prior infection would be protective against subsequent reinfection, at least in the short-term [6, 195]. It was only recently that neutralizing antibodies against the spike protein became well-established as a correlate of protection against SARS-CoV-2 infection [155], although the neutralizing antibody titer needed for protection is unknown.

There has also been much speculation about the longevity of immune protection against SARS-CoV-2. Many scientists turned to seasonal coronaviruses as a harbinger of short-lived immune protection: humans are repeatedly re-infected with the "common-cold" coronaviruses every few years [263, 87, 134]. The duration of protection against infection or re-infection with SARS-CoV-2 following a primary infection or vaccination remains somewhat unclear. Neutralizing antibody titers decline modestly in a pattern that is similar for other respiratory viruses [66, 145], memory B cells persist for at least 12 months [106, 253, 298], but protection against infection seems to wane modestly over the course of several months following vaccination [15, 241]. Many questions remain about the nature of the antibody response to SARS-CoV-2, its longevity, and correlates of protection.

Early SARS-CoV-2 antibody-isolation studies discovered that while only a small minority of anti-spike antibodies target the RBD, RBD-binding antibodies tend to be the most potently neutralizing, often competing with its binding to the ACE2 receptor [33, 38, 149, 188, 255, 267, 318, 356, 357, 336]. In addition, anti-RBD antibodies often dominate the neutralizing activity of the polyclonal antibody response elicited by natural infection [16, 282, 320]. Thus, RBD-binding neutralizing antibodies became leading therapeutic candidates [64] and one of the goals of vaccine development [165].

Further studies isolating, characterizing, and sequencing RBD-specific B cell clones found that naive common germline heavy-chain genes, including VH1-2, VH3-53/66, V3-30, and VH1-69, can bind the spike RBD [98], and that as a result, many individuals develop "convergent" antibody responses to SARS-CoV-2 infection [335, 251, 166, 218, 33, 38, 318, 346,

345, 335]. One consequence of this convergent antibody response is that SARS-CoV-2 infection elicits a neutralizing antibody response that is highly focused on one RBD epitope [120, 122, 320, 51]. A similarly unequal immunodominance of different epitopes has also been described for other viral antigens, such as influenza hemagglutinin [12, 13]. I explore these ideas further in **Chapters 6–8**.

My dissertation work focuses on antibodies that bind the spike's RBD. While the approaches used in my thesis work primarily measure antibody *binding* to the RBD, we have concentrated this effort on *neutralizing* monoclonal antibodies and understanding how these binding results relate to the neutralizing activity of polyclonal serum. However, neutralizing antibodies are only one type of immunity. Non-neutralizing antibodies are also elicited against the SARS-CoV-2 virus [33, 38, 149, 188, 255, 267, 318, 356, 357, 336], and antibody-dependent cell-mediated immunity is also elicited in response to infection [344, 55, 297] with SARS-CoV-2 and ADCC and Fc-mediated effector functions may be important for protection against the virus [326]. Our results are therefore also relevant for considering how RBD-binding antibodies that are non-neutralizing antibodies or that protect via cell-mediated mechanisms are affected by viral mutations. In addition, T cell immunity is an important component of the immune response to SARS-CoV-2 [266, 284, 203, 146], but our work does not examine the effects of mutations on T cell epitopes. In fact, T cell epitopes are often derived from proteins that are more conserved than spike, such as the N protein or ORF1a-encoded proteins [175], and thus may be relatively less affected by mutations that occur in viral variants [114].

1.2.1 The SARS-CoV-2 spike protein and RBD as potential vaccine antigens

The SARS-CoV-2 pandemic led to the development of vaccines with unprecedented speed [63, 165]. The first clinical trial of a vaccine candidate for SARS-CoV-2 began in March 2020 (NCT04283461 for Moderna's mRNA-1273 vaccine), only three months after SARS-CoV-2

infection of humans was first described [147, 63]. The mRNA vaccine platform, encoding the spike ectodomain with two stabilizing proline mutations in S2 (S-2P) [63], was the basis for the Moderna mRNA-1273 and Pfizer/BioNTech BNT162b2 vaccines that became the first SARS-CoV-2 vaccines authorized for emergency use in the United States [45, 46].

To date, there are 128 vaccine candidates in clinical development, and 194 in preclinical development [328]. These include protein subunit, viral vector, DNA, inactivated virus, mRNA, and other vaccine platforms [328]. While there was some early speculation that vaccines that exclusively targeted the spike protein might either fail to elicit strong protection or provide pathways for viral evolution to erode immunity [165], all SARS-CoV-2 vaccines currently in use in the United States target the spike protein via either an mRNA- or non-replicating adenoviral vector [45, 46, 44]. Early leading vaccine candidates such as the Moderna mRNA-1273 and Pfizer/BioNTech BNT162b2 mRNA-based vaccines, the Novavax spike protein nanoparticle vaccine, the Johnson&Johnson/Janssen Ad26-adenovirus vectored vaccine, and the ChAdOx1-adenovirus vectored vaccine were found to induce levels of neutralizing antibodies against SARS-CoV-2 comparable to that of natural infection [147, 310, 153, 100, 257].

Once vaccines became readily available in the United States, mechanistic studies of the B cell response to vaccination began. The response to mRNA vaccination is so far the most-studied, due to the availability of mRNA vaccines in the United States. Key work from the Nussenzweig lab has found that B cell antibody genes continue to evolve 6 months after infection and can increase in affinity for the RBD as well as sometimes ability to bind variants that have point mutations in the antibody epitope [106]. In the 6 months following vaccination, however, B cell genes do not evolve to increase in affinity or breadth [57]. For individuals who were previously infected and then later vaccinated, there is ongoing antibody somatic mutation and antibody evolution and the selective retention of B cell clones expressing broad and potent antibodies [314].

Important questions about the specificity of the antibody response elicited by vaccination and how this compares to infection, and whether SARS-CoV-2 would evolve to evade vaccine-elicited immunity remained. I work to address these questions in **Chapter 7**.

1.2.2 The possibility for antigenic evolution of SARS-CoV-2

When SARS-CoV-2 first emerged in the human population, one of the most pressing questions was whether the virus would eventually evolve to erode immunity. In the first months of the pandemic, the majority of the population was naive to the virus, so immune selection was not a major pressure driving the virus' evolution. Given the virus' continued spread, however, it was possible that immune selection could become an important driver of SARS-CoV-2 evolution.

This was such an open question because some viruses undergo rapid antigenic evolution while others are more antigenically stable. Influenza and measles viruses are both RNA viruses with similarly high mutation rates [259]. However, influenza evolves antigenically to erode antibody immunity [23], whereas measles is antigenically stable such that a vaccine developed almost 60 years ago still provides full protection against all currently circulating measles strains.

One hypothesis is that the surface proteins of some viruses are more functionally tolerant of mutations than others. Indeed, the surface proteins of rapidly evolving viruses such as influenza hemagglutinin are quite tolerant of mutations [84], whereas the measles virus hemagglutinin (H) protein is less mutationally tolerant [105]. However, one can readily select measles virus mutants that escape neutralization by individual monoclonal antibodies [25], so clearly the virus is tolerant to individual monoclonal antibody-escape mutations. Indeed, our work in **Chapter 2** suggested that the SARS-CoV-2 RBD would be relatively tolerant to single mutations, including within known antibody epitopes [280].

Antigenic evolution is influenced by more than mutational tolerance, however. The neu-

tralizing activity of the polyclonal antibody response to infection or vaccination can be narrowly focused on one or a few immunodominant epitopes, or broadly reactive to multiple codominant epitopes. For influenza virus, the polyclonal neutralizing antibody response is narrowly focused, such that single viral mutations can reduce neutralization by 10-fold or more [176]. Measles virus, on the other hand, elicits a neutralizing antibody response to numerous codominant epitopes [217] and is thus less readily evaded by mutations.

Whether SARS-CoV-2 would behave more like influenza or measles virus, however, was unclear. It was conjectured that coronaviruses in general may not undergo antigenic evolution due to their proofreading polymerase and lower mutation rate [79]. Evidence from before the COVID-19 pandemic suggested that antigenic evolution of coronaviruses was a real possibility. For instance, decades-old human-challenge studies found that humans infected with one strain of the 229E seasonal coronavirus were resistant to reinfection with that strain, but partially susceptible to a different strain [247]. This, rather than waning antibody immunity, could help to explain the frequency of reinfection with seasonal coronaviruses [263, 87, 134]. Recently, clear evidence for positive selection [158] and antigenic evolution [89] of seasonal coronaviruses has been demonstrated. Together, these findings indicated that there was a need to assess the potential for SARS-CoV-2 to evolve antigenically, and to consider antigenic evolution and immune escape in ongoing surveillance efforts. This was the focus of the majority of my dissertation work described in **Chapters 3-7**.

1.3 Methods to identify mutations that affect antibody binding

Given the possibility that SARS-CoV-2 may evolve over time to erode antibody immunity, there is a critical need to identify which mutations might reduce antibody binding to inform viral surveillance.

1.3.1 Genomic surveillance has enabled the real-time identification of viral variants

SARS-CoV-2 genomic surveillance has been performed with an intensity unseen for any previous virus [222]. These data are largely collected in the GISAID database, initially developed to prepare for a possible future influenza pandemic [91]. Early molecular epidemiology uncovered the first clear evidence of community transmission of SARS-CoV-2 in the United States [22], and subsequent surveillance efforts have yielded key insights into the introductions and spread of the virus through various parts of the world [325, 78, 330, 215]. Interactive tools to understand SARS-CoV-2 viral evolution and variants, including Nextstrain (nextrain.org) [126], CoVariants (covariants.org) [138], COVID-CG (covidcg.org) [50], and outbreak.info (outbreak.info) [214] have been useful for basic scientists and policy makers alike.

In the early days of the pandemic, genetic variation was used as molecular markers to track the spread of the virus. Today, we are increasingly concerned about the functional consequences of SARS-CoV-2 genetic variation. Molecular epidemiology has been used to identify variants with increased transmissibility, including the D614G variant [161], the B.1.1.7 (Alpha) variant [71, 237], and the B.1.617.2 (Delta) variant [239]. However, the experimental data needed to interpret the phenotypic consequences of viral evolution cannot be performed at the rapid pace of genomic sequencing.

For instance, existing methods to identify SARS-CoV-2 escape mutations by passaging virus in the presence of antibodies are incomplete in the sense that they only find one or a few of the possible escape mutations. Structural biology can more comprehensively define how an antibody physically contacts the virus, but structures are time consuming to determine and still do not directly report which viral mutations escape from antibody binding [70, 81, 148]. Instead, we chose to use the Bloom lab's expertise in high-throughput and prospective methods to comprehensively identify the mutations that might affect RBD function and antibody binding.

1.3.2 High-throughput methods to identify antibody-escape mutations

Deep mutational scanning is a powerful, high-throughput technique to quantify the fitness effects of thousands of mutations to a protein in parallel [101]. Rather than testing each mutant individually, all mutants are assayed together in a single experiment in bulk culture. The phenotype of each mutant is quantified by comparing its frequency in the bulk culture pre- vs. post-selection. In this manner, thousands of phenotypic measurements can be made simultaneously [101].

The Bloom lab has extended this approach to study how mutations to viral proteins such as influenza HA [288, 84, 85, 86, 177, 176] and HIV-Env [125, 81] affect viral growth and escape from neutralizing antibodies. This latter method of "mutational antigenic profiling" has dramatically increased the throughput of viral immune studies. These high-throughput *in vitro* immune selections can shed light on viral evolution in nature. Immune escape mutations identified in HA and HIV-Env *in vitro* also rose to high frequency in nature [176] and in antibody therapy clinical trials [81], respectively. These studies suggest that selection *in vitro* can mimic selection in humans.

Previous work from the Bloom lab has measured the phenotypic effects of mutations on influenza and HIV-1 viral proteins in the context of replicative virus. There would be, however, multiple technical challenges to applying a similar technique to the SARS-CoV-2 spike protein. First, SARS-CoV-2 must be studied under high biosafety conditions. Additionally, deep mutational scanning requires generating a high-diversity mutant library. For viruses, this requires a reverse-genetics system that can generate high-diversity libraries with high efficiency. While reverse-genetics systems to generate infectious clones for SARS-CoV-2 have been developed [338, 9, 140], these systems are relatively complex. Given the technical challenges and safety concerns of working with a large-scale library of SARS-CoV-2 infectious viruses, we decided to take an approach that would allow us to study the effects of mutations in high-throughput without working with infectious virus.

1.3.3 Yeast display as a deep mutational scanning platform

One method that avoids the concerns of live-virus work is phage display [275]. Phage display combined with deep mutational scanning has been used to identify antibody epitopes and mutations that reduce antibody binding for HIV-1 [113]. While this method is extremely high-throughput, a major limitation is that only relatively short, linear peptides can be assayed. Phage display has been used to identify linear SARS-CoV-2 spike epitopes bound by antibodies, largely in the fusion peptide, which contains linear epitopes [272, 112, 111]. Much of the spike protein and RBD, however, is comprised of conformational epitopes that would not be captured by this approach.

Yeast display has been used for protein design and high-throughput experiments for about 25 years, initially to engineer antibody variable domains with increased ligand affinity [28, 49, 108]. While yeast display has a lower throughput than phage display, and only about 1 million variants can be assessed in a reasonable experiment, the displayed protein of interest folds intracellularly. Thus, conformational epitopes can be assayed, and proteins that are several hundred amino acids in length can be used. Yeast-displayed proteins can be post-translationally modified, although yeast glycans are more mannose-rich than mammalian glycans [127]. Yeast display can also be used with fluorescence-activated cell sorting (FACS) for selection. Yeast display and FACS sorting of variants based on ligand binding signal intensity over a concentration gradient can also facilitate quantitative measurement of binding affinities [5].

A yeast display deep mutational scanning approach has also been extended to isolated domains of viral surface proteins. Deep mutational scanning of the stem region of influenza hemagglutinin has identified broadly reactive stem-binding antibodies and their escape mutations [109]. Yeast-expressed SARS-CoV-1 spike RBD had been used as a candidate vaccine antigen previously [53], and a soluble RBD-SD1 had been produced for SARS-CoV-2 [318], so we believed that it would be a yeast-display deep mutational scanning method might be

a feasible approach to study the effects of mutations to the SARS-CoV-2 RBD.

1.4 Layout of dissertation

In the following chapters, I describe my work uncovering the effects of mutations to the SARS-CoV-2 RBD on antibody binding and neutralization. The Bloom lab has previously used deep mutational scanning and mutational antigenic profiling to characterize the effects of mutations on the function and antigenicity of other viral surface proteins, such as influenza hemagglutinin [288, 84, 85, 86, 177, 176] and HIV-1 env [125, 81] proteins. When the SARS-CoV-2 virus emerged as a pandemic human pathogen, we were in a unique position to apply these high-throughput and prospective techniques to study the effects of mutations on the SARS-CoV-2 spike protein.

In **Chapter 2**, together in collaboration with Dr. Tyler Starr, a post-doctoral researcher in the Bloom lab, we measure the effects of all single amino-acid mutations to the SARS-CoV-2 spike RBD on its expression and ability to bind to the ACE2 receptor. We find that while most mutations are deleterious for RBD expression or ACE2 binding, there are a substantial number of mutations that are well-tolerated or even enhance ACE2 binding. At the time that the study was performed, there was relatively little genetic variation among human SARS-CoV-2 isolates. As with all SARS-CoV-2-related knowledge, however, things changed quickly. In the subsequent months, multiple viral lineages containing ACE2-enhancing mutations emerged and rose to high frequencies. This results of this first study continue to be a major resource for the SARS-CoV-2 community and are currently utilized by scientists involved in SARS-CoV-2 surveillance and decision-making. It also demonstrated the utility of the yeast-displayed deep mutational scanning platform for measuring the ability of RBD mutants to bind to soluble ligands beyond the ACE2 receptor.

In **Chapter 3**, Dr. Starr and I extended this platform to identify mutations to the RBD that can evade binding by monoclonal antibodies. We performed this proof-of-principle

high-throughput antibody-escape mapping with a panel of monoclonal antibodies that were isolated and first characterized by Seth Zost in James Crowe's lab [356, 357]. In collaboration with Pavlo Gilchuk from the Crowe lab, we used our deep mutational scanning data to inform the selection of a cocktail of two antibodies with non-overlapping escape mutations that was not escaped by any mutation in *in vitro* viral selection experiments.

Dr. Starr and I then applied this new approach in **Chapter 4** to map mutations that escape binding by antibodies in development for clinical use, the Regeneron antibody cocktail (comprised of REGN10933 and REGN10987, also known as casirivimab and imdevimab) [129] and the Eli Lily antibody LY-CoV016 (etesevimab) [271]. Importantly, these data were used over the subsequent months to evaluate the significance of the emergence of new viral variant lineages and to predict whether these emerging lineages would remain sensitive to neutralization by these clinical antibodies.

The long-term evolution of SARS-CoV-2, however, is far more likely to be shaped by the human polyclonal antibody response elicited by infection and vaccination than by monoclonal antibody therapeutics administered to a small number of patients. Thus, in **Chapter 5**, I extend this high-throughput approach that Dr. Starr and I developed for monoclonal antibodies to map mutations that reduce binding by polyclonal plasmas from individuals who were previously infected with SARS-CoV-2 in the early days of the pandemic. We found that while there was heterogeneity in the specificity of the antibody response among individuals and over time, in general, mutations to the class 2 antibody epitope [16] and in particular, site 484, had some of the largest effects on antibody binding and neutralization. It was rather remarkable that the antibody response was rather monoclonal-like, as has been observed for influenza [176]. This was an especially important finding in the context of our finding that the neutralizing response of SARS-CoV-2 convalescent plasmas was highly focused on the spike RBD, further cementing the importance of genomic surveillance to monitor for mutations to the RBD that might reduce antibody binding. Coincidentally,

shortly before this study was completed, several SARS-CoV-2 variant lineages emerged with mutations to site 484 [285, 94]. Our results were among the first studies to indicate that these variants would be substantially less susceptible to neutralization by convalescent plasmas. These neutralization experiments were critically performed by Dr. Andrea Loes in the Bloom lab.

In **Chapter 6**, I further explore the relationship between the specificities of monoclonal antibodies and polyclonal plasmas. I map mutations that reduce binding by a panel of monoclonal antibodies of various structural classes and compare these results to those for polyclonal convalescent plasmas. I find further evidence that infection with SARS-CoV-2 results in an antibody response that is relatively focused towards the class 2 site, and in particular, site 484.

One of the major scientific and technological advances that emerged as a direct result of the COVID-19 pandemic was the rapid development and deployment of remarkably effective mRNA vaccines targeting the SARS-CoV-2 spike protein. As these vaccines began to be distributed, we believed that it would be important to understand the specificity of the vaccine-elicited antibody response, and which viral mutations might evade this response. Thus, we mapped mutations that reduce vaccine-elicited antibody binding in **Chapter 7**. We found that at approximately 3 months post-vaccination or post-symptom onset, while vaccination elicited a neutralizing antibody response that was even *more* targeted towards the RBD than infection, within the RBD, the vaccine-elicited response was broader than the infection-elicited response, and relatively less affected by single mutations to the RBD. These results were obtained with key contributions by Dr. Loes with neutralization assays.

Finally, in the spring of 2021, it became clear that SARS-CoV-2 was evolving on a global scale, with the emergence of many new viral lineages, many of which had mutations with either enhanced ACE2 affinity, reduced antibody susceptibility, or both. Thus, in **Chapter 8**, we wanted to understand how mutations *in the background of these new variants* would

affect antibody binding and neutralization. Dr. Starr and I created a new deep mutational scanning library in the B.1.351 (or Beta) background and in collaboration with Dr. Alex Sigal's lab, mapped mutations that reduce binding of B.1.351-elicited plasmas. We found that the B.1.351 variant induces an antibody response with a shifted immunodominance hierarchy. Experiments in this chapter were performed with assistance from Rachel Eguia. While the B.1.351 variant has since been displaced by the Delta variant, these results suggest that as the SARS-CoV-2 virus continues to evolve, it will be necessary to understand which mutations might evade immunity elicited by new variants. As individuals gain increasingly complex immune histories from serial infections and/or vaccinations, these results have important implications for interpreting the consequences of viral antigenic evolution.

Chapter 2

DEEP MUTATIONAL SCANNING OF SARS-COV-2 RECEPTOR BINDING DOMAIN REVEALS CONSTRAINTS ON FOLDING AND ACE2 BINDING

A version of this chapter has been previously published as:

Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, Bloom JD. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell. 2020 Sep 3;182(5):1295-1310.e20. doi: 10.1016/j.cell.2020.08.012. Epub 2020 Aug 11. PMID: 32841599; PMCID: PMC7418704.

2.1 Abstract

The receptor binding domain (RBD) of the SARS-CoV-2 spike glycoprotein mediates viral attachment to ACE2 receptor, and is a major determinant of host range and a dominant target of neutralizing antibodies. Here we experimentally measure how all amino-acid mutations to the RBD affect expression of folded protein and its affinity for ACE2. Most mutations are deleterious for RBD expression and ACE2 binding, and we identify constrained regions on the RBD's surface that may be desirable targets for vaccines and antibody-based therapeutics. But a substantial number of mutations are well tolerated or even enhance ACE2 binding, including at ACE2 interface residues that vary across SARS-related coronaviruses. However, we find no evidence that these ACE2-affinity enhancing mutations have been selected in current SARS-CoV-2 pandemic isolates. We present an interactive visualization

and open analysis pipeline to facilitate use of our dataset for vaccine design and functional annotation of mutations observed during viral surveillance.

2.2 Introduction

The SARS-related (sarbecovirus) subgenus of betacoronaviruses comprises a diverse lineage of viruses that circulate in bat reservoirs and spill over into other mammalian species [30]. Sarbecoviruses initiate infection by binding to receptors on host cells via the viral spike protein. The entry receptor for both SARS-CoV-2 and the original SARS-CoV (which we refer to here as SARS-CoV-1) is the human cell-surface protein angiotensin converting enzyme 2 (ACE2). The receptor binding domain (RBD) of spike from both these viruses binds ACE2 with high affinity [139, 179, 306, 332, 183]. Because of its role in viral entry, the RBD is a major determinant of cross-species transmission and evolution [184, 269, 179, 249]. In addition, the RBD is the target of the most potent anti-SARS-CoV-2 neutralizing antibodies identified to date [38, 149, 230, 255, 267, 271, 336, 356]), and several promising vaccine candidates use RBD as the sole antigen.

Despite its important function, the RBD is highly variable among sarbecoviruses [141], reflecting the complex selective pressures shaping its evolution [196, 76, 103]. Furthermore, RBD mutations have already appeared among SARS-CoV-2 pandemic isolates, including some near the ACE2-binding interface—but their impacts on receptor recognition and other biochemical phenotypes remain largely uncharacterized. Therefore, comprehensive knowledge of how mutations impact the SARS-CoV-2 RBD would aid efforts to understand viral evolution and guide the design of vaccines and other countermeasures.

To address this need, we used a quantitative deep mutational scanning approach [101, 5] to experimentally measure how all possible SARS-CoV-2 RBD amino-acid mutations affect ACE2-binding affinity and protein expression (a correlate of protein folding stability). The resulting sequence-phenotype maps illuminate the forces that shape RBD evolution, quan-

tify constraint on antibody epitopes, and suggest that purifying selection is the main force acting on RBD mutations observed in human SARS-CoV-2 isolates to date. To facilitate use of our measurements in immunogen design and viral surveillance, we provide interactive visualizations, an open analysis pipeline, and complete raw and processed data.

2.3 Results

2.3.1 Yeast display of RBDs from SARS-CoV-2 and related sarbecoviruses

To enable rapid functional characterization of thousands of RBD variants, we developed a yeast surface-display platform for measuring expression of folded RBD protein and its binding to ACE2 [5, 28]. This platform enables RBD expression on the cell surface of yeast (Figure 2.1B), where it can be assayed for ligand-binding affinity or protein expression levels, a close correlate of protein folding efficiency and stability [163, 164]. Because yeast have protein-folding quality control and glycosylation machinery similar to mammalian cells, they add N-linked glycans at the same RBD sites as human cells [53], although these glycans are more mannose-rich than mammalian-derived glycans [127]. Yeast-expressed RBD from SARS-CoV-1 has similar antigenic and structural properties to RBD expressed in mammalian cells and binds to ACE2 as expected [53].

To validate the yeast-display platform, we selected RBDs from the Wuhan-Hu-1 SARS-CoV-2 isolate and six related sarbecoviruses (Figure 2.1A). These other sarbecoviruses include the closest known relatives of SARS-CoV-2 from bats and pangolins (RaTG13 and GD-Pangolin), SARS-CoV-1 (Urbani strain) and a close bat relative (LYRa11), and two more distantly related bat sarbecoviruses (BM48-31 and HKU3-1). Based on prior work, all these RBDs are expected to bind human ACE2 except those from BM48-31 and HKU3-1 [171, 179, 268]. We cloned the RBDs into a vector for yeast-display, induced RBD expression, and incubated with varying concentrations of fluorescently labeled human ACE2 (Figure 2.1B). We then used flow cytometry to measure ACE2 binding across 11 ACE2 con-

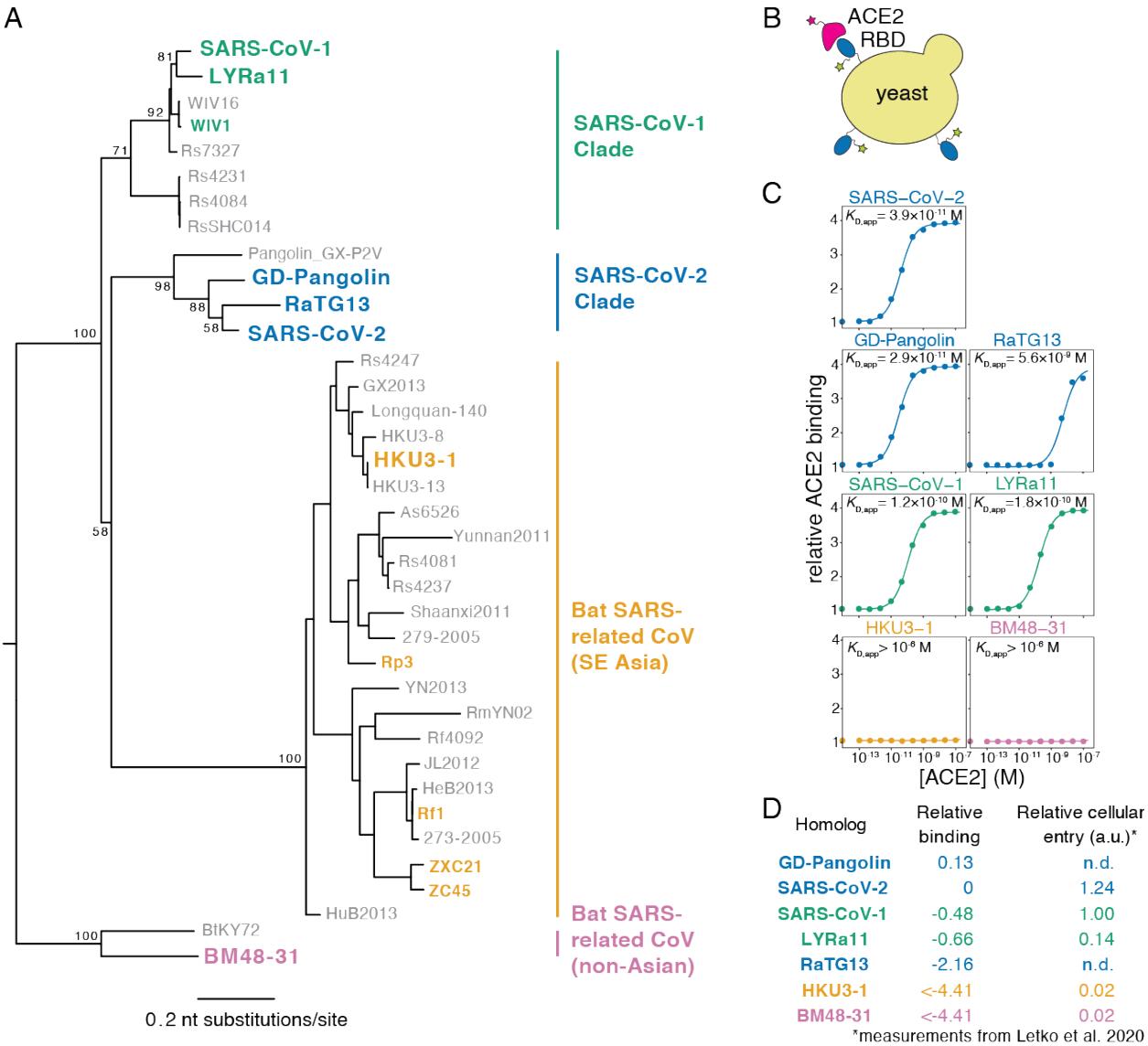
centrations, enabling the calculation of a dissociation constant for the binding of each RBD to ACE2 (Figure 2.1C). Because we used ACE2 in its native dimeric form [341], we refer to the measured constants as apparent dissociation constants ($K_{\mathrm{D,app}}$) which are affected by binding avidity. We report log binding constants $\Delta \log_{10}(K_{\mathrm{D,app}})$ relative to the wildtype SARS-CoV-2 RBD, polarized such that a positive value reflects stronger binding (Figure 2.1D).

All RBDs exhibited ACE2 binding affinities consistent with prior knowledge. We measure $K_{\mathrm{D,app}} = 3.9 \times 10^{-11}$ M for the SARS-CoV-2 RBD (Figure 2.1C), which is tighter than affinities reported for monomeric ACE2 [268, 306, 332] due to avidity effects caused by our use of native dimeric ACE2. Consistent with previous studies [268, 306, 332], the SARS-CoV-1 RBD binds ACE2 with lower affinity than SARS-CoV-2 (Figures 2.1C,D). The SARS-CoV-1-related bat strain LYRa11 binds with even lower affinity, while the more distant bat RBDs (HKU3-1 and BM48-31) have no detectable binding. These measurements are consistent with the ability of these RBDs to enable viral particles to enter cells expressing human ACE2 [179] (Figure 2.1D). Within the newly described SARS-CoV-2 clade, GD-Pangolin binds ACE2 with slightly higher affinity than SARS-CoV-2, while the bat isolate RaTG13 binds with two orders of magnitude lower affinity, consistent with prior reports [268, 333]. These results validate our yeast surface-display platform for RBD affinity measurements, and map variation in ACE2 affinity within the SARS-CoV-2 clade and the broader sarbecovirus subgenus.

2.3.2 Deep mutational scanning of all amino-acid mutations to the SARS-CoV-2 RBD

We next integrated the yeast-display platform with deep mutational scanning to determine how all amino-acid mutations to the SARS-CoV-2 RBD impact expression and binding affinity for ACE2. We constructed two independent mutant libraries of the RBD using a PCR-based mutagenesis method that introduces all 19 mutant amino acids at each position

Figure 2.1: **Yeast display of RBDs from SARS-CoV-2 and related sarbecoviruses**. **(A)** Maximum likelihood phylogeny of sarbecovirus RBDs. RBDs included in the present study are in bold colored text. Node labels indicate bootstrap support. **(B)** RBD yeast surface-display enables fluorescent detection of RBD expression and ACE2 binding. **(C)** Yeast displaying the indicated RBD were incubated with varying concentrations of human ACE2, and binding was measured via flow cytometry. Binding constants are reported as $K_{\mathrm{D,app}}$ from the illustrated titration curve fits. **(D)** Comparison of yeast display binding with previous measurements of the capacity of viral particles to enter ACE2-expressing cells. Relative binding is $\Delta \log_{10}(K_{\mathrm{D,app}})$ measured in the current study; relative cellular entry is infection of ACE2-expressing cells by VSV pseudotyped with spike containing the indicated RBD, reported by [179] in arbitrary luciferase units relative to SARS-CoV-1 RBD; n.d. indicates not determined by Letko et al. [179].



D

Homolog	Relative binding	Relative cellular entry (a.u.)*
GD-Pangolin	0.13	n.d.
SARS-CoV-2	0	1.24
SARS-CoV-1	-0.48	1.00
LYRa11	-0.66	0.14
RaTG13	-2.16	n.d.
HKU3-1	<-4.41	0.02
BM48-31	<-4.41	0.02

*measurements from Letko et al. 2020

[26]. To facilitate sequencing and obtain linkage among amino-acid mutations within a single variant, we appended 16-nucleotide barcode sequences downstream of the coding sequence [136], bottlenecked each library to approximately 100,000 barcoded variants, and linked each RBD variant to its barcode via long-read PacBio SMRT sequencing [198]. By examining the concordance of RBD variant sequences for barcodes sampled by multiple PacBio reads, we validated that this process correctly determined the sequence of $> 99.8\%$ of the variants. RBD variants contained an average of 2.7 amino-acid mutations, with the number of mutations per variant roughly following a Poisson distribution. Our libraries covered 3,804 of the 3,819 possible RBD amino-acid mutations, of which 95.7% were present as the sole amino-acid mutation in at least one barcoded variant. To provide internal standards for our measurements, we spiked the mutant libraries with a barcoded panel of 11 unmutated sarbecovirus RBD homologs (strains in color in Figure 2.1A), including those tested in the isogenic assays in Figure 2.1C.

To determine how mutations affect RBD expression and ACE2 binding, we combined fluorescence-activated cell sorting (FACS) with deep sequencing of variant barcodes [5]. To measure expression, we fluorescently labeled RBD protein on the yeast surface via a C-terminal epitope tag and used FACS to collect about 15 million cells from each library, partitioned into four bins from low to high expression (Figure 2.2A). We sequenced the barcodes from each bin and reconstructed each variant's mean fluorescence intensity (MFI) from its distribution of reads across bins. We represent expression as $\Delta \log$ (MFI) relative to the unmutated SARS-CoV-2 RBD, such that a positive $\Delta \log$(MFI) indicates increased expression. To measure ACE2-binding affinity, we incubated yeast libraries that had been pre-sorted for RBD expression with 16 concentrations of fluorescently labeled ACE2 (10^{-6} to 10^{-13} M, plus 0M ACE2), and used FACS to collect > 5 million RBD+ yeast cells at each concentration, partitioned into 4 bins from low to high ACE2 binding (Figures 2.2B). We again sequenced the barcodes from each bin, reconstructed the mean ACE2 binding of each

variant at each concentration, and used the resulting titration curves to infer dissociation constants $K_{\mathrm{D,app}}$ which we represent as $\Delta \log_{10}(K_{\mathrm{D,app}})$ relative to the unmutated SARS-CoV-2 RBD, with positive values indicating stronger binding.

These high-throughput measurements of expression and ACE2 binding were consistent with expectations about the effects of mutations. RBD variants containing premature stop codons universally failed to express folded full-length protein (Figure 2.2C). Unmutated variants and those with synonymous mutations had a tight distribution of neutral expression and binding measurements (Figure 2.2C,D). Variants containing amino-acid mutations had a wide range of expression and binding phenotypes, with variants containing just one mutation tending to have more mild functional defects than those with multiple mutations (Figure 2.2C,D). These trends are consistent with the fact that most mutations are deleterious to protein folding or function [277] – however, some mutated variants exhibit expression or binding that is comparable or even slightly higher than the parental SARS-CoV-2 RBD. The panel of RBD homologs from other sarbecovirus strains all expressed well but exhibited a wide range of ACE2 binding affinities (Figure 2.2C,D), as expected since only some are derived from viruses that can enter cells using human ACE2 [179].

These measurements show that the RBD possesses considerable mutational tolerance (Figure 2.2C,D). For instance, 46% of single amino-acid mutations to SARS-CoV-2 RBD maintain an affinity to ACE2 at least as high as that of SARS-CoV-1, suggesting that there is a substantial mutational space consistent with sufficient affinity to maintain human infectivity. Many single amino-acid mutants also maintain expression comparable to unmutated SARS-CoV-2, indicating that a large mutational space is compatible with properly folded RBD protein.

We next aggregated the measurements on all variants to quantify the effects of individual amino-acid mutations. Because many variants contain multiple mutations, we used global epistasis models to determine the effects of individual mutations from both singly and mul-

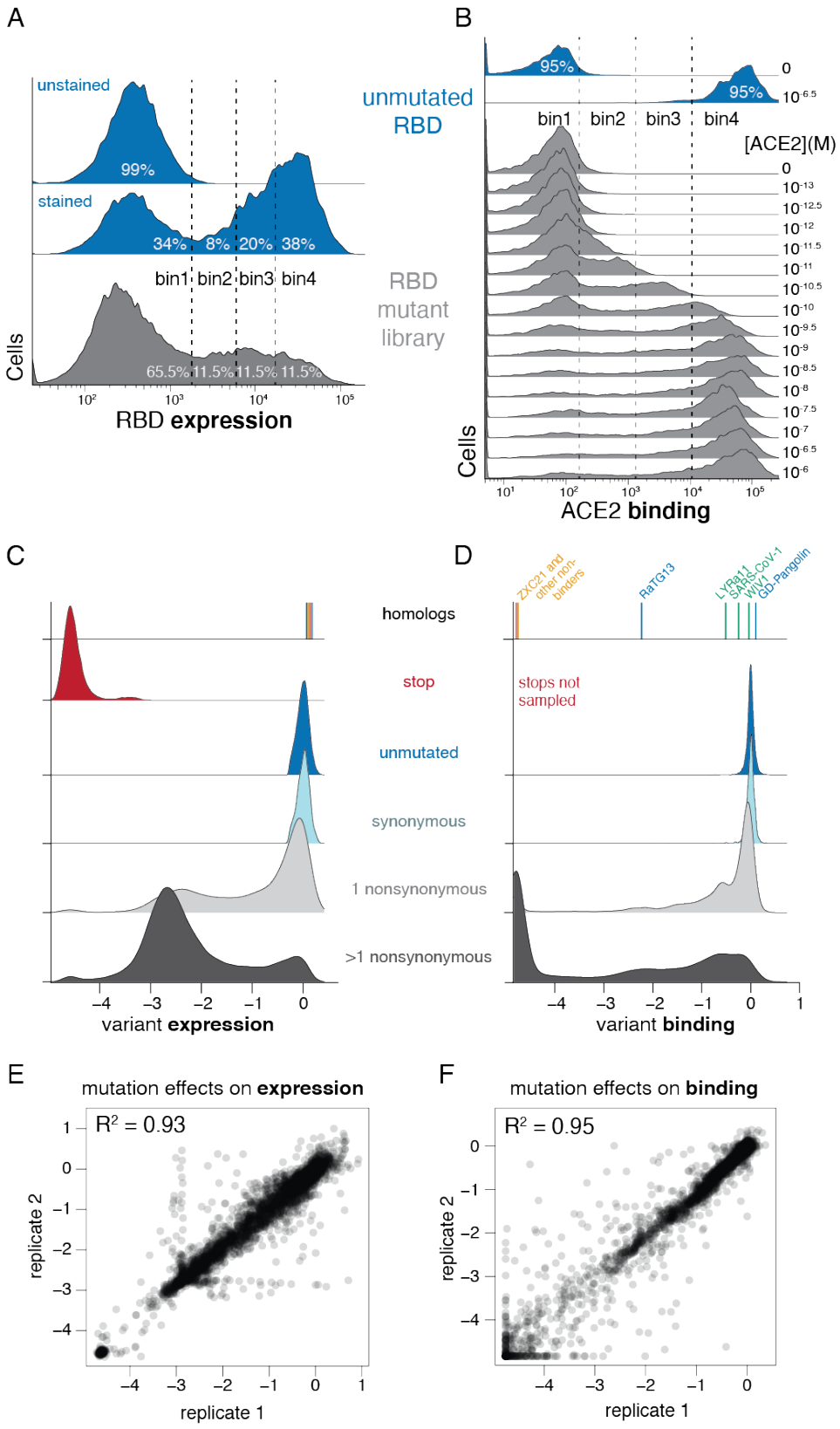
tiply mutated variants [220]. The resulting single-mutant $\Delta \log(\text{MFI})$ and $\Delta \log_{10}(K_{\text{D,app}})$ measurements correlated well between the independent library duplicates ($R^2 = 0.93$ and 0.95, respectively; Figures 2.2E,F). Throughout the rest of this paper, we report single mutant effects as the average of the duplicate measurements. Overall, we obtained expression measurements for 99.5% and binding measurements for 99.6% of all 3,819 single amino-acid mutations.

Visualization of sequence-to-phenotype maps

The complete measurements of how amino-acid mutations affect expression and ACE2 binding represent rich sequence-to-phenotype maps for the RBD. We visualize the data in several ways. Figure 2.3 provides heatmaps that show how each mutation affects expression or ACE2 binding, with sites annotated by whether they contact ACE2, their relative solvent accessibility, and their amino-acid identities in SARS-CoV-2 and SARS-CoV-1. Interactive versions of these heatmaps are at `https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS`, and enable zooming, subsetting by functional annotations, and mouse-selection based readouts of numerical measurements for individual mutations. Finally, interactive structure-based visualizations using dms-view [137] are linked at `https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS/structures/`, and project the effects of mutations onto a crystal structure of the ACE2-bound RBD [172] and a cryo-EM structure of the full spike ectodomain [306].

The sequence-phenotype maps reveal tremendous heterogeneity in mutational constraint across the RBD. Many sites are highly tolerant of mutations with respect to one or both of expression and ACE2 binding, while other sites are constrained to the wildtype amino acid. A substantial number of sites (e.g., 382 to 395) are quite tolerant of mutations with respect to ACE2 binding, but are constrained with respect to expression—consistent with folding and stability being global constraints common to many sites [93, 235]. There are also a handful of sites where ACE2 binding imposes strong constraints but expression does not

Figure 2.2: **Deep mutational scanning of all amino-acid mutations to the SARS-CoV-2 RBD**. **(A,B)** FACS approach for deep mutational scans for expression (A) and binding (B). Cells were sorted into four bins from low to high expression or binding, with separate sorts for each ACE2 concentration. The frequency of each library variant in each bin was determined by Illumina sequencing of the barcodes of cells collected in that bin, enabling reconstruction of per-variant expression and binding phenotypes. Bin boundaries were drawn based on distributions of expression or binding for unmutated SARS-CoV-2 controls (blue), and gray shows the distribution of library variants for library replicate 1 in these bins. **(C, D)** Distribution of library variant phenotypes for expression (C) and binding (D), with variants classified by the types of mutations they contain. Internal control RBD homologs are indicated with vertical lines, colored by clade as in Figure 2.1A. Stop-codon-containing variants were purged by an RBD+ pre-sort prior to ACE2 binding measurements, and so are not sampled in (D). **(E, F)** Correlation in single-mutant effects on expression (E) and binding (F), as determined from independent mutant library replicates.


A
unstained
99%
stained
34%
8%
20%
38%
bin1 bin2 bin3 bin4
Cells
65.5%
11.5%
11.5%
11.5%
RBD expression
B
unmutated RBD
95%
95%
0
bin1 bin2 bin3 bin4
[ACE2](M)
RBD mutant library
Cells
ACE2 binding
C
homologs
stop
unmutated
synonymous
1 nonsynonymous
>1 nonsynonymous
variant expression
D
ZXC21 and other non-binders
RaTG13
LYRa11
SARS-CoV-1
WIV1
GD-Pangolin
stops not sampled
variant binding
E
mutation effects on expression
R² = 0.93
replicate 2
replicate 1
F
mutation effects on binding
R² = 0.95
replicate 2
replicate 1

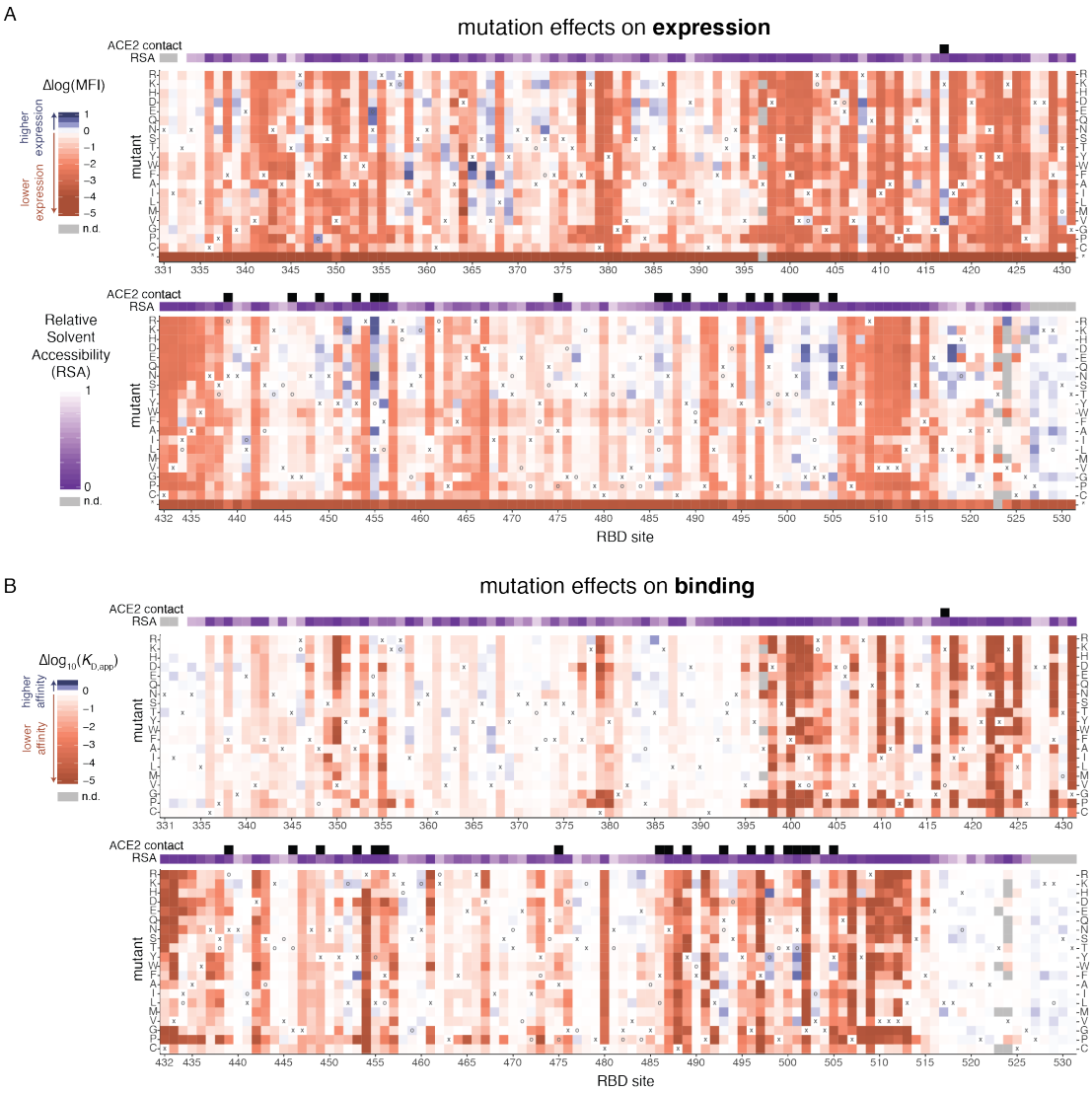


Figure 2.3: **Sequence-to-phenotype maps of the SARS-CoV-2 RBD**. **(A, B)** Heatmaps illustrating how all single mutations affect RBD expression (A) and ACE2 binding affinity (B). Interactive versions of these heatmaps are at `https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS`. Squares are colored by mutational effect according to scale bars on the left, with red indicating deleterious mutations. The SARS-CoV-2 amino acid is indicated with an 'x', and the SARS-CoV-1 amino acid, if different, is indicated with an 'o'. Black boxes in top overlay indicate residues that contact ACE2 in the SARS-CoV-2 (PDB 6M0J) or SARS-CoV-1 (PDB 2AJF) crystal structures. The purple overlay represents the relative solvent accessibility (RSA) of a residue in the ACE2-bound SARS-CoV-2 crystal structure.

(e.g. 489, 502, and 505). Moreover, at some sites there are mutations that clearly enhance expression or ACE2-binding affinity (blue colors in Figure 2.3).

2.3.3 Validation of deep mutational scanning measurements

We performed a series of experiments to confirm the dynamic range of our assays and their relevance for RBD expressed in mammalian cells or full spike trimer on pseudotyped lentiviral particles (Figures 2.4). To validate the dynamic range of our deep mutational scanning, we first re-cloned and tested RBD mutants in isogenic yeast-display assays. These experiments recapitulated the deep mutational scanning (Figures 2.4A-C), including confirmation that some mutations enhance expression (V367F and G502D) or ACE2 affinity (N501F, N501T, and Q498Y) in the context of yeast-expressed RBD.

We next compared our deep mutational scanning to measurements on mammalian-expressed RBDs. We purified mammalian- expressed RBDs from six sarbecoviruses (SARS-CoV-2, SARS-CoV-1, WIV1, RaTG13, ZXC21, and ZC45), and measured their 1:1 binding affinities for monomeric human ACE2 using biolayer interferometry, which agreed with the measurements from our deep mutational scan (Figures 2.4D). Moreover, we observed that using a natively dimeric ACE2 enables detection of binding by the RaTG13 RBD, which can support ACE2-mediated cell entry [268] even though the 1:1 affinity is too weak to detect.

Finally, we validated the deep mutational scanning measurements in the context of spike-pseudotyped lentiviral particles (Figure 2.4E) [67]. The trends observed for entry by the spike-pseudotyped lentiviral particles generally confirmed the deep mutational scanning: three of four mutations that were detrimental for RBD expression or ACE2 binding reduced pseudovirus entry, while a mutation that had little phenotypic effect in the deep mutational scan did not affect viral entry. We also tested two ACE2 affinity-enhancing mutations and found that both increased pseudovirus entry. Note that this result with single-cycle pseudovirus does not necessarily imply that these mutations would increase growth of authentic

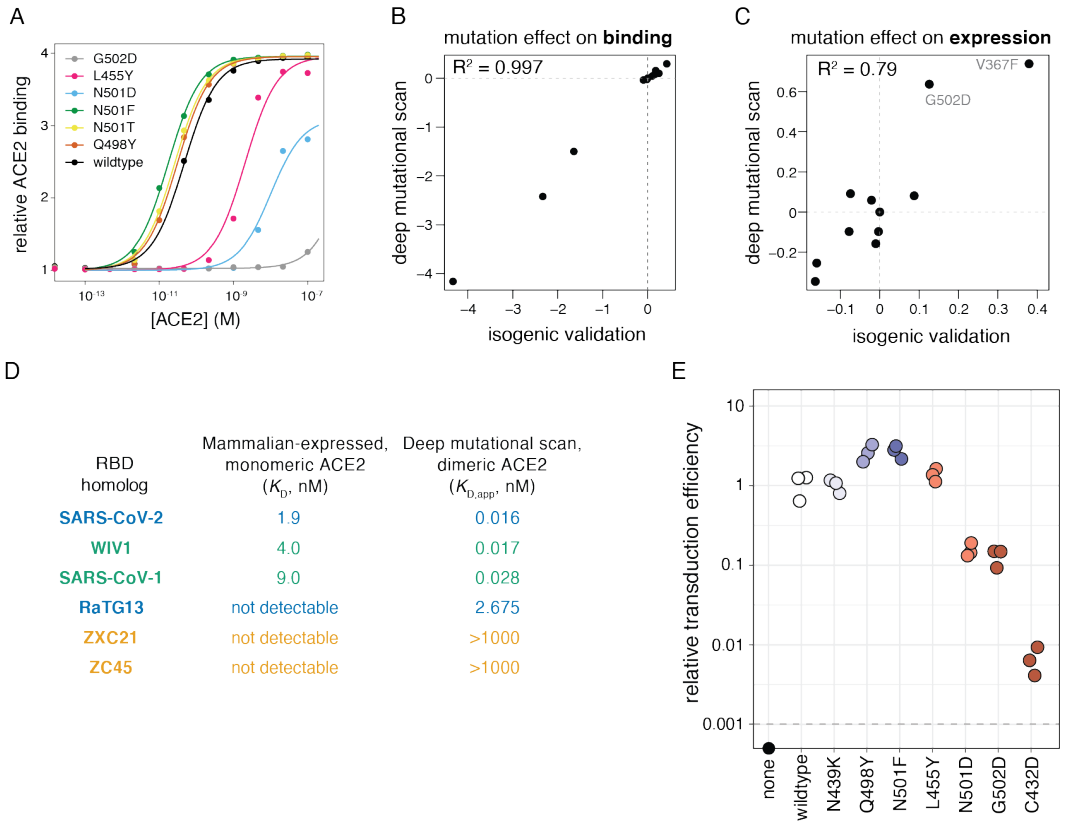


Figure 2.4: **Validation of deep mutational scanning measurements**. **(A)** Titration curves for select mutations that were re-cloned and validated in isogenic yeast cultures, as in Figure 2.1C. **(B, C)** Correlation in binding (B) and expression (C) effects between deep mutational scanning and isogenic yeast validations, including mutants shown in (A) and Figure 2.6C. **(D)** Comparisons of dissociation constants measured for mammalian-expressed purified RBD binding to monomeric human ACE2 and yeast displayed RBD binding to natively dimeric ACE2 from our deep mutational scan. **(E-G)** Validation of expression-enhancing mutations. **(E)** Effects of mutations on transduction of ACE2-expressing cells by lentiviral particles pseudotyped with SARS-CoV-2 spike. Mutants are colored by their effects on ACE2 binding as measured in the deep mutational scan (Figure 2.3B). Titers that fell below the limit of detection (dashed horizontal line) are plotted on the x-axis. Measurements were made in biological triplicate, and reflect the integrated effects of mutations on pseudovirus production and cellular entry.

SARS-CoV-2, since multi-cycle viral replication often involves tuning of receptor affinity to simultaneously optimize viral attachment and release [135, 37, 173]. Taken together, these experiments help validate the accuracy and relevance of the deep mutational scanning.

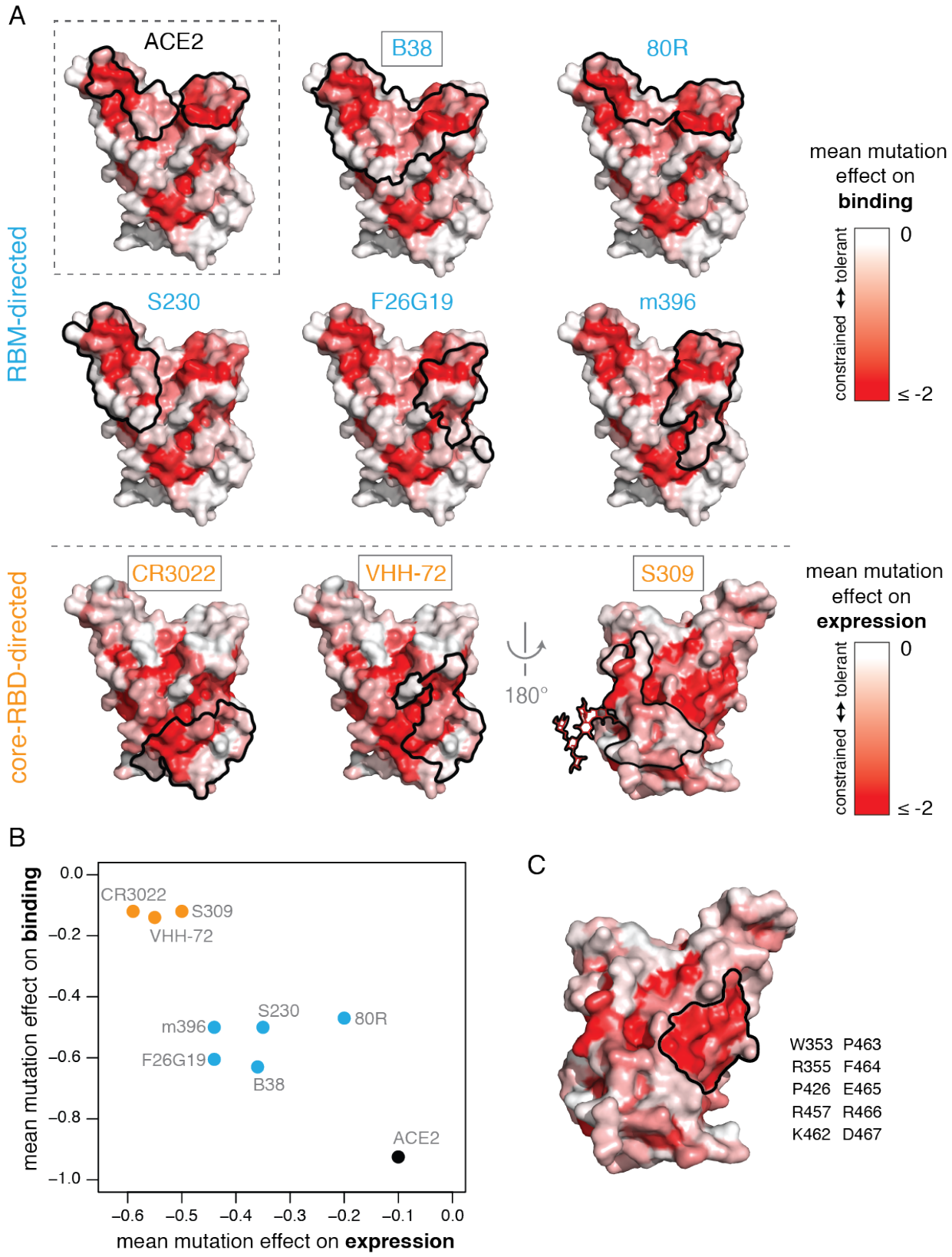
2.3.4 Mutational constraint of antibody epitopes

The RBD is the dominant target of neutralizing antibodies to SARS-CoV-2 [33, 38, 149, 230, 255, 356]. It is unclear to what extent the RBD will evolve to escape such antibodies in a manner reminiscent of some other viruses [274, 296], although *in vitro* studies suggest that SARS-CoV-2 and SARS-CoV-1 RBDs are capable of fixing mutations that escape neutralizing antibodies [19, 252]. To better define the RBD's evolutionary capacity for antibody escape, we examined mutational constraint in the epitopes of antibodies that bind the SARS-CoV-1 or SARS-CoV-2 RBD (Figures 2.5A) [144, 224, 230, 236, 309, 331, 346, 336].

Many antibodies have epitopes that overlap the RBD ACE2 contact interface, and are therefore strongly constrained by mutation effects on binding. For instance, antibodies B38 and 80R engage the two constrained patches that comprise the ACE2-binding interface, while S230, F26G19, and m396 engage either one of these ACE2-binding sub-regions. However, none of the currently characterized antibodies have epitopes as constrained as the ACE2-contact surface itself (Figure 2.5B), suggesting further epitope focusing could be achieved. The importance of such focusing is demonstrated by a recent study that identified RBD mutations enabling escape from RBM-directed neutralizing antibodies [19]—our data indicate that the escape occurs at sites that have high mutational tolerance.

Epitopes of core-RBD-directed antibodies tend to be mutationally constrained with respect to expression rather than binding (Figures 2.5A,B). These core-RBD epitopes are conserved across the sarbecovirus alignment, explaining the possible cross-reactivity of these antibodies between SARS-CoV-1 and SARS-CoV-2 [143, 331, 230]. Although residues in these epitopes are constrained for stability even in our measurements on the isolated RBD,

Figure 2.5: **Mutational constraint of antibody epitopes**. **(A)** For each of 8 RBD-directed antibodies, black outlines indicate the epitope structural footprint, with surfaces colored by mutational constraint (red indicates more constrained). The direct ACE2 interface is shown in the upper-left, for reference. Names of antibodies capable of neutralizing SARS-CoV-2 are boxed. (Others neutralize SARS-CoV-1 but have not been demonstrated to neutralize SARS-CoV-2.) Constraint is illustrated as mutational effects on binding for RBM-directed antibodies (blue, top), and expression for core-RBD-directed antibodies (orange, bottom). The N343 glycan, which is present in the S309 epitope and is constrained with respect to expression, is shown only on this surface for clarity. **(B)** Average mutational constraint for binding and expression within each epitope. Points are colored according to the RBM versus core-RBD designation in (A). **(C)** Identification of a patch of mutational constraint surrounding RBD residue E465 which has not yet been targeted by any described antibodies. Surface is colored according to mutational effects on expression, as in (A, bottom). Residues in this constrained E465 patch are listed.


A
RBM-directed
ACE2
B38
80R
S230
F26G19
m396
mean mutation effect on binding
constrained ↔ tolerant
0
≤ -2
core-RBD-directed
CR3022
VHH-72
S309
180°
mean mutation effect on expression
constrained ↔ tolerant
0
≤ -2
B
mean mutation effect on binding
mean mutation effect on expression
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
−0.6
−0.5
−0.4
−0.3
−0.2
−0.1
0.0
CR3022
S309
VHH-72
S230
m396
80R
F26G19
B38
ACE2
C
W353 P463
R355 F464
P426 E465
R457 R466
K462 D467

some of them likely exhibit additional constraint due to quaternary contacts in the full spike trimer [306, 331, 347]. We identified an additional core-RBD patch centered on residue E465 that is also mutationally constrained (Figure 2.5C) and evolutionarily conserved, but is not targeted by any currently known antibody and might represent a promising target.

Taken together, our results identify multiple mutationally constrained patches on the RBD surface that can be targeted by antibodies. These findings provide a framework that could inform the formulation of antibody cocktails aiming to limit the emergence of viral escape mutants [19, 230, 335, 356], particularly if deep mutational scanning approaches like our own are extended to define RBD antibody epitopes in functional as well as structural terms [81].

Using sequence-phenotype maps to interpret genetic variation in SARS-CoV-2

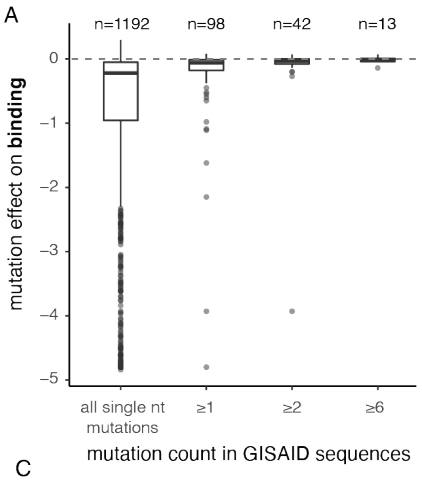
An important question is whether any mutations that have appeared in circulating SARS-CoV-2 isolates have functional consequences. Despite intense interest in this question, experimental work to characterize the effects of SARS-CoV-2 mutations has lagged far behind their identification in viral sequences. Our comprehensive maps of the phenotypic effects of mutations provide a direct way to interpret the impact of current and future genetic variation in the SARS-CoV-2 RBD.

To assess the phenotypic impacts of mutations that have appeared in the SARS-CoV-2 RBD to date, we downloaded all 31,570 spike sequences available from GISAID [91] on May 27, 2020, and identified RBD amino-acid mutations present in high-quality clinical isolates. All observed RBD mutations are at low frequency, with 56 of the 98 observed mutations present only in a single GISAID sequence. The observed mutations are significantly less deleterious for both ACE2 binding and RBD expression than random single-nucleotide-accessible mutations (Figures 2.6AP-value $< 10^{-}6$ for binding and expression, permutation tests), consistent with the action of purifying selection. Purifying selection against dele-

terious mutations is especially apparent for mutations that are observed multiple times in circulating variants, with a substantial number of singletons being mildly or moderately deleterious whereas mutations observed multiple times are largely neutral. This general pattern of increased purifying selection on more common mutations is consistent with theoretical expectation and empirical patterns observed for other viruses [242, 339].

Our discovery of multiple strong affinity-enhancing mutations to the SARS-CoV-2 RBD raises the question of whether positive selection favors such mutations, since the relationship between receptor affinity and fitness can be complex for viruses that are well-adapted to their hosts [37, 135, 173]. Affinity-enhancing mutations are accessible via single-nucleotide mutation from SARS-CoV-2, but none are observed among circulating viral sequences in GISAID (Figure 2.6A), and the observed mutations do not enhance ACE2 affinity more than randomly drawn samples of single nucleotide mutations. Taken together, we see no clear evidence of selection for stronger ACE2 binding, consistent with SARS-CoV-2 already possessing adequate ACE2 affinity at the beginning of the pandemic.

Last, we validated our deep mutational scanning for mutations that are especially prevalent among naturally occurring sequences in GISAID. The deep mutational scanning suggests small phenotypic effects for the most prevalent mutations, with the exception of V367F, which substantially enhances expression (Figure 2.6B). We re-cloned and tested most of these prevalent mutations for expression and ACE2 binding in isogenic yeast display assays. Consistent with the deep mutational scanning, the only large phenotypic effect was increased expression of V367F (Figure 2.6C,D). The relevance of V367F's stability-enhancing effect for viral fitness is unclear, though this mutation has independently arisen multiple times [300]. We also validated that N439K, the most prevalent RBD mutation which may have a very slight affinity-enhancing effect (Figures 2.6B,C), has no measurable impact on entry of spike-pseudotyped lentiviral particles (Figure 2.4). Taken together, our results suggest that there is little phenotypic diversity in ACE2 binding among circulating variants at this early



B

mutation	effect on **binding** $\Delta\log_{10}(K_{D,app})$	effect on **expression** $\Delta\log(MFI)$	number GISAID sequences	number geographic regions
V367F	0.07	0.74	24	8
N439K	0.04	-0.35	131	2
S477N	0.06	0.06	31	2
T478I	-0.04	-0.16	49	1
V483A	-0.03	0.09	30	2

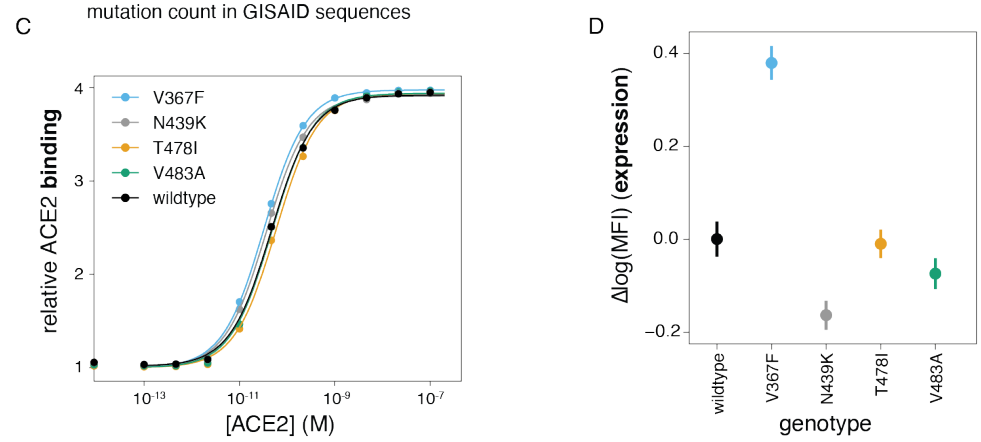


Figure 2.6: **Phenotypic impacts of genetic variation in the SARS-CoV-2 RBD**. **(A)** Distribution of effects on ACE2 binding of mutations observed among circulating SARS-CoV-2 isolates. The distribution of mutation effects is shown for all amino-acid mutations accessible via single-nucleotide mutation from the SARS-CoV-2 Wuhan-Hu-1 gene sequence, compared to the distributions for subsets of mutations that are observed in sequenced SARS-CoV-2 isolates deposited in GISAID at increasing observation count thresholds. n, number of mutations in each subset. **(B)** Summary of most frequent mutations among GISAID sequences, reporting our deep mutational scanning measured effect on binding and expression, the number of GISAID sequences containing the mutation, and the number of geographic regions from which a mutation has been reported. **(C, D)** Validation of the mutational effects on binding (C) and expression (D) for 4 of the 5 most frequent circulating RBD variants. S477N rose to high frequency after we began our validation experiments, and so was not included. Error bars in (D) are standard error from 11 samples.

stage of the pandemic—although it will be interesting to use our maps to continually assess the phenotypic effects of future mutations as the virus evolves.

2.4 Discussion

Vast numbers of viral genomes have been sequenced in almost real-time during the SARS-CoV-2 pandemic. These genomic sequences have been useful for understanding viral emergence and spread [96, 22], but the lack of corresponding high-throughput functional characterization means that speculation has outpaced experimental data when it comes to understanding the phenotypic consequences of mutations. Here, we take a step toward providing phenotypic maps commensurate with the scale of genomic data by experimentally characterizing how all amino-acid mutations to the RBD affect the expression of folded protein and its affinity for ACE2, two key factors for viral fitness. These maps show that RBD mutations that have appeared in SARS-CoV-2 to date are nearly neutral with respect to these two biochemical phenotypes, with the exception of one mutation (V367F) that increases RBD stability. Notably, there has been no selection to date for any of the evolutionarily accessible mutations that enhance ACE2 binding affinity. The genetic diversity of SARS-CoV-2 is likely to increase as it continues to circulate in the human population, and so our phenotypic maps should become increasingly valuable for viral surveillance as mutations accumulate over time.

It is important to remember that our maps define biochemical phenotypes of the RBD, not how these phenotypes relate to viral fitness. There are many complexities in the relationship between biochemical phenotypes of yeast-displayed RBD and viral fitness. First, there are subtle differences in glycan structures between yeast versus human cells [127], though the overall role of glycans in RBD stability is preserved in yeast systems [53]. Second, the RBD is just one domain of the viral spike, which engages in complex dynamic movements to mediate viral entry [143, 309, 306, 332]. Finally, spike-mediated entry is just one component of fitness,

which involves a myriad of incompletely understood factors that determine how well a virus spreads from one human to another [169]. To some degree, these caveats are universal of experimental studies, as even sophisticated animal models are imperfect proxies for true fitness [191]–but they are especially true for basic biochemical phenotypes like the ones we measure. However, on a hopeful note, our measurements correlate well with cellular entry by spike-pseudotyped viral particles expressing sarbecovirus RBD homologs (Figures 2.1D) and single mutants of the SARS-CoV-2 RBD (Figure 2.4). Furthermore, fitness ultimately arises from the concerted action of biochemical phenotypes, which are in turn determined by genotype [256, 73, 130]. By making the first link from mutations to biochemical phenotypes, we have taken a step towards enabling better interpretation of viral genetic variation.

One important area where our maps do have clear relevance is assessing the potential for SARS-CoV-2 to undergo antigenic drift by fixing mutations at sites targeted by antibodies, as occurs for some other viruses such as influenza [274]. The RBD is the dominant target of neutralizing antibodies [38, 149, 231, 255, 267, 271, 356], and so any antigenic drift will be constrained by its mutational tolerance. Our results show that many mutations to the RBD are well-tolerated with respect to both protein folding and ACE2 binding. However, the ACE2 binding interface is more constrained than most of the RBD's surface, which could limit viral escape from antibodies that target this interface [252]. In this respect, our maps enable several important observations. First, no characterized antibodies have epitopes as constrained as the actual RBD surface that contacts ACE2, suggesting that there is room for epitope focusing to minimize viral escape. Second, there are a number of RBD mutations that enhance ACE2 affinity, which implies ample evolutionary potential for compensation of deleterious mutations in the ACE2 interface in a manner reminiscent of multi-step escape pathways that have been described for other viruses [27]. It should be possible to shed further experimental light on the potential for antigenic drift by extending our deep mutational scanning methodology to directly map immune-escape mutations as has

been done for other viruses [81, 176].

RBD-based antigens represent a promising vaccine approach. Our sequence-phenotype maps can directly inform efforts to engineer such vaccines in several ways. First, we identify many mutations that enhance RBD expression and thermal stability, a desirable property in vaccine immunogens. Second, our maps show which mutations can be introduced into the RBD without disrupting key biochemical phenotypes, thereby opening the door to resurfacing immunogens to focus antibodies on specific epitopes [319, 88]. Finally, our maps show which surfaces of the RBD are under strong constraint and might thereby be targeted by structure-guided vaccines to stimulate immunity with breadth across the sarbecovirus clade: in addition to the ACE2 interface itself, these surfaces include several core-RBD surface patches targeted by currently described antibodies and a previously undescribed core-RBD surface patch surrounding residue E465.

Finally, our work should be useful for understanding the evolution of sarbecoviruses more broadly, including the potential for more spillovers into the human population. There is a dizzying diversity of RBD genotypes and phenotypes among sarbecoviruses within bat reservoirs [31, 76, 103, 141, 179, 196]. A prerequisite for these viruses to jump to humans is the ability to efficiently bind human receptors [21, 205, 206, 179]. Our maps are immediately useful in assessing the effects on ACE2-binding of mutations to viruses within the SARS-CoV-2 clade, and extensions to account for epistasis and genetic background could further inform understanding of the evolutionary trajectories that enable sarbecoviruses to efficiently infect human cells.

2.5 Materials and Methods

Data and Code Availability

The complete computational workflow to generate and analyze these data, including reproducible code within a programmatically constructed computational environment are available

at: `https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS`. All raw sequencing data are uploaded to the NCBI Short Read Archive (BioProject PRJNA639956).

Experimental Model and Subject Details

Saccharomyces cerevisiae strain AWY101 (Wentz and Shusta, 2007) was cultured at 30°C (except where indicated) in baffled flasks while shaking at 275rpm. Selective media contained 6.7 g/L Yeast Nitrogen Base, 5.0 g/L Casamino acids, 1.065 g/L MES, and 2% w/v carbon source (dextrose for routine maintenance, galactose supplemented with 0.1% dextrose for RBD induction). HEK-293T cells (ATCC CRL-3216) were cultured in D10 growth media (DMEM with 10% heat-inactivated FBS, 2 mM l-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin) at 37°C in a humidified 5% CO2 incubator. Expi293F (Thermo Fisher Cat No. A14527) suspension cells were grown at at 37°C in a humidified 8% CO2 incubator rotating at 130 rpm. Cell lines were not authenticated.

RBD cloning

The Spike receptor binding domain (RBD) from SARS-CoV-2 (isolate Wuhan-Hu-1, Genbank accession number MN908947, residues N331-T531) and additional sarbecovirus homologs (RaTG13, Genbank MN996532; GD-Pangolin consensus from [171]; SARS-CoV-1 Urbani, Genbank AY278741; WIV1, Genbank KF367457 (identical RBD sequence to WIV16); LYRa11, Genbank KF569996; Rp3, Genbank DQ071615; HKU3-1, Genbank DQ022305; Rf1, Genbank DQ412042; ZXC21, Genbank MG772934; ZC45, Genbank MG772933; and BM48-31, Genbank NC014470) were ordered as yeast codon-optimized gBlocks (IDT) and cloned into the pETcon yeast surface-display expression vector. The destination vector was modified downstream from the yeast surface-display fusion construct to include a barcode landing pad for subsequent library generation, along with Illumina sequencing priming handles for downstream barcode sequencing and NotI digestion sites for downstream

PacBio sequencing preparation. This plasmid sequence is provided on GitHub at `https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/tree/master/data/plasmid_maps/2649_pETcon-SARS-CoV-2-RBD-201aa.gb`.

Isogenic yeast display induction and titration

RBD variant plasmids were transformed into the AWY101 Saccharomyces cerevisiae strain [322], selecting for the plasmid Trp auxotrophic marker on SD-CAA selective plates (6.7g/L Yeast Nitrogen Base, 5.0g/L Casamino acids, 1.065g/L MES acid, and 2% w/v dextrose). Single colonies were inoculated into 1.5mL liquid SD-CAA media, and grown overnight at 30°C. Then 1 OD unit of yeast were back-diluted into 1.5mL SG-CAA+0.1%D induction media (2% w/v galactose supplemented with 0.1% dextrose), and incubated for 16-18 hours at room temperature.

Induced cells were spun down at 250,000 cells per sample and washed in PBS-BSA (0.2 mg/mL). Samples were resuspended in primary labeling solutions across a range of concentrations of biotinylated human ACE2 ectodomain (ACROBiosystems AC2-H82E6), which contains its natural dimerization domain. Primary labeling reactions were conducted in sufficient reaction volumes for each concentration to avoid ligand depletion effects of greater than 10%. For instance, the lowest sample concentration of 10^{-13} M was scaled to 25mL, at which volume 2.9% of total ligand molecules are estimated to be titrated in RBD:ACE2 complexes given the wildtype $K_{\mathrm{D,app}}$ and an estimated 50,000 surface RBDs per cell [28]. Following overnight equilibration of ACE2 binding at room temperature, cells were washed in ice-cold PBS-BSA, and resuspended in PBS-BSA containing 1:200 diluted FITC-conjugated anti c-Myc antibody (Immunology Consultants Lab, CMYC-45F) to label for RBD surface expression via a C-terminal c-Myc epitope tag, and 1:200 diluted PE-conjugated streptavidin (Thermo Fisher S866) to detect bound biotinylated ACE2 ligand. Following 1 hour of secondary labeling at 4 degrees C, cells were washed twice in ice-cold PBS-BSA, and

resuspended in PBS.

RBD surface expression and ACE2-binding levels were determined via flow cytometry using a BD LSRFortessa X-50. For flow cytometry, 10,000 cells were analyzed at each ACE2 concentration across a titration series. Cells were gated to select for singleton events, FITC labeling was used to subset RBD+ cells, and PE labeling was measured within this FITC+ population. To mimic the subsequent library sorting experiments in which we are blinded to exact PE fluorescence within a given PE fluorescence bin (since we only sequence barcodes within a bin), we analyzed isogenic titration data by drawing equivalent bins of PE fluorescence that capture 95% of unbound unmutated SARS-CoV-2 cells (bin1), 95% of saturated SARS-CoV-2 cells (bin4), and a bin2/bin3 boundary evenly spaced on the log-scale between the boundaries of the bin1 and bin4 partitions (see Figure 2.2B). For each ACE2 concentration, we determine the mean bin of PE fluorescence as a simple weighted mean value across integer-weighted bins:

$$\overline{\text{bin}_{[\text{ACE2}]}} = \frac{\sum_{i=1}^{4} n_{i,[\text{ACE2}]} \times i}{\sum_{i=1}^{4} n_{i,[\text{ACE2}]}}$$

where $n_{i,[\text{ACE2}]}$ is the number of cells that fall into bin i at a given ACE2 concentration, and i is the simple integer value of a bin from 1 to 4.

We determined the binding constant $K_{\text{D,app}}$ describing the affinity of each RBD variant for human ACE2 ligand along with free parameters a (titration response range) and b (titration curve baseline) via nonlinear least squares regression using a standard non-cooperative Hill equation relating the mean bin response variable to the ACE2 labeling concentration:

$$\overline{\text{bin}_{[\text{ACE2}]}} = a \times \frac{[\text{ACE2}]}{[\text{ACE2}] + K_{\text{D,app}}} + b$$

We report apparent K_{D} values ($K_{\text{D,app}}$) that do not take into account the stoichiometry of the multivalent yeast-displayed RBD interaction with dimeric ACE2. Following this

"apparent" nomenclature, we report ACE2 concentrations as molarity of the monomeric subunit. Computational notebooks detailing the fits of all isogenic RBD titrations is provided on GitHub (https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/data/isogenic_titrations/homolog_validations.md) and https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/data/isogenic_titrations/point-mut-validations.md.

Library mutagenesis

Mutagenesis of the SARS-CoV-2 RBD was performed in two independent replicates via the method described in [26] and we used NNS rather than NNN primers. Briefly, we designed mutagenic primers containing degenerate NNS codons that tile across the SARS-CoV-2 RBD, which were ordered as oPools from Integrated DNA Technologies. The script used to design the mutagenic primers and the resulting primers are available at https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/tree/master/data/primers/mutational_lib. We conducted three rounds of mutagenesis, each consisting of 7 mutagenic PCR cycles and 20 joining PCR cycles. The final joined products were amplified for 10 cycles with primers that append a unique identifier N16 barcode sequence to the 3' end of each mutagenized insert, downstream from the RBD stop codon and mRNA 3' UTR. Barcodes were also PCR appended to the un-mutagenized RBD homologs via the same primer addition PCR. Primers used in library assembly are provided on GitHub https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/tree/master/data/primers.

Mutagenized SARS-CoV-2 libraries and pooled wildtype homolog RBDs were cloned into EcoRI-HF/SacI-HF digested pETcon 2649 vector (sequence linked above) using NEBuilder HiFi DNA Assembly (NEB E2621). Assembled products were Ampure purified and electroporated into electrocompetent NEB10-beta cells. Electroporated cells were plated on 15cm LB+ampicillin plates at an estimated bottleneck of 100,000 (SARS-CoV-2 mutant libraries)

or 1,000 (pooled RBD homologs) colony forming units to limit library size. After approximately 18 hours of outgrowth, colonies were scraped into liquid LB+ampicillin, and grown for 2.5 hours in liquid culture prior to plasmid purification.

Plasmid pools were transformed into the AWY101 strain of Saccharomyces cerevisiae via the protocol of Gietz and Schiestl [115]. SARS-CoV-2 mutant libraries were transformed at 50 μg scale and the pooled RBD homolog controls were transformed at 10 μg scale. Colony forming unit counts from plated serial dilutions indicate transformation yield of > 1 million cfus. Transformed yeast grew for 14 hours post-transformation in 100 mL selective SD-CAA media, and were subsequently back-diluted into 100 mL fresh SD-CAA at 1 OD600 for an additional 9 hours, to enable further resolution of multiple vector transformants. Transformed yeast libraries were flash frozen in 1e8 cfu aliquots and stored -80 degrees C.

PacBio library sequencing and analysis

PacBio sequencing was used to acquire long sequence reads spanning the N16 barcode and the RBD gene sequence. PacBio sequencing inserts were prepared from bacterially-purified plasmid pools via NotI-HF restriction digest followed by gel purification and SMRTbell ligation. The use of restriction digest rather than PCR eliminates the possibility of PCR strand exchange scrambling barcodes. Each SARS-CoV-2 RBD mutant library was spiked to 1% frequency with the internal standard pool of RBD homologs. Each replicate library was sequenced in two SMRT Cells on a PacBio Sequel using 20-hour movie collection times. PacBio circular consensus sequences (CCSs) were generated from the raw subreads using the ccs program (`https://github.com/PacificBiosciences/ccs`, version 4.2.0), setting the parameters to require 99.9% accuracy and a minimum of 3 passes. The resulting CCSs are available on the NCBI Sequence Read Archive at `https://www.ncbi.nlm.nih.gov/bioproject/PRJNA639956`.

We then processed the CCSs to identify the RBD sequence (SARS-CoV-2 or one of the

11 homologs), call any mutations in the RBD sequence, and determine the associated 16-nucleotide barcode. To do this, we used alignparse [65], version 0.1.3. We only retained CCSs that matched the parental RBD sequence with no more than 45 nucleotide mutations (corresponding to up to 15 codon mutations), had a barcode of the expected 16 nucleotide length, and had no more than one mismatch in the flanking regions expected in the sequenced amplicon. A computational notebook providing full details is available on GitHub at `https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/results/summary/process_ccs.md`.

We next used these processed CCSs to generate a codon-variant lookup table that links each barcode to its associated codon mutations in the RBD sequence. To do this, we first filtered only for CCSs where the PacBio ccs-reported accuracy was at least 99.99% in both the RBD gene sequence and the barcode (the vast majority of CCSs passed this filter). We then determined the empirical accuracy of the CCSs by determining the concordance between the RBD gene sequence called by CCSs with the same barcode using the method implemented at `https://jbloomlab.github.io/alignparse/alignparse.consensus.html#alignparse.consensus.empirical_accuracy`. For both libraries, the empirical accuracy of the entire region of the CCS covering the RBD sequence was 99.8% if we ignored those with indels. Most barcodes were covered by multiple CCSs, and in that case we built a consensus of these CCSs after discarding any barcodes for which the CCSs differed often or at many sites using the method implemented at `https://jbloomlab.github.io/alignparse/alignparse.consensus.html#alignparse.consensus.simple_mutconsensus`. Finally, we discarded any variants with indels in the RBD. Therefore, more than 99.8% of the final barcode-linked variants should have the correctly determined RBD sequence, since 99.8% is the accuracy for those covered by just one CCS and most variants were called by the consensus of multiple CCSs. For further analysis of the barcoded variants, we then created a codon variant table using dms_variants (`https://jbloomlab.github.io/dms_variant`

s/,version0.6.0). The final barcode-variant lookup table (which associates each barcode with its RBD sequence) is at https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/raw/master/results/variants/codon_variant_table.csv. Some summary statistics about the final composition of the libraries are in Figure S1, and the complete code used to generate the barcode-variant lookup table and many additional plots characterizing the composition of the libraries are on GitHub at https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/results/summary/build_variants.md.

Deep mutational scanning library yeast surface-display induction and labeling

Yeast libraries were thawed and grown overnight at 30 degrees C in 180mL SD-CAA media at an initial OD600 of 0.1. We spiked our SARS-CoV-2 mutant libraries with the barcoded RBD homolog pool at a total fraction of 0.6% yeast density, such that each RBD homolog barcode should be present at a frequency on the same order of magnitude as the typical SARS-CoV-2 variant barcode. To induce RBD surface expression, yeast were back-diluted to 50 mL (expression experiments) or 200 mL (binding experiments) SG-CAA+0.1%D induction media at 0.67 OD600 and incubated at room temperature for 16-18 hours with mild agitation.

For library expression experiments, 45 OD units yeast were washed twice with PBS-BSA and labeled in 3 mL 1:100 diluted anti-Myc-FITC antibody for 1hr at 4 degrees C with gentle mixing. Labeled cells were washed twice in PBS-BSA and resuspended in 5mL PBS for FACS. For library binding experiments, 8 OD units yeast per titration concentration (10^{-13} M to 10^{-6} M ACE2 at half-log intervals, plus a 0M ACE2 sample) were washed twice with PBS-BSA, and incubated with ACE2 ligand overnight at room temperature with gentle agitation. Labeling volumes were scaled at low ACE2 concentration to limit ligand depletion effects, as with isogenic titrations described above. Following equilibration of ACE2 labeling, cells were kept chilled while washing once with PBS-BSA, labeling for one hour in 1 mL PBS-BSA with 1:100 diluted Myc-FITC and 1:200 Streptavidin-PE, washed two more times with

PBS-BSA, and resuspended to 1 mL in PBS.

Fluorescence activated cell sorting (FACS) of yeast libraries

Yeast libraries were sorted into bins of FITC or PE fluorescence using a BD FACS Aria II. Cells were sorted into 5mL FACS tubes containing 1mL of 2xYPAD supplemented with 1% BSA. Tubes were pre-wet with collection media prior to sample collection, to reduce sticking and improve post-sort yield.

For expression sorts, cells were gated for singleton events (Figure S2A), followed by partitioning into four bins of FITC fluorescence (Figures 2A): bin 1 captures 99% of unstained cells, and bins 2-4 split the remaining library fraction into tertiles. We sorted > 50 million cells from each library into these bins. From these same inductions, we also sorted 15 million RBD+ cells from each library, to enrich RBD-expressing cells within our libraries for our titration sorting experiments.

For ACE2-binding titrations, we gated cells for singleton events and RBD+ expression. For each ACE2 concentration sample, we sorted cells into four bins of PE fluorescence as described above: bin1 captures 95% of unmutated SARS-CoV-2 cells incubated with 0M ACE2, bin4 captures 95% of unmutated cells at saturating ACE2 ligand, and the bin2/bin3 boundary evenly splits the log-MFI scale between the bin1 and bin4 boundaries (Figure 2.2B). We sorted each ACE2 concentration sample into these four bins for approximately 15 minutes, capturing 5-6 million cells per ACE2 concentration.

Following each sort, cells from each collection tube were spun for 5 min at 3,000 g in a tabletop centrifuge, yielding a visible pellet for any sample with at least 500,000 collected cells. Collection supernatant was removed, and cells were resuspended in SD-CAA media supplemented with 1:100 penicillin-streptomycin. Cells were resuspended to an estimated 2e6 cells/mL in 15mL culture tubes or baffled flasks for expresion post-sort samples, 5e5 cells/mL in baffled flasks for RBD+ sort samples, and 1mL (<1e6 cells) or 1.5mL (>1e6

cells) in 96-deep-well plates for titration samples. For expression FACS experiments, total cell recovery from all samples was measured via serial dilution and plating on YPD and SD-CAA plates for each sample, which showed average cellular recovery of 85% (range 79-94%), with 62% (range 52-77%) of cells retaining plasmid, with exception of the FITC-negative bin 1 populations, which showed 20% plasmid retention. These per-sample cell recovery counts were used to calibrate downstream sequencing numbers for the actual number of cells that grew out from each sort bin. For titration sorts, we did not titer all 64 post-sort samples, but instead spot checked 6 samples to ensure normal levels of cell recovery, which showed an average 66% cell recovery and 46% plasmid retention. As we did not titer all samples, we use the FACS log cell count as the estimate of number of cells collected in each bin, which makes the assumption that there are no systematic differences in post-sort cell yield across bins, which is more appropriate for these titration sorts where the ACE2 binding gates are nested within an overall RBD+ selection gate that selects for even plasmid retention.

Post-sort samples were grown overnight in liquid media at 30°C. Plasmids were purified from post-sort yeast samples of <4e7 cfu using Zymo Yeast Miniprep kits (single column or 96-well plate formats) according to kit instructions, but with the addition of $>$ 2 hours Zymolyase treatment and a -80 degrees C freeze/thaw cycle prior to cell lysis.

Illumina Sequencing

Post-sort plasmid samples were PCR amplified from 10 μL plasmid template input using primers flanking the N16 barcode that append remaining Illumina sequencing handles that are not already plasmid encoded, and unique NextFlex sample indices `https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/tree/master/data/primers`. PCRs were conducted with KOD polymerase for 20 cycles, except for titration sort samples of less than 10,000 cells, where 28 cycles were necessary to obtain sufficient PCR product due to low sample input:

1. 95°C, 2min
2. 95°C, 20s
3. 58°C, 10s
4. 70°C, 10s
5. Return to 2, 19× (27× for low-input samples)

PCR products were Ampure purified, quantified via PicoGreen, and pooled to mirror desired sample frequencies given cell counts in each FACS sample. Pooled samples were gel purified, Ampure purified, and submitted for 2 lanes of 50bp single end Illumina HiSeq sequencing per library.

Demultiplexed reads were aligned to library barcodes determined from PacBio sequencing, yielding a count of the number of times each library barcode was sequenced within each FACS partition. Read counts for each FACS sample were downweighted by the ratio of total reads from a bin compared to the number of cells that were actually sorted into that bin. For one bin in which the number of HiSeq reads was less than the number of cells sorted into a bin, we re-amplified PCR product from a newly purified plasmid aliquot, and obtained reads via a single lane of MiSeq 50bp single end sequencing. Computational notebooks providing additional details on our Illumina sequencing processing and statistics are provided on GitHub `https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/results/summary/count_variants.md`and`https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/results/summary/analyze_counts.md`.

Calculating variant phenotypes for expression

For each library variant, we estimated mean expression based on its distribution of cell counts across FITC sort bins and the known censored fluorescence boundaries of each sort

bin using a maximum likelihood approach [228], enacted in the fitdistrplus R package [75], assuming the uncensored log-transformed fluorescence values for a genotype follow a normal distribution. Expression measurements were retained for barcodes for which at least 20 cells were sampled across the four sort bins, resulting in measured expression phenotypes for 92.9 and 90.5% of variants in libraries 1 and 2, respectively.

Expression measurements were represented as the difference in log-mean fluorescence intensity (MFI) relative to wildtype ($\Delta\text{logMFI} = \text{logMFI}_{\text{variant}} - \text{logMFI}_{\text{wildtype}}$), such that a positive value indicates higher RBD expression. A computational notebook presenting our calculation of expression phenotypes and results is included on GitHub `https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/results/summary/compute_expression_meanF.md`.

Calculating variant phenotypes for ACE2-binding affinity

For each library barcode at each ACE2 sample concentration, we determined its simple mean bin of ACE2-binding via the equation used above in isogenic titrations. We fit titration curves as above to determine barcode-specific $K_{\text{D,app}}$ from the series of FACS-seq derived mean bin measurements across ACE2 concentration. Because a barcode's mean bin might be measured with varying certainty across different bins, we used weighted least squares nonlinear regression, weighing each mean bin estimate by an empirical variability estimate based on the per-sample cell count, derived from estimates of variability in repeated wild-type/synonymous barcode measurements grouped by sampling depth. To avoid fits of errant titration curves, we constrained the baseline parameter b to be fit between 1 and 1.5, and the response parameter a to be fit between 1.5 and 3. Through initial curve fit constraints and subsequent QC filtering, our fit KD,app binding constants were constrained to be within the concentration range of our titration (10^{-13} – $10^{-6}M$), and therefore many barcodes are censored at the upper limit with true $K_{\text{D,app}} \geq 10^{-6}$M . We filtered out titration curves fit

for variants with an average cell count < 5 across sample concentrations, or with cell count < 2 in 7 or more of the 16 ACE2 concentration samples. Finally, we filtered out the 5% of curves with the highest normalized mean square residual, where residuals are normalized from 0 to 1 by the fit response parameter a, such that titration curves that plateau at lower levels of saturated binding don't have systematically smaller mean square residuals. This process yielded $K_{\mathrm{D,app}}$ estimates for 75.2 and 75.4% of variants in libraries 1 and 2, respectively. Binding measurements were represented as the difference in $\log10(K_{\mathrm{D,app}})$ relative to wildtype ($\Delta \log_{10}(K_{\mathrm{D,app}}) = \log_{10}(K_{\mathrm{D,app}})^{\mathrm{wildtype}} - \log_{10}(K_{\mathrm{D,app}})^{\mathrm{variant}}$), polarized such that a positive value indicates higher variant ACE2 affinity. A computational notebook presenting our calculation of expression phenotypes and results is included on GitHub `https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/results/summary/compute_binding_Kd.md`.

Analysis of circulating variants

All 31,570 spike sequences on GISAID as of 27 May 2020 were downloaded and aligned via mafft (Katoh and Standley, 2013). Sequences from non-human origins and sequences containing any gap characters were removed. All amino-acid mutations among GISAID sequences were enumerated. Some low-coverage spike sequences contain undetermined 'X' characters. We excluded any mutation from our curated set of GISAID mutations if it was solely observed on sequence backgrounds containing at least one undetermined X character in the RBD sequence; however, sequences with X characters were allowed to contribute to observations of mutation count for mutations that were observed on at least one other high-coverage RBD sequence. To characterize patterns of selection on amino-acid mutations observed among GISAID sequences, we conducted permutation tests as described in the Figure S7 legend. Our full analysis of mutational effects of circulating variants is provided on GitHub `https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/res`

ults/summary/circulating_variants.md. We acknowledge all GISAID contributors for their sharing of sequencing data (https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/data/alignments/Spike_GISAID/gisaid_hcov-19_acknowledgement_table.xls).

Pseudotyped lentiviral particle infection assays

We selected seven single mutations from our deep mutational scanning measurements for validation of phenotypic effects in a spike-pseudotyped lentivirus assay [67]. Mutations were selected that exhibited deleterious effects on RBD expression (C432D) or ACE2 binding (L455Y, N501D and G502), no strong phenotypic effect on either binding or expression (N439K), and affinity-enhancing effects (Q498Y and N501F). These point mutations were introduced via site-directed mutagenesis (New England Biolabs E0554S) into the HDM vector containing codon-optimized SARS-CoV-2 Spike from Wuhan-Hu-1, with an upstream Kozak sequence. The full sequence of this plasmid is available at https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/data/plasmid_maps/2736_HDM_IDTSpike_EcoKozak.gb.

Pseudotyped lentiviral particles were generated as previously described [67]. Viruses were rescued in triplicate (i.e. independent transfections), which should average out variation in transfection efficiency such that viral entry phenotypes are reflective of both pseudovirus production and entry efficiency. Briefly, 2.5e5 293T cells per well were seeded in 12-well plates in 1 mL D10 growth media (DMEM with 10% heat-inactivated FBS, 2 mM l-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin). 24h later, cells were transfected using BioT transfection reagent (Bioland Scientific, Paramount, CA, USA) with 0.5 μg of the ZsGreen lentiviral backbone pHAGE2-CMV-ZsGreen-W (BEI Resources NR-52520), 0.11 μg each of the lentiviral helper plasmids HDM-Hgpm2 (BEI Resources NR-52517), pRC-CMV-Rev1b (BEI Resources NR-52519), and HDM-tat1b (BEI Resources NR-52518), and

0.17 μg wildtype or mutant SARS-CoV-2 Spike plasmids. Media was changed to fresh D10 at 24 h post-transfection. At 60 hours post transfection, the viral supernatant was collected, filtered through a 0.45 μm SFCA low protein-binding filter, and stored at -80°C. To quantify efficiency of pseudovirus production, we quantified p24 levels (in pg/mL) in viral transfection supernatants via ELISA, in technical duplicate, per kit instructions (Advanced Bioscience Laboratories Cat. 5421).

The resulting viruses were titered as previously described [67]. 293T cells stably expressing ACE2 (BEI NR-52511) were seeded at 1e4 cells per well in poly-L-lysine coated 96-well plates (Greiner 655930). 24 h later, 3 wells were counted and averaged to determine the number of cells per well at time of infection. Media was removed from the 293T-ACE2 cells and replaced with fresh D10 containing 50 μL of pseudovirus supernatant in a final volume of 150 μL. 60 h post-infection, cells were analyzed by flow cytometry. Titers are reported relative to the mean of the wildtype.

2.6 Notes

Acknowledgments

We thank Keara Malone for experimental assistance, Katherine Xue for helpful suggestions, and Frederick Matsen for intellectual support and hospitality. We thank the Flow Cytometry and Genomics core facilities at the Fred Hutchinson Cancer Research Center for experimental support, and Mike Murphy, Deleah Pettie, and the Mammalian Production Core at the Institute for Protein Design for assistance with protein purification. This work was supported by the NIAID / NIH (R01AI141707 and R01AI12893 to J.D.B., HHSN272201700059C to D.V., F30AI149928 to K.H.D.C., and T32AI083203 to A.J.G.), NIGMS / NIH (R01GM120553 to D.V.), a Pew Biomedical Scholars Award to D.V., Burroughs Wellcome Investigators in the Pathogenesis of Infectious Diseases awards to D.V. and J.D.B., the Bill & Melinda Gates Foundation (OPP1156262 to D.V. and N.P.K.), Fast Grants to N.P.K. and D.V., and a gen-

erous gift from the Open Philanthropy Project to N.P.K. T.N.S. is a Washington Research Foundation Innovation Fellow at the University of Washington Institute for Protein Design and a Howard Hughes Medical Institute Fellow of the Damon Runyon Cancer Research Foundation. J.D.B. is an Investigator of the Howard Hughes Medical Institute.

Declarations of Interests

N.P.K. is a co-founder, shareholder, and chair of the scientific advisory board of Icosavax, Inc.

Chapter 3

COMPLETE MAPPING OF MUTATIONS TO THE SARS-COV-2 SPIKE RECEPTOR-BINDING DOMAIN THAT ESCAPE ANTIBODY RECOGNITION

A version of this chapter has been previously published as:

Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, et al. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. Cell Host Microbe. 2021 Jan 13;29(1):44-57.e9. doi:10.1016/j.chom.2020.11.007.

3.1 Abstract

Antibodies targeting the SARS-CoV-2 spike receptor-binding domain (RBD) are being developed as therapeutics and make a major contribution to the neutralizing antibody response elicited by infection. Here, we describe a deep mutational scanning method to map how all amino-acid mutations in the RBD affect antibody binding, and apply this method to 10 human monoclonal antibodies. The escape mutations cluster on several surfaces of the RBD that broadly correspond to structurally defined antibody epitopes. However, even antibodies targeting the same RBD surface often have distinct escape mutations. The complete escape maps predict which mutations are selected during viral growth in the presence of single antibodies, and enable us to design escape-resistant antibody cocktails–including cocktails of antibodies that compete for binding to the same surface of the RBD but have different escape mutations. Therefore, complete escape-mutation maps enable rational design of antibody

therapeutics and assessment of the antigenic consequences of viral evolution.

3.2 Introduction

The COVID-19 pandemic has generated urgent interest in antibody therapeutics and vaccines that induce antibodies to SARS-CoV-2. Many of the most potently neutralizing anti-SARS-CoV-2 antibodies target the receptor-binding domain (RBD) of the viral spike protein, often competing with its binding to the ACE2 receptor [33, 38, 149, 188, 255, 267, 318, 356, 357, 336]. In addition, anti-RBD antibodies often dominate the neutralizing activity of the polyclonal antibody response elicited by natural infection [16, 282, 320]. Both passively-administered and vaccine-induced anti-RBD neutralizing antibodies protect against SARS-CoV-2 in animals [7, 38, 131, 255, 304, 336, 356], and preliminary evidence suggests the presence of neutralizing antibodies also correlates with protection in humans [6].

Determining which viral mutations escape from antibodies is crucial for designing therapeutics and vaccines and assessing the antigenic implications of viral evolution. Escape mutants can be selected by passaging virus expressing the SARS-CoV-2 spike protein in the presence of anti-RBD antibodies in the laboratory [19, 320], and some RBD mutations that alter antibody binding are already present at very low levels in SARS-CoV-2 circulating in the human population [182]. It seems plausible that such mutations could become prevalent over longer evolutionary time, given that the seasonal coronavirus 229E has accumulated genetic variation in its RBD in the last few decades that is sufficient to ablate antibody binding [327].

However, current methods to identify SARS-CoV-2 escape mutations by passaging virus in the presence of antibodies are incomplete in the sense that they only find one or a few of the possible escape mutations. Structural biology can more comprehensively define how an antibody physically contacts the virus, but structures are time consuming to determine and still do not directly report which viral mutations escape from antibody binding [70, 81, 148].

Here we overcome these limitations by developing a high-throughput approach to completely map mutations in the SARS-CoV-2 RBD that escape antibody binding, and apply this approach to 10 human antibodies. The resulting escape maps reveal the extent to which different antibodies are escaped by mutations at overlapping or orthogonal sites, and show that antibodies targeting structurally similar regions sometimes have escape mutations at entirely distinct residues. Furthermore, we show that the escape maps predict which mutations are selected when spike-expressing virus is passaged in the presence of neutralizing antibodies, and can inform the design of antibody cocktails that resist escape. Therefore, complete escape-mutation maps can be used to assess the antigenic consequences of viral genetic variation and the potential for viral escape from specific antibodies or antibody cocktails.

3.3 Results

3.3.1 A yeast-display system to completely map SARS-CoV-2 RBD antibody-escape mutations

To map antibody-escape mutations in a high-throughput manner, we leveraged a system for expressing conformationally-intact RBD on the surface of yeast cells (Figure 3.1A). As described previously [280], we created duplicate mutant libraries of the RBD from the Wuhan-Hu-1 strain of SARS-CoV-2 that together contained nearly all possible amino-acid mutations in the 201-residue RBD (they contain 3,804 of the 3,819 possible mutations, with ¿95% present as single mutants). Each yeast cell carries a short 16-nucleotide barcode that identifies the RBD mutant it expresses, enabling us to rapidly characterize the composition of the RBD mutant libraries via deep sequencing of the DNA barcodes.

Here, we developed a method to use these libraries to comprehensively identify mutations in the RBD that allow it to escape binding by antibodies. To eliminate RBD mutants that were completely misfolded or unable to bind ACE2, we first used fluorescence-activated cell

sorting (FACS) to eliminate RBD variants with $< 0.01\times$ the affinity for ACE2 compared to that of the unmutated RBD (Figure 3.7A,B). We reasoned this sorting would purge the libraries of completely nonfunctional RBD mutants, but retain mutants with decreased ACE2 affinity that might enable antibody escape [19, 252]. We then incubated the ACE2-sorted yeast libraries with an anti-RBD antibody (see next section) and sorted for cells that expressed RBD mutants that bound substantially less antibody than unmutated SARS-CoV-2 RBD (Figure 3.1C, Figure 3.7C). We deep-sequenced the nucleotide barcodes to quantify RBD variant frequencies in the initial ACE2+ population and the antibody-escape population (Figure 3.1C). We quantified the effect of each RBD mutation by estimating the fraction of cells expressing that mutation that fell into the antibody-escape sort bin, and termed this quantity the mutation's "escape fraction". We represented the escape fractions using logo plots (Figure 3.1C).

3.3.2 Mapping escape from each of 10 human monoclonal antibodies

We applied our escape-mutation mapping to 10 human monoclonal antibodies: 9 neutralizing antibodies isolated from SARS-CoV-2 convalescent patients [357], and a recombinant form of one cross-reactive non-neutralizing antibody isolated from a convalescent SARS-CoV-1 patient (CR3022) [347, 143, 286, 290]. All 10 antibodies bind the SARS-CoV-2 RBD with high affinity, but they differ in their neutralization potencies, extent to which they compete with ACE2 for RBD binding, and cross-reactivity with SARS-CoV-1 (Figure 3.2A) [347, 356].

We mapped escape mutations for each of the 10 antibodies in biological duplicate by applying the workflow in Figure 3.1C to each of our two independently generated RBD mutant libraries (Figures 3.7C, 3.8). We determined the effect of each mutation on antibody escape (the "escape fraction", Figure 3.1C) after applying quality-control filters to remove RBD mutants with low expression, ACE2 binding, or sequencing counts (see Methods for details). The resulting escape fraction measurements correlated strongly between the duplicate mu-

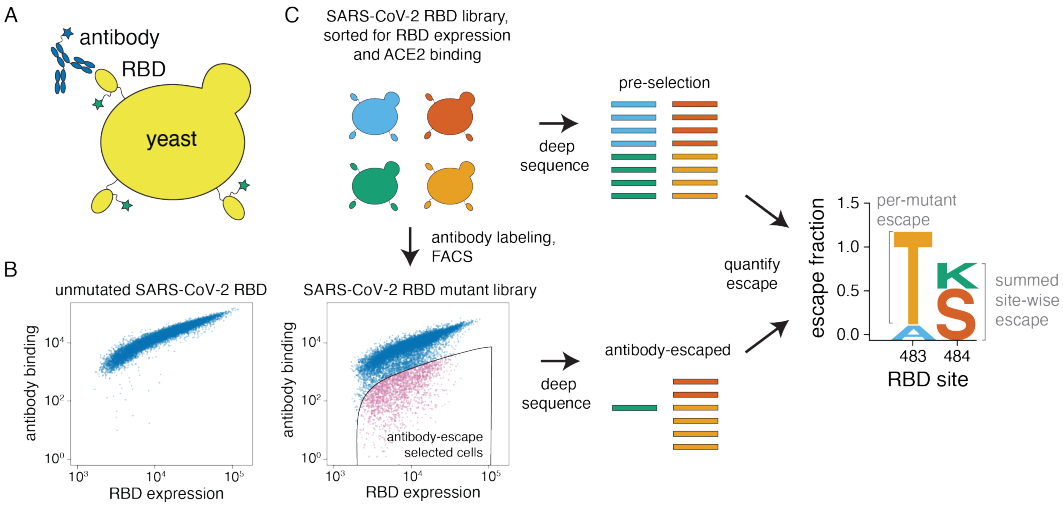


Figure 3.1: **A yeast-display system to completely map SARS-CoV-2 RBD antibody escape mutations.** **(A)** Yeast display RBD on their surface. The RBD contains a c-myc tag, enabling dual-fluorescent labeling to quantify both RBD expression and antibody binding of RBD by flow cytometry. **(B)** Per-cell RBD expression and antibody binding as measured by flow cytometry for yeast expressing unmutated RBD and one of the RBD mutant libraries. **(C)** Experimental workflow. Yeast expressing RBD mutant libraries are sorted to purge RBD mutations that abolish ACE2 binding or RBD folding. These mutant libraries are then labeled with antibody, and cells expressing RBD mutants with decreased antibody binding are enriched using FACS (the "antibody-escape" bin; see Figure 3.7 for gating details). Both the initial and antibody-escape populations are deep sequenced to identify mutations enriched in the antibody-escape population. The deep-sequencing counts are used to compute the "escape fraction" for each mutation, which represents the fraction of yeast cells with a given RBD mutation that falls into the antibody-escape sort bin. The escape fractions are represented in logo plots, with tall letters indicating mutations that strongly escape antibody binding.

tant libraries (Figure 3.8), and for the rest of this paper we report the average measurements across libraries. Note that the magnitude of the measured effects of mutations on antibody escape depends on the antibody concentration and the flow cytometry gates applied (Figure 3.7C), meaning that the escape fractions are comparable across sites for any given antibody, but are not precisely comparable among antibodies without external calibration.

The effects of mutations on antibody escape are summarized in Figure 3.8C. Each antibody is escaped by mutations at just a small subset of residues in the RBD. In general, CR3022 and the three antibodies that compete with CR3022 for RBD binding are escaped by mutations in the core RBD distal from the ACE2 receptor binding motif (RBM) (Figure 3.8A,B,C). The remaining antibodies are escaped primarily by mutations in the RBM of the RBD, including at ACE2 contact residues (Figure 3.8C). Notably, the escape mutations for the most potently neutralizing antibodies fall mostly in the RBM (Figure 3.8A,C), consistent with prior studies showing that potent anti-RBD neutralizing antibodies often strongly compete with ACE2 binding [33, 38, 142, 149, 188, 255, 267, 318, 336, 356].

However, the escape-mutation maps are far more nuanced than can be represented by simply grouping the RBD into broad antigenic regions. While a few antibodies have extremely similar escape mutations (e.g., COV2-2082 is similar to COV2-2094, and COV2-2479 is similar to COV2-2050), antibodies that target the same broad region of the RBD often have distinct escape mutations (e.g., COV2-2832 and COV2-2499 are escaped by entirely non-overlapping sets of escape mutations in the RBM). There is also heterogeneity in which specific amino acid mutations mediate escape. At some selected sites, many mutations confer escape (e.g., site 378 for COV2-2677 or site 490 for COV2-2096). But at other sites, only certain mutations confer escape: for instance, only negatively-charged amino acids at site 408 escape COV2-2082, and only mutations at site 372 that introduce a serine or threonine (creating an N-linked glycosylation motif at site 370) escape COV2-2677.

To better compare the escape maps across antibodies, we used multidimensional scaling

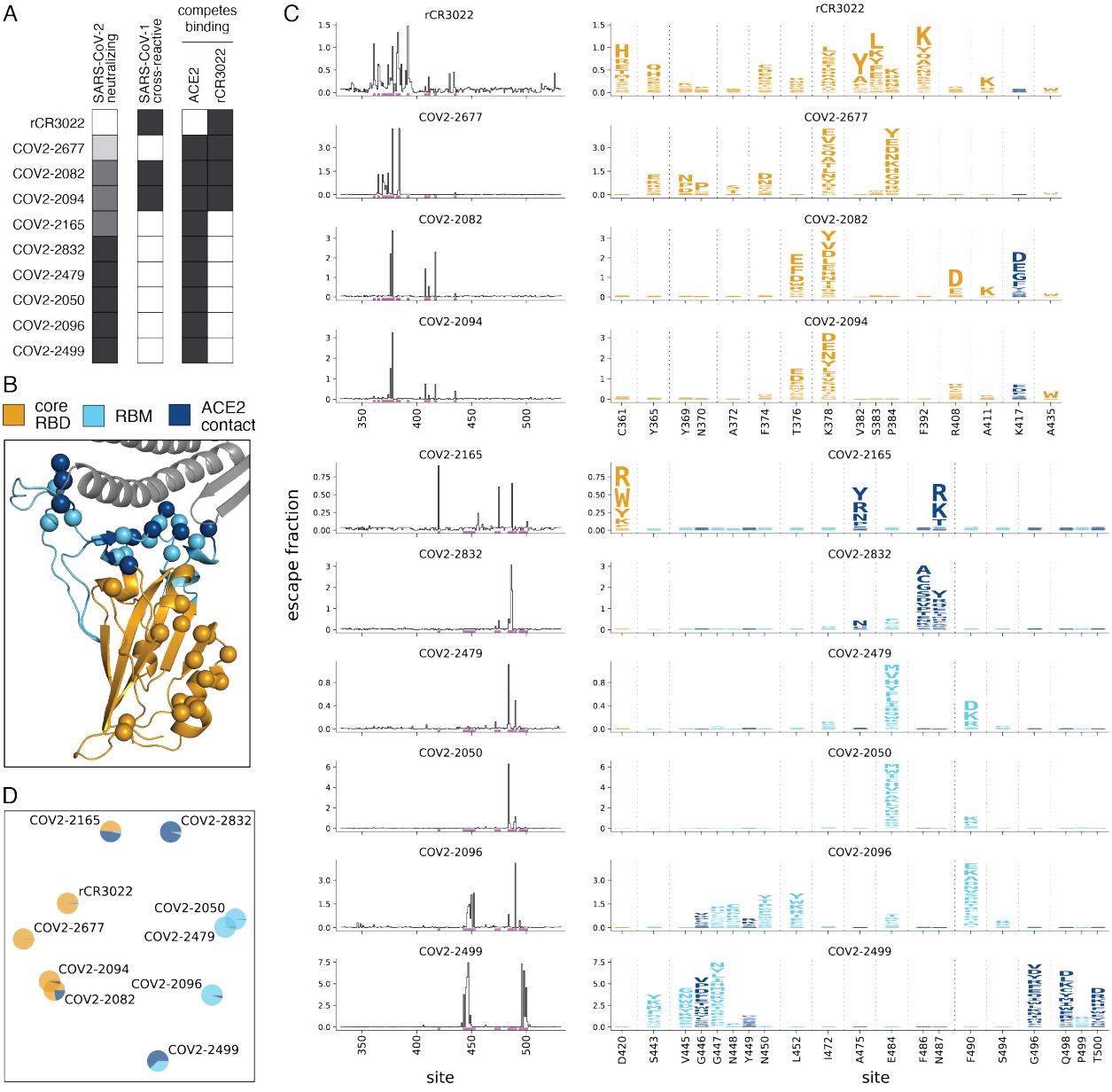
to project the similarity in escape mutations into a two-dimensional plot (Figure 3.8D). In this plot, the distance between antibodies increases as their escape mutations become more distinct. In addition, the pie chart colors in Figure 3.8D indicate the regions of the RBD where mutations confer escape. This plot makes clear that antibodies that target similar regions of the RBD sometimes but not always have similar escape mutations: for instance, COV2-2479, COV2-2050, and COV2-2096 all target the RBM–but only the first two of these antibodies cluster closely in Figure 3.8D. Overall, the two-dimensional projection in Figure 3.8D provides an intuitive way to visualize the relationships among antibodies in the space of immune-escape mutations, similar to how dimensionality reduction techniques such as t-SNE or UMAP help visualize high-dimensional single-cell transcriptomic data [10, 20].

To independently validate the escape maps, we tested key escape mutations from four antibodies in neutralization assays using spike-pseudotyped lentiviral particles [67]. The agreement between the escape maps and neutralization assays was excellent (Figure 3.3, Figure 3.9A) and validated the subtle differences between antibodies. For instance, as indicated by the maps, a mutation at site 487 escapes both COV2-2165 and COV2-2832, but a mutation at site 486 only escapes COV2-2832 (Figure 3.3). Because the antibody escape selections were performed at a single concentration, the magnitude of escape measured in the mapping cannot be directly converted to a quantitative change in binding affinity. However, all tested escape mutations identified by the escape maps increased IC50 by 10-fold or greater. We also validated the map for the non-neutralizing antibody CR3022 by showing that mutations had the expected effects on binding of this antibody to mammalian-cell expressed RBD (Figure 3.9B-E).

3.3.3 Structural data partially but not completely explain the escape maps

We next examined the extent to which the escape maps could be rationalized in terms of the three-dimensional structures of the antibody-RBD complexes. We used negative-stain

Figure 3.2: **Complete maps of escape mutations from 10 human monoclonal antibodies. (A)** Properties of the antibodies as reported by [356]. SARS-CoV-2 neutralization potency is represented as a gradient from black (most potent) to white (non-neutralizing). Antibodies that bind SARS-CoV-1 spike or compete with RBD binding to ACE2 or CR3022 are indicated in black. **(B)** Structure of the SARS-CoV-2 RBD (PDB: 6M0J, [172]) with residues colored by whether they are in the core RBD distal from ACE2 (orange), in the receptor-binding motif (RBM, light blue), or directly contact ACE2 (dark blue). ACE2 is in gray. RBD sites where any antibody in the panel selects escape mutations are indicated with spheres at their alpha carbons. **(C)** Maps of escape mutations from each antibody. The line plots show the total escape at each RBD site (sum of escape fractions of all mutations at that site). Sites with strong escape mutations (indicated by purple at bottom of the line plots) are shown in the logo plots. Sites in the logo plots are colored by RBD region as in (B), with the height of each letter representing the escape fraction for that mutation. Note that different sites are shown for the CR3022-competing antibodies (top four) and all other antibodies (bottom six). For interactive zoom-able versions of these plots, see `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Crowe_antibodies`. **(D)** Multidimensional scaling projection of the escape-mutant maps, with antibodies having similar escape mutations drawn close together. Each antibody is shown with a pie chart that uses the color scale in (B) to indicate the RBD regions where it selects escape mutations.


A
SARS-CoV-2 neutralizing
SARS-CoV-1 cross-reactive
competes binding
ACE2
rCR3022
rCR3022
COV2-2677
COV2-2082
COV2-2094
COV2-2165
COV2-2832
COV2-2479
COV2-2050
COV2-2096
COV2-2499
B
core RBD
RBM
ACE2 contact
C
escape fraction
site
D
COV2-2165
COV2-2832
rCR3022
COV2-2050
COV2-2677
COV2-2479
COV2-2094
COV2-2096
COV2-2082
COV2-2499

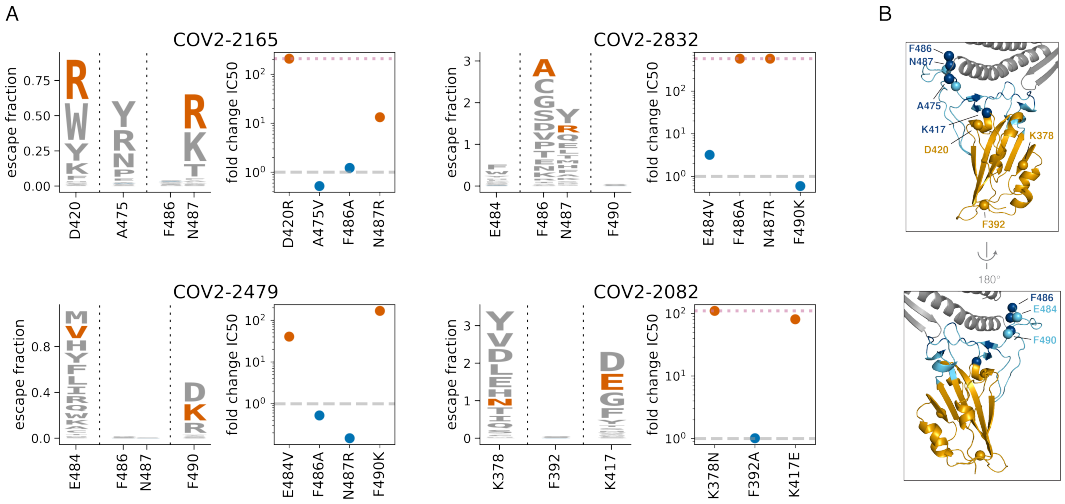


Figure 3.3: **Neutralization assays validate antibody escape maps. (A)** For each of the four indicated antibodies, we validated two mutations that our maps indicated should escape antibody binding, and one or two mutations that should not escape binding. Logo plots show the escape maps for all tested sites, with the tested mutations expected to escape antibody binding in red. Dot plots show the fold change in neutralization (inhibitory concentration 50%, IC50) relative to the unmutated (wildtype) spike measured using spike-pseudotyped lentiviral particles for all tested mutations. Fold changes greater than one (dashed gray line) mean a mutation escapes antibody neutralization. Points in red correspond to the mutations expected to mediate escape, and those in blue correspond to mutations not expected to escape (blue letters are not visible in the logo plots as they do not have substantial effects in the mapping). The dotted pink line at the top of some plots indicates the upper limit to the dynamic range; points on the line indicate a fold change greater than or equal to this value. See Figure 3.9A for the raw neutralization curves, and Figures 3.9B,C for similar validation for the non-neutralizing antibody CR3022. For these validation experiments, we chose mutations that had among the largest effects for each of the four antibodies, with additional prioritization of mutations that escaped from multiple antibodies, are present in circulating strains (e.g. A475V) or other sarbecoviruses (e.g. F490K, E484V), or were surprising in their lack of escape (e.g. confirming that COV2-2165, -2832, and -2479 each selected different subsets of escape mutations at sites 484, 486, 487, and 490). **(B)** RBD structure colored as in Figure 3.2A, with labeled spheres indicating sites where mutation effects on neutralization were validated.

electron microscopy (EM) to obtain structures of five of the antibodies in complex with the RBD, and analyzed an existing structure of CR3022 bound to RBD [347](Yuan et al., 2020). To enable structural interpretation of the escape mutations, we juxtaposed these structures of antibody-bound RBD with structural projections of our escape maps (Figure 3.4). We also created interactive structure-based visualizations of the escape maps using dms-view [137] that are available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Crowe_antibodies/` and enable facile analysis of escape mutations on the RBD structure.

Both the antibody-RBD structures and the escape maps highlight several antigenic regions on the RBD (Figure 3.4). The first region, targeted by four antibodies, is on the internal face of the core RBD (Figure 3.4A), which is only accessible in the context of full spike protein when the RBD transitions into the "open" conformation to engage ACE2 [143, 332, 306, 347]. Antibodies targeting regions of this surface more proximal to the RBM tend to more potently compete with ACE2 binding and neutralize virus (e.g. compare Figure 3.4A to Figure 3.2A). This observation is consistent with structural studies on another panel of antibodies targeting this core RBD surface [229].

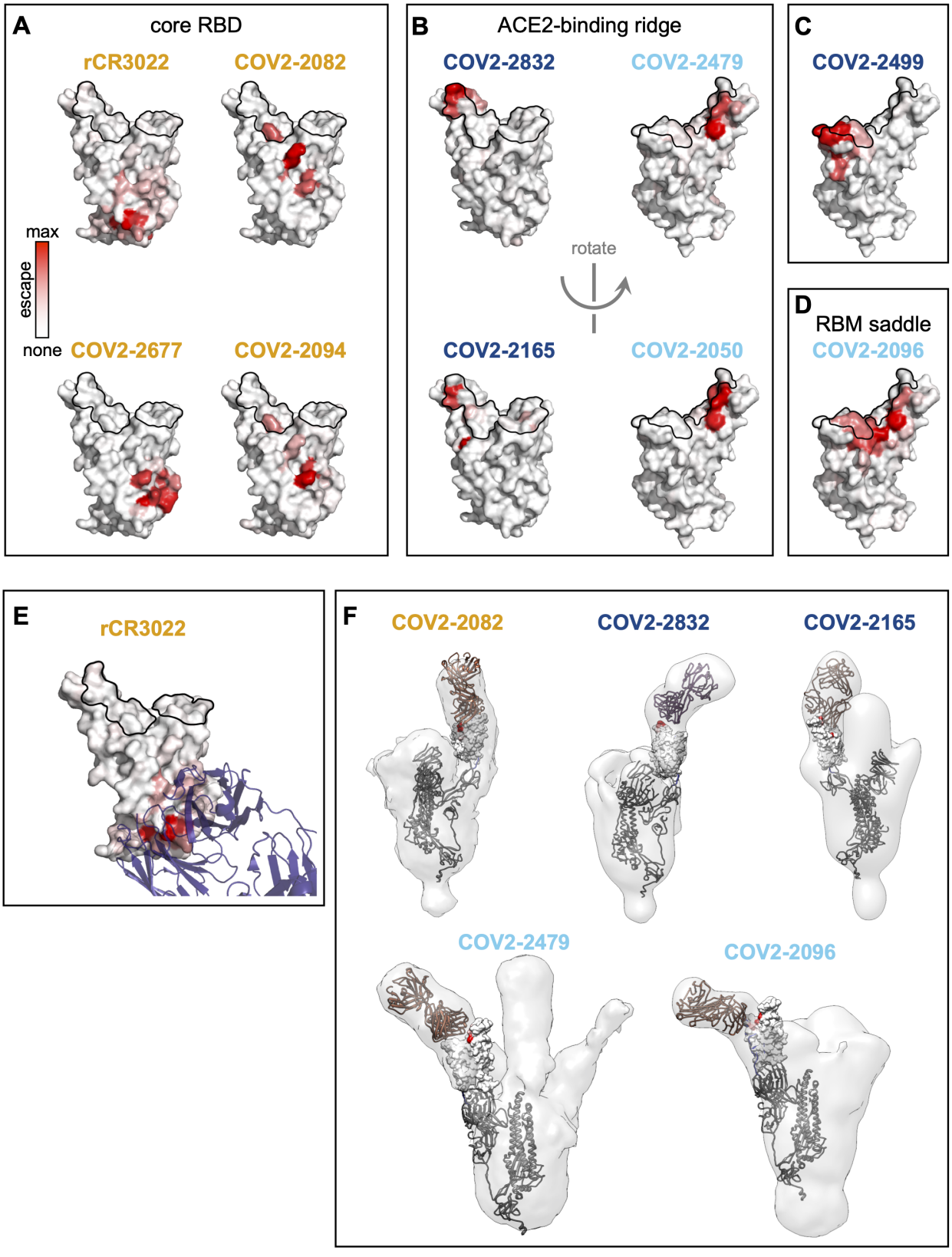
The remaining antibodies target several distinct regions on the RBM: four antibodies are escaped by mutations on the internal or external face of one lateral edge of the RBM (the "ACE2-binding ridge", Figure 3.4B), one antibody is escaped by mutations on the external face at the opposite edge of the RBM (Figure 3.4C), and one antibody is escaped by mutations that bridge the exterior surface of the central concave "saddle" of the RBM (Figure 3.4D). In all cases, the escape mutations fall in or near the structurally defined contact surface between the antibody and RBD (Figure 3.4E,F). In some cases, the negative-stain EM explains specific features of the escape-mutant maps (Figure 3.4F). For instance, COV2-2165 is strongly escaped by mutations at site D420 in addition to the ACE2-binding ridge, suggesting a binding footprint that extends beyond the ACE2-binding ridge. This hypothesis is supported by negative-stain EM data, which shows differences in the binding approach

of COV2-2165 relative to that of COV2-2832, another ACE2-binding ridge antibody that is not escaped by mutations at D420 (Figure 3.4F).

However, the escape-mutation maps contain substantial information beyond what can be gleaned from structure alone. For example, COV2-2832 and COV2-2479 both target the ACE2-binding ridge, but have non-overlapping escape mutations on different faces of the ridge (Figure 3.2C, Figure 3.4F). Similarly, while the negative-stain EM structures show that COV2-2165 and COV2-2832 both bind the ACE2-binding ridge, and the two antibodies select escape mutations very close to one another in the three-dimensional structure (Figure 3.4), there are important differences. For instance, COV2-2832 is escaped by mutations at sites F486 and N487, while COV2-2165 is only escaped by mutations at site N487 (Figure 3.2, Figure 3.4; validated by neutralization assays in Figure 3.3). In addition, while some antibodies (e.g., COV2-2096) can be escaped by mutations across a wide swath of the RBD surface, others (e.g., COV-2050) are only sensitive to mutations at a handful of sites. Analysis of the high-resolution CR3022-bound RBD crystal structure further emphasizes heterogeneity in escape across antibody-contact residues, and suggests particular mechanisms by which mutations to the most strongly selected sites mediate escape (Figure 3.9F-J).

The fact that the escape mutations occur at only a subset of sites in the antibody-RBD interfaces is consistent with classical biochemical studies showing that protein-protein binding interfaces can be dominated by "hot spots" that contribute most of the binding energy [61, 68], and more recent work showing that the functional and structural epitopes of anti-viral antibodies are often distinct [81]. From a therapeutic standpoint, these results emphasize the value of directly mapping escape mutations when considering the potential for viral antibody escape. For instance, our results suggest that it should be possible to make effective cocktails of antibodies with similar structural epitopes but orthogonal escape mutations, such as COV2-2165 + COV2-2479 or COV2-2499 + COV2-2050.

Figure 3.4: **Structural mapping of antibody binding and escape. (A-D)** For each antibody, the structure shows the RBD surface (PDB 6M0J, [172]) colored by the largest-effect escape mutation at each site, with white indicating no escape and red indicating the strongest escape mutation for that antibody. ACE2 contact residues are outlined in black. Antibodies are arranged so that those with similar structural epitopes are in the same panel, namely by whether their epitopes are in (A) the core of the RBD, (B) the ACE2-binding ridge, (C) the opposite edge of the RBM, or (D) the saddle of the RBM surface. **(E)** Crystal structure of the CR3022-bound RBD (PDB 6W41 [347]), with Fab in purple, ACE2 contact sites outlined, and RBD colored according to sites of escape as in (A). **(F)** For 5 monoclonal antibodies, Fab bound to SARS-CoV-2 spike ectodomain trimer was visualized by negative-stain electron microscopy (EM). The RBD is modeled as a surface representation, colored according to sites of escape as in (A). Fab chains are modeled in gold. Detailed EM collection statistics are in Table S1. Antibody names are colored according to Figure 3.2B: core-binding, orange; RBM-binding, cyan; ACE2 contact site-binding, dark blue. See `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Crowe_antibodies` for interactive versions of the escape-colored structures in (A-D).


A
core RBD
rCR3022
COV2-2082
max
escape
none
COV2-2677
COV2-2094
B
ACE2-binding ridge
COV2-2832
COV2-2479
rotate
COV2-2165
COV2-2050
C
COV2-2499
D
RBM saddle
COV2-2096
E
rCR3022
F
COV2-2082
COV2-2832
COV2-2165
COV2-2479
COV2-2096

3.3.4 Functional and evolutionary constraint on antibody-escape mutations

Our complete maps of escape mutations enable us to assess the potential for SARS-CoV-2 to evolve to escape antibodies targeting the RBD. We first examined whether the antibody-escape mutations identified in our study are present in viruses circulating in the human population. Of 93,858 SARS-CoV-2 sequences in the GISAID database as of September 6, 2020, there were 5 or more naturally occurring mutants at 14 of the 36 RBD sites where mutations escape at least one antibody (Figure 3.5A, Figure 3.10). However, mutations at all these sites are present only at very low frequency ($< 0.1\%$ of viral sequences). The antibody-escape sites with naturally occurring mutations include sites 484 and 490, where other studies have recently reported selecting mutations that escape monoclonal antibodies or sera containing polyclonal antibodies [19, 320, 182]. Overall, these results show that while the vast majority of viruses remain susceptible to all antibodies examined here, there is nascent low-level viral genetic variation at some key sites of escape mutations.

To better assess the potential for future viral genetic variation, we quantified the functional constraint on sites of escape using existing deep mutational scanning measurements of how RBD mutations affect ACE2-binding and expression of properly folded RBD protein [280]. Figure 3.5B shows the escape maps for two antibodies colored by the functional effects of mutations (comparable data for all antibodies are in Figure 3.11). It is obvious from Figure 5B that some escape mutations from the core-RBD-directed antibody COV2-2094 are deleterious for expression of properly folded RBD (e.g., mutations at site 435), whereas some escape mutations from the RBD-directed antibody COV2-2165 are deleterious for ACE2 binding (e.g., mutations at site 487). To quantify this trend, we determined the mean functional effect of all mutations at each site of escape from each antibody (Figure 3.5C). At a broad level, sites of escape from antibodies targeting the RBM and especially ACE2-contact residues are often constrained by how mutations affect ACE2 binding. On the other hand, sites of escape from antibodies targeting the core RBD are often constrained by

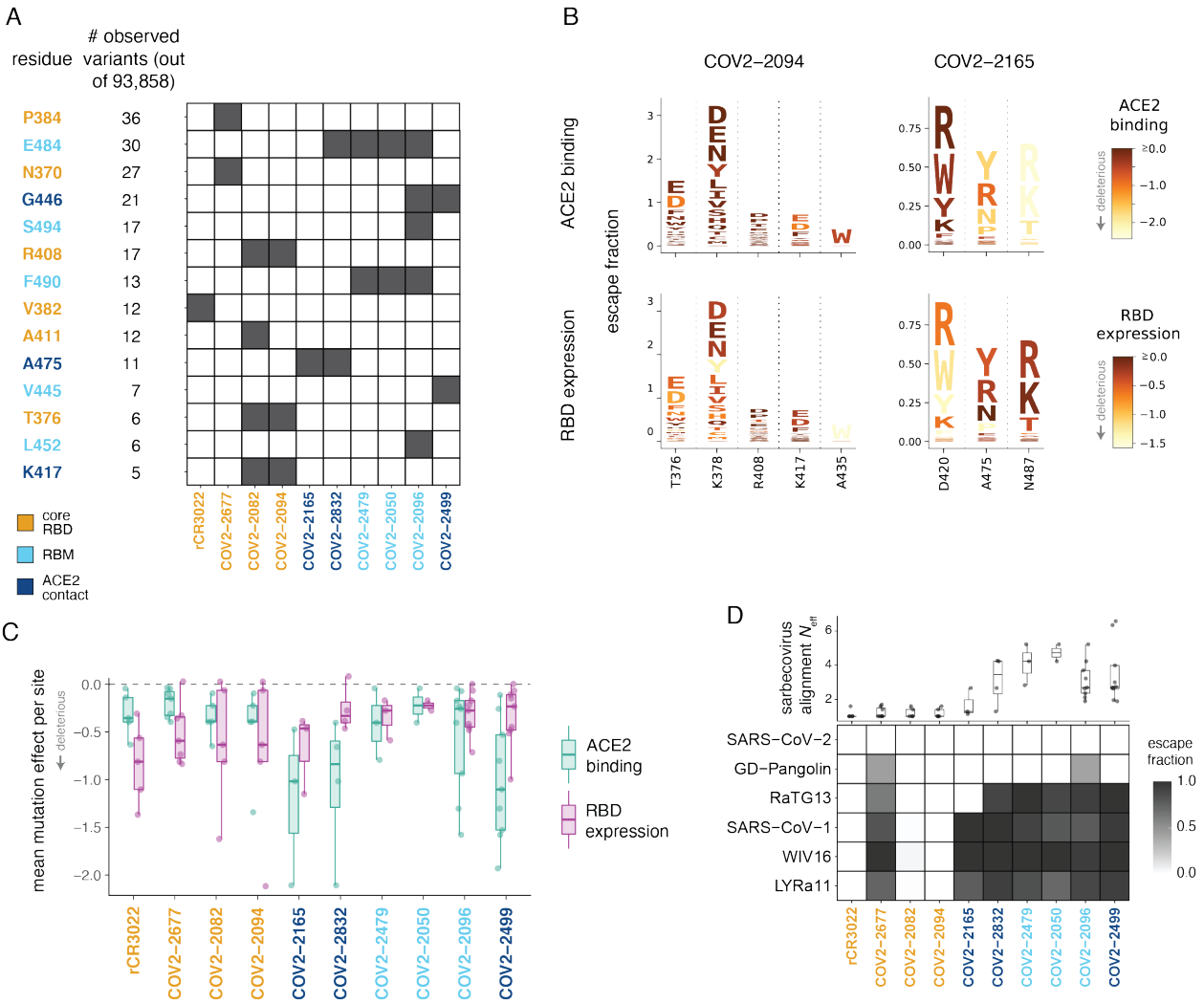
how mutations affect RBD folding and expression (Figure 3.5B). These observations highlight how some antibodies target RBD sites that are functionally constrained and thus may have reduced potential for evolution.

We also examined the ability of each antibody to bind RBDs from other SARS-related coronaviruses (sarbecoviruses). To do this, we included in our libraries the unmutated RBDs from two close relatives of SARS-CoV-2 (RaTG13 [352] and GD-Pangolin [171]), along with SARS-CoV-1 and two of its close relatives (WIV16 [343] and LYRa11 [132]). Using the same approach employed to measure the effects of mutations to SARS-CoV-2, we quantified the ability of each antibody to bind these RBD homologs. We found a stark difference in cross-sarbecovirus reactivity between antibodies targeting the core RBD and those targeting the RBM (Figure 3.5D). Three of four antibodies targeting the core RBD bound to all five RBD homologs, whereas RBM-directed antibodies only bound the two homologs most closely related to SARS-CoV-2 (GD-Pangolin and RaTG13). This pattern is explained by the evolutionary conservation at sites of escape (Figure 3.5D, top): in general, sites of escape from antibodies targeting the RBD core are mostly conserved across sarbecoviruses, while sites of escape from RBM-directed antibodies are highly variable across sarbecoviruses. The only exception is COV2-2677, which does not bind any other RBD homologs despite targeting conserved sites in the core RBD: this discrepancy is explained by the A372T escape mutation, which restores an N370 glycosylation motif that is present in all sarbecoviruses except SARS-CoV-2. These results show that antibodies targeting the conserved core RBD are more likely than antibodies targeting the RBM to provide pan-sarbecovirus immunity.

3.3.5 Escape maps predict results of antibody selection experiments and inform design of cocktails

We next examined if the escape maps accurately predicted the mutants selected when virus is grown in the presence of antibody. To investigate this, we used a recombinant replication-

Figure 3.5: **Functional and evolutionary constraint on antibody escape mutations.** **(A)** Variation at sites of antibody escape among currently circulating SARS-CoV-2 viruses. For each site of escape from at least one antibody, we counted the sequences in GISAID with an amino-acid change (there were 93,858 sequences at the time of the analysis). Sites with at least 5 GISAID variants are shown ordered by count; black cells indicate antibodies with escape mutations at that site. Sites are in orange for the core RBD, light blue for the RBM, and dark blue for ACE2 contact residues. Antibodies are colored according to where the majority of their sites of escape fall. Figure 3.104 shows similar data broken down by amino-acid change and without count thresholding. **(B)** Escape maps (as in Figure 3.2C), with letters colored according to how deleterious mutations are for ACE2 binding or RBD expression effects [280]. Only sites of escape mutations for each antibody are depicted. Similar logo plots for all antibodies are shown in Figure 3.11. **(C)** Mutational constraint on sites of escape. For each antibody, the mean effects of all 19 possible amino acid mutations at sites of escape on ACE2 binding and RBD expression are shown. **(D)** Top: effective number of amino acids (N_{eff}) in the sarbecovirus RBD alignment at sites of escape for each antibody. N_{eff} is a measure of the variability of a site (the exponentiated Shannon entropy), and ranges from 1 for a position that is conserved across all sequences to an upper limit of 20 for a site where all amino acids are present at equal frequency. Bottom: escape fraction for each sarbecovirus RBD homolog from the yeast display selections, with an escape fraction of 1 indicating a RBD homolog completely escapes (no binding), and an escape fraction of 0 indicating no escape (antibody binds).


A
residue
observed variants (out of 93,858)
P384 36
E484 30
N370 27
G446 21
S494 17
R408 17
F490 13
V382 12
A411 12
A475 11
V445 7
T376 6
L452 6
K417 5
core RBD
RBM
ACE2 contact
rCR3022
COV2−2677
COV2−2082
COV2−2094
COV2−2165
COV2−2832
COV2−2479
COV2−2050
COV2−2096
COV2−2499
B
COV2−2094
COV2−2165
ACE2 binding
RBD expression
escape fraction
deleterious
T376
K378
R408
K417
A435
D420
A475
N487
C
mean mutation effect per site
deleterious
ACE2 binding
RBD expression
D
sarbecovirus alignment N_{eff}
SARS−CoV−2
GD−Pangolin
RaTG13
SARS−CoV−1
WIV16
LYRa11
escape fraction

competent vesicular stomatitis virus (VSV) expressing the SARS-CoV-2 spike in place of the endogenous VSV glycoprotein (G) [40]. Such viruses provide a facile system to select for spike mutations that evade antibody neutralization [40, 320, 80]. We chose five potently neutralizing antibodies (IC50 values ranged from 15 to 150 ng/mL), and used a high-throughput quantitative real-time cell analysis assay [117, 116] to select viral mutants that could escape each individual antibody at a single saturating concentration of 5 μg/mL, performing between 16 and 56 individual replicates for each antibody (Figure 3.6A and 3.12A,B,C). For four of the five antibodies, this process selected viral variants that we confirmed resisted neutralization by 10 μg/mL of the antibody used for the selection (Figure 3.6A). For one antibody (COV2-2165), no escape mutants were detected even in 56 attempted replicates (Figure 3.6A). We sequenced the antibody-selected escape viruses, and in all cases they carried RBD mutations that the escape maps indicated mediate strong escape (examine the mutations in Figure 3.6A on the maps in Figure 3.2C).

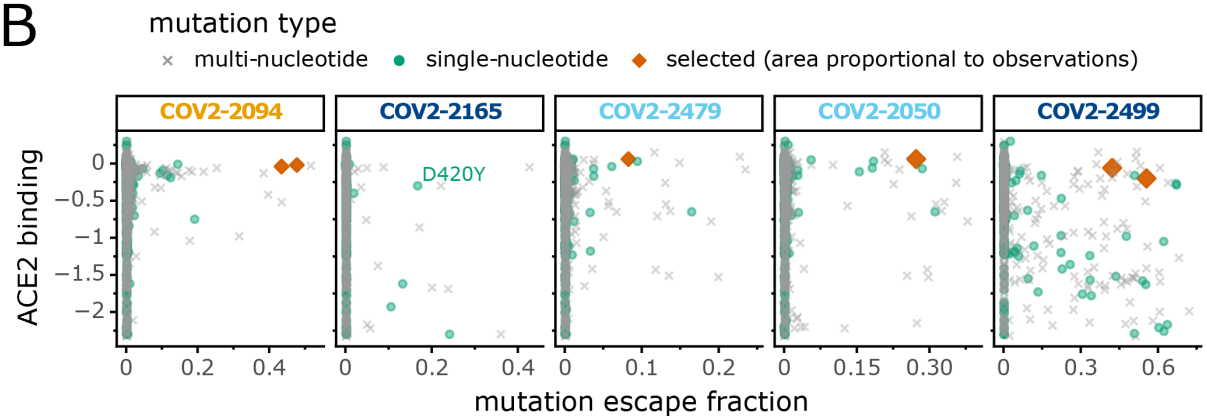
We next sought to understand why the antibodies selected the viral mutations that they did–and why it was not possible to select any viral mutants that escaped one of the antibodies. To do this, we considered two additional factors: which mutations are tolerated for protein function, and which mutations are accessible by single-nucleotide changes. We assessed how well mutations are tolerated functionally using deep mutational scanning measurements of how all RBD mutations affect ACE2 binding [280]. We plotted all mutations in scatter plots to examine their impact on antibody escape and ACE2 binding, further stratifying by whether mutations were accessible by single-nucleotide changes to the spike gene encoded in the VSV (Figure 3.6B). The mutations selected by the antibodies were consistently among the ones with the largest effects on antibody escape that also did not greatly impair ACE2 binding and were accessible by single-nucleotide changes (red diamonds in Figure 3.6B). The antibody for which we could not select any viral escape mutants (COV2-2165) only had a single escape mutation (D420Y) that was accessible by a single-nucleotide change and not

Figure 3.6: **Viral escape-mutant selections with individual antibodies and antibody cocktails**. **(A)** Results of viral selections with five individual monoclonal antibodies. The number of replicates where escape variants were selected are indicated, color coded according to whether escape was selected frequently (red) or rarely (white). The mutations present in the RBD of the selected escape variants are indicated. **(B)** Each point represents a different amino-acid mutation to the RBD, with the x-axis indicating how strongly the mutation ablates antibody binding in our escape maps (larger values indicate more escape from binding) and the y-axis indicating how the mutation affects ACE2 binding (negative values indicate impaired ACE2 binding). The point shapes indicate whether or not mutations are accessible by single-nucleotide changes, and whether they were selected in viral escape experiments. All selected mutations were accessible by single-nucleotide changes. Note that the only accessible escape mutation from COV2-2165 that is not deleterious to ACE2 binding is D420Y, but this mutation is highly deleterious to expression of properly folded RBD (Figure 3.5B and 3.11). **(C)** Results of viral selections with antibody cocktails, with the last three columns showing the number of replicates with escape out of the total tested. The data for the single antibodies are repeated from (A). In all panels, antibody names are colored according to where in the RBD the majority of their sites of escape fall: orange for the core RBD, light blue for the RBM, and dark blue for ACE2 contact residues. See Figure 3.12 for additional data relevant to this figure.

A

antibody	replicates with escape / total replicates (%)*	confirmed escape at 10 µg/mL antibody	mutations selected (number of times)**	antibody that neutralized escape mutants at 10 µg/mL
COV2-2094	8 / 24 (33%)	yes (all tested)	K378E (1), K378N (1)	COV2-2499
COV2-2165	0 / 56 (0%)	not applicable	not applicable	not applicable
COV2-2479	3 / 56 (5%)	yes (all tested)	E484K (1)	COV2-2165
COV2-2050	4 / 30 (13%)	yes (1 tested)	E484K (2)	COV2-2499
COV2-2499	7 / 16 (44%)	yes (all tested)	G446D (2), Q498R (2)	COV2-2094

B



C

first antibody	second antibody	antibodies compete for binding	first antibody escape	second antibody escape	cocktail escape
COV2-2499	COV2-2094	no	7 / 16 (44%)	8 / 24 (33%)	0 / 80 (0%)
COV2-2479	COV2-2165	yes	3 / 56 (5%)	0 / 56 (0%)	0 / 56 (0%)
COV2-2499	COV2-2050	yes	7 / 16 (44%)	4 / 30 (13%)	0 / 80 (0%)

combined data from two experiments

highly deleterious for ACE2 binding. However, D420Y is extremely deleterious for expression of properly folded RBD protein (Figure 3.5B and 3.11), explaining why it was not possible to select any viral escape mutants from COV2-2165. Therefore, the escape maps can be combined with deep mutational scanning of functional constraint and basic knowledge of the genetic code to predict which viral mutations are likely to arise under antibody pressure–and to identify antibodies for which escape mutations are unlikely.

One approach to thwart the risk of viral escape that is inherent in monotherapy approaches is to use antibody cocktails [150, 317]. In the context of SARS-CoV-2, recent work has demonstrated that cocktails of two antibodies that do not compete for binding to the same region of spike may offer higher resistance to escape mutations [19] while protecting animals from SARS-CoV-2 challenge [18, 356]. We hypothesized that we could leverage our escape maps to rationally design more nuanced cocktails of antibodies with distinct escape mutations, even if the antibodies recognize overlapping antigenic regions and compete for binding to spike.

We created three different two-antibody cocktails: one "conventional" cocktail of antibodies that did not compete for binding to spike protein (COV2-2499 + COV2-2094), and two cocktails of antibodies that competed for binding to the RBM region of the spike protein RBD but that our maps indicated were escaped by distinct mutations (COV2-2479 + COV2-2165 and COV2-2499 + COV2-2050) (Figure 3.6C and 3.12D). Each cocktail contained a 1:1 mix of the two constituent antibodies at a total concentration of 5 μg/mL of antibody, so that the total antibody concentration was the same as in the single-antibody selections described above. We performed between 80 and 104 escape-selection replicates with each cocktail. No cocktail escape mutants were identified in any of these replicates, despite the fact that two of the cocktails were composed of antibodies for which substantial numbers of escape mutants were selected by the individual antibodies (Figure 3.6C and 3.12C). The lack of cocktail escape mutants is likely due to the "orthogonality" of the escape mutations

for the individual antibodies, as viruses with the mutations selected by each single antibody were sensitive to the other antibody in the cocktail (Figure 3.6A). A caveat is that we performed the selections for 72 h days at high antibody concentrations: it is possible that multi-mutation escape pathways might be selected in longer experiments at sub-neutralizing antibody concentrations. Overall, these results demonstrate how complete escape maps can inform the design of "non-conventional" cocktails of antibodies that compete for binding to the antigen but are nonetheless resistant to viral escape because they have orthogonal escape mutations (e.g., the cocktail COV2-2499 + COV2-2050).

3.4 Discussion

We have described an approach to completely map mutations to the SARS-CoV-2 RBD that escape antibody binding. Unlike traditional selection experiments that only identify a handful of the possible escape mutations, our method completely maps mutations that escape antibody binding. These maps complement structure-based approaches that define the physical interface between an antibody and virus but do not directly measure how mutations affect antibody binding.

The escape maps reveal remarkable nuance in which mutations escape individual antibodies. Our maps show that at a superficial level, the antibodies target just a few patches on the surface of the RBD that likely correspond to "antigenic regions" that have been defined using other approaches [16, 33, 188, 229, 251, 255, 318, 356]. However, the fine details of the escape maps show that the effects of specific mutations can vary dramatically even among antibodies that superficially target the same region. For instance, antibodies COV2-2479, COV2-2050, and COV2-2832 all target the RBD ACE2-binding ridge–but while the first two have nearly identical escape mutations, the escape mutations for COV2-2832 are almost completely distinct. We speculate that these differences arise from the fact that even antibodies that physically contact a large surface area on the RBD are often only escaped by

mutations at a few residues, a vivid illustration of the classically defined importance of "hot spots" in antibody-antigen binding [29, 70, 148]).

We also overlaid the escape maps with existing deep mutational scanning data on the functional consequences of mutations for the expression of properly folded RBD and its affinity for ACE2 [280]. In general, the sites of escape from antibodies directed to the core RBD are constrained with respect to their effects on expression of properly folded RBD, whereas sites of escape from antibodies directed to the RBD's receptor-binding motif are more constrained with respect to their effects on ACE2 binding. While these analyses come with the caveat that no experimental measures of the effects of mutations fully capture how they affect true viral fitness, it is nonetheless informative to assess how mutations that escape antibody binding impact the key biochemical functions of the RBD.

Remarkably, combining the escape maps with these functional measurements predicts which mutations are selected when spike-expressing virus is grown in the presence of individual antibodies. The selected viral escape mutations are consistently those that have large effects on antibody escape but little negative impact on ACE2 binding and RBD folding, and are also accessible by single-nucleotide mutations. Furthermore, one of the antibodies was highly resistant to viral escape–and we showed this could be explained by the fact that the virus has no escape mutations from this antibody that are both tolerable for RBD function and accessible by single-nucleotide changes. Therefore, complete measurements of both the antigenic and functional consequences of viral mutations provide the phenotypic data necessary to assess both the likelihood of viral escape under antibody pressure and the specific mutations that arise when escape occurs.

One immediate implication of our results is that counter to prevailing wisdom, antibody cocktails do not have to target distinct regions of the RBD in order to resist viral escape. Simple inspection of the escape maps reveals pairs of antibodies targeting the RBD's ACE2-binding interface that share no common escape mutations, and so could be good candidates

for therapeutic cocktails. Indeed, we combined our escape maps with selections on spike-expressing viruses to show that cocktails of antibodies that compete for binding to spike but have different escape mutations still resist viral escape. It is possible that such cocktails could even be preferable to cocktails of antibodies targeting distinct regions [262, 264], since acquiring multiple different escape mutations in the ACE2 binding interface could impose an intolerable loss of receptor binding on the virus.

Our results are also of utility for assessing if ongoing viral evolution is likely to be of antigenic consequence. The escape maps enable immediate assessment of whether mutations to the RBD alter antigenicity. At over a dozen of the sites of escape that we mapped for these antibodies, there is already low-level genetic variation among circulating SARS-CoV-2 strains. Furthermore, the high-throughput nature of our experimental approach should make it possible to rapidly generate similar maps for other monoclonal antibodies or polyclonal antibodies in sera, thereby providing quantitative experimental data that can be cross-referenced to mutations observed during genomic surveillance of circulating SARS-CoV-2 strains [161]. Note however that we only examined the effects of single mutants, as mutations typically fix in a stepwise manner [276], and processes that require multiple mutations tend to occur more slowly in viral evolution [104]. However, as the SARS-CoV-2 virus continues to circulate in humans, multiple mutations will eventually fix, and so the examination of the antigenic effects of multiple mutations is an important area for future work.

It is important to note that our approach maps how mutations affect antibody binding to yeast-displayed RBD, which comes with several caveats. First, our approach can only map escape from antibodies that target epitopes entirely within the RBD, and will not identify mutations that mediate escape by altering the relative positioning of the RBD in the context of full spike protein [321, 354]. Second, our method assesses binding to RBD displayed in monomeric form, which means it is likely unable to map escape from antibodies that bind quaternary epitopes spanning multiple RBD protomers in the spike trimer [16, 188, 293], and

definitely unable to map antibodies that bind epitopes outside of the RBD [56, 188]. Last, although yeast do add N-linked glycans to the RBD at the same sites as human cells [53], these glycans are more mannose-rich [127], which could affect binding by antibodies with glycan-rich epitopes. However, despite these potential caveats, all the mapped escape mutations that we tested had the expected effects in the context of spike-pseudotyped lentiviral or VSV particles. In addition, our approach can map mutations that escape binding by non-neutralizing as well as neutralizing antibodies, and we successfully validated mutations that ablated binding by a non-neutralizing antibody using mammalian-cell produced RBD.

Some viruses, such as measles, are antigenically stable such that immunity from an initial infection or vaccination typically provides life-long protection [186, 225]. Others, such as influenza virus, undergo rapid antigenic drift, such that immunity elicited against one viral strain can be ineffective against that strain's descendents just a few years later [176, 274]. The extent to which mutations that substantially affect the antigenicity of SARS-CoV-2 will fix during viral evolution remains an open question. The escape-mutation maps we have generated, as well our methodology for rapidly creating such maps for additional antibodies and sera, should help answer this question by facilitating assessment of the antigenic consequences of mutations observed during viral surveillance.

3.5 Materials and Methods

Data and code availability

We provide data and code in the following ways:

- Raw data tables of single-mutation escape fractions, averaged across libraries (GitHub: `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/results/supp_data/MAP_paper_antibodies_raw_data.csv`)

- Raw data table of single-mutation escape fractions, measurements for individual library

replicates (GitHub: `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/results/escape_scores/escape_fracs.csv`)

- Illumina sequencing counts for each barcode in each antibody escape bin (GitHub: `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/results/counts/variant_counts.csv`)

- The complete computational pipeline to analyze these data (GitHub: `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies`)

- A Markdown summary of the organization of analysis steps, with links to key data files and Markdown summaries of each step in the analysis pipeline (Github: `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/results/summary/summary.md`)

- All raw sequencing data are uploaded to the NCBI Short Read Archive (BioProject: PRJNA639956, BioSample: SAMN16054076)

- Electron density maps for the Fab/SARS-CoV-2 S complex are available from the Electron Microscopy Data Bank under the following accession codes: EMD-22627 and EMD-22628.

Description of RBD deep mutational scanning library

The yeast-display RBD mutant libraries are identical to those previously described [280]. Briefly, mutant libraries containing an average of 2.7 amino-acid mutations per variant were constructed in the spike receptor binding domain (RBD) from SARS-CoV-2 (isolate Wuhan-Hu-1, Genbank accession number MN908947, residues N331-T531). Duplicate mutant libraries were generated, and contain 3,804 of the 3,819 possible amino-acid mutations, with

¿95% present as single mutants. Each RBD variant was linked to a unique 16-nucleotide barcode sequence to facilitate downstream sequencing. The RBD mutant library also contained non-mutated sarbecovirus RBD homologs, RaTG13, Genbank MN996532; GD-Pangolin consensus from [171]; SARS-CoV-1 Urbani, Genbank AY278741; WIV16, Genbank KT444582; and LYRa11, Genbank KF569996.

Human monoclonal antibodies targeting SARS-CoV-2 RBD

The 9 human monoclonal antibodies isolated from SARS-CoV-2 convalescent patients were produced as described in [357] The recombinant CR3022 antibody (CR3022), was kindly provided by Neil King and Mike Murphy, University of Washington, Institute for Protein Design, based on the sequence reported by [286]. All antibodies were expressed as human IgG. Properties of the ten antibodies represented in Figure 3.2A were reported by [356]: SARS-CoV-2 neutralization potency (black, IC50 $<$ 150 ng/mL; dark gray, 150-1,000; light gray, 1,000-1:10,000; white, no detectable inhibition); SARS-CoV-1 spike binding via ELISA (black, detectable; white, no detectable binding); potency of ACE2 competition via ACE2-blocking ELISA (black, IC50 $<$ 150 ng/mL; white, no competition); and CR3022 competition via ELISA (black, $<$ 25% baseline CR3022 binding when pre-incubating with saturating antibody; white, $>$ 60% of baseline CR3022 binding).

Fluorescence activated cell sorting (FACS) of yeast libraries to eliminate mutants that are completely non-folded or do not bind ACE2

Libraries were sorted for RBD expression and ACE2 binding to eliminate RBD variants that are completely misfolded or non-functional (Figure 3.7A,B). We chose staining and sorting conditions that would select for variants with ACE2 affinity comparable to or better than RaTG13, the homolog with the lowest affinity that still marginally mediates cell entry [268]. Yeast library aliquots of 18 OD units (approx. 1e8 cfus) were thawed into 180 mL SD-

CAA (6.7 g/L Yeast Nitrogen Base, 5.0 g/L Casamino acids recipe, 1.065 g/L MES, and 2% w/v dextrose) and grown overnight shaking at 30°C, 280rpm. 33.3 OD units were back-diluted into 50 mL SG-CAA+0.1% dextrose (SD-CAA with 2% w/v galactose and 0.1% w/v dextrose in place of 2% dextrose) to induce RBD surface expression. Yeast were induced for 16-18 h at 23 degrees C with mild agitation. 25 OD units of cells were washed twice with PBS-BSA (1x PBS with 0.2 mg/mL BSA), and incubated with 1e-8 M biotinylated ACE2 (ACROBiosystems AC2-H82E6) for 1 h at room temperature. Cells were washed with ice-cold PBS-BSA before secondary labeling for 1 h at 4 degrees C in 3 mL 1:200 PE-conjugated streptavidin (Thermo Fisher S866) to label for bound ACE2, and 1:100 FITC-conjugated anti-Myc (Immunology Consultants Lab, CYMC-45F) to label for RBD surface expression. Labeled cells were washed twice with PBS-BSA and resuspended in 2.5 mL PBS. FACS was used to enrich RBD libraries for cells capable of binding ACE2, via a selection gate drawn to capture unmutated SARS-CoV-2 cells labeled at 1% the ACE2 concentration of the library samples (i.e., 1e-10 M ACE2) (Figure 3.7B). 15 million ACE2+ cells were collected for each library, grown overnight in SD-CAA medium, and stored at -80°C in 9 OD unit (approx. 5e7 cfus) aliquots.

Sorting of yeast libraries to select mutants that escape binding by antibodies

Antibody selection experiments were performed in biological duplicate using the independently generated mutant RBD libraries. One 9 OD unit aliquot of each ACE2+-enriched RBD library was thawed and grown overnight in 45 mL SD-CAA. Libraries were induced as described above. Induced cultures were washed and incubated with 400 ng/mL antibody for 1 h at room temperature with gentle agitation, followed by secondary labeling with 1:100 FITC-conjugated anti-Myc to label for RBD expression and 1:200 PE-conjugated goat anti-human-IgG (Jackson ImmunoResearch 109-115-098) to label for bound antibody. A flow cytometric selection gate was drawn to capture unmutated SARS-CoV-2 cells labeled at 1%

the antibody concentration of the library samples (Figure 3.7C). Libraries were sorted to select RBD variants that reduce antibody binding and fall into this selection gate. For each sample, approximately 10 million RBD+ cells were processed on the cytometer, with between 4e5 and 2.6e6 antibody-escaped cells collected per sample (see percentages in Figure 3.7C for what fraction of the library had reduced binding to each antibody). Antibody-escaped cells were grown overnight in SD-CAA to expand cells prior to plasmid extraction.

DNA extraction and Illumina sequencing

Plasmid samples were prepared from overnight cultures of antibody-escaped and 30 OD units (1.6e8 cfus) of pre-selection yeast populations (Zymoprep Yeast Plasmid Miniprep II). The 16-nucleotide barcode sequences identifying each RBD variant were amplified by PCR and prepared for Illumina sequencing exactly as described in Starr et al. (2020). Barcodes were sequenced on an Illumina HiSeq 3500 with 50 bp single-end reads. To minimize noise from inadequate sequencing coverage, we ensured that each antibody-escape sample had at least 3x as many post-filtering sequencing counts as FACS-selected cells, and reference populations had at least 2.5e7 post-filtering sequencing counts.

Analysis of deep sequencing data to compute antibody escape fraction for each mutation

We computed escape fractions for each mutation from the counts in the Illumina deep sequencing of the 16-nucleotide barcodes as schematized in Figure 3.1C. We first used the dms_variants package (`https://jbloomlab.github.io/dms_variants/`, version 0.8.2) to process the Illumina sequences into counts of each barcoded RBD variant in each condition using the barcode / RBD-variant look-up table described in Starr et al. (2020). A rendering of the code that performs this variant counting is at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/results/summary/count_variants.md`.

We then computed the "escape fraction" for each barcoded variant in each antibody-

selected library, which we define as:

$$E_v = F \times \frac{({n_v}^{\text{post}}/N_{\text{post}})}{({n_v}^{\text{pre}}/N_{\text{pre}})}$$

where F is the total fraction of the library that escapes antibody binding (these fractions are given as percentages in the bottom two rows of Figure 3.7C), ${n_v}^{\text{post}}$ and ${n_v}^{\text{pre}}$ are the counts of variant v in the RBD library after and before enriching for antibody-escape variants with a pseudocount of 0.5 added to all counts, and $N_{\text{post}} = \sum_v {n_v}^{\text{post}}$ and $N_{\text{pre}} = \sum_v {n_v}^{\text{pre}}$ are the total counts of all variants before and after the antibody-escape enrichment. These escape fractions represent the fraction of a given variant that escape antibody binding, and should in principle range from 0 to 1. But due to statistical fluctuations in the counts sometimes the escape fractions E_v can be greater than one: any values of $E_v > 1$ were set to 1.

We then computationally applied two filters to remove variants that fail to express properly folded RBD and so escape antibody binding for that trivial reason rather than antibody-specific escape mutations. In principle, such variants should have been fully removed by the initial sort that only retained yeast cells with appreciable RBD expression and ACE2 binding, but in practice a small background remained as demonstrated by the fact that stop-codon variants were present at very low but still non-zero levels. For the first filter, we removed all variants with pre-selection counts lower than the counts in the 99th percentile of stop-codon-containing variant ordered by count. The logic was that this filter removed nearly all variants that were observed less frequently than stop-codon variants, which are assumed to not express properly folded RBD. For the second filter, we removed any variants that had ACE2-binding scores < -2.35 or RBD expression scores < -1.5 using the scores measured in Starr et al. (2020). In addition, we removed any variants that had single mutations with scores less than either of these thresholds (again using the single-mutation scores determined in Starr et al. (2020)) even if the variant score itself was above this threshold. The logic was that this filter removed any variants that fail to express at least low levels of properly folded ACE2. A rendering of the code that performs the computation of the escape fractions and

this subsequent filtering is at https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/results/summary/counts_to_scores.md.

We next deconvolved the variant-level escape fractions into escape fraction estimates for individual mutations. To do this we used global epistasis models (Otwinowski et al., 2018) as implemented in the the dms_variants package as detailed at https://jbloomlab.github.io/dms_variants/dms_variants.globalepistasis.html, using the same Gaussian likelihood function as in Otwinowski et al. (2018). In order to make the fitting more reliable, we removed any variants with mutations not seen in at least one single-mutant variant and at least two variants overall. We report the escape fraction on the "observed phenotype" scale: that is, we use the global epistasis models to transform the variant-level escape fractions to estimated latent phenotypes for each mutation, and then re-transform those latent phenotype estimates back through the global epistasis model. If any of these re-transformed escape fractions were not in the range between 0 and 1, they were adjusted to a minimum value of 0 or a maximum value of 1. The end result of this process was a separate estimate for each library and antibody of the escape fraction for each mutation that was not highly deleterious for expression of properly folded RBD. The correlation between these estimates for the different libraries is in Figure 3.8. In this paper, we report the average of the two libraries, and in the rare cases a mutation is only sampled in one library then we report the value for just that library. These values are reported in https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/results/supp_data/MAP_paper_antibodies_raw_data.csv. The code that performs this global epistasis decomposition of escape scores for individual mutations is at https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/results/summary/scores_to_frac_escape.md.

In some places in this paper, we report site-level measurements in addition to mutation-level escape scores. The first measure of site-level escape is the total site escape (total height

of letter stacks, e.g. in Figure 3.1C), and simply represents the sum of all mutation-level escape fractions at a site. The second measure of site-level escape is the maximum escape at a site, which is just the maximum of all of the mutation-level escape fractions at the site.

Classification of sites of escape from each antibody

For certain visualizations or analyses, it was necessary to classify which sites mediated escape from each antibody. To do this, for each antibody we identified those sites where the total site escape was $> 10\times$ the median across all sites, and was also at least 10% of the maximum total site escape for any site for that antibody. We found that this heuristic reliably separated sites of clear antibody escape from other sites. This approach was used to determine which sites to display in the logo plots, and which sites to include in the analysis of natural sequence variation.

Data visualization

The static logo plot visualizations of the escape maps in the paper figures were created using the dmslogo package (`https://jbloomlab.github.io/dmslogo/`, version 0.3.2) and in all cases the height of each letter indicates the escape fraction for that amino-acid mutation calculated as described above. In Figure 3.2, we have separated the antibodies into two groups, and for each group the logo plots show all sites of escape from any antibody in that group according to the classification scheme described above. The code that generates these logo plot visualizations is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/results/summary/analyze_escape_profiles.md`.

In many of the visualizations (e.g., Figure 3.2A), the RBD sites are categorized as falling into one of three structural regions (core RBD, RBM, or ACE2-contact residue) and colored accordingly. The RBM is defined as residues 437-508 [180] with remaining residues comprising the core RBD. ACE2 contacts are defined as RBD residues with non-hydrogen atoms

within 4 Angstrom of ACE2 atoms in the PDB: 6M0J crystal structure [172]. In Figures 3.5B and 3.11, the letters in the escape maps are colored according to the effects of mutations on ACE2 binding or RBD expression as measured in [280].

The multidimensional scaling in Figure 3.2D that projects the antibodies into a two-dimensional space of escape mutations was performed using the Python scikit-learn package. We first computed the similarity in the escape maps between each pair of antibodies as follows. Let x_a1 be the vector of the total site escape values at each site for antibody *a1*. Then the similarity in escape between antibodies *a1* and *a2* is simply calculated as the dot product of the total site escape vectors after normalizing each vector to have a Euclidean norm of one; namely, the similarity is:

$$\frac{x_a1}{||x_a1||} \cdot \frac{x_a2}{||x_a2||}$$

With this definition, the similarity is one if the total site escape is identical for the two antibodies, and zero if the escape is at completely distinct sites. We then calculated a dissimilarity for each pair of antibodies as simply one minus the similarity, and performed metric multidimensional scaling with two components on the dissimilarity matrix. The result is shown in Figure 3.2D, with antibodies shown in pie charts that are colored proportional to total squared site escape that falls into that RBD structural region. The code that generates these logo plot visualizations is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/results/summary/mds_escape_profiles.md`.

For the static structural visualizations in the paper figures, the RBD surface (PDB: 6M0J, [172]) was colored by the largest-effect escape mutation at each site, with white indicating no escape and red indicating the strongest escape mutation for that antibody.

We created interactive structure-based visualizations of the escape maps using dms-view [137] that are available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Crowe_antibodies/`. The logo plots in these escape maps can be colored according to the deep mutational scanning measurements of how mutations affect ACE2 binding or RBD

expression as described above.

Analysis of circulating variants and evolutionary conservation of antibody epitopes

All 94,233 spike sequences on GISAID as of 6 September 2020 were downloaded and aligned via mafft [151]. Sequences from non-human origins and sequences containing gap characters were removed, leaving 93,858 sequences. All RBD amino-acid mutations among GISAID sequences were enumerated, retaining only mutations that were sampled on at least one high-coverage sequence lacking undetermined 'X' characters within the RBD. All GISAID mutations at sites of escape from antibodies in our panel (using the method described above to define sites of escape) are shown in Figure 3.10. Counts were collapsed by site, and sites with at least 5 circulating mutations on GISAID are shown in Figure 3.5A. We acknowledge all GISAID contributors for their sharing of sequencing data (`https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/data/GISAID/gisaid_hcov-19_acknowledgement_table_2020_09_06.pdf`).

To compute conservation of positions among sarbecoviruses, we used the RBD sequence set from [280], which includes all unique RBD sequences curated by [179], in addition to the non-Asian sarbecovirus BtKy72 [292] and newly described RBD sequences RaTG13 [352], RmYN02 [352], and GD-Pangolin and GX-Pangolin [171]. RBD sequences were aligned at the amino-acid level via mafft with a gap opening penalty of 4.5. Alignment is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Crowe_antibodies/blob/master/data/RBDs_aligned.fasta`. Shannon entropy of each alignment position was calculated using the bio3d package in R [119] as:

$$h = -\sum_i p_i \times \log_2 p_i$$

where p_i is the proportion of sequences with amino acid i. The effective number of amino acids at each position (N_{eff}) was calculated as 2^h.

Pseudotyped lentiviral particles for neutralization assays and quantification of cellular entry

For neutralization assays, we used spike pseudotyped lentiviral particles that were generated essentially as described in [67], using a codon-optimized SARS-CoV-2 spike from Wuhan-Hu-1 that contains a 21-amino-acid deletion at the end of the cytoplasmic tail that improves viral titers [66] along with the D614G mutation that is now prevalent in human SARS-CoV-2 [161]. The plasmid encoding this spike, HDM_Spikedelta21_D614G, is available from Addgene (#158762), and the full sequence is at `https://www.addgene.org/158762/`. Point mutations were introduced into the RBD of this plasmid via site-directed mutagenesis.

Pseudotyped lentiviral particles were generated as previously described [67]. Viruses were rescued in biological duplicate (i.e., independent transfections). Briefly, 6e5 293T cells per well were seeded in 6-well plates in 2 mL D10 growth media (DMEM with 10% heat-inactivated FBS, 2 mM l-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin). 24h later, cells were transfected using BioT transfection reagent (Bioland Scientific, Paramount, CA, USA) with a Luciferase_IRES_ZsGreen backbone, Gag/Pol lentiviral helper plasmid, and wildtype or mutant SARS-CoV-2 spike plasmids. Media was changed to fresh D10 at 24 h post-transfection. At 60 h post-transfection, viral supernatants were collected, filtered through a 0.45 μm SFCA low protein-binding filter, and stored at -80 degrees C.

The resulting viruses were titered as previously described [67]. 293T-ACE2 cells (BEI NR-52511) were seeded at 1.25e4 cells per well in 50 μL D10 in poly-L-lysine coated 96-well plates (Greiner 655930). After 24 h, 100 μL of diluted viral supernatants were added to cells across a dilution range of 4 serial 4-fold dilutions (i.e., 0.52 to 33.3 μL of virus were ultimately added to each well). Approximately 70 h post-infection, viral entry was quantified Bright-Glo Luciferase Assay System (Promega, E2610) as described in Crawford et al. (2020a). The relative titers reported in Figure 3.9D were calculated as the fold-change of relative luciferase units per microliter of each mutant RBD virus compared to unmutated RBD virus.

For the neutralization assays, the ACE2-293T cells were plated as described above for

viral titering. 24 h later, pseudotyped lentivirus supernatants were diluted 1:6 and incubated with antibodies across a concentration range for 1 h at 37 degrees C, at a final concentration of antibody between 0.366 and 6,000 ng/mL. 100 μL of the virus-antibody mixture then was added to cells.

At approx. 70 h post-infection, luciferase activity was measured as described above. Fraction infectivity of each antibody-containing well was calculated relative to a "no-antibody" well inoculated with the same initial viral supernatant (containing wildtype or mutant RBD) in the same row of the plate. We used the neutcurve package (`https://jbloomlab.github.io/neutcurve/`) to calculate the inhibitory concentration 50% (IC50) of each antibody against each virus by fitting a Hill curve with the bottom fixed at 0 and the top fixed at 1. The IC50 fold-change relative to unmutated RBD was calculated for each mutant for each antibody.

293T mammalian cell-surface RBD display system

To validate the effects of individual mutations on antibody binding to the non-neutralizing antibody CR3022 in a mammalian system as shown in Figure 3.9B,C, the RBD sequence used in yeast display was modified for mammalian surface display to create the HDM_Spike_RBD_B7-1 plasmid described in [190]. Site-directed mutagenesis was used to introduce single amino-acid substitutions into this plasmid.

293T cells were seeded at 6e5 cells per well in a 6-well plate. After 24 h, duplicate wells were transfected with 1 µg HDM_Spike_RBD_B7-1 plasmids and 1 µg of Transfection Carrier DNA (Promega, E4881) using BioT reagent (Bioland Sci, B01-02), according to manufacturer's protocol. At 18 to 20 h post-transfection, cells were washed with phosphate buffered saline (PBS), dissociated from the plate with enzyme-free dissociation buffer (ThermoFisher, 13151014), harvested by centrifugation at 1,200 × g for 3 min, and washed in FACS buffer (PBS+1% bovine serum albumin). Cells were stained with recombinant biotinylated ACE2

(ACROBiosystems, AC2-H82E6) and serial dilutions of CR3022 antibody for 1 h at room temperature, washed with FACS buffer, resuspended in a 1:200 dilution of PE-conjugated streptavidin (ThermoFisher, S866) and APC-conjugated Goat Anti-Human IgG (Jackson Labs, 109-115-098), and incubated on ice for 1 h. Cells were then washed twice in the FACS buffer and resuspended in PBS. CR3022 antibody and ACE2-binding levels were determined via flow cytometry using a BD LSRFortessa X-50. 10,000 cells were analyzed at each CR3022 concentration. Cells were gated to select for singleton events, ACE2 labeling was used to subset RBD+ cells and measure RBD expression, and CR3022 labeling was measured within this RBD+ population. Compensation and gating was performed using FlowJo v10.7. EC50s were computed using the neutcurve package to fit four-parameter Hill curves (both baselines free) and the midpoint is reported as the EC50. The assays were performed on two separate days, and fold changes were computed relative to the unmutated (wildtype) RBD from that day.

Antibody competition-binding analysis

For the competition experiments reported in Figure 3.12D, wells of 384-well microtiter plates were coated with 1 µg/mL of purified SARS-CoV-2 S6Pecto protein at 4°C overnight. Plates were blocked with 2% BSA in DPBS containing 0.05% Tween-20 (DPBS-T) for 1 h. Purified unlabeled antibodies were diluted to 20 µg/mL in blocking buffer, added to the wells (20 μL/well) in triplicate, and incubated for 1 h at ambient temperature. SARS-CoV-2 antibodies were added to each of three wells with the respective antibody at 2.5 μg/mL in a 5 μL/well volume (final 0.1 μg/mL concentration of biotinylated antibody) without washing of unlabeled antibody and then incubated for 1 h at ambient temperature. Plates were washed, and bound antibodies were detected using HRP-conjugated avidin (Sigma) and TMB substrate. The signal obtained for binding of the biotin-labeled reference antibody in the presence of the unlabeled tested antibody was expressed as a percentage of the binding

of the reference antibody alone after subtracting the background signal. Tested antibodies were considered competing if their presence reduced the reference antibody binding to less than 30% of its maximal binding and non-competing if the signal was greater than 75%. A level of 30–75% was considered intermediate competition.

VSV viruses expressing SARS-CoV-2 spike protein

The generation of a replication-competent vesicular stomatitis virus (VSV) expressing SARS-CoV-2 S protein that replaces VSV G protein (VSV-SARS-CoV-2) has been described previously (Case et al., 2020). This virus encodes the spike protein from SARS-CoV-2 with a 21 amino-acid C-terminal deletion. The spike-expressing VSV virus was propagated in MA104 cells (ATCC CRL-2378.1) as described previously (Case et al., 2020), and viral stocks were titrated on Vero E6 cell monolayer cultures. Plaques were visualized using neutral red staining.

Selection of escape mutants using the spike-expressing VSV

To screen for escape mutations selected in the presence of individual antibodies or antibody cocktails, we used a real-time cell analysis assay (RTCA) and xCELLigence RTCA MP Analyzer (ACEA Biosciences Inc.) with modification of previously described assays (Gilchuk et al., 2020a; Weisblum et al., 2020). Fifty (50) μL of cell culture medium (DMEM supplemented with 2% FBS) was added to each well of a 96-well E-plate to obtain a background reading. Eighteen thousand (18,000) Vero E6 cells in 50 μL of cell culture medium were seeded per each well, and plates were placed on the analyzer. Measurements were taken automatically every 15 min and the sensograms were visualized using RTCA software version 2.1.0 (ACEA Biosciences Inc). VSV-SARS-CoV-2 virus (5e3 plaque forming units [PFU] per well, approx. 0.3 MOI) was mixed with a saturating neutralizing concentration of individual antibody (5 μg/mL) or two-antibody cocktail (1:1 antibody ratio, 5 μg/mL total

antibody concentration) in a total volume of 100 μL and incubated for 1 h at 37°C. At 16-20 h after seeding the cells, the virus-antibody mixtures were added into 8 to 96 replicate wells of 96-well E-plates with cell monolayers. Wells containing only virus in the absence of antibody and wells containing only Vero E6 cells in medium were included on each plate as controls. Plates were measured continuously (every 15 min) for 72 h. The escape mutants were identified by delayed CPE in wells containing antibody. To verify escape from antibody selection, isolated viruses were assessed in a subsequent RTCA experiment in the presence of 10 μg/mL of mAb as used for the escape virus selection and a partner mAb recognizing non-overlapping epitope residues (see Figure 3.6A).

Sequence analysis of the gene encoding spike protein from spike protein-expressing VSV escape mutants

To identify escape mutations present in spike protein-expressing VSV antibody-selected escape variants, the escape viruses isolated after RTCA escape screening were propagated in 6-well culture plates with confluent Vero E6 cells in the presence of 10 μg/mL of the corresponding antibody. Viral RNA was isolated using a QiAmp Viral RNA extraction kit (Qiagen) from aliquots of supernatant containing a suspension of the selected virus population. The spike protein gene cDNA was amplified with a SuperScript IV One-Step RT-PCR kit (ThermoFisher Scientific) using primers flanking the S gene. The amplified PCR product (approx. 4,000 bp) was purified using SPRI magnetic beads (Beckman Coulter) at a 1:1 ratio and sequenced by the Sanger sequence technique using primers giving forward and reverse reads of the RBD.

3.6 Notes

Acknowledgements

We thank Keara Malone for experimental assistance, Mike Murphy and Neil King for CR3022 protein, and the Flow Cytometry and Genomics core facilities at the Fred Hutchinson Cancer Research Center for experimental support, especially Dolores Covarrubias, Andy Marty, and MinJae Kim. EM data collections were conducted at the Center for Structural Biology Cryo-EM Facility at Vanderbilt University.

Funding

This work was supported by the NIAID / NIH (R01AI141707 and R01AI12893 to J.D.B., T32AI083203 to A.J.G., and F30AI149928 to K.H.D.C.) and the Gates Foundation. T.N.S. is a Washington Research Foundation Innovation Fellow at the University of Washington Institute for Protein Design and a Howard Huges Medical Institute Fellow of the Damon Runyon Cancer Research Foundation (DRG-2381-19). This work was also supported by Defense Advanced Research Projects Agency (DARPA) grants HR0011-18-2-0001 and HR0011-18-3-0001; NIH contracts 75N93019C00074 and 75N93019C00062; NIH grants U01 AI150739, R01 AI130591 and R35 HL145242, the Dolly Parton COVID-19 Research Fund at Vanderbilt, and a grant from Fast Grants, Mercatus Center, George Mason University. S.J.Z. was supported by NIH T32 AI095202. J.E.C. is a recipient of the 2019 Future Insight Prize from Merck KGaA, which supported this work with a grant. J.D.B. is an Investigator of the Howard Hughes Medical Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the US government or the other sponsors.

Conflicts of Interest

J.E.C. has served as a consultant for Sanofi; is on the Scientific Advisory Boards of CompuVax and Meissa Vaccines; is a recipient of previous unrelated research grants from Moderna and Sanofi; and is a founder of IDBiologics. Vanderbilt University has applied for patents concerning SARS-CoV-2 antibodies analyzed in this work. S.P.J.W. and P.W.R. have filed a disclosure with Washington University for the recombinant VSV. The other authors declare no competing interests.

Author contributions

Conceptualization, A.J.G., T.N.S., S.Z., J.E.C., and J.D.B.; Methodology, A.J.G., T.N.S., S.Z., E.B., P.G., R.S.N., R.E.S., N.S., P.W.R., Z.L., S.P.J.W, R.H.C, and J.D.B.; Investigation, A.J.G., T.N.S., S.Z., P.G., and E.B.; Code, A.J.G., T.N.S., and J.D.B.; Formal Analysis, A.J.G., T.N.S., and J.D.B.; Validation, A.J.G., A.N.L., R.E., and K.H.D.C.; Writing – Original Draft, A.J.G., T.N.S. and J.D.B.; Writing – Review and Editing, all authors; Supervision, J.E.C. and J.D.B.

3.7 Supplementary Figures

Figure 3.7: **FACS gating**. **(A)** Representative hierarchical gates drawn to isolate RBD+ single cells as the parent population for FACS gates in (B, C). First, hierarchical gates were drawn to select single-cell events: forward scatter (FSC) versus side scatter (SSC, top left), SSC width versus height (bottom left), and FSC width versus height (top right). Next, FITC+ labeling of a C-terminal epitope tag on the RBD was used to identify RBD+ cells (purple, bottom right). Selection gates for ACE2+ and antibody-negative sorts (B, C) are nested within this RBD+ population. **(B)** RBD mutant libraries were first sorted for variants that could bind ACE2 with at least 0.01x the affinity of unmutated SARS-CoV-2 RBD, the approximate affinity of RaTG13, the homolog with the lowest affinity that still marginally mediates cell entry (Shang et al., 2020). Top three plots show unmutated SARS-CoV-2 labeled at 0 M, 1e-10 M, and 1e-8 M ACE2. A selection gate was drawn to capture unmutated cells labeled at 1e-10 M ACE2. The bottom two plots show the application of this selection gate to the duplicate RBD mutant libraries labeled at 1e-8 M ACE2. Percentages of RBD+ cells (yellow) in each control and library sample that fall into the ACE2+ sort bin are shown in the upper-right of each FACS plot. These ACE2+ sorted libraries were grown overnight and used for subsequent antibody-escape selections. **(C)** Selection gates for the antibody-escape sorts. Unmutated SARS-CoV-2 RBD was labeled at 400 ng/mL (1x) and 4 ng/mL (0.01x) with each antibody. Antibody-escape selection gates were drawn to capture 0.2% or less of the 1x and up to 95% of the 0.01x antibody-labeled unmutated RBD control cells. Each mutant RBD library was labeled with 400 ng/mL (1x) antibody, and cells that were captured in the "antibody-escape bin" were sorted and their barcodes were sequenced. Percentages of RBD+ cells in each control and library sample that fall into the antibody-escape bin are shown in the bottom-right of each FACS plot.

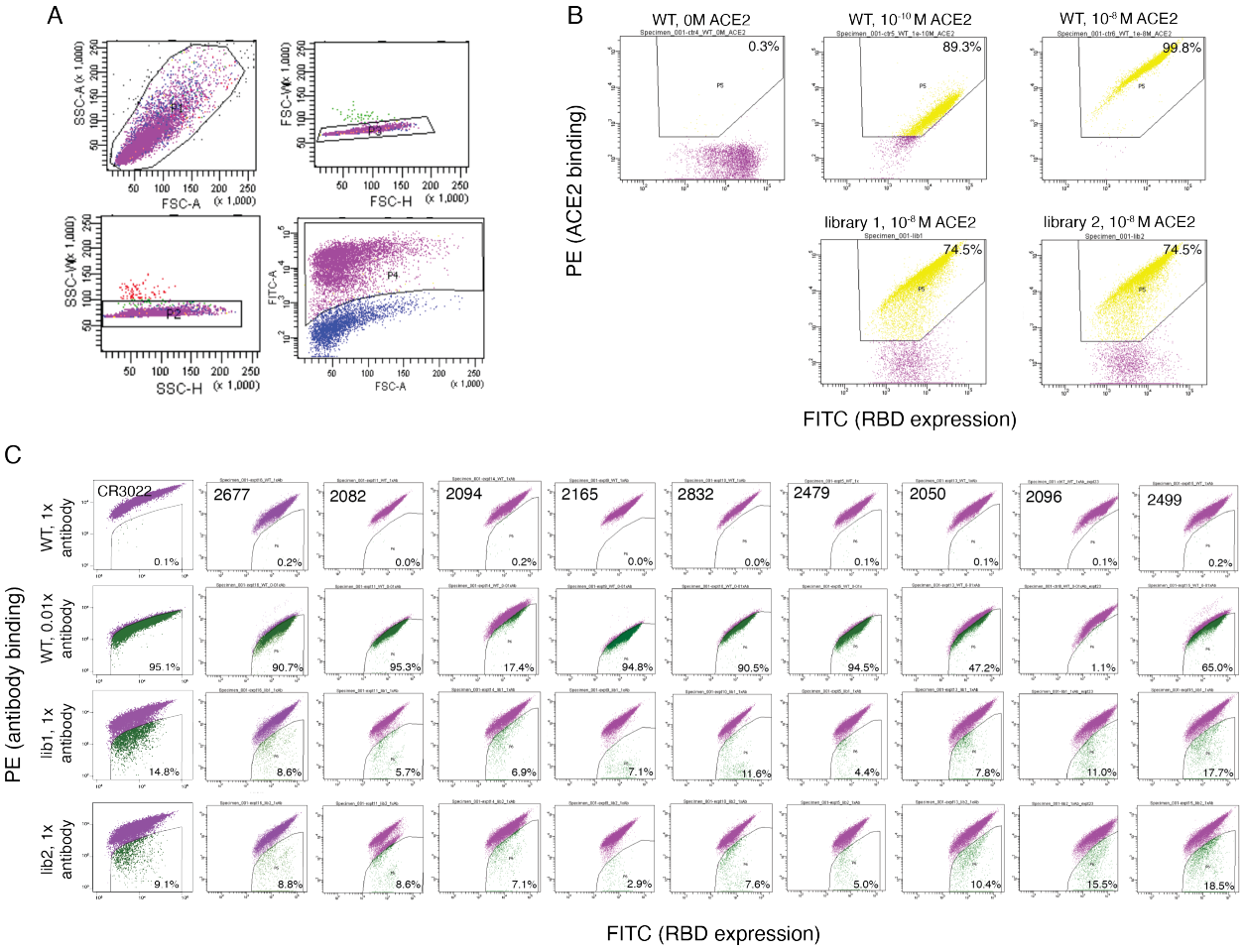

A
SSC-A (x 1,000)
FSC-A (x 1,000)
P1
FSC-H (x 1,000)
P3
SSC-H (x 1,000)
P2
FITC-A
P4
B
WT, 0M ACE2
0.3%
WT, 10-10 M ACE2
89.3%
WT, 10-8 M ACE2
99.8%
library 1, 10-8 M ACE2
74.5%
library 2, 10-8 M ACE2
74.5%
PE (ACE2 binding)
FITC (RBD expression)
C
WT, 1x antibody
WT, 0.01x antibody
lib1, 1x antibody
lib2, 1x antibody
PE (antibody binding)
CR3022
2677
2082
2094
2165
2832
2479
2050
2096
2499
FITC (RBD expression)

Figure 3.8: **Correlation between the duplicate mappings of escape mutations made with the independently generated mutant virus libraries ("lib1" and "lib2").** **(A)** Correlation between the total escape at each site. **(B)** Correlation between the escape fraction measured for each individual mutation. The text insets in each plot give the Pearson's correlation coefficient and the number of sites or mutations for which measurements were made for both libraries. The data shown in the rest of the paper are the average of those from the two libraries. Though there is some variation between replicates in the escape fraction of individual mutations (particularly for antibodies with smaller-magnitude escape) (B), correlation in the site-wise sum escape (A) is reasonable for all antibodies.

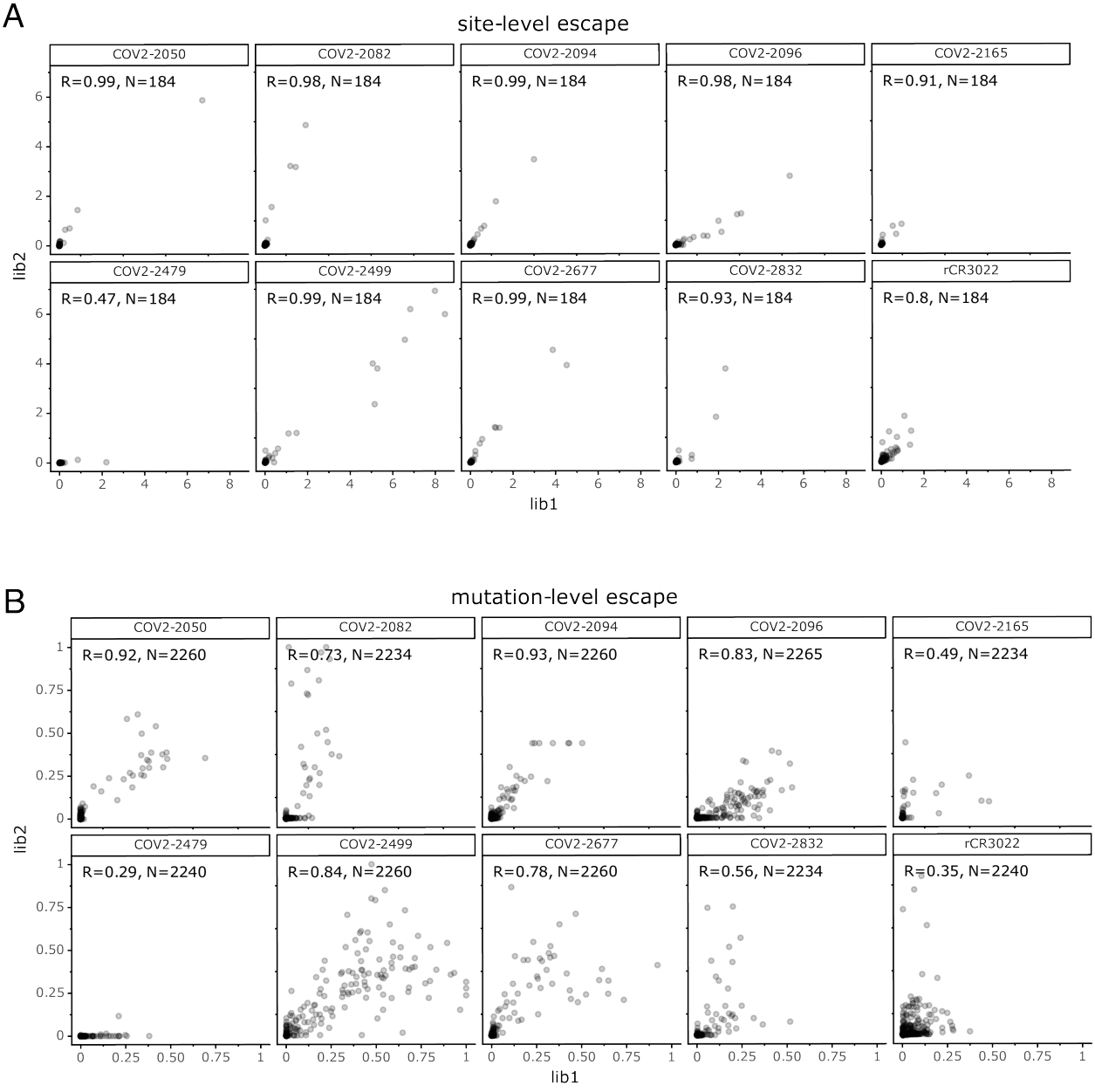

A
site-level escape
COV2-2050
R=0.99, N=184
COV2-2082
R=0.98, N=184
COV2-2094
R=0.99, N=184
COV2-2096
R=0.98, N=184
COV2-2165
R=0.91, N=184
COV2-2479
R=0.47, N=184
COV2-2499
R=0.99, N=184
COV2-2677
R=0.99, N=184
COV2-2832
R=0.93, N=184
rCR3022
R=0.8, N=184
lib2
lib1
B
mutation-level escape
COV2-2050
R=0.92, N=2260
COV2-2082
R=0.73, N=2234
COV2-2094
R=0.93, N=2260
COV2-2096
R=0.83, N=2265
COV2-2165
R=0.49, N=2234
COV2-2479
R=0.29, N=2240
COV2-2499
R=0.84, N=2260
COV2-2677
R=0.78, N=2260
COV2-2832
R=0.56, N=2234
rCR3022
R=0.35, N=2240
lib2
lib1

Figure 3.9: **Validation of the functional and structural relevance of antibody-escape maps**. **(A)** Neutralization curves with the spike-pseudotyped lentiviral particles used to determine IC50 values plotted in Figure 3.3. Each point represents the mean and standard error of 2 independent measurements. The IC50s were computed using the neutcurve package (`https://jbloomlab.github.io/neutcurve/`) to fit two-parameter Hill curves (with the baselines fixed to 0 and 1). IC50s outside the range of tested antibody concentrations are reported as upper bounds. **(B)** Antibody CR3022 is non-neutralizing, so we instead used flow cytometry to measure CR3022 binding to RBD expressed on the surface of mammalian cells (see Methods for details), with the values representing the fold change in effective concentration 50% (EC50) for antibody binding to each mutant RBD relative to wildtype. **(C)** The binding curves summarized in (B), with the y-axis representing binding as measured by flow cytometry. EC50s are computed using the neutcurve package to fit four-parameter Hill curves (both baselines free) and the midpoint is reported as the EC50. The assays were performed on two separate days, and fold changes are computed relative to the unmutated (wildtype) RBD from that day. **(D)** CR3022 escape mutations are compatible with function in spike-pseudotyped lentiviral particles. The infectious titer of spike-pseudotyped lentivirus mutants in transfection supernatants as quantified by fold change in relative luciferase units (RLUs) compared to virus pseudotyped with the unmutated (wildtype) spike. All titers were measured in biological duplicate transfections (two jittered points) except K417E. **(E)** To estimate RBD expression on the surface of 293T cells in the CR3022 binding assays in panels B and C, cells were also labeled with biotinylated ACE2 and fluorophore-conjugated streptavidin. ACE2 binding levels, a proxy for RBD expression, were measured by flow cytometry. Box plots represent the median and 25th and 75th percentiles, whiskers are 1.5 * interquartile range, and outliers are shown individually. For each condition, n=12-24. **(F-J)** Structural analysis of CR3022 escape in the high-resolution CR3022-bound RBD structure (PDB: 6W41 (Yuan et al., 2020)). **(F)** Escape at CR3022-contact residues. RBD residues are colored by total site-wise escape, from white (0 total escape) to red (maximum total escape). CR3022 CDR loops that mediate RBD contacts are shown in blue and labeled. Side chains are shown as sticks and C-alpha spheres are shown for RBD residues defined as CR3022 structural contacts (non-hydrogen atoms within 4 angstrom distance) or sites of strong selection (defined in Methods). RBD sites are labeled by sub-panel of zoomed-in structural views. **(G)** RBD residue K378, most mutations to which mediate CR3022 escape, forms polar contacts with CR3022 residues D54HC and E56HC. **(H)** RBD residue S383, where mutations to bulky charged or hydrophobic residues escape CR3022 binding, forms a polar contact with K386RBD which in turn coordinates D101HC. S383 is also sterically constrained by close packing of the CR3022 CDRH3. **(I)** RBD residues V382 and F392, where mutations that alter side chain volume (V382 and F392) or hydrophobicity (F392) mediate CR3022 escape, pack with hydrophobic residues from the CR3022 CDRL1 and CDRL2 loops. **(J)** RBD residue T430, where mutations do not facilitate strong escape, forms a polar contact with S27FLC at the periphery of the CR3022-RBD interface.

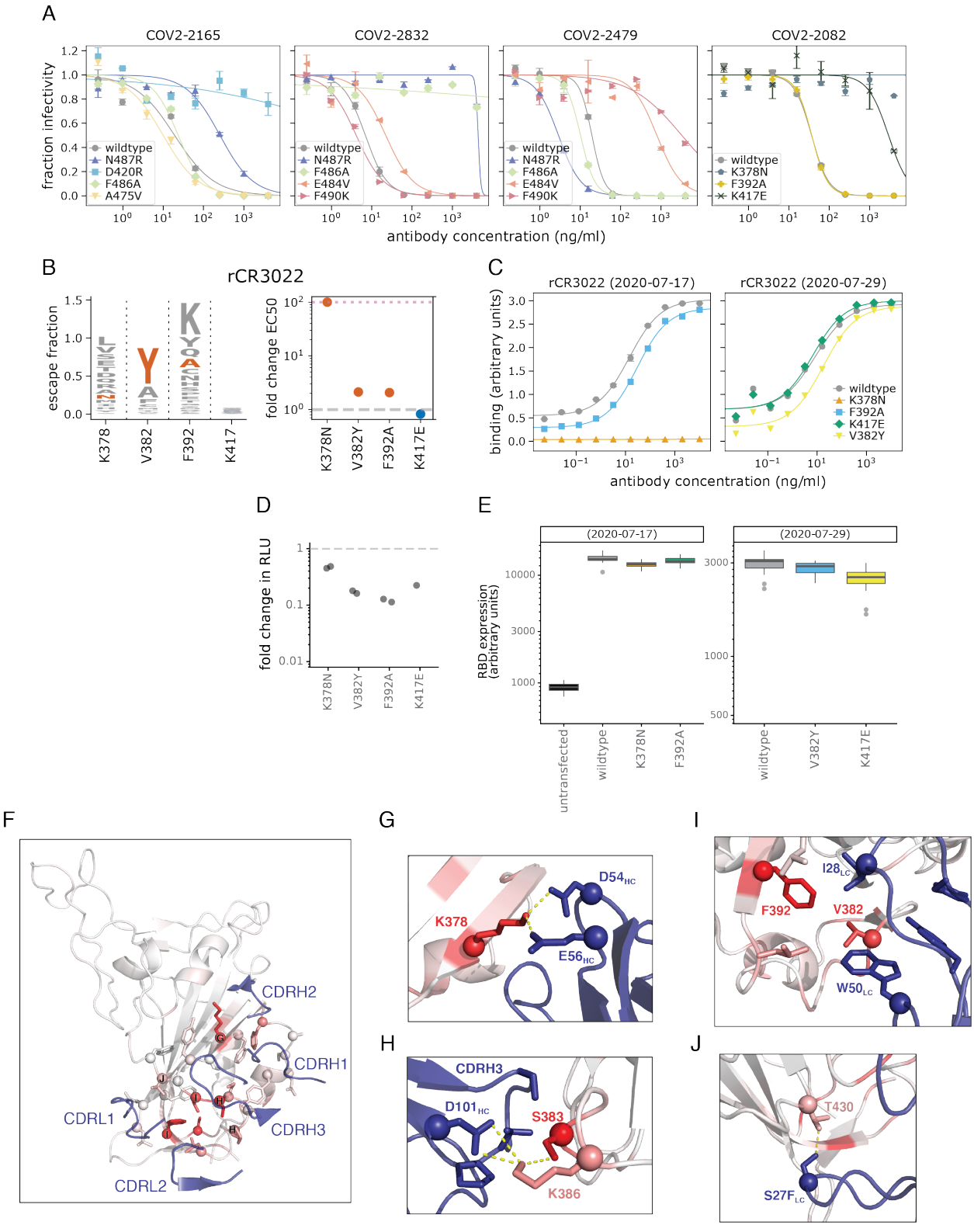

A
COV2-2165
COV2-2832
COV2-2479
COV2-2082
fraction infectivity
antibody concentration (ng/ml)
wildtype
N487R
D420R
F486A
A475V
E484V
F490K
K417E
K378N
F392A
B
rCR3022
escape fraction
fold change EC50
K378
V382
F392
K417
C
rCR3022 (2020-07-17)
rCR3022 (2020-07-29)
binding (arbitrary units)
V382Y
D
fold change in RLU
E
(2020-07-17)
(2020-07-29)
RBD expression (arbitrary units)
untransfected
F
CDRH2
CDRH1
CDRL1
CDRH3
CDRL2
G
D54HC
K378
E56HC
H
CDRH3
D101HC
S383
K386
I
I28LC
F392
V382
W50LC
J
T430
S27FLC

Figure 3.10: **Variation at sites of antibody escape among currently circulating SARS-CoV-2 viruses**. Table shows all RBD mutations sampled among sequences in GISAID as of 6 September 2020 at sites of escape from at least one antibody. Cells are colored by escape fraction of the individual circulating mutant for each antibody: white cells indicate sites that are not sites of escape from an antibody; for sites of escape, per-mutation escape fraction is colored from light to dark gray, with any mutation conferring > 0.1 escape fraction colored equally dark. Sites are in orange for the core RBD, light blue for the RBM, and dark blue for ACE2 contact residues. Antibodies are colored according to where the majority of their sites of escape fall. These per-mutation counts are collapsed into the site-wise table presented in Figure 3.5

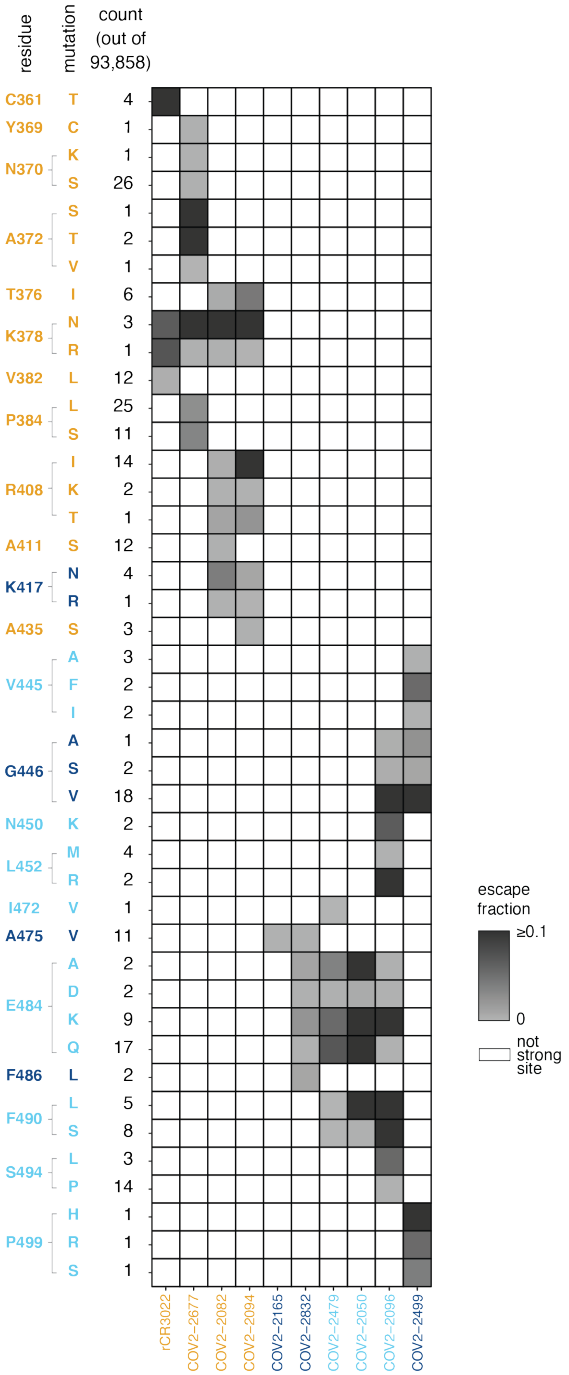

residue
mutation
count (out of 93,858)
C361 T 4
Y369 C 1
N370 K 1
N370 S 26
A372 S 1
A372 T 2
A372 V 1
T376 I 6
K378 N 3
K378 R 1
V382 L 12
P384 L 25
P384 S 11
R408 I 14
R408 K 2
R408 T 1
A411 S 12
K417 N 4
K417 R 1
A435 S 3
V445 A 3
V445 F 2
V445 I 2
G446 A 1
G446 S 2
G446 V 18
N450 K 2
L452 M 4
L452 R 2
I472 V 1
A475 V 11
E484 A 2
E484 D 2
E484 K 9
E484 Q 17
F486 L 2
F490 L 5
F490 S 8
S494 L 3
S494 P 14
P499 H 1
P499 R 1
P499 S 1
rCR3022
COV2−2677
COV2−2082
COV2−2094
COV2−2165
COV2−2832
COV2−2479
COV2−2050
COV2−2096
COV2−2499
escape fraction
≥0.1
0
not strong site

Figure 3.11: **Logo plots of antibody escape accounting for mutation effects on ACE2-binding affinity and RBD folding**. Logo plots as in Figure 3.2C. Mutations are colored according to their effects on ACE2-binding affinity (left) or RBD folding and expression (right), as measured previously (Starr et al., 2020). Some mutations annotated as escape in our main display impair ACE2 binding or RBD folding, which may limit their fitness in the context of virus particles.

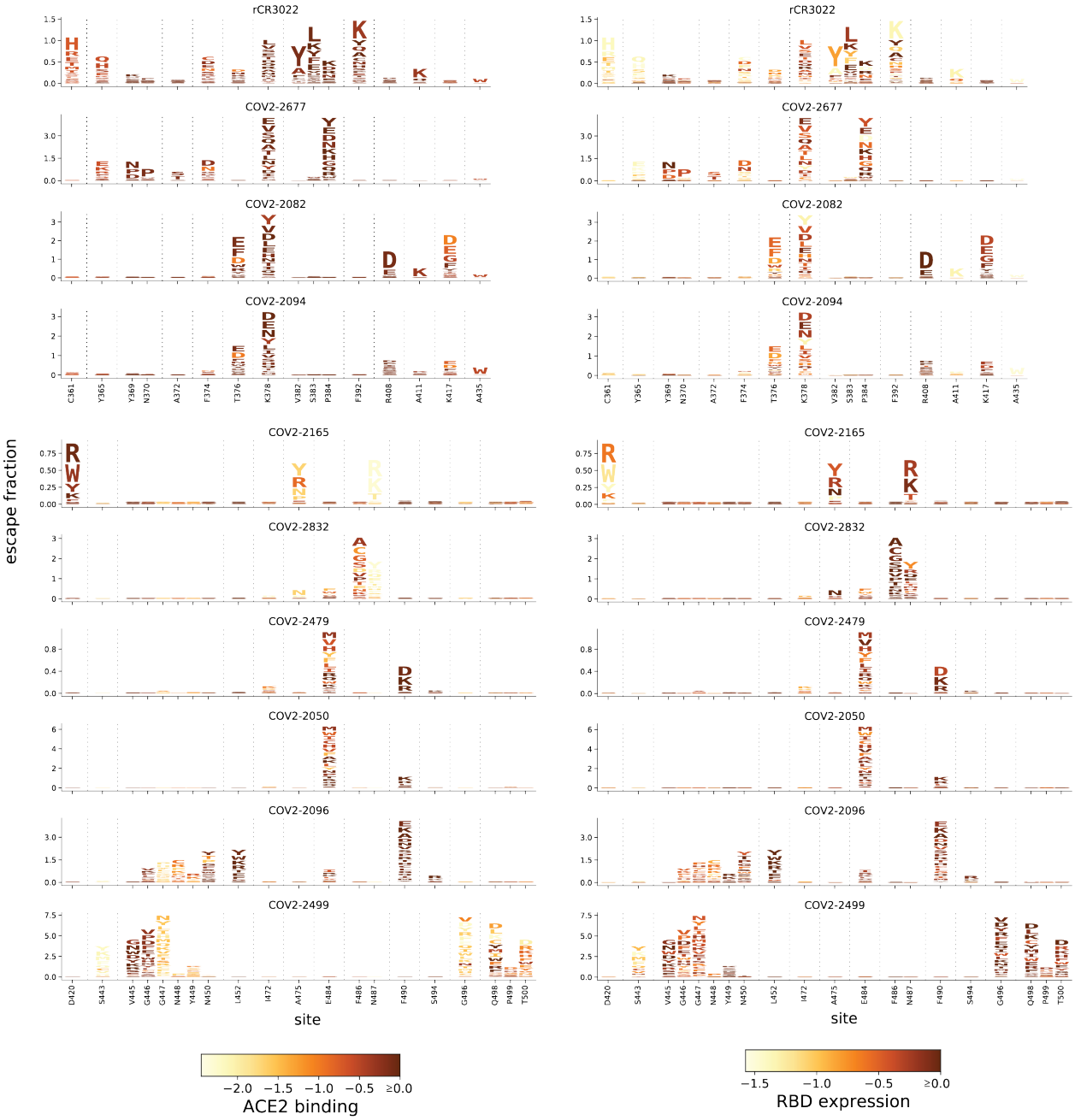
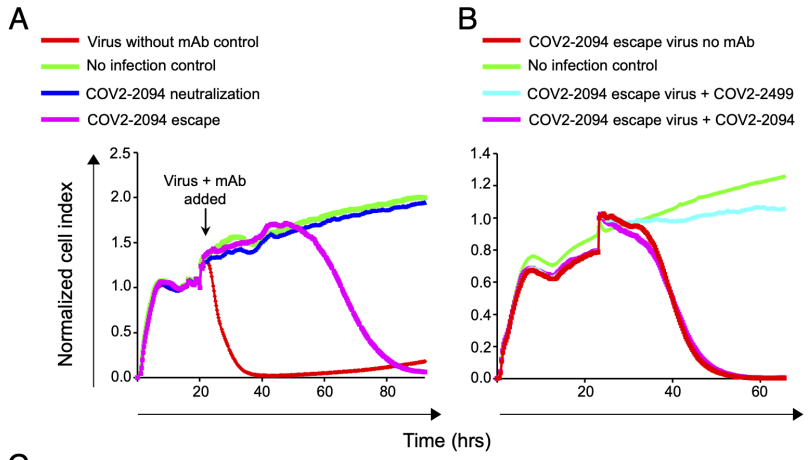

rCR3022
COV2-2677
COV2-2082
COV2-2094
COV2-2165
COV2-2832
COV2-2479
COV2-2050
COV2-2096
COV2-2499
escape fraction
site
C361 Y365 Y369 N370 A372 F374 T376 K378 V382 S383 P384 F392 R408 A411 K417 A435
D420 S443 V445 G446 G447 N448 Y449 N450 L452 I472 A475 E484 F486 N487 F490 S494 G496 Q498 P499 T500
−2.0 −1.5 −1.0 −0.5 ≥0.0
ACE2 binding
−1.5 −1.0 −0.5 ≥0.0
RBD expression

Figure 3.12: **Real-time cell analysis (RTCA) to select for spike-expressing VSV viruses that escape antibody neutralization, and antibody competition for binding to RBD**. **(A)** Representative RTCA sensograms showing virus that escaped antibody neutralization. Cytopathic effect (CPE) was monitored kinetically in Vero E6 cells inoculated with virus in the presence of a saturating concentration of antibody COV2-2094 (5 μg/mL). Escape (magenta) or lack of escape (blue) are shown. Uninfected cells (green) or cells inoculated with virus without antibody (red) serve as controls. Magenta and blue curves represent a single representative well; the red and green controls are mean of technical quadruplicates. **(B)** Representative RTCA sensograms validating that the virus selected by COV2-2094 in panel (A) indeed escaped COV2-2094 (magenta) but was neutralized by COV2-2499 (light blue). **(C)** Example sensograms from individual wells of 96-well E-plate analysis showing viruses that escaped neutralization (noted with *) by indicated antibodies. Escape in the illustrated replicates 6 and 7 for COV2-2499 was confirmed in validation neutralization assays but was not sequence-verified due to delayed CPE and not included in the counts in Figure 3.6. **(D)** Competition assays for RBD binding, with percentages showing binding of a second labeled antibody to the RBD after pre-binding with the first antibody. Values close to 0% indicate complete competition, and values close to 100% indicate lack of competition.



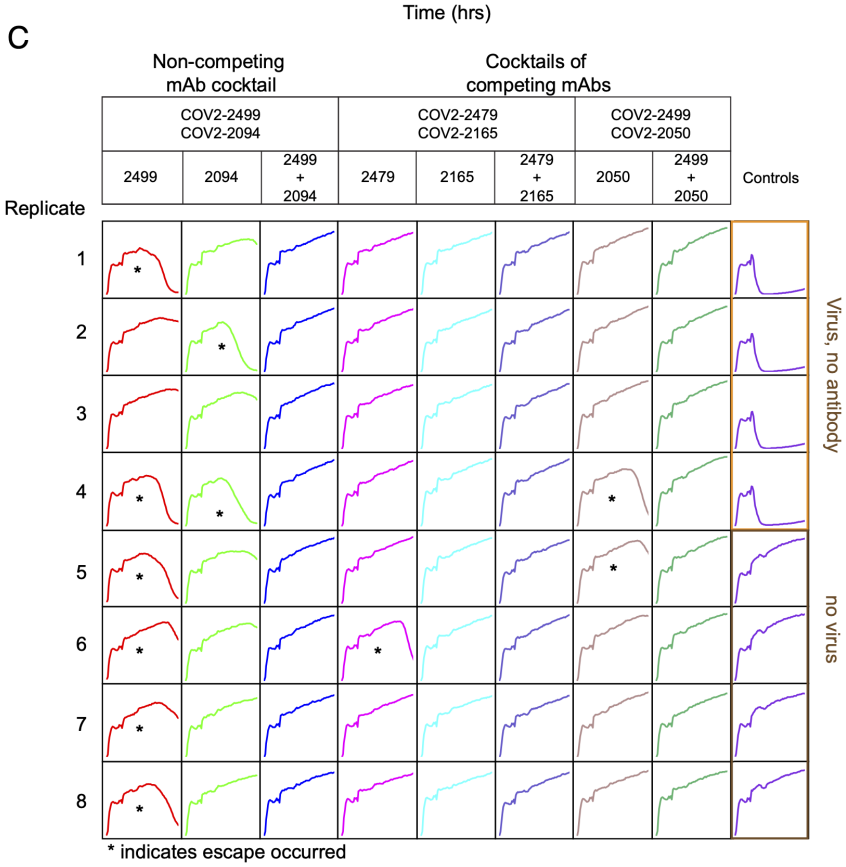
D

	Cocktail 1	
	2499	2094
2499	2%	89%
2094	78%	0%

	Cocktail 2	
	2479	2165
2479	0%	1%
2165	13%	1%

	Cocktail 3	
	2499	2050
2499	2%	36%
2050	0%	-1%

C



Chapter 4

PROSPECTIVE MAPPING OF VIRAL MUTATIONS THAT ESCAPE ANTIBODIES USED TO TREAT COVID-19

A version of this chapter has been previously published as:

Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, Li JZ, Bloom JD. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. Science. 2021 Feb 19;371(6531):850-854. doi: 10.1126/science.abf9302. Epub 2021 Jan 25. PMID: 33495308; PMCID: PMC7963219.

4.1 Abstract

Antibodies are a potential therapy for SARS-CoV-2, but the risk of the virus evolving to escape them remains unclear. Here we map how all mutations to SARS-CoV-2's receptor-binding domain (RBD) affect binding by the antibodies in the REGN-COV2 cocktail and the antibody LY-CoV016. These complete maps uncover a single amino-acid mutation that fully escapes the REGN-COV2 cocktail, which consists of two antibodies targeting distinct structural epitopes. The maps also identify viral mutations that are selected in a persistently infected patient treated with REGN-COV2, as well as during in vitro viral escape selections. Finally, the maps reveal that mutations escaping the individual antibodies are already present in circulating SARS-CoV-2 strains. Overall, these complete escape maps enable interpretation of the consequences of mutations observed during viral surveillance.

4.2 Introduction

Antibodies are being developed as therapeutics to combat SARS-CoV-2 [250]. Antibodies against some other viruses can be rendered ineffective by viral mutations that are selected during treatment of infected patients [41, 167] or that spread globally to confer resistance on entire viral clades [273]. Therefore, determining which SARS-CoV-2 mutations escape key antibodies is essential for assessing how mutations observed during viral surveillance may impact the efficacy of antibody treatments.

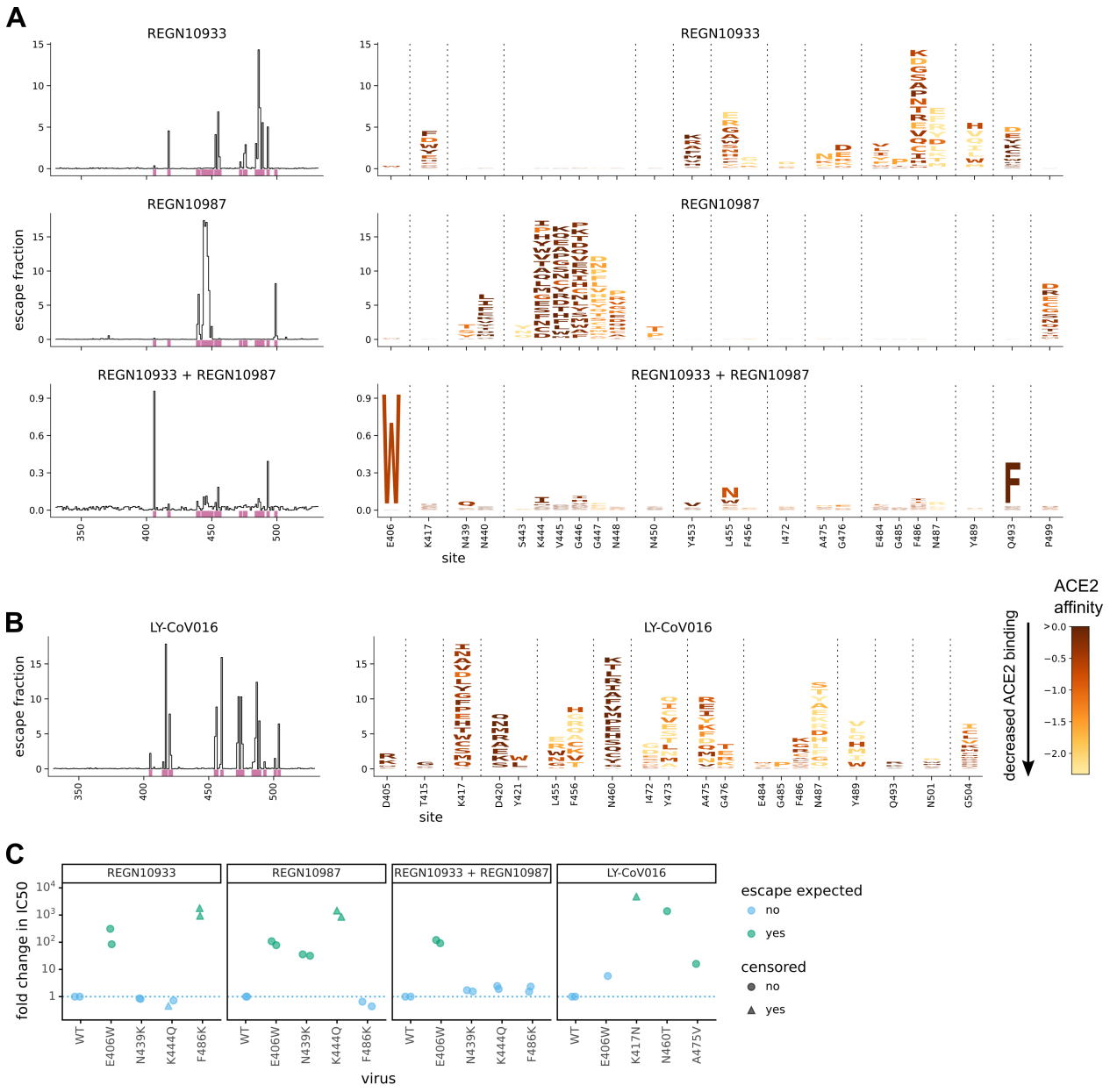
Most leading anti-SARS-CoV-2 antibodies target the viral receptor-binding domain (RBD), which mediates binding to the ACE2 receptor [332, 306]. We recently developed a deep mutational scanning method to map how all mutations to the RBD affect its function and recognition by antiviral antibodies [280, 124]. This method involves creating libraries of RBD mutants, expressing them on the surface of yeast, and using fluorescence-activated cell sorting and deep sequencing to quantify how each mutation affects RBD folding, ACE2 affinity (measured across a titration series), and antibody binding (Fig. 4.5A). Here we used the duplicate mutant libraries described in [280], which consist of barcoded RBD variants that cover 3,804 of the 3,819 possible amino acid mutations. Our libraries were made in the context of the RBD from the early isolate Wuhan-Hu-1, which still represents the most common RBD sequence although several mutants are currently increasing in frequency [244]. We mapped how the 2,034 mutations that do not strongly disrupt RBD folding and ACE binding [280] affected binding by recombinant forms of the two antibodies in Regeneron's REGN-COV2 cocktail (REGN10933 and REGN10987) [129, 19], and Eli Lilly's LY-CoV016 antibody (also known as CB6 or JS016) [271] (Fig. 4.5B). REGN-COV2 was recently granted an emergency use authorization for treatment of COVID-19 [3], while LY-CoV016 is currently in phase 3 clinical trials [4].

4.3 Results

We completely mapped RBD mutations that escape binding by the three individual antibodies as well as the REGN10933 + REGN10987 cocktail (Fig. 4.1A,B and zoomable maps at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_clinical_Abs/`). REGN10933 and REGN10987 are escaped by largely non-overlapping sets of mutations in the RBD's receptor-binding motif (Fig. 4.1A), consistent with structural work showing that these antibodies target distinct epitopes in this motif [129]. But surprisingly, one mutation (E406W) strongly escapes the cocktail of both antibodies (Fig. 4.1A). The escape map for LY-CoV016 also reveals escape mutations at a number of different sites in the RBD (Fig. 4.1B). Although some escape mutations impair the RBD's ability to bind ACE2 or be expressed in properly folded form, many come at little or no cost to these functional properties according to prior deep mutational scanning measurements using yeast-displayed RBD [280] (color gradient in Fig. 4.1A,B indicates loss of ACE2 affinity and in Fig. 4.6 indicates decrease in RBD expression).

To validate the antigenic effects of key mutations, we performed neutralization assays using spike-pseudotyped lentiviral particles, and found concordance between the antibody-binding escape maps and neutralization assays (Fig. 4.1C and Fig. 4.7). As expected from the maps for the REGN-COV2 antibodies, a mutation at site 486 escaped neutralization only by REGN10933, whereas mutations at sites 439 and 444 escaped neutralization only by REGN10987—and so none of these mutations escaped the cocktail. But E406W escaped both individual REGN-COV2 antibodies, and thus also strongly escaped the cocktail. Structural analyses and viral-escape selections led Regeneron to posit that no single amino-acid mutation could escape both antibodies in the cocktail [129, 19], but our complete maps identify E406W as a cocktail escape mutation. Importantly, E406W impacts the REGN-COV2 antibodies in a relatively specific way and does not grossly perturb the function of the RBD, since it only mildly reduces neutralization by LY-CoV016 (Fig. 4.1C) and the titers

Figure 4.1: **Complete maps of mutations that escape binding by the REGN-COV2 antibodies and Ly-CoV016**. **(A)** Maps for antibodies in REGN-COV2. Line plots at left show escape at each site in the RBD (summed effects of all mutations at each site). Sites of strong escape (purple underlines) are shown in logo plots at right. The height of each letter is proportional to how strongly that amino-acid mutation mediates escape, with a per-mutation "escape fraction" of 1 corresponding to complete escape. The y-axis scale is different for each row, so for instance E406W escapes all REGN antibodies but it is most visible for the cocktail as it is swamped out by other sites of escape for the individual antibodies. See `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_clinical_Abs/` for zoomable versions. Letters are colored by how mutations affect the RBD's affinity for ACE2 as measured via yeast display [280], with yellow indicating poor affinity and brown indicating good affinity; see Fig. 4.6A,B for maps colored by how mutations affect expression of folded RBD and Fig. 4.6C,D for distribution of effects on ACE2 affinity and RBD expression across all mutations observed among circulating viral isolates. **(B)** Map for LY-CoV016. **(C)** Validation of key mutations in neutralization assays using spike-pseudotyped lentiviral particles. We chose to validate mutations predicted to have large effects or that are present at high frequency among circulating SARS-CoV-2 isolates (e.g., N439K). Each point indicates the fold-increase in inhibitory concentration 50% (IC50) for a mutation relative to the unmutated "wildtype" (WT) spike, which contains D614G. The dotted blue line at 1 indicates wildtype-like neutralization, and values > 1 indicate increased neutralization resistance. Point colors indicate if escape was expected at that site from the maps. Point shapes indicate that the fold change is censored (an upper or lower limit) due to the IC50 being outside the dilution series used. Most mutants were tested in duplicate, and so have two points. Full neutralization curves are presented in Fig. 4.7.


A
REGN10933
REGN10933
REGN10987
REGN10987
REGN10933 + REGN10987
REGN10933 + REGN10987
escape fraction
site
B
LY-CoV016
LY-CoV016
escape fraction
site
ACE2 affinity
decreased ACE2 binding
C
REGN10933
REGN10987
REGN10933 + REGN10987
LY-CoV016
fold change in IC50
virus
escape expected
no
yes
censored
no
yes

of spike-pseudotyped lentiviral particles (Fig. 4.7F).

To explore if our escape maps are consistent with how the virus evolves under antibody selection, we first examined data from Regeneron's viral escape-selection experiments in which spike-expressing VSV was grown in cell culture in the presence of REGN10933, REGN10987, or the cocktail [19]. That work identified five escape mutations from REGN10933, two from REGN10987, and none from the cocktail (Fig. 4.6A). All seven cell-culture-selected mutations were prominent in our escape maps while also being accessible by just a single-nucleotide change to the wildtype codon in the Wuhan-Hu-1 RBD sequence (Fig. 4.6B), demonstrating concordance between the escape maps and viral evolution under antibody pressure in cell culture. Notably, E406W is not accessible by a single-nucleotide change, which may explain why it was not identified by the Regeneron cocktail selections despite being relatively well tolerated for RBD folding and ACE2 affinity.

To determine if the escape maps could inform analysis of viral evolution in infected humans, we examined deep sequencing data from a persistently infected immunocompromised patient who was treated with REGN-COV2 at day 145 after diagnosis with COVID-19 [59]. The late timing of treatment allowed ample time for the patient's viral population to accumulate genetic diversity, some of which could have been driven by immune pressure since the patient mounted a weak autologous neutralizing antibody response prior to treatment [59]. Administration of REGN-COV2 was followed by rapid changes in the frequencies of five amino-acid mutations in the RBD (Fig. 4.2C and Fig. 4.8). Our escape maps showed that three of these mutations escaped REGN10933, and one escaped REGN10987 (Fig. 4.2B). Notably, the mutations did not all sweep to fixation after antibody treatment: instead, there were competing rises and falls (Fig. 4.2C). This pattern has been observed in the adaptive within-host evolution of other viruses [340, 97], and can arise from genetic hitchhiking and competition among viral lineages. Both these forces appear to be at play in the persistently infected patient (Fig. 4.2C and Fig 4.8C): E484A (not an escape mutation in our

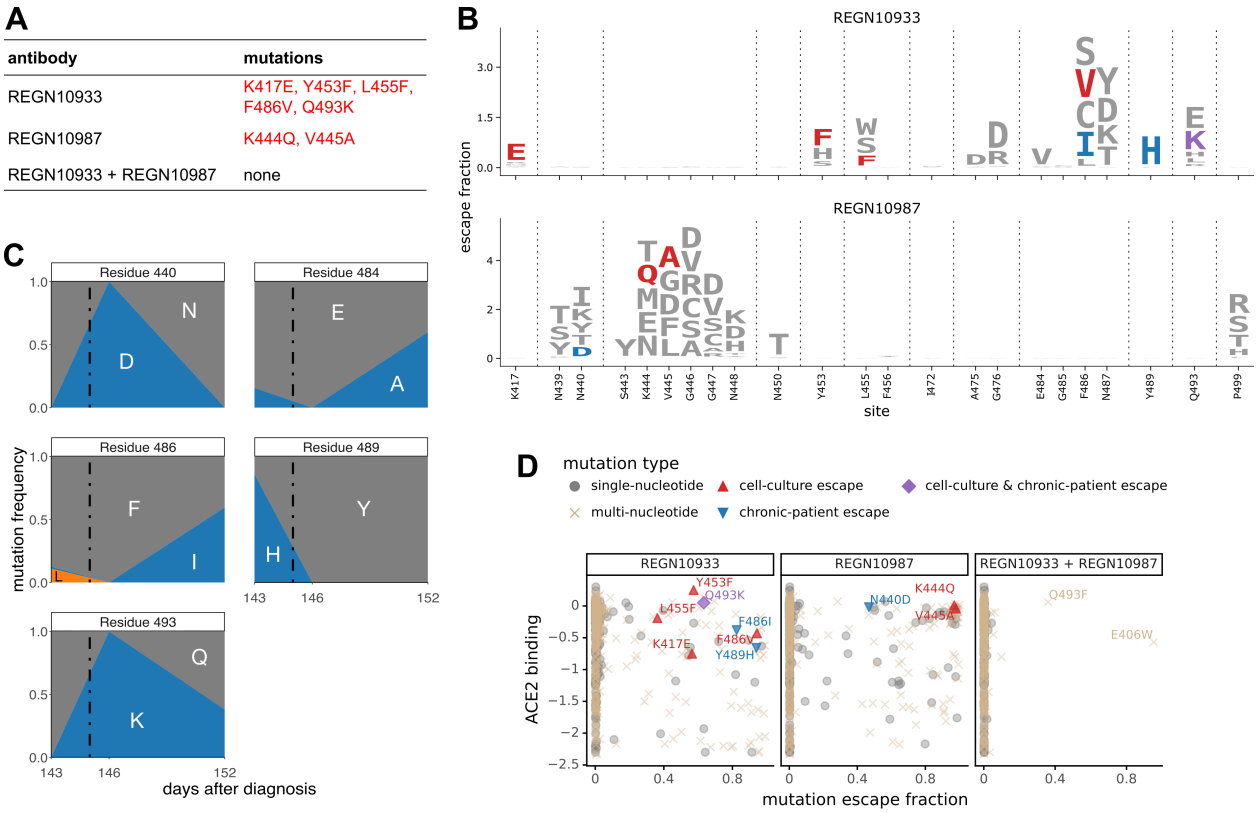


Figure 4.2: **Escape maps are consistent with viral mutations selected in cell culture and a persistently infected patient**. **(A)** Viral escape mutations selected by Regeneron with spike-pseudotyped VSV in cell culture in the presence of antibody [19]. **(B)** Escape maps like those in Fig. 4.1A but showing only mutations accessible by single-nucleotide changes to the Wuhan-Hu-1 sequence, with non-gray colors indicating mutations in cell culture (red), in the infected patient (blue), or both (purple). Fig. 4.9 shows these maps colored by how mutations affect ACE2 affinity or RBD expression. **(C)** Dynamics of RBD mutations in a patient treated with REGN-COV2 at day 145 of his infection (black dashed vertical line). E484A rose in frequency in linkage with F486I, but since E484A is not an escape mutation in our maps it is not shown in other panels. See also Fig. 4.8. **(D)** The escape mutations that arise in cell culture and the infected patient are single-nucleotide accessible and escape antibody binding without imposing a large cost on ACE2 affinity (as measured using yeast display [280]). Each point is a mutation with shape / color indicating whether it is accessible and selected during viral growth. Points further to the right on the x-axis indicate stronger escape from antibody binding; points further up on the y-axis indicate higher ACE2 affinity.

maps) hitchhikes with F486I (which escapes REGN10933) after treatment, and the viral lineage carrying N440D and Q493K (which escape REGN10987 and REGN10933, respectively) competes first with the REGN10933 escape-mutant Y489H, and then with the E484A / F486I lineage and Q493K-alone lineage.

Three of the four escape mutations in the REGN-COV2-treated patient were not identified in Regeneron's viral cell-culture selections (Fig. 4.2B), illustrating an advantage of complete maps. Viral selections are "incomplete" in the sense that they only identify whatever mutations are stochastically selected in that particular cell-culture experiment. In contrast, complete maps annotate all mutations, which could include mutations that arise for reasons unrelated to treatment but incidentally affect antibody binding.

Of course, viral evolution is shaped by functional constraints as well as pressure to evade antibodies. The mutations selected in cell culture and the patient consistently met the following criteria: they escaped antibody binding, were accessible via a single-nucleotide change, and imposed little or no cost on ACE2 affinity (as measured by prior deep mutational scanning using yeast-displayed RBD [280]; Fig. 4.2D, Fig. 4.9). Therefore, complete maps of how mutations affect key biochemical phenotypes of the RBD (e.g., ACE and antibody binding) can be used to assess likely paths of viral evolution. A caveat is that over longer evolutionary timeframes, the space of tolerated mutations could shift due to epistatic interactions, as has been observed in viral immune and drug escape [27, 118, 349].

The complete maps enable us to assess what escape mutations are already present among circulating SARS-CoV-2. We examined all human-derived SARS-CoV-2 sequences available as of January 11, 2021, and found a substantial number of RBD mutations that escaped one or more of the antibodies (Fig. 4.3). However, the only escape mutations present in ¿0.1% of sequences were the REGN10933 escape-mutant Y453F (0.3% of sequences; see also (12)), the REGN10987 escape-mutant N439K (1.7% of sequences; see also Fig. 4.1C and (22)), and the LY-CoV016 escape-mutant K417N (0.1% of sequences; see also Fig. 4.1C).

Y453F is associated with independent outbreaks linked to mink farms in the Netherlands and Denmark [287, 221]; notably the mink sequences themselves sometimes also contain other escape mutations such as F486L [221]. N439K is prevalent in Europe, where it has comprised a large percentage of sequences from regions including Scotland and Ireland [287, 50]. K417N is present in the B.1.351 lineage first identified in South Africa [285]. Another mutation of current interest is N501Y, which is present in B.1.351 and also the B.1.1.7 lineage originally identified in the United Kingdom [244]. Our maps indicate that N501Y has no effect on either REGN-COV2 antibody and only a modest effect on LY-CoV016 (Fig. 4.3).

To determine if the escape maps could be rationalized from the structural interfaces of the antibodies and RBD, we projected the maps onto crystal or cryo-EM structures (Fig. 4.4A; interactive versions at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_clinical_Abs/`). As might be expected, escape mutations generally occur in the antibody-RBD interface. However, structures alone are insufficient to predict which mutations mediate escape. For example, LY-CoV016 uses both its heavy and light chains to bind a wide epitope overlapping the ACE2-binding surface, but escape is dominated by mutations at RBD residues that contact the heavy chain CDRs (Figs. 4.4A). In contrast, escape from REGN10933 and REGN10987 mostly occurs at RBD residues that pack at the antibody heavy/light-chain interface (Fig. 4.4A). The E406W mutation that escapes the REGN-COV2 cocktail occurs at a residue not in contact with either antibody (Fig. 4.4A, B). Although E406 is in closer structural proximity to LY-CoV016 (Figs. 4.4B), the E406W mutation has a much smaller impact on this antibody (Fig. 4.1B,C), suggesting a long-range structural mechanism specific to the REGN-COV2 antibodies. Taken together, mutations at RBD residues that contact antibody do not always mediate escape, and several prominent escape mutations occur at residues not in contact with antibody (Fig. 4.4B).

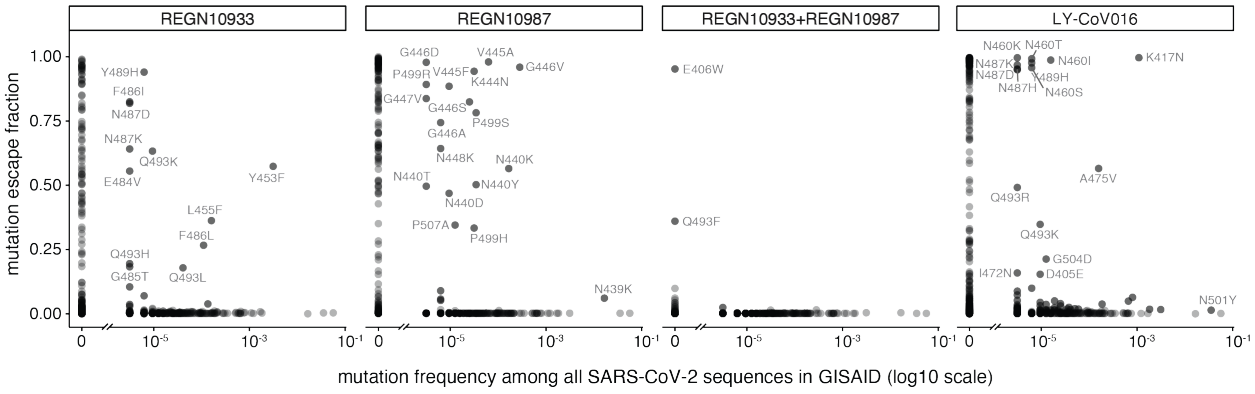


Figure 4.3: **Antibody escape mutations in circulating SARS-CoV-2**. For each antibody or antibody combination, the escape score for each mutation is plotted versus its frequency among the 317,866 high-quality human-derived SARS-CoV-2 sequences on GISAID [91] as of January 11, 2021. Escape mutations with notable GISAID frequencies are labeled. The REGN-COV2 cocktail escape mutation E406W requires multiple nucleotide changes from the Wuhan-Hu-1 RBD sequence and is not observed among sequences in GISAID. Other mutations to residue E406 (E406Q and E406D) are observed with low frequency counts, but neither of these mutant amino acids is a single-nucleotide mutation away from W, either.

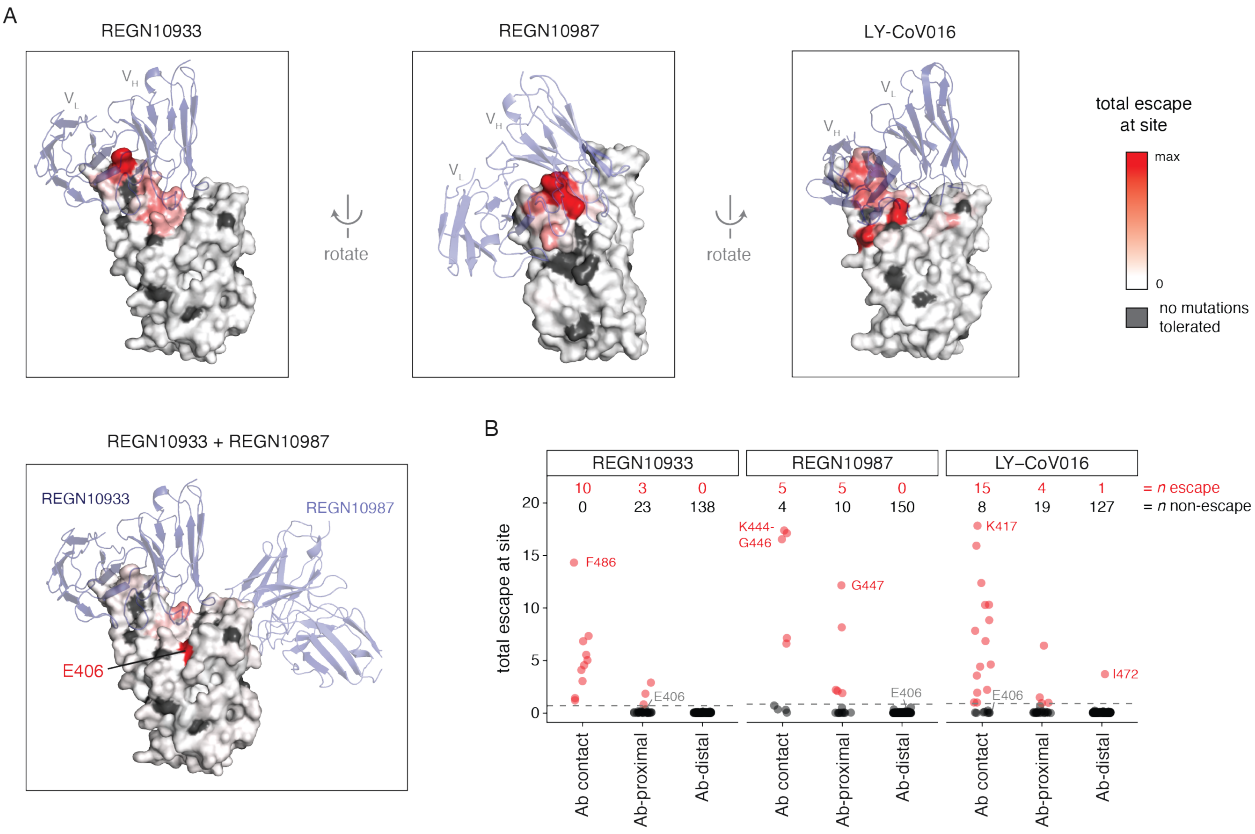


Figure 4.4: **Structural context of escape mutations**. **(A)** Escape maps projected on antibody-bound RBD structures. (REGN10933 and REGN10987: PDB 6XDG [129]; LY-CoV016: PDB 7C01 [271]). Antibody heavy- and light-chain variable domains are shown as blue cartoons, and the RBD surface is colored to indicate how strongly mutations at that site mediate escape (white indicates no escape, red indicates strongest escape site for that antibody / cocktail). Sites where no mutations are functionally tolerated are colored gray. **(B)** For each antibody, sites were classified as direct antibody contacts (non-hydrogen atoms within 4 angstroms of antibody), antibody-proximal ($4-8$ angstroms), or antibody-distal (>8 angstroms). Each point indicates a site, classified as escape (red) or non-escape (black). The dashed gray line indicates the cutoff used to classify sites as escape or non-escape (see Methods for details). Red and black numbers indicate how many sites in each category are escape or non-escape sites, respectively. Interactive visualizations are at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_clinical_Abs/`.

4.4 Discussion

Overall, we have completely mapped mutations that escape three leading anti-SARS-CoV-2 antibodies. These maps demonstrate that prior characterization of escape mutations was incomplete, identifying neither a single amino-acid mutation that escapes both antibodies in the REGN-COV2 cocktail nor most mutations that arose in a persistently infected patient treated with the cocktail. Of course, our maps still do not answer the most pressing question: will SARS-CoV-2 evolve widespread resistance to these antibodies? But certainly, it is concerning that so many escape mutations impose little cost on RBD folding or receptor affinity, and that some are already present at low levels among circulating viruses. Ultimately, it will be necessary to wait and see what mutations spread as SARS-CoV-2 circulates in the human population. Our work will help with the "seeing," by enabling immediate interpretation of the effects of the mutations catalogued by viral genomic surveillance.

4.5 Materials and Methods

Data and Materials Availability

Raw sequencing data are on the NCBI SRA under BioProject PRJNA639956 / BioSample SAMN16850904 (escape mapping) and Bioproject PRJNA681234 (patient sequencing). Computer code and processed data for the escape mapping are at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_clinical_Abs`. Code and data for the patient sequencing are at `https://github.com/jbloomlab/SARS-CoV-2_chronic-infection-seq`. The sequences of the antibodies are provided via citations in the Material and Methods, the cells and plasmids for the neutralization assays are available in BEI Resources and AddGene (see Materials and Methods for details), and the yeast mutant libraries are available to academic researchers upon request with a completed Materials Transfer Agreement.

Antibodies

Publicly available antibody variable domain sequences were acquired for REGN10933, REGN-10987, and LY-CoV016 (also known as JS016, LY3832479, or CB6). Specifically, REGN10933 and REGN10987 variable domain sequences were reported by Hansen et al. [129] in supplemental Data S1. LY-CoV016 (CB6) sequence was reported by Shi et al. [271], Genbank Accessions MT470196 and MT470197.

Recombinant antibodies were cloned and produced by Genscript. Specifically, antibody variable domains were cloned with the human IgG1 heavy chain and human IgK (REGN10933 and LY-CoV016) or human IgL2 (REGN10987) constant regions into pcDNA3.4 vector, and transfected into HD 293F cells maintained at 37°C with 8% CO2 on an orbital shaker. Cell culture supernatants were collected, and affinity purified over RoboColumn Eshmuno A 0.6mL columns.

Antibody-escape mapping

Antibody selection experiments were performed in biological duplicate using a deep mutational scanning (mutational antigenic profiling) approach [124] using previously described duplicate mutant RBD libraries within a yeast-surface display vector [280]. These libraries were generated via a codon tiling PCR mutagenesis approach, which introduced an average of 2.7 amino acid mutations per library variant. RBD variants were linked to unique 16-nucleotide barcode sequences to facilitate downstream sequencing and bottlenecked to library sizes of about 100,000 uniquely barcoded variants. The libraries contain virtually all possible amino-acid mutations to the SARS-CoV-2 RBD, with ¿95% of possible RBD amino acid mutations present as a single mutation on at least one barcode in the libraries. We previously used these libraries together with FACS and deep sequencing to measure the effect of all RBD mutations on yeast-surface expression levels and ACE2 affinity across an ACE2 titration series [280]. As described in [124], these libraries were sorted to eliminate

variants that lose ACE2 binding prior to mapping the antibody-escape variants.

Antibody labeling and selection was performed essentially as described in [124]. Specifically, 9 OD aliquots of RBD libraries were thawed and grown overnight at 30°C 275 rpm in 45mL SD-CAA (6.7 g/L Yeast Nitrogen Base, 5.0 g/L Casamino acids, 1.065 g/L MES, and 2% w/v dextrose). Libraries were back-diluted to an OD of 0.67 in SG-CAA+0.1% dextrose (SD-CAA with 2% w/v galactose and 0.1% w/v dextrose in place of 2% dextrose), and incubated for 16-18 hours at room temperature with mild agitation to induce RBD surface expression. For each antibody selection, 20 OD units of induced cells were washed twice with PBS-BSA (0.2 mg/mL), and incubated in 4mL PBS-BSA with 400 ng/mL antibody (monoclonal REGN10933, REGN10987, LY-CoV016, or REGN10933+REGN10987 pooled at 1:1 w/w ratio at total 400 ng/mL) for 1 h at room temperature with gentle agitation. Labeled cells were washed with ice-cold PBS-BSA followed by secondary labeling for 1 h at 4°C in 2.5 mL 1:200 PE-conjugated goat anti-human-IgG (Jackson ImmunoResearch 109-115-098) to label for bound antibody, and 1:100 FITC-conjugated anti-Myc (Immunology Consultants Lab, CYMC-45F) to label for RBD surface expression. Labeled cells were washed twice with PBS-BSA and resuspended in 2.5mL PBS. Yeast expressing the unmutated SARS-CoV-2 RBD were prepared in parallel to library samples, labeled at the same 400 ng/mL and 100x reduced 4 ng/mL antibody concentrations.

Antibody-escape cells were selected via fluorescence-activated cell sorting (FACS) on a BD FACSAria II. FACS selection gates were drawn to capture 95% of yeast expressing unmutated SARS-CoV-2 RBD labeled at 4 ng/mL antibody (100x reduced antibody concentration relative to library samples, see Figure 4.5B,C). For each library sample, approximately 6-8 million RBD+ cells were processed on the cytometer, with between 5.9e5 and 1.9e6 antibody-escaped cells collected per sample into SD-CAA supplemented with 1% w/v BSA (see selection percentages in Figure 4.5C). Antibody-escaped cells were grown overnight in 1.5mL SD-CAA + 100 U/mL penicillin + 100 μg/mL streptomycin at 30°C 275 rpm.

Plasmid samples were prepared from up to 7.5 OD units of overnight cultures of antibody-escaped cells, and 30 OD units of pre-selection yeast populations (Zymoprep Yeast Plasmid Miniprep II) per manufacturer instructions, with the addition of a -80°C freeze-thaw step prior to cell lysis. The 16-nucleotide barcode sequences identifying each RBD variant were amplified by PCR and prepared for Illumina sequencing exactly as described by Starr et al. [280]. Barcodes were sequenced via 50 bp single-end reads on an Illumina HiSeq 2500, targeting at least 3x as many sequencing reads as FACS-selected cells, and pre-sort reference populations of at least 2.5e7 reads.

Analysis of mutant library deep sequencing and computation of per-mutant escape fractions

Escape fractions were computed as described in [124], with minor modifications as noted below. Specifically, we used the dms_variants package (`https://jbloomlab.github.io/dms_variants/`, version 0.8.2) to process Illumina sequences into counts of each barcoded RBD variant in each pre-sort and antibody-escape population using the barcode/RBD look-up table from [280]. Markdown renderings of these steps in the computational analysis are at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_clinical_Abs/blob/main/results/summary/aggregate_variant_counts.md` and `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_clinical_Abs/blob/main/results/summary/counts_to_cells_ratio.md`.

For each antibody selection, we then computed the "escape fraction" for each barcoded variant using the deep sequencing counts for each variant in the original and antibody-escape populations and the total fraction of the library that escaped antibody binding via the formula provided in Greaney et al [124]. These escape fractions represent the estimated fraction of cells expressing that specific variant that fall in the antibody escape bin, so a value of 0 means the variant is always bound by antibody and a value of 1 means that it always escapes antibody binding. Because our escape bin captures cells labeled at 0.01x the concentration of the library selection, this means that an escape fraction of 1 indicates a

mutant that decreases antibody binding by 100x or greater. We then applied a computational filter to remove variants with low sequencing counts or highly deleterious mutations that might cause antibody escape simply by leading to poor expression of properly folded RBD on the yeast cell surface. Specifically, we ignored all variants with pre-selection sequencing counts that were lower than the counts for the 99th percentile of the stop-codon containing variants–the logic here being that stop codon variants are largely purged by the earlier sorts for RBD expressing and ACE2-binding variants and so any residual presence provides an indication of low-count "noise." Next, we removed any variants that had poor RBD expression or ACE2 binding, or contained mutations that individually cause poor RBD expression and ACE2 binding, the logic being that this would eliminate misfolded or non-expressing RBDs. Specifically, we removed variants that had (or contained mutations with) ACE2 binding scores < -2.35 or expression scores < -1, using the variant- and mutation-level deep mutational scanning scores from Starr et al [280]. Note that these filtering criteria are slightly more stringent than those used in Greaney et al [124]. The ACE2 binding cutoff of -2.35 is used to represent the binding of RaTG13 to human ACE2 [280], which possesses the lowest known affinity capable of mediating cell entry (29). The RBD expression cutoff of -1 is used to eliminate mutations that have as large an expression deficit as mutations to core disulfide residues. 2,034 of the 3,819 possible RBD amino acid mutations passed these filtering steps and were included in our escape maps. A markdown rendering of the computation of the variant-level escape fractions and the variant filtering is at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_clinical_Abs/blob/main/results/summary/counts_to_scores.md`.

Because some library variants contain multiple amino acid mutations, we next deconvolved variant-level escape scores into escape fraction estimates for single mutations using global epistasis models [220] implemented in the dms_variants package, as detailed at `https://jbloomlab.github.io/dms_variants/dms_variants.globalepistasis.html`.

In this fitting, we excluded variants that contained mutations that were not seen as either single mutants or in at least two multiple-mutant variants. We then computed the estimated effect of each mutation as the impact of that mutation on the "observed phenotype" scale transformation of its "latent phenotype" as computed using the global epistasis models, and applied a floor of zero and a ceiling of 1 to these escape fractions. All of the above analysis steps were performed separately for each of the duplicate mutant libraries. We then only retained mutations that passed all of the above filtering and were measured in both libraries or had at least two-single mutant variant measurements in one library. The reported scores throughout the paper are the average across the libraries. Correlations in final single-mutant escape scores are shown in Figure 4.51D. A markdown rendering of the computation that computes these mutation-level escape fractions is at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_clinical_Abs/blob/main/results/summary/scores_to_frac_escape.md`.

For plotting and analyses that required identifying RBD sites of "strong escape" (e.g., choosing which sites to show in logo plots in Fig 4.1A,B or label in Figure 4.4B), we considered a site to mediate strong escape if the total escape (sum of mutation-level escape fractions) for that site exceeded the median across sites by $>$ 5-fold, and was at least 5% of the maximum for any site. A markdown rendering of the identification of these sites of strong escape is at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_clinical_Abs/blob/main/results/summary/call_strong_escape_sites.md`.

Pseudotyped lentiviral particle neutralization assays

We performed neutralization assays by infecting 293T-ACE2 cells with lentiviral particles carrying the luciferase gene and pseudotyped with the SARS-CoV-2 spike essentially as described in Crawford et al [67] with the following two modifications: we introduced into the Wuhan-Hu-1 spike protein sequence a deletion of the final 21 amino acids in the cytoplasmic tail (which increases viral titers (32)) and the D614G mutation (which further increases

viral titers and makes the sequence better match currently circulating viruses [124]). The spike plasmid used for these experiments, HDM-SARS2-spike-del21-D614G, is available on AddGene as plasmid #158762 (`https://www.addgene.org/158762/`). Point mutants described in this paper were introduced into that plasmid. The 293T-ACE2 cells and the other lentiviral plasmids are available from BEI Resources at NR-52511,NR-52517, NR-52518, NR-52519, and NR-52520. Note that as described in Crawford et al [67] the virions were incubated with antibody for 60 minutes, and then 100 ul of the virus + antibody mix was added to 50 ul of 293T-ACE2. The antibody concentrations reported in Fig. 4.7 and used to compute the IC50s are the concentrations in the virus + antibody mix prior to adding to the target cells.

Deep-sequencing analysis of within-host viral genetic diversity in persistently infected patient

The persistently infected patient and his clinical time course are described in detail in [59]. That paper also describes the Illumina deep sequencing of that patient at nine time points (days 18, 25, 78, 81, 128, 130, 143, 146, and 152 after the initial positive PCR test). All sequencing is from nasal swab samples. The deep sequencing data have been deposited on the Sequence Read Archive under BioProject accession PRJNA681234.

Intra-patient single-nucleotide polymorphisms (SNPs) were identified with an automated variant-calling pipeline (`https://github.com/jbloomlab/SARS-CoV-2_chronic-infection-seq`) created with Snakemake [162]. Briefly, paired-end reads were filtered, and sequencing adaptors were removed with fastp [52]. Reads from SARS-CoV-2 were enriched by kmer matching to the Wuhan-Hu-1 reference genome (NC_045512.2) using BBDuk (`https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/`). Following filtering, reads were aligned to the Wuhan-Hu-1 reference with BWA-MEM [181]. Variants were identified by counting the coverage of each base at every position in the reference genome using a custom Python script. These variants were filtered based on a minimum allele

frequency of > 0.01, a PHRED quality threshold of > 25, and coverage of more than 100 reads. The coverage pattern over the spike gene was plotted by averaging the number of reads over every base meeting the minimum PHRED score of 25 in 10 bp bins (Fig. 4.8A).

To visualize the change in allele frequencies over time (Fig. 4.2C & 4.8B), we identified sites in the spike gene with nonsynonymous mutations that rose above 10% frequency at any sampled time point (note that we ignore any mutations relative to Wuhan-Hu-1 that are fixed at all time points as these are not intra-host variants). Using this list of high-confidence polymorphisms, we selected any other nonsynonymous mutations annotated at those sites, regardless of frequency, to get a full picture of allelic variation in putatively selected residues. For the analysis of just the RBD mutations between days 143 and 152 (Fig. 4.2C), we excluded any mutations that were either fixed or absent over the timeframe of interest (T478K, S494P, and N501Y).

To phase the variant alleles in the RBD (Fig. 4.8C) over the last three time points, we used a custom Python script that counted the co-occurrence of nonsynonymous variants in read-pairs. To maximize the number of informative reads for each time point, we only required that reads cover segregating sites in each time point based on analysis of the mutation frequencies in Fig. 4.2C. In other words, for the day 143 sample, we required reads to cover sites 484, 486, and 489, but not sites 440 or 493. For the day 146 sample, only one haplotype was possible (N440D/Q493K); thus, its frequency was assumed to be 100%. Finally, for the day 152 sample, we required reads to cover sites 484, 486, 489, and 493, but not site 440. Of these informative reads, those with SAM flags indicating quality failure or secondary mapping were excluded. To estimate the frequency of the identified haplotypes, we divided each haplotype's count by the total number of unique haplotypes at each time point. Despite the lower number of supporting reads for each haplotype than for individual variants (527 reads for Day 143; 732 reads for Day 146; 1091 reads for Day 152), each haplotype's frequencies were consistent with the frequencies of the individual variants of which they were comprised. Finally, we

filtered out any haplotypes present at a frequency of less than 0.01.

Analysis of mutations in circulating human SARS-CoV-2 strains

For the analysis in Fig. 4.3, all 343,236 spike sequences on GISAID [91] as of 11-January-2021 were downloaded and aligned via mafft [151]. Sequences from non-human origins and sequences containing gap or ambiguous characters were removed, as were sequences with extremely high numbers of RBD mutations relative to other sequences, leaving 317,866 retained sequences. All RBD amino-acid mutations were enumerated compared to the reference Wuhan-Hu-1 SARS-CoV-2 RBD sequence (Genbank MN908947, residues N331-T531). To explore the prevalence of mutations such as Y453F, N439K, and K417N with finer-scale geographic resolution, we used the COVID-19 CG resource (covidcg.org) [50]. We acknowledge all contributors to the GISAID EpiCoV database for their sharing of sequence data (all contributors listed at: `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_clinical_Abs/blob/main/data/gisaid_hcov-19_acknowledgement_table_2021_01_11.pdf`).

Data visualization

The static logo plots in the paper were created using dmslogo (`https://jbloomlab.github.io/dmslogo/`) version 0.5.0. Coloring of mutations by prior deep mutational scanning measurements of yeast-displayed RBD ACE2 affinity and RBD expression used the data from [280].

The interactive visualizations of the escape maps and their projections on the RBD-antibody structures available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_clinical_Abs/` were created using dms-view (`https://dms-view.github.io/docs/`) [137].

The static structural views in the paper were rendered in PyMOL using antibody-bound RBD structures PDB 6XDG [129] and PDB 7C01 [271]. The E406W mutation was modelled into the 6XDG PDB using PyMOL without further structural modification. Structural

distances were computed using the bio3d package in R [119].

4.6 Notes

Acknowledgments

We thank Katharine Crawford for help with neutralization assays, Alison Feder, David Veesler, Neil King, and Daniel Ellis for helpful comments, and the Fred Hutch Flow Cytometry and Genomics facilities for assistance.

Funding

This work was supported by the NIAID (R01AI127893 and R01AI141707 to JDB and T32AI083203 to AJG), the Gates Foundation (INV-004949 to JDB), and the Massachusetts Consortium for Pathogen Readiness through grants from the Evergrande Fund (to JZL). Scientific computing at the Fred Hutch is supported by ORIP grant S10OD028685. TNS is a Washington Research Foundation Innovation Fellow at the University of Washington Institute for Protein Design and an HHMI Fellow of the Damon Runyon Cancer Research Foundation (DRG-2381-19). JDB is an Investigator of the Howard Hughes Medical Institute.

Author contributions:

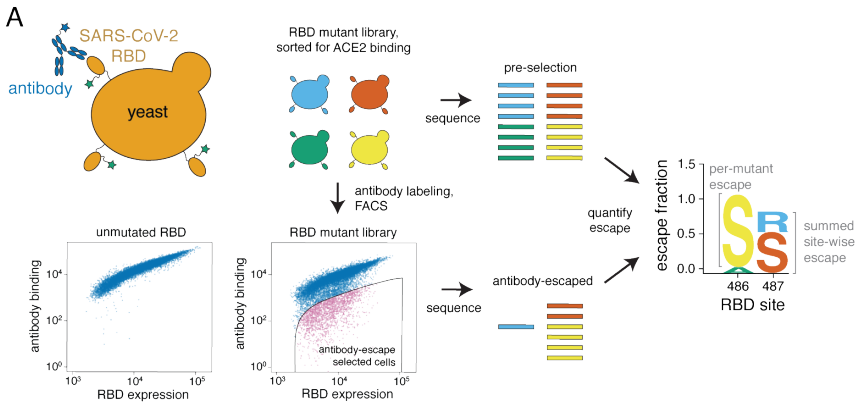
TNS, AJG, ASD, and JDB designed the study. TNS, AJG, AA, and ASD performed the experiments. TNS, AJG, and JDB analyzed the experimental data. JZL and MCC sequenced the persistent infection, and WWH analyzed that data. TNS, AA, WWH, and JDB wrote the initial draft, and all authors edited the final version.

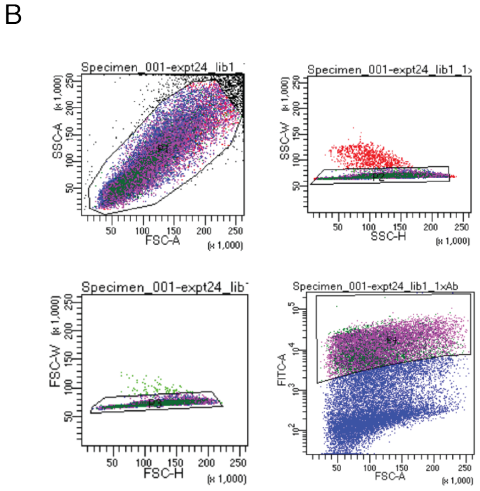
Competing Interests

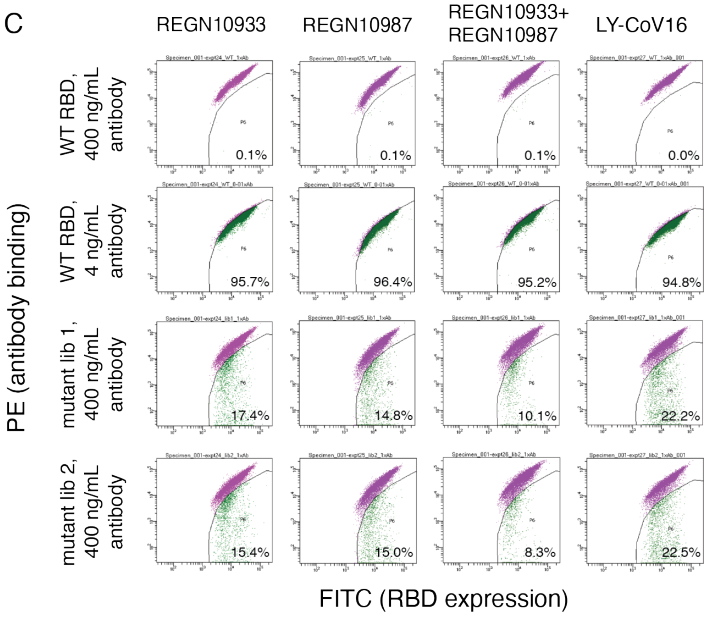
JZL has consulted for Abbvie and Jan Biotech. The other authors declare no competing interests.

4.7 Supplemental Figures

Figure 4.5: **Deep mutational scanning method to map antibody-escape mutations**. **(A)** Experimental approach to map antibody-escape mutations [124]. SARS-CoV-2 RBD is expressed on the yeast cell surface [280], where fluorescent labeling detects RBD surface expression and antibody binding. A library of the SARS-CoV-2 RBD variants, previously sorted to purge non-functional variants [124], is labeled with antibody. Individual yeast cells expressing antibody-escape RBD variants are isolated via fluorescence-activated cell sorting (FACS). Deep sequencing quantifies variant frequencies before and after FACS, enabling the calculation of an "escape fraction" for each RBD mutation, which describes the fraction of cells containing a mutation that fall into the antibody-escape FACS bin. Escape fractions are illustrated in logo plots, where the height of a letter indicates the escape fraction for an individual mutation, and the sum of letter heights at a position indicates the total escape at a site. **(B)** Representative FACS gates used to select single yeast cells (nested SSC/FSC, SSC-W/SSC-H, and FSC-W/FSC-H gates) that express RBD on the cell surface (FITC/FSC). **(C)** Among RBD+ cells, antibody-escape bins were drawn on antibody-binding versus RBD expression scatterplots, with gate stringency determined from unmutated RBD controls. Antibody-escape sort gates were drawn to capture approximately 95% of cells expressing unmutated SARS-CoV-2 RBD when labeled at 0.01x the concentration of antibody used to label mutant libraries, such that an escape fraction of 1 indicates a 100-fold or greater decrease in antibody binding. The percentage of cells that fall in the antibody-escape bin in controls and independent library replicates are shown. **(D)** Correlations in deep mutational scanning scores between independent library duplicates. For each antibody, the escape fraction of individual mutations (top) and total escape per site (bottom) is shown for two independently generated and assayed mutant libraries. R, Pearson correlation coefficient. N, number of mutations or sites. Virtually all of the 3,819 possible RBD mutations are present in our libraries, but mutations that completely disrupt folding or binding are purged prior to antibody selections (see Methods).


A
SARS-CoV-2
RBD
antibody
yeast
RBD mutant library,
sorted for ACE2 binding
pre-selection
sequence
antibody labeling,
FACS
quantify
escape
per-mutant
escape
summed
site-wise
escape
escape fraction
RBD site
unmutated RBD
RBD mutant library
antibody binding
RBD expression
antibody-escape
selected cells
sequence
antibody-escaped


B
Specimen_001-expt24_lib1
Specimen_001-expt24_lib1_1
Specimen_001-expt24_lib
Specimen_001-expt24_lib1_1xAb
SSC-A
SSC-W
SSC-H
FSC-A
FSC-H
FITC-A


C
REGN10933
REGN10987
REGN10933+
REGN10987
LY-CoV16
WT RBD,
400 ng/mL
antibody
WT RBD,
4 ng/mL
antibody
mutant lib 1,
400 ng/mL
antibody
mutant lib 2,
400 ng/mL
antibody
0.1%
0.1%
0.1%
0.0%
95.7%
96.4%
95.2%
94.8%
17.4%
14.8%
10.1%
22.2%
15.4%
15.0%
8.3%
22.5%
PE (antibody binding)
FITC (RBD expression)

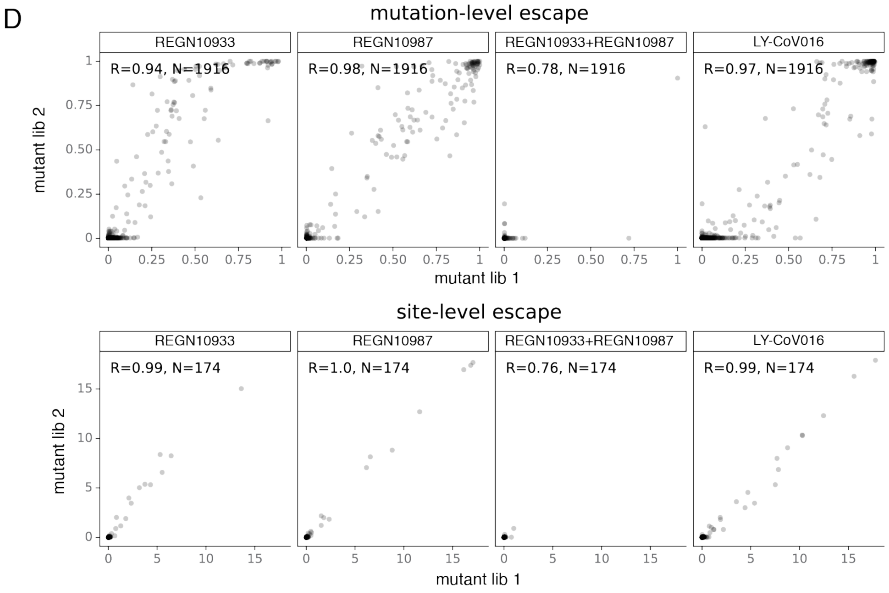

D
mutation-level escape
REGN10933
REGN10987
REGN10933+REGN10987
LY-CoV016
R=0.94, N=1916
R=0.98, N=1916
R=0.78, N=1916
R=0.97, N=1916
mutant lib 2
mutant lib 1
site-level escape
R=0.99, N=174
R=1.0, N=174
R=0.76, N=174
R=0.99, N=174

Figure 4.6: **Complete escape maps colored by effects of mutations on RBD expression in the context of yeast surface display**. **(A)** (A,B) The escape maps shown here are identical to those in Fig. 4.1A,B except that the letters are colored by how mutations affect RBD expression as measured in prior deep mutational scanning using a yeast display system [280], rather than by how they affect the RBD's affinity for ACE2. **(C,D)** To contextualize the range of ACE2 binding and RBD expression values shown by the color scales in the logo plots in panels (A) and (B) of this figure and Fig. 4.1A,B, we plot the distribution of mutational effects on these properties among all possible amino-acid mutations, and among all mutations observed at least 1, 10, or 25 times among sequences deposited in GISAID as of January 11th, 2021. The dotted gray horizontal lines illustrate the computational filters used to eliminate poorly folded or functional mutants (ACE2 binding < -2.35 and RBD expression < -1, see Methods for details) for the current study. The plots show that virtually all mutations observed among circulating isolated pass the computational filters, especially if we require 10 or more observations. Of note, the three high-count mutations with the largest ACE2-binding deficits are all antibody-escape mutations mapped in this study.

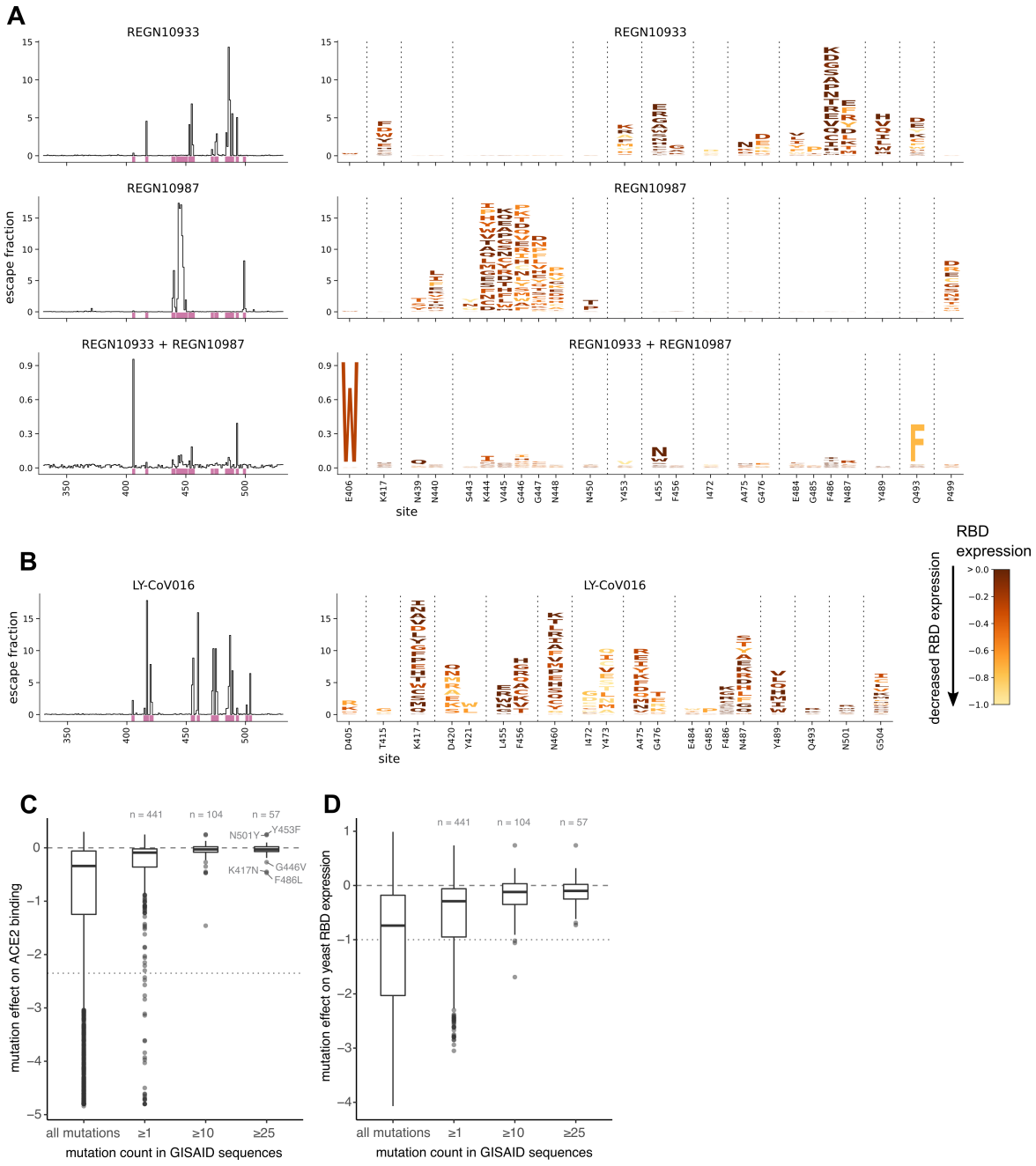

A
REGN10933
REGN10933
REGN10987
REGN10987
REGN10933 + REGN10987
REGN10933 + REGN10987
escape fraction
site
B
LY-CoV016
LY-CoV016
escape fraction
site
RBD expression
decreased RBD expression
C
n = 441
n = 104
n = 57
N501Y
Y453F
K417N
G446V
F486L
mutation effect on ACE2 binding
all mutations
≥1
≥10
≥25
mutation count in GISAID sequences
D
n = 441
n = 104
n = 57
mutation effect on yeast RBD expression
all mutations
≥1
≥10
≥25
mutation count in GISAID sequences

Figure 4.7: **Pseudovirus neutralization curves validating the escape mutant mapping**. **(A-E)** Neutralization of spike mutants against the indicated antibody or antibody cocktail in the context of spike-pseudotyped lentiviral particles. Each panel represents a different batch of neutralization assays performed on a different day. The antibody concentrations shown on the x-axis are the concentration of antibody at which the virus was incubated for 60 minutes prior to being added to target cells (see Methods for details). We determined IC50s by fitting two-parameter Hill curves (with baselines fixed to 1 and 0) using the neutcurve package (`https://jbloomlab.github.io/neutcurve/`, version 0.5.2). See `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_clinical_Abs/tree/main/exptl_neut_data` for the raw data and code used for the fitting. The fold changes in IC50 plotted in Figure 4.1C are the IC50 for the mutant divided by the IC50 for the wildtype (WT) measured on the same day. The numerical values for the IC50s plotted here are available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_clinical_Abs/blob/main/exptl_neut_data/ic50s.csv`. **(F)** Titers of spike pseudotyped lentiviral particles produced with the indicated mutation. The titers are determined by luciferase luminescence after infection of the 293T-ACE2 target cells, and are normalized so that the wildtype spike has a value of one. All titers were measured at least in duplicate (hence the two or more points). C432D is a negative control as it mutates a key cysteine essential for RBD folding; no-spike is another negative control that involves producing pseudotyped lentiviral particles without adding any spike protein expression plasmid.

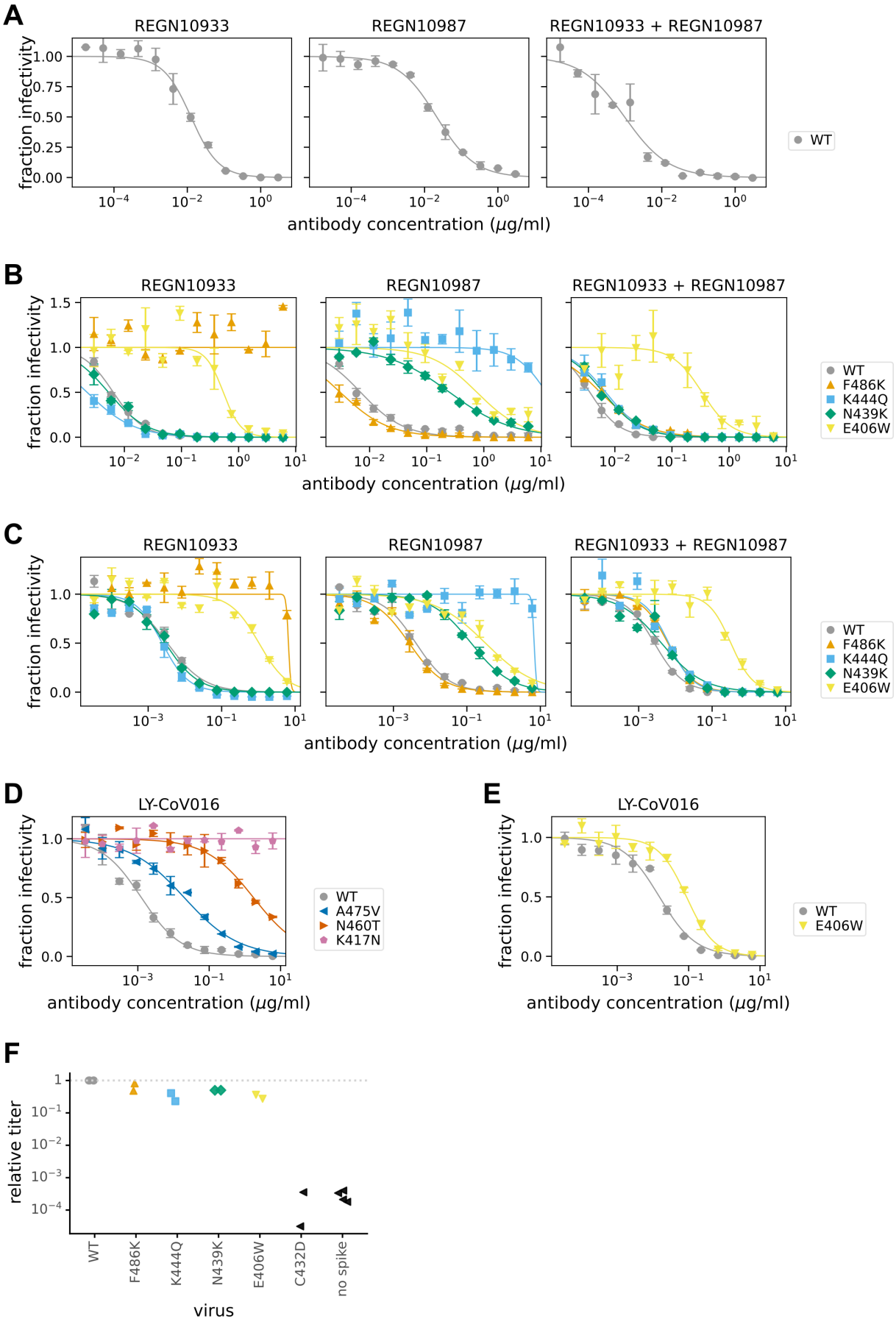
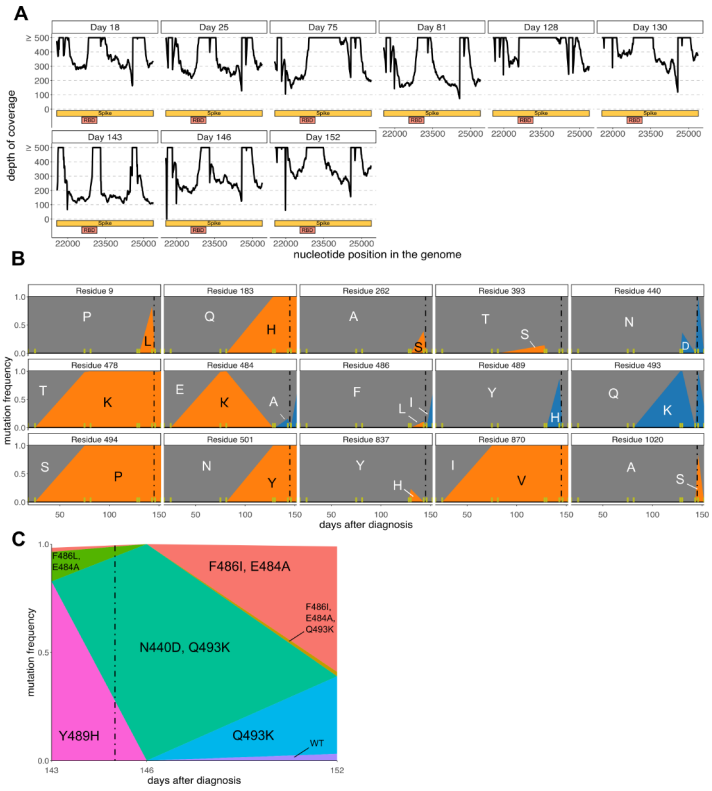

A
REGN10933
REGN10987
REGN10933 + REGN10987
fraction infectivity
antibody concentration (μg/ml)
WT
B
REGN10933
REGN10987
REGN10933 + REGN10987
fraction infectivity
antibody concentration (μg/ml)
WT
F486K
K444Q
N439K
E406W
C
REGN10933
REGN10987
REGN10933 + REGN10987
fraction infectivity
antibody concentration (μg/ml)
WT
F486K
K444Q
N439K
E406W
D
LY-CoV016
fraction infectivity
antibody concentration (μg/ml)
WT
A475V
N460T
K417N
E
LY-CoV016
fraction infectivity
antibody concentration (μg/ml)
WT
E406W
F
relative titer
WT
F486K
K444Q
N439K
E406W
C432D
no spike
virus

Figure 4.8: **Spike mutations in a persistently infected patient treated with REGN-COV2 as determined by Illumina deep sequencing**. **(A)** Coverage at each site in spike for each time point, calculated as the average number of aligned reads with a Q-score ≥ 25 in 10 bp bins. The x-axis shows the nt position in the genome coordinates of Wuhan-Hu-1 (NC_045512.2). The bars underneath show spike and its RBD domain. Coverage > 500 is clipped on the y-axis. **(B)** Dynamics of amino-acid mutations in spike across all time points. Yellow vertical lines on the x-axis indicate sampling times, and the dashed black line indicates administration of REGN-COV2 (145 days). Fig. 4.1C is a subset of this plot that just shows RBD mutations in the time point immediately before and then after REGN-COV2 administration; those mutations are indicated in blue while all others are in orange. None of the RBD mutations that fixed prior to the administration of REGN-COV2 (T478K, S494P, N501Y) impact binding by either of the REGN antibodies according to our maps, although combined mutations could have synergistic effects beyond what we measure in our single-mutant escape maps. However, note that the patient did mount a weak autologous neutralizing antibody response that was detectable by day 31 (16). Therefore, it is possible that both these fixed RBD mutations and the ones that were present at lower frequencies prior to REGN-COV2 treatment could have been selected in part by pressure from the patient's own immune response. However, we do not have access to data on the specificity of the patients' serum neutralizing antibodies. **(C)** Frequencies of different haplotypes in the RBD at the last three time points suggest competition among viral lineages. Note that it is possible that rare haplotypes (such as F468I / E484A / Q493K haplotype) represent library preparation artifacts that arise due to PCR strand exchange between molecules from more common haplotypes (e.g., F486I / E484A haplotype and Q493K haplotype).


A
Day 18
Day 25
Day 75
Day 81
Day 128
Day 130
Day 143
Day 146
Day 152
depth of coverage
nucleotide position in the genome
B
Residue 9
Residue 183
Residue 262
Residue 393
Residue 440
Residue 478
Residue 484
Residue 486
Residue 489
Residue 493
Residue 494
Residue 501
Residue 837
Residue 870
Residue 1020
mutation frequency
days after diagnosis
C
F486L, E484A
F486I, E484A
F486I, E484A, Q493K
N440D, Q493K
Y489H
Q493K
WT
mutation frequency
days after diagnosis

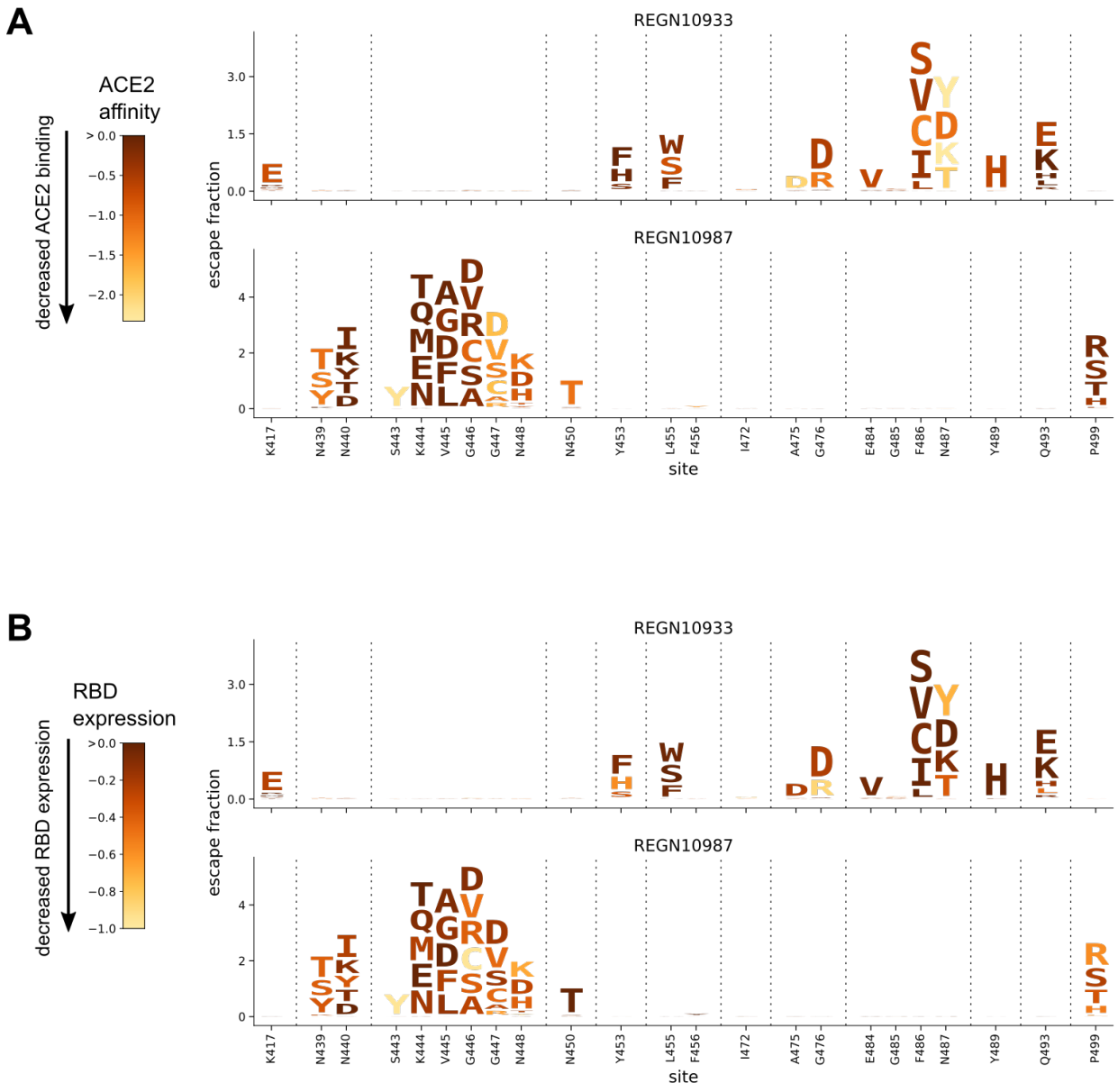


Figure 4.9: **Maps of single-nucleotide accessible escape mutations from REGN10933 and REGN10987 colored by how mutations affect the RBD's affinity for ACE2 or expression of folded protein. (A-B)** These plots show the same mutations as in Figure 4.2B (only those accessible by single-nucleotide changes to Wuhan-Hu-1), but colored according to the schemes in Fig 4.1A and Fig 4.6.

Chapter 5

COMPREHENSIVE MAPPING OF MUTATIONS TO THE SARS-COV-2 RECEPTOR-BINDING DOMAIN THAT AFFECT RECOGNITION BY POLYCLONAL HUMAN PLASMA ANTIBODIES

A version of this chapter has been previously published as:

Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host Microbe. 2021 Mar 10;29(3):463-476.e6. doi: 10.1016/j.chom.2021.02.003. Epub 2021 Feb 8. PMID: 33592168; PMCID: PMC7869748.

5.1 Abstract

The evolution of SARS-CoV-2 could impair recognition of the virus by human antibody-mediated immunity. To facilitate prospective surveillance for such evolution, we map how convalescent plasma antibodies are impacted by all mutations to the spike's receptor-binding domain (RBD), the main target of plasma neutralizing activity. Binding by polyclonal plasma antibodies is affected by mutations in three main epitopes in the RBD, but there is substantial variation in the impact of mutations both among individuals and within the same individual over time. Despite this inter- and intra-person heterogeneity, the mutations that most reduce antibody binding usually occur at just a few sites in the RBD's receptor binding motif. The most important site is E484, where neutralization by some plasma is

reduced >10-fold by several mutations, including one in the emerging 20H/501Y.V2 and 20J/501Y.V3 SARS-CoV-2 lineages. Going forward, these plasma escape maps can inform surveillance of SARS-CoV-2 evolution.

5.2 Introduction

Neutralizing antibodies against the SARS-CoV-2 spike are associated with protection against infection in both humans [6, 195] and animals [7, 304, 356]. However, other human coronaviruses undergo antigenic evolution that erodes neutralizing antibody immunity [89]. This antigenic evolution is driven by positive selection for mutations in the viral spike, particularly in regions involved in receptor binding [158, 327]. To monitor for similar antigenic evolution of SARS-CoV-2, it is important to determine which viral mutations impact human polyclonal antibody immunity.

A multitude of recent studies have identified viral mutations that escape monoclonal antibodies targeting the SARS-CoV-2 spike [19, 124, 278, 320, 182, 189]. However, it remains unclear how mutations that escape specific monoclonal antibodies will affect the polyclonal antibody response elicited by infection. Several recent studies have identified viral mutations that impact neutralization by polyclonal human plasma or sera. So far, these studies have relied on either selecting viral escape mutants with reduced neutralization sensitivity [11, 320], or characterizing the antigenic effects of specific mutations such as those observed in circulating viral isolates [182, 287, 207, 189, 324, 315]. This work has shown that single mutations to the spike's receptor-binding domain (RBD) or N-terminal domain (NTD) can appreciably reduce viral neutralization by polyclonal plasma, sometimes by as much as 10-fold. However, these studies characterize an incomplete subset of all possible mutations, and thus do not completely describe the effects of viral mutations on recognition by polyclonal antibodies.

Here we comprehensively map how all amino-acid mutations to the SARS-CoV-2 spike

RBD affect binding by the antibodies in plasma collected from convalescent individuals ~1 to ~3 months post-symptom onset. We focus on the RBD because prior studies have reported that RBD-binding antibodies contribute the majority of the neutralizing activity of most human plasma [229, 282], a result we confirm. Our complete maps of how mutations impact plasma antibody binding identify three major epitopes in the RBD. However, antibody binding from different individuals is impacted differently by mutations in these epitopes, and sometimes the impacts of mutations also change over time for longitudinal samples from the same individual. Some mutations that reduce antibody binding also reduce viral neutralization by >10-fold. The site where mutations tend to have the largest effect on binding and neutralization is E484, which unfortunately is a site where mutations are present in several emerging SARS-CoV-2 lineages [285, 302]. However, some plasma are more affected by mutations at other sites, while others are largely unaffected by any single mutation. Overall, these systematic maps of how mutations to the SARS-CoV-2 RBD affect recognition by human antibody immunity can inform surveillance of ongoing viral evolution.

5.3 Results

5.3.1 RBD antibodies dominate the neutralizing activity of most convalescent plasmas

We characterized 35 plasma samples longitudinally collected from 17 different SARS-CoV-2-infected individuals between 15 and 152 days post-symptom onset (5.7A). Prior work has shown that these samples all have RBD-binding antibodies and neutralizing activity, with a median neutralization titer 50% (NT50) of ~250 (range of 34 to >10,000) against lentiviral particles pseudotyped with the D614 variant of the SARS-CoV-2 spike. For most of the 17 individuals, both the RBD binding and the neutralizing activity decreased moderately from one to four months post-infection [66] (see `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/supplementary_tables/TableS1.csv`). We previously characterized binding of the IgG, IgA, and IgM antibody isotypes to RBD for all the plasma

samples [66], and those data are re-plotted in 5.7B.

Several recent studies have reported that RBD-binding antibodies contribute the majority of the neutralizing activity in most convalescent human plasma [229, 282]. To confirm the importance of anti-RBD antibodies for the samples in our study, we used RBD-conjugated beads to deplete RBD-binding antibodies, and compared the neutralizing activity pre- and post-depletion. First, we validated that these depletions effectively removed RBD-directed but not other anti-spike antibodies. To do this, we created "synthetic" sera by combining non-neutralizing pre-pandemic sera with either an RBD-binding or N-terminal domain (NTD)-binding monoclonal antibody. As expected, RBD antibody depletion completely eliminated binding of the anti-RBD synthetic sera to both RBD and spike, but did not reduce the spike binding activity of the anti-NTD synthetic sera (5.7C).

We then validated that the depletion removed all RBD-binding antibodies from the convalescent human plasma, and examined how depletion of RBD-binding antibodies affected total plasma binding to the spike ectodomain (Figures 5.1A, 5.7D). To do this, we performed ELISAs for binding to the RBD and spike for each sample both pre- and post-depletion, using an anti-IgG secondary antibody. The depletion removed essentially all RBD-binding IgG antibodies but only modestly decreased the amount of IgG that bound to spike (Figures 5.1A, 5.7D). This result suggests that RBD-binding antibodies comprise a relatively modest proportion of all spike-binding IgG plasma antibodies in naturally infected individuals, consistent with studies reporting that less than half of spike-reactive B cells and monoclonal antibodies bind to RBD [33, 142, 267, 303].

We next measured how depletion of RBD-binding antibodies affected neutralization of lentiviral particles pseudotyped with the now-predominant G614 variant of the SARS-CoV-2 spike, and found that RBD-binding antibodies usually dominated the neutralizing activity 5.1B, 5.7C, `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/supplementary_tables/TableS1.csv`.Specifically, the majority of the neutralizing

activity was due to RBD-binding antibodies in nearly all samples (33 of 35 tested), and >90% of neutralizing activity was due to RBD-binding antibodies in over a third of the samples (13 of 35 tested) (5.1B, 5.7E).

Notably, RBD-binding antibodies dominated the plasma neutralizing activity both at early (30 day) and late (100 day) time points post-symptom onset. For many individuals, the contribution of RBD-binding antibodies to plasma neutralizing activity increased over time, although this was not always the case (5.1B,5.7F). For instance, the contribution of RBD-binding antibodies to neutralization increased over time for subjects E and J, but not subjects L or M. The strong contribution of RBD-binding antibodies to plasma neutralization demonstrates that mapping mutations that escape these antibodies is crucial for understanding the potential for SARS-CoV-2 antigenic evolution.

5.3.2 Complete mapping of RBD mutations that reduce binding by plasma collected 1 month post-symptom onset

To completely map RBD mutations that reduce binding by polyclonal plasma antibodies, we extended a deep-mutational scanning method previously developed to identify mutations that escape binding by monoclonal antibodies [124]. Briefly, we used libraries of yeast that each expressed a different RBD variant on their surface. The library covered nearly all possible single amino-acid mutations to the RBD [280]. We incubated these yeast libraries with polyclonal human plasma, and used fluorescence-activated cell sorting (FACS) with an IgG/IgA/IgM secondary antibody to enrich for yeast expressing RBD mutants that bound appreciably less plasma antibodies than unmutagenized RBD (Figure 5.8A–C). We then used deep sequencing to measure the frequency of each RBD mutation in the initial population and the antibody-escape FACS bin. We quantified the effect of each RBD mutation on plasma antibody binding as that mutation's "escape fraction," which is the fraction of all yeast cells expressing RBD with that mutation that fall into the FACS escape bin. These

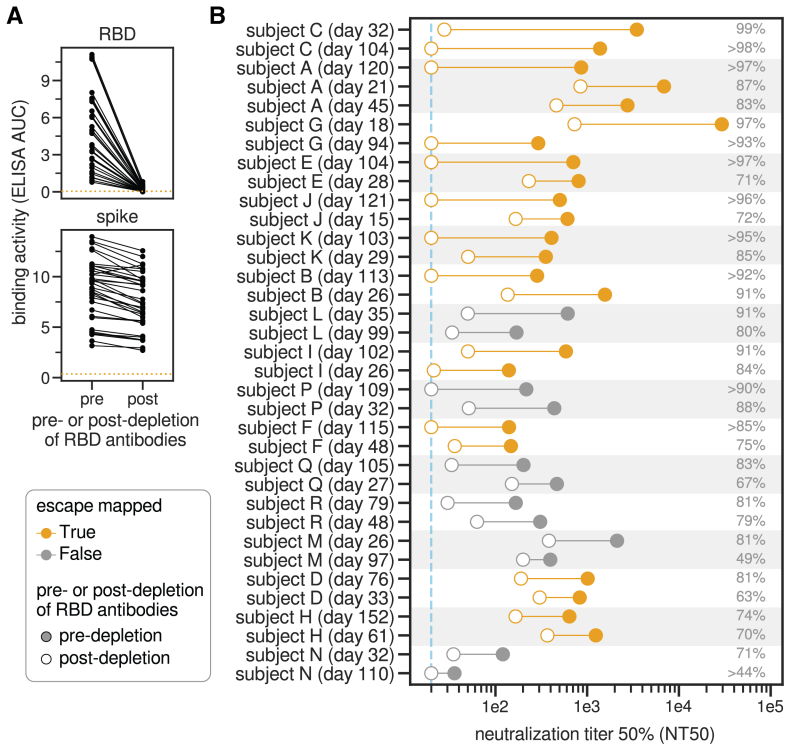


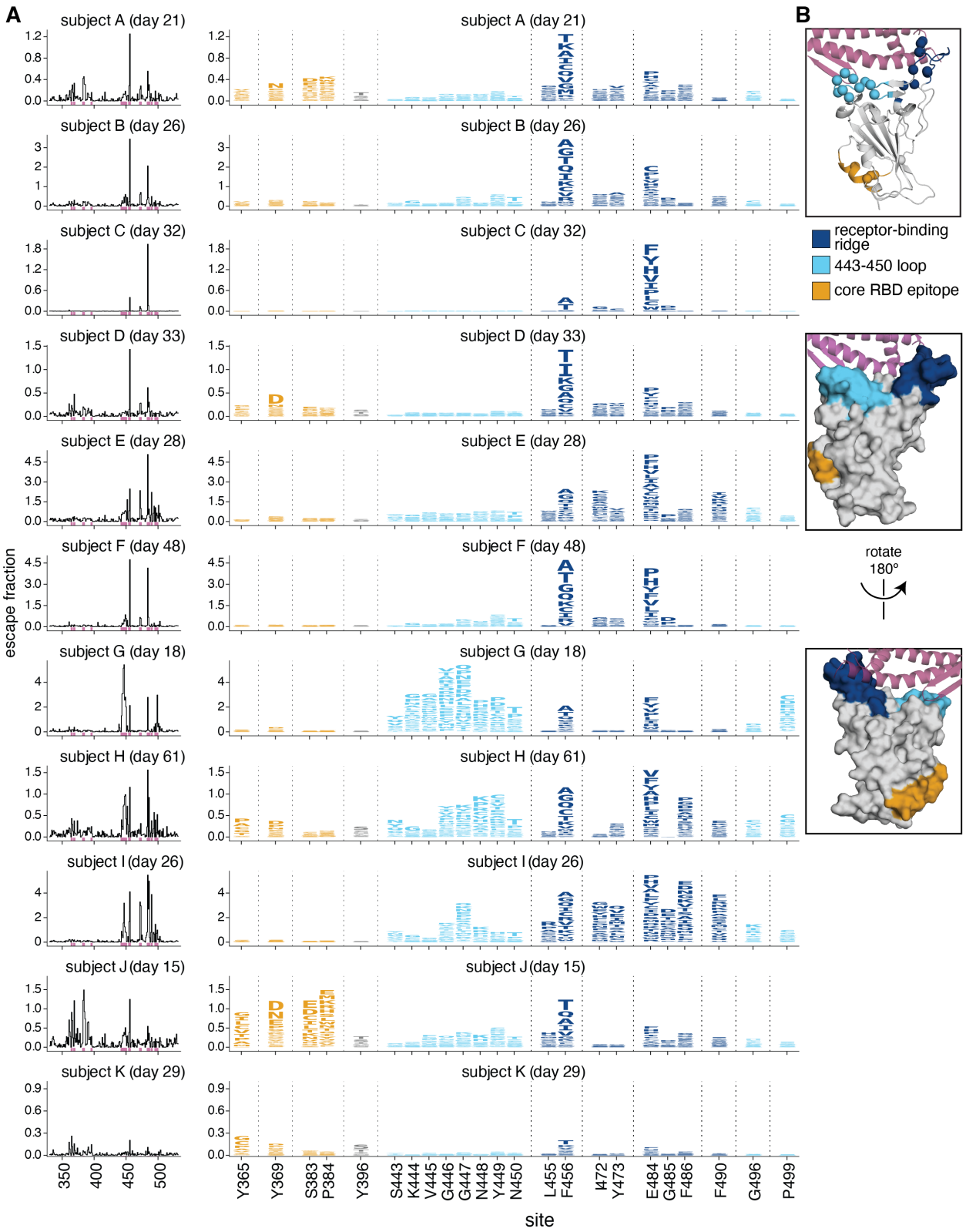
Figure 5.1: **RBD-binding antibodies are responsible for most of the neutralizing activity of human polyclonal plasma**. **(A)** Change in binding of plasma to RBD and spike before and after depletion of RBD antibodies, measured by ELISA area under the curve (AUC). The dashed orange line is binding of pre-pandemic pooled sera collected in 2017-2018. Raw ELISA binding curves in Figure 5.7A. **(B)** Neutralization titer 50% (NT50) of human plasma before and after depletion of RBD-binding antibodies. Legend is at left: filled and open circles are pre- and post-depletion samples, respectively, connected by a line. Orange indicates plasma for which we subsequently mapped mutations that reduce binding. The numbers at right indicate the percent of all neutralizing activity attributable to RBD-binding antibodies. Plasma are sorted in descending order of percent of neutralization due to RBD-binding antibodies, first by subject and then within subject. The dashed blue line is the limit of detection (NT50 of 20). Points on this line have an NT50 $\leq$ 20, so the percent of neutralization due to RBD-binding antibodies may be an underestimate for these plasma. See Figure 5.71 and `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/supplementary_tables/TableS1.csv` for additional data including sample metadata, full ELISA and neutralization curves, and numerical values plotted here.

escape fractions range from 0 (no effect on plasma antibody binding) to 1 (all cells with this mutation are in the antibody-escape bin) (Figure 5.8A–C) [124, 278]. All mapping experiments were performed in biological duplicate using independently constructed RBD mutant libraries; the replicates were highly correlated (Figure 5.8D,E), and we report the average measurements across the two libraries throughout.

We began by mapping mutations that reduced binding by plasma antibodies in samples collected from 11 individuals at approximately 30 days post-symptom onset (range 15–61 days). These samples had neutralizing titers against lentiviral particles pseudotyped with the G614 variant of the SARS-CoV-2 spike that ranged from 140 to 30,000, with the extent of neutralization attributable to RBD-targeting antibodies ranging from 63% to 99% (first time point for subjects shown in orange in Figure 5.1B). We quantified each RBD mutation's escape fraction and visualized the results using "escape maps," which are logo plots where the height of each letter is proportional to that mutation's escape fraction (Figure 5.2A, 5.8A). Interactive versions of these escape maps are at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_HAARVI_sera`. The total height of the letter stacks for each site represent the sum of the escape fractions for all mutations at that site, and so can range from 0 (no effect of any mutation) to 19 (all mutations at the site have an escape fraction of 1).

Although the effects of mutations on plasma antibody binding varied widely across individuals, the escape maps revealed several common patterns. Mutations that strongly reduced binding fell in one of three discrete regions of the RBD: the receptor-binding ridge within the receptor-binding motif (RBM), a loop in the RBM opposite the ridge (spanning sites 443–450, and the structurally adjacent sites at 494–501), or a surface patch in the core RBD (Figure 5.2B). The receptor-binding ridge and 443–450 loop are also targeted by many potently neutralizing antibodies, including the two antibodies in the REGN-COV2 cocktail [124, 129, 278]. The core RBD epitope is targeted by monoclonal antibodies that tend to be less potently neutralizing but more broadly cross-reactive to SARS-like coronaviruses

Figure 5.2: **Complete maps of RBD mutations that reduce binding by polyclonal plasma antibodies from 11 individuals**. **(A)** The line plots at left indicate the total effect of all mutations at each site in the RBD on plasma antibody binding, with larger values indicating a greater reduction in antibody binding. The logo plots at right zoom in on individual mutations at key sites (indicated by purple highlighting on the x-axis of the line plots). In these logo plots, the height of each letter is that mutation's escape fraction, so larger letters indicate mutations that cause a greater reduction in antibody binding. Escape fractions are comparable across sites within a sample, but not necessarily between samples due to the use of sample-specific FACS gates—therefore, for each plasma, the y-axis is scaled independently (see Methods). Sites in the logo plots are colored by RBD epitope. **(B)** For coloring of the logo plots, we designated three RBD epitopes based on the structural locations where mutations had large effects on plasma antibody binding. The images show the structure of the RBD bound to ACE2 (PDB 6M0J, [172]) in several representations. The receptor-binding-ridge epitope is dark blue, the epitope containing the 443–450 loop is cyan, the core-RBD epitope is orange, the rest of the RBD is gray, and ACE2 is purple. For the cartoon rendering in the top structure, alpha carbons for sites of strong binding-escape for any of the 11 plasmas (i.e., all sites shown in the logo plots) are represented as spheres. Interactive versions of these escape maps are available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_HAARVI_sera/`.


A
B
subject A (day 21)
subject B (day 26)
subject C (day 32)
subject D (day 33)
subject E (day 28)
subject F (day 48)
subject G (day 18)
subject H (day 61)
subject I (day 26)
subject J (day 15)
subject K (day 29)
escape fraction
site
receptor-binding ridge
443-450 loop
core RBD epitope
rotate 180°

[16, 229, 347, 356]. In particular, binding by all 11 samples was reduced by mutations at site F456, and binding by most samples (9 of 11) was reduced by mutations at site E484 (Figure 5.2A). Both of these sites are within the receptor-binding ridge epitope. Notably, E484 is a site at which mutations have recently been demonstrated to reduce neutralization by both monoclonal antibodies and human sera or plasma [11, 124, 278, 320, 315].

We grouped the samples into several classes based on which mutations most strongly reduced plasma antibody binding (Figure 5.3 and the interactive visualizations at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_HAARVI_sera/`; sera were grouped manually based on qualitative examination of escape profiles). Binding by 6 of the 11 samples was reduced primarily by mutations in the receptor-binding ridge. For instance, binding by plasma antibodies from subject B (day 26) was most strongly affected by mutations at sites F456 and E484 (Figure 5.2, 5.3A). Binding by three samples was strongly reduced by mutations across a broader swath of the RBM, including the 443–450 loop (Figure 5.3B). An example is subject G (day 18), which was strongly affected by mutations at sites 443–450 in addition to F456 and E484. Binding by two samples was most affected by mutations in the core RBD epitope (Figure 5.3C). The sites where mutations reduced binding by these core-RBD targeting plasma clustered around the lipid-binding pocket in the RBD, where binding of free fatty acids may contribute to locking spike into a "closed" conformation [39, 291]. Notably, for the sample from subject K (day 29), no single RBD mutation had more than a small effect on plasma antibody binding (Figures 5.2, 5.3D).

There are some regions of the RBD where mutations did not strongly affect plasma antibody binding for any sample in our panel. These regions include the sites near the 343 glycan that are targeted by the SARS-CoV-1 cross-reactive antibody S309 [230], and the region near residue E465 on the "lateral" side of the RBD, which to our knowledge is not an epitope for any known neutralizing antibodies (Figure 5.3E) [16, 124, 229, 280]. Antibodies targeting these two regions may be rare, have low binding avidity in the context of polyclonal

plasma, or be subdominant relative to other RBD epitopes [17, 229, 320].

5.3.3 How mutations affect plasma antibody binding can shift over time in the same individual

Next, we examined how the RBD mutations that affect plasma antibody binding change over time as an individual's immune response matures. We speculated that such changes might occur because other studies have shown that anti-SARS-CoV-2 antibodies become more somatically hypermutated and less clonal in the months following recovery from infection [106, 218, 253]. Moreover, we reasoned that mapping mutations that affect plasma antibody binding several months after infection would be relevant for determining which viral mutations might alter the effectiveness of immunity if these individuals were re-exposed to a distinct SARS-CoV-2 variant in the future. We performed escape mapping for samples collected at later time points (76–152 days post-symptom onset) from all 11 individuals for whom we had characterized plasma antibody binding at the 1 month time point. For some but not all individuals there were substantive changes over time in how binding was affected by RBD mutations (Figures 5.4, 5.9, 5.10 and interactive visualizations at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_HAARVI_sera`) [137].

Specifically, for over half of the 11 individuals there was relatively little change in which RBD mutations affected plasma antibody binding 5.4A, 5.9, 5.10. For two individuals, antibody binding became strikingly more broad and less affected by any single RBD mutation (5.4B, 5.9, 5.10). For one individual, mutations in the 443–450 loop had a much larger effect on binding by plasma antibodies from the later time compared to the earlier time (5.4C, 5.9, 5.10). Finally, for one individual, there was a strong narrowing of the response, with no single RBD mutation having a large effect on binding by plasma from the early time point, but mutations at F456 and to a lesser extent E484 having large effects at the later time point (5.4D, 5.9, 5.10). In summary, while the specificity of plasma antibody binding is

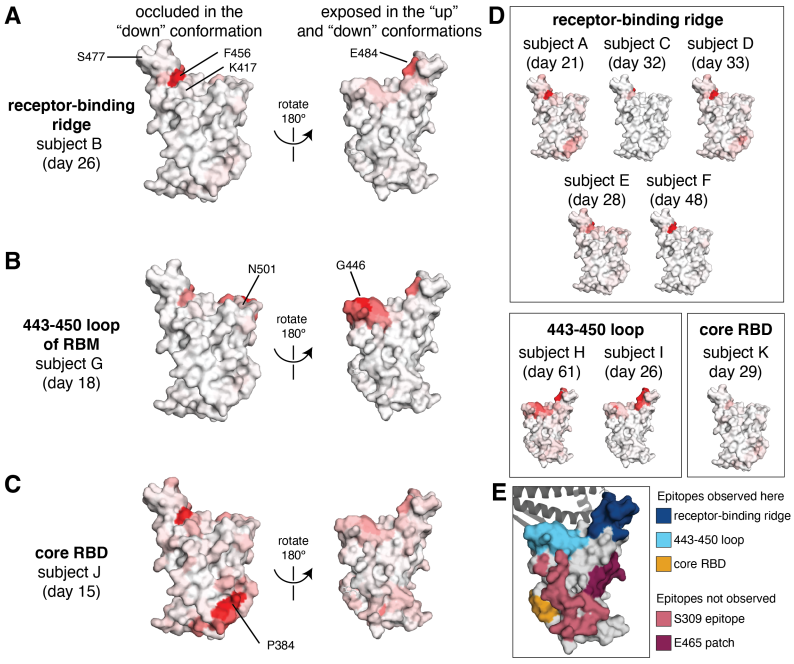


Figure 5.3: **Regions of the RBD where mutations strongly reduced binding by the antibodies in plasmas collected from 11 individuals**. The total effect of mutations at each site (sum of escape fractions) are projected onto the structure of the RBD (PDB 6M0J), with white indicating no effect of mutations at that site and red indicating a large reduction in antibody binding. Two views of the RBD are shown: the surface of the RBD that is buried in the "down" conformation, and the surface that is always exposed and accessible [306, 332]. **(A)** For some individuals (typified by subject B), antibody binding is predominantly reduced by mutations in the receptor-binding ridge, particularly at sites F456 and E484. **(B)** For some individuals (typified by subject G), antibody binding is strongly reduced by mutations in the 443–450 loop of the RBM in addition to the receptor-binding ridge. **(C)** For a few individuals (typified by subject J), antibody binding is affected by mutations in the core RBD epitope around site P384. **(D)** Samples from the other eight individuals fall in one of the three classes detailed in panels (A) to (C). For panels (A) to (D), the white-to-red coloring scale is set to span the same range as the y-axis limits for that plasma in Figure 5.2. **(E)** Mutations in two major surface regions (the S309 epitope and the sites near E465) do not strongly affect plasma antibody binding for any of the subjects. Shown is a surface representation of the RBD, with the 3 polyclonal plasma epitopes colored as in Figure 5.2. The S309 epitope and region near E465 ("E465 patch") are shown in pink and maroon. ACE2 is shown in a dark gray cartoon representation. Interactive versions of these structural visualizations are available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_HAARVI_sera/`.

often maintained over time, in some individuals the specificity broadens to become relatively unaffected by any single RBD mutation, while in other individuals the specificity narrows so that single mutations have a greater impact.

However, there was no clear relationship between changes in the fine specificity of antibody binding and overall plasma neutralizing activity. For instance, subject B and subject C maintained similar binding specificities over time (Figure 5.4A), even though the neutralization titers of both subjects' plasma decreased (Figure 5.1B). Similarly, subject D showed major changes in binding specificity over time (Figure 5.4B), although this change in specificity was not accompanied by a substantial change in overall plasma neutralization titer.

5.3.4 For some plasma, RBD mutations that reduce antibody binding strongly reduce neutralization

To determine how mutations that reduced plasma antibody binding to the RBD affected viral neutralization, we characterized a subset of mutations in neutralization assays with spike-pseudotyped lentiviral particles. For these assays, we chose mutations that our mapping showed had substantial effects on plasma antibody binding by samples from multiple individuals, and prioritized mutations present in circulating isolates of SARS-CoV-2.

In many cases, single mutations that were mapped to strongly reduce plasma antibody binding also greatly reduced viral neutralization. The effect of mutations at site E484 were particularly striking (Figure 5.5A,B). For several plasma, the neutralization titer dropped by over an order of magnitude against viruses pseudotyped with spikes with E484 mutated to K, Q, or P. For instance, these three mutations to E484 caused 35- to 115-fold decreases in the neutralization titer of the plasma collected from subject C (day 32) (Figure 5.5A,D). As another example, both E484K and E484Q reduced neutralization by the plasma from subject B (day 26) by 10-fold, the same reduction achieved by depleting the plasma of all RBD-binding antibodies (Figure 5.5A,D).

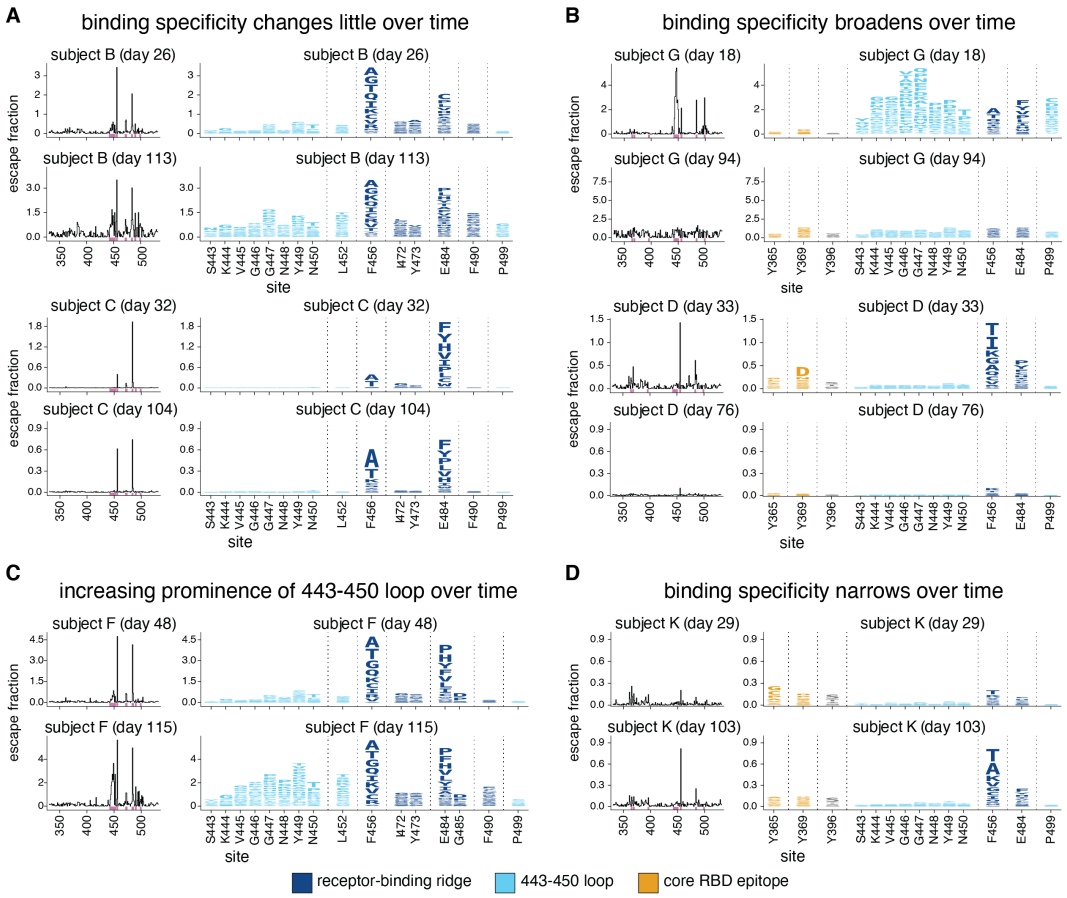
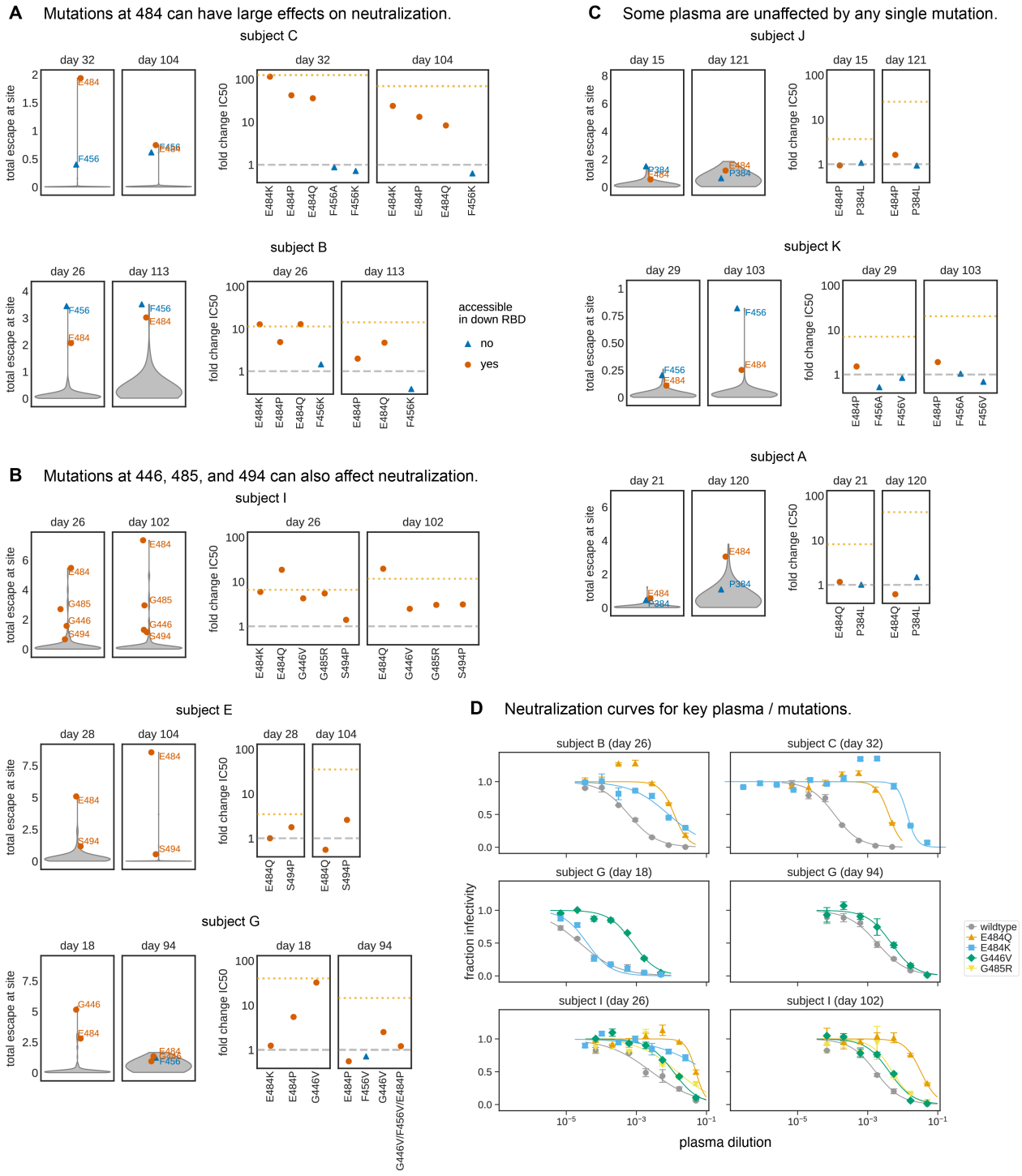


Figure 5.4: **The RBD mutations that affect plasma antibody binding change over time for some individuals**. Escape maps, colored as in Figure 5.2, demonstrating temporal patterns: **(A)** no change over time, **(B)** broadening over time, **(C)** increasing prominence of one antigenic region, the 443–450 loop, or **(D)** narrowing over time. This figure shows the escape maps over time for 6 of the 11 individuals to illustrate representative trends; see Figure 5.9 for escape maps for all individuals at all time points. Figure 5.10 shows the effects of mutations at each site projected onto the RBD structure. Different sets of sites are shown in the logo plots in panels A and C, and in panels B and D. Sites highlighted in the logo plots are indicated in purple on the x-axes of the associated line plots. The y-axis limits were set as in Figure 5.2A (see Methods). Interactive versions of these visualizations are available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_HAARVI_sera/`.

While mutations at E484 generally caused the largest drops in neutralization, other mutations mapped to decrease antibody binding for specific plasma also affected neutralization. A dramatic example was G446V, which caused a 30-fold decrease in the neutralization titer of subject G (day 18) (Figure 5.5B,D). Mutations G485R and S494P also caused lesser but still appreciable (3 to 5-fold) decreases in neutralization titer for a few plasma (Figure 5.5B). However, no single mutation completely abrogated neutralization for any of the plasma samples (Figure 5.11C).

In general, there was good concordance between the mapping of how mutations affected plasma antibody binding and their impact on viral neutralization. This concordance can be seen in Figure 5.5A-C, where the violin plots show the distributions of the effects of mutations on plasma antibody binding across all sites. The sites in the upper tails of these violin plots are ones where mutations had large effects on binding, and mutations to such sites usually reduced neutralization. The one major exception was site F456, where mutations often caused large reductions in binding to yeast-displayed RBD but never appreciably affected neutralization (Figure 5.2A, 5.5A-C). This discrepancy is not because antibodies targeting this region are inherently non-neutralizing or unaffected by mutations at site 456, as F456A and F456K disrupt neutralization by two monoclonal antibodies with epitopes that include F456 [124, 271, 278, 356] (Figure 5.11B). Rather, we hypothesize that the discrepancy is because we mapped how mutations affected binding using isolated RBD, but in the native viral context of full spike, RBD can be positioned in either a "down" or "up" conformation [306, 332]. All sites where mutations that reduced binding also affected neutralization are accessible in both conformations, but F456 is only accessible in the "up" conformation. Because the RBD is usually in the "down" conformation [36, 152, 306], we speculate that sites accessible only in the "up" conformation may be subdominant in the context of polyclonal plasma neutralization of full spike even if they are dominant when assaying binding to isolated yeast-displayed RBD.

Figure 5.5: **Mutations mapped to reduce plasma antibody binding often reduce viral neutralization**. **(A–C)** Violin plots at left show the distribution of how mutations at all sites in the RBD affect plasma binding in the mapping experiments. The plots at right then show the effects of tested mutations on neutralization (the fold-change in neutralization inhibitory concentration 50% (IC50)). For instance, the top row in (A) shows that mutations at E484 and F456 are mapped to reduce plasma antibody binding for subject C at both day 32 and day 104, and that multiple different mutations at E484 but not F456 greatly reduced plasma neutralization (e.g., a greater than 100-fold increase in IC50 for E484K for the day 32 plasma). Sites that are accessible in the down conformation of the RBD in the context of full spike are indicated by red circles (e.g., E484), and sites that are inaccessible in the RBD's down conformation are indicated by blue triangles, (e.g., F456). In the plots showing the fold-change in IC50s, the dashed gray line indicates a value of one (no change in neutralization), and the dotted orange line indicates the change in inhibitory concentration if all RBD-binding antibodies are removed (see Figure 5.1B). **(D)** Full neutralization curves for a subset of plasma and viral mutants demonstrating how E484Q, E484K, G446V, and G485R substantially reduce viral neutralization for some plasma. For all neutralization curves used to determine changes in neutralization plotted in (A–C), see Figure 5.11. The y-axis limits in the violin plots are set as the maximum of the y-axis limit for all time points of a subject in the escape maps in Figure 5.2A and 5.9. Numerical IC50 values and fold-change IC50 relative to wildtype are listed at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/experimental_validations/results/mutant_neuts_results/mutants_foldchange_ic50.csv`.


A
Mutations at 484 can have large effects on neutralization.
subject C
day 32
day 104
total escape at site
fold change IC50
E484
F456
E484K
E484P
E484Q
F456A
F456K
subject B
day 26
day 113
accessible in down RBD
no
yes
B
Mutations at 446, 485, and 494 can also affect neutralization.
subject I
G485
G446
S494
G446V
G485R
S494P
subject E
day 28
subject G
day 18
day 94
G446
C
Some plasma are unaffected by any single mutation.
subject J
day 15
day 121
P384
P384L
subject K
day 29
day 103
F456V
subject A
day 21
day 120
D
Neutralization curves for key plasma / mutations.
subject B (day 26)
subject C (day 32)
subject G (day 18)
subject G (day 94)
subject I (day 26)
subject I (day 102)
fraction infectivity
plasma dilution
wildtype
E484Q
E484K
G446V
G485R

The neutralization assays also validated one of the most notable findings from the mapping: that the antigenic effects of mutations varied markedly across samples from different individuals. For several samples, the maps of binding escape were relatively "flat" with no mutations having large effects, and for these samples no tested mutations substantially affected neutralization (Figure 5.5C). Additionally, sometimes the effects of mutations changed over time for the same individual. Such a temporal change was especially notable for subject G: mapping of the day 18 sample showed a strong effect of mutations centered around G446, but by day 94 the escape map had flattened (Figure 5.4B). Concordant with the maps, G446V greatly decreased neutralization by subject G's day 18 plasma, but had only a modest effect on the day 94 plasma, even when combined with mutations at several other key sites (Figure 5.5B,D). These facts highlight how the antigenic effects of mutations vary across people and time, and suggest that some plasma are more resistant than others to erosion by viral evolution.

5.3.5 RBD mutations that reduce plasma binding and neutralization in circulating SARS-CoV-2 isolates

To determine the extent that mutations we mapped to affect plasma binding are present among circulating SARS-CoV-2 isolates, we determined the frequency of mutations at each RBD site among all SARS-CoV-sequences in GISAID as of Dec-23-2020 [91]. We then compared these frequencies to the effects of mutations at each site on plasma antibody binding, averaged across all samples (Figure 5.6A).

The most concerning site of mutations is E484 (Figure 5.6A). E484 is the site where mutations tend to have the largest effect on plasma antibody binding to the RBD, and our neutralization assays (Figure 5.6A,D) and similar experiments by others [11, 189, 320] show that mutations to site E484 reduce the neutralization potency of some human plasma by $>$ 10-fold, although other plasma are unaffected by mutations at this site. Over 0.1%

of all sequenced isolates have mutations at this site. Of note, E484K is present in the 20H/501Y.V2 viral lineage (also known as B.1.351, originally identified in South Africa) and the 20J/501Y.V3 viral lineage (also known as P.1, originally identified in Brazil) [94, 285, 302]; another mutation at the same site (E484Q) has also been found in a smaller number of human isolates (Figure 5.6B). Consistent with the observation of E484K/Q mutation in naturally occurring SARS-CoV-2 isolates, these mutations have neutral-to-very mildly beneficial effects on RBD affinity for ACE2 [280]. The 20H/501Y.V2 and 20J/501Y.V3 lineages have two other RBD mutations, N501Y and K417N or K417T, respectively, that co-occur with E484K [94, 285]. K417N escapes neutralization by some monoclonal antibodies [124, 278], but mutations to site 417 only modestly affected binding by a few of the samples we assayed (the largest effects were for the last time point for subjects A and J, see Figure 5.9). N501Y increases affinity for ACE2 [280] and is also present in the 20I/501Y.V1 lineage (also known as B.1.1.7, originally identified in the U.K.) that may have increased transmissibility [154, 237, 244]. Although other mutations at N501 have modest effects on binding by some monoclonal antibodies [124, 278], mutations at N501 do not strongly affect binding by any plasma we tested (Figures 5.2A, 5.93). However, several recent studies suggest that the combination of K417N + E484K + N501Y may cause a larger decrease in neutralization than any of these mutations alone [42, 324, 315].

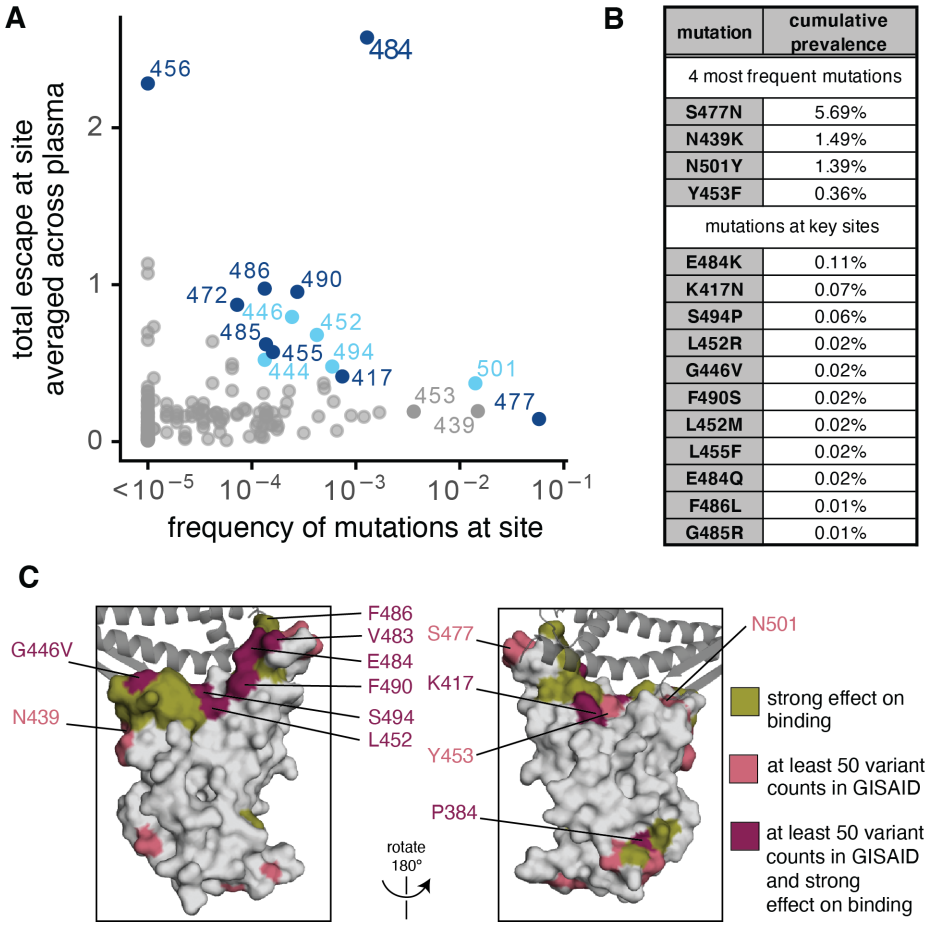
Several other sites where we mapped mutations to affect plasma antibody binding for a few samples also have low-level variation ($< 0.1\%$) among circulating viruses (Figure 5.6A–C). These include site G446, where the G446V mutation reduced neutralization by one sample by $>$ 10-fold (Figure 5.5B,D). Other key sites with circulating variation where mutations impact binding by some samples are indicated in Figure 5.6. Notably, site F456, where mutations consistently affect plasma antibody binding but not neutralization, has little variation among circulating viruses (Figure 5.6A)–perhaps in part because most mutations to site F456 decrease RBD affinity for ACE2 [280].

The four mutations at the highest frequency among sequenced viruses (S477N, N439K, N501Y, and Y453F; see Figure 5.6B) do not strongly affect plasma antibody binding to yeast-displayed RBD by any samples we tested. As mentioned above, N501Y increases affinity for ACE2, is present in the 20I/501Y.V1 (B.1.1.7) lineage [154], and is in the epitope defined by the "443–450 loop" (Figure 5.2B,5.3B)—but does not impact binding by any samples we tested, a result corroborated by live-virus neutralization assays [213, 248]. Y453F and N439K both also increase affinity for ACE2 [280, 287], and both escape some monoclonal antibodies [19, 278, 287] but neither greatly impact plasma antibody binding by the samples we tested. Finally, S477N also reduces neutralization by some monoclonal antibodies [189], but did not greatly affect binding by the samples we tested.

In summary, our results suggest that E484 is the site of most concern for viral mutations that impact binding and neutralization by polyclonal plasma antibodies targeting the RBD. However, mutations at the other plasma antibody epitopes (e.g., the 443–450 loop, site 417, and residues around 484 such as 455, 485, 486, and 490) also have antigenic impacts.

5.4 Discussion

We comprehensively mapped how mutations to the SARS-CoV-2 RBD affected binding by the antibodies in convalescent human plasma. One major result is that plasma antibody binding is predominantly affected by mutations at just a few dominant epitopes in the RBD. In particular, E484 is the site in the RBD where mutations usually have the largest effect on binding and neutralization—possibly because E484 is often targeted by antibodies that utilize heavy-chain germline genes that are common among anti-SARS-CoV-2 RBD antibodies, IGHV3-53 and IGHV3-66 [17, 124, 251, 320, 346, 357]. Mutations at other structurally adjacent sites in the RBD's receptor binding ridge (e.g., L455, F456, G485, F486, and F490) can also have substantial antigenic effects. Another major epitope is centered on the loop formed by residues 443–450 in the RBD's receptor-binding motif, and mutations in this



mutation	cumulative prevalence
4 most frequent mutations	
S477N	5.69%
N439K	1.49%
N501Y	1.39%
Y453F	0.36%
mutations at key sites	
E484K	0.11%
K417N	0.07%
S494P	0.06%
L452R	0.02%
G446V	0.02%
F490S	0.02%
L452M	0.02%
L455F	0.02%
E484Q	0.02%
F486L	0.01%
G485R	0.01%

Figure 5.6: **Frequencies of mutations that affect plasma antibody binding among circulating SARS-CoV-2 isolates**. **(A)** Effects of mutations at each RBD site on plasma antibody binding versus frequency of mutations at each site among all SARS-CoV-2 sequences in GISAID as of Dec. 23, 2020. Key sites (see Methods) are labeled and colored according to epitope region as in Figure 5.2. **(B)** Cumulative prevalence for the 4 most frequent mutations and also any mutations at sites labeled in (A) with at least 10 counts in GISAID. **(C)** Surface representations of the RBD (PDB 6M0J). Sites where mutations have a strong effect on binding, have circulating variation with > 50 total counts in GISAID, or both, are colored in olive, pink, or maroon, respectively. See Methods for precise description of highlighted sites. ACE2 is shown as a dark gray cartoon.

epitope sometimes strongly affect plasma antibody neutralization. A third epitope is in the core of the RBD distal from the receptor-binding motif, although mutations here tend to have smaller effects on plasma antibody binding. Notably, RBD mutations reported by other studies to have large effects on plasma neutralization are also in the epitope centered around E484 or in the 443–450 loop [11, 182, 315, 320, 324].

While the major plasma epitopes are targeted by many characterized monoclonal antibodies [16, 19, 124, 129, 278], there are also sites where mutations that escape monoclonal antibodies have little effect on plasma antibody binding for any sample we tested. For instance, mutations in the S309 epitope footprint [230] and at sites of escape from antibody C135 (e.g., R346 and N440) [16, 320] had minimal effects on plasma antibody binding (Figures 3, S3). This lack of concordance between the epitopes of plasma and monoclonal antibodies is consistent with other studies reporting that the specificities of potent monoclonal antibodies often do not recapitulate the plasma from which they were isolated [16, 320]. These antibodies may be rare in polyclonal plasma or the epitopes they target may be subdominant [229]. However, subdominant epitopes may become more important as SARS-CoV-2 evolves: after mutations at immunodominant sites such as E484 partially erode plasma antibody neutralization, the remaining neutralization is presumably due to antibodies targeting previously subdominant epitopes.

Another key finding is that there is extensive person-to-person variation in how mutations affect plasma antibody binding and neutralization. For instance, the neutralizing activity of several samples was reduced by $>$ 10-fold by single mutations to site E484, but a few samples were essentially unaffected by E484 mutations. Similarly, mutations at sites in the 443–450 loop (e.g., G446V) caused a large drop in plasma antibody binding and neutralization for some samples, but had little effect on others. This inter-individual heterogeneity is further compounded by the fact that the effects of mutations sometimes changed over time for samples longitudinally collected from the same individual. These temporal changes could be

due to a disproportionate decay in one dominant antibody clonotype, or a relative increase in antibodies targeting other epitopes [106].

There are several limitations to our study. Most importantly, we only examined mutations to the RBD. While we and others [229, 282] have shown that RBD-binding antibodies contribute the majority of the plasma neutralizing activity of most convalescent human sera and plasma, antibodies also target other regions of the spike. For example, mutations and deletions in the NTD can affect plasma antibody neutralization [11, 207, 189, 202, 303]. In addition, we only mapped samples from 11 individuals at two time points. Given the substantial inter- and intra-individual heterogeneity, mapping more samples may identify additional sites of importance. On a technical level, we assayed binding of antibodies to isolated RBD expressed by yeast, which implies several limitations. First, we are unable to map the effects of mutations that alter the spike's overall conformation or affect antibodies spanning quaternary epitopes [16]. Second, our mapping likely overestimates the contributions of antibodies that bind epitopes that are more accessible on isolated RBD than in the context of full spike (e.g., F456). Finally, the N-linked glycans on yeast-expressed proteins are more mannose-rich than those on mammalian-expressed proteins [127], which could affect measurements of how N-linked glycans affect antibody binding. However, the general consistency of our mapping with our pseudovirus neutralization assays and the plasma-escape mutations reported by others suggest that our study successfully defines the major RBD epitopes of convalescent human plasma antibodies.

The comprehensive nature of our mapping makes it possible to assess which circulating RBD mutations are likely to have the greatest impact on human immunity. In particular, our results predict that the emerging 20H/501Y.V2 and 20J/501Y.V3 viral lineages (originally identified South Africa and Brazil, respectively) carrying the E484K mutation will have reduced susceptibility to neutralization by the polyclonal plasma antibodies of some individuals. In contrast, the N501Y mutation present in the 20I/501Y.V1 lineage (B.1.1.7) is

unlikely to greatly affect neutralization by most human plasma, although it could contribute to increased viral titer or enhanced transmissibility [156, 237]. The NTD deletions in this lineage, however, may have an antigenic effect [11, 202, 207]. Notably, very recent studies on viruses from these lineages largely confirm these predictions from our mapping [42, 324, 248]. More generally, our mapping can be used to assess the likely antigenic impacts of additional viral mutations that emerge in the future.

Our mapping also reveals broader features of antibody immunity that are relevant to SARS-CoV-2 evolution. One reason that influenza virus undergoes such rapid antigenic evolution is that neutralizing human immunity often focuses on just a few residues in hemagglutinin, such that a single mutation can dramatically reduce neutralization [176]. In contrast, antibody immunity to measles virus targets multiple co-dominant measles epitopes, meaning that no single mutation has a large effect on neutralization [217]. Our results show that polyclonal antibody immunity to the SARS-CoV-2 RBD is sometimes focused as for influenza, but in other cases more broadly targets the RBD in a way that mitigates the effect of any single mutation. This heterogeneity in the antigenic impacts of RBD mutations implies that the immunity of different individuals will be impacted differently by viral evolution. It also suggests that an important area for future work is understanding how viral mutations impact vaccine-elicited immunity, and using this knowledge to design vaccines that are robust to viral antigenic evolution.

5.5 Materials and Methods

Data and Code Availability

We provide data and code in the following ways:

- The complete code for the full computational data analysis pipeline of the mapping experiments is available on GitHub at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera`.

- The escape fraction measured for each mutation are at `https://raw.githubusercontent.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/main/results/supp_data/human_sera_raw_data.csv`.

- All raw sequencing data are available on the NCBI Short Read Archive at BioProject PRJNA639956, BioSample SAMN17185313.

SARS-CoV-2 convalescent human plasmas

Plasma samples were previously described ([66]) and collected as part of a prospective longitudinal cohort study of individuals with SARS-CoV-2 infection in Seattle, WA February–July 2020. See `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/supplementary_tables/TableS1.csv` for the sample metadata, which is also described in ([66]). That table also links the sample IDs used in ([66]) to the names used for the plasmas in this paper. All plasmas were heat-inactivated prior to use by treatment at 56 degrees C for 60 minutes. Prior to use in each assay, plasma samples were centrifuged for 15 min at 2000 $\times g$ to pellet platelets.

RBD deep mutational scanning library

The yeast-display RBD mutant libraries are previously described ([124, 280]). Briefly, duplicate mutant libraries were constructed in the spike receptor binding domain (RBD) from SARS-CoV-2 (isolate Wuhan-Hu-1, Genbank accession number MN908947, residues N331–T531) and contain 3,804 of the 3,819 possible amino-acid mutations, with ¿95% present as single mutants. Each RBD variant was linked to a unique 16-nucleotide barcode sequence to facilitate downstream sequencing. As previously described, libraries were sorted for RBD expression and ACE2 binding to eliminate RBD variants that are completely misfolded or non-functional (i.e., lacking modest ACE2 binding affinity) ([124]).

FACS sorting of yeast libraries to select mutants with reduced binding by polyclonal plasmas

Plasma mapping experiments were performed in biological duplicate using the independent mutant RBD libraries, as previously described for monoclonal antibodies ([124]), with the following modifications: Mutant yeast libraries induced to express RBD were washed and incubated with plasma at a range of dilutions for 1 h at room temperature with gentle agitation. For each plasma, we chose a sub-saturating dilution such that the amount of fluorescent signal due to plasma antibody binding to RBD was approximately equal across plasma. The exact dilution used for each plasma is given in `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/supplementary_tables/TableS2.xlsx`. After the plasma incubations, the libraries were secondarily labeled with 1:100 FITC-conjugated anti-MYC antibody (Immunology Consultants Lab, CYMC-45F) to label for RBD expression and 1:200 Alexa-647- or DyLight-405-conjugated goat anti-human-IgA+IgG+IgM (Jackson ImmunoResearch 109-605-064 or 109-475-064, respectively) to label for bound plasma antibodies. A flow cytometric selection gate was drawn to capture 3–6% of the RBD mutants with the lowest amount of plasma binding for their degree of RBD expression (Figure 5.7A-C). We also measured what fraction of cells expressing unmutated RBD fell into this gate when stained with 1x and 0.1x the concentration of plasma. For each sample, approximately 10 million RBD+ cells (range 7.4e6 to 1.7e7 cells) were processed on the cytometer, with between 2e5 and 8e5 plasma-escaped cells collected per sample (see percentages at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/supplementary_tables/TableS2.xlsx`). Antibody-escaped cells were grown overnight in SD-CAA (6.7g/L Yeast Nitrogen Base, 5.0g/L Casamino acids, 1.065 g/L MES acid, and 2% w/v dextrose) to expand cells prior to plasmid extraction.

DNA extraction and Illumina sequencing

Plasmid samples were prepared from 30 OD units (1.6e8 cfus) of pre-selection yeast populations and approximately 5 OD units (3.2e7 cfus) of overnight cultures of plasma-escaped cells (Zymoprep Yeast Plasmid Miniprep II) as previously described ([124]). The 16-nucleotide barcode sequences identifying each RBD variant were amplified by PCR and prepared for Illumina sequencing as described in ([280]). Barcodes were sequenced on an Illumina HiSeq 3500 with 50 bp single-end reads. To minimize noise from inadequate sequencing coverage, we ensured that each antibody-escape sample had at least 2.5x as many post-filtering sequencing counts as FACS-selected cells, and reference populations had at least 2.5e7 post-filtering sequencing counts.

Analysis of deep sequencing data to compute each mutation's plasma escape fraction

Escape fractions were computed as described in ([124]), with minor modifications as noted below. We used the dms_variants package (`https://jbloomlab.github.io/dms_variants/`, version 0.8.2) to process Illumina sequences into counts of each barcoded RBD variant in each pre-sort and antibody-escape population using the barcode/RBD look-up table from [280].

For each plasma selection, we computed the "escape fraction" for each barcoded variant using the deep sequencing counts for each variant in the original and plasma-escape populations and the total fraction of the library that escaped antibody binding via the formula provided in ([124]). These escape fractions represent the estimated fraction of cells expressing that specific variant that fall in the plasma escape bin, such that a value of 0 means the variant is always bound by plasma and a value of 1 means that it always escapes plasma binding. We then applied a computational filter to remove variants with low sequencing counts or highly deleterious mutations that might cause antibody escape simply by leading to poor expression of properly folded RBD on the yeast cell surface ([124, 280]). Specifically,

we removed variants that had (or contained mutations with) ACE2 binding scores < -2.35 or expression scores < -1, using the variant- and mutation-level deep mutational scanning scores from ([280]). Note that these filtering criteria are slightly more stringent than those used in ([124]) but are identical to those used in [278].

We next deconvolved variant-level escape scores into escape fraction estimates for single mutations using global epistasis models [220] implemented in the dms_variants package, as detailed at (`https://jbloomlab.github.io/dms_variants/dms_variants.globalepistasis.html`) and described in ([124]). The reported scores throughout the paper are the average across the libraries; these scores are also in `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/supplementary_tables/TableS3.csv`. Correlations in final single-mutant escape scores are shown in Figure 5.8D,E.

For plotting and analyses that required identifying RBD sites of "strong escape" (e.g., choosing which sites to show in logo plots in Figures 5.2A,B, and 5.10 or label in Figure 5.4B), we considered a site to mediate strong escape if the total escape (sum of mutation-level escape fractions) for that site exceeded the median across sites by $>$ 10-fold, and was at least 10% of the maximum for any site. We also included site K417, which did not meet this threshold but was of interest due to its frequency among circulating viruses.

Full documentation of the computational analysis is at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera`.

Generation of pseudotyped lentiviral particles

We used spike-pseudotyped lentiviral particles that were generated essentially as described in [67], using a codon-optimized SARS-CoV-2 spike from Wuhan-Hu-1 that contains a 21-amino-acid deletion at the end of the cytoplasmic tail ([66]) and the D614G mutation that is now predominant in human SARS-CoV-2 [161]. The plasmid encoding this spike, HDM_Spikedelta21_D614G, is available from Addgene (#158762), and the full sequence is

at (`https://www.addgene.org/158762`). Point mutations were introduced into the RBD of this plasmid via site-directed mutagenesis. Therefore, all mutations tested in this paper are in the G614 background, and are compared to a "wildtype" spike with G614. The only exception is the C432D mutation, which was made in the D614 background and is included in Figure 5.11D to show a baseline titer for lentiviral particles pseudotyped with a non-functional spike variant.

To generate these spike-pseudotyped lentiviral particles ([67]), 6e5 293T cells per well were seeded in 6-well plates in 2 mL D10 growth media (DMEM with 10% heat-inactivated FBS, 2 mM l-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin). 24h later, cells were transfected using BioT transfection reagent (Bioland Scientific, Paramount, CA, USA) with a Luciferase_IRES_ZsGreen backbone, Gag/Pol lentiviral helper plasmid, and wildtype or mutant SARS-CoV-2 spike plasmids. Media was changed to fresh D10 at 24 h post-transfection. At 60 h post-transfection, viral supernatants were collected, filtered through a 0.45 μm SFCA low protein-binding filter, and stored at -80 °C.

Titering and p24 ELISAs of pseudotyped lentiviral particles

Titers of spike-pseudotyped lentiviral particles were determined as described in ([67]) with the following modifications. Spike-pseudotyped lentiviral supernatants were diluted in D10 growth media starting with a 1:10 dilution followed by 7 serial 2-fold dilutions. 100 μL of each dilution was added to 1.25e4 293T-ACE2 cells (BEI NR-52511) grown overnight in 50 μL of D10 growth media in a 96-well black-walled poly-L-lysine coated plate (Greiner Bio-One, 655936). Relative luciferase units (RLU) were measured 60 h post-infection (Promega Bright-Glo, E2620) in the infection plates with a black back-sticker (Fisher Scientific, NC9425162) added to minimize background. Titers in RLU per mL were calculated for each dilution, and the median of each technical replicate normalized to p24 concentration (in pg/mL) is plotted in Figure 5.11D.

p24 ELISAs were conducted following the manufacturers instructions (Advanced Bioscience Laboratories Cat. #5421). All lentiviral supernatants were diluted 1:100,000 and measured in technical duplicate. In Figure 5.11D, technical duplicate p24 concentrations (in pg/mL) were averaged and used to normalize RLU values for each lentiviral supernatant preparation.

Neutralization assays

293T-ACE2 cells (BEI NR-52511) were seeded at 1.25e4 cells per well in 50 μL D10 in poly-L-lysine coated, black-walled, 96-well plates (Greiner 655930). 24 h later, pseudotyped lentivirus supernatants were diluted to approximately 200,000 RLU per well (determined by titering as described above and shown in Figure 5.11D and incubated with a range of dilutions of plasma for 1 h at 37 °C. 100 μL of the virus-antibody mixture was then added to cells. The following dilutions were used for each virus: WT 1:50; P384L 1:20; F456K 1:10; E484Q 1:20; G485R 1:32; S494P 1:50; G446V/F456V/E484P 1:6; F456V 1:40; E484P 1:6; G446V 1:25; E484K 1:50; F456A 1:6.

At 70 h post-infection, luciferase activity was measured using the Bright-Glo Luciferase Assay System (Promega, E2610). Fraction infectivity of each plasma antibody-containing well was calculated relative to a "no-plasma" well inoculated with the same initial viral supernatant (containing wildtype or mutant RBD) in the same row of the plate. We used the neutcurve package (`https://jbloomlab.github.io/neutcurve`) to calculate the inhibitory concentration 50% (IC50) and the neutralization titer 50% (NT50), which is simply 1/IC50, of each plasma against each virus by fitting a Hill curve with the bottom fixed at 0 and the top fixed at 1. The full neutralization curves are in Figure 5.11.

Depletion of RBD-binding antibodies from polyclonal plasma

Magnetic beads conjugated to the SARS-CoV-2 RBD (AcroBiosystems, MBS-K002) were prepared according to the manufacturer's protocol. Beads were resuspended in ultrapure water at 1 mg beads/mL and a magnet was used to wash the beads 3 times in PBS with 0.05% BSA. Beads were then resuspended in PBS with 0.05% BSA at 1 mg beads per mL. Beads (manufacturer-reported binding capacity of 10-40 μg/mL anti-RBD antibodies) were incubated with human plasma at a 3:1 ratio beads:plasma (150 μL beads + 50 μL plasma), rotating overnight at 4°C. A magnet was used to separate antibodies that bind RBD from the supernatant, and the supernatant (the post-RBD antibody depletion sample) was removed. A mock depletion (pre-depletion sample) was performed by adding 150 μL of PBS + 0.05% BSA and incubating rotating overnight at 4°C. For the neutralization assays on these plasma depleted of RBD-binding antibodies shown in Figure 5.7E; the reported plasma dilution is corrected for the dilution incurred by the depletion process.

Measurement of plasma binding to RBD or spike by ELISA

The IgG ELISAs for spike protein and RBD were conducted as previously described [82]. Briefly, ELISA plates were coated with recombinant spike and RBD antigens described in [82] at 2 μg/mL. Five 3-fold serial dilutions of plasma beginning at 1:100 were performed in phosphate-buffered saline with 0.1% Tween in 1% (w/v) nonfat dry milk. Dilution series of the "synthetic" sera comprised of the anti-RBD antibody REGN10987 [129] or anti-NTD antibody 4A8 [56] and pooled pre-pandemic human plasma from 2017–2018 (Gemini Biosciences; nos. 100–110, lot H86W03J; pooled from 75 donors) were performed such that the anti-spike antibody was present at a highest concentration of 0.25 μg/mL. Both antibodies were recombinantly produced by Genscript. The REGN10987 is that used in [278] and the variable domain heavy and light chain sequences for 4A8 were obtained from Genbank GI 1864383732 and 1864383733 [56] and produced on a human IgG1 and IgK background, re-

spectively. Pre-pandemic plasma alone, without anti-RBD antibody depletion, was used as a negative control, averaged over 2 replicates. The area under the curve (AUC) was calculated as the area under the titration curve with the serial dilutions on a log-scale.

Analysis of RBD mutations among circulating SARS-CoV-2 isolates

All 283,908 spike sequences on GISAID as of Dec-23-2020 were downloaded and aligned via mafft [151]. Sequences from non-human origins and sequences containing gap characters or excessive mutations were removed, leaving 263,217 sequences. The code that performs this alignment and filtering is at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/results/summary/gisaid_rbd_mutations.md`. The counts and frequencies of mutations at each RBD site were then computed using this filtered sequence set. We acknowledge all GISAID contributors for sharing sequencing data (`https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/data/gisaid_hcov-19_acknowledgement_table_2020_12_30.pdf`).

Sites and mutations highlighted in Figure 5.6 were chosen as follows. Sites in the RBM containing the 4 RBD mutations with the highest cumulative frequency (S477N, N439K, N501Y, and Y453F), the two sites with the highest total escape (F456 and E484), and sites that have ≥ 30 variant counts in GISAID and are sites of strong escape for any plasma, are labeled in Figure 5.6A. The labeled sites are colored according to epitope region as in Figure 5.2. Figure 5.6B highlights the 4 most frequent mutations and also any mutations at the other sites labeled in Figure 5.6A with at least 10 or more counts in GISAID. Figure 5.6C highlights sites where mutations have a strong effect on binding of at least 1 plasma or have circulating variation with > 50 counts in GISAID. Site K417 was also of interest due to the presence of the K417N/T mutations in recently identified lineages [94, 285], and thus is also highlighted in each panel in Figure 5.6.

Data visualization

The static logo plot visualizations of the escape maps in the paper figures were created using the dmslogo package (`https://jbloomlab.github.io/dmslogo`, version 0.3.2) and in all cases the height of each letter indicates the escape fraction for that amino-acid mutation calculated as described above. For each plasma, the y-axis is scaled to be the greatest of (a) the maximum site-wise escape metric observed for that plasma, (b) 20x the median site-wise escape fraction observed across all sites for that plasma, or (c) an absolute value of 1.0 (to appropriately scale plasma that are not "noisy" but for which no mutation has a strong effect on plasma binding). Site C361 has been removed from the plots, because while mutations at this site reduce plasma binding, these mutations ablate a disulfide bond in the core RBD that is important for proper folding of the RBD and likely result in a grossly misfolded RBD and do not represent specific plasma escape mutations. Sites N501 and K417 have been added to Figure 5.9 due to their frequencies among circulating viruses. The code that generates these logo plot visualizations is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/results/summary/escape_profiles.md`.

In many of the visualizations (e.g., Figures 5.2, 5.4, 5.6A, 5.9, and 5.10), the RBD sites are categorized by epitope region (core-RBD epitope, receptor-binding ridge, or 443–450 loop) and colored accordingly. We define the core-RBD epitope as residues 365–372+382–386, the receptor-binding ridge epitope to be residues 417+455+456+471–490, and the 443–450 loop epitope to be residues 443–452+494–501. In Figure 5.4E, the S309 epitope is defined as RBD non-hydrogen atoms within 4Å of antibody in PDB 6WPS, [230].

For the static structural visualizations in the paper figures, the RBD surface (PDB 6M0J, [172]) was colored by the site-wise escape metric at each site, with white indicating no escape and red scaled to be the same maximum used to scale the y-axis in the logo plot escape maps, determined as described above.We created interactive structure-based visualizations of the escape maps using dms-view [137] that are available at `https://jbloomlab.github.io/`

`SARS-CoV-2-RBD_MAP_HAARVI_sera`. The logo plots in these escape maps can be colored according to the deep mutational scanning measurements of how mutations affect ACE2 binding or RBD expression as described above.

5.6 Notes

Acknowledgments

We thank Adam Dingens for experimental assistance; the Flow Cytometry and Genomics core facilities at the Fred Hutchinson Cancer Research Center for experimental support, especially Dolores Covarrubias and Andy Marty; Neil King, Alexandra Walls, David Veesler, and the UW Institute for Protein Design for purified RBD and spike proteins; and Seth Zost and James Crowe of Vanderbilt University for monoclonal antibodies. We also thank all research participants in the Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) study for their generous participation and all HAARVI study researchers and staff, especially Caitlin Wolf.

Funding

This work was supported by the NIAID / NIH (R01AI141707 and R01AI127893 to J.D.B., T32AI083203 to A.J.G., and F30AI149928 to K.H.D.C.) and the Gates Foundation (INV-004949). The Scientific Computing Infrastructure at Fred Hutch is funded by ORIP grant S10OD028685. T.N.S. is a Washington Research Foundation Innovation Fellow at the University of Washington Institute for Protein Design and a Howard Hughes Medical Institute Fellow of the Damon Runyon Cancer Research Foundation (DRG-2381-19). J.D.B. is an Investigator of the Howard Hughes Medical Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the US government or the other sponsors.

Author contributions

: Conceptualization, A.J.G., H.Y.C., and J.D.B.; Methodology, A.J.G., K.H.D.C., T.N.S., and J.D.B.; Investigation, A.J.G., A.L., K.H.D.C., and K.M.; Code, A.J.G., T.N.S., and J.D.B.; Formal Analysis, A.J.G. and J.D.B.; Validation, A.J.G., A.N.L., and K.H.D.C.; Resources, H.Y.C.; Writing – Original Draft, A.J.G. and J.D.B.; Writing – Review and Editing, all authors; Supervision, H.Y.C. and J.D.B.

Declaration of Interests

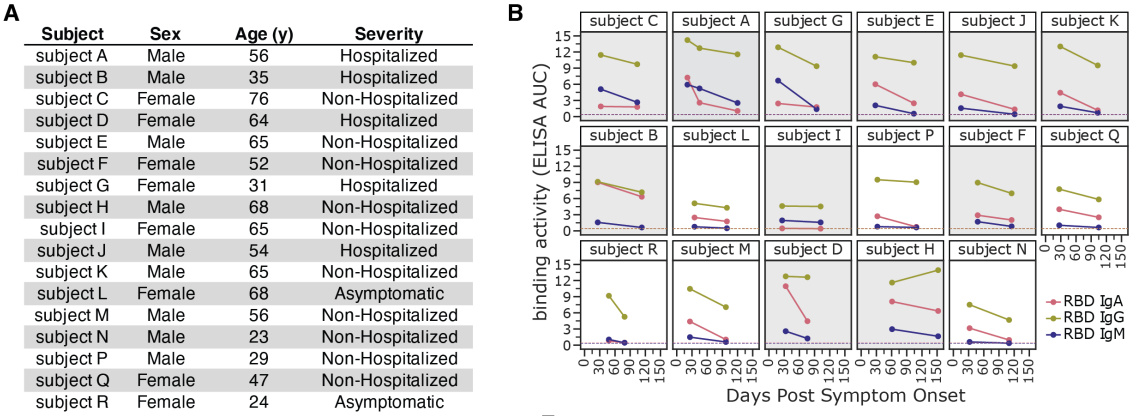
H.Y.C. is a consultant for Merck, Pfizer, Ellume, Bill and Melinda Gates Foundation and has received support from Cepheid and Sanofi-Pasteur. The other authors declare no competing interests.

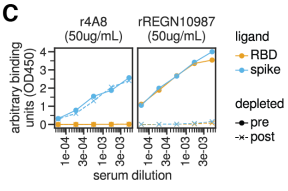
5.7 Supplementary Figures

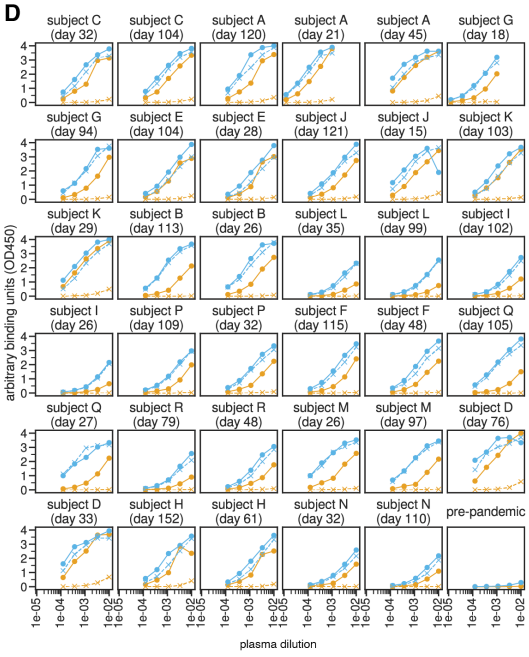
Figure 5.7: **Raw ELISA and neutralization curves of plasma pre- and post-depletion of RBD-targeting antibodies**. **(A)** Participant sex, age (y), and disease severity. **(B)** Previously measured RBD binding for three antibody isotypes for these plasma samples as measured by ELISA area under the curve (AUC), taken from [66]. Gray background indicates plasma for which we subsequently mapped mutations that reduce binding. Similar data across additional time points not used in the current study are available in [66]; see `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/supplementary_tables/TableS1.csv` to map between the sample IDs used in the current study and [66]. **(C)** Effect of RBD antibody depletion on binding to RBD and spike by "synthetic sera" comprised of pre-pandemic pooled plasma with the NTD-targeting antibody 4A8 [56] or RBD-targeting antibody REGN10987 [129]. Antibodies were added to pre-pandemic plasma at 50 μg/mL. The x-axis indicates the dilution factor of the plasma/antibody mix, and the y-axis is the ELISA reading at each dilution. **(D)** Raw ELISA binding curves of plasma to RBD and spike before and after depletion of RBD-binding antibodies. Legend for panels (C) and (D): orange is RBD binding, blue is spike binding; filled circles with solid lines represent pre-depletion, and x's with dashed lines represent post-depletion of anti-RBD antibodies. **(E)** Raw neutralization curves for plasma before (gray) and after (orange) depletion of RBD-binding antibodies. Neutralization assays were performed with lentiviruses pseudotyped with spike D614G, the predominant SARS-CoV-2 circulating variant. **(F)** Change in the amount of neutralizing activity that is due to RBD-binding antibodies over time for each individual. Each point gives the fold-change in neutralization inhibitory concentration 50% (IC50) post- versus pre-depletion for plasma isolated at the indicated time, such that larger values indicate that more of the neutralizing activity is due to RBD-binding antibodies. Open circles represent samples for which the post-depletion NT50 was at the limit of detection, i.e., less than 20 (see Figure 5.1B; these circles are therefore lower bounds in the fold-change in IC50. **(G)** Correlation between previously measured neutralization titers 50% (NT50) with spike D614-spike-pseudotyped lentivirus [66] and pre-depletion neutralization titers measured with G614-spike-pseudotyped lentivirus (present study), Pearson's R = 0.88.

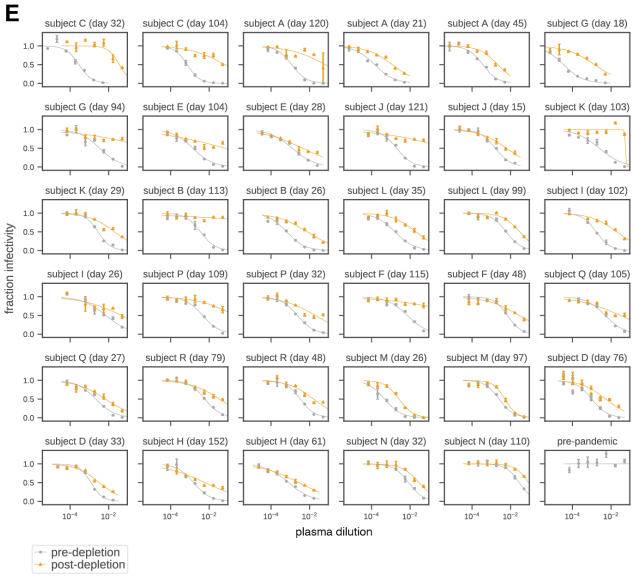
A

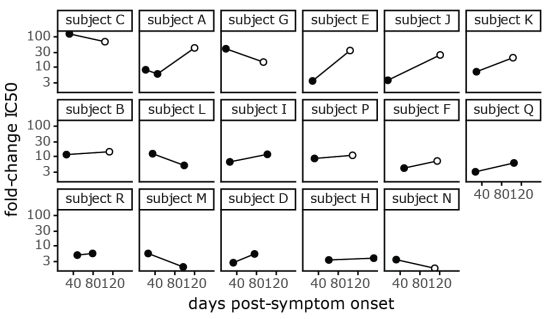
Subject	Sex	Age (y)	Severity
subject A	Male	56	Hospitalized
subject B	Male	35	Hospitalized
subject C	Female	76	Non-Hospitalized
subject D	Female	64	Hospitalized
subject E	Male	65	Non-Hospitalized
subject F	Female	52	Non-Hospitalized
subject G	Female	31	Hospitalized
subject H	Male	68	Non-Hospitalized
subject I	Female	65	Non-Hospitalized
subject J	Male	54	Hospitalized
subject K	Male	65	Non-Hospitalized
subject L	Female	68	Asymptomatic
subject M	Male	56	Non-Hospitalized
subject N	Male	23	Non-Hospitalized
subject P	Male	29	Non-Hospitalized
subject Q	Female	47	Non-Hospitalized
subject R	Female	24	Asymptomatic











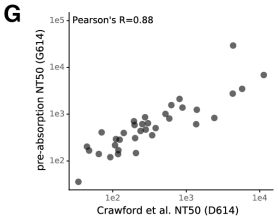
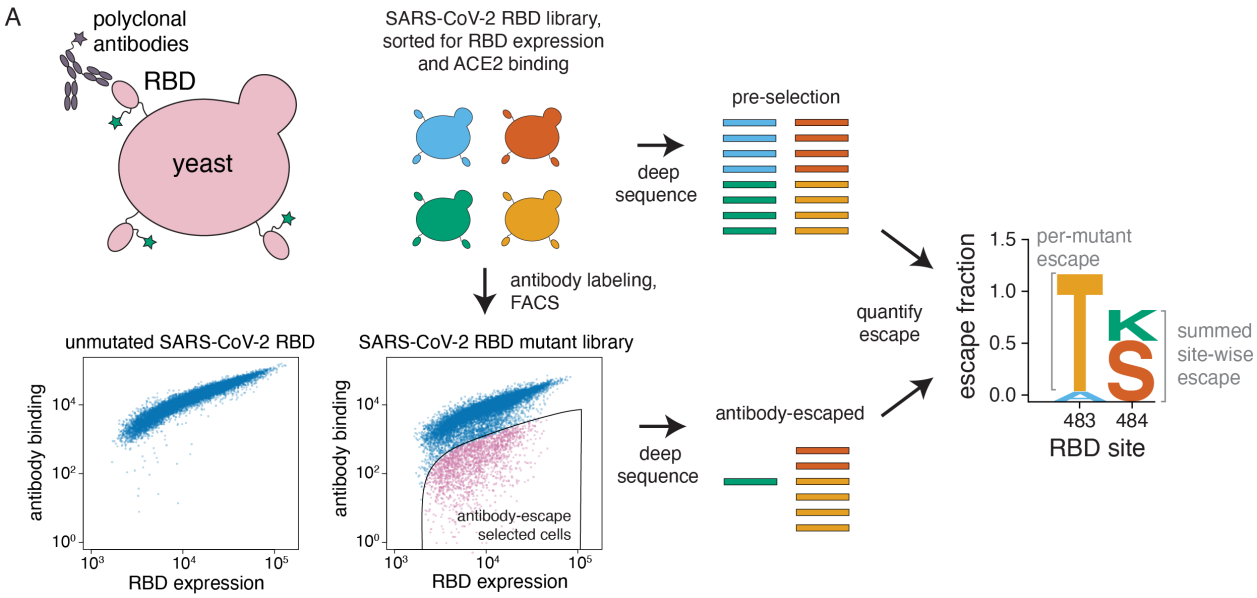
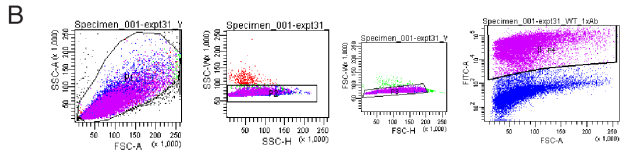
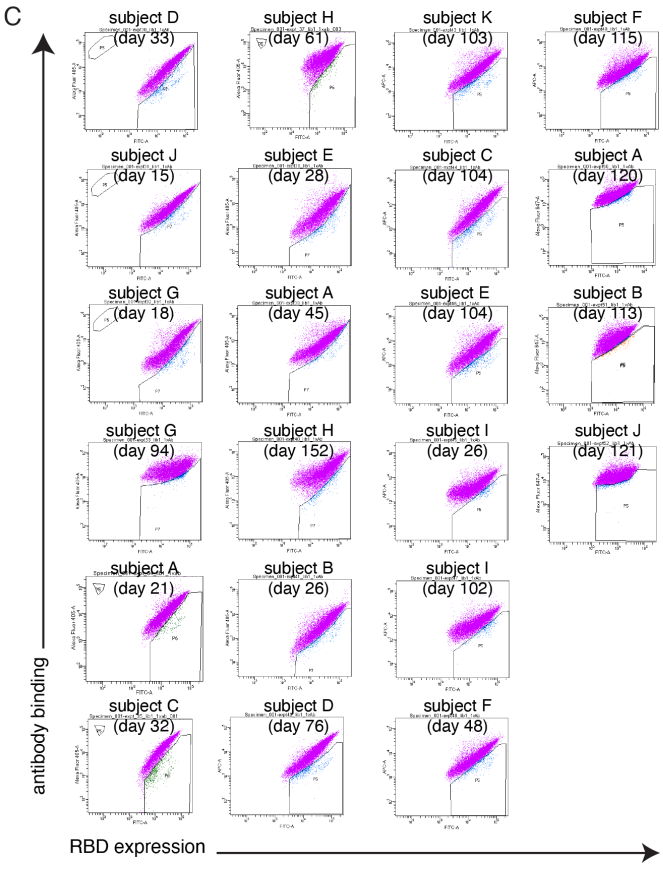
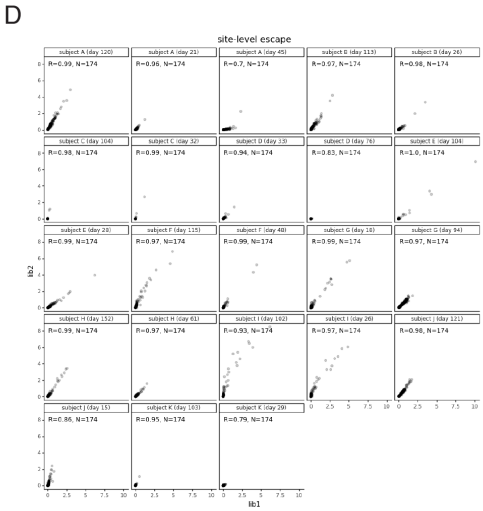


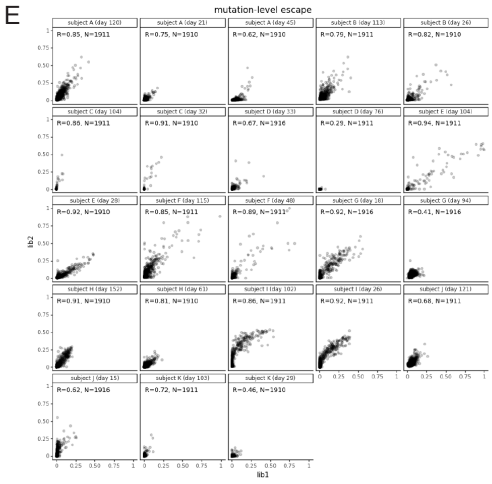
Figure 5.8: **Approach for mapping RBD mutations that reduce binding by polyclonal plasma**. **(A)** The RBD is expressed on the surface of yeast. Flow cytometry can be used to quantify both RBD expression (via a C-terminal MYC tag) and antibody binding to the RBD protein expressed on the surface of each yeast cell. A library of yeast expressing different RBD mutants were incubated with polyclonal plasma and plasma antibody binding was detected using a IgA+IgG+IgM secondary antibody. We then used FACS to enrich for cells expressing RBD that bound reduced levels of antibody, and used deep sequencing to quantify the frequency of each mutation in the initial and "antibody escape" cell populations. We quantified the effect of each mutation as the "escape fraction," which represents the fraction of cells expressing RBD with that mutation that fell in the "antibody escape" FACS bin. Escape fractions are represented in logo plots, with the height of each letter proportional to the effect of that amino acid mutation on antibody binding. The site-level escape metric is the sum of the escape fractions of all mutations at a site. Note that both experimental and computational filtering steps were used to remove RBD mutants that were misfolded or completely unable to bind the ACE2 receptor (see Methods). **(B)** Representative plots of nested FACS gating strategy used for all plasma selection experiments to select for single cells (SSC-A vs. FSC-A, and FSC-W vs. FSC-H) that also express RBD (FITC-A vs. FSC-A). **(C)** FACS gating strategy for one of two independent libraries to select cells expressing RBD mutants with reduced binding by polyclonal plasma. Gates were set manually during sorting, aiming for 3-6% of the RBD+ library to fall into the selection gate (cells in blue). The same gate was set for both independent libraries stained with each plasma, and the FACS scatter plots looked qualitatively similar between the two libraries. For information on the fraction of library cells that fall into each selection gate, see `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/supplementary_tables/TableS2.xlsx`. **(D)** Correlation plots of site-level escape for each of the two independent RBD mutant libraries for each plasma. Site-level escape is the sum of escape fraction for each mutation at a site. **(E)** Correlation plots of mutation-level escape for each of the two independent RBD mutant libraries for each plasma.


A
polyclonal antibodies
RBD
yeast
SARS-CoV-2 RBD library, sorted for RBD expression and ACE2 binding
pre-selection
deep sequence
antibody labeling, FACS
quantify escape
unmutated SARS-CoV-2 RBD
SARS-CoV-2 RBD mutant library
antibody binding
RBD expression
antibody-escaped
antibody-escape selected cells
escape fraction
per-mutant escape
summed site-wise escape
RBD site
483
484


B


C
subject D (day 33)
subject H (day 61)
subject K (day 103)
subject F (day 115)
subject J (day 15)
subject E (day 28)
subject C (day 104)
subject A (day 120)
subject G (day 18)
subject A (day 45)
subject E (day 104)
subject B (day 113)
subject G (day 94)
subject H (day 152)
subject I (day 26)
subject J (day 121)
subject A (day 21)
subject B (day 26)
subject I (day 102)
subject C (day 32)
subject D (day 76)
subject F (day 48)
antibody binding
RBD expression


D
site-level escape


E
mutation-level escape

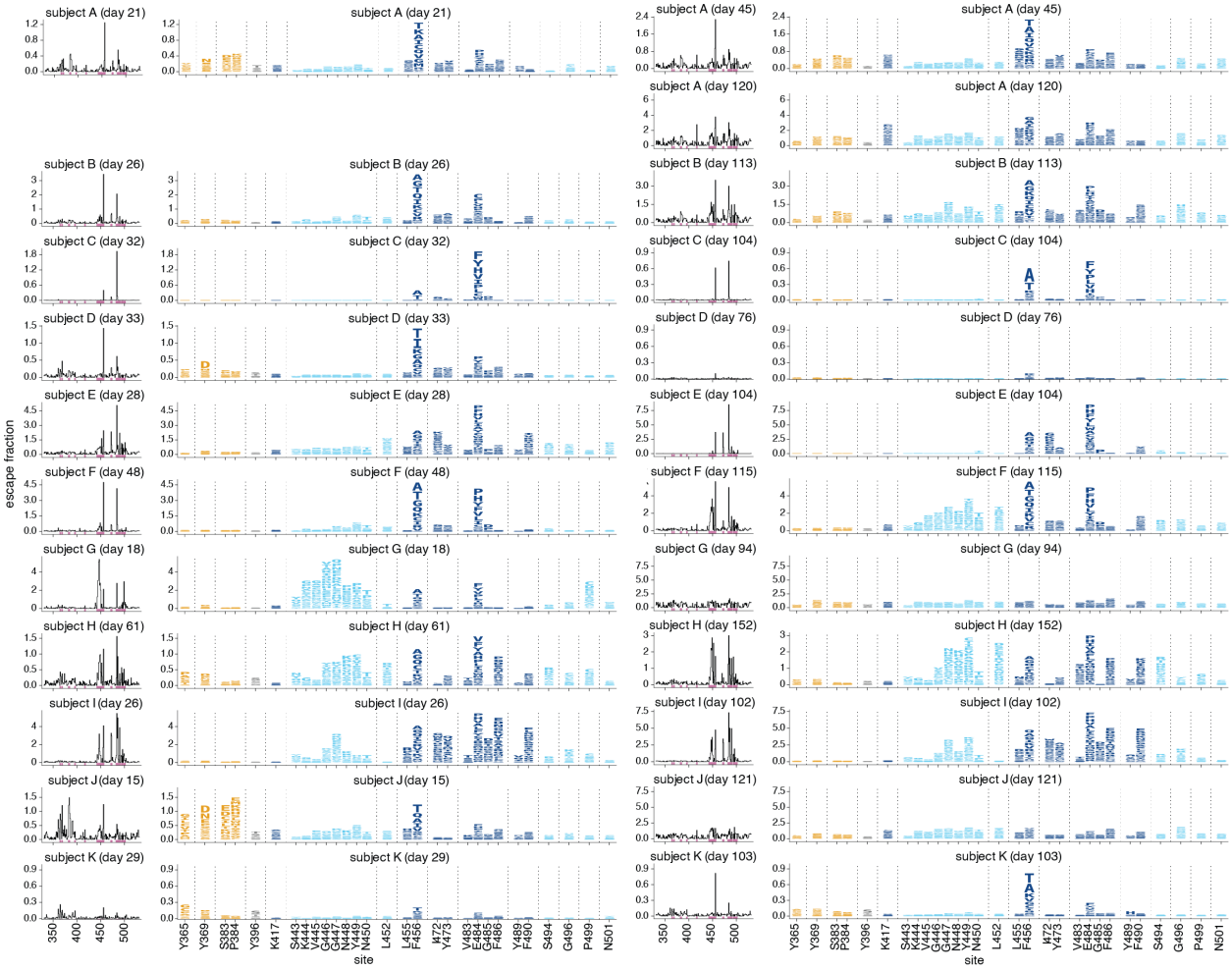


Figure 5.9: **Escape maps over time for all study subjects and time points**. Escape maps for all individuals and time points, with 2 time points shown side-by-side, ordered as in Figure 5.2. Escape fractions are comparable across sites within a sample, but not necessarily between samples due to the use of sample-specific FACS gates—therefore, for each sample, the y-axis is scaled independently (see Methods). Sites are colored by RBD epitope region as in Figure 5.2. Sites shown in logo plots, highlighted in purple in line plots at left, are sites of strong escape for any of the 23 plasma, plus sites K417 and N501. Interactive versions of these escape maps are available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_HAARVI_sera/`.

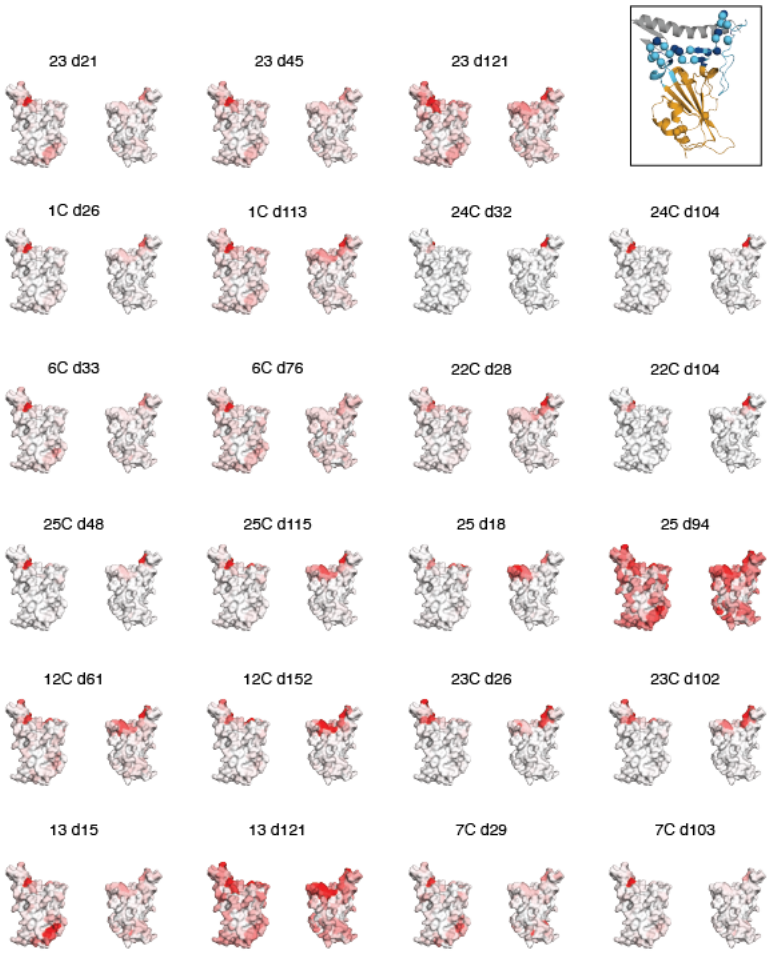
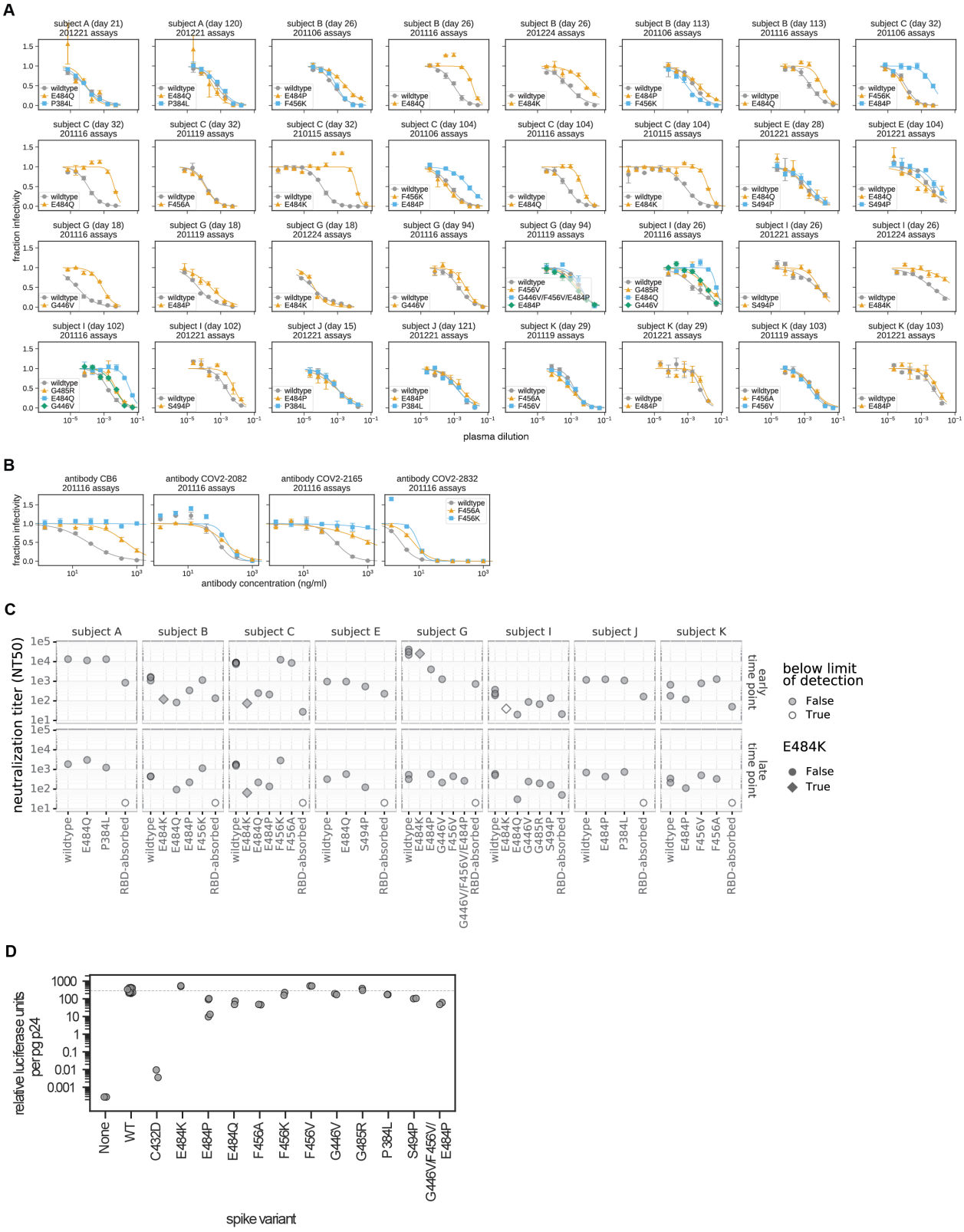


Figure 5.10: **Regions in the RBD where mutations reduce binding by plasma antibodies for all study subjects and samples over time**. The structures show the effects of mutations at each site projected onto the RBD structure using a white-to-red color scale as in 5.3A-D. The color scale for each plasma is scaled to span the same range as the y-axis for that plasma in Figure 5.9. Top right inset: the alpha-carbon of any site of strong escape (all sites shown in the logo plots in Figure 5.9) is shown as a sphere on a cartoon representation of the RBD (PDB 6M0J). The RBD is colored as in Figure 5.2B. Interactive versions of these structural visualizations are available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_HAARVI_sera/`.

Figure 5.11: **Full curves for all assays testing how RBD mutations affected viral neutralization**. **(A)** The x-axis gives the plasma dilution, and the y-axis gives the fraction of viral infectivity remaining at that dilution. A different plot facet is shown for each plasma (labeled by subject and day of collection) and assay date. The neutralization curves were fit and plotted using neutcurve (`https://jbloomlab.github.io/neutcurve/`, version 0.5.1) and fitting 2-parameter Hill curves with the baselines fixed at one and zero to calculate IC50s. These IC50s were then used to determine the fold-change values plotted in Figure 5.5A-C, comparing each mutant to the wildtype run on the same assay date. Note that NT50 is the reciprocal of the IC50. The curves plotted in Figure 5.5D recapitulate data plotted in this panel, but aggregate mutants across several assay dates and show the wildtype curve for just the first assays date. This aggregation across assay dates is well supported since the wildtype was re-run on each assay date and always yielded very similar IC50s for any given plasma. **(B)** Neutralization curves for monoclonal antibodies run against mutations to F456. Our previous escape mapping showed that F456A/K mutations escape binding by the anti-SARS-CoV-2 RBD monoclonal antibodies COV2-2165 and CB6 (also known as LY-CoV016), but not by COV2-2082 or COV2-2832 [124, 271, 278, 356]. The neutralization assays shown here supported this mapping, and demonstrated that mutations at F456 can indeed greatly reduce neutralization by monoclonal antibodies. **(C)** Absolute neutralization titer (NT50) for each tested plasma and RBD mutant. The numerical IC50s from all curves in both panels are available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_HAARVI_sera/blob/main/experimental_validations/results/mutant_neuts_results/mutants_foldchange_ic50.csv`. **(D)** Viral entry titers for key RBD mutants. Titers were measured as relative luciferase units (RLU) normalized to p24 (in picograms) measured by ELISA. The median wildtype titer was 291 RLU/pg p24 or 3.82e8 RLU/mL, and is shown with a dotted horizontal line. "None" is virus-like particles with no spike protein, and C432D disrupts a critical disulfide bond and RBD folding [280].


A
subject A (day 21) 201221 assays
subject A (day 120) 201221 assays
subject B (day 26) 201106 assays
subject B (day 26) 201116 assays
subject B (day 26) 201224 assays
subject B (day 113) 201106 assays
subject B (day 113) 201116 assays
subject C (day 32) 201106 assays
subject C (day 32) 201116 assays
subject C (day 32) 201119 assays
subject C (day 32) 210115 assays
subject C (day 104) 201106 assays
subject C (day 104) 201116 assays
subject C (day 104) 210115 assays
subject E (day 28) 201221 assays
subject E (day 104) 201221 assays
subject G (day 18) 201116 assays
subject G (day 18) 201119 assays
subject G (day 18) 201224 assays
subject G (day 94) 201116 assays
subject G (day 94) 201119 assays
subject I (day 26) 201116 assays
subject I (day 26) 201221 assays
subject I (day 26) 201224 assays
subject I (day 102) 201116 assays
subject I (day 102) 201221 assays
subject J (day 15) 201221 assays
subject J (day 121) 201221 assays
subject K (day 29) 201119 assays
subject K (day 29) 201221 assays
subject K (day 103) 201119 assays
subject K (day 103) 201221 assays
fraction infectivity
plasma dilution
B
antibody CB6 201116 assays
antibody COV2-2082 201116 assays
antibody COV2-2165 201116 assays
antibody COV2-2832 201116 assays
fraction infectivity
antibody concentration (ng/ml)
C
subject A
subject B
subject C
subject E
subject G
subject I
subject J
subject K
neutralization titer (NT50)
early time point
late time point
below limit of detection
False
True
E484K
False
True
D
relative luciferase units per pg p24
spike variant
None
WT
C432D
E484K
E484P
E484Q
F456A
F456K
F456V
G446V
G485R
P384L
S494P
G446V/F456V/E484P

Chapter 6

MAPPING MUTATIONS TO THE SARS-COV-2 RBD THAT ESCAPE BINDING BY DIFFERENT CLASSES OF ANTIBODIES

A version of this chapter has been previously published as:

Greaney AJ, Starr TN, Barnes CO, Weisblum Y, Schmidt F, Caskey M, Gaebler C, Cho A, Agudelo M, Finkin S, Wang Z, Poston D, Muecksch F, Hatziioannou T, Bieniasz PD, Robbiani DF, Nussenzweig MC, Bjorkman PJ, Bloom JD. Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. Nat Commun. 2021 Jul 7;12(1):4196. doi: 10.1038/s41467-021-24435-8. PMID: 34234131; PMCID: PMC8263750.

6.1 Abstract

Monoclonal antibodies targeting a variety of epitopes have been isolated from individuals previously infected with SARS-CoV-2, but the relative contributions of these different antibody classes to the polyclonal response remains unclear. Here we use a yeast-display system to map all mutations to the viral spike receptor-binding domain (RBD) that escape binding by representatives of three potently neutralizing classes of anti-RBD antibodies with high-resolution structures. We compare the antibody-escape maps to similar maps for convalescent polyclonal plasmas, including plasmas from individuals from whom some of the antibodies were isolated. While the binding of polyclonal plasma antibodies are affected by mutations across multiple RBD epitopes, the plasma-escape maps most resemble those

of a single class of antibodies that target an epitope on the RBD that includes site E484. Therefore, although the human immune system can produce antibodies that target diverse RBD epitopes, in practice the polyclonal response to infection is skewed towards a single class of antibodies targeting an epitope that is already undergoing rapid evolution.

6.2 Introduction

Control of the SARS-CoV-2 pandemic will depend on widespread population immunity acquired through infection or vaccination. But a little over a year into the pandemic, a proliferating number of new viral lineages are rising in frequency [14, 94, 285, 302, 323, 350]. These emerging lineages have mutations at $< 1\%$ of all residues in the viral spike, and at no more than 3 of the 201 residues in the spike receptor-binding domain (RBD)–yet these handfuls of mutations often substantially erode and in some cases even ablate the polyclonal neutralizing antibody response elicited by infection [11, 42, 51, 170, 189, 232, 315, 313, 320, 324].

A substantial fraction of the neutralizing activity of polyclonal antibody response to SARS-CoV-2 infection is due to antibodies that target the RBD [74, 120, 229, 251, 282], although antibodies that target the NTD also contribute to neutralization [11, 207, 189, 202, 303, 320]. Structural and binding competition studies have shown that the most potently neutralizing anti-RBD antibodies target several distinct epitopes on the RBD's receptor-binding motif [16, 229, 74, 142, 188]. However, the contributions of these different classes of RBD-targeting antibodies to the overall activity of the polyclonal antibody response remains less clear. It is therefore important to systematically determine both how viral mutations impact each antibody class, and how these antibody-specific effects shape the overall effects of viral mutations in a polyclonal context.

Here, we comprehensively map RBD mutations that reduce binding by structurally characterized representatives of three classes of neutralizing monoclonal antibodies that target the RBD's receptor binding motif, as well as polyclonal plasmas from convalescent individ-

uals from whom some of the antibodies were isolated [17, 16, 251, 271]. We make these measurements by using a deep mutational scanning approach to systematically map how all RBD amino-acid mutations affect binding to yeast-displayed RBDs ([124]). The resulting escape maps allow us to systematically compare how RBD mutations affect binding by the monoclonal antibodies, and we find that the antibodies cluster in the space of viral escape in a way that largely recapitulates prior classifications based on structural analyses of the antibody epitopes. However, some of the potently neutralizing monoclonal antibodies contribute very little to the escape maps of the polyclonal plasmas, even for individuals from whom the antibodies were isolated. Instead, the plasma-escape maps usually most resemble a single antibody class (class 2 in the Barnes et al. classification [16]) that targets the face of the receptor-binding ridge that is accessible in both up and down RBD conformations. Unfortunately, a mutation that escapes this antibody class (E484K) is present in many emerging viral lineages, including B.1.351, P.1, P.2, and B.1.526 [14, 94, 285, 302, 323]. We suggest that domination of the RBD-targeting polyclonal response by a single antibody class is a major factor in enabling a small number of viral mutations to sometimes substantially erode neutralizing antibody immunity.

6.3 Results

6.3.1 Mapping all mutations that escape binding by key classes of RBD-targeting monoclonal antibodies

Most potent neutralizing antibodies against the SARS-CoV-2 RBD target the receptor-binding motif, where they compete for binding to ACE2 [16, 229, 356]. Antibodies targeting the RBD have been divided into four major classes based on structural analyses of their epitopes; two with epitopes overlapping with the ACE2 binding site (class 1 and class 2), potent neutralizers that do not directly bind to the ACE2 contact surface (class 3), and antibodies that target a cryptic epitope outside of the receptor-binding motif and are generally

less potent (class 4) [16]. We focused our studies on several antibodies representative of the first three classes of potently neutralizing receptor-binding motif-targeting antibodies. Class 1 antibodies bind the face of the receptor-binding motif that is accessible only when the RBD is in the up conformation (Figure 6.1A); the antibodies from this class in our study are C105 and LY-CoV016. Class 2 antibodies bind a face of the receptor-binding ridge that is accessible in both the up and down conformations (Figure 6.1A); the antibodies from this class in our study are C144, C002, and C121. Class 1 and 2 antibodies compete with ACE2 for RBD binding and have some overlap in their structural footprints particularly at ACE2 contact sites at the top of the receptor-binding ridge. Class 3 antibodies bind the opposite side of the receptor-binding motif (including the 443–450 loop), which like the class 2 epitope is accessible in both the up and down conformations, but has less overlap with the ACE2 binding footprint [16] (Figure 6.1A); the antibodies from this class in our study are C135 and C110. All antibodies were isolated from humans previously infected with SARS-CoV-2 [251, 271] and have high-resolution structures [17, 16, 271].

We used a previously described yeast-display deep mutational scanning approach ([124]) to map all amino-acid mutations to the RBD that escape binding by each antibody. Briefly, we incubated libraries of yeast expressing nearly all possible single amino-acid mutations in the RBD [280] with each antibody, and then used fluorescence-activated cell sorting (FACS) to enrich for cells expressing RBD mutants that escaped antibody binding (Figure 6.7). We used deep sequencing to quantify each RBD mutation's "escape fraction," which is the fraction of cells with that mutation that fall into the FACS antibody-escape bin. These escape fractions range from 0 (no cells with the mutation fall into the antibody-escape bin) to 1 (all cells with the mutation fall into the antibody-escape bin). The escape fractions are well-correlated between independent libraries, and we report the average of duplicate measurements throughout (Figure 6.8). We represent the escape maps as logo plots, where the height of each letter is proportional to its escape fraction (Figure 6.1B). The escape

map for LY-CoV016 has been previously described [278]; all other data using this assay were newly generated in this study. Interactive versions of these maps are available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Rockefeller`.

The escape maps show that antibodies within the same class are generally escaped by mutations at similar sites in the RBD, with large differences between classes (Figure 6.1B). For example, some class 1 but not class 2 antibodies were escaped by mutations to sites K417, D420, N460, and A475 (Figure 6.1B). Class 2 but not class 1 antibodies, on the other hand, were escaped by mutations at sites E484, F490, and Q493 (Figure 6.1B). Class 3 antibodies, which predominantly bind to the part of the receptor-binding motif opposite that of class 1 and class 2 antibodies (Figure 6.1A), tend to be escaped by a different set of mutations, including those at sites R346, K444, and G446–N450 (Figure 6.1B).

However, our maps emphasize that the antibody classes are approximate groupings, with overlap at some sites between classes. For example, class 1 and class 2 antibodies overlap in their binding footprints at the top of the receptor-binding ridge that contacts ACE2 (Figure 6.1A), and mutations to some of these overlapping sites (i.e., L455, F456, F486, and Y489) escape binding by some antibodies from both classes (Figure 6.1B). Similarly, the class 3 antibody C110 and the class 2 antibodies can both be escaped by mutations to site F490. There are also differences in escape mutations within antibody classes. For instance, the class 3 antibody C135 is strongly affected by mutations at site R346 but not L452, whereas the opposite is true for C110 (Figure 6.1B). Other examples can be found for the class 2 antibodies: C002 is escaped by some mutations to site L452, whereas C144 and C121 are less affected by these mutations (Figure 6.1B).

Our escape maps make it possible to group the antibodies using a more continuous approach, without strict divisions into three classes. Specifically, we used multidimensional scaling to project the escape maps into a two-dimensional space of binding escape (note that this computational technique has previously been used to visualize the antigenic relationships

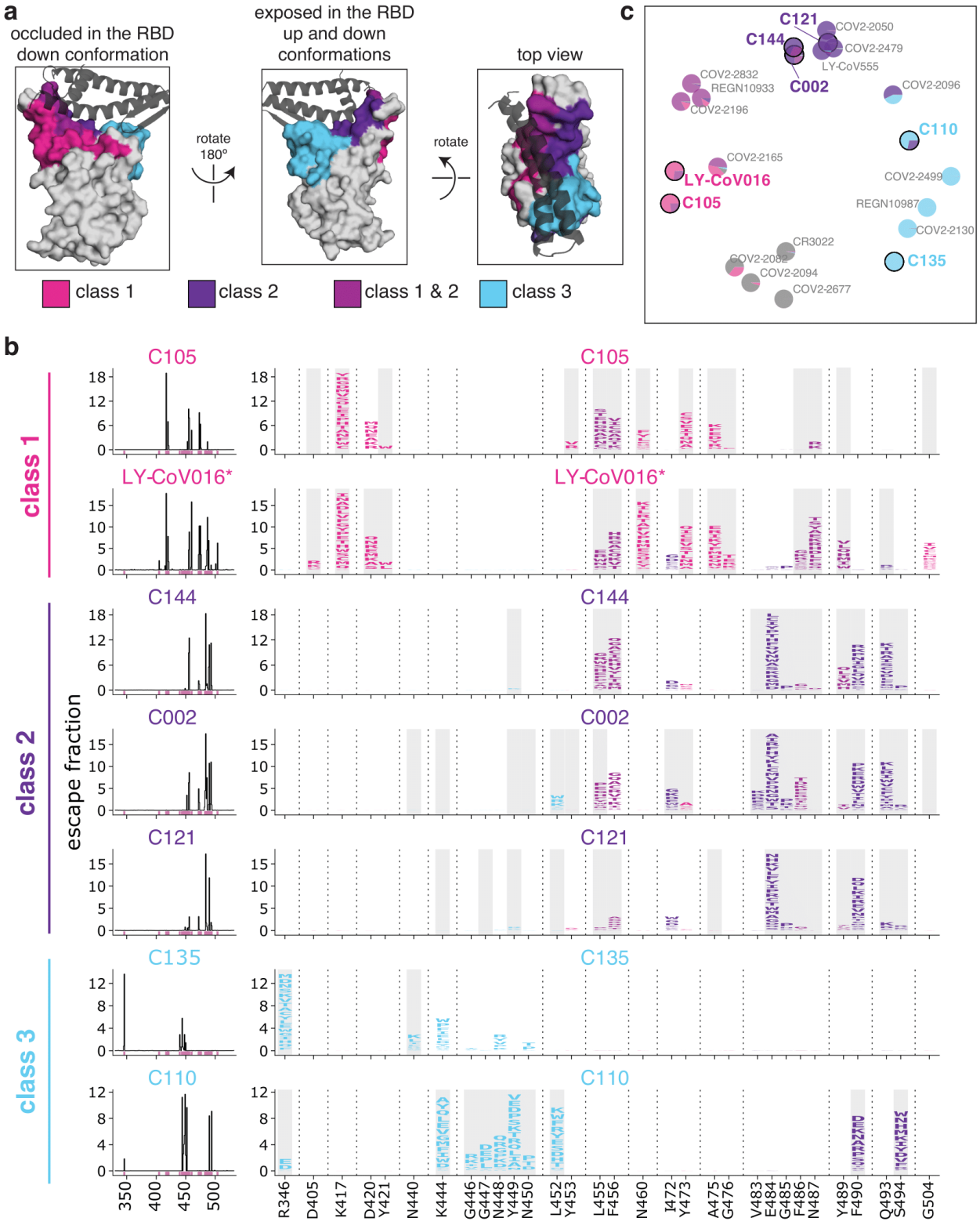
among strains of influenza virus [274]). We did this for the seven antibodies shown in Figure 6.1B, as well as 15 other antibodies for which we have previously determined escape maps [83, 124, 278, 279], including class 4 antibodies, which bind to the RBD outside the receptor-binding motif and are generally less potently neutralizing [74, 188, 229, 356]. Antibodies with similar escape mutations are located close to one another in the multidimensional scaling projection, and antibodies with very distinct escape mutations are far apart (Figure 6.1C). The projection shows that while there is some clustering by structural class in space of escape, the antibodies are continuously distributed. For instance, the class 3 antibody C110 also selects escape at some class 2 sites, such as F490 and S494, and so C110 is located somewhat between the class 2 and 3 antibodies in the multidimensional scaling projection (Figure 6.1C).

6.3.2 RBD mutations reduce antibody binding at only a subset of contact sites

We took advantage of the availability of high-resolution structures to compare the sites of escape mutations to the structural contacts between the antibodies and RBD. Most mutations that escape antibody binding are at sites in the RBD that directly contact the antibody; these sites are highlighted in gray in Figure 6.1B, with a structural contact defined as any non-hydrogen atom within 4Å. To visualize the escape mutations in a structural context, we mapped the extent of escape at each site to the structure of the antibody-bound RBD [17, 16, 271] (Figure 6.2A, 6.11A; interactive, zoomable versions at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Rockefeller`). All sites at which mutations strongly escape binding are in direct ($<$ 4Å) or proximal (4–8Å) contact with antibody in the resolved structures (Figure 6.2B).

However, not all antibody-contact sites had mutations that strongly escaped antibody binding (Figure 6.2B, 6.11A). There are several explanations. First, our approach maps functional antibody escape mutants that retain proper RBD folding and bind to ACE2 with

Figure 6.1: **Maps of mutations to the RBD that escape binding by three classes of monoclonal antibodies that target the receptor-binding motif**. **(A)** Epitopes for each of the three antibody classes [16]. ACE2 is shown as a gray cartoon. Some sites fall under both class 1 and class 2 and are shown as an intermediate pink-purple. **(B)** Escape maps for monoclonal antibodies from each of the three classes. The line plots at left indicate the summed effect of all mutations at each site in the RBD, with larger values indicating a greater reduction in antibody binding. The logo plots at right show the effects of individual mutations at key sites (indicated by purple highlighting on the x-axis of the line plots). In these logo plots, the height of each letter is that mutation's escape fraction, so larger letters indicate mutations that cause a greater reduction in antibody binding. Sites in the logo plots are colored by RBD epitope. Sites that contact antibody (non-hydrogen atoms within 4Åin high-resolution structures) are highlighted with gray backgrounds. For C110, sites 444, 446, and 447 are unresolved in the structure but are likely in close contact with the antibody, and so are highlighted in gray. The data for LY-CoV016 were previously reported [278] and are replotted here. **(C)** Multidimensional scaling projection of the escape maps, such that antibodies with similar escape mutations are drawn close together. Each antibody is represented as a pie chart colored according to the amount of escape in each RBD epitope. Antibodies for which escape maps are shown in panel (B) have black outlines and colored names. The other antibodies were profiled previously [83, 124, 278, 279]. Escape maps for all class 1, 2, and 3 antibodies in the plot are shown in Figure 6.9; for the class 4 antibody-escape maps, see [124]. Colors used in all panels: sites are colored according to epitopes, as defined in [16]. Bright pink for class 1, dark purple for class 2, medium pink-purple for class 1/2 overlap sites (455, 456, 486, 487, 489), cyan for class 3, and gray for all other RBD sites. Escape maps colored according to mutation effects on RBD expression and ACE2 binding are in Figure 6.10. The ACE2-bound RBD structure in (A) is from PDB 6M0J [172].



*LY-CoV016 escape mapping was previously performed in Starr, et al. (2021)

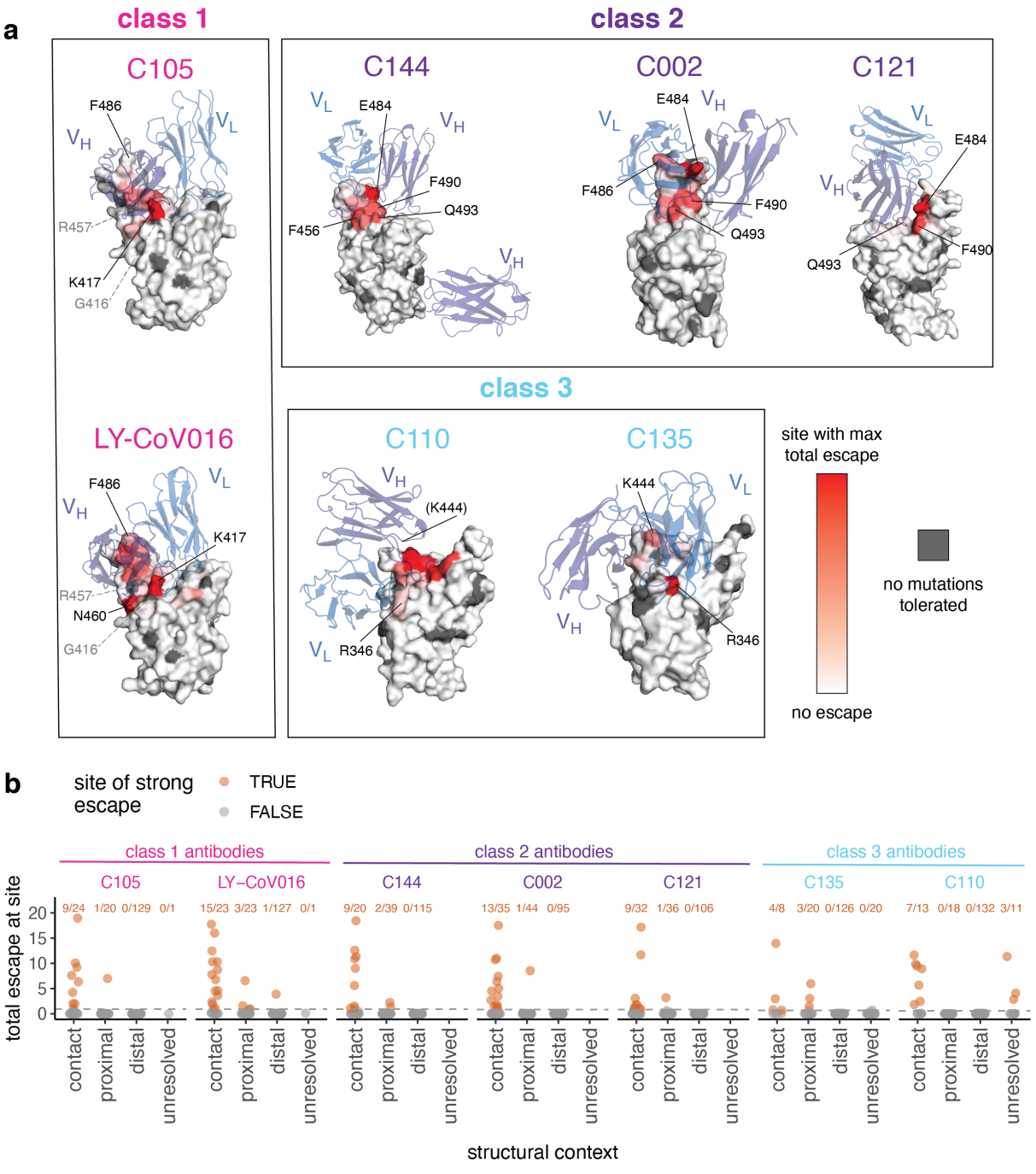
$\geq 1\%$ the affinity of the unmutated RBD [280] (see Methods). Only 2,304 of the 3,819 possible amino-acid mutations to the RBD meet these criteria, and some sites have no tolerated mutations. For instance, G416 and R457 are both in the structural epitope of C105, but these sites have no tolerated mutations and thus do not appear in the escape map (sites with no tolerated mutations are indicated in dark gray in Figure 6.2A, 6.11). Second, sometimes mutations at antibody-contact sites simply do not strongly disrupt antibody binding [68]. For instance, site F486 is in structural contact with both C105 and LY-CoV016 and has many well-tolerated mutations, but mutations at this site more strongly affect the binding of LY-CoV016 than C105 (Figure 6.1B, 6.2A, 6.11A). Other examples include site R346, where nearly all mutations escape C135 but only charge-reversal mutations escape C110 (Figure 6.1B, 6.2A, 6.11A). Similarly, at site Q493, C144 and C002 are escaped by many mutations, but C121 is only escaped by Q493K/R (Figure 6.1B, 6.2A, 6.11A).

For some of the class 2 antibodies, the antibody makes a quaternary contact with an adjacent RBD in the context of spike trimer [16] (Figure 6.11B). Our yeast-display system assays antibody binding to isolated RBD, and so does not map escape mutations to quaternary contact sites and cannot inform on their importance for antibody binding.

6.3.3 The escape maps of polyclonal plasmas often differ from those of monoclonal antibodies isolated from the same individual

The six antibodies newly mapped in this study were isolated from four different SARS-CoV-2 convalescent individuals (Figure 6.3A). Plasma was collected from these individuals at the same time that blood was collected for antibody isolation [251]. Because our escape-mutant mapping approach works for polyclonal sera or plasmas in addition to monoclonal antibodies [120], we mapped mutations that reduced binding by each of the four plasmas plus one plasma sample without corresponding antibodies (Figure 6.3B; the plasmas are prefixed with "COV-" to distinguish them from the antibodies which are prefixed with "C").

Figure 6.2: **Mutations that escape antibody binding are usually in the direct structural footprint**. **(A)** The total escape at each site is mapped onto the surface of the Fab-bound RBD, with white indicating no escape and red indicating the site with the most escape from that antibody. Sites where no mutations are tolerated for RBD folding or ACE2 binding are indicated in dark gray. For C105 and LY-CoV016, gray labels with dashed lines indicate example contact sites with no tolerated mutations. For C110, the general area where site 444 (unresolved in structure) would be located is indicated. **(B)** Total escape at each site in the RBD, with sites classified according to whether they are an antibody contact (within 4Å), antibody-proximal (4 to 8Å), antibody-distal (> 8Å), or unresolved in the Fab-trimer structure. Text indicates the number of sites in each structural category that are sites of strong escape, (n / total) shown in orange. See Methods for details and PDB accessions.


a
class 1
class 2
C105
F486
VH
VL
R457
K417
G416
C144
VL
E484
VH
F490
Q493
F456
VH
C002
E484
VH
VL
F486
F490
Q493
C121
VL
E484
VH
Q493
F490
class 3
LY-CoV016
F486
VL
VH
K417
R457
N460
G416
C110
VH
(K444)
VL
R346
C135
K444
VL
VH
R346
site with max total escape
no escape
no mutations tolerated
b
site of strong escape
TRUE
FALSE
class 1 antibodies
class 2 antibodies
class 3 antibodies
C105
LY−CoV016
C144
C002
C121
C135
C110
9/24 1/20 0/129 0/1
15/23 3/23 1/127 0/1
9/20 2/39 0/115
13/35 1/44 0/95
9/32 1/36 0/106
4/8 3/20 0/126 0/20
7/13 0/18 0/132 3/11
total escape at site
20
15
10
5
0
contact
proximal
distal
unresolved
structural context

The plasma-escape maps shared many commonalities across all five individuals (Figure 6.3B), and generally resembled those from our prior study of a larger cohort of convalescent individuals [120]. In particular, mutations to sites F456 and E484 reduced binding for all five plasma samples (Figure 6.3B). Mutations to site E484 are of special note as the E484K mutation is present in the emerging B.1.351, P.1, P.2, and B.1.526 SARS-CoV-2 lineages [14, 94, 285, 302, 323] and can reduce the neutralization titer of convalescent plasmas by 3-fold or more [42, 120, 189, 315, 324]. Mutations to other sites, such as G446–N450 and F486 also reduce binding for some of the plasma samples profiled here (Figure 6.3B), consistent with our prior study of a larger cohort [120]. These findings suggest that while there is some heterogeneity among which mutations reduce binding of different individuals' polyclonal plasma antibodies, there are also sites that are commonly targeted and should be monitored for antigenic evolution.

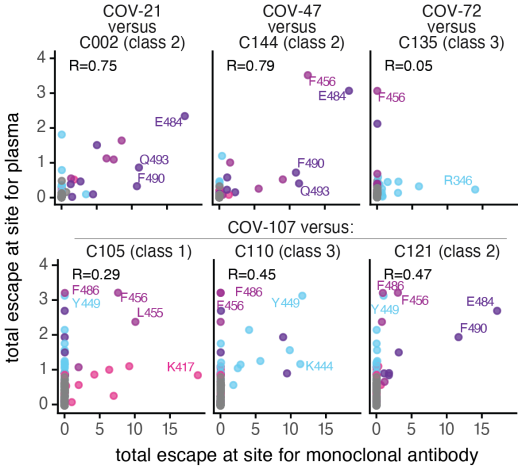
While the escape maps for the different plasma samples shared broad similarities, they often starkly differed from the escape maps of monoclonal antibodies isolated from the same individuals (Figure 6.3B, compare plasma escape maps in logo plots with overlay bars showing sites of escape for antibodies from the same individual). For instance, mutations to R346 had the largest effects on binding by the class 3 antibody C135, but had little effect on the same individual's plasma (COV-72). Similarly, mutations at K417 had the largest effects on binding by the class 1 antibody C105, but had little effect on the corresponding plasma (COV-107). Conversely, mutations to site G496 reduced binding by the COV-21 plasma, but did not strongly affect any of the monoclonal antibodies in this study (Figure 6.3B). Overall, the correlations between the sites at which mutations escaped binding for the monoclonal antibodies and their corresponding polyclonal plasmas were highest for the class 2 antibodies, and lower for the other antibody classes (Figure 6.3C).

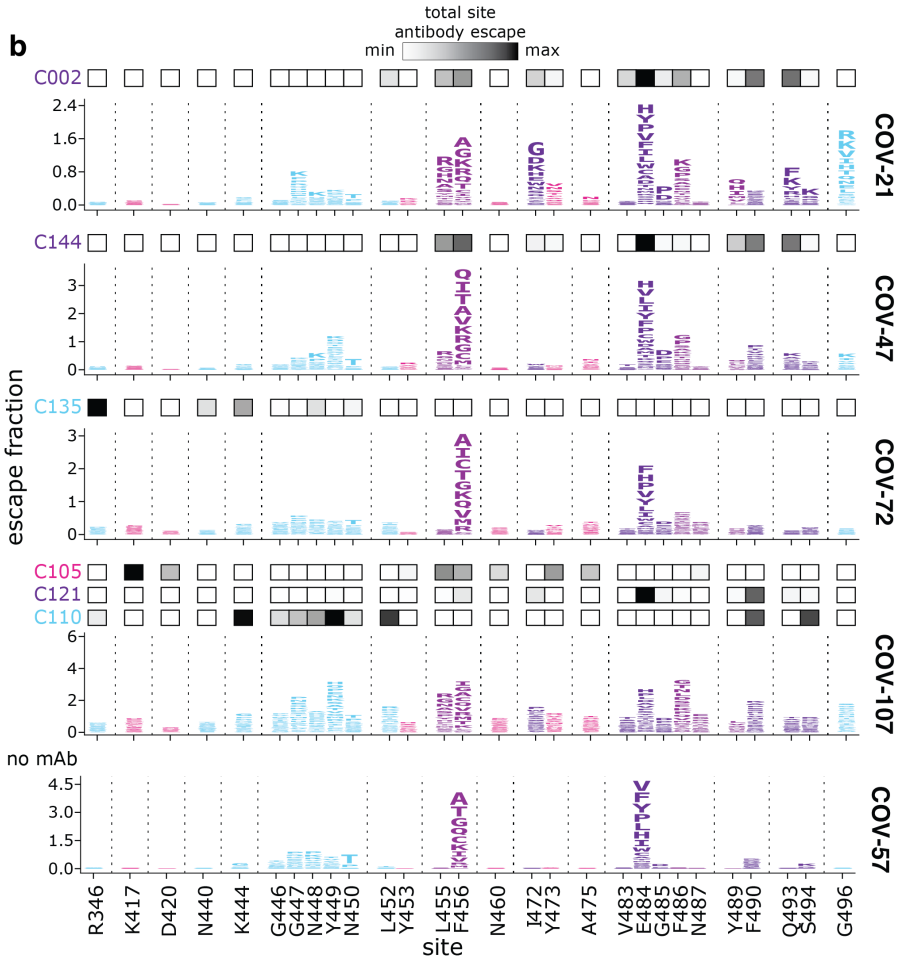
Figure 6.3: **The mutations that reduce binding of polyclonal plasmas often differ from those that reduce binding by monoclonal antibodies isolated from the same individual**. **(A)** Table indicating which plasmas and antibodies were derived from the same individual. **(B)** Escape maps for the polyclonal plasma antibodies, as in Figure 6.1B. The y-axis is scaled separately for each plasma (see Methods). When there are monoclonal antibodies isolated from the same individual, the total monoclonal antibody escape at each site is shown using the heat maps above the escape maps, with white indicating no effect and black indicating strong escape. **(C)** Correlation of plasma and monoclonal antibody escape for each plasma / antibody pair from the same individual. Each point in the scatter plots is a site, with the x-axis indicating the total escape at that site for the antibody and the y-axis indicating the total escape at that site for the plasma. Key sites are labeled. Pearson's R shown above each plot. Colors in B, C reflect antibody classes as in Figure 6.1.

a **polyclonal plasma & corresponding monoclonal antibodies**

plasma	antibody	antibody class
COV-021	C002	class 2
COV-047	C144	class 2
COV-072	C135	class 3
COV-107	C105	class 1
	C121	class 2
	C110	class 3
COV-057	none	none

class 1
class 2
class 3
class 1 & 2





6.3.4 Class 2 antibodies contribute the most to the RBD escape maps of polyclonal plasmas

To more broadly compare how antibodies of different classes contribute to the convalescent plasma escape maps, we used multidimensional scaling to project 22 antibodies and 28 polyclonal plasmas into a two-dimensional space of binding-escape mutations (Figure 6.4A; the projection shows the 22 antibodies in Figure 6.1C, the 5 plasmas from Figure 6.3, and 23 plasmas from a previously characterized larger cohort [120]). The plasmas from both cohorts cluster together in the space of binding escape, but far from some of the antibodies (Figure 6.4A). In particular, while the plasmas are often clustered towards the middle of the plot, suggesting contributions from multiple antibody classes, most plasmas are positioned closest to the class 2 antibodies in the space of binding escape (Figure 6.4A; an interactive version that includes additional antibodies and vaccine sera is available at `https://jbloomlab.github.io/SARS2_RBD_Ab_escape_maps/`).

To visualize the escape maps in terms of the RBD's three-dimensional structure, we projected onto the surface of the RBD the total escape at each site averaged across all antibodies in a class, or all convalescent plasmas in a cohort (Figure 6.4B). Again, the polyclonal plasmas most closely resembled the class 2 antibodies. For instance, mutations to site E484 greatly reduced binding of both class 2 antibodies and polyclonal plasmas (Figure 6.4B). Most of the class 1 antibody contributions to the plasma binding-escape maps came from sites shared with class 2 antibodies, such as F456 (Figure 6.4B), although note that mutations to F456 often do not strongly reduce plasma neutralization [120]. Consistent with the lesser contributions of class 1 antibodies, mutations at the class 1 site K417 had little effect on plasma binding, and others have found that the K417N mutation alone has a minimal-to-modest effect on plasma neutralization [51, 170, 315, 313]. Sites of escape from class 3 antibodies (e.g., G446) had visible effects on the polyclonal plasmas, but again less so than for class 2 antibody sites (Figure 6.4B). However, in a prior larger study [120], we found that mutations to the 443–450 loop strongly reduced binding of plasmas from a minority of

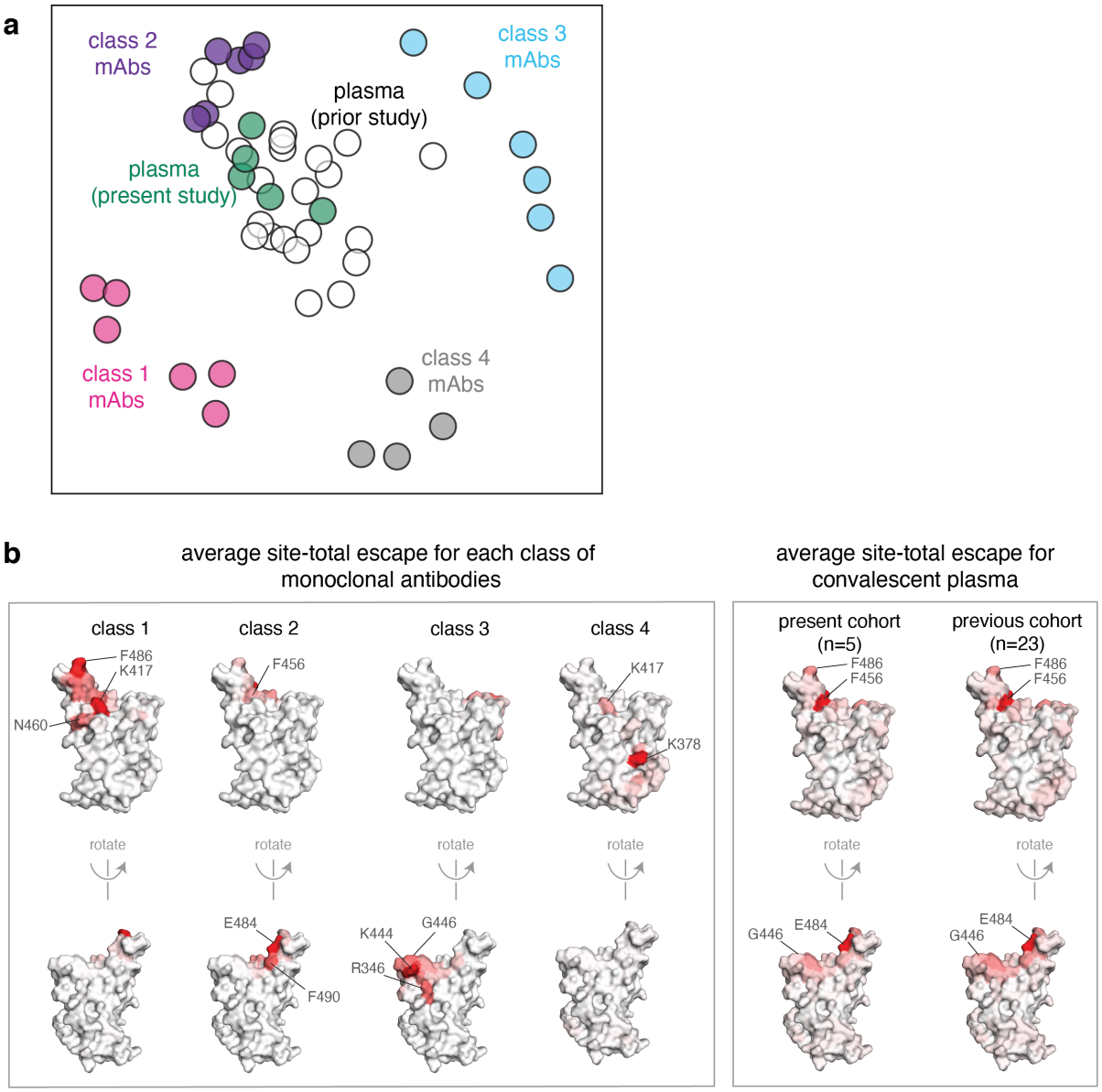
individuals, consistent with a few plasmas falling closer to class 3 than class 2 antibodies in the space of binding-escape mutations (Figure 6.4A). Overall, these results show that while multiple antibody specificities contribute, class 2 antibodies usually dominate convalescent polyclonal plasmas–although once a virus has accumulated mutations in class 2 epitopes (as has already occurred in some emerging lineages [94, 285, 302, 323]), then class 1 or 3 antibodies might dominate for the remaining anti-RBD antibody activity.

6.3.5 Escape maps are consistent with the RBD mutations that arise when virus is grown in the presence of monoclonal antibodies

We assessed how well our escape maps predicted the actual antibody-escape mutations that arose when virus was grown in the presence of the antibodies. Prior work selected viral escape mutants by passaging chimeric vesicular stomatitis virus (VSV) encoding the SARS-CoV-2 spike in the presence of several of the monoclonal antibodies we mapped in this study [320]. We hypothesized that the mutations selected during viral passage would reduce antibody binding without impairing ACE2 binding affinity. Accordingly, we examined how all of the selected mutations affected both antibody binding (as measured in the current study) and ACE2 affinity (as measured in our prior deep mutational scanning [280]). Figure 6.5A,B shows that in every case, the antibody-escape mutations selected in the virus were indeed among the single-nucleotide-change accessible amino-acid mutations that mediated the strongest escape from antibody binding without strongly impairing ACE2 affinity. Conversely, mutations that escaped antibody binding but were deleterious for ACE2 binding or RBD expression (e.g., E484V/A and mutations to sites 455 and 456) were not selected in the viral passaging. Therefore, our escape maps can be used in conjunction with prior data on the functional effects of RBD mutations to largely predict which escape mutations will arise when virus is grown in the presence of antibodies.

Weisblum and colleagues also tested many RBD point mutations for their effects on neu-

Figure 6.4: **The escape maps of convalescent polyclonal plasmas most resemble class 2 antibodies**. **(A)** Multidimensional scaling projection of the escape maps of polyclonal plasmas and monoclonal antibodies of each class. Antibodies or plasmas that are nearby in the plot have their binding affected by similar RBD mutations. The antibodies are those in Figure 6.1C, colored according to antibody class, as in Figure 6.1. The five plasmas newly mapped in this study are shown in green, and the previously mapped 23 plasmas [120] are shown in white. **(B)** Structural projection of sites where mutations reduce binding by each class of monoclonal antibodies (left) or polyclonal plasmas (right). The RBD surface coloring is scaled from white to red, with white indicating no escape, and red indicating the site with the greatest average site-total escape for all antibodies or plasmas in that group. Mutations to sites such as E484, F456, and F486 have some of the largest effects on binding by polyclonal plasmas and class 2 antibodies. An interactive version of (A) that includes additional antibodies and vaccine sera is available at `https://jbloomlab.github.io/SARS2_RBD_Ab_escape_maps/`.

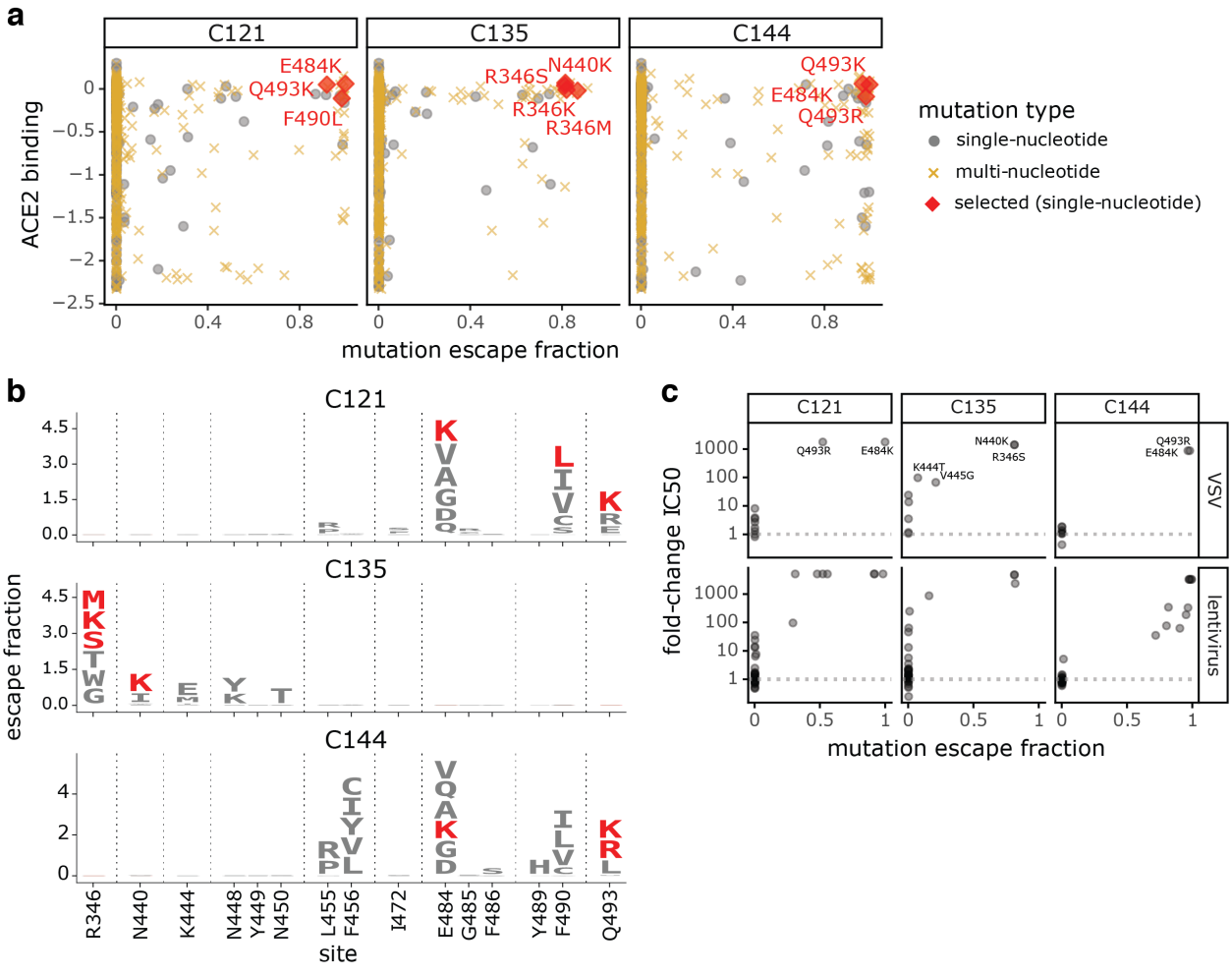

a
class 2
mAbs
class 3
mAbs
plasma
(prior study)
plasma
(present study)
class 1
mAbs
class 4
mAbs
b
average site-total escape for each class of
monoclonal antibodies
average site-total escape for
convalescent plasma
class 1
class 2
class 3
class 4
F486
K417
N460
F456
K417
K378
rotate
E484
K444
G446
R346
F490
present cohort
(n=5)
previous cohort
(n=23)
F486
F456
G446
E484

tralization of chimeric VSV or lentiviral particles by the antibodies studied here [320]. There was generally good agreement between our escape maps and these previously measured effects of mutations on viral neutralization. Nearly all mutations with a >100-fold reduction in neutralization also had large effects in our escape maps, although in a few cases mutations with more moderate effects on neutralization were not prominent in the escape maps (Figure 6.5C, S6A). Previously, we and others have reported that single point mutations can reduce the neutralization of some plasmas by >10-fold, although other plasmas are largely unaffected by any single mutation [120, 189, 320]. For the plasmas in this study, prior work found that no tested mutation had such large effects on neutralization [320]. However, the class 2 antibody-escape mutation E484K did reduce neutralization by COV47 plasma by approximately 5-fold [320], concordant with the prominence of site 484 in that plasma's escape map (Figure 6.3B, S6B).

6.3.6 Mutations that reduce binding by class 1, 2, and 3 antibodies are present in emerging viral lineages

To assess the extent that SARS-CoV-2 has already acquired mutations that reduce binding by each antibody class, we compared the total escape at each site averaged across all antibodies of that class to the frequency of mutations at the site among sequenced SARS-CoV-2 isolates in GISAID as of May 11, 2021 [91]. Figure 6.6A shows that mutations at sites targeted by each antibody class are present at appreciable frequencies among sequenced SARS-CoV-2 isolates. In particular, sites K417 and E484 are the strongest sites of escape for class 1 and class 2 antibodies, respectively—and mutations at both these sites are present in a substantial number of sequenced viral isolates (Figure 6.6A). In contrast, mutations that most strongly escape class 3 antibodies are currently not as prevalent among sequenced isolates (Figure 6.6A). While mutations have been observed at site K444 (the strongest site of escape from class 3 antibodies), these are at lower frequency than site 417 or 484 mutations (Figure

Figure 6.5: **Escape maps predict mutations that are selected during viral growth in the presence of monoclonal antibodies**. **(A)** Mutations selected when chimeric VSV encoding the SARS-CoV-2 spike was grown in the presence of each of the three indicated antibodies by Weisblum et al. (2020) [320]. Each point represents a different amino-acid mutation, with the x-axis indicating how strongly the mutation escapes antibody binding (measured in the current study) and the y-axis indicating how well the mutant binds to ACE2 (measured in [280]). The red diamonds indicate the mutations selected in VSV-spike by Weisblum et al., the gray circles indicate all other amino-acid mutations accessible by a single-nucleotide change, and the gold x's indicate amino-acid mutations that require multiple nucleotide changes to the codon. **(B)** Logo plots showing the effects of only single-nucleotide accessible amino-acid mutations on antibody binding. Mutations selected in VSV-spike virus by Weisblum et al. are colored red. **(C)** The correlation of the effects of mutations on antibody binding measured in the current study and effects on viral neutralization previously measured by Weisblum et al. [320] using chimeric VSV (top) or lentiviral particles (bottom). The x-axis shows the escape fraction measured in the current study, and the y-axis shows the fold-change in inhibitory concentration 50% (IC50) for viral neutralization caused by that mutation, such that larger numbers correspond to greater reductions in neutralization sensitivity. For effects of all antibody- and plasma-binding-escape mutations on ACE2 binding and RBD expression, see Figure 6.10. For each mutation's escape fraction compared to fold-change IC50 against each monoclonal antibody or polyclonal plasma tested in Weisblum et al. [320], see Figure 6.12.


a
C121
C135
C144
E484K
Q493K
F490L
R346S
N440K
R346K
R346M
Q493K
E484K
Q493R
ACE2 binding
mutation escape fraction
mutation type
single-nucleotide
multi-nucleotide
selected (single-nucleotide)
b
C121
C135
C144
escape fraction
site
R346
N440
K444
N448
Y449
N450
L455
F456
I472
E484
G485
F486
Y489
F490
Q493
c
C121
C135
C144
VSV
lentivirus
fold-change IC50
mutation escape fraction
Q493R
E484K
N440K
R346S
K444T
V445G

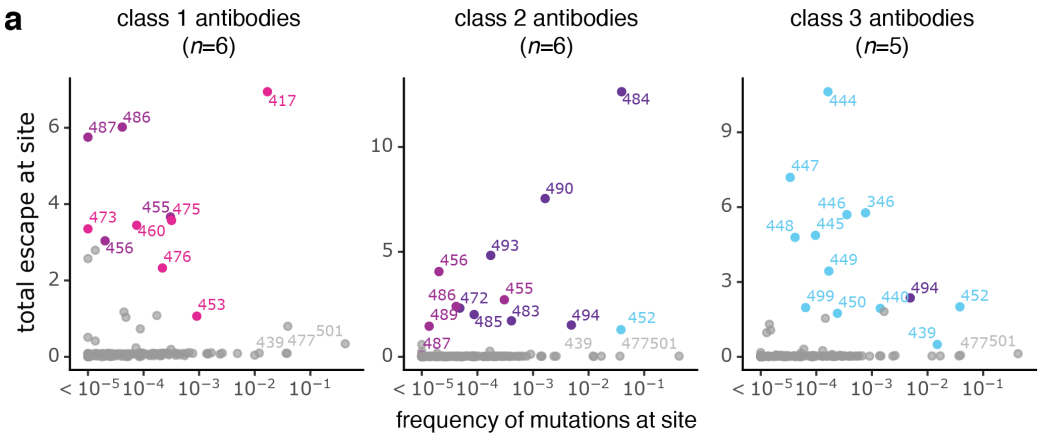
6.6A). Mutations at the class 3 site L452 are present at higher frequency, but L452 is only a moderate site of escape for this antibody class. For instance, mutations to L452 escape binding of C135 but do not strongly escape the other class 3 antibodies. However, L452 mutations also escape binding of some class 2 antibodies. As expected from the fact that class 2 antibodies dominate convalescent polyclonal plasma, the natural frequency versus escape plots for the plasma closely resemble those for class 2 antibodies, with mutations at site E484 having the largest effect (Figure 6.6B).

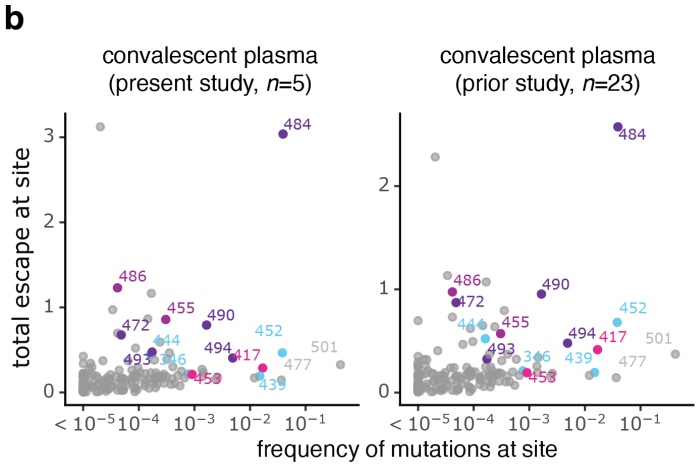
We also examined the presence of escape mutations for each antibody class in some key emerging viral lineages (Figure 6.6C). All these emerging viral lineages except B.1.1.7 [232, 270, 313, 51, 312] have a mutation that escapes some antibodies from at least one class. The class 2 antibody escape mutation E484K is present in the B.1.351, P.1, P.2 and B.1.526 lineages (Figure 6.6C) [14, 94, 285, 302, 323]. Two of these lineages, B.1.351 and P.1, also have a class 1 antibody escape mutation, K417N or K417T, respectively (Figure 6.6C). The B.1.427/429 and B.1.617 lineages carry a class 3 antibody escape mutation (L452R; Figure 6.6C). No viral lineages currently combine mutations that escape all three antibody classes—but the future emergence of a class 3 escape mutation in one of the lineages that already escape class 1 and 2 antibodies (B.1.351 or P.1) would be a worrying development, and should be monitored for closely.

6.4 Discussion

We comprehensively mapped all mutations that escape binding by three major classes of antibodies targeting the SARS-CoV-2 RBD, and compared these escape maps to those for convalescent polyclonal plasma. We find that while multiple antibody specificities contribute, a single antibody class (class 2) largely shapes how RBD mutations affect binding by polyclonal plasma, even for individuals from whom potent neutralizing antibodies of other classes were isolated. A similarly unequal immunodominance of different epitopes has also been de-

Figure 6.6: **Mutations that escape binding by antibodies and plasmas among sequenced SARS-CoV-2 isolates**. **(A)** The total escape at each site averaged across antibodies in each class versus frequency of mutations at each site in GISAID sequences as of May 11, 2021. (B) The total escape at each site averaged across the polyclonal plasmas versus frequency of mutations at each site in GISAID sequences. Left: plasma samples profiled in this study, right: plasma samples profiled previously [120]. (C) Antibody-escape mutations found in emerging viral lineages. RBD mutations in each lineage are assigned to the class of antibody they most strongly escape (e.g., E484K most strongly escapes class 2 antibodies but may also affect some class 1 antibodies). Other RBD mutations present in each viral lineage with negligible effects on binding of these antibodies are listed at right. For numbers of antibodies in each class, see n indicated in each panel in A and B. In each plot, key sites are labeled and colored according to RBD epitope using the same color scheme as in Figure 6.1.





c

viral lineage	escape mutations: class 1	class 2	class 3	other RBD mutations
B.1.1.7	–	–	–	N501Y
B.1.351	K417N	E484K	–	N501Y
P.1	K417T	E484K	–	N501Y
P.2	–	E484K	–	–
B.1.427/ 429	–	(+/-E484K)	L452R	–
B.1.525		E484K		
B.1.526	–	E484K	–	S447N
B.1.61.7.1	–	E484Q	L452R	
B.1.617.2	–	–	L452R	T478R

scribed for other viral antigens, such as influenza hemagglutinin [12, 13]. However, the immunodominance of different epitopes can vary depending on the antigen formulation and delivery [12], and there is preliminary evidence that some vaccines could elicit antibody responses to SARS-CoV-2 with immunodominance hierarchies that are less skewed than those elicited by infection [51, 120, 121, 315].

The importance of class 2 antibodies in the RBD-targeting portion of polyclonal plasma could be due in part to the twin facts that their epitope is exposed in both up and down conformations of the RBD, and that such antibodies are often generated from frequently observed germline genes (including VH1-2, VH3-53, VH1-69) [33, 38, 166, 251, 318, 346, 345, 335]. Consistent with our work here showing that class 2 antibodies shape how mutations affect polyclonal plasma binding, mutations at the site that most strongly affects binding by this antibody class (E484) have arisen multiple times in emerging viral lineages [14, 94, 285, 302, 323]. Therefore, our results show the importance of thinking about antigenic evolution in the context of different classes of antibodies that recognize different epitopes on the RBD.

Our work also sheds light on the extent to which it is functionally meaningful to subdivide anti-RBD antibodies into distinct classes based on their structurally defined epitopes. These structurally defined classes are inherently approximate groupings, since even antibodies with superficially similar structural epitopes bind to their antigens in subtly distinct ways [16]. Our comprehensive maps of binding escape mutations capture these subtle differences, and show that antibodies in the same structurally defined class can be differentially affected by the same mutation. Using the escape maps, we can visualize how the antibodies are related in terms of how their binding is functionally impacted by mutations at different RBD sites. These visualizations show that the arrangement of antibodies in the space of viral escape is indeed continuous, but that the class definitions based on structural analyses capture the high-level features of this arrangement since the structural footprint of an antibody largely determines which mutations most impact its binding. However, we suggest that in some

cases, more continuous visualizations of the arrangements of antibodies in the space of viral escape such as the ones we present here could have benefits over structural classification schemes for analyzing the impacts of viral mutations.

There are several limitations to our study. First, we did not examine effects of mutations or deletions to the NTD, which can also affect neutralization [11, 207, 189, 202, 303, 320]. Our experiments assayed binding of antibodies to monomeric yeast-expressed RBD, and so cannot capture mutational effects on trimer conformation or antibodies with quaternary epitopes [16]. Additionally, our experiments were designed to identify mutations with large effects on antibody binding, and may overlook mutations with more subtle effects. Finally, the N-linked glycans on yeast-expressed proteins are more mannose-rich than those on mammalian-expressed proteins [127].

Framing viral antigenic evolution in terms of different classes of antibody epitopes is useful for interpreting the impacts of viral mutations and forecasting where in the RBD new escape mutations may arise in the future. Consistent with our work showing that mutations at class 2 antibody epitopes generally cause the largest reductions in the binding of polyclonal anti-RBD plasma antibodies elicited by infection, many emerging SARS-CoV-2 lineages have already acquired a mutation (E484K) at the site that most potently escapes antibodies of this class. Once viruses have escaped class 2 antibodies, antibodies of other classes (i.e., class 1 and 3) will contribute most remaining RBD-targeted antibody immunity. In this respect, it is noteworthy that some of the most prominent emerging viral lineages with E484K have also acquired a mutation (K417N/T in B.1.351 and P.1, respectively) at the site that most potently escapes binding by class 1 antibodies. Moreover, the B.1.427/429 and B.1.617.2 viral lineages contain a moderate class 3 escape mutation (L452R) [77, 350], and B.1.617.1 contains both E484Q and L452R [238]. Other clusters of sequences containing mutations to class 2 and 3 epitopes (E484K, R346K) have been reported [72]. Fortunately, no major viral lineages currently contain mutations to all three epitopes. However, we suggest the

appearance of such a variant would be a worrying development, and should be monitored closely.

6.5 Materials and Methods

SARS-CoV-2 convalescent human plasma samples

Plasma samples were previously described and collected as part of a prospective longitudinal cohort study of SARS-CoV-2 convalescent individuals in New York, NY [251]. Plasma samples profiled in this study were obtained 21–35 days post-symptom onset [251]. Samples were obtained upon written consent from community participants under protocols approved by the Institutional Review Board of the Rockefeller University (DRO-1006). All samples were heat-inactivated prior to use by treatment at 56 C for 60 minutes. Prior to use in each assay, plasma samples were centrifuged for 15 min at 2000 $\times g$ to pellet platelets.

Neutralizing monoclonal antibodies binding the SARS-CoV-2 spike RBD

Antibodies were isolated from individuals in the cohort above as previously described [251]. Briefly, the IGH, IGL, and IGK genes were sequenced using IgG-specific primers from single-cell sorted RBD$^+$, CD20$^+$ memory B cells (CD3$^-$CD8$^-$CD14$^-$CD16$^-$CD20$^+$Ova$^-$RBD-PE$^+$RBD-AF647$^+$). Recombinant monoclonal antibodies were produced and purified as previously described [160, 265]. Antibodies were produced with a human IgG1 heavy chain and human IgK (C002, C110, C135) or human IgL2 (C105, C121, C144) constant regions. These antibodies were previously structurally characterized in complex with SARS-CoV-2 S trimer [17, 16]. A subset was functionally characterized in [261, 320]. The PBD accessions for the antibody-S complex structures are: 6XCM and 6XCN for C105, 7C01 for LY-CoV016, 7K8S and 7K8T for C002, 7K8X and 7K8Y for C121, 7K90 for C144, 7K8Z for C135, and 7K8V for C110 [17, 16, 271]. The protein sequences for these antibodies are available at the aforementioned PDB accession numbers.

RBD deep mutational scanning library

Monoclonal antibody and polyclonal clonal plasma selection experiments were performed in biological duplicate using a deep mutational scanning approach ([124]) with previously described duplicate yeast-displayed mutant RBD libraries [280]. These libraries were generated in the RBD background of the SARS-CoV-2 isolate Wuhan-Hu-1 (Genbank accession number MN908947, residues N331–T531) via NNS codon tiling PCR mutagenesis, which introduced an average of 2.7 amino acid mutations per library variant. RBD variants were linked to unique 16-nucleotide barcode sequences to facilitate downstream sequencing and bottlenecked to library sizes of approximately 100,000 uniquely barcoded variants. The libraries contain 3,804 of the 3,819 possible amino-acid mutations, with $>$ 95% present as single mutants on at least one barcode in the libraries. We previously used these libraries to measure the effect of all RBD mutations on yeast-surface RBD expression and ACE2 affinity [280]. These libraries were sorted to eliminate variants that lose ACE2 binding prior to mapping the antibody-escape variants ([124]).

FACS sorting of yeast libraries to select mutants with reduced binding by polyclonal plasmas

Antibody labeling and selection was performed essentially as described in [124]. Specifically, 9 OD aliquots of RBD libraries were thawed and grown overnight at 30°C 275 rpm in 45mL SD-CAA (6.7 g/L Yeast Nitrogen Base, 5.0 g/L Casamino acids, 1.065 g/L MES, and 2% w/v dextrose). Libraries were diluted to an OD of 0.67 in SG-CAA+0.1% dextrose (SD-CAA with 2% w/v galactose and 0.1% w/v dextrose in place of 2% dextrose), and incubated for 16–18 hours at room temperature with mild agitation to induce RBD surface expression. For each antibody selection, 20 OD units of induced cells were washed twice with PBS-BSA (0.2 mg/mL), and incubated in 4mL of PBS-BSA with monoclonal antibody or plasma for 1 h at room temperature with gentle agitation. Incubations were performed with 400 ng/mL for each monoclonal antibody (C105, C144, C002, C121, C135, or C110) or with a sub-

saturating dilution of polyclonal plasma such that the amount of fluorescent signal due to plasma antibody binding to RBD was approximately equal across plasmas (COV-021, 1:500; COV-047, 1:200; COV-057, 1:50; COV-072, 1:200; COV-107, 1:80). Labeled cells were washed with ice-cold PBS-BSA followed by secondary labeling for 1 h at 4°C in 2.5 mL 1:200 PE-conjugated goat anti-human-IgG (Jackson ImmunoResearch 109-115-098) to label for bound monoclonal antibody or 1:200 Alexa-647-conjugated goat anti-human-IgA+IgG+IgM (Jackson ImmunoResearch 109-605-064) to label for bound plasma antibodies, and 1:100 FITC-conjugated anti-Myc (Immunology Consultants Lab, CYMC-45F) to label for RBD surface expression. Labeled cells were washed twice with PBS-BSA and resuspended in 2.5mL PBS. Yeast expressing the unmutated SARS-CoV-2 RBD were prepared in parallel to library samples, labeled at the same 400 ng/mL and 100x reduced 4 ng/mL antibody concentrations for the monoclonal antibodies, and with 1x and 10x reduced plasma concentrations for the polyclonal plasmas.

Yeast cells expressing RBD variants with substantially reduced antibody binding were selected via fluorescence-activated cell sorting (FACS) on a BD FACSAria II, using FACSDiva software, version 6.1. For monoclonal antibody selections, FACS selection gates were drawn to capture 95% of yeast expressing unmutated SARS-CoV-2 RBD labeled at 100x reduced antibody concentration relative to library samples. For polyclonal plasma selections, FACS selection gates were drawn to capture 2.8–5% of the RBD mutants with the lowest amount of plasma binding for their degree of RBD expression (Figure 6.7A–C). Nearly zero ($< 0.1\%$) and 0.2 to 27.2% of cells expressing unmutated RBD fell into this gate when stained with 1x and 0.1x the concentration of plasma, respectively. For each sample, approximately 10 million RBD+ cells (range 8.7e6 to 1.5e7 cells) were processed on the cytometer, with between 1.5e6 and 2.0e6 monoclonal antibody-escaped cells and 3.2e5 and 5.3e5 plasma-escaped cells collected per sample. Antibody-escaped cells collected per sample into SD-CAA supplemented with 1% w/v BSA and grown overnight in 1.5mL SD-CAA + 100 U/mL

penicillin + 100 µg/mL streptomycin at 30°C 275 rpm.

DNA extraction and Illumina sequencing

Plasmid samples were prepared from up to 7.5 OD units (4e7 CFUs) of overnight cultures of antibody-escaped cells, and 30 OD units (1.6e8 CFUs) of pre-selection yeast populations (Zymoprep Yeast Plasmid Miniprep II) per manufacturer instructions, with the addition of a -80°C freeze-thaw step prior to cell lysis. The 16-nucleotide barcode sequences identifying each RBD variant were amplified by PCR and prepared for Illumina sequencing exactly as described previously [280]. Specifically, a primer with the sequence 5'-AATGATACGGCGACCACCGAGA-3' was used to anneal to the Illumina P5 adaptor sequence, and the PerkinElmer NextFlex DNA Barcode adaptor primers with the sequence 5'-CAAGCAGAAGACGGCATACGAGATxxxxxxxxGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3' (where xxxxxxxx indicates the sample index sequence) were used to anneal to the Illumina P7 adaptor sequence and append sample indexes for sample multiplexing. Barcodes were sequenced via 50 bp single-end reads on an Illumina HiSeq 2500, targeting at least 2.5x as many sequencing reads as FACS-selected cells, and pre-sort reference populations of at least 2.5e7 reads.

Analysis of mutant library deep sequencing and computation of per-mutant escape fractions

Escape fractions were computed as described in [124], with minor modifications as noted below. Specifically, we used the dms_variants package (`https://jbloomlab.github.io/dms_variants/`, version 0.8.5) to process Illumina sequences into counts of each barcoded RBD variant in each pre-sort and antibody-escape population using the barcode/RBD look-up table from [280]. Markdown renderings of these steps in the computational analysis are at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/results/summary/aggregate_variant_counts.md` and `https://github.com/jbloomlab/SARS-`

CoV-2-RBD_MAP_Rockefeller/blob/main/results/summary/counts_to_cells_ratio.md.

For each antibody selection, we then computed the escape fraction for each barcoded variant using the deep sequencing counts for each variant in the original and antibody-escape populations and the total fraction of the library that escaped antibody binding via the formula provided in [124]. These escape fractions represent the estimated fraction of cells expressing that specific variant that falls in the antibody escape bin, so a value of 0 means the variant is always bound by antibody and a value of 1 means that it always escapes antibody binding. We then applied a computational filter to remove variants with low sequencing counts or highly deleterious mutations that might cause antibody escape simply by leading to poor expression of properly folded RBD on the yeast cell surface. Specifically, we ignored all variants with pre-selection sequencing counts that were lower than the counts for the 99th percentile of the stop-codon containing variants because stop codon variants are largely purged by the earlier sorts for RBD expressing and ACE2-binding variants and so any residual presence provides an indication of low-count noise. Next, we removed any variants that had poor RBD expression or ACE2 binding, or contained mutations that individually cause poor RBD expression and ACE2 binding to eliminate misfolded or non-expressing RBDs. Specifically, we removed variants that had (or contained mutations with) ACE2 binding scores < -2.35 or expression scores < -1, using the variant- and mutation-level deep mutational scanning scores [280]. Note that these filtering criteria are slightly more stringent than those used in [124] but are identical to those used in [120, 278]. The ACE2 binding cutoff of -2.35 is used to represent the binding of RaTG13 to human ACE2 [280], which possesses the lowest known affinity capable of mediating cell entry [179]. The RBD expression cutoff of -1 is used to eliminate mutations that have as large an expression deficit as mutations to core disulfide residues. 2,034 of the 3,819 possible RBD amino acid mutations passed these filtering steps and were included in our escape maps. All previously reported escape mapping data [124, 120, 278, 279, 83] were reanalyzed in this study with

the parameters listed above. A markdown rendering of the computation of the variant-level escape fractions and the variant filtering is at https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/results/summary/counts_to_scores.md.

Because some library variants contain multiple amino acid mutations, we next deconvolved variant-level escape scores into escape fraction estimates for single mutations using global epistasis models [220] implemented in the dms_variants package, as detailed at (https://jbloomlab.github.io/dms_variants/dms_variants.globalepistasis.html). In this fitting, we excluded variants that contained mutations that were not seen as either single mutants or in at least two multiple-mutant variants. We then computed the estimated effect of each mutation as the impact of that mutation on the observed phenotype scale transformation of its latent phenotype as computed using the global epistasis models, and applied a floor of zero and a ceiling of 1 to these escape fractions. All of the above analysis steps were performed separately for each of the duplicate mutant libraries. We then only retained mutations that passed all of the above filtering and were measured in both libraries or had at least two-single mutant variant measurements in one library. The reported scores throughout the paper are the average across the libraries; these scores are also at https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/results/supp_data/all_samples_raw_data.csv. Correlations in final single-mutant escape scores are shown in Figure 6.8. A markdown rendering of the computation that computes these mutation-level escape fractions is at https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/results/summary/scores_to_frac_escape.md.

For plotting and analyses that required identifying RBD sites of strong escape (e.g., choosing which sites to show in logo plots in Figure 6.1A or Figure 6.3B or label in Figure 6.2B), we considered a site to mediate strong escape if the total escape (sum of mutation-level escape fractions) for that site exceeded the median across sites by > 10 fold, and was at least 10% of the maximum for any site. A markdown rendering of the identification of

these sites of strong escape is at https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/results/summary/call_strong_escape_sites.md.

Comparison of mutation escape fractions to previously measured neutralization concentrations

In Figure 6.5C and Figure 6.12, mutation-level antibody-escape fractions measured in this study are compared to previously measured neutralization titers (inhibitory concentration 50%, IC50) of the same monoclonal antibodies and polyclonal plasmas against some RBD point-mutants [320]. The numerical IC50 values were extracted from figures in Weisblum et al. (2020) [320] using the WebPlotDigitizer tool v4.4 (https://apps.automeris.io/wpd/). We convert those previously measured IC50s to fold-change IC50 relative to wildtype RBD. The numerical IC50s found in [320] are also tabulated at https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/experimental_data/data/weisblum_ic50.csv.

Analysis of mutations in circulating human SARS-CoV-2 strains

For the analysis in Figure 6, all 765,455 spike sequences on GISAID(Elbe and Buckland-Merrett, 2017) as of May 11, 2021 were downloaded and aligned via mafft(Katoh and Standley, 2013) Sequences from non-human origins and sequences containing gap or ambiguous characters were removed, as were sequences with extremely high numbers of RBD mutations relative to other sequences, leaving 679,502 retained sequences. All RBD amino-acid mutations were enumerated compared to the reference Wuhan-Hu-1 SARS-CoV-2 RBD sequence (Genbank MN908947, residues N331–T531). We acknowledge all contributors to the GISAID EpiCoV database for their sharing of sequence data (all contributors listed at: https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/data/gisaid_hcov-19_acknowledgement_table.pdf).

Data visualization

The static logo plots in the paper were created using dmslogo (`https://jbloomlab.github.io/dmslogo/`) version 0.6.2; a markdown rendering of the code that creates these logo plots is at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/results/summary/escape_profiles.md`. For each plasma, the y-axis is scaled to be the greatest of (a) the maximum site-wise escape metric observed for that plasma, (b) 20x the median site-wise escape fraction observed across all sites for that plasma, or (c) an absolute value of 1.0 (to appropriately scale plasmas that are not "noisy" but for which no mutation has a strong effect on plasma binding).

In Figure 6.10, mutations are colored by prior deep mutational scanning measurements of yeast-displayed RBD ACE2 affinity and RBD expression from [280], which are available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/results/single_mut_effects/single_mut_effects.csv`.

The multidimensional scaling in Figure 6.1C and Figure 6.4A that projects the antibodies into a two-dimensional space of escape mutations was performed using the Python scikit-learn package, version 0.23.2. We computed the similarity and dissimilarity in the escape maps between each pair of antibodies, then performed metric multidimensional scaling with two components on the dissimilarity matrix exactly as defined in [124]. In Figure 6.1C, the multidimensional scaling shows antibodies as pie charts colored proportionally to the total squared site escape that falls into that RBD structural region. The code that generates these logo plot visualizations is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/results/summary/mds_escape_profiles.md`.

The interactive visualizations of the escape maps and their projections on the RBD-antibody structures available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Rockefeller` were created using dms-view (`https://dms-view.github.io/docs/`)(Hilton et al., 2020).

The static structural views (Figure 6.2A, 6.11) in the paper were rendered in PyMOL using antibody-bound RBD structures. PBDs are as follows: C105, 6XCN; LY-CoV016, 7C01; C110, 7K8V; C135, 7K8Z; C144, 7K90; C002, 7K8S; C121, 7K8Y. In Figure 6.11B, 7K8T is used instead of 7K8S to illustrate the quaternary antibody epitope. For identifying contact sites to highlight in Figure 6.1B logo plots or to classify sites in Figure 6.2B as contact sites (within 4Åof antibody) or antibody-proximal sites within 4–8Å, the following PDBs were used: 6XCM and 6XCN for C105, 7K8S and 7K8T for C002, 7K8X and 7K8Y for C121, 7K90 for C144, 7K8Z for C135, and 7K8V for C110)(Barnes et al., 2020a, 2020b). Structural distances were computed using the bio3d package version 2.4.0 in R version 3.6.2 [119]. Surface representations of the RBD for non-antibody-bound structures utilize PDB 6M0J [172].

In many of the visualizations, the RBD sites are categorized by epitope region (class 1, class 2, or class 3), defined by [16] and colored accordingly. We define the class 1 epitope as residues 403+405+406+417+420+421+453+455–460+473–476+486+487+489+504, the class 2 epitope to be residues 455+456+472+483–487+489+490+491+492+493+494, and the class 3 epitope to be residues 345+346+437–452+496+498–501. There are 5 residues that overlap between the class 1 and class 2 epitopes: 455+456+486+487+489.

Data Availability

The raw Illumina sequencing data generated in this study have been deposited in the NCBI SRA database under accession code BioProject: PRJNA639956, BioSample SAMN18148595. The data are available without restriction. The processed Illumina sequencing counts for the escape mapping are available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/results/counts/variant_counts.csv`. The raw data tables of mutant escape fractions are at: `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/results/supp_data/all_samples_raw_data.csv`. These

data are the escape fractions shown in Figures 1–6 and Figures 6.8–6.12. The following publicly available datasets were used in the analysis in this study: The PDB accessions for the antibody-S complex structures are: 6XCM and 6XCN for C105, 7C01 for LY-CoV016, 7K8S and 7K8T for C002, 7K8X and 7K8Y for C121, 7K90 for C144, 7K8Z for C135, and 7K8V for C110 [17, 16, 271]. The protein sequences for these antibodies are available at the aforementioned PDB accession numbers. Surface representations of the RBD for non-antibody-bound structures utilize PDB 6M0J [172]. The unmutated SARS-CoV-2 RBD sequence is that from isolate Wuhan-Hu-1 (Genbank accession number MN908947, residues N331–T531). The analysis in Figure 6.6 uses all 765,455 spike sequences on the GISAID EpiCoV database (`https://www.gisaid.org/`) as of May 11, 2021. Together, these are the minimal set of data required to replicate the analysis. All data are available without restriction.

Code Availability

The complete custom code computational pipeline for escape-mapping data analysis is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller` and archived in the Zenodo repository (DOI `10.5281/zenodo.4901733`). Markdown summaries of the escape-mapping data analysis steps are at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/results/summary/summary.md`.

Materials Availability

The SARS-CoV-2 RBD mutant libraries (#1000000172) and unmutated parental plasmid (#166782) are available on Addgene.

6.6 Notes

Acknowledgments

We thank Andrea Loes for experimental assistance and Cathy Lin for administrative support; Dolores Covarrubias, Andy Marty, and the Genomics and Flow Cytometry core facilities at the Fred Hutchinson Cancer Research Center for experimental support; J. Vielmetter, P. Hoffman, and the Protein Expression Center in the Beckman Institute at Caltech for expression assistance.

Funding

This work was supported by the NIAID / NIH (R01AI141707 and R01AI127893 to J.D.B., T32AI083203 to A.J.G., P01 AI138398-S1 to M.C.N. and P.J.B.) and the Gates Foundation (INV-004949). Support was also provided by the Caltech Merkin Institute for Translational Research (P.J.B.), a George Mason University Fast Grant (P.J.B.), and ATAC consortium EC 101003650 (D.F.R.); NIH grants U19 AI111825 and NIH U01 AI151698 for the United World Antiviral Research Network, UWARN (M.C.N., D.F.R.). The Scientific Computing Infrastructure at Fred Hutch is funded by ORIP grant S10OD028685. T.N.S. is a Washington Research Foundation Innovation Fellow at the University of Washington Institute for Protein Design and a Howard Hughes Medical Institute Fellow of the Damon Runyon Cancer Research Foundation (DRG-2381-19). C.O.B was supported by the Hanna Gray Fellowship Program from the Howard Hughes Medical Institute and the Postdoctoral Enrichment Program from the Burroughs Wellcome Fund. J.D.B., P.D.B., and M.C.N. are Investigators of the Howard Hughes Medical Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the US government or the other sponsors.

Author contributions

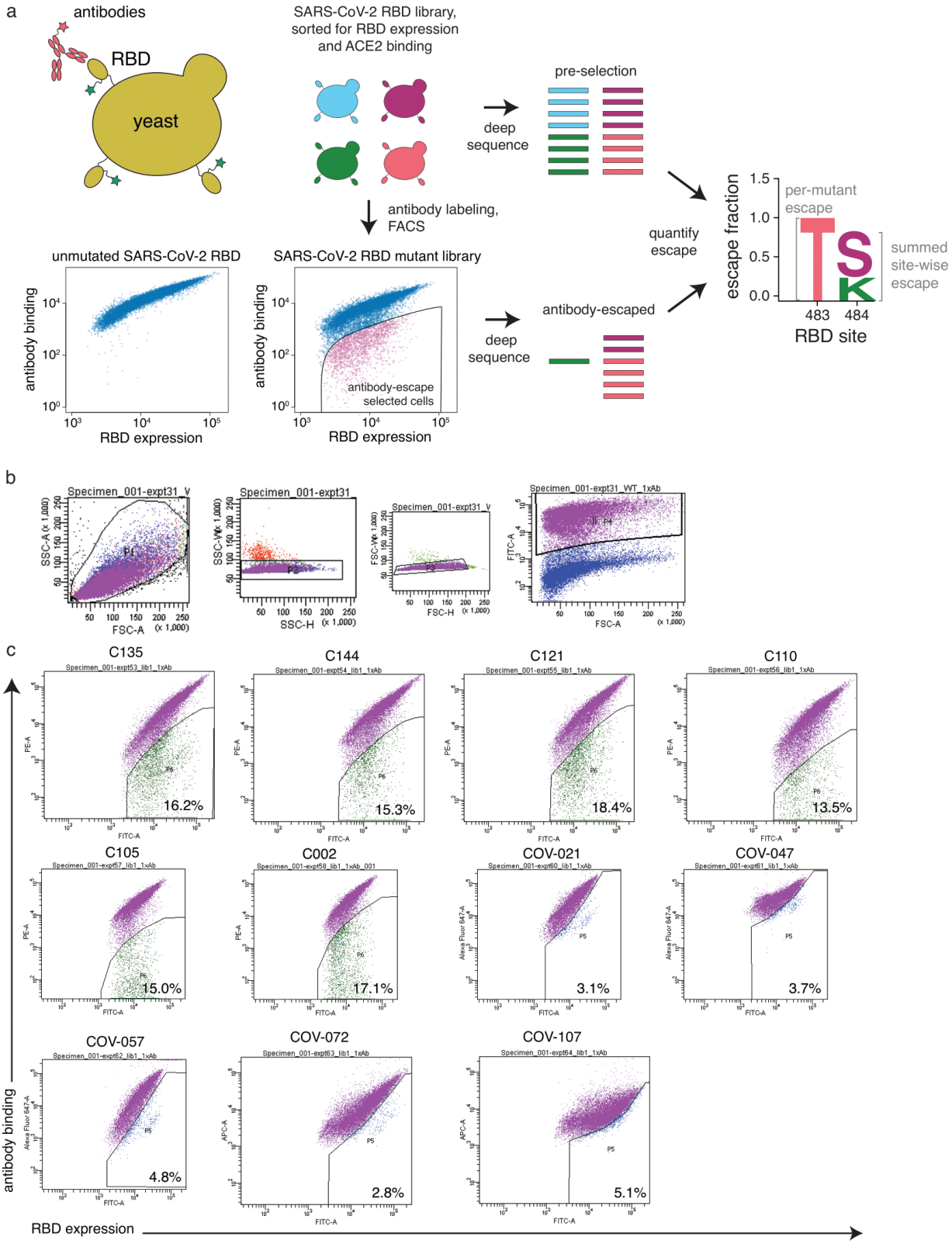
Conceptualization, A.J.G., C.O.B., M.C.N., P.J.B., and J.D.B.; Methodology, A.J.G., T.N.S., and J.D.B.; Investigation, A.J.G.; Software, A.J.G., T.N.S., and J.D.B.; Formal Analysis, A.J.G. and J.D.B.; VSV escape data, Y.W., F.S., D.P.; Resources, M.C.N., D.F.R., M.C., C.G., A.C., M.A., S.F., Z.W., F.M.; Writing – Original Draft, A.J.G. and J.D.B.; Writing – Review and Editing, all authors; Supervision, T.H., P.D.B., M.C.N., P.J.B., and J.D.B.

Declaration of Interests

Subsequent to completion and submission of the initial version of this study, J.D.B. began consulting for Moderna on viral evolution and epidemiology. J.D.B. has the potential to receive a share of IP revenue as an inventor on a Fred Hutch optioned technology/patent (application WO2020006494) related to deep mutational scanning of viral proteins. The Rockefeller University has filed a provisional patent application related to SARS-CoV-2 monoclonal antibodies on which D.F.R. and M.C.N. are inventors. The Rockefeller University has applied for a patent relating to the replication-competent VSV/SARS-CoV-2 chimeric virus on which Y.W, F.S., T.H. and P.B. are inventors (US patent 63/036,124). The remaining authors declare no competing interests.

6.7 Supplementary Figures

Figure 6.7: **Approach for mapping RBD mutations that reduce binding by monoclonal antibodies or polyclonal plasmas**. **(A)** The RBD is expressed on the surface of yeast. Flow cytometry is used to quantify both RBD expression (via a C-terminal MYC tag) and antibody binding to the RBD protein expressed on the surface of each yeast cell. A library of yeast expressing different RBD mutants were incubated with antibodies or plasmas and binding was detected using a IgG or IgA+IgG+IgM secondary antibody for monoclonal antibodies or polyclonal plasmas, respectively. We then used FACS to enrich for cells expressing RBD that bound reduced levels of antibody, and deep sequencing to quantify the frequency of each mutation in the initial and antibody-escape cell populations. We quantified the effect of each mutation as the "escape fraction," which represents the fraction of cells expressing RBD with that mutation that fell in the antibody escape FACS bin. Escape fractions are represented in logo plots, with the height of each letter proportional to the effect of that amino-acid mutation on antibody binding. The site-level escape metric is the sum of the escape fractions of all mutations at a site. Note that both experimental and computational filtering steps were used to remove RBD mutants that were misfolded or completely unable to bind the ACE2 receptor (see Methods). **(B)** Representative plots of nested FACS gating strategy used for all plasma-selection experiments to select for single cells (SSC-A vs. FSC-A, SSC-W vs. SSC-H, and FSC-W vs. FSC-H) that also express RBD (FITC-A vs. FSC-A). **(C)** FACS gating strategy for one of two independent libraries to select cells expressing RBD mutants with reduced binding by monoclonal antibodies or polyclonal plasmas (cells in blue). Gates were set manually during sorting. Different strategies were used for monoclonal antibodies vs. polyclonal plasmas. For monoclonal antibodies, selection gates were set to capture up to 95% of yeast cells expressing unmutated RBD, stained with an antibody concentration 100x lower than that used for library staining. For polyclonal plasmas, selection gates were set to capture 3–6% of the RBD+ library. The same gate was set for both independent libraries stained with each plasma, and the FACS scatter plots looked qualitatively similar between the two libraries. For information on the fraction of library cells that fall into each selection gate, see `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Rockefeller/blob/main/data/SupplementaryTable2.xlsx`.


a
antibodies
RBD
yeast
SARS-CoV-2 RBD library, sorted for RBD expression and ACE2 binding
deep sequence
pre-selection
antibody labeling, FACS
unmutated SARS-CoV-2 RBD
SARS-CoV-2 RBD mutant library
antibody binding
RBD expression
antibody-escape selected cells
deep sequence
antibody-escaped
quantify escape
escape fraction
per-mutant escape
summed site-wise escape
RBD site
b
c
C135
C144
C121
C110
C105
C002
COV-021
COV-047
COV-057
COV-072
COV-107
16.2%
15.3%
18.4%
13.5%
15.0%
17.1%
3.1%
3.7%
4.8%
2.8%
5.1%
antibody binding
RBD expression

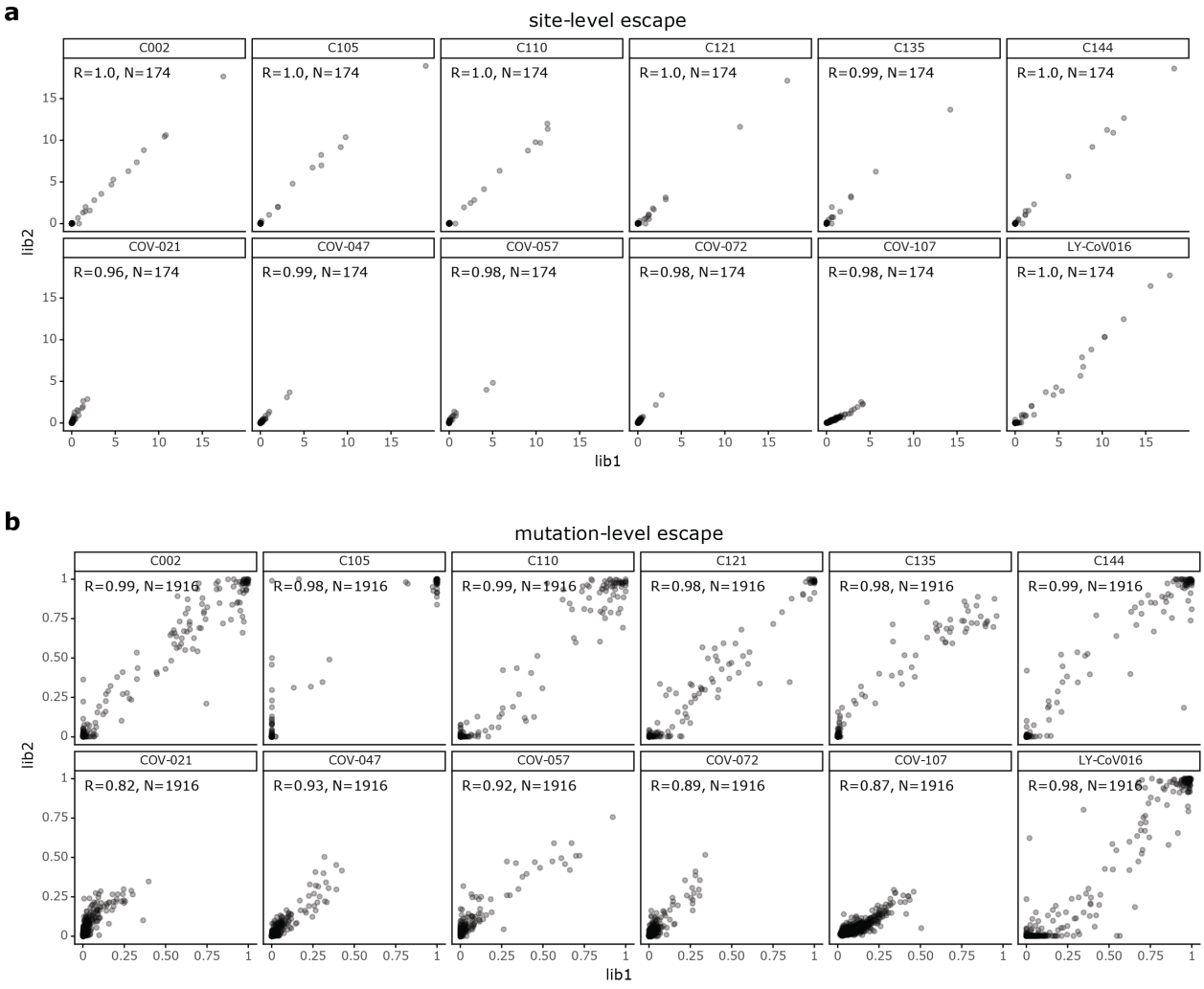


Figure 6.8: **Correlations between replicates for site- and mutation-level escape metrics**. **(A)** Correlation plots of site-level escape for each of the two independent RBD mutant libraries for each antibody or plasma. Each point represents one site in the RBD. **(B)** Correlation plots of mutation-level escape for each of the two independent RBD mutant libraries for each antibody or plasma. In this plot, each point represents a different mutation.

Figure 6.9: **Escape maps for class 1, 2, or 3 antibodies we have profiled here or in previous studies**. Escape maps for **(A)** class 1, **(B)** class 2, or **(C)** class 3 antibodies shown in Figure 6.1C. All escape maps were previously generated [83, 124, 278, 279] except for C105, C144, C002, C121, C135, and C110 which are new to this study. Different sets of key sites are shown for each of the three antibody classes (see Methods). Sites are colored by RBD epitope as in Figure 6.1, also shown in panel **(D)**. For the escape maps of class 4 antibodies shown in Figure 6.1C, see [124].

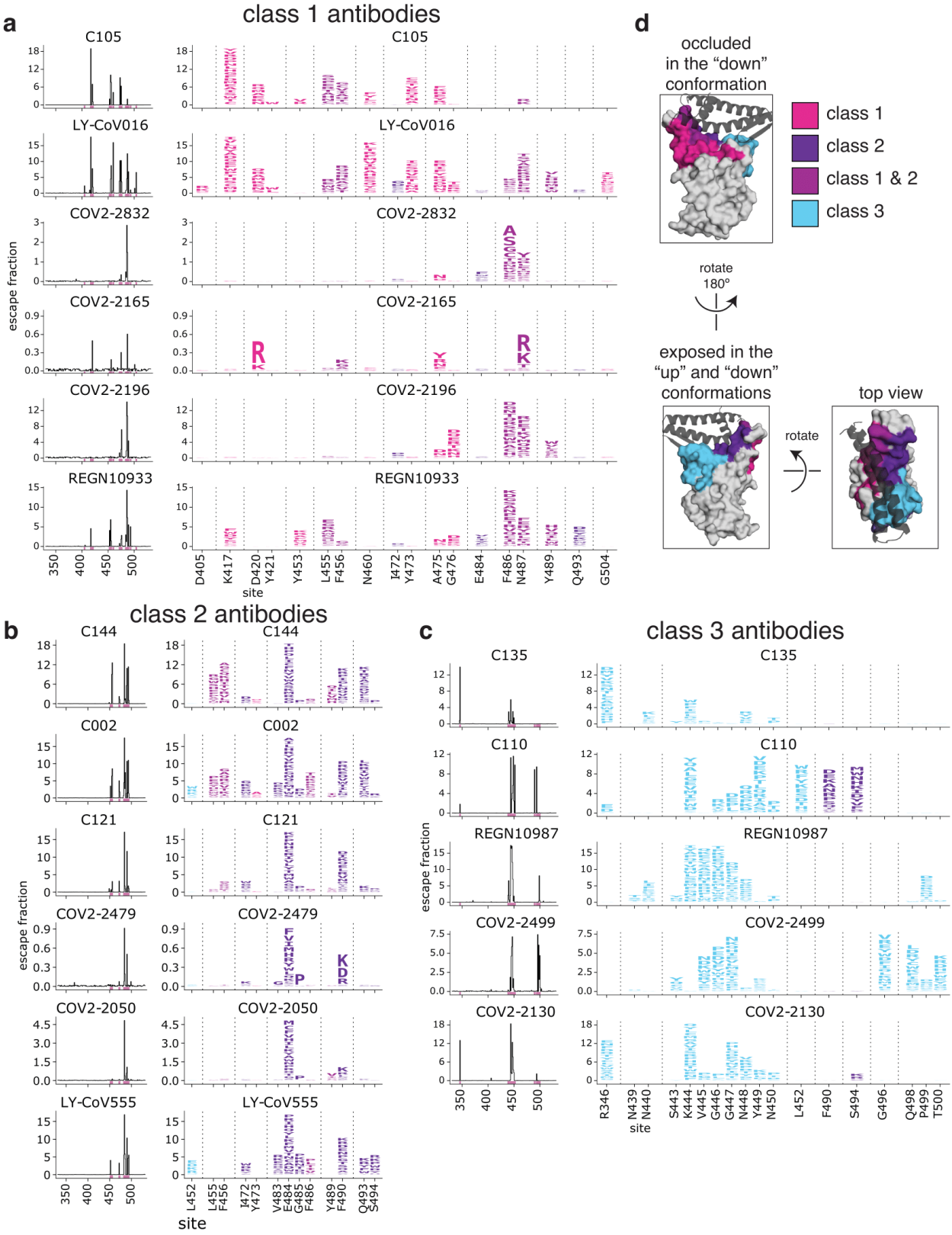

a
class 1 antibodies
C105
LY-CoV016
COV2-2832
COV2-2165
COV2-2196
REGN10933
escape fraction
site
D405 K417 D420 Y421 Y453 L455 F456 N460 I472 Y473 A475 G476 E484 F486 N487 Y489 Q493 G504
d
occluded in the "down" conformation
class 1
class 2
class 1 & 2
class 3
rotate 180°
exposed in the "up" and "down" conformations
rotate
top view
b
class 2 antibodies
C144
C002
C121
COV2-2479
COV2-2050
LY-CoV555
L452 L455 F456 I472 Y473 V483 E484 G485 F486 Y489 F490 Q493 S494
c
class 3 antibodies
C135
C110
REGN10987
COV2-2499
COV2-2130
R346 N439 N440 S443 K444 V445 G446 G447 N448 Y449 N450 L452 F490 S494 G496 Q498 P499 T500

Figure 6.10: **Logo plots colored according to effects of mutations on ACE2 binding and RBD expression**. Logo plots show the escape fractions of each mutation at key sites (any site called as a site of strong escape for any antibody or plasma). Letters are colored by how mutations affect RBD affinity for ACE2 or RBD expression as measured via yeast display [280], with yellow indicating poor affinity or expression and brown indicating good affinity and expression. The top 6 rows of logo plots are for monoclonal antibodies; the bottom 5 rows are for polyclonal plasmas.

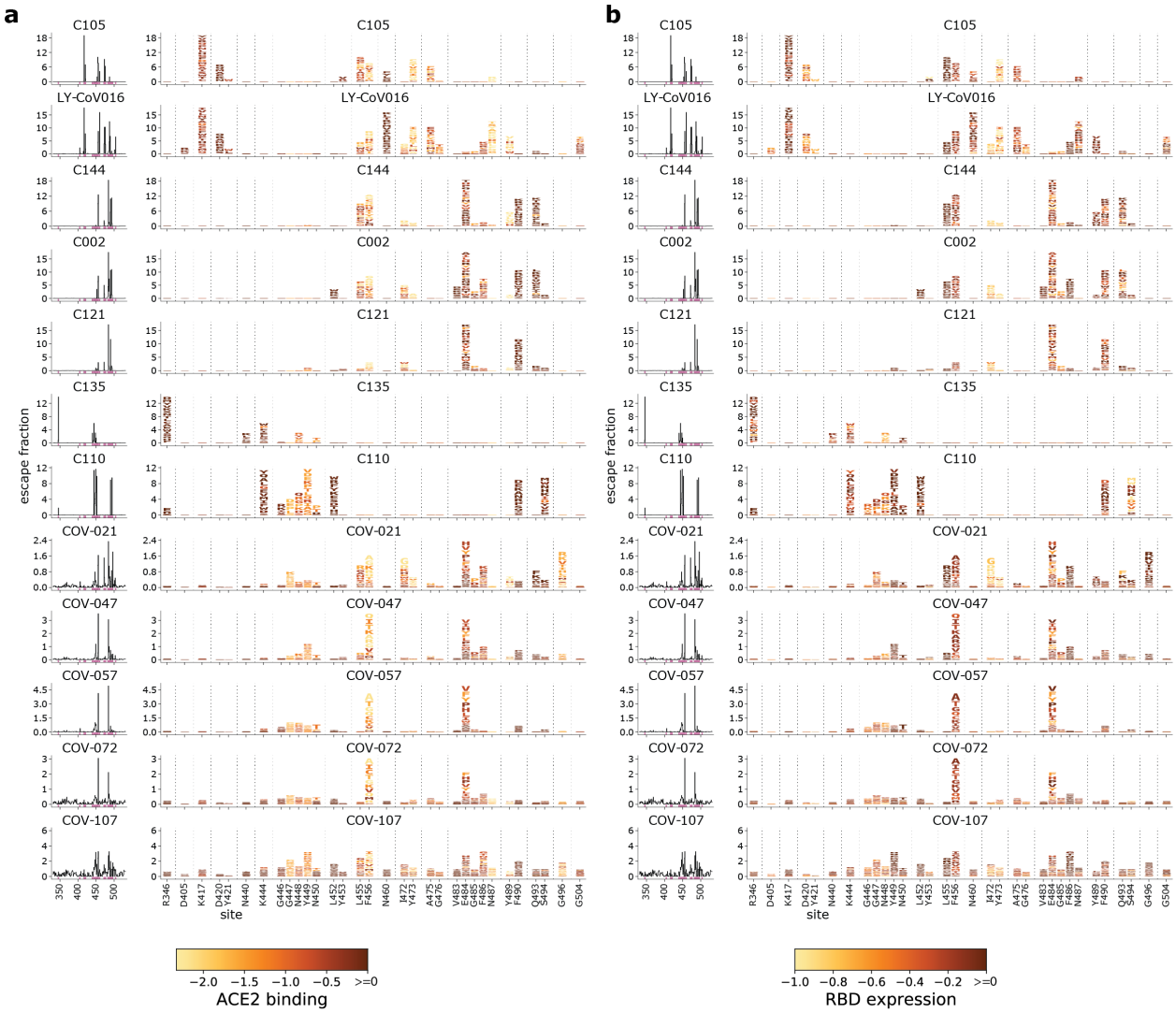

a
b
C105
LY-CoV016
C144
C002
C121
C135
C110
COV-021
COV-047
COV-057
COV-072
COV-107
escape fraction
site
ACE2 binding
−2.0 −1.5 −1.0 −0.5 >=0
RBD expression
−1.0 −0.8 −0.6 −0.4 −0.2 >=0

Figure 6.11: **Visualization of the maximum escape at a site mapped onto cartoon representations of antibody-bound RBD**. **(A)** Mapping of maximum antibody escape at a site to the antibody-bound RBD. Antibody-contact sites on the RBD (within 4Å) are shown as spheres. Sites with no escape measurements due to excessive functional constraint on the site are shown in dark gray. Each site is colored according to the maximum escape fraction of any mutation at that site (whereas Figure 6.2 shows site total escape), scaled from white (no escape) to red (maximum escape for any mutation for that antibody). Inset panels at right indicate key RBD-antibody interactions where mutations to the indicated RBD site disrupt antibody binding. RBD color scale indicates site total escape, as in Figure 6.2A. **(B)** Visualization of class 2 antibody quaternary epitopes. The total escape at each site is mapped onto the surface of the Fab-bound RBD as in Figure 6.2A, with white indicating no escape, and red indicating the site with the most escape from that antibody. Sites where no mutations are tolerated are indicated in dark gray. Antibody quaternary contact sites are shown in periwinkle. The C144 antibody binds to spike trimer in the "all RBDs down" conformation and forms a quaternary epitope that bridges across two adjacent RBDs by binding to a hydrophobic RBD cavity at the base of the N343 N-linked glycan. The C002 and C121 antibodies, when bound to a downRBD, can form a quaternary epitope with an adjacent up-RBD. The up-RBD also contacts another C121 Fab [16]. Our yeast-display system utilizes monomeric RBD and therefore does not map escape mutations to quaternary contact sites. These results thus cannot be used to determine the importance of the quaternary sites for antibody binding. Previous work, however, has shown that the V367F RBD mutation to the C144 quaternary epitope does not affect neutralization of pseudotyped lentivirus by C144 [320]. See Methods for PDB accession codes used to generate structural representations. Antibody heavy chains are shown in dark blue, and light chains are in light blue.

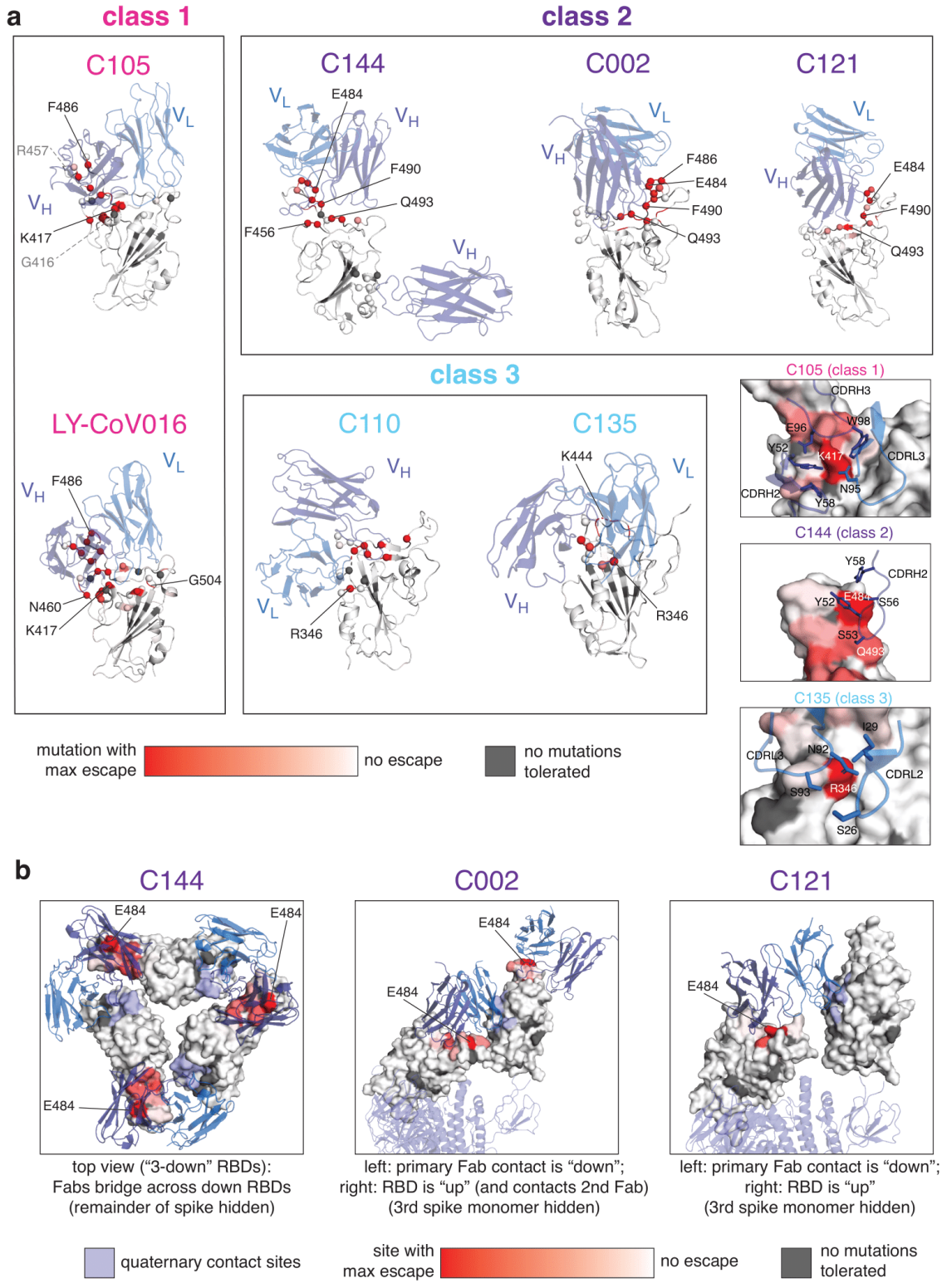
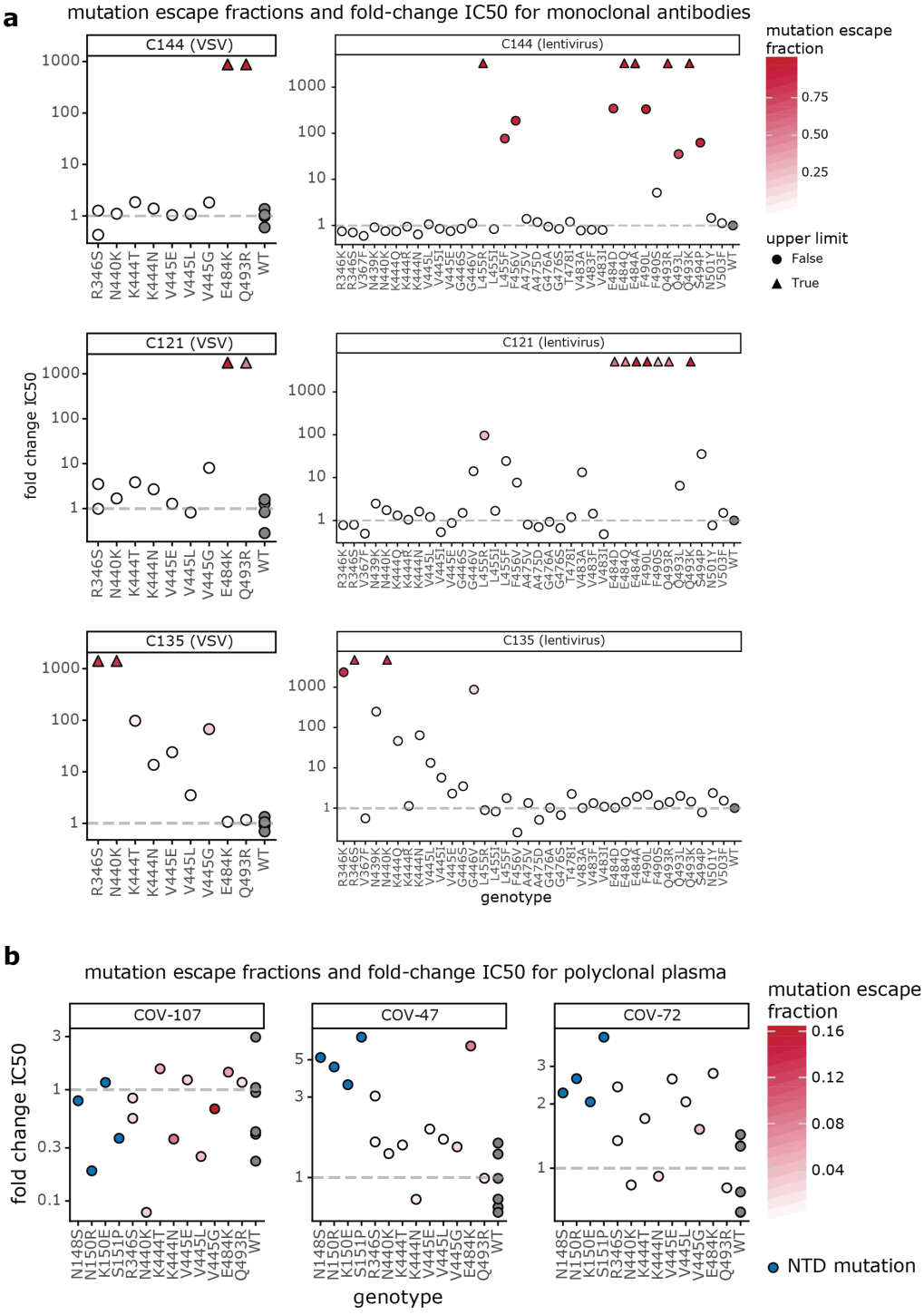

a
class 1
class 2
C105
F486
R457
VL
VH
K417
G416
C144
E484
VL
VH
F490
Q493
F456
VH
C002
VL
VH
F486
E484
F490
Q493
C121
VL
VH
E484
F490
Q493
class 3
LY-CoV016
VL
VH
F486
G504
N460
K417
C110
VH
VL
R346
C135
K444
VL
VH
R346
C105 (class 1)
CDRH3
W98
E96
Y52
K417
CDRL3
CDRH2
N95
Y58
C144 (class 2)
Y58
CDRH2
Y52
E484
S56
S53
Q493
C135 (class 3)
I29
CDRL3
N92
CDRL2
S93
R346
S26
mutation with max escape
no escape
no mutations tolerated
b
C144
E484
E484
E484
C002
E484
E484
C121
E484
top view (“3-down” RBDs): Fabs bridge across down RBDs (remainder of spike hidden)
left: primary Fab contact is “down”; right: RBD is “up” (and contacts 2nd Fab) (3rd spike monomer hidden)
left: primary Fab contact is “down”; right: RBD is “up” (3rd spike monomer hidden)
quaternary contact sites
site with max escape
no escape
no mutations tolerated

Figure 6.12: **The previously measured effects of spike mutations on neutralization for 3 monoclonal antibodies and 3 polyclonal plasmas**. **(A)** The effects of mutations on neutralization of chimeric VSV encoding the SARS-CoV-2 (left) or spike-pseudotyped lentivirus particles (right) by monoclonal antibodies from [320]. The y-axis shows the fold-change in IC50 compared to the Wuhan-Hu-1-like spike, such that larger numbers are greater reductions in neutralization sensitivity. Mutations that had IC50s at or above the limit of detection are indicated as triangles. Points are colored according to their mutation escape fraction (Figure 6.1, Supplementary Data 1). Wildtype is in gray. **(B)** The effects of mutations on neutralization of spike-encoding chimeric VSV by polyclonal antibodies from [320]. Wildtype spike IC50 values are shown in gray. NTD mutations are shown in blue. Escape fraction color scales are independent in (A) and (B).


a
mutation escape fractions and fold-change IC50 for monoclonal antibodies
C144 (VSV)
C144 (lentivirus)
C121 (VSV)
C121 (lentivirus)
C135 (VSV)
C135 (lentivirus)
fold change IC50
genotype
mutation escape fraction
upper limit
False
True
b
mutation escape fractions and fold-change IC50 for polyclonal plasma
COV-107
COV-47
COV-72
fold change IC50
genotype
mutation escape fraction
NTD mutation

Chapter 7

ANTIBODIES ELICITED BY MRNA-1273 VACCINATION BIND MORE BROADLY TO THE RECEPTOR BINDING DOMAIN THAN DO THOSE FROM SARS-COV-2 INFECTION

A version of this chapter has been previously published as:

Greaney AJ, Loes AN, Gentles LE, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD. Antibodies elicited by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from SARS-CoV-2 infection. Sci Transl Med. 2021 Jun 30;13(600):eabi9915. doi: 10.1126/scitranslmed.abi9915. Epub 2021 Jun 8. PMID: 34103407; PMCID: PMC8369496.

7.1 Abstract

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants with mutations in key antibody epitopes has raised concerns that antigenic evolution will erode adaptive immunity elicited by prior infection or vaccination. The susceptibility of immunity to viral evolution is shaped in part by the breadth of epitopes targeted. To investigate how vaccine responses may be influenced by viral evolution, we used deep mutational scanning to compare the specificity of antibodies elicited by the mRNA-1273 vaccine versus natural infection with SARS-CoV-2. The neutralizing activity of vaccine-elicited antibodies is even more focused on the spike receptor-binding domain (RBD) than for infection-elicited antibodies. However, within the RBD, binding of vaccine-elicited antibodies is more broadly distributed across epitopes than for infection-elicited antibodies. This greater bind-

ing breadth suggests that single RBD mutations have less impact on neutralization by vaccine sera than convalescent sera. Therefore, antibody immunity acquired by infection or different modes of vaccination may have differing susceptibility to erosion by viral evolution.

One Sentence Summary: Deep mutational scanning shows the mRNA-1273 RBD-binding antibody response is less affected by single viral mutations than the infection response.

7.2 Introduction

Mitigation of the coronavirus disease 2019 (COVID-19) pandemic will depend on population immunity acquired via infection with or vaccination against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Unfortunately, humans are repeatedly re-infected with the endemic "common-cold" coronaviruses [263], at least in part because these viruses evolve to escape neutralizing antibody immunity elicited by prior infection [89]. SARS-CoV-2 is already undergoing similar antigenic evolution, with the recent emergence of new viral lineages with reduced neutralization by antibodies elicited by infection and vaccination [42, 51, 110, 313, 315, 324]. Preliminary results suggest that immunity still provides substantial protection against infection and severe disease [2, 1] caused by these new viral lineages; however, if SARS-CoV-2 is similar to other human coronaviruses, then, at minimum, the protection against reinfection will eventually be eroded by viral evolution.

However, unlike for other human coronaviruses, a large fraction of the population is acquiring SARS-CoV-2 immunity from vaccination rather than infection. The first two vaccines approved for emergency use in the United States were Moderna's mRNA-1273 and Pfizer/BioNTech's BNT162b2. Both mRNA vaccines encode the full SARS-CoV-2 spike ectodomain with a transmembrane anchor and stabilizing S-2P mutations [63]. It is possible that these vaccines could elicit antibodies with distinct specificities compared to natural infection due to variation in the spike (such as the S-2P mutations) or divergent immune responses to a two-dose mRNA vaccine versus infection. If the specificities differ, this could

influence the impact of viral evolution on SARS-CoV-2 immunity.

To address this question, we used a combination of serological assays and deep mutational scanning to map the specificity of the human polyclonal antibody response after two doses of the mRNA-1273 vaccine. The vaccine elicited neutralizing activity that is even more focused on the spike receptor-binding domain (RBD) than infection-elicited immunity. However, within the RBD, binding by vaccine-elicited antibodies was often less affected by single mutations. As a result, common RBD mutations sometimes eliminated less of the neutralizing activity of mRNA-1273 vaccine sera than convalescent sera, and vaccine sera retained substantial RBD-directed neutralization even in the presence of mutations to three major RBD neutralizing epitopes.

7.3 Results

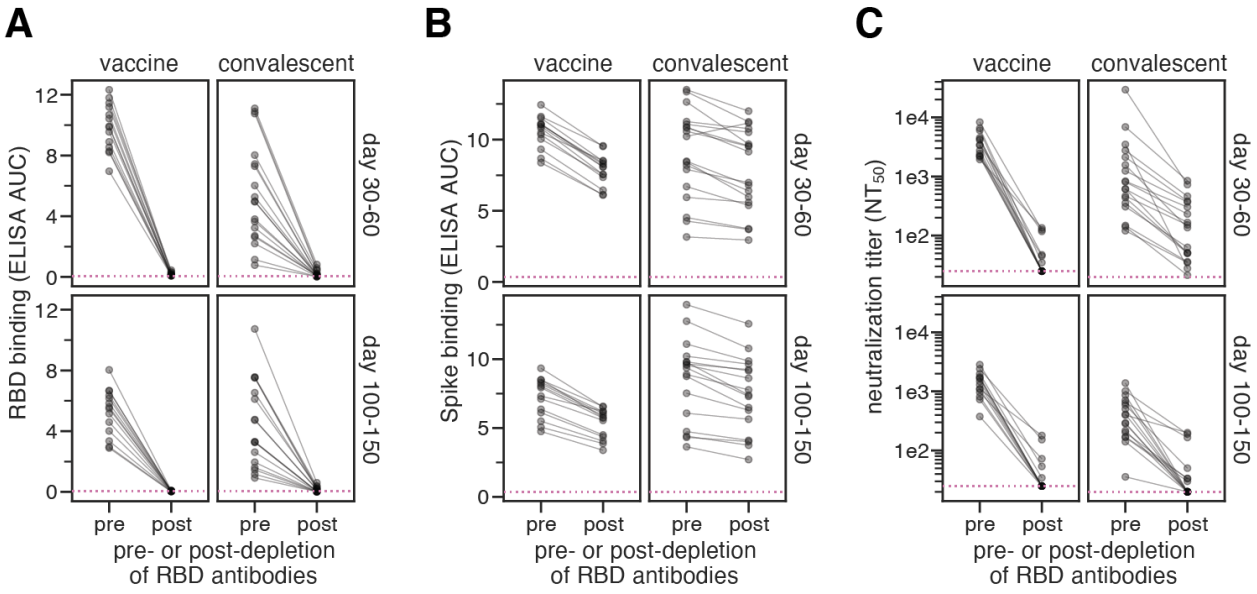
7.3.1 The neutralizing activity of mRNA-1273 vaccine-elicited antibodies is more RBD-focused than for infection-elicited antibodies

We studied sera from adults (ages 18–55 years) who received two doses of the Moderna mRNA-1273 vaccine in phase 1 clinical trials [147]. The majority of our study focused on 14 individuals who received the 250 μg dose, although we validated key conclusions with a smaller subset of eight trial participants who received the 100 μg dose. The sera were collected at 36 and 119 days after the first vaccine dose, corresponding to 7 and 90 days after the second dose. It was previously shown that these individuals had high amounts of binding and neutralizing antibodies against SARS-CoV-2, with neutralizing antibody titers within the upper quartile of sera from SARS-CoV-2 convalescent individuals [147]. Throughout, we compared vaccine sera to convalescent plasma or serum samples from two independent cohorts [66, 251]. The convalescent plasma samples were characterized in earlier studies [66, 122, 120, 251], and grouped into an early time point of 15–60 days post-symptom onset and a late time point of 100–150 days post-symptom onset.

The majority of the neutralizing activity of convalescent sera and plasma is due to RBD-binding antibodies [74, 120, 229]. To determine if neutralization by vaccine sera is similarly RBD-targeted, we depleted RBD-binding antibodies from the day 36 and 119 sera isolated from 14 individuals who received the 250 μg dose of the mRNA-1273 vaccine. We then measured serum IgG binding to the RBD and full spike ectodomain before and after depletion. As expected, depletion removed all RBD-binding antibodies (Fig. 7.1A, Fig. 7.6A and B). However, depleting RBD-binding antibodies only moderately decreased spike-binding activity in either vaccine sera or convalescent plasma (Fig. 7.1B, Fig. 7.6B), consistent with studies showing that a minority of spike-binding vaccine-elicited B cells target the RBD [8, 315].

To determine the contribution of RBD-binding antibodies to neutralization, we measured the neutralizing activity of vaccine sera before and after depleting RBD-binding antibodies using spike-pseudotyped lentiviral particles. For samples isolated from 13 of 14 vaccinated individuals, greater than 90% of the neutralizing activity at both time points was dependent upon RBD-binding antibodies (Fig. 7.1C and D, `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna/blob/main/experimental_data/results/rbd_absorptions/data_file_s1.csv`). For 17 of 28 vaccine sera, depletion of RBD-binding antibodies reduced the neutralization titer (reciprocal IC50) from > 1000 to < 25 (Fig. 7.1C and D, Fig. 7.6C and D). The percent neutralizing activity due to RBD-binding antibodies was higher for vaccine sera than for convalescent plasma samples collected between day 15 and 60 ($p=1.0 \times 10^{-6}$, Fig. 7.1C and D) [120]. These assays were performed in 293T cells over-expressing human angiotensin-converting enzyme 2 (ACE2), which may underestimate contributions of non-RBD-binding antibodies to viral neutralization [51, 56, 283]. Nonetheless, because the same assay was used for vaccine and convalescent samples, we conclude that the neutralizing activity of the antibody response elicited by the mRNA-1273 vaccine is more focused on the RBD than for infection-elicited antibodies.

Figure 7.1: **RBD-binding antibodies are responsible for most neutralizing activity of mRNA-1273 vaccine-elicited sera**. **(A)** Binding of serum antibodies to SARS-CoV-2 RBD, as measured by ELISA area-under-the-curve (AUC), for vaccine-elicited sera and convalescent plasma before and after depletion of RBD-binding antibodies. The dashed pink line indicates binding of pre-pandemic sera. **(B)** Binding of serum antibodies to the full spike ectodomain. The y-axis scale units in (A) and (B) are not comparable between samples from vaccinated and convalescent individuals due to different dilution factors (beginning at 1:500 for vaccine sera and 1:100 for convalescent plasma samples). **(C)** Neutralization titer (NT50) of vaccine-elicited sera and convalescent plasma samples before and after depletion of RBD-binding antibodies. The limit of detection is shown as a dashed horizontal pink line. **(D)** Percent of neutralizing activity of vaccine-elicited sera and convalescent plasma samples due to RBD-binding antibodies. P-values are from a log-rank test accounting for censoring. n=17 for each time point for convalescent plasma samples, and n=14 for each time point for vaccine sera. ns, not significant. All measurements of convalescent plasma binding and neutralization were previously reported in [120]


A
vaccine
convalescent
RBD binding (ELISA AUC)
day 30-60
day 100-150
pre
post
pre- or post-depletion
of RBD antibodies
B
Spike binding (ELISA AUC)
C
neutralization titer (NT50)
D
day 15-60
day 100-150
percent neutralizing potency
due to RBD antibodies
p=1e-6
p=0.15 (ns)

7.3.2 Complete mapping of RBD mutations that reduce binding by vaccine-elicited sera at 119 days post-vaccination reveals broad binding specificity across multiple RBD epitopes

We used deep mutational scanning [124, 120] to map all mutations to yeast-displayed RBD that reduced vaccine serum antibody binding. Our experiments utilized duplicate libraries containing 3,804 of the 3,819 possible single amino-acid mutations to the RBD of the Wuhan-Hu-1 strain of SARS-CoV-2, 2,034 of which are tolerated for proper protein folding and at least modest ACE2 binding [280]. We incubated the yeast-displayed libraries with each serum, and used fluorescence-activated cell sorting (FACS) to enrich for the 3–5% of cells expressing RBD mutants with the lowest amount of serum binding (Fig. 7.7, 7.8, `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna/blob/main/data/TableS2_FACSinfo.csv`). The degree to which mutations reduce serum binding varies across samples, so the FACS gates were set separately for each sample. We used deep sequencing to quantify the "escape fraction" for each of the 2,034 tolerated RBD mutations against each serum by determining the frequency of each mutant in the serum-escape bin versus the original unsorted population. These escape fractions range from 0 (no cells with the mutation in the serum-escape bin) to 1 (all cells with the mutation in the serum-escape bin) (`https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna/blob/main/results/supp_data/moderna_convalescent_all_raw_data.csv`). Correlations between escape fractions measured for independent biological replicate libraries are shown in Fig. 7.9. We represent the escape maps as logo plots, where the height of each letter is proportional to its escape fraction (Fig. 7.2, Fig. 7.10 and S6).

The escape maps for sera collected at day 119 from individuals who received the 250 μg vaccine dose fell into four qualitative categories [16] (Fig. 7.2A and B). For 5 of 14 individuals, escape from antibody binding was focused on RBD sites 456 and 484 (Fig. 7.2B and C, Fig. 7.10). These two sites are on the receptor-binding ridge in the neutralizing

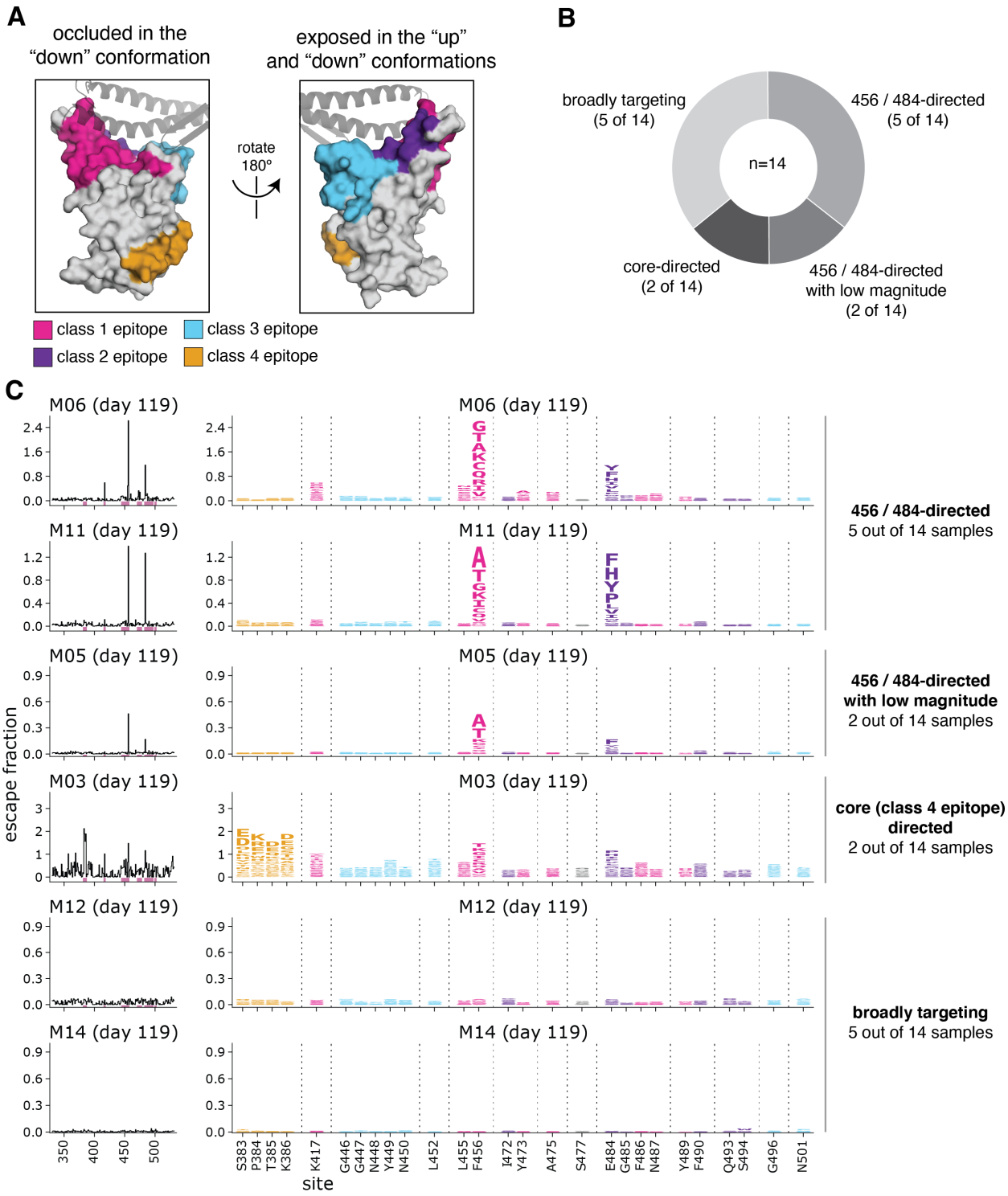
"class 1" and "class 2" RBD epitopes, respectively [16] (Fig. 7.2A). Two more individuals also had escape maps that were focused on sites 456 and 484, but with a very low overall magnitude of escape (Fig. 7.2B and C, Fig. 7.10). For 2 of 14 individuals, serum binding was most affected by mutations in the "class 4" epitope located in the core RBD, including sites 383 to 386 (Fig. 7.2, Fig. 7.10). Antibodies targeting the class 4 epitope are often non-neutralizing or less potently neutralizing than antibodies targeting the receptor-binding motif [74, 188, 229, 356]. The escape maps for the remaining 5 individuals were "flat," meaning that no single mutation had a large effect on serum binding, suggestive of broad binding to multiple RBD epitopes (Fig. 7.2B and C, Fig. 7.10).

To determine if the vaccine dose affected the RBD binding specificity of the polyclonal antibody response, we mapped binding escape from the day 119 sera from 8 individuals vaccinated with 100 μg rather than 250 μg doses. The escape maps of the 100 μg cohort resembled those of the 250 μg cohort and fell into the 456/484-targeting, core-targeting, or "flat" categories (Fig. 7.11). Although the sample sizes are small, and a higher fraction of the 100 μg dose escape maps were "flat" than for the 250 μg cohort (4/8 versus 5/14, respectively), this suggests 100 and 250 ug doses elicit antibody responses similar in the breadth of their RBD binding specificity.

7.3.3 Binding escape maps become more focused to specific sites in the RBD from 36 days to 119 days post-vaccination

To examine longitudinal changes in binding specificity of vaccine-elicited serum antibodies to the RBD, we also determined binding-escape maps for sera collected at day 36 post-vaccination from five individuals who received the 250 μg dose (Fig. 7.3). All of these day-36 sera had relatively "flat" escape maps, meaning that no single mutation had a large effect on serum binding (Fig. 7.3A). However, by day 119, the escape maps for most individuals were more focused on specific sites in the RBD (Fig. 7.3B). Specifically, for four of five individuals,

Figure 7.2: **Complete maps of RBD mutations that reduce binding by serum collected 119 days post-vaccination with the 250 μg dose of mRNA-1273**. **(A)** The epitopes of four major classes [16] of RBD-binding antibodies are colored on the RBD surface (PDB 6M0J). ACE2 is shown as a gray ribbon diagram. **(B)** Number of sera that fell into each of the four major categories of binding-escape maps as categorized by subjective visual inspection. **(C)** Escape maps for six representative sera are shown. The line plots at left indicate the sum of effects of all mutations at each RBD site on serum antibody binding, with larger values indicating more escape. The logo plots at right show key sites (highlighted in purple on the line plot x-axes). The height of each letter is that mutation's escape fraction; larger letters indicate a greater reduction in binding. Escape fractions are not strictly comparable between samples due to the use of sample-specific FACS selection gates—therefore, for each sample, the y-axis is scaled independently. RBD sites are colored by epitope as in (A). The escape fractions were correlated between independent libraries, and we report the average of duplicate measurements throughout. Interactive versions of logo plots and structural visualizations are at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Moderna/`.


A
occluded in the "down" conformation
exposed in the "up" and "down" conformations
rotate 180°
class 1 epitope
class 3 epitope
class 2 epitope
class 4 epitope
B
broadly targeting (5 of 14)
456 / 484-directed (5 of 14)
n=14
core-directed (2 of 14)
456 / 484-directed with low magnitude (2 of 14)
C
M06 (day 119)
M11 (day 119)
M05 (day 119)
M03 (day 119)
M12 (day 119)
M14 (day 119)
escape fraction
site
456 / 484-directed
5 out of 14 samples
456 / 484-directed with low magnitude
2 out of 14 samples
core (class 4 epitope) directed
2 out of 14 samples
broadly targeting
5 out of 14 samples

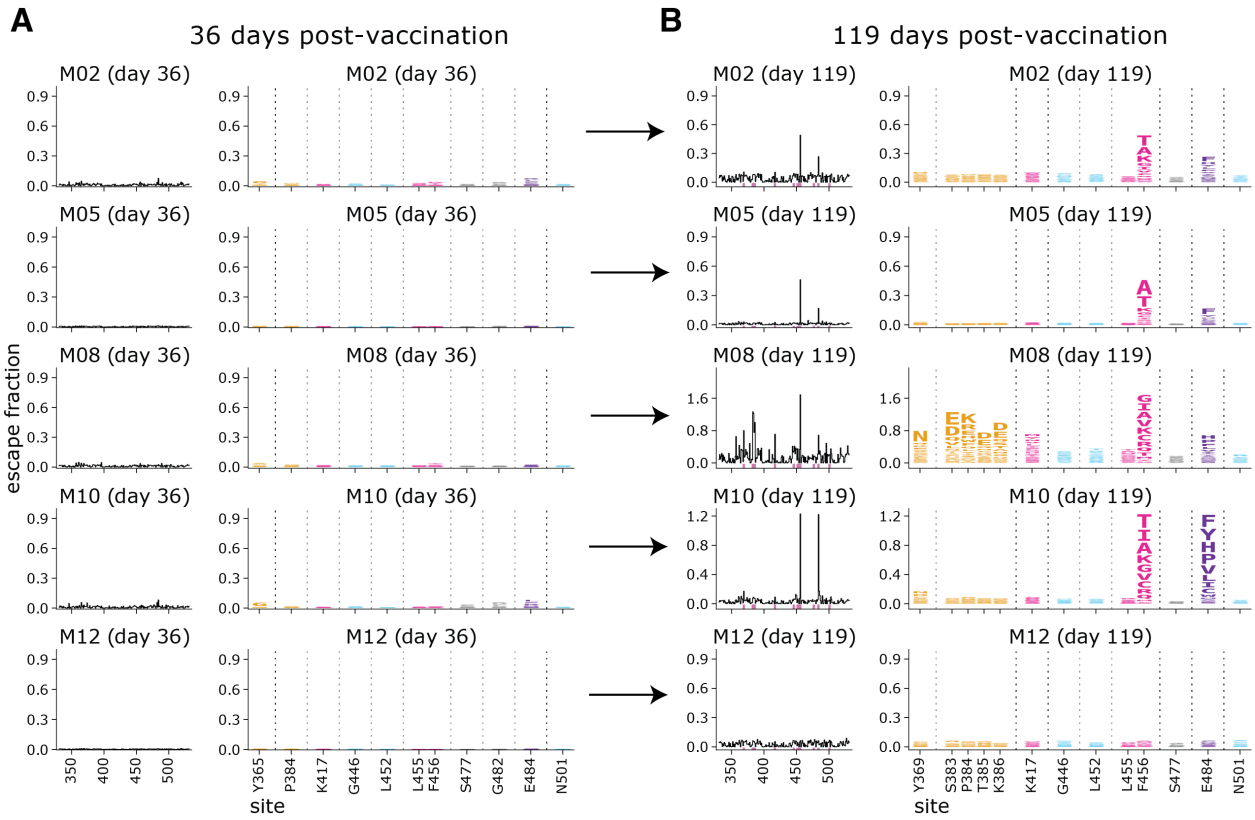
the escape maps became focused on RBD sites 456 and 484 (Fig. 7.3B). For one of these individuals, the focusing on sites 456 and 484 was accompanied by increased focusing on the class 4 epitope, including sites 383–386. Only one individual, M12, had a day-119 escape map as flat as the day-36 escape map. These results suggest that, as the vaccine-induced RBD-binding antibody response matures over time, it becomes more focused on specific sites in the RBD.

7.3.4 RBD binding by vaccine-elicited serum samples is broader than for convalescent plasma samples

To elucidate differences in the specificity of the RBD-binding antibody response elicited by vaccination versus infection, we compared the vaccine-sera escape maps to ones that we previously determined for convalescent plasma samples [120, 122]. At both 15–60 day and 100–150 day ranges, the convalescent escape maps were more focused on specific RBD sites than the vaccine escape maps (Fig. 7.4A). The difference was especially striking at the early time point, where the day 36 vaccine samples all had flat escape maps, whereas the convalescent samples often had escape maps indicating that antibody binding was strongly affected by mutations at specific RBD sites such as 456 and 484 (Fig. 7.4A). The difference between the vaccine and convalescent samples was less striking at the later time point, but the convalescent maps were still more focused than the vaccine maps, as demonstrated by the lower magnitude of the escape fractions. There were also differences in the RBD sites where mutations affected binding for the vaccine versus convalescent samples. Although most samples of both types were affected by mutations at sites 456 and 484, the convalescent samples tended to also be affected by mutations to the 443–450 loop in the class 3 epitope, whereas mutations in the class 4 epitope spanning sites 383–386 sometimes had a more pronounced effect on the vaccine samples (Fig. 7.2 and 7.4A, Fig. 7.10).

To visualize relationships between vaccine- and infection-elicited antibody responses, we

Figure 7.3: **Comparison of escape maps for sera collected at day 36 and day 119 post-vaccination shows that the RBD-binding response focuses over time. (A,B)** Escape maps for sera at day 36 (A) and day 119 (B) from 5 individuals who received the 250 μg vaccine dose are shown. The day 36 maps are all relatively flat, indicating no RBD mutation has a large effect on serum antibody binding. By day 119, the maps are often more focused on sites 456 and 484. The y-axis is scaled separately for each serum sample. Interactive versions are at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Moderna/`.


A
36 days post-vaccination
B
119 days post-vaccination
M02 (day 36)
M05 (day 36)
M08 (day 36)
M10 (day 36)
M12 (day 36)
M02 (day 119)
M05 (day 119)
M08 (day 119)
M10 (day 119)
M12 (day 119)
escape fraction
site
350
400
450
500
Y365
P384
K417
G446
L452
L455
F456
S477
G482
E484
N501
Y369
S383
T385
K386

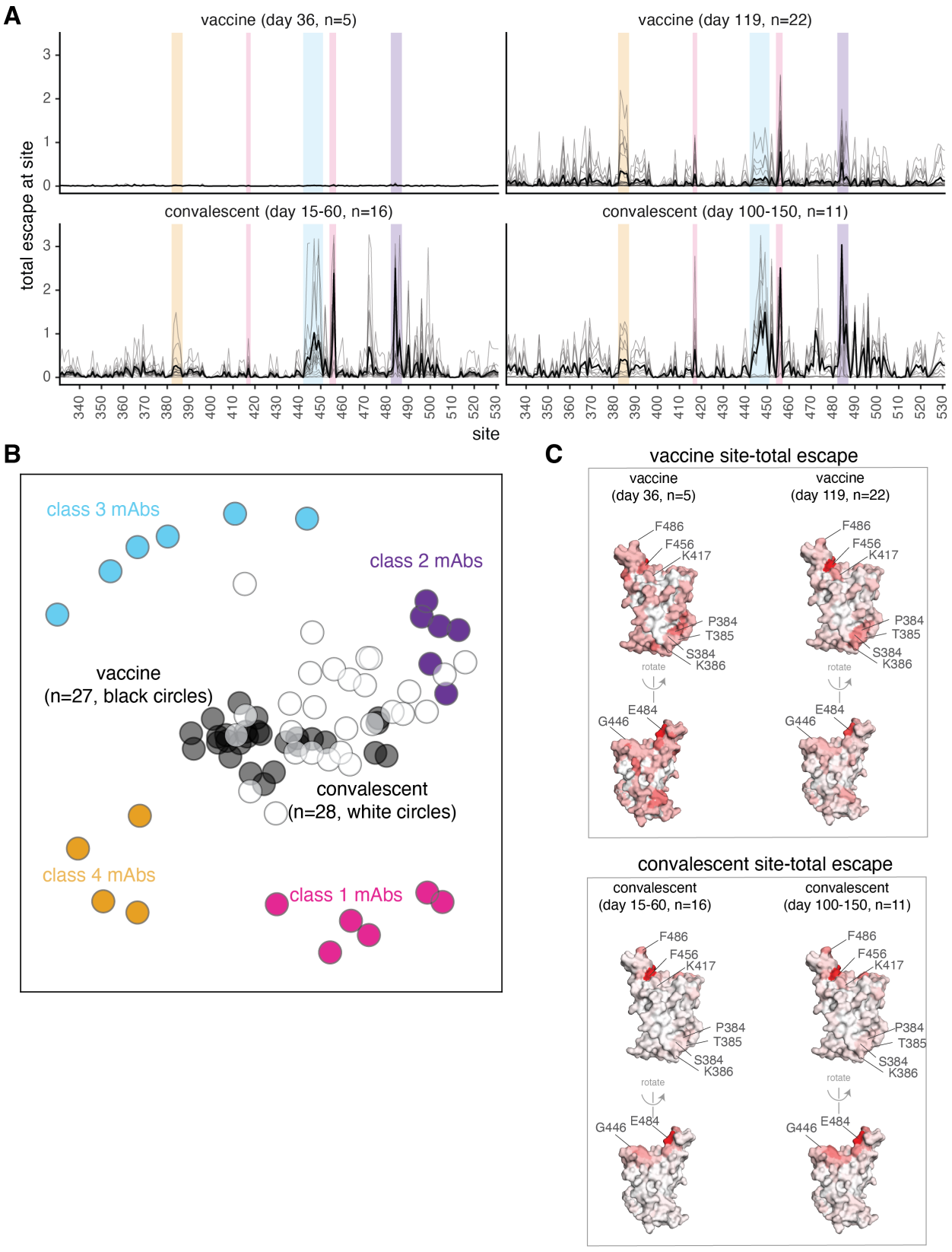
used multidimensional scaling to create a two-dimensional projection of the escape maps for the vaccine serum samples, convalescent plasma samples [120, 122], and previously characterized monoclonal antibodies [83, 124, 122, 279, 278] (Fig. 7.4B, an interactive version where you can mouse over points for details is at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Moderna/mds.html`). In this projection, monoclonal antibodies, sera samples, or plasma samples with similar binding-escape mutations are located close together, whereas those affected by distinct mutations are far apart. As previously reported [120], convalescent plasma samples clustered closest to class 2 antibodies (Fig. 7.4B), which are generally most affected by mutations to site 484. In contrast, the vaccine sera were more centrally located in the middle of the antibodies of all four classes, reflecting their flatter binding-escape maps that were less dominated by mutations that escape any single antibody class (Fig. 7.4B).

To examine sites of binding-escape mutations in the context of the RBD's structure, we projected the total escape at each site averaged across all vaccine or convalescent samples at each time point onto the surface of the RBD (Fig. 7.4C). The sites where mutations affected binding of vaccine sera were broadly distributed across the RBD surface (Fig. 7.4C), whereas convalescent plasma samples were most affected by mutations at just a few key regions (sites 456 and 484, and to a lesser degree the 443–450 loop) (Fig. 7.4C). However, as noted above, binding escape from the vaccine sera was somewhat more focused at day 119 relative to day 36, including at sites 456, 484, and 383–386.

7.3.5 Single RBD mutations have less impact on vaccine-elicited antibody neutralizing activity than infection-elicited antibody neutralizing activity

We tested key RBD mutations in spike-pseudotyped lentiviral neutralization assays against a subset of vaccine and convalescent sera. We used the binding-escape maps to choose six representative samples each from the day 100–150 vaccine and convalescent sera for which $> 90\%$ of the neutralizing activity was due to RBD-binding antibodies (Fig. 7.1, Fig.

Figure 7.4: **The binding of vaccine-elicited polyclonal antibodies is more broadly distributed across the RBD than for infection-elicited antibodies**. **(A)** Escape from RBD-binding antibodies at each site in the RBD was mapped for vaccine sera or convalescent plasma samples collected at early or late time points. Thin gray lines show individual serum or plasma samples, and the thick black line shows the mean (number of samples is indicated in the plot titles). Key sites within the epitopes of each major RBD antibody class are highlighted with the colors defined in Fig. 7.2A and in panel (B). **(B)** Relationships among escape maps of vaccine sera, convalescent plasma samples, and monoclonal antibodies visualized with a multidimensional scaling projection. Vaccine sera include both doses and time points. Convalescent plasma samples include all time points. **(C)** Total binding escape at each site mapped onto the RBD surface after averaging across all serum or plasma in each group. The RBD surface coloring is scaled from white to red, with white indicating no escape, and red indicating the site with the greatest escape. The color scaling spans the full range of white to red for each serum or plasma group, so a quantitative scale is not comparable across groups. Escape maps for monoclonal antibodies previously described in [83, 124, 122, 279, 278], and convalescent plasma samples in [120, 122]. An interactive version of panel (B) where you can mouse over points for details is at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Moderna/mds.html`.


A
vaccine (day 36, n=5)
vaccine (day 119, n=22)
convalescent (day 15-60, n=16)
convalescent (day 100-150, n=11)
total escape at site
site
B
class 3 mAbs
class 2 mAbs
vaccine
(n=27, black circles)
convalescent
(n=28, white circles)
class 4 mAbs
class 1 mAbs
C
vaccine site-total escape
vaccine
(day 36, n=5)
vaccine
(day 119, n=22)
F486
F456
K417
P384
T385
S384
K386
rotate
E484
G446
convalescent site-total escape
convalescent
(day 15-60, n=16)
convalescent
(day 100-150, n=11)

7.6)[120]. The escape maps for the vaccine and convalescent samples chosen for these assays are summarized in Fig. 7.5A and detailed in Fig. 7.2 and Fig. 7.11.

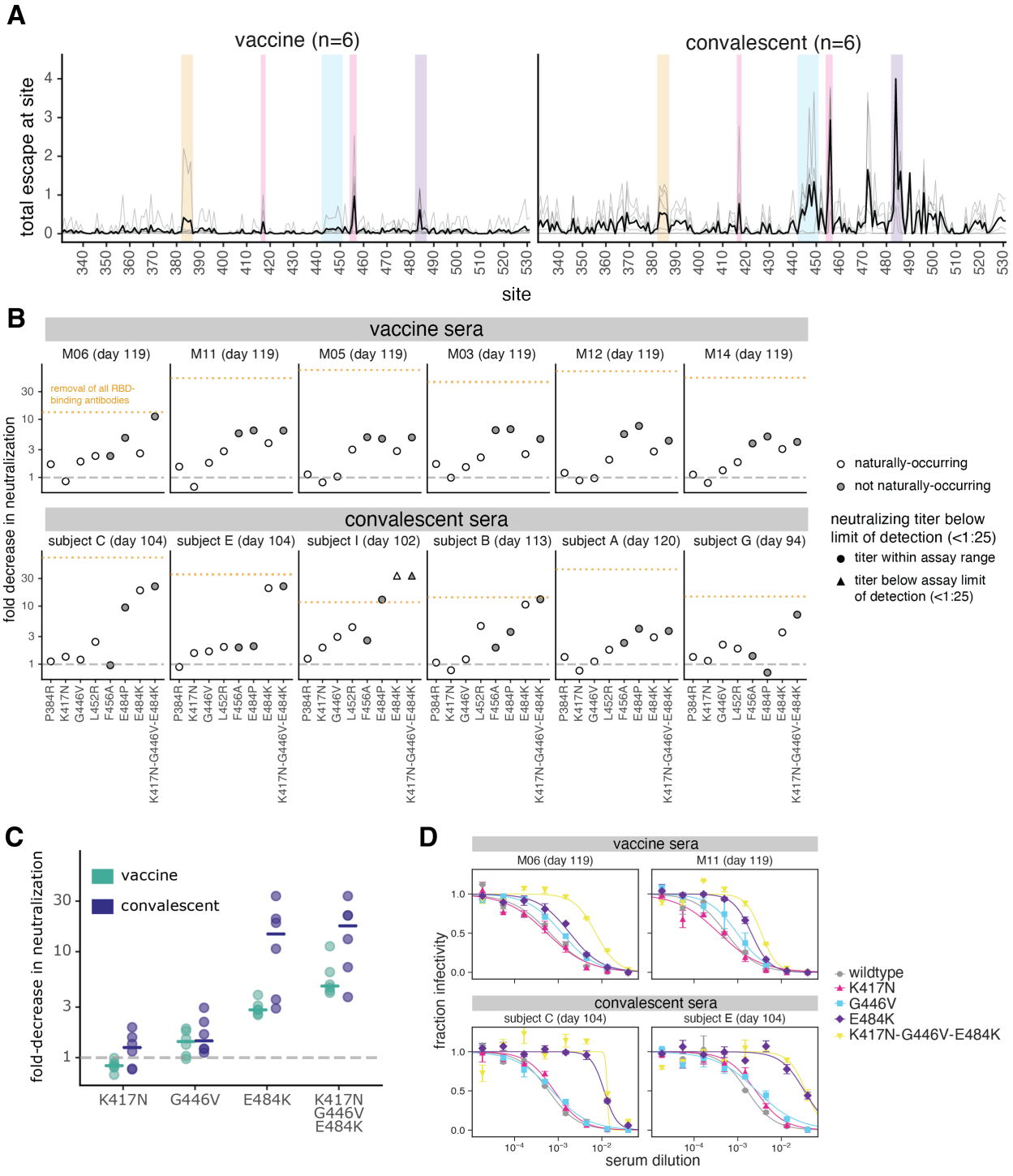
We performed neutralization assays on mutants in each of the four major RBD epitopes (class 1, K417N and F456A; class 2, E484P and E484K; class 3, G446V and L452R; class 4, P384R). Among these mutations, K417N, L452R, and E484K are present in emerging viral lineages, including B.1.351, P.1, B.1.427/429, B.1.526, and B.1.617 [14, 94, 285, 302, 323, 350] that have been shown to have reduced neutralization [42, 51, 199, 315, 313, 324, 232]. We also tested a triple mutant, K417N-G446V-E484K, with mutations in the class 1, 2, and 3 epitopes. For many convalescent sera, single RBD mutations reduced neutralization by approximately the same amount as removing all RBD-binding antibodies (Fig. 7.5B, Fig. 7.11 and 7.14, `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna/blob/main/experimental_data/results/mutant_neuts_results/frac_eroded.csv`). However, no single RBD mutation we tested had a comparably large effect on vaccine sera (Fig. 7.5B). This result is consistent with the binding-escape maps, which generally indicate that vaccine sera have a broader RBD-binding specificity than convalescent sera.

The mutations that most impacted neutralization also differed between vaccine and convalescent sera (Fig. 7.5B). For convalescent sera, the largest reduction in neutralization was consistently caused by mutations to site E484 in the class 2 epitope [120, 122], including the E484K mutation present in multiple emerging viral lineages [94, 285, 323]. In contrast, E484K generally caused a more moderate decrease in neutralization for vaccine sera. For some vaccine sera, another mutation at site E484 (E484P) caused a larger loss of neutralization, but E484P has not been found in any sequenced isolates of SARS-CoV-2 and has been shown to reduce both ACE2 binding affinity [280] and viral entry titers (Fig. 7.11D). The F456A mutation to the class 1 epitope often reduced neutralization by vaccine sera, although it had little effect on convalescent sera; this mutation is also not observed in natural sequences and reduces viral entry titers (Fig. 7.11D). Mutations to the class 3 epitope (G446V, L452R)

modestly reduced neutralization by some vaccine and convalescent sera (Fig. 7.5B). However, P384R in the less-neutralizing core RBD class 4 epitope [74, 188, 229, 356] and K417N in the class 1 epitope had little effect on neutralization by any sera, consistent with previous reports [51, 170, 313, 315]. Importantly, although single mutations sometimes caused large decreases in neutralization by convalescent sera, in no case did they reduce neutralization by vaccine sera $>$ 10-fold or to a titer $<$ 100 (Fig. 7.5B, Fig. 7.11).

The fact that single mutations ablated the anti-RBD neutralizing activity of some convalescent sera, but only modestly eroded the activity of vaccine sera, suggests that the vaccine elicits neutralizing antibodies with a greater number of RBD specificities. To test this idea, we performed neutralization assays with a triple mutant (K417N-G446V-E484K) containing a mutation in each of the class 1, 2, and 3 epitopes. For convalescent sera, the E484K mutation alone often caused a decrease in neutralization comparable to the triple mutant (Fig. 7.5C and D, Fig. 7.11), consistent with the convalescent escape maps showing a strong focus on site E484. In contrast, for vaccine sera, the triple mutant always reduced neutralization more than any of its constituent single mutations (Fig. 7.5C and D, Fig. 7.11). Moreover, the triple mutant decreased neutralization to the same extent as removing all RBD-binding antibodies for only one out of the six vaccine sera samples tested (Fig. 7.5B), indicating that the vaccine usually induces some neutralizing antibodies not escaped by mutations to sites K417, G446, and E484. These results are consistent with the escape maps indicating that the vaccine sera often have a broader RBD-binding specificity. Of note, infection also elicited very broad anti-RBD neutralizing activity in some cases; for instance, serum from the convalescent individual with the broadest escape map (participant G, day 94) was substantially more affected by the triple mutant than any of its constituent single mutants (Fig. 7.5B, Fig. 7.11 and S8).

Figure 7.5: **Effects of RBD mutations on neutralization by day 100–150 sera from vaccinated and convalescent individuals**. **(A)** Total binding escape at each RBD site is shown for the samples from vaccinated (n=6) or convalescent (n=6) individuals tested in neutralization assays. The thin gray lines show individual samples, and the dark black line shows the mean. Key sites within each epitope are highlighted using the same color scheme as in Fig. 7.2A. **(B)** Neutralization of G614 spike-pseudotyped lentiviral particles with the indicated RBD mutations, shown as the fold-decrease in NT50 compared to G614 spike with no additional mutations. Mutations that have been observed in human SARS-CoV-2 isolates are colored in white, and non-naturally-occurring mutations in gray. The orange dashed line represents the effect of depleting all RBD-binding antibodies. **(C)** The fold decrease in neutralization titer caused by individual mutations in each of the three major neutralizing epitopes of the RBD: K417 in the class 1 epitope, E484K in the class 2 epitope, and G446V in the class 3 epitope. The combination of all three mutations is also shown. Horizontal lines represent the median. In (B) and (C), the dashed gray line indicates no change in neutralization relative to unmutated spike. **(D)** Representative neutralization curves from two vaccine and two convalescent samples against the triple mutant and its composite single mutations.


A
vaccine (n=6)
convalescent (n=6)
total escape at site
site
B
vaccine sera
M06 (day 119)
M11 (day 119)
M05 (day 119)
M03 (day 119)
M12 (day 119)
M14 (day 119)
removal of all RBD-binding antibodies
convalescent sera
subject C (day 104)
subject E (day 104)
subject I (day 102)
subject B (day 113)
subject A (day 120)
subject G (day 94)
fold decrease in neutralization
P384R
K417N
G446V
L452R
F456A
E484P
E484K
K417N-G446V-E484K
naturally-occurring
not naturally-occurring
neutralizing titer below limit of detection (<1:25)
titer within assay range
titer below assay limit of detection (<1:25)
C
vaccine
convalescent
fold-decrease in neutralization
K417N
G446V
E484K
K417N G446V E484K
D
vaccine sera
M06 (day 119)
M11 (day 119)
convalescent sera
subject C (day 104)
subject E (day 104)
fraction infectivity
serum dilution
wildtype
K417N
G446V
E484K
K417N-G446V-E484K

7.4 Discussion

In this study, we have shown differences in the specificity of polyclonal serum antibodies acquired by infection versus vaccination with mRNA-1273. The neutralizing activity of vaccine sera is more focused on the RBD than for convalescent sera, with the majority of vaccine sera losing all detectable neutralization at a 1:25 cutoff after depletion of RBD-directed antibodies. This fact is surprising, since the mRNA-1273 vaccine encodes the full spike ectodomain [63], and one conjectured benefit of full-spike versus RBD-only vaccines was elicitation of neutralizing antibodies targeting non-RBD subdomains.

At first glance, the RBD focusing of the vaccine sera neutralization might seem likely to increase susceptibility to viral mutations, but the rest of our results suggest that this may not be the case. Our comprehensive maps of how RBD mutations reduce serum antibody binding show that vaccine-elicited antibodies are usually less affected by any single RBD mutation than infection-elicited antibodies. Whereas infection-elicited RBD antibodies are often strongly focused on an epitope including site E484, vaccine-elicited antibodies bind more broadly across the RBD, including to the more conserved "core" regions. This broader binding makes neutralization by vaccine sera more resistant to mutations within the RBD. For instance, RBD-directed neutralization by convalescent sera was greatly reduced or even eliminated by a combination of key mutations at the three major epitopes in the RBD's receptor-binding motif, but all vaccine sera that we tested retained substantial neutralization against this triple mutant. This result implies that either vaccination induces an antibody response more broadly distributed across the RBD surface, or that the individual antibodies elicited by vaccination are more robust to these mutations [106, 212]. Our results are consistent with a recent study by Amanat et al., which reported that several single RBD mutations reduce binding of serum from individuals vaccinated with the Pfizer mRNA vaccine less than for serum from previously infected individuals [8].

We found that the specificity of the mRNA-1273 vaccine-induced RBD-binding antibody

response often narrows over time. In contrast, the infection-elicited RBD-binding antibody response often broadens over time [106, 120]. However, because the early vaccine-induced RBD-binding antibody response is so broad compared to that induced by infection, even with these contrasting dynamics, the vaccine response remains broader than the convalescent response at late time points of 3–4 months. Additionally, the overall antibody response is more homogeneous for vaccinated than convalescent individuals. For instance, the RBD binding titers, neutralizing titers, amount of neutralization derived from RBD-binding antibodies, and effects of mutations on neutralization were more uniform for the vaccinated cohort than the convalescent cohort.

Our results do not explain why there are differences between the vaccine- and infection-elicited antibody responses, but we note two possibilities. First, the vaccine encodes a stabilized S-2P spike, which could present some epitopes in slightly different conformations and lead to less S1 shedding. Second, the vaccine is delivered in a two-dose schedule by an mRNA-lipid nanoparticle, which may lead to different kinetics of antigen presentation than viral infection [69, 226]. Indeed, another recent study suggests that mRNA vaccination elicits a different distribution of isotypes and fewer antibodies that cross-react to common-cold coronaviruses as compared to infection [254].

There are several limitations to our study. The vaccinated individuals in our study were relatively young (18–55 years) and healthy, whereas the convalescent individuals were older (23–76 years, median 56) with a range of comorbidities [66]. Additionally, we did not examine effects of mutations or deletions to the N-terminal domain of the spike protein, which can also affect neutralization by vaccine sera [315]. Our experiments assayed binding of antibodies to isolated RBD expressed by yeast, and so cannot capture mutational effects on trimer conformation or antibodies with quaternary epitopes [16]. Finally, the N-linked glycans on yeast-expressed proteins are more mannose-rich than those on mammalian-expressed proteins [127].

Despite these limitations, our results in conjunction with other recent studies [8] suggest that mRNA vaccines and infection elicit somewhat distinct anti-spike antibody responses. Therefore, it is important to differentiate antibody immunity acquired by different means when assessing the impact of viral evolution. Considerable effort is being expended to identify emerging antigenic variants of SARS-CoV-2 and determine which ones might evade immunity [42, 285, 315, 324]. Our findings suggest that the results could vary depending on the source of immunity. Furthermore, carefully characterizing and comparing the specificity of antibody immunity elicited by additional vaccine modalities could provide a basis for determining whether some vaccine responses will be more resistant to viral evolution.

7.5 Materials and Methods

Data and materials availability

All data associated with this study are in the paper or supplementary materials. The SARS-CoV-2 RBD mutant libraries (#1000000172) and unmutated parental plasmid (#166782) are available on Addgene. The plasmid encoding the SARS-CoV-2 spike gene used to generate pseudotyped lentiviral particles, HDM_Spikedelta21_D614G, is available from Addgene (#158762) and BEI Resources (NR-53765). Further information and requests for reagents and resources should be directed to and will be fulfilled by Jesse Bloom (jbloom@fredhutch.org) upon completion of a materials transfer agreement. The complete code for the full computational data analysis pipeline of the mapping experiments is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna`. The escape fraction measured for each mutation in `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna/blob/main/results/supp_data/moderna_convalescent_all_raw_data.csv`. All raw sequencing data are available on the NCBI Short Read Archive at BioProject PRJNA639956, BioSample SAMN18683769. The neutralization titers of vaccine- and infection-elicited sera against the tested RBD point mutants is at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_M`

oderna/blob/main/experimental_data/results/mutant_neuts_results/fitparams.csv.

Study design

De-identified post-vaccination sera were obtained as secondary research samples from the National Institutes of Allergy and Infectious Diseases-sponsored mRNA-1273 phase 1 clinical trial (NCT04283461) [147]. We obtained samples from 14 individuals who received two 250 μg doses of the mRNA-1273 vaccine, and 8 individuals who received two 100 μg doses. All individuals were between ages 18 and 55 years old. The study size was determined by the number of samples that were available from the phase 1 clinical trial, and not based on any power calculations. Experiments described in this manuscript were not performed blinded. The samples were collected under the human subject approvals described in [147]. Due to the de-identified nature of the samples, the work described in this paper was deemed non-human subjects research by the Fred Hutchinson Cancer Research Center Institutional Review Board.

Previously reported results from samples from two cohorts of SARS-CoV-2 convalescent individuals are reanalyzed here [120, 122]. One cohort of convalescent plasma samples were previously described [66, 120] and collected as part of a prospective longitudinal cohort study of individuals with SARS-CoV-2 infection in Seattle, WA between February and July 2020. The plasma samples from 17 individuals were examined here (8/17 female; age range 23–76 years, mean 51.6 years, median 56 years). All data from this cohort, including the neutralization and RBD- and spike-binding activity of plasma samples pre- and post-depletion of RBD-binding antibodies in Fig. 7.1 and RBD-binding escape maps in Fig. 7.4, Fig. 7.11B, and Fig. 7.11, were previously reported [120] with the exception of neutralization assays in Fig. 7.5, Fig. 7.11, and Fig. 7.14, which were newly performed in this study. This work was approved by the University of Washington Institutional Review Board.

All data from the second cohort of plasma samples (n=5), including the aggregated escape maps in Fig. 7.4, were previously reported [120] and are reanalyzed here. The plasma samples were originally collected 21–35 days post-symptom onset as part of a prospective longitudinal cohort study of SARS-CoV-2 convalescent individuals in New York, NY, under the human subject approvals described in [251].

RBD deep mutational scanning library

The yeast-display RBD mutant libraries were previously described [280, 124]. Briefly, duplicate mutant libraries were constructed in the spike receptor binding domain (RBD) from SARS-CoV-2 (isolate Wuhan-Hu-1, Genbank accession number MN908947, residues N331–T531) and contain 3,804 of the 3,819 possible amino-acid mutations, with $> 95\%$ present as single mutants. Each RBD variant was linked to a unique 16-nucleotide barcode sequence to facilitate downstream sequencing. As previously described, libraries were sorted for RBD expression and ACE2 binding to eliminate RBD variants that are completely misfolded or non-functional, such as those lacking modest ACE2 binding affinity [124].

FACS sorting of yeast libraries to select mutants with reduced binding by polyclonal post-vaccination sera

Serum mapping experiments were performed in biological duplicate using the independent mutant RBD libraries, similarly to as previously described for monoclonal antibodies [124] and exactly as previously described for polyclonal plasma samples [120]. Briefly, mutant yeast libraries induced to express RBD were washed and incubated with serum at a range of dilutions for 1 hour at room temperature with gentle agitation. For each serum, we chose a sub-saturating dilution such that the amount of fluorescent signal due to serum antibody binding to RBD was approximately equal across samples. The exact dilution used for each serum is given in `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna/blob/`

`main/data/TableS2_FACSinfo.csv`. After the serum incubations, the libraries were secondarily labeled for 1 hour with 1:100 fluorescein isothiocyanate-conjugated anti-MYC antibody (Immunology Consultants Lab, CYMC-45F) to label for RBD expression and 1:200 Alexa Fluor-647-conjugated goat anti-human-IgA+IgG+IgM (Jackson ImmunoResearch 109-605-064) to label for bound serum antibodies. A flow cytometric selection gate was drawn to capture 3–6% of the RBD mutants with the lowest amount of serum binding for their degree of RBD expression (Fig. 7.7 and 7.8). We also measured what fraction of cells expressing unmutated RBD fell into this gate when stained with 1x and 0.1x the concentration of serum. For each sample, approximately 10 million RBD+ cells (range 7.3e6 to 1.4e7 cells) were processed on the BD FACSAria II cell sorter, with between 3e5 and 6e5 plasma-escaped cells collected per sample (`https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna/blob/main/data/TableS2_FACSinfo.csv`). Antibody-escaped cells were grown overnight in synthetic defined medium with casamino acids (6.7g/L Yeast Nitrogen Base, 5.0g/L Casamino acids, 1.065 g/L MES acid, and 2% w/v dextrose) to expand cells prior to plasmid extraction.

DNA extraction and Illumina sequencing

Plasmid samples were prepared from 30 optical density (OD) units (1.6e8 colony forming units (cfus)) of pre-selection yeast populations and approximately 5 OD units (3.2e7 cfus) of overnight cultures of serum-escaped cells (Zymoprep Yeast Plasmid Miniprep II) as previously described [124]. The 16-nucleotide barcode sequences identifying each RBD variant were amplified by polymerase chain reaction (PCR) and prepared for Illumina sequencing as described in [280]. Barcodes were sequenced on an Illumina HiSeq 2500 with 50 bp single-end reads. To minimize noise from inadequate sequencing coverage, we ensured that each antibody-escape sample had at least 2.5x as many post-filtering sequencing counts as FACS-selected cells, and reference populations had at least 2.5e7 post-filtering sequencing

counts.

Analysis of deep sequencing data to compute each mutation's escape fraction

Escape fractions were computed as described in [124], with minor modifications as noted below. We used the dms_variants package (`https://jbloomlab.github.io/dms_variants/`, version 0.8.5) to process Illumina sequences into counts of each barcoded RBD variant in each pre-sort and antibody-escape population using the barcode/RBD look-up table from [280]. For each serum selection, we computed the "escape fraction" for each barcoded variant using the deep sequencing counts for each variant in the original and serum-escape populations and the total fraction of the library that escaped antibody binding via the formula provided in [124]. These escape fractions represent the estimated fraction of cells expressing that specific variant that falls in the escape bin, such that a value of 0 means the variant is always bound by serum and a value of 1 means that it always escapes serum binding. We then applied a computational filter to remove variants with low sequencing counts or highly deleterious mutations that might cause antibody escape simply by leading to poor expression of properly folded RBD on the yeast cell surface [124, 280]. Specifically, we removed variants that had (or contained mutations with) ACE2 binding scores < -2.35 or expression scores < -1, using the variant- and mutation-level deep mutational scanning scores from [280]. Note that these filtering criteria are slightly more stringent than those used in [124] but are identical to those used in [278, 279, 122, 120].

We next deconvolved variant-level escape scores into escape fraction estimates for single mutations using global epistasis models [220] implemented in the dms_variants package, as detailed at (`https://jbloomlab.github.io/dms_variants/dms_variants.globalepistasis.html`) and described in [124]. The reported scores throughout the paper are the average across the libraries; these scores are also in `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna/blob/main/results/supp_data/moderna_convalescent_all_raw_dat`

`a.csv`. Correlations in final single-mutant escape scores are shown in Fig. 7.9.

For plotting and analyses that required identifying RBD sites of "strong escape", we considered a site to mediate strong escape if the total escape (sum of mutation-level escape fractions) for that site exceeded the median across sites by $>$ 5-fold, and was at least 5% of the maximum for any site. Full documentation of the computational analysis is at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna`.

Generation of pseudotyped lentiviral particles

HEK-293T (American Type Culture Collection, CRL-3216) cells were used to generate SARS-CoV-2 spike-pseudotyped lentiviral particles and 293T-ACE2 cells (Biodefense and Emerging Infectious Research Resources Repository (BEI Resources), NR-52511) were used to titer the SARS-CoV-2 spike-pseudotyped lentiviral particles and to perform neutralization assays (see below). We used spike-pseudotyped lentiviral particles that were generated essentially as described in [67], using a codon-optimized SARS-CoV-2 spike from Wuhan-Hu-1 strain that contains a 21-amino-acid deletion at the end of the cytoplasmic tail [66] and the D614G mutation that is now predominant in human SARS-CoV-2 [161]. The plasmid encoding this spike, HDM_Spikedelta21_D614G, is available from Addgene (#158762) and BEI Resources (NR-53765), and the full sequence is at (`https://www.addgene.org/158762`). Point mutations were introduced into the RBD of this plasmid via site-directed mutagenesis. Therefore, all mutations tested in this paper are in the G614 background, and are compared to a "wild-type" spike with G614.

To generate these spike-pseudotyped lentiviral particles [67], 6e5 HEK-293T (ATCC CRL-3216) cells per well were seeded in 6-well plates in 2 mL D10 growth media (Dulbecco's Modified Eagle Medium with 10% heat-inactivated fetal bovine serum, 2 mM l-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin). 24 hours later, cells were transfected using BioT transfection reagent (Bioland Scientific) with a Luciferase_IRES_ZsGreen back-

bone, Gag/Pol lentiviral helper plasmid (BEI Resources NR-52517), and wild-type or mutant SARS-CoV-2 spike plasmids. Media was changed to fresh D10 at 24 hours post-transfection. At 60 hours post-transfection, viral supernatants were collected, filtered through a 0.45 μm surfactant-free cellulose acetate low protein-binding filter, and stored at -80°C.

Titering of pseudotyped lentiviral particles

Titers of spike-pseudotyped lentiviral particles were determined as described in [67] with the following modifications. 100 μL of diluted spike-pseudotyped lentiviral particles was added to 1.25e4 293T-ACE2 cells (BEI Resources NR-52511), grown overnight in 50 μL of D10 growth media in a 96-well black-walled poly-L-lysine coated plate (Greiner Bio-One, 655936). Relative luciferase units (RLU) were measured 65 hours post-infection (Promega Bright-Glo, E2620) in the infection plates with a black back-sticker (Thermo Fisher Scientific, NC9425162) added to minimize background. Titers were first estimated from the average of 8 two-fold serial dilutions of virus starting at 25 μL virus in a total volume of 150 μL, performed in duplicate, and normalized to a wild-type D614G variant harvested on the same day. Quantitative titering was then performed at a single virus dilution, targeting 200,000 RLU per well. Values in Fig. 7.11D are an average RLUs per μL measured across 16 technical replicates at a single dilution.

Neutralization assays

293T-ACE2 cells (BEI Resources NR-52511) were seeded at 1.25e4 cells per well in 50 μL D10 in poly-L-lysine coated, black-walled, 96-well plates (Greiner 655930). 24 hours later, pseudotyped lentivirus supernatants were diluted to approximately 200,000 RLU per well (determined by titering as described above and incubated with a range of dilutions of serum for 1 hour at 37 °C. 100 μL of the virus-antibody mixture was then added to cells. At about 50 or 70 hours post-infection, luciferase activity was measured using the Bright-Glo

Luciferase Assay System (Promega, E2610). Fraction infectivity of each serum antibody-containing well was calculated relative to a "no-serum" well inoculated with the same initial viral supernatant (containing wild-type or mutant RBD) in the same row of the plate. We used the neutcurve package (`https://jbloomlab.github.io/neutcurve` version 0.5.2) to calculate the inhibitory concentration 50% (IC50) and the neutralization titer 50% (NT50), which is 1/IC50, of each serum against each virus by fitting a Hill curve with the bottom fixed at 0 and the top fixed at 1.

Depletion of RBD-binding antibodies from polyclonal sera

Two rounds of sequential depletion of RBD-binding antibodies were performed for vaccine-elicited sera. Magnetic beads conjugated to the SARS-CoV-2 RBD (AcroBiosystems, MBS-K002) were prepared according to the manufacturer's protocol. Beads were resuspended in ultrapure water at 1 mg beads/mL and a magnet was used to wash the beads 3 times in phosphate-buffered saline (PBS) with 0.05% bovine serum albumin (BSA). Beads were then resuspended in PBS with 0.05% BSA at 1 mg beads per mL. Beads (manufacturer-reported binding capacity of 10–40 μg/mL anti-RBD antibodies) were incubated with human sera at a 3:1 ratio beads:serum (150 μL beads + 50 μL serum), rotating overnight at 4°C. A magnet (MagnaRack Magnetic Separation Rack, Thermo Fisher Scientific, CS15000) was used to separate antibodies that bind RBD from the supernatant, and the supernatant (the post-RBD antibody depletion sample) was removed. A mock depletion (pre-depletion sample) was performed by adding 150 μL of PBS + 0.05% BSA and incubating rotating overnight at 4°C. A second round of depletion was then performed to ensure full depletion of RBD-binding antibodies. For the neutralization assays on these sera depleted of RBD-binding antibodies, the reported serum dilution is corrected for the dilution incurred by the depletion process.

Measurement of serum binding to RBD or spike by enzyme-linked immunosorbent assay (ELISA)

The IgG ELISAs for spike protein and RBD were conducted as previously described [82]. Briefly, ELISA plates were coated with recombinant spike and RBD antigens described in [82] at 2 μg/mL. Five 3-fold serial dilutions of sera beginning at 1:500 were performed in PBS with 0.1% Tween with 1% Carnation nonfat dry milk. Dilution series of the "synthetic" sera comprised of the anti-RBD antibody REGN10987 [129] or anti-N-terminal domain antibody 4A8 [56] and pooled pre-pandemic human serum from 2017–2018 (Gemini Biosciences; nos. 100–110, lot H86W03J; pooled from 75 donors) were performed such that the anti-spike antibody was present at a highest concentration of 0.25 μg/mL. Both antibodies were recombinantly produced by Genscript. The REGN10987 is that used in [278] and the variable domain heavy and light chain sequences for 4A8 were obtained from Genbank GI 1864383732 and 1864383733 [56] and produced on a human IgG1 and IgK background, respectively. Pre-pandemic serum alone, without anti-RBD antibody depletion, was used as a negative control, averaged over 2 replicates. Secondary labeling was performed with goat anti-human IgG-Fc horseradish peroxidase (HRP) (1:3000, Bethyl Labs, A80-104P). Antibody binding was detected with TMB/E HRP substrate (Millipore Sigma, ES001) and 1 N HCl was used to stop the reaction. OD450 was read on a Tecan infinite M1000Pro plate reader. The area under the curve (AUC) was calculated using the scikit-learn python package, version 0.23.2 (`https://scikit-learn.org/stable/`) as the area under the titration curve with the serial dilutions on a log-scale.

Data visualization

The static logo plot visualizations of the escape maps in the paper figures were created using the dmslogo package (`https://jbloomlab.github.io/dmslogo`, version 0.6.2) and in all cases the height of each letter indicates the escape fraction for that amino-

acid mutation calculated as described above. For each sample, the y-axis is scaled to be the greatest of (a) the maximum site-wise escape metric observed for that sample, (b) 20x the median site-wise escape fraction observed across all sites for that serum, or (c) an absolute value of 1.0 (to appropriately scale samples that are not "noisy" but for which no mutation has a strong effect on antibody binding). Sites K417, L452, S477, E484, and N501 have been added to logo plots due to their frequencies among circulating viruses. The code that generates these logo plot visualizations is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_MAP_Moderna/blob/main/results/summary/escape_profiles.md`. In many of the visualizations, the RBD sites are categorized by epitope region [16] and colored accordingly. We define the class 1 epitope as residues 403+405+406+417+420+421+453+455–460+473–476+486+487+489+504, the class 2 epitope as residues 472+483–485+490–494, the class 3 epitope to be residues 345+346+437-452+496+498–501, and the class 4 epitope as residues 365–372+382–386.

For the static structural visualizations in the paper figures, the RBD surface (PDB 6M0J, [172]) was colored by the site-wise escape metric at each site, with white indicating no escape and red scaled to be the same maximum used to scale the y-axis in the logo plot escape maps, determined as described above. We created interactive structure-based visualizations of the escape maps using dms-view [137] that are available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_MAP_Moderna/`. The logo plots in these escape maps can be colored according to the deep mutational scanning measurements of how mutations affect ACE2 binding or RBD expression as described above.

For the composite line plots shown in Fig. 7.4, the convalescent (day 15–60) group includes two independent cohorts of individuals, one recruited in New York, NY (n=5) [251], and another recruited in Seattle, WA (n=11) [66]. The convalescent (day 100–150) group is from the longitudinal cohort recruited in Seattle, WA (n=11). The escape maps for convalescent individuals were previously reported in [120, 122]. The mRNA-1273 (day 119)

group includes individuals who were vaccinated with either the 100 or 250 μg vaccine dose (n=8 and n=14, respectively). The y-axis maximum is scaled to 1.1 times the maximum group mean site-total escape among all groups, so outlier points exceeding this value are not shown.

Statistical Analysis

The percent of neutralizing activity of vaccine-elicited sera and convalescent plasma due to RBD-binding antibodies is plotted with the plotnine python package, version 0.7.1 (`https://plotnine.readthedocs.io/en/stable/index.html`), shown as a Tukey boxplot (middle line indicating median, box limits indicating interquartile range) with individual measurements overlaid as points. P-values are from a log-rank test accounting for censoring, calculated with the lifelines python package, version 0.25.10 (`https://lifelines.readthedocs.io/en/latest/`).

7.6 Notes

Acknowledgments

We thank Cathy Lin for administrative support; Dolores Covarrubias, Andy Marty, and the Genomics and Flow Cytometry core facilities at the Fred Hutchinson Cancer Research Center for experimental support. We thank Lisa Jackson (Kaiser Permanente), Chris Roberts, Catherine Luke, and Rebecca Lampley (National Institute of Allergy and Infectious Diseases, National Institutes of Health) for assistance with obtaining the mRNA-1273 phase 1 trial vaccine samples. We also thank all research participants and study staff of the Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) study.

Funding

This study was funded by the National Institutes of Allergy and Infectious Diseases, National Institutes of Health (R01AI141707 and R01AI127893 to JDB, T32AI083203 to AJG, and F30AI149928 to KDHC); the Gates Foundation (INV-004949 to JDB and INV-016575 to HYC); an Emergent Ventures Award (to HYC); the National Institutes of Health Office of Research Infrastructure Programs S10OD028685 (to the Scientific Computing Infrastructure at the Fred Hutchinson Cancer Research Center); and the Howard Hughes Medical Institute (to JDB). TNS is a Howard Hughes Medical Institute Fellow of the Damon Runyon Cancer Research Foundation (DRG-2381-19). The mRNA-1273 phase 1 study was sponsored and primarily funded by the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, MD. This trial has been funded in part with federal funds from the NIAID under grant awards UM1AI148373, to Kaiser Washington; UM1AI148576, UM1AI148684, and NIH P51 OD011132, to Emory University; NIH AID AI149644, and contract award HHSN272201500002C, to Emmes. Funding for the manufacture of mRNA-1273 phase 1 material was provided by the Coalition for Epidemic Preparedness Innovation.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the US government or the other sponsors.

Author contributions

AJG and JDB conceptualized and designed the study. AJG and TNS developed the yeast-display deep mutational scanning antibody-escape mapping method. KDHC developed the spike-pseudotyped lentiviral system and neutralization assays. AJG performed the serum-escape mapping. KDM assisted with serum-escape mapping; ANL cloned spike point mutants and generated pseudotyped lentiviral particles. ANL, AJG, and LEG performed pseudotyped neutralization assays. AJG and JDB wrote the code and performed the formal analysis. HYC

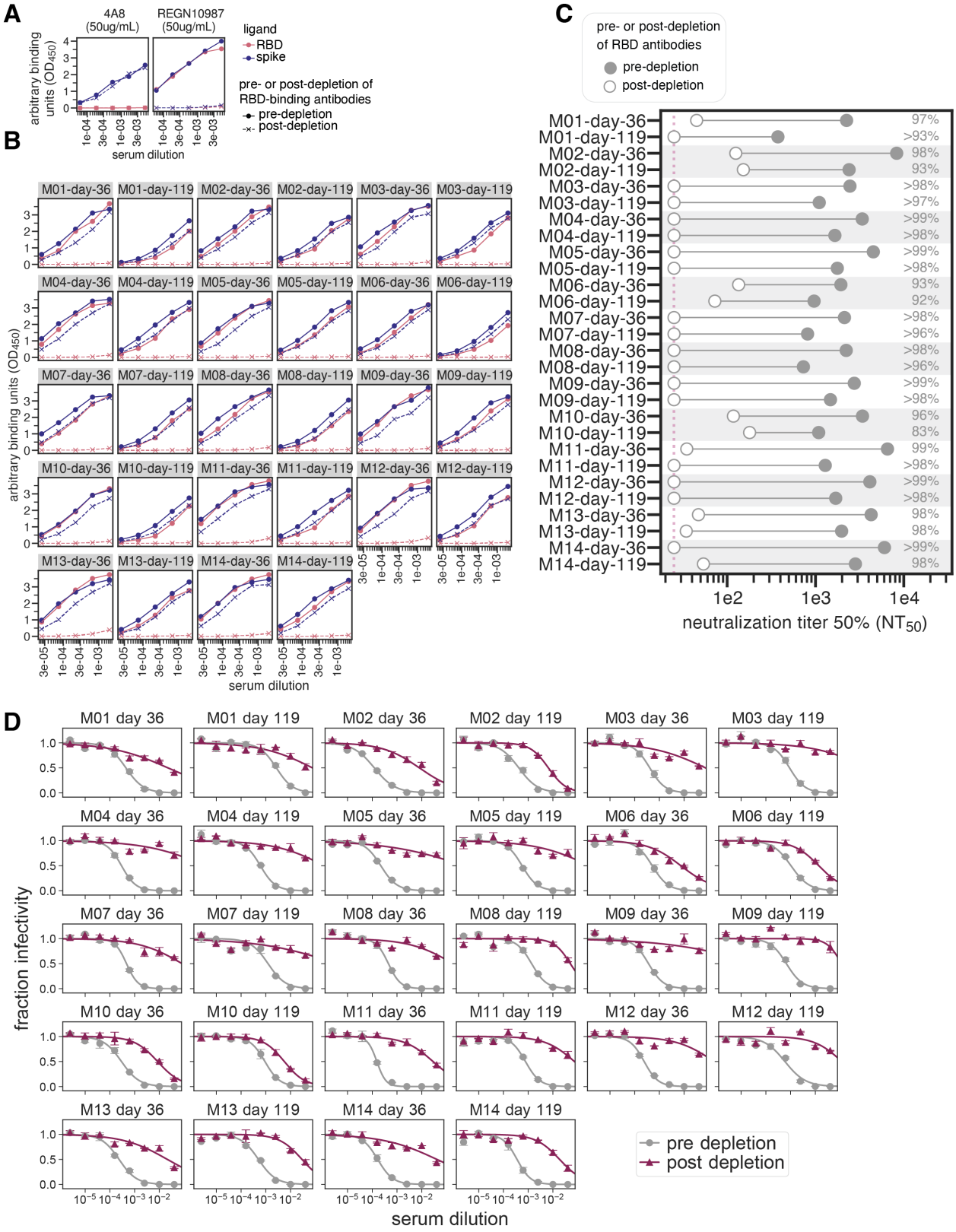
provided the convalescent plasma and serum samples. AJG and JDB wrote the original draft and all authors reviewed and edited the manuscript.

Declaration of interests

Subsequent to completion and submission of the initial version of this study, JDB began consulting for Moderna on viral evolution and epidemiology. JDB and KHD have the potential to receive a share of IP revenue as an inventor on a Fred Hutchinson Cancer Research Center-optioned technology/patent (application WO2020006494) related to deep mutational scanning of viral proteins. HYC is a consultant for Merck, Pfizer, Ellume, and the Bill and Melinda Gates Foundation and has received support from Cepheid and Sanofi-Pasteur. The other authors declare no competing interests.

7.7 Supplementary Figures

Figure 7.6: **Raw enzyme-linked immunosorbent assay (ELISA) and neutralization curves of mRNA-1273 serum samples before and after depletion of receptor binding domain (RBD)-binding antibodies**. **(A)** Effect of RBD antibody depletion on binding to RBD and spike by "synthetic sera" comprised of pre-pandemic pooled serum with the N-terminal domain (NTD)-targeting antibody 4A8 [56] or RBD-targeting antibody REGN10987 [129]. Antibodies were added to pre-pandemic serum at 50 μg/mL. The x-axis indicates the dilution factor of the serum+antibody mix, and the y-axis is the optical density at wavelength 450 (OD450) reading at each dilution. These controls were previously used in [120], and demonstrate that the depletions effectively remove RBD-targeting antibodies but not antibodies targeting other epitopes such as the NTD. **(B)** Raw ELISA binding curves of sera to RBD and spike before and after depletion of RBD-binding antibodies. **(C)** Neutralization titer 50% (NT50) of vaccine-elicited sera pre- and post-depletion of RBD-binding antibodies, shown in filled and open circles, respectively. All neutralization assays were performed with SARS-CoV-2 spike D614G-pseudotyped lentiviral particles. Two time points were assessed per individual, at day 36 and day 119 post-dose 1 of vaccination. The limit of detection is shown as a dashed pink vertical line. The percent neutralization due to RBD-binding antibodies are shown at right. **(D)**) Raw neutralization curves for sera before (gray circles) and after (maroon triangles) depletion of RBD-binding antibodies. Each assay was performed in technical duplicate, and points show the mean and standard error of the replicates.


A
4A8 (50ug/mL)
REGN10987 (50ug/mL)
arbitrary binding units (OD450)
serum dilution
ligand
RBD
spike
pre- or post-depletion of RBD-binding antibodies
pre-depletion
post-depletion
B
arbitrary binding units (OD450)
serum dilution
C
pre- or post-depletion of RBD antibodies
pre-depletion
post-depletion
neutralization titer 50% (NT50)
D
fraction infectivity
serum dilution
pre depletion
post depletion

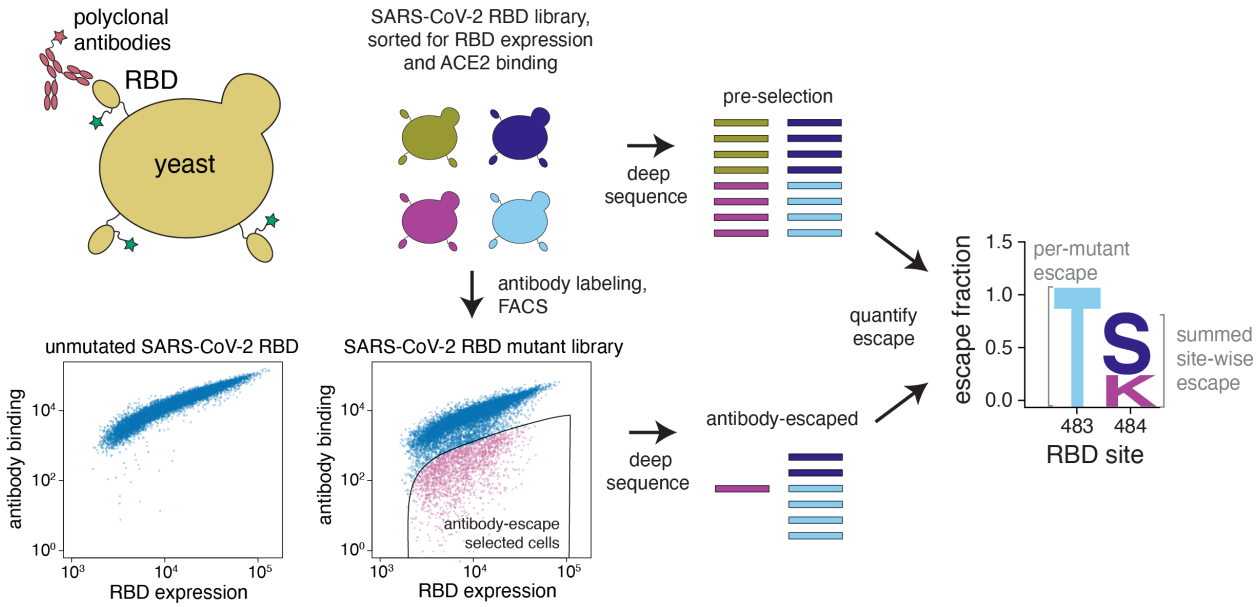
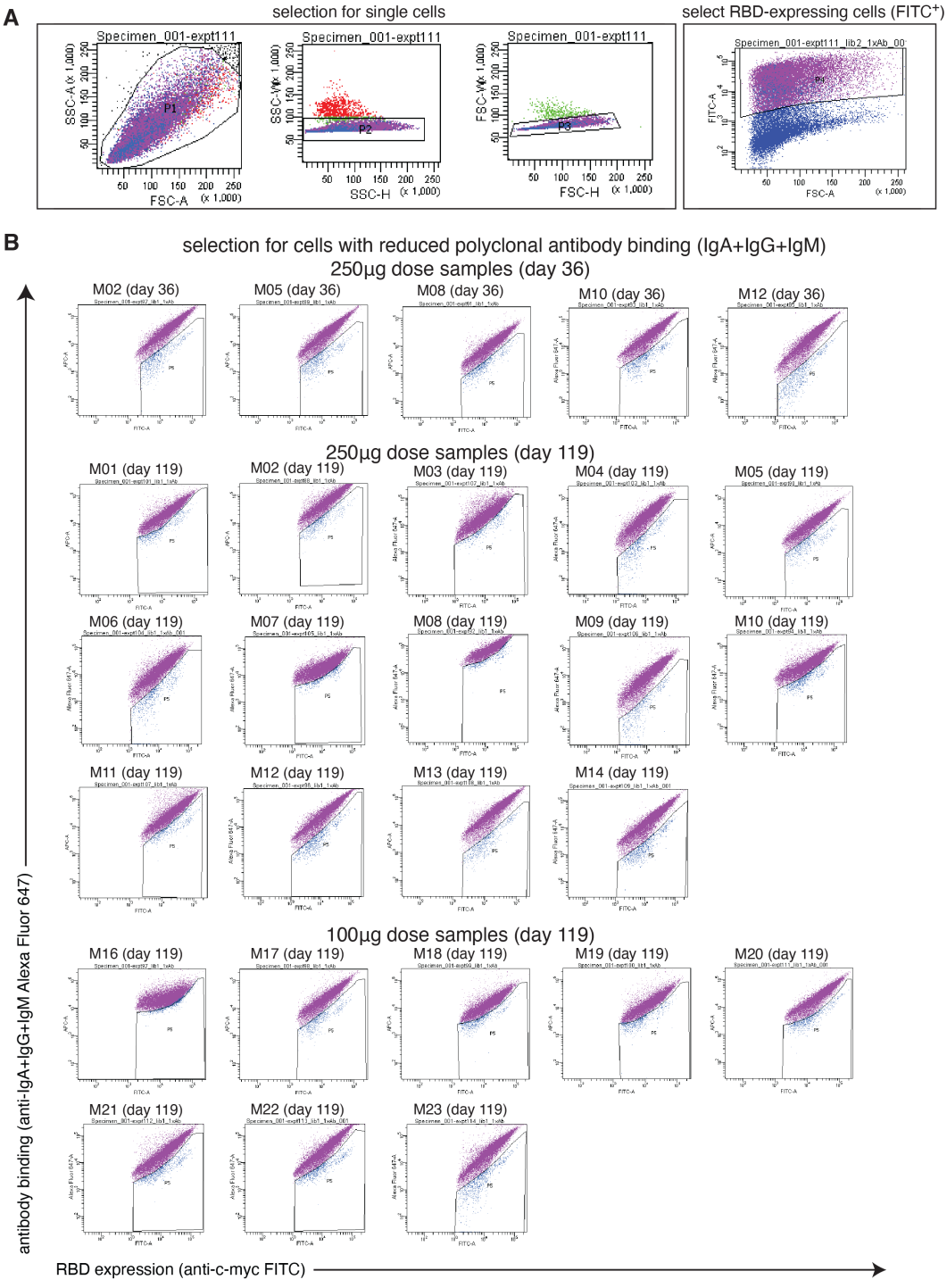


Figure 7.7: **Schematic of the deep mutational scanning approach used to quantify the effects of RBD mutations on antibody escape**. The RBD is expressed on the surface of yeast (top left). Flow cytometry is used to quantify both RBD expression (via a C-terminal MYC tag, green star) and antibody binding to the RBD protein expressed on the surface of each yeast cell (bottom left). A library of yeast expressing RBD mutants was incubated with polyclonal serum and fluorescence-activated cell sorting (FACS) was used to enrich for cells expressing RBD that bound reduced amounts of serum antibodies, as detected using an IgA+IgG+IgM secondary antibody. Deep sequencing was used to quantify the frequency of each mutation in the initial and "antibody escape" cell populations. We quantified the effect of each mutation as the "escape fraction," which represents the fraction of cells expressing RBD with that mutation that fell in the "antibody escape" FACS bin. Escape fractions are represented in logo plots, with the height of each letter proportional to the effect of that amino-acid mutation on antibody binding. The site-level escape metric is the sum of the escape fractions of all mutations at a site. Experimental and computational filtering was used to remove RBD mutants that were misfolded or unable to bind the angiotensin-converting enzyme (ACE2) receptor.

Figure 7.8: **FACS gating strategy to select yeast cells that express RBD mutants with reduced binding by serum antibodies**. **(A)** Representative plots of nested FACS gating strategy used for all serum selection experiments to select for single cells. Samples were gated by SSC-A versus FSC-A, SSC-W versus SSC-H, and FSC-W versus FSC-H) that also express RBD (FITC-A vs. FSC-A). **(B)** FACS gating strategy for one of two independent libraries to select cells expressing RBD mutants with reduced binding by polyclonal sera (cells in blue). Gates were set manually during sorting. Selection gates were set to capture approximately 5% of the RBD+ library. The same gate was set for both independent libraries stained with each serum sample, and the FACS scatter plots looked qualitatively similar between the two libraries. For information on the fraction of library cells that fall into each selection gate, see table S2. SSC-A, side scatter-area; FSC-A, forward scatter-area; SSC-W, side scatter-width; SSC-H, side scatter-height; FSC-W, forward scatter-width; FSC-H, forward scatter height; FITC-A, fluorescein isothiocyanate-area.


A
selection for single cells
select RBD-expressing cells (FITC+)
Specimen_001-expt111
SSC-A (x 1,000)
FSC-A (x 1,000)
SSC-H (x 1,000)
FSC-H (x 1,000)
P1
P2
P3
Specimen_001-expt111_lib2_1xAb_00
FITC-A
B
selection for cells with reduced polyclonal antibody binding (IgA+IgG+IgM)
250µg dose samples (day 36)
M02 (day 36)
M05 (day 36)
M08 (day 36)
M10 (day 36)
M12 (day 36)
250µg dose samples (day 119)
M01 (day 119)
M02 (day 119)
M03 (day 119)
M04 (day 119)
M05 (day 119)
M06 (day 119)
M07 (day 119)
M08 (day 119)
M09 (day 119)
M10 (day 119)
M11 (day 119)
M12 (day 119)
M13 (day 119)
M14 (day 119)
100µg dose samples (day 119)
M16 (day 119)
M17 (day 119)
M18 (day 119)
M19 (day 119)
M20 (day 119)
M21 (day 119)
M22 (day 119)
M23 (day 119)
antibody binding (anti-IgA+IgG+IgM Alexa Fluor 647)
RBD expression (anti-c-myc FITC)

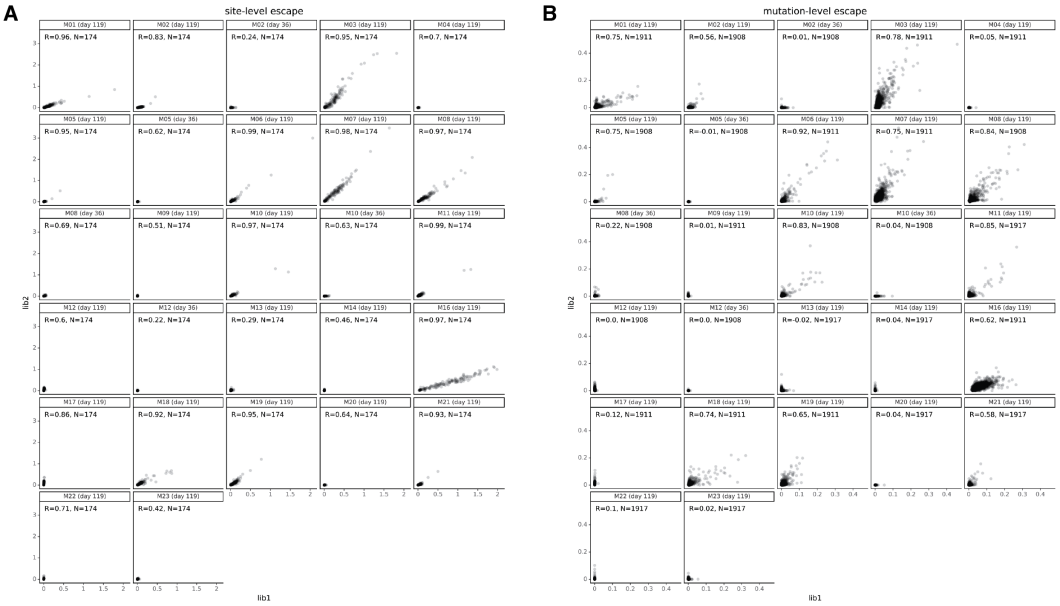


Figure 7.9: **Site- and mutation-level correlations between serum-escape measurements for each replicate library**. Correlation plots of site- or mutation-level escape for each of the two independent RBD mutant libraries for each serum sample are shown. Each point represents one site in the RBD in **(A)** or a different mutation in **(B)**.

Figure 7.10: **Complete binding-escape maps for the day 119 sera from all 14 individuals who received the 250 μg mRNA-1273 vaccine dose**. A subset of these sera are also shown in Fig. 7.2C. RBD sites are colored according to epitope, as in Fig. 7.2A.

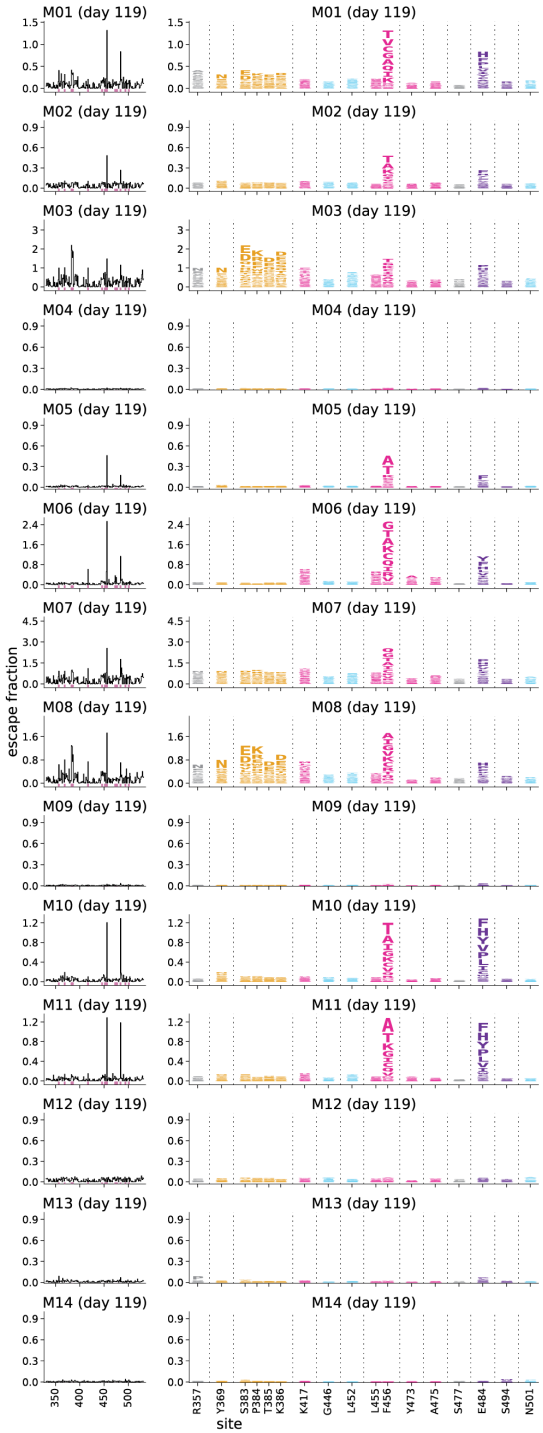
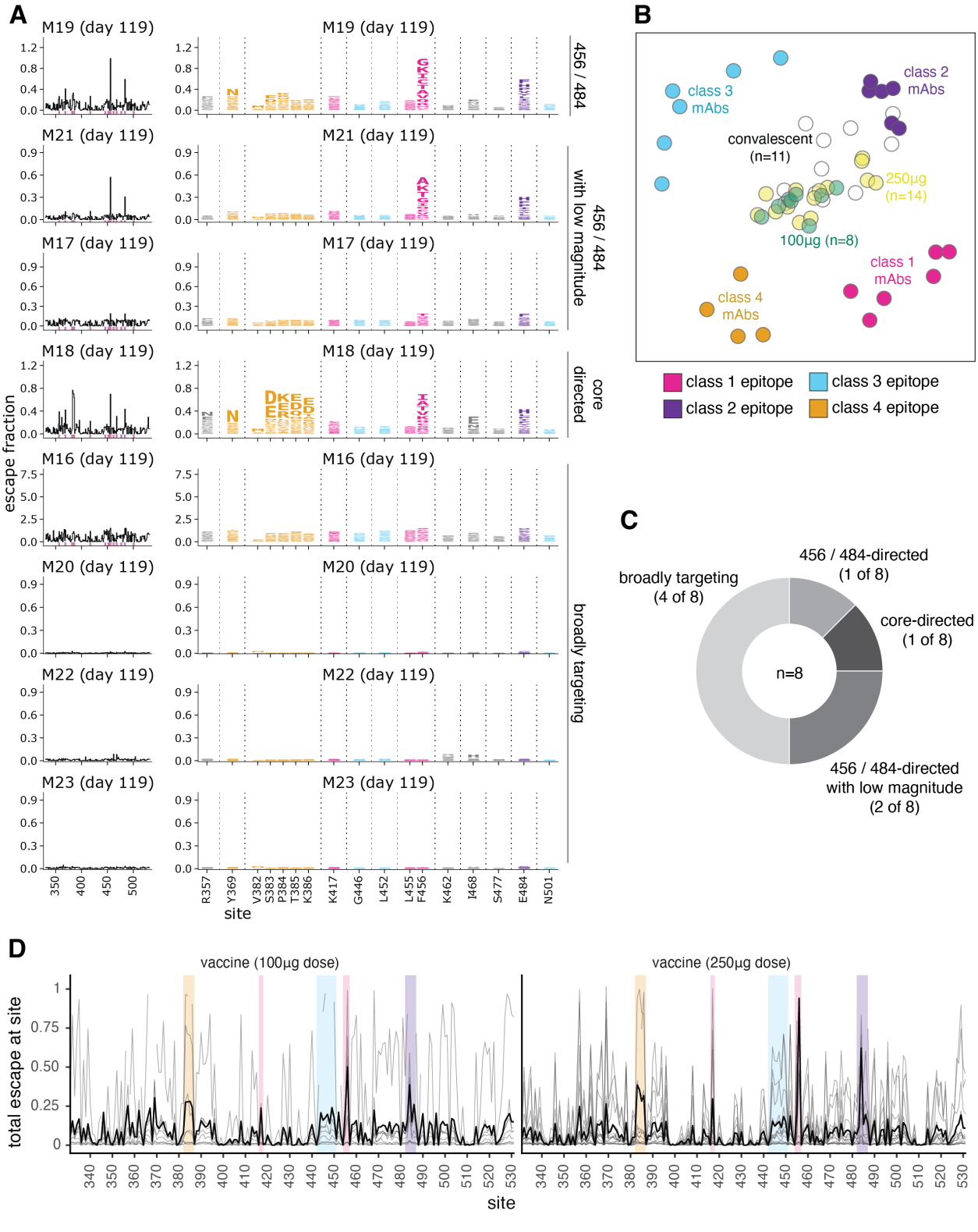

M01 (day 119)
M02 (day 119)
M03 (day 119)
M04 (day 119)
M05 (day 119)
M06 (day 119)
M07 (day 119)
M08 (day 119)
M09 (day 119)
M10 (day 119)
M11 (day 119)
M12 (day 119)
M13 (day 119)
M14 (day 119)
escape fraction
350
400
450
500
R357
Y369
S383
P384
T385
K386
K417
G446
L452
L455
F456
Y473
A475
S477
E484
S494
N501
site

Figure 7.11: **Escape maps from individuals who received the 100 μg dose of mRNA-1273 at 119 days post-vaccination largely resemble those of individuals who received the 250 μg dose**. **(A)** Logo plots representing the complete escape maps for the day 119 sera from 8 individuals who received the 100 μg dose of mRNA-1273. RBD sites are colored according to epitope, as in Fig. 7.2A. **(B)** Multidimensional scaling projection that illustrates relationships among escape maps of sera and monoclonal antibodies in two dimensions, similarly to the projection shown in Fig. 7.4. Similar mutations affect the binding of antibodies or sera located near one another in the plot. Here, the only serum samples shown are the day 119 samples for individuals who received the 100 or 250 μg vaccine dose. The 100 μg dose samples are shown in green and the 250 μg dose samples are shown in yellow (day 119 for all). The projection includes the escape maps of 22 monoclonal antibodies (escape maps first described in [83, 124, 122, 279, 278] of the 4 major structural classes to orient the plot. Antibodies are colored according to epitope, as in Fig. 7.2A. **(C)** Number of sera that fell into each of the four major categories of binding-escape maps as categorized by subjective visual inspection. **(D)** Composite line plots showing the total binding escape at each site in the RBD for each sample in gray, and the mean site-total escape for each group (100 or 250 μg vaccine doses) as a thicker black line. 100 μg, n=8; 250 μg, n=14. The same key sites within each epitope are highlighted, as in Fig. 7.4.


A
M19 (day 119)
M21 (day 119)
M17 (day 119)
M18 (day 119)
M16 (day 119)
M20 (day 119)
M22 (day 119)
M23 (day 119)
escape fraction
site
456 / 484
456 / 484 with low magnitude
core directed
broadly targeting
B
class 3 mAbs
class 2 mAbs
convalescent (n=11)
250µg (n=14)
100µg (n=8)
class 1 mAbs
class 4 mAbs
class 1 epitope
class 3 epitope
class 2 epitope
class 4 epitope
C
broadly targeting (4 of 8)
456 / 484-directed (1 of 8)
core-directed (1 of 8)
n=8
456 / 484-directed with low magnitude (2 of 8)
D
vaccine (100µg dose)
vaccine (250µg dose)
total escape at site
site

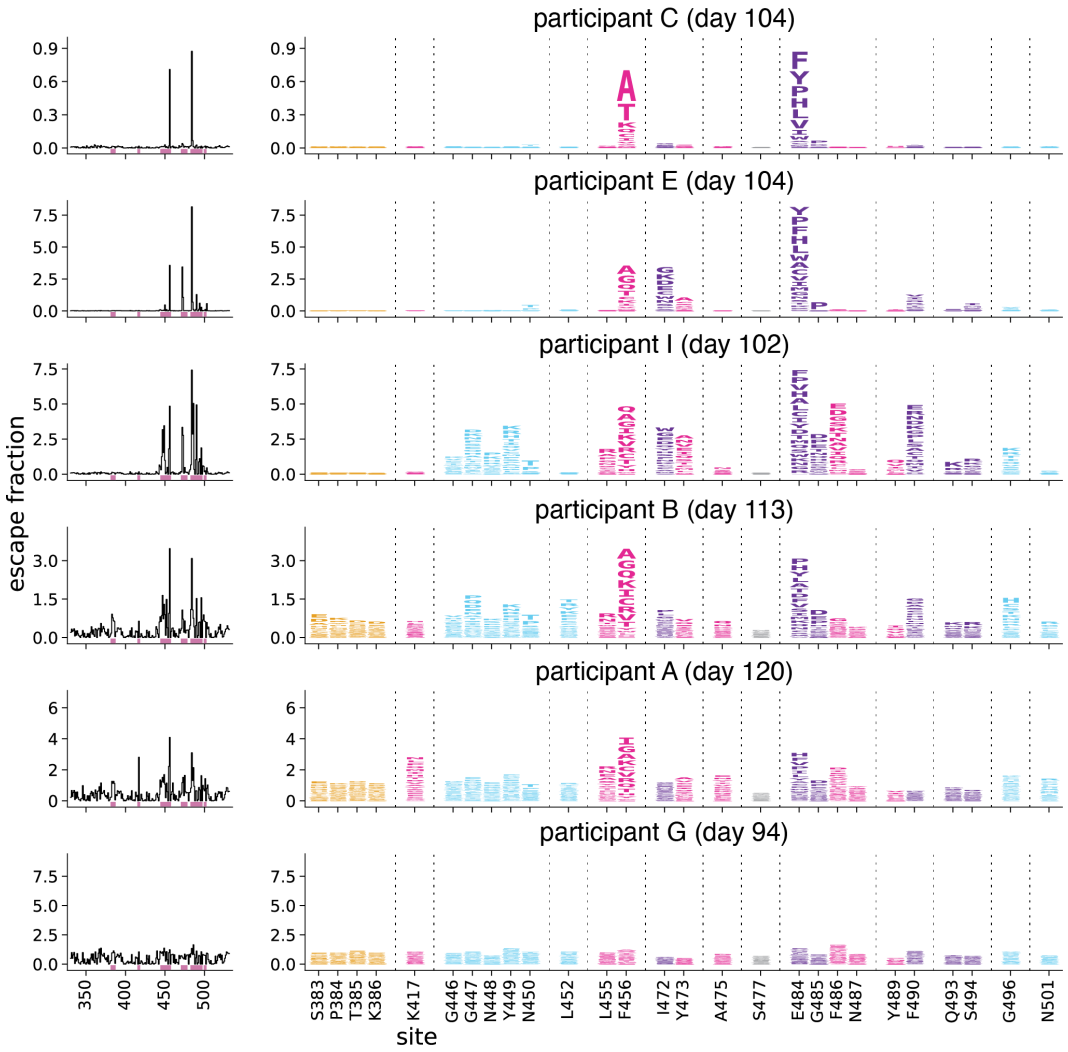


Figure 7.12: **Complete escape maps six representative convalescent plasma samples from the day 100–150 cohort**. RBD sites are colored according to epitope, as in Fig. 7.2A, with the same sites shown (highlighted in pink on the x-axis in the line plots at left). These six plasma samples are those used in neutralization assays shown in Fig. 7.5 and Fig. 7.11 and 7.14. Escape maps were first reported in [120].

Figure 7.13: **Escape maps and effects of individual RBD mutations on neutralization for representative samples from vaccinated and convalescent individuals**. **(A)** The site-wise antibody-binding escape for each of the vaccine and convalescent samples tested in neutralization assays is shown. **(B)** The effects of RBD mutations on neutralization of G614 spike-pseudotyped lentiviral particles with the indicated mutations are shown as the inhibitory concentration 50% (IC50). Naturally-occurring mutations are colored in white, and non-naturally-occurring mutations in gray. **(C)** The fold-change in IC50 compared to wild type spike, grouped by vaccine or convalescent sera is shown as in Fig. 7.5C, but shown here for all RBD mutants. Dashed line indicates no change in neutralization relative to wild type spike. Horizontal bars represent the group median fold-change IC50. In (B) and (C), each point represents the IC50 from one individual calculated from technical duplicates. The highest two points for E484K and K417N-G446V-E484K, and the highest 4 points for "all RBD antibodies depleted" are at the limit of detection. **(D)** Spike-pseudotyped lentiviral particle entry titers for RBD mutants tested in neutralization assays, calculated as the mean relative luciferase units per μL from 16 technical replicates. Mutations that are observed in at least one SARS-CoV-2 sequence in the Global Initiative on Sharing All Influenza Data (GISAID) are colored in white, and non-naturally-occurring mutations in gray. All spike sequences contained G614, which fixed in circulating sequences in 2020 [161].

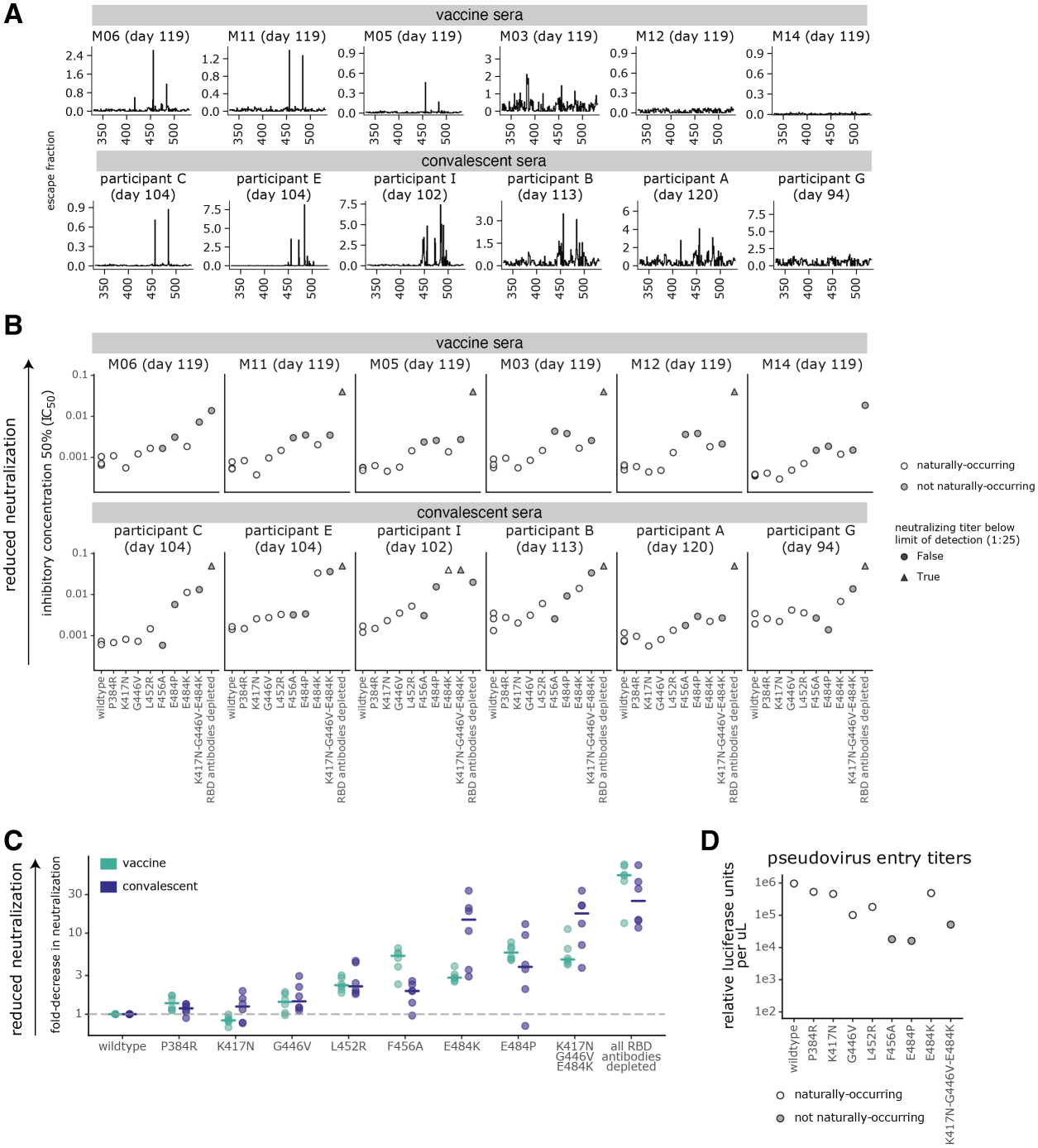
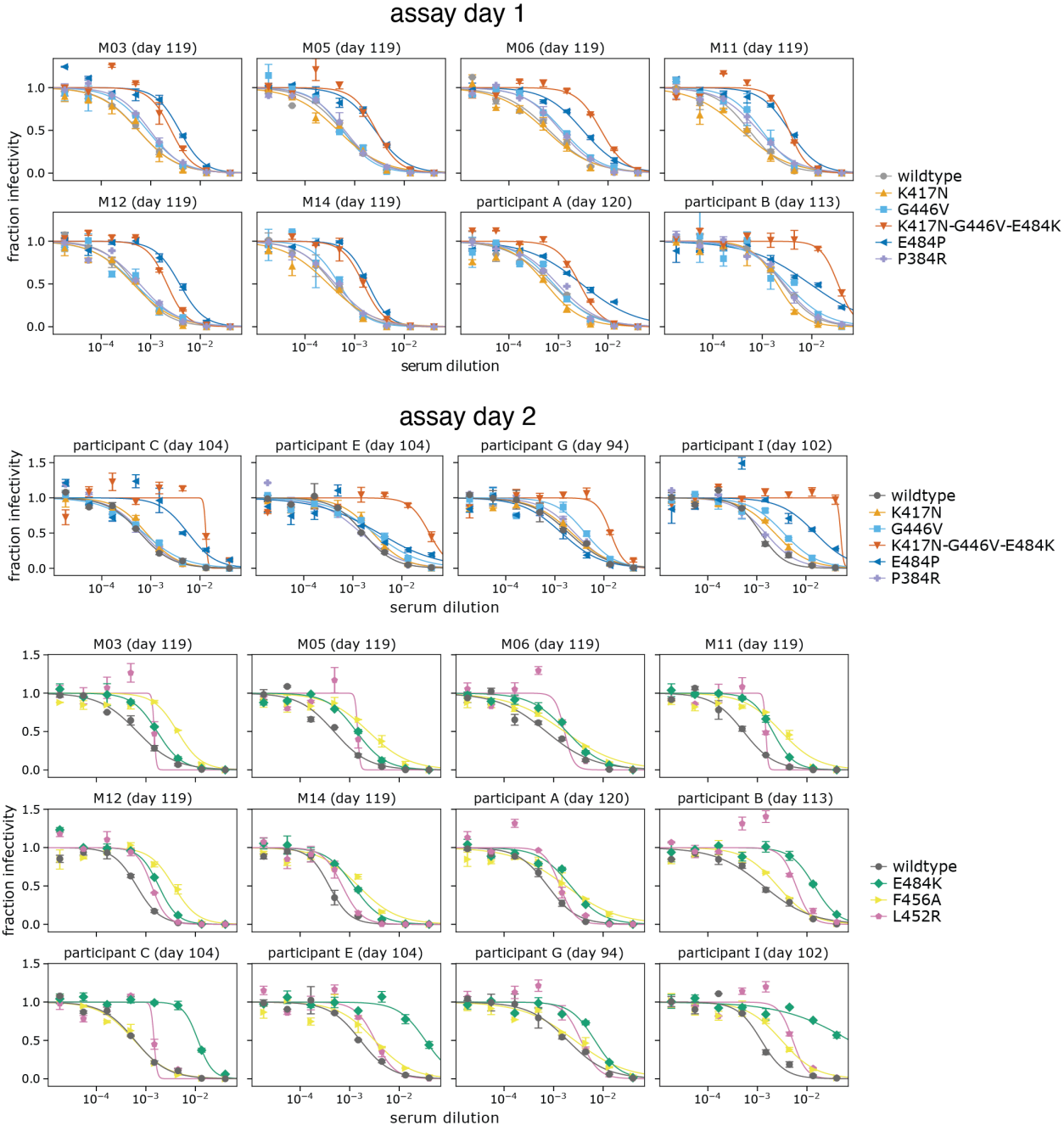

A
vaccine sera
M06 (day 119)
M11 (day 119)
M05 (day 119)
M03 (day 119)
M12 (day 119)
M14 (day 119)
convalescent sera
participant C (day 104)
participant E (day 104)
participant I (day 102)
participant B (day 113)
participant A (day 120)
participant G (day 94)
escape fraction
B
reduced neutralization
inhibitory concentration 50% (IC50)
naturally-occurring
not naturally-occurring
neutralizing titer below limit of detection (1:25)
False
True
wildtype
P384R
K417N
G446V
L452R
F456A
E484P
E484K
K417N-G446V-E484K
RBD antibodies depleted
C
reduced neutralization
fold-decrease in neutralization
vaccine
convalescent
all RBD antibodies depleted
D
pseudovirus entry titers
relative luciferase units per uL
naturally-occurring
not naturally-occurring

Figure 7.14: **Full neutralization curves for all assays evaluating the impact of RBD mutations on viral neutralization**. **(A)** Vaccine samples are prefixed with "M" and convalescent samples are prefixed with "participant". The x-axis shows the serum dilution, and the y-axis shows the fraction of viral infectivity remaining at that dilution. IC50 values were calculated by fitting 2-parameter Hill curves with the baselines fixed at one and zero. These IC50 values were used to determine the fold-change values in Fig. 7.5 and Fig. 7.11. In each plot, mutants are shown with the wild-type tested on the same date. Error bars represent the standard error of n=2 replicates. For readability, plots are arranged by assay date and no more than 6 curves are shown per plot. All points on the same plot are from assays performed on the same day and changes in neutralizing titer are computed relative to wild-type run on that day. In Fig. 7.5D, the wild-type curve from the first assay date is shown.


assay day 1
M03 (day 119)
M05 (day 119)
M06 (day 119)
M11 (day 119)
M12 (day 119)
M14 (day 119)
participant A (day 120)
participant B (day 113)
fraction infectivity
serum dilution
wildtype
K417N
G446V
K417N-G446V-E484K
E484P
P384R
assay day 2
participant C (day 104)
participant E (day 104)
participant G (day 94)
participant I (day 102)
fraction infectivity
serum dilution
wildtype
K417N
G446V
K417N-G446V-E484K
E484P
P384R
M03 (day 119)
M05 (day 119)
M06 (day 119)
M11 (day 119)
M12 (day 119)
M14 (day 119)
participant A (day 120)
participant B (day 113)
participant C (day 104)
participant E (day 104)
participant G (day 94)
participant I (day 102)
fraction infectivity
serum dilution
wildtype
E484K
F456A
L452R

Chapter 8

A SARS-COV-2 VARIANT ELICITS AN ANTIBODY RESPONSE WITH A SHIFTED IMMUNODOMINANCE HIERARCHY

A version of this chapter has been previously published as:

Allison J. Greaney, Tyler N. Starr, Rachel T. Eguia, Andrea N. Loes, Khadija Khan, Farina Karim, Sandile Cele, John E. Bowen, Jennifer K. Logue, Davide Corti, David Veesler, Helen Y. Chu, Alex Sigal, Jesse D. Bloom. A SARS-CoV-2 variant elicits an antibody response with a shifted immunodominance hierarchy. (2021) bioRxiv 2021.10.12.464114; doi: https://doi.org/10.1101/2021.10.12.464114

8.1 Abstract

Many SARS-CoV-2 variants have mutations at key sites targeted by antibodies. However, it is unknown if antibodies elicited by infection with these variants target the same or different regions of the viral spike as antibodies elicited by earlier viral isolates. Here we compare the specificities of polyclonal antibodies produced by humans infected with early 2020 isolates versus the B.1.351 variant of concern (also known as Beta or 20H/501Y.V2), which contains mutations in multiple key spike epitopes. The serum neutralizing activity of antibodies elicited by infection with both early 2020 viruses and B.1.351 is heavily focused on the spike receptor-binding domain (RBD). However, within the RBD, B.1.351-elicited antibodies are more focused on the "class 3" epitope spanning sites 443 to 452, and neutralization by these antibodies is notably less affected by mutations at residue 484. Our results show that SARS-

CoV-2 variants can elicit polyclonal antibodies with different immunodominance hierarchies.

8.2 Introduction

Over the past year, SARS-CoV-2 viral variants have emerged with mutations that alter the antigenicity of spike and erode neutralization of the virus by infection- and vaccine-elicited polyclonal antibodies [14, 42, 51, 77, 187, 313, 200, 201, 199, 209, 324, 62]. While it is well established that many SARS-CoV-2 variants are less susceptible to antibody immunity generated by early 2020 infections, it is unknown if the antibodies elicited by infection with these variants have different specificities and epitope immunodominance hierarchies. For influenza virus, it has been demonstrated that immunodominance of different epitopes changes over time as the virus evolves antigenically [219, 234, 260]. If a similar phenomenon occurs for SARS-CoV-2, then the sites of important antigenic mutations will change over time.

Here we address this question by combining serology and deep mutational scanning to compare the specificity of the polyclonal antibody response elicited by infection with early 2020 viruses versus the B.1.351 variant (also referred to as Beta or 20H/501Y.V2). The B.1.351 variant was first detected in Nelson Mandela Bay, South Africa and likely emerged in August 2020 after the country's first epidemic wave [285]. B.1.351 was the dominant lineage in South Africa by the end of 2020, although it has subsequently been displaced by the B.1.617.2 (Delta) lineage [138]. B.1.351 has mutations throughout the spike protein, including at key epitopes in both the RBD and NTD [42, 200, 202, 324]. The B.1.351 variant has among the largest reductions in neutralization by convalescent plasmas of any SARS-CoV-2 variant to date [51, 194, 232, 315, 313]. Additionally, prior work has demonstrated that B.1.351 convalescent plasmas can neutralize early 2020 viruses better than early 2020 plasmas can neutralize B.1.351 viruses [42, 211], suggesting that there may be a shift in the specificity of the antibody response. Our results described below expand this understanding

by showing that while neutralization by B.1.351-elicited plasma antibodies is still heavily focused on the RBD, their site-specificity within the RBD is somewhat shifted compared to antibodies elicited by early 2020 viruses. Specifically, within the RBD, B.1.351-elicited sera is relatively more targeted to the class 3 epitope (in the classification scheme of [16]) and relatively less targeted to the class 1 and 2 epitopes.

8.3 Results

8.3.1 The B.1.351 SARS-CoV-2 variant lineage has mutations in multiple spike epitopes

The B.1.351 spike used in our experiments contained the following mutations relative to the Wuhan-Hu-1 strain: D80A, D215G, del242–244, K417N, E484K, N501Y, D614G, and A701V (Fig. 8.1); note that some B.1.351 viruses also contain L18F. Three of these mutations are in the RBD (K417N, E484K, and N501Y). K417N and E484K strongly disrupt binding of class 1 and class 2 antibodies, respectively [122]. N501Y is in or proximal to the class 3 epitope, but does not strongly affect the binding or neutralization of polyclonal convalescent or vaccine-elicited antibodies [16, 315], although it enhances the RBD's affinity for its receptor, angiotensin converting enzyme 2 (ACE2) [62, 280, 289].

8.3.2 Convalescent plasma samples from individuals infected with B.1.351 or an early 2020 virus

We obtained plasma samples collected approximately 30 days post-symptom onset (mean 33, range 27–40 days) from 9 individuals infected with SARS-CoV-2 during the "second wave" of COVID-19 in South Africa from late December 2020 through late January 2021. During this timeframe, B.1.351 virus accounted for $> 90\%$ of sequenced infections in the area [42, 138, 285]. None of the individuals had evidence of prior SARS-CoV-2 infection, so we presume these individuals experienced a primary B.1.351 infection.

To enable comparison of B.1.351-elicited antibodies to those elicited by infection with an

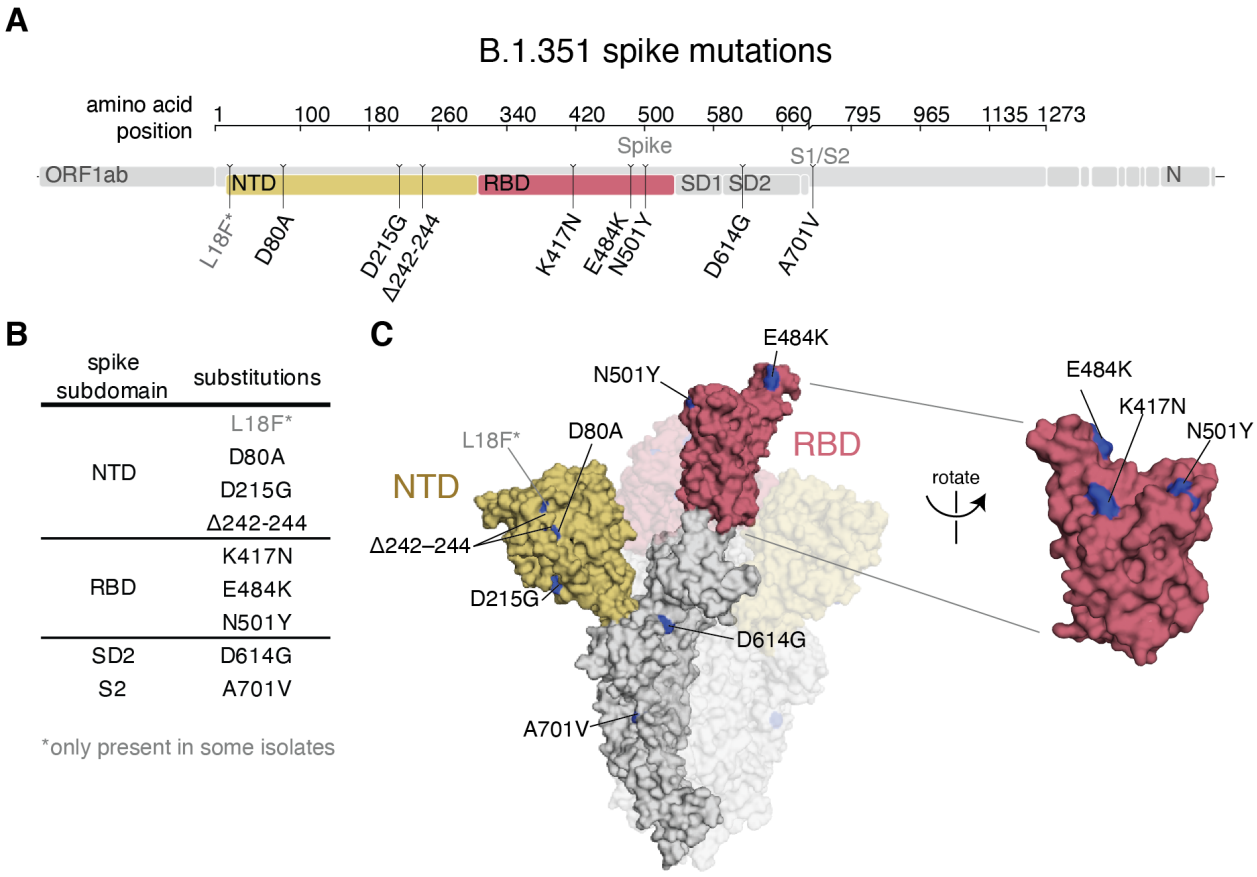


Figure 8.1: **B.1.351 spike mutations**. **(A,B)** Mutations in the B.1.351 spike relative to Wuhan-Hu-1 [285]. L18F is only present in some B.1.351 isolates. Visualization generated by `https://covdb.stanford.edu/sierra/sars2/by-patterns/`. **(C)** Sites where mutations occur in the spike ectodomain are highlighted in blue on the Wuhan-Hu-1 one-RBD open spike trimer (left, PDB 6ZGG) [333] or RBD (right) (PDB 6M0J) [172]. The surface of one spike monomer is shown; the other two protomers are transparent.

early 2020 virus, we reexamined a set of convalescent plasma samples collected approximately 30 days post-symptom onset (mean 32, range 15–61 days) from 17 individuals with symptom onset on or prior to March 15, 2020 in Washington state, USA [66, 120]. At that time, most sequenced viral isolates in Washington state had spike sequences identical to Wuhan-Hu-1, although D614G viruses were also present at a low level [161, 215]. No other spike mutations were present at appreciable frequencies at that time.

8.3.3 Infection with B.1.351 elicits a neutralizing antibody response at least as RBD-focused as early 2020 viruses

Early 2020 viruses induce a neutralizing antibody response that largely targets the RBD [74, 120, 229], although some neutralizing antibodies also bind the NTD [188, 200, 283, 303]. Because B.1.351 has mutations in both the RBD and NTD, it is important to determine if the specificity of the neutralizing antibody response elicited by this virus is similarly RBD-focused.

We depleted plasmas from B.1.351-infected individuals of B.1.351 RBD-binding antibodies, or performed a mock depletion, and measured neutralization of B.1.351 spike-pseudotyped lentiviral particles (Fig. 8.2A, 8.6, data file S1). The median neutralization titer (NT50) of these plasmas against the B.1.351-spike-pseudotyped lentiviral particles for the mock depletion was 2,459 (range 259–5,081). For 7 out of 9 samples, greater than 90% of neutralizing activity was ablated by removal of RBD-binding antibodies (Fig. 8.2A).

We compared these B.1.351 results to previous measurements of the RBD-focused neutralizing activity of plasmas from individuals infected with early 2020 viruses. These prior measurements were made using Wuhan-Hu-1 RBD depletions and D614G spike-pseudotyped lentiviral particles [120]. The neutralizing activity of the B.1.351 plasmas was at least as RBD-focused as the early 2020 virus plasmas, with most neutralizing activity of most plasmas from both cohorts attributable to RBD-binding antibodies (Fig. 8.2B,C). There was a

slight trend for the neutralizing activity of the B.1.351 plasmas to be more RBD-focused than the early 2020 plasmas, but the difference was not statistically significant (Fig. 8.2C). One caveat is that all neutralization assays were performed in 293T cells overexpressing ACE2, which tend to emphasize the effect of RBD-binding, ACE2-competitive antibodies more than assays performed on cells with lower levels of ACE2 expression [51, 283, 178].

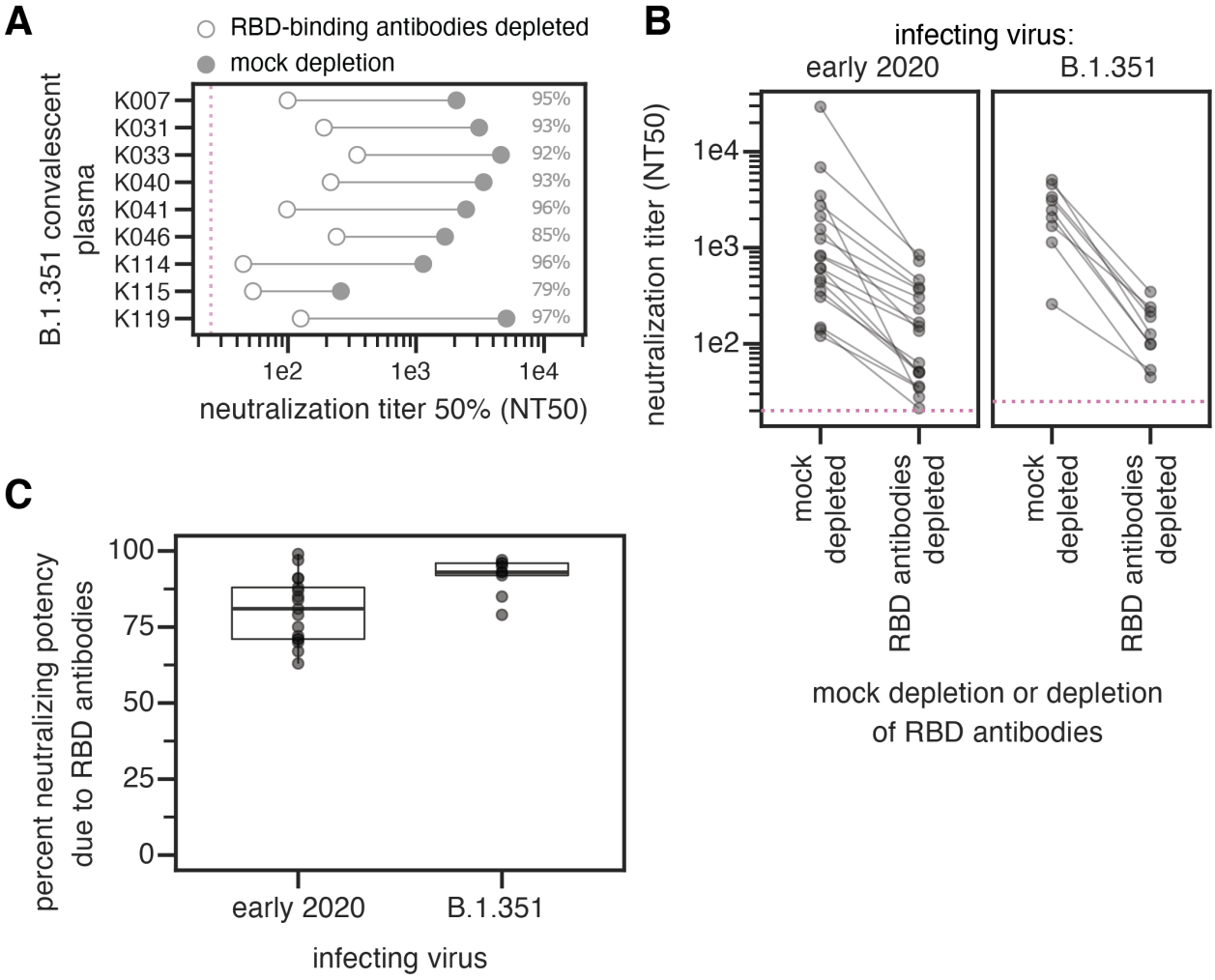
8.3.4 Complete mapping of mutations in the B.1.351 RBD that reduce binding by polyclonal plasma antibodies elicited by B.1.351 infection

To determine how mutations within the RBD affect plasma antibody binding, we used a previously described deep mutational scanning approach. Briefly, this approach involves generating comprehensive mutant libraries of the RBD, displaying the mutant RBDs on the surface of yeast, and using fluorescence-activated cell sorting (FACS) and deep sequencing to quantify how mutations impact antibody binding [124, 120].

Previously, we have performed such deep mutational scanning using the RBD from the Wuhan-Hu-1 isolate to map mutations that affect binding by polyclonal antibodies elicited by infection or vaccination that involves a RBD identical to that in Wuhan-Hu-1 [120, 121, 122]. However, for the current work we wanted to determine the specificity of antibodies elicited by B.1.351 infection to the B.1.351 RBD. Therefore, we generated new duplicate libraries containing 99.7% (3,807 of 3,819) of the possible single amino-acid mutations in the B.1.351 RBD. We displayed these libraries on the surface of yeast, and measured the effects of mutations on RBD expression and binding to ACE2 (Fig. 8.7, data file S2 [280]). We used computational filters based on these measurements as well as a pre-sort of the library for RBDs that bind ACE2 with at least 1% the avidity of the unmutated B.1.351 RBD to filter spurious antibody-escape mutations that were highly deleterious or led to gross unfolding of the RBD.

We then measured how all the single RBD mutations affected the binding of polyclonal

Figure 8.2: **The neutralizing activity of plasma antibodies elicited by B.1.351 infection is heavily focused on the RBD**. **(A)** The neutralizing titer (NT50) of plasmas from B.1.351-infected individuals against B.1.351 spike-pseudotyped lentiviral particles, following mock depletion or depletion of B.1.351 RBD-binding antibodies. **(B)** Comparison of neutralization titer following mock depletion or depletion of B.1.351 RBD-binding antibodies for early 2020 (n=17) [120] and B.1.351 convalescent plasmas (n=9). The pink dashed line in A,B indicates the limit of detection (NT50 of 25 for B.1.351 plasmas, and 20 for early 2020 plasmas). **(C)** Percent loss of neutralization after removal of RBD-binding antibodies for early 2020 and B.1.351 convalescent plasmas. The difference is not significant (Cox proportional-hazards test, accounting for censoring, $p = 0.12$). Experiments with B.1.351 infection-elicited plasmas were performed with B.1.351 RBD proteins and spike-pseudotyped lentiviruses, and experiments with early 2020 plasmas were performed with Wuhan-Hu-1 RBD proteins and D614G spike-pseudotyped lentiviruses. The data for the early 2020 viruses are reprinted from [120]. Neutralization titers are in data file S1 and at `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/experimental_data/results/rbd_depletion_neuts/RBD_depletion_NT50_b1351_haarvi.csv`. Full neutralization curves for the B.1.351 plasmas are in Fig. 8.6, and the full curves for the early 2020 plasmas are shown in the supplement of [120].


A
RBD-binding antibodies depleted
mock depletion
B.1.351 convalescent plasma
K007
K031
K033
K040
K041
K046
K114
K115
K119
95%
93%
92%
93%
96%
85%
96%
79%
97%
1e2
1e3
1e4
neutralization titer 50% (NT50)
B
infecting virus:
early 2020
B.1.351
neutralization titer (NT50)
1e4
1e3
1e2
mock depleted
RBD antibodies depleted
mock depleted
RBD antibodies depleted
mock depletion or depletion of RBD antibodies
C
percent neutralizing potency due to RBD antibodies
100
75
50
25
0
early 2020
B.1.351
infecting virus

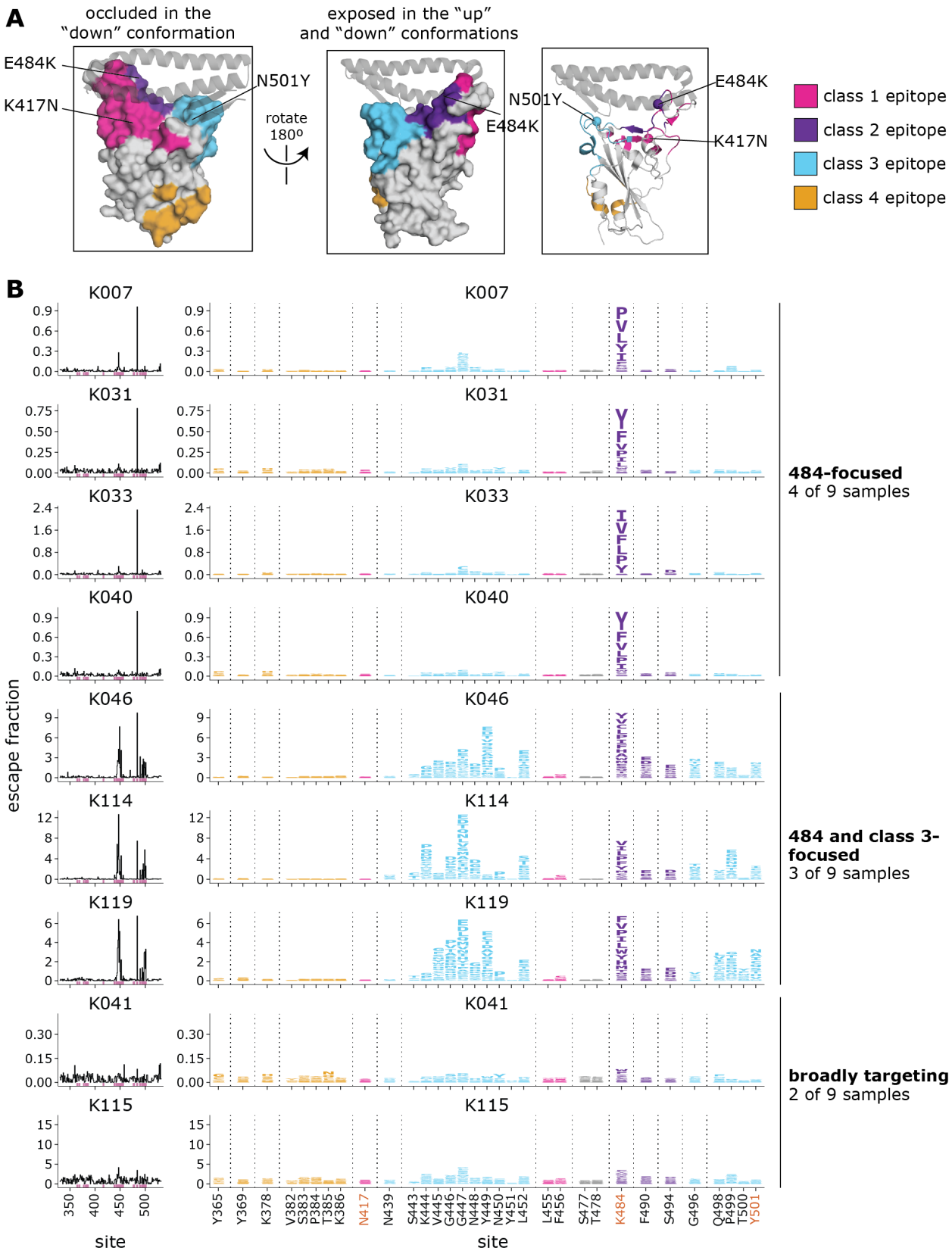
antibodies in the B.1.351 convalescent plasmas to the B.1.351 RBD. To do this, we incubated the yeast-displayed B.1.351 libraries with each plasma and used fluorescence-activated cell sorting (FACS) to enrich for RBD mutants with reduced antibody binding (Fig. 8.8A-C). We deep sequenced the pre- and post-enrichment populations to quantify each mutation's "escape fraction". These escape fractions range from 0 (no cells with the mutation in the escape bin) to 1 (all cells with the mutation in the plasma-escape bin) (data file S3). The escape fractions measured for independent biological replicate libraries were well-correlated (Fig. 8.8D), and in the sections below we report the average across the two replicate libraries. We represent the escape maps as logo plots, where the height of each letter is proportional to its escape fraction (Fig. 8.8A).

8.3.5 B.1.351-elicited antibodies focus on different epitopes than early 2020 convalescent samples

We examined the sites and epitopes to which mutations had the greatest effect on antibody binding. We use the Barnes, et al. [16] antibody epitope classification scheme, in which there are antibody classes 1 through 4 (Fig. 8.3A). The class 1, 2, and 3 antibodies are often potently neutralizing, while the class 4 antibodies are usually less potently neutralizing in vitro [74, 188, 229, 230, 356]. Relative to Wuhan-Hu-1, B.1.351 contains mutations in or proximal to the class 1, 2, and 3 epitopes (K417N, E484K, and N501Y, respectively) (Fig. 8.3A), although the N501Y mutation has little effect on polyclonal convalescent antibody binding or neutralization for Wuhan-Hu-1-like viruses [51, 313, 337].

For the B.1.351 plasmas, in 4 of 9 cases, mutations to site 484 within the class 2 epitope had the largest effects on antibody binding and the K484E reversion mutation had little effect (Fig. 8.3B, data file S3). In 3 of 9 cases, mutations to the class 3 epitope (sites 443–450, 498–501, shown in cyan) and the class 2 site 484 had comparably large effects on antibody binding. In two cases, no mutation had a particularly large effect on binding. Mutations to

Figure 8.3: **Complete maps of mutations in the B.1.351 RBD that reduce binding by B.1.351 convalescent plasmas**. **(A)** The Wuhan-Hu-1 RBD (PDB 6M0J) colored by antibody epitope [16]. The three sites where mutations distinguish the Wuhan-Hu-1 and B.1.351 RBDs are labeled. ACE2 is shown as a gray ribbon diagram. **(B)** Escape maps for B.1.351 convalescent plasmas. The line plots at left indicate the sum of effects of all mutations at each RBD site on plasma antibody binding, with larger values indicating more escape. The logo plots at right show key sites (highlighted in purple on the line plot x-axes). The height of each letter is that mutation's escape fraction; larger letters indicate a greater reduction in binding. For each sample, the y-axis is scaled independently. RBD sites are colored by epitope as in (A). Sites 417, 484, and 501 are labeled with red text on the x-axis. All escape scores are in data file S3 and at `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/results/supp_data/B1351_raw_data.csv`. Interactive versions of logo plots and structural visualizations are at `https://jbloomlab.github.io/SARS-CoV-2-RBD_B.1.351/`.


A
occluded in the "down" conformation
exposed in the "up" and "down" conformations
E484K
K417N
N501Y
rotate 180°
N501Y
E484K
E484K
K417N
class 1 epitope
class 2 epitope
class 3 epitope
class 4 epitope
B
escape fraction
site
484-focused
4 of 9 samples
484 and class 3-focused
3 of 9 samples
broadly targeting
2 of 9 samples

the class 1 and 4 antibody epitopes did not have large effects on plasma binding.

There are clear differences in the RBD epitope targeting of the B.1.351 plasmas versus previously characterized plasmas from a cohort of individuals (n=11) infected with early 2020 viruses in Washington state, USA [120]. These 11 samples are a subset of the 17 whose RBD-targeting neutralizing activity is described above (Fig. 8.2B,C). Specifically, binding of the early 2020 plasmas were most affected by mutations to the class 1 and 2 epitopes, with mutations to sites 456, 486, and 484 having some of the largest effects on binding to the RBD (Fig. 8.4, 8.9, data file S4), although mutation to site 456 have little effect on neutralization in vitro reflecting the common hyperfocusing of neutralizing antibody responses [120, 122]. While the B.1.351 plasmas were also strongly affected by mutations to the class 2 epitope and site 484, mutations to the class 1 epitope had little effect. Moreover, while both groups of plasmas are affected by class 3 epitope mutations, the relative importance of class 3 mutations is greater for the B.1.351 plasmas (Fig. 8.4A,B).

8.3.6 Class 3 epitope mutations have a larger effect on neutralization for B.1.351 plasmas, while mutations at the class 2 site 484 have a larger effect for early 2020 plasmas

To test if the differences in plasma antibody binding specificity described above lead to different effects of mutations on neutralization, we performed neutralization assays on key mutants using spike-pseudotyped lentiviral particles. For these experiments, we chose the eight B.1.351 samples with the highest neutralizing potency (there was not enough residual sample volume to perform neutralization assays with the lowest-potency sample). We also chose four early 2020 samples with substantial RBD-focused neutralizing activity and with antibody-binding escape maps representative of the early 2020 cohort as a whole (Fig. 8.9). In all assays, we tested neutralization by B.1.351 and early 2020 plasmas against point mutants in the homologous B.1.351 or D614G spikes.

Mutations to site 484 had strikingly different effects on neutralization by B.1.351 versus

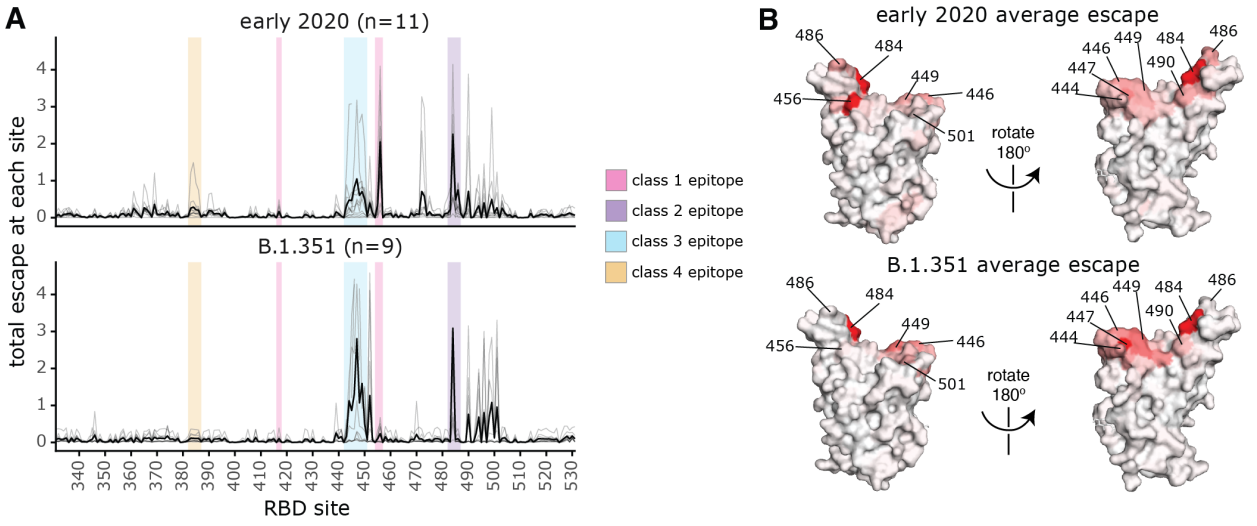


Figure 8.4: **Comparison of binding escape mutations between plasmas elicited by infection with B.1.351 versus early 2020 viruses**. **(A)** The total escape at each site is shown as a light gray line for each plasma in the early 2020 or B.1.351 cohorts. The thicker black line indicates the average for each cohort. Key antibody epitopes are highlighted, colored as in Fig. 8.2A. **(B)** The total escape at each site averaged across each cohort is mapped to the Wuhan-Hu-1 RBD surface (PDB 6M0J [172]), with sites colored from white to red, with white indicating no escape, and red being the site with the most escape. Interactive versions of logo plots and structural visualizations are at `https://jbloomlab.github.io/SARS-CoV-2-RBD_B.1.351/`. The early 2020 escape-mapping data in this figure were originally published in [120] and are reanalyzed here. The full escape maps for the early 2020 samples are shown in Fig. 8.9 and the full escape maps for the B.1.351 samples are shown in Fig. 8.3.

early 2020 plasmas. For the early 2020 plasmas, both E484K/Q mutations, as well as the K417N-E484K-N501Y triple mutation, reduced neutralization by > 10-fold, which is comparable to the reduction caused by removing all RBD-binding antibodies from the plasmas (Fig. 8.5). Therefore, the neutralizing activity of early 2020 plasmas is often highly focused on site 484, as has been described previously [51, 120, 122, 189, 313, 320, 324, 337, 351]. In contrast, mutations to site 484 had much smaller effects on neutralization by B.1.351 plasmas. The K484E reversion had little effect on neutralization by B.1.351 plasmas, which was striking given the large effect of E484K on early 2020 plasma neutralization. While the K484Q mutation had the largest effect on B.1.351 plasmas of any of the single mutations we tested (geometric mean of 3.0-fold change), the effect was smaller than that for the early 2020 plasmas (geometric mean of 18.3-fold change).

The class 3 epitope was a slightly more important target of neutralization for the B.1.351 plasmas than for early 2020 plasmas, consistent with the deep mutational scanning escape maps. The G446V mutation to the class 3 epitope had a slightly larger, but still modest, effect on neutralization for the B.1.351 plasmas than for most of the early 2020 plasmas (Fig. 8.5, 8.10). No tested mutation, nor the 417-484-501 triple mutant, reduced neutralization by the B.1.351 plasmas as much as removing all RBD-binding antibodies (Fig. 8.5), a result in stark contrast to that observed for the early 2020 plasmas.

8.4 Discussion

We found that a SARS-CoV-2 variant induces antibody responses with different immunodominance hierarchies than early SARS-CoV-2 viral isolates. Changes in immunodominance hierarchies over time and asymmetric antigenic drift have also been observed for influenza virus [219, 234, 258, 260]. Such changes can have important consequences, as they can contribute to individuals with different exposure histories having different susceptibilities to viral mutants [128, 329]. Although the changes in immunodominance we have observed

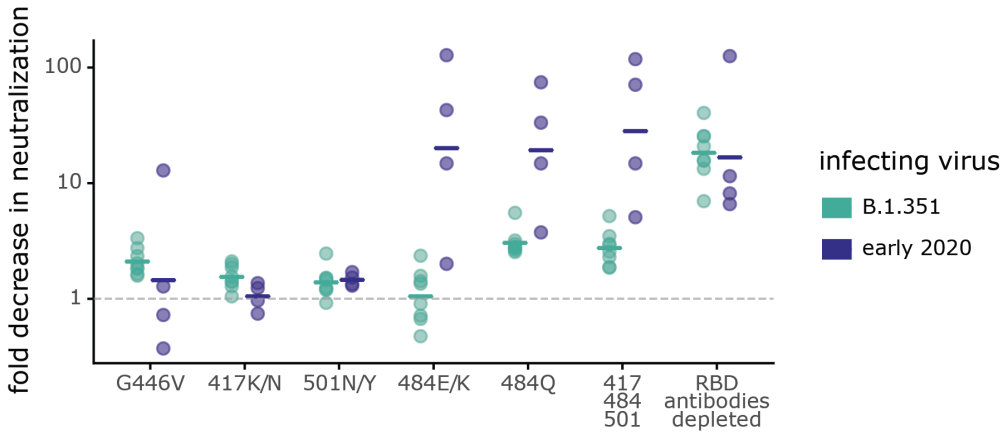


Figure 8.5: **Some mutations have different effects on neutralization by B.1.351 and early 2020 plasmas**. Plasmas from B.1.351- or early 2020-convalescent individuals were tested for neutralization of wildtype B.1.351 or D614G spike-pseudotyped lentiviral particles, respectively, and against the indicated point mutants in their respective parental backgrounds. The y-axis indicates the fold-change in neutralization caused by the mutations, with larger values indicating less neutralization. Each point is the average of two technical replicates for one individual. The crossbars indicate the group geometric mean. The dashed gray line is at 1 (i.e., mutation causes no change in neutralization). Sites 417, 484, and 501 differ between B.1.351 and early 2020 viruses, and so mutations are tested in each background that changes the identity to that in the other virus (e.g., E484K in early 2020 viruses, and K484E in B.1.351). Full neutralization curves and effects of mutations for each individual are shown in Fig. 8.10, and the numerical values and IC50s are given in data file S5 and at `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/experimental_data/results/neut_titers/neut_titers.csv`.

here are relatively modest, they could become larger as the virus continues to evolve and different individuals accumulate increasingly disparate exposure histories through infection and vaccination [107, 358].

We suggest several speculative hypotheses about several reasons why B.1.351 might elicit different hierarchies of antibodies. Although the B.1.351 spike protein has multiple mutations in key antigenic sites in the RBD and NTD [42, 324], the neutralizing antibody response elicited by B.1.351 infection is at least as RBD-focused as for early 2020 infections, suggesting that none of the RBD mutations have reduced the antigenicity of that spike subdomain. But within the RBD, site 484 is less immunodominant for B.1.351-elicited plasmas. Infection with early 2020 viruses frequently leads to the development of neutralizing class 2 antibodies that target an epitope containing site 484 [106, 122, 279, 320], and are derived from common antibody germline genes (e.g., IGHV3-53/66, IGHV3-30, IGHV1-2) [16, 245, 251]. We speculate that viruses containing K484 rather than E484 (such as B.1.351) might less readily elicit such neutralizing antibodies [210], or might elicit antibodies that draw less of their binding energy from site 484. Furthermore, if the class 2 epitope (containing site 484) is less immunogenic in B.1.351, that could lead to relatively stronger targeting of the class 3 epitope for B.1.351-elicited sera. Note that such phenomena could be human-specific, since the class 2 epitope containing site 484 is not as immunodominant in other species with different germline antibody genes (i.e., rhesus macaques) [133].

Changing immunodominance hierarchies could explain previous reports that polyclonal antibodies elicited by infection with different SARS-CoV-2 variants can have differing neutralizing breadths and specificities [34, 95, 174, 187]. For instance, prior studies of individuals infected with B.1.351 demonstrated that the convalescent plasmas from B.1.351-infected individuals neutralized early 2020 viruses better than early 2020 convalescent plasmas neutralized B.1.351 viruses [42, 211]. Our results help mechanistically explain this finding by showing that one of the key epitopes that differs between early 2020 viruses and B.1.351

(the class 2 epitope centered on site 484) is more immunodominant for early 2020 infections. Such changes in immunodominance hierarchies could also explain recent results suggesting that polyclonal antibodies elicited by B.1.351 infection are less effective at neutralizing the Delta (B.1.617.2) variant than antibodies elicited by early 2020 viruses [43, 187].

Our study has several limitations. The cohorts of individuals infected with early 2020 and B.1.351 viruses are small, and are geographically and temporally distinct. Nevertheless, the two cohorts are relatively well-matched with respect to age, sex, and days-post symptom onset of sample collection and assays were performed under comparable conditions. Our deep mutational scanning measured binding to yeast-displayed RBD, which may not capture all relevant features of full-length spike in the context of virus. Finally, our neutralization assays used pseudotyped lentiviral particles and ACE2-overexpressing cells, and some recent works suggest that the relative importance of different spike epitopes for neutralization can depend on the viral system and target cell line used [51, 56, 178, 283].

Although the B.1.351 variant has now been displaced by the Delta variant, our results illustrate the need to understand immunity elicited by different SARS-CoV-2 variants. As population immunity due to infection or vaccination increases, preexisting immunity is becoming an increasingly important driver of SARS-CoV-2 evolution [159], as has shown to be the case for seasonal coronaviruses [89, 158]. Moreover, as individuals begin to accumulate more complex SARS-CoV-2 immune histories due to multiple infections and/or vaccinations, the effects of immune imprinting or original antigenic sin [102, 316] may start to interact with the variant-specific immunodominance hierarchies we have described to create increasingly diverse antibody specificities in the human population.

8.5 Materials and Methods

Description of cohort and ethics statement

Samples were collected from participants enrolled in a prospective cohort study approved by the Biomedical Research Ethics Committee (BREC) at the University of KwaZulu–Natal (reference BREC/00001275/2020). Written informed consent was obtained from each participant. The mean age was 54 years (median 53; range 26–78 years). Four were males and 5 were females. All participants had symptomatic SARS-CoV-2 infection and a positive SARS-CoV-2 qPCR from a swab of the upper respiratory tract, and all participants required hospitalization. All 9 participants were HIV-negative. None of the participants had evidence of prior SARS-CoV-2 infection. Blood was sampled approximately 30 days post-symptom onset (mean 32.9, range 27–40 days) from 9 individuals infected with SARS-CoV-2 during the "second wave" of infections in South Africa from late December 2020 through late January 2021, when the B.1.351 virus was detected in $> 90\%$ of sequenced infections in the area [42, 285, 138]. B.1.351 infection was corroborated by the experimental findings in this paper that all plasmas bound to B.1.351 spike and RBD, had reduced binding to DMS library variants with mutations to site 484, and better neutralized B.1.351 spike-pseudotyped lentiviral particles relative to D614G particles. All participant samples had detectable antibody binding and neutralizing titers against B.1.351 SARS-CoV-2 spike.

Early-2020 convalescent plasma samples were previously described [66, 120] and collected as part of a prospective longitudinal cohort study of individuals with SARS-CoV-2 infection in Seattle, WA, between February and July 2020. The plasma samples from 17 individuals were examined here (8 of 17 females; age range 23 to 76 years, mean 51.6 years, median 56 years). These samples were collected approximately 30 days post-symptom onset (mean 31.6 days, median 29 days, min 15 days, max 61 days). Five cases were hospitalized, 2 were asymptomatic, and the remainder were symptomatic non-hospitalized. The neutralization

activity of plasma samples before and after depletion of RBD-binding antibodies in Fig. 8.2 and RBD binding-escape maps in Fig. 8.9 were previously reported [120], but neutralization assays for all 30-days post-symptom onset plasmas in Fig. 8.5, Fig. 8.10 were newly performed in this study. The neutralization assays on the 100-day early 2020 samples in Fig. 8.10 were previously reported [121]. This work was approved by the University of Washington Institutional Review Board.

Plasma separation from whole blood

Plasma was separated from EDTA-anticoagulated blood by centrifugation at 500 rcf for 10 min and stored at −80 °C. Aliquots of plasma samples were heat-inactivated at 56 °C for 30 min and clarified by centrifugation at 10,000 rcf for 5 min, after which the clear middle layer was used for experiments. Inactivated plasma was stored in single-use aliquots to prevent freeze–thaw cycles.

Construction of B.1.351 RBD yeast-displayed DMS library

Duplicate single-mutant site-saturation variant libraries were designed in the background of the spike receptor binding domain (RBD) from SARS-CoV-2 B.1.351 (identical to that from Wuhan-Hu-1, Genbank accession number MN908947, residues N331–T531, with the addition of the following amino-acid substitutions: K417N, E484K, N501Y), and produced by Twist Bioscience. The Genbank map of the plasmid encoding the unmutated SARS-CoV-2 B.1.351 RBD in the yeast-display vector is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/data/plasmid_maps/3021_pETcon-SARS-CoV-2-RBD_K417N_E484K_N501Y.gb`. The site-saturation variant libraries were delivered as double-stranded DNA fragments by Twist Bioscience.

This sequence has 5' and 3' flanking sequences that are unmutated in the variant libraries (lower case). The uppercase portion is the RBD coding sequence, amino acids N331–T531

(Wuhan-Hu-1 spike numbering). The libraries were designed to contain all 19 amino acids at each site in the RBD, without stop codons, with no more than one amino-acid mutation per variant. The variant gene fragments were PCR-amplified with these primers: 5'-tctgcaggctagtggtggag-3' and 5'-agatcggaagagcgtcgtgtagggaaagagtgtagatctcggtggtcgccgtatcattaattctcttaggattcgattcacattc-3'. A second round of PCR was performed using the same forward primer (5'-tctgcaggctagtggtggag-3') and the reverse primer 5'-ccagtgaattgtaatacgactcactatagggcgaattggagctcgcggccgcnnnnnnnnnnnnnnnnnagatcggaagagcgtcgtgtag-3' to append the Nx16 barcodes and add the overlapping sequences to clone into the recipient vector backbone as described in [280, 281].

Failed positions in the Twist-delivered library (sites 362, 501, and 524 in Wuhan-Hu-1 numbering) were mutagenized in-house using a PCR-based method with NNS degenerate primers and cloned into the unmutated wildtype backbone plasmid using NEB HiFi assembly, exactly as described in [281]. These were then PCR-amplified using the same 5'-tctgcaggctagtggtggag-3' and 5'-ccagtgaattgtaatacgactcactatagggcgaattggagctcgcggccgcnnnnnnnnnnnnnnagatcggaagagcgtcgtgtag-3' primers to pool with the barcoded Twist library gene fragments.

The barcoded variant gene fragments were cloned in bulk into the NotI/SacI-digested unmutated wildtype plasmid, as described in [280, 281]. The Genbank plasmid map for the fully assembled, barcoded B.1.351 RBD libraries (with the unmutated B.1.351 RBD sequence) is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/data/plasmid_maps/pETcon-SARS-CoV-2-RBD-B1351_lib-assembled.gb`. The pooled, barcoded mutant libraries were electroporated into E. coli (NEB 10-beta electrocompetent cells, New England BioLabs C3020K) and plated at a target bottleneck of 50,000 variants per duplicate library, corresponding to > 10 barcodes per mutant within each library. Colonies from bottlenecked transformation plates were scraped and plasmid purified. Plasmid libraries (10 μg plasmid per replicate library) were transformed into the

AWY101 yeast strain [322] according to the protocol of Gietz and Schiestl [115].

PacBio sequencing to link variant mutations and barcodes

As described by Starr et al. [280], PacBio sequencing was used to generate long sequence reads spanning the Nx16 barcode and RBD coding sequence. PacBio sequencing amplicons were prepared from library plasmid pools via NotI digestion, gel purification, and Ampure XP bead clean-up. Sample-specific barcodes and SMRTbells were ligated using the HiFi Express v2 kit. The multiplexed libraries were sequenced on a PacBio Sequel II with a 15-hour movie collection time. Demultiplexed PacBio HiFi circular consensus sequences (CCSs) were generated using the SMRT Link GUI, version 10.1.0.119588. HiFi reads are CCSs with ≥ 3 full passes and a mean quality score $Q \geq 20$. The resulting CCSs are available on the NCBI Sequence Read Archive, BioProject PRJNA770094, BioSample SAMN22208699.

HiFi reads were processed using alignparse (version 0.2.6) [65] to determine each variant's mutations and the associated Nx16 barcode sequence, requiring no more than 45 nucleotide mutations from the intended target sequence, an expected 16-nt length barcode sequence, and no more than 4 mismatches across the sequenced portions of the vector backbone. Attribution of barcodes to library variants determined that the libraries contained 3,807 of the 3,819 possible single amino-acid mutations to the B.1.351 RBD. Approximately 26% of barcodes in the duplicate libraries corresponded to wildtype B.1.351 RBD (Fig. 8.7A). The libraries were designed to contain only wildtype and 1-amino acid mutations, but some multiple mutations and stop codons were stochastically introduced during the library generation process. These mutations were excluded from downstream analysis of the effects of mutations on ACE2 binding, RBD expression, and plasma antibody binding, except when used in quality control checks (i.e., that most variants containing premature stop codons should not be expressed on the yeast cell surface and thus should have very low expression scores).

Determining the effects of mutations on RBD expression and ACE2 binding to filter the library for functional variants

The effects of each mutation on RBD expression on the surface of yeast and on ACE2 binding were measured essentially as described previously for the Wuhan-Hu-1 RBD [280]. Specifically, each biological replicate library was grown overnight at 30°C in 45mL SD-CAA media (6.7g/L Yeast Nitrogen Base, 5.0g/L Casamino acids, 1.065 g/L MES acid, and 2% w/v dextrose) at an initial OD600 of 0.4. To induce RBD surface expression, yeast were back-diluted in SG-CAA+0.1%D (2% w/v galactose supplemented with 0.1% dextrose) induction media at 0.67 OD600 and incubated at room temperature for 16–18 hours with mild agitation. For RBD expression experiments, 45 OD units of yeast were labeled in 1:100 diluted chicken-anti-Myc-FITC antibody (Immunology Consultants CMYC45F) to detect the RBD's C-terminal Myc tag. For ACE2-binding experiments, 12 OD units of yeast were incubated overnight at room temperature with monomeric biotinylated ACE2 (ACROBiosystems AC2-H82E8) across a concentration range of 10^{-13} M to 10^{-6} M at 1-log intervals. Labeling volumes were increased at low ACE2 concentration to limit ligand depletion effects. Cells were then labeled with 1:100 diluted Myc-FITC to detect RBD expression and 1:200 Streptavidin-PE (Invitrogen S866) to detect binding of biotinylated ACE2.

Cells were processed on a BD FACSAria II and sorted into four bins from low to high RBD expression (measured by myc-FITC staining) or ACE2 binding (measured by streptavidin-PE fluorescence). The RBD expression sort bins were set such that bin 1 would capture 99% of unstained cells, and the remaining 3 bins divide the remainder of each mutant RBD library into equal tertiles. For ACE2 binding, bin 1 captured 95% of cells expressing unmutated RBD incubated with no ACE2 (0 M), and bin 4 captured 95% of cells expressing unmutated RBD incubated with a saturating amount of ACE2 (10^{-6} M). Bins 2 and 3 equally divided the distance between the bin 1 upper and bin 4 lower fluorescence boundaries on a log scale. The frequency of each variant in each bin was determined by Illumina sequencing of RBD

variant barcodes.

The effects of each mutation on RBD expression and ACE2 binding were determined as described in [280]. RBD mutant expression and ACE2 binding scores were calculated according to the equations in [280]. For ACE2 binding, a score of -1.0 corresponds to a 10-fold loss in affinity (K_d) compared to the wildtype RBD. For RBD expression, a score of -1.0 corresponds to a 10-fold reduction in mean RBD-myc-FITC fluorescence intensity. These measurements were used to computationally filter library variants that were highly deleterious for RBD expression or ACE2 binding and would likely represent spurious antibody-escape mutations (see below for details). The ACE2 binding and RBD expression scores for the single amino-acid mutations in the B.1.351 RBD are available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/data/final_variant_scores.csv`.

As previously described, prior to performing the antibody-escape experiments, the yeast libraries were pre-sorted for RBD expression and binding to dimeric ACE2 (ACROBiosystems AC2-H82E6) to eliminate RBD variants that are completely misfolded or non-functional, such as those lacking modest ACE2 binding affinity [124]. Specifically, unmutated B.1.351 RBD and each RBD mutant library were incubated with dimeric ACE2 at 10^{-8} M (a saturating concentration of ACE2 for unmutated B.1.351 RBD). A FACS selection gate was set to capture 98% of cells expressing unmutated B.1.351 RBD that were incubated with 10^{-10} M ACE2, to purge the mutant libraries of highly deleterious mutations (i.e., those that have $< 1\%$ the affinity of unmutated B.1.351 RBD). These pre-sorted yeast libraries containing RBD variants with at least nominal expression and ACE2 binding were used in downstream antibody-escape experiments (see below).

Depleting plasma of nonspecific yeast-binding antibodies prior to antibody-escape experiments

Prior to the yeast-display deep mutational scanning, plasma samples were twice-depleted of nonspecific yeast-binding antibodies. AWY101 yeast containing a negative control (contain-

ing an empty vector pETcon plasmid) were grown overnight at 30°C in galactose-containing media. Then, up to 50 μliters of plasma samples were incubated, rotating, with 40 OD units of the yeast for 2 hours at room temperature in a total volume of 1mL. The yeast cells were pelleted by centrifugation, and the supernatant was transferred to an additional 40 OD units of yeast cells, and the incubation was repeated overnight at 4°C. Before beginning the plasma-escape mapping experiments, the negative control yeast were pelleted by centrifugation and the supernatant (containing serum antibodies but not negative control yeast or yeast-binding antibodies) was used in plasma-escape mapping.

FACS sorting of yeast libraries to select B.1.351 mutants with reduced binding by polyclonal plasmas from B.1.351-convalescent individuals

Plasma mapping experiments were performed in biological duplicate using the independent mutant RBD libraries, similarly to as previously described for monoclonal antibodies [124] and polyclonal plasma samples [120]. Mutant yeast libraries induced to express RBD were washed and incubated with plasma at a range of dilutions for 1 hour at room temperature with gentle agitation. For each plasma, we chose a sub-saturating dilution such that the amount of fluorescent signal due to plasma antibody binding to RBD was approximately equal across samples. The exact dilution used for each plasma is given in Fig. 8.8. After the plasma incubations, the libraries were secondarily labeled for 1 hour with 1:100 fluorescein isothiocyanate-conjugated anti-MYC antibody (Immunology Consultants Lab, CYMC-45F) to label for RBD expression and 1:200 Alexa Fluor-647-conjugated goat anti-human-IgA+IgG+IgM (Jackson ImmunoResearch 109-605-064) to label for bound plasma antibodies. A flow cytometric selection gate was drawn to capture 3–6% of the RBD mutants with the lowest amount of plasma binding for their degree of RBD expression (Fig. 8.8). For each sample, approximately 10 million RBD+ cells (range 10^7 to 1.5×10^7 cells) were processed on the BD FACSAria II cell sorter, with between 4×10^5 and 2×10^6 plasma-escaped

cells collected per sample. Antibody-escaped cells were grown overnight in synthetic defined medium with casamino acids (6.7g/L Yeast Nitrogen Base, 5.0g/L Casamino acids, 1.065 g/L MES acid, and 2% w/v dextrose + 100 U/mL penicillin + 100 μg/mL streptomycin) to expand cells prior to plasmid extraction.

DNA extraction and Illumina sequencing

Plasmid samples were prepared from 30 optical density (OD) units (1.6e8 colony forming units (cfus)) of pre-selection yeast populations and approximately 5 OD units (3.2e7 cfus) of overnight cultures of plasma-escaped cells (Zymoprep Yeast Plasmid Miniprep II) as previously described [124]. The 16-nucleotide barcode sequences identifying each RBD variant were amplified by polymerase chain reaction (PCR) and prepared for Illumina sequencing as described in [280]. Specifically, a primer with the sequence 5'-AATGATACGGCGACCACCGAGA-3' was used to anneal to the Illumina P5 adaptor sequence, and the PerkinElmer NextFlex DNA Barcode adaptor primers with the sequence 5'-CAAGCAGAAGACGGCATACGAGATxxxxxxxxGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3' (where xxxxxxxx indicates the sample index sequence) were used to anneal to the Illumina P7 adaptor sequence and append sample indexes for sample multiplexing. Barcodes were sequenced on an Illumina HiSeq 2500 with 50 bp single-end reads. To minimize noise from inadequate sequencing coverage, we ensured that each antibody-escape sample had at least 2.5x as many post-filtering sequencing counts as FACS-selected cells, and reference populations had at least 2.5e7 post-filtering sequencing counts.

Analysis of deep sequencing data to compute each mutation's escape fraction

Escape fractions were computed as described in [124], with minor modifications as noted below. We used the dms_variants package (`https://jbloomlab.github.io/dms_variants/`, version 0.8.10) to process Illumina sequences into counts of each barcoded RBD variant in

each pre-selection and antibody-escape population. For each plasma selection, we computed the escape fraction for each barcoded variant using the deep sequencing counts for each variant in the original and plasma-escape populations and the total fraction of the library that escaped antibody binding via the formula provided in [124]. These escape fractions represent the estimated fraction of cells expressing that specific variant that falls in the escape bin, such that a value of 0 means the variant is always bound by plasma and a value of 1 means that it always escapes plasma binding.

We then applied a computational filter to remove variants with > 1 amino-acid mutation, low sequencing counts, or highly deleterious mutations that might cause antibody escape simply by leading to poor expression of properly folded RBD on the yeast cell surface [124, 280]. Specifically, we removed variants that had ACE2 binding scores < -3.0 or expression scores < -1.0, after calculating mutation-level deep mutational scanning scores for this library as in [280]. An ACE2 binding score threshold of -3.0 retained 99.4% and an RBD expression score threshold of -1.0 retained 93.8% of all RBD mutations observed $\geq 50\times$ in GISAID as of Aug. 1, 2021 Fig. (Fig. 8.7C).

We also removed all mutations where the wildtype residue was a cysteine. There were 2,014 out of the possible 3,653 mutations to non-disulfide bond residues in the RBD that passed these computational filters.

The reported antibody-escape scores throughout the paper are the average across the libraries; these scores are also in data file S3. Correlations in final single-mutant escape scores are shown in Fig. 8.8D.

For plotting and analyses that required identifying RBD sites of strong escape, we considered a site to mediate strong escape if the total escape (sum of mutation-level escape fractions) for that site exceeded the median across sites by $>$ 5-fold, and was at least 5% of the maximum for any site. Full documentation of the computational analysis is at `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351`.

Differences between composition and analysis of B.1.351 RBD libraries and Wuhan-Hu-1 libraries

Importantly, because the B.1.351 libraries were generated using a different method than the Wuhan-Hu-1 RBD libraries, which is fully described in [280], the analysis of deep sequencing data to compute each mutation's escape fraction is also different. The newly generated B.1.351 libraries were ordered from Twist Bioscience to have one amino-acid mutation per variant, whereas the Wuhan-Hu-1 libraries were generated in-house with a PCR-based approach, with an average of 2.7 mutations per variant [280]. Because there were often multiple mutations per variant for the Wuhan-Hu-1 libraries, global epistasis modeling was used to deconvolve the effects of single amino-acid mutations on antibody binding [124, 120], whereas for the B.1.351 libraries, the measurements for single-mutant variants were used directly (occasional variants with multiple mutations were discarded) to calculate antibody escape.

Generation of pseudotyped lentiviral particles

HEK-293T (American Type Culture Collection, CRL-3216) cells were used to generate SARS-CoV-2 spike-pseudotyped lentiviral particles and 293T-ACE2 cells (Biodefense and Emerging Infectious Research Resources Repository (BEI Resources), NR-52511) were used to titer the SARS-CoV-2 spike-pseudotyped lentiviral particles and to perform neutralization assays (see below).

For experiments involving D614G spike, we used spike-pseudotyped lentiviral particles that were generated essentially as described in [67], using a codon-optimized SARS-CoV-2 spike from Wuhan-Hu-1 strain that contains a 21-amino-acid deletion at the end of the cytoplasmic tail [66] and the D614G mutation that is now predominant in human SARS-CoV-2 [161]. The plasmid encoding this spike, HDM_Spikedelta21_D614G, is available from Addgene (#158762) and BEI Resources (NR-53765), and the full sequence is at (`https://www.addgene.org/158762`). Point mutations were introduced into the RBD of this

plasmid via site-directed mutagenesis.

For experiments involving B.1.351 spike, we introduced the following mutations into the HDM_Spikedelta21_D614G plasmid to match the amino acid sequence of EPI_ISL_700420: 80A, D215G, L242-244del, K417N, E484K, N501Y, and A701V. This plasmid map is available online at `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/data/plasmid_maps/2957_HDM_Spikedelta21_B.1.351.gb`.

To generate spike-pseudotyped lentiviral particles [67], 6e105 HEK-293T (ATCC CRL-3216) cells per well were seeded in 6-well plates in 2 mL D10 growth media (Dulbecco's Modified Eagle Medium with 10% heat-inactivated fetal bovine serum, 2 mM l-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin). 24 hours later, cells were transfected using BioT transfection reagent (Bioland Scientific) with a Luciferase_IRES_ZsGreen backbone, Gag/Pol lentiviral helper plasmid (BEI Resources NR-52517), and wild-type or mutant SARS-CoV-2 spike plasmids. Media was changed to fresh D10 at 24 hours post-transfection. At 60 hours post-transfection, viral supernatants were collected, filtered through a 0.45 μm surfactant-free cellulose acetate low protein-binding filter, and stored at -80 °C.

Titering of pseudotyped lentiviral particles

Titers of spike-pseudotyped lentiviral particles were determined as described in [67] with the following modifications. 100 μL of diluted spike-pseudotyped lentiviral particles was added to 1.25e4 293T-ACE2 cells (BEI Resources NR-52511), grown overnight in 50 μL of D10 growth media in a 96-well black-walled poly-L-lysine coated plate (Greiner Bio-One, 655936). Relative luciferase units (RLU) were measured 65 hours post-infection (Promega Bright-Glo, E2620) in the infection plates with a black back-sticker (Thermo Fisher Scientific, NC9425162) added to minimize background. Titers were first estimated from the average of 8 two-fold serial dilutions of virus starting at 10 μL virus in a total volume of 150 μL, performed in duplicate.

Neutralization assays

293T-ACE2 cells (BEI Resources NR-52511) were seeded at 1.25e4 cells per well in 50 μL D10 in poly-L-lysine coated, black-walled, 96-well plates (Greiner 655930). 24 hours later, pseudotyped lentivirus supernatants were diluted to approximately 200,000 RLU per well (determined by titering as described above) and incubated with a range of dilutions of plasma for 1 hour at 37 °C. 100 μL of the virus-antibody mixture was then added to cells. At about 50 or 70 hours post-infection, luciferase activity was measured using the Bright-Glo Luciferase Assay System (Promega, E2610). Fraction infectivity of each plasma antibody-containing well was calculated relative to a no-plasma well inoculated with the same initial viral supernatant in the same row of the plate. We used the neutcurve package (`https://jbloomlab.github.io/neutcurve` version 0.5.7) to calculate the inhibitory concentration 50% (IC50) and the neutralization titer 50% (NT50), which is 1/IC50, of each plasma against each virus by fitting a Hill curve with the bottom fixed at 0 and the top fixed at 1.

Depletion of RBD-binding antibodies from polyclonal sera

Two rounds of sequential depletion of RBD-binding antibodies were performed for vaccine-elicited sera. Magnetic beads conjugated to the SARS-CoV-2 B.1.351 RBD (ACROBiosystems, MBS-K032) were prepared according to the manufacturer's protocol. Beads were resuspended in ultrapure water at 1 mg beads/mL and a magnet was used to wash the beads 3 times in phosphate-buffered saline (PBS) with 0.05% bovine serum albumin (BSA). Beads were then resuspended in PBS with 0.05% BSA at 1 mg beads per mL. Beads (manufacturer-reported binding capacity of 10–40 μg/mL anti-RBD antibodies) were incubated with human plasma at a 2:1 ratio beads:plasma, rotating overnight at 4°C or for 2 hours at room temperature. A magnet (MagnaRack Magnetic Separation Rack, Thermo Fisher Scientific, CS15000) was used to separate antibodies that bind RBD from the supernatant, and the supernatant (the post-RBD antibody depletion sample) was removed. A mock depletion (pre-depletion

sample) was performed by adding an equivalent volume of PBS + 0.05% BSA and rotating overnight at 4°C or for 2 hours at room temperature. Up to three rounds of depletions were performed to ensure full depletion of RBD-binding antibodies. For the neutralization assays on these plasmas depleted of RBD-binding antibodies, the reported plasma dilution is corrected for the dilution incurred by the depletion process. Note that these assays were performed in 293T cells over-expressing human ACE2, which may underestimate contributions of non-RBD-binding antibodies to viral neutralization [51, 56, 283].

Measurement of plasma binding to RBD or spike by enzyme-linked immunosorbent assay (ELISA)

The IgG ELISAs for spike protein and RBD were conducted as previously described [82]. Briefly, ELISA plates were coated with recombinant B.1.351 spike (purified and prepared as described in [82]) and RBD (ACROBiosystems, SPD-C52Hp) antigens described in at 2 μg/mL. Five 3-fold serial dilutions of sera beginning at 1:500 were performed in PBS with 0.1% Tween with 1% Carnation nonfat dry milk. Dilution series of the synthetic sera comprised of the anti-RBD antibody REGN10987 [129], which binds to both Wuhan-1-like RBD and B.1.351 RBD, and pooled pre-pandemic human serum from 2017–2018 (Gemini Biosciences; nos. 100–110, lot H86W03J; pooled from 75 donors) were performed such that the anti-spike antibody was present at a highest concentration of 0.25 μg/mL. REGN10987 was recombinantly produced by Genscript. The REGN10987 is the same as that used in [278]. Pre-pandemic serum alone, without anti-RBD antibody depletion, was used as a negative control, averaged over 2 replicates. Secondary labeling was performed with goat anti-human IgG-Fc horseradish peroxidase (HRP) (1:3000, Bethyl Labs, A80-104P). Antibody binding was detected with TMB/E HRP substrate (Millipore Sigma, ES001) and 1 N HCl was used to stop the reaction. OD450 was read on a Tecan infinite M1000Pro plate reader.

Data visualization

The static logo plot visualizations of the escape maps in the paper figures were created using the dmslogo package (`https://jbloomlab.github.io/dmslogo`, version 0.6.2) and in all cases the height of each letter indicates the escape fraction for that amino-acid mutation calculated as described above. For each sample, the y-axis is scaled to be the greatest of (a) the maximum site-wise escape metric observed for that sample, (b) 20x the median site-wise escape fraction observed across all sites for that plasma, or (c) an absolute value of 1.0 (to appropriately scale samples that are not noisy but for which no mutation has a strong effect on antibody binding). Sites K417, L452, S477, T478, E484, and N501 have been added to logo plots due to their frequencies among circulating viruses. The code that generates these logo plot visualizations is available at `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/results/summary/escape_profiles.md`. In many of the visualizations, the RBD sites are categorized by epitope region [16] and colored accordingly. We define the class 1 epitope as residues 403+405+406+417+420+421+453+455–460+473–476+486+487+489+504, the class 2 epitope as residues 472+483–485+490–494, the class 3 epitope to be residues 345+346+437-452+496+498–501, and the class 4 epitope as residues 365–372+378+382–386.

For the static structural visualizations in the paper figures, the RBD surface (PDB 6M0J) was colored by the site-wise escape metric at each site, with white indicating no escape and red scaled to be the same maximum used to scale the y-axis in the logo plot escape maps, determined as described above. We created interactive structure-based visualizations of the escape maps using dms-view [137] that are available at `https://jbloomlab.github.io/SARS-CoV-2-RBD_B.1.351/`. The logo plots in these escape maps can be colored according to the deep mutational scanning measurements of how mutations affect ACE2 binding or RBD expression as described above.

Statistical Analysis

The percent of neutralizing activity of early-2020 and B.1.351-convalescent plasmas due to RBD-binding antibodies is plotted with the plotnine python package, version 0.8.0 (`https://plotnine.readthedocs.io/en/stable/index.html`), shown as a Tukey boxplot (middle line indicating median, box limits indicating interquartile range) with individual measurements overlaid as points. P-values are from a log-rank test accounting for censoring, calculated with the lifelines python package, version 0.25.10 (`https://lifelines.readthedocs.io/en/latest/`).

8.6 Notes

Acknowledgments

We thank Cathy Lin for administrative support; Dolores Covarrubias, Andy Marty, the Genomics and Flow Cytometry core facilities at the Fred Hutchinson Cancer Research Center, and Katy Munson at the University of Washington PacBio Sequencing Services for experimental support. We also thank all study participants for their generous participation and contribution to this work.

Funding

This project has been funded in part with federal funds from the NIAID/NIH under contract numbers HHSN272201400006C and 75N93021C00015. This work was supported by grants from the NIAID / NIH (R01AI141707 to J.D.B., and T32AI083203 to A.J.G.., DP1AI158186 and HHSN272201700059C to D.V.), a Pew Biomedical Scholars Award (D.V.), an Investigators in the Pathogenesis of Infectious Disease Awards from the Burroughs Wellcome Fund (D.V.), and the Gates Foundation (INV-004949 to J.D.B, INV-016575 to H.Y.C. OPP1156262 to D.V., and INV-018944 to A.S.). The Scientific Computing Infrastructure

at Fred Hutch is funded by ORIP grant S10OD028685. T.N.S. is a Howard Hughes Medical Institute Fellow of the Damon Runyon Cancer Research Foundation (DRG-2381-19). J.D.B. is an Investigator of the Howard Hughes Medical Institute. H.Y.C. is also funded by an Emergent Ventures Award. The content is solely the responsibility of the authors and does not necessarily represent the official views of the US government or the other sponsors.

Author contributions

A.J.G., A.S., and J.D.B. conceptualized and designed the study. A.J.G. and T.N.S. developed the yeast-display deep mutational scanning antibody-escape mapping method. A.J.G. performed the serum-escape mapping. T.N.S. and A.J.G. performed the deep mutational scanning experiments for ACE2 binding and RBD expression. A.N.L. designed the B.1.351 spike plasmid. R.T.E. and A.N.L. cloned spike point mutants and generated pseudotyped lentiviral particles. R.T.E. and A.J.G. performed pseudotyped neutralization assays. A.J.G., T.N.S., and J.D.B. wrote the code and performed the formal analysis. H.Y.C., J.K.L., K.K., F.K., S.C., and A.S. collected and provided the convalescent plasma samples. J.E.B., D.C., and D.V. provided the B.1.351 spike and RBD for ELISA experiments. A.J.G. and J.D.B. wrote the original draft and all authors reviewed and edited the manuscript.

Declaration of interests

J.D.B. consults for Moderna and Flagship Labs 77 on topics related to viral evolution, and has the potential to receive a share of IP revenue as an inventor on a Fred Hutch optioned technology/patent (application WO2020006494) related to deep mutational scanning of viral proteins. 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. D.V. is named as an inventor on a patent application filed by the University of Washington related to SARS-CoV-2 vaccines and has received an unrelated sponsored research agreement from Vir

Biotechnology Inc. D.C. is an employee of and may hold shares in Vir Biotechnology. The other authors declare no competing interests.

8.7 Supplementary Figures

8.8 Supplementary Data Files

- **Data file S1**. Neutralization titers for B.1.351 and early 2020 infection-elicited sera before and after depletion of homologous RBD-binding antibodies: `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/experimental_data/results/rbd_depletion_neuts/RBD_depletion_NT50_b1351_haarvi.csv`

- **Data file S2**. The effects of all single amino-acid mutations in the B.1.351 RBD on ACE2 binding and RBD expression: `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/data/final_variant_scores.csv`

- **Data file S3**. Plasma-escape scores for B.1.351 plasmas against the B.1.351 RBD deep mutational scanning library: `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/results/supp_data/B1351_raw_data.csv`

- **Data file S4**. Plasma-escape scores for early 2020 plasmas against the Wuhan-Hu-1 RBD deep mutational scanning library: `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/results/prior_DMS_data/early2020_escape_fracs.csv`

- **Data file S5**. Neutralization titers of early 2020 and B.1.351 plasmas against spike-pseudotyped lentiviral particles in the homologous spike background: `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/experimental_data/results/neut_titers/neut_titers.csv`

Figure 8.6: **Enzyme-linked immunosorbent assay (ELISA) and neutralization curves of B.1.351 convalescent plasmas before and after depletion of B.1.351 RBD-binding antibodies**. **(A)** Controls showing that the RBD antibody depletion completely removes a RBD-targeting neutralizing antibody. Effect of two rounds of RBD antibody depletion on binding to B.1.351 RBD and spike (left) and neutralization of B.1.351 spike-pseudotyped lentiviral particles (right) by synthetic serum. The synthetic serum was made by adding the RBD-targeting antibody REGN10987 that binds both Wuhan-Hu-1 and B.1.351 RBD [129] to pre-pandemic pooled serum at 50 μg/mL. The x-axis indicates the antibody concentration (μg/mL), and the y-axis is the optical density at wavelength 450 (OD450) reading at each dilution (left) or fraction infectivity (right). **(B)** Binding of B.1.351 convalescent plasmas to B.1.351 RBD and spike for mock depletion (gray lines) and depletion of RBD-binding antibodies (orange lines). Some samples were depleted three times (dashed lines and open circles) if two rounds of depletion did not abrogate binding to RBD. There were not substantial reductions in OD450 after the third round of depletions, so we reasoned that the samples were maximally depleted, and no further rounds of depletions were performed. Removal of RBD-binding antibodies only modestly reduces spike binding, consistent with prior findings that the majority of anti-spike antibodies do not bind the RBD [33, 120, 149, 255, 267]. **(C)** Neutralization curves for plasma mock depletion (gray circles) and depletion of RBD-binding antibodies (orange triangles). Each assay was performed in technical duplicate, and points show the mean and standard error of the replicates. Pre-pandemic pooled serum was included in (B) and (C) as a negative control for binding and neutralization. RBD-binding antibodies were removed from the plasma using streptavidin magnetic beads conjugated to biotinylated B.1.351 RBD. All binding assays were performed with B.1.351 RBD and spike, and all neutralization assays were performed with B.1.351 spike-pseudotyped lentiviral particles. This figure shows the underlying measurements for all of the B.1.351 plasmas in Fig. 8.2; the underlying measurements for the early 2020 plasmas in Fig. 8.2 are shown in [120].

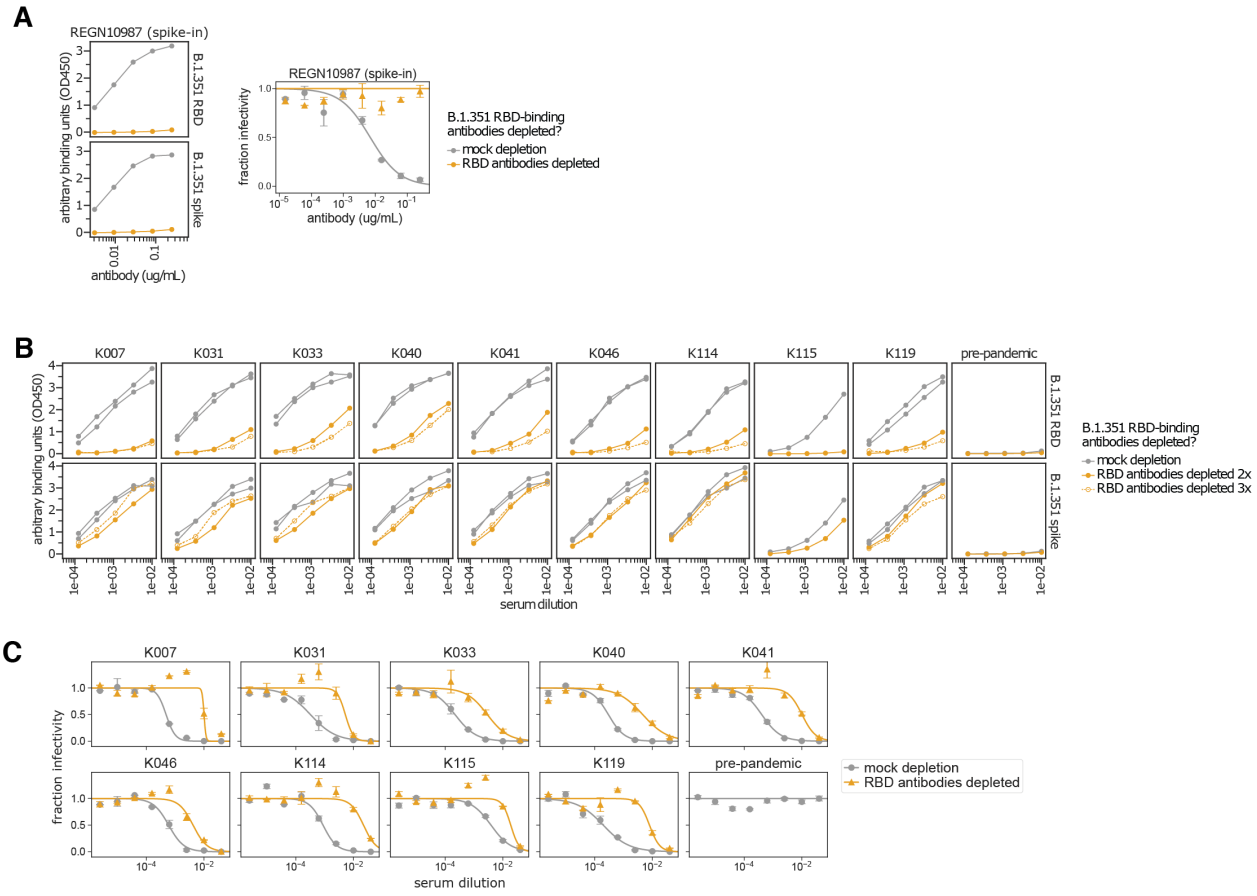
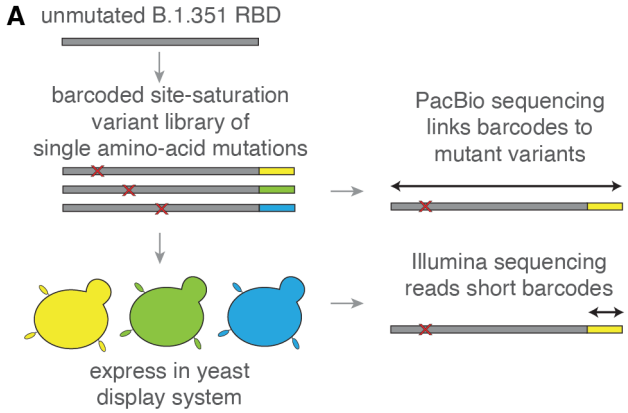
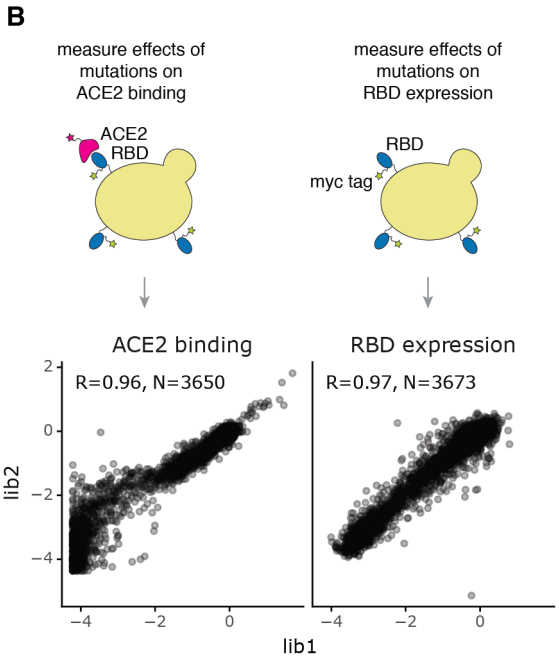

A
REGN10987 (spike-in)
arbitrary binding units (OD450)
B.1.351 RBD
B.1.351 spike
antibody (ug/mL)
REGN10987 (spike-in)
fraction infectivity
antibody (ug/mL)
B.1.351 RBD-binding antibodies depleted?
mock depletion
RBD antibodies depleted
B
K007
K031
K033
K040
K041
K046
K114
K115
K119
pre-pandemic
arbitrary binding units (OD450)
B.1.351 RBD
B.1.351 spike
serum dilution
B.1.351 RBD-binding antibodies depleted?
mock depletion
RBD antibodies depleted 2x
RBD antibodies depleted 3x
C
K007
K031
K033
K040
K041
K046
K114
K115
K119
pre-pandemic
fraction infectivity
serum dilution
mock depletion
RBD antibodies depleted

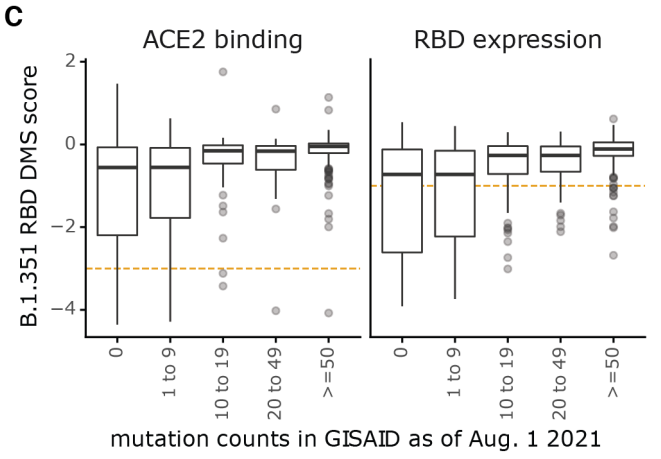
Figure 8.7: **Generation of the B.1.351 RBD mutant libraries and measurements of effects of mutations on ACE2 binding and RBD expression**. **(A)** Schematic showing the B.1.351 RBD mutant library design. A site-saturation variant library was generated in the B.1.351 RBD background, targeting one amino-acid mutation per variant. Nx16 unique DNA barcodes were added to the variant gene fragments. The Nx16 barcodes were linked to their associated RBD mutations by PacBio circular consensus sequencing (CCS). The plasmid library DNA was transformed into yeast cells. In downstream experiments, the Nx16 barcodes are sequenced by short-read Illumina sequencing. The tables at right indicate key library statistics. **(B)** Correlations between biological independent replicate library measurements of the effects of single mutations on ACE2 binding and RBD expression, measured as described in [280]. See Methods for experimental details. **(C)** Thresholds on the ACE2 binding and RBD expression scores (dashed orange lines) for the B.1.351 mutant library to computationally filter highly deleterious variants that may represent spurious antibody-escape mutations. Importantly, we aimed to retain most mutations that have been observed $\geq 50\times$ in sequenced SARS-CoV-2 isolates. The x-axis categorizes mutations by their number of observations in GISAID [91] as of Aug. 1, 2021. An ACE2 binding score threshold of $\geq 50 - 3.0$ (1,000-fold loss in binding affinity) and an RBD expression score of ≥ -1.0 (10-fold loss in RBD expression) were chosen, which filter comparable numbers of mutations as in prior Wuhan-Hu-1 experiments [120, 280]. These filters retain 99.4 and 93.8% of mutations, respectively, that have been observed ≥ 50 times in sequenced SARS-CoV-2 isolates. **(D)** Relationship between the ACE2 binding and RBD expression scores for the B.1.351 RBD library compared to those previously published for the Wuhan-Hu-1 library [280]. The computational filters used for antibody-escape experiments for the Wuhan-Hu-1 [120] and B.1.351 libraries are dashed orange lines. Each dot is one mutation, and mutations to disulfide bonds are shown in red. A key difference is that for the previously published Wuhan-Hu-1 experiments, dimeric rather than monomeric ACE2 was used [280].



library	variant class	n barcodes	% of library
lib1 (44,115 barcodes)	wildtype	11,416	25.9
	1 nonsynonymous	26,837	60.8
	>1 nonsynonymous	5,770	13.1
	stop	92	0.2
lib2 (38,901 barcodes)	wildtype	10,131	26.0
	1 nonsynonymous	23,567	60.6
	>1 nonsynonymous	5,103	13.1
	stop	100	0.3

library	n barcodes	amino-acid mutations missing	mutations with >=1 barcode	mutation with >=2 barcodes	% coverage of possible mutations
1	44,115	50	3,769	3,656	98.7
2	38,901	68	3,751	3,604	98.2
both	83,016	12	3,807	3,783	99.7





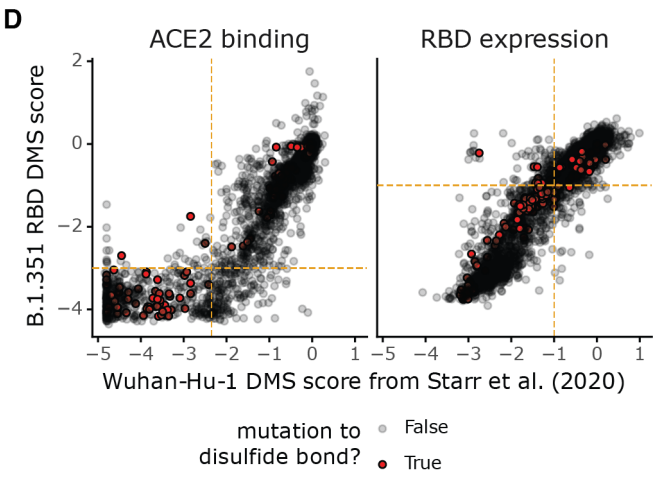
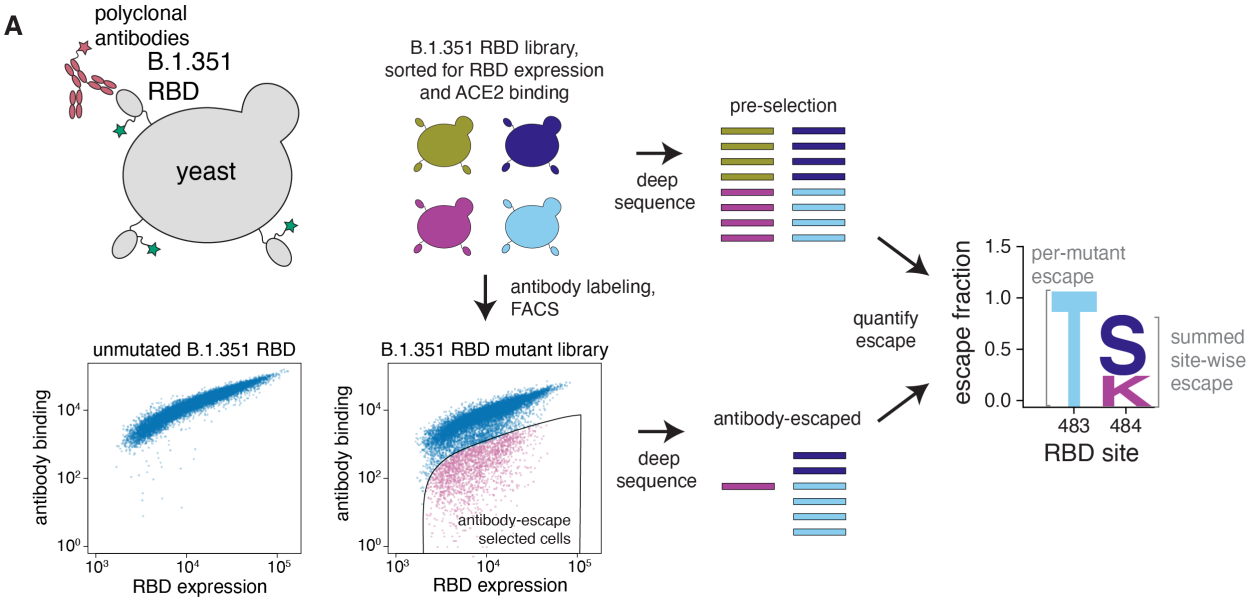
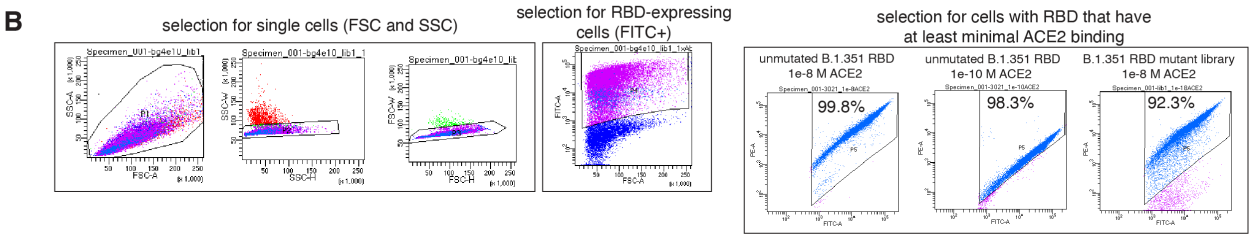
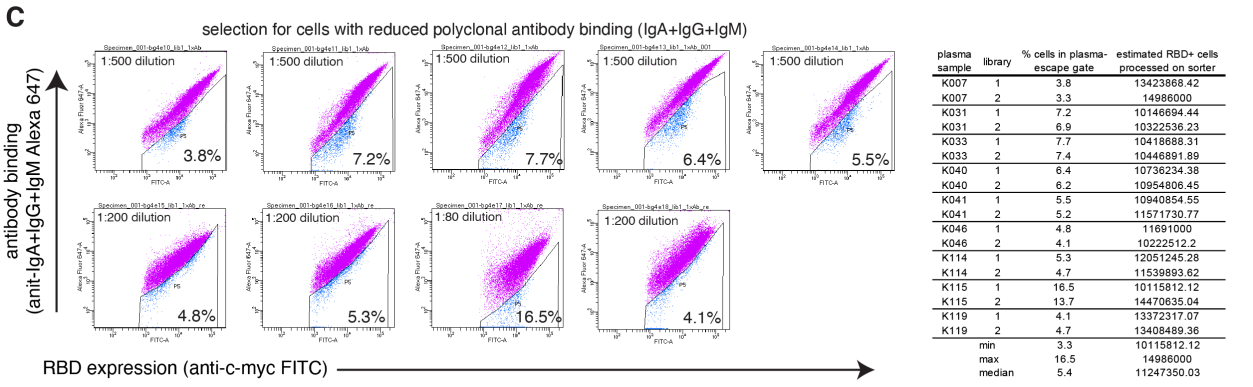


Figure 8.8: **Deep mutational scanning approach to map mutations that reduce binding of B.1.351 infection-elicited polyclonal plasma antibodies to the B.1.351 RBD**. **(A)** Schematic of the approach. The RBD is expressed on the surface of yeast (top left). Flow cytometry is used to quantify both RBD expression (via a C-terminal MYC tag, green star) and antibody binding to the RBD protein expressed on the surface of each yeast cell (bottom left). A library of yeast expressing B.1.351 RBD mutants was incubated with convalescent plasmas and fluorescence-activated cell sorting (FACS) was used to enrich for cells expressing RBD that bound reduced amounts of plasma antibodies, as detected using an IgA+IgG+IgM secondary antibody. Deep sequencing was used to quantify the frequency of each mutation in the initial and antibody-escape cell populations. We quantified the effect of each mutation as the escape fraction, which represents the fraction of cells expressing RBD with that mutation that fell in the antibody escape FACS bin. Escape fractions are represented in logo plots, with the height of each letter proportional to the effect of that amino-acid mutation on antibody binding. The site-level escape metric is the sum of the escape fractions of all mutations at a site. Experimental and computational filtering were used to remove RBD mutants that were misfolded or unable to bind the ACE2 receptor. **(B)** Left: Representative plots of nested FACS gating strategy used for all plasma selection experiments to select for single cells. Samples were gated by SSC-A versus FSC-A, SSC-W versus SSC-H, and FSC-W versus FSC-H) that also express RBD (FITC-A vs. FSC-A). Right: The RBD mutant libraries were sorted to retain cells expressing variants that bound to ACE2 with at least nominal affinity. Unmutated B.1.351 RBD and each RBD mutant library was incubated with dimeric ACE2 at 10^{-8} M. A FACS selection gate was set to capture 98% of cells expressing unmutated B.1.351 RBD that were incubated with 10^{-10} M ACE2, to purge the mutant libraries of highly deleterious mutations (i.e., those that have $< 1\%$ the affinity of unmutated B.1.351 RBD). **(C)** Left: FACS gating strategy for one of two independent libraries to select cells expressing RBD mutants with reduced binding by polyclonal sera (cells in blue). Gates were set manually during sorting. Selection gates were set to capture approximately 5% of the RBD+ library. The same gate was set for both independent libraries stained with each plasma sample, and the FACS scatter plots looked qualitatively similar between the two libraries. Right: the fraction of library cells that fall into each selection gate. **(D)** Mutation- and site-level correlations of escape scores between biologically independent library replicates. SSC-A, side scatter-area; FSC-A, forward scatter-area; SSC-W, side scatter-width; SSC-H, side scatter-height; FSC-W, forward scatter-width; FSC-H, forward scatter height; FITC-A, fluorescein isothiocyanate-area.







plasma sample	library	% cells in plasma-escape gate	estimated RBD+ cells processed on sorter
K007	1	3.8	13423868.42
K007	2	3.3	14986000
K031	1	7.2	10146694.44
K031	2	6.9	10322536.23
K033	1	7.7	10418688.31
K033	2	7.4	10446891.89
K040	1	6.4	10736234.38
K040	2	6.2	10954806.45
K041	1	5.5	10940854.55
K041	2	5.2	11571730.77
K046	1	4.8	11691000
K046	2	4.1	10222512.2
K114	1	5.3	12051245.28
K114	2	4.7	11539893.62
K115	1	16.5	10115812.12
K115	2	13.7	14470635.04
K119	1	4.1	13372317.07
K119	2	4.7	13408489.36
	min	3.3	10115812.12
	max	16.5	14986000
	median	5.4	11247350.03

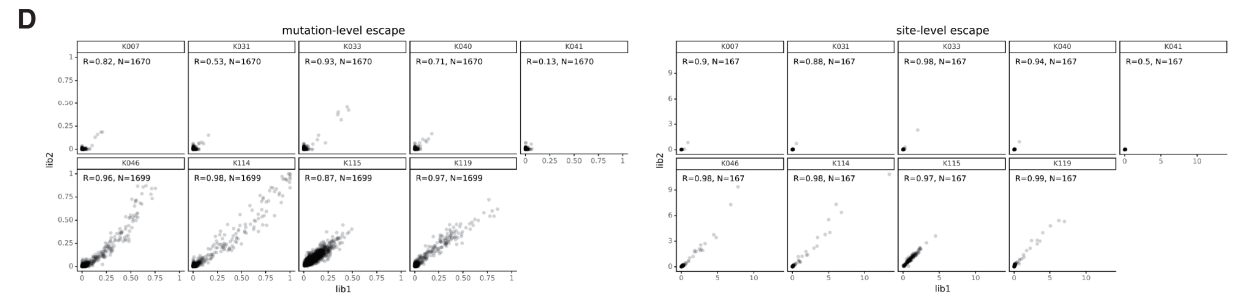


Figure 8.9: **Escape maps for the early 2020 convalescent plasmas, as measured using a deep mutational scanning approach in the Wuhan-Hu-1 RBD background**. The line plots at left indicate the site-level antibody escape for all RBD sites, and the logo plots at right zoom in on key sites (highlighted in purple on the line plot x-axes). For each sample, the y-axis is scaled independently. RBD sites are colored by antibody epitope. Sites 417, 484, and 501 are labeled with red text on the x-axis. All 11 samples from the Washington state early 2020 cohort [120] are shown here and averaged in Fig. 8.4. Interactive versions of logo plots and structural visualizations are at `https://jbloomlab.github.io/SARS-CoV-2-RBD_B.1.351/`. These data were originally published in [120] and are reanalyzed here. The numerical antibody-escape scores are in data file S4 and at `https://github.com/jbloomlab/SARS-CoV-2-RBD_B.1.351/blob/main/results/prior_DMS_data/early2020_escape_fracs.csv`.

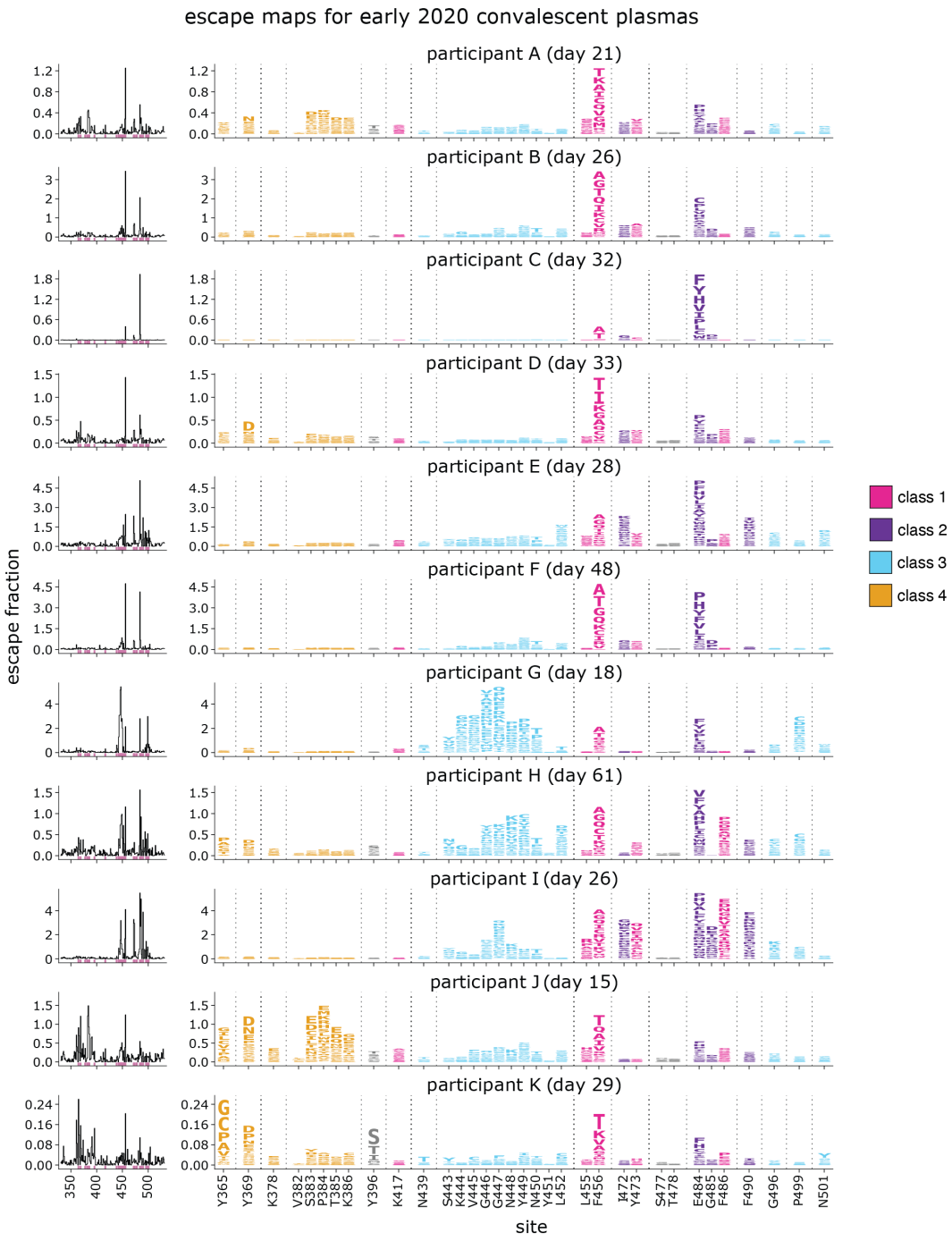
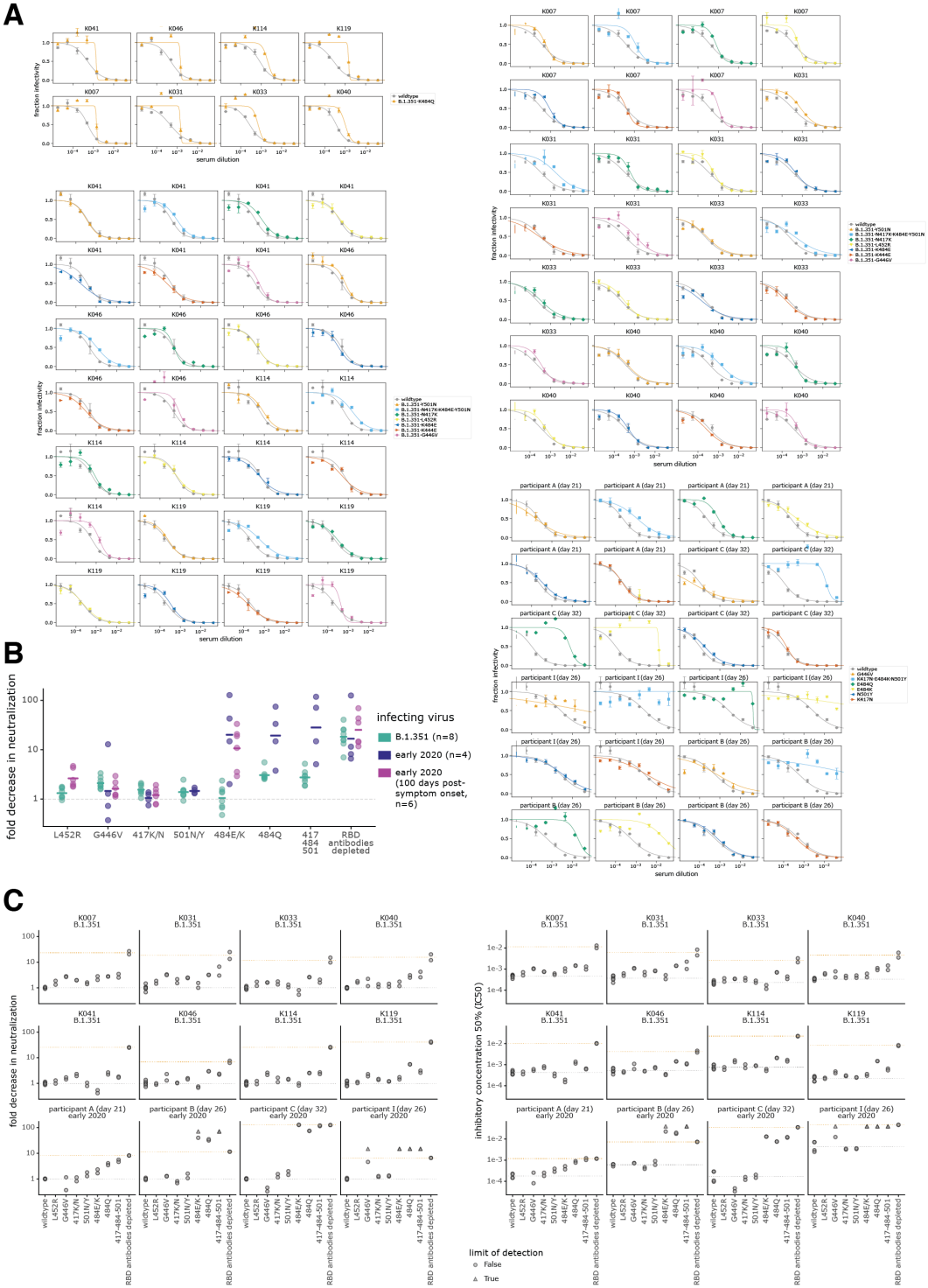

escape maps for early 2020 convalescent plasmas
participant A (day 21)
participant B (day 26)
participant C (day 32)
participant D (day 33)
participant E (day 28)
participant F (day 48)
participant G (day 18)
participant H (day 61)
participant I (day 26)
participant J (day 15)
participant K (day 29)
escape fraction
site
class 1
class 2
class 3
class 4

Figure 8.10: **Neutralization of point mutants of B.1.351 and D614G spike-pseudotyped lentiviral particles by convalescent plasmas from B.1.351 and early 2020-infected individuals**. **(A)** Each plot shows the neutralization curves of one point mutant and the wildtype measured on the same assay date for each plasma (the same wildtype curve is repeated on multiple plots for comparison). Plots are grouped by assay date. Each point is the average of two technical replicates. **(B)** The fold-decrease in neutralization for samples shown in Fig. 8.5, with the addition of previously measured neutralization by samples from 6 early 2020 convalescent individuals collected approximately 100 days post-symptom onset [121]. **(C)** The fold-change in IC50 (left) or the absolute IC50 (right) for the neutralization of each point mutant by each plasma. The fold-change IC50 is calculated relative to the geometric mean of two wildtype technical replicates performed on the same assay date. Each point is one technical replicate. The dashed gray line indicates the geometric mean of all wildtype measurements for that plasma, and the orange line indicates the geometric mean of the effect of removing all RBD-binding antibodies. B.1.351 plasma names are prefixed with K*, and early 2020 plasmas are prefixed with "participant". All assays were performed with the "homologous" virus: B.1.351 spike for B.1.351 plasmas, and D614G spike for early 2020 plasmas. Mutations are given the same names for B.1.351 and D614G spikes, so 417K/N is 417N in the B.1.351 background and 417K in the D614G background; 484E/K is 484E in B.1.351 and 484K in D614G; 501N/Y is 501N in B.1.351 and 501Y in D614G; and 417-484-501 is 417K-484E-501N in B.1.351 and 417N-484K-501Y in D614G.


A
B
C
fold decrease in neutralization
infecting virus
B.1.351 (n=8)
early 2020 (n=4)
early 2020 (100 days post-symptom onset, n=6)
L452R
G446V
417K/N
501N/Y
484E/K
484Q
417 484 501
RBD antibodies depleted
serum dilution
fraction infectivity
inhibitory concentration 50% (IC50)
limit of detection
False
True

Chapter 9

CONCLUSIONS

The SARS-CoV-2 virus has been circulating among humans for less than two years, and yet even in this brief period of time, we have seen that SARS-CoV-2 has the potential to evolve antigenically. Immune pressure is likely to become a major driving force of the virus' evolution as larger fractions of the population gain immunity through infection or vaccination. In my dissertation work, I have developed a system to comprehensively measure the effects of all possible mutations to the RBD on binding of monoclonal antibodies and polyclonal serum antibodies from individuals previously infected with or vaccinated against SARS-CoV-2. Here I review some of the major findings of my work and potential future avenues of research.

9.1 Summary

We developed a method to comprehensively identify mutations that reduce antibody binding to the RBD [124]. There are two key features of our approach to study the antigenic effects of mutations to the RBD: (1) it is *complete*, and (2) it is *prospective*. Unlike traditional selection experiments that only identify a handful of the possible escape mutations, our method completely maps mutations that escape antibody binding. Moreover, our results are also of utility for assessing if ongoing viral evolution is likely to be of antigenic consequence. Beginning in late 2020, multiple SARS-CoV-2 variant lineages emerged with mutations in the RBD [285, 94]. It can take weeks to clone and test point mutants when a new mutations arises. Our escape maps enable immediate assessment of whether mutations to the RBD alter antigenicity. These quantitative experimental data can be cross-referenced to mutations

observed during genomic surveillance of circulating SARS-CoV-2 strains. To facilitate this rapid assessment of the antigenic consequences of mutations, we have created an interactive visualization of all of our of antibody-escape mapping, for dozens of antibodies and > 60 plasmas or sera to date at: `https://jbloomlab.github.io/SARS2_RBD_Ab_escape_maps/`.

These methods have immediate practical applications. We used this method to map mutations that reduce binding of monoclonal antibodies, including those in clinical use [124, 122, 279, 278]. We found that mutations present in circulating SARS-CoV-2 variants can escape binding of some of the therapeutic antibodies in clinical use. For example, we showed that the E484K mutation present in the Beta, Gamma, Mu, and Iota variants, and the L452R mutation in the Delta, Kappa, and Epsilon variants each escape binding of LY-CoV555 (bamlamivimab) [279]. On the other hand, we found that some mutations found at high frequency, such as N501Y, S477N, and T478K, have little effect on the binding of any tested antibodies, although the N501Y mutation increases ACE2 affinity [280].

We applied this method to map mutations that reduce binding of convalescent plasmas [120, 122, 123] and vaccine-elicited sera [121]. We found that mutations to site 484 could strongly reduce the binding and neutralization of convalescent plasmas from individuals infected with SARS-CoV-2 in early 2020 [120, 122]. This demonstrates that the infection-elicited neutralizing antibody response is highly focused on one immunodominant epitope. This implies that single viral mutations may substantially erode infection-elicited immunity, and could be a concern for vaccine-elicited immunity, too [358]. We found that the vaccine-elicited response, however, has broader binding of antibodies across the RBD and is relatively less affected by single mutations than infection-elicited polyclonal plasmas [121]. Single viral mutations present in circulating SARS-CoV-2 viral variants can still reduce antibody binding and neutralization, although usually to a lesser degree than is seen for infection-elicited antibodies [121].

Finally, we knew that SARS-CoV-2 would continue to evolve, and that eventually, a

new viral variant would become the dominant circulating strain. It was unknown how the specificity of the antibody response elicited by these variants would be different. Thus, we extended our deep mutational scanning approach to map mutations that reduce binding and neutralization of polyclonal plasmas from individuals who were infected with SARS-CoV-2 variants. To do this, we created a new deep mutational scanning library in the B.1.351 variant background and identified mutations in this background that reduced binding of plasmas from B.1.351-convalescent individuals. We found a shifted immunodominance hierarchy, with a relatively greater importance of the class 3 epitope than for early 2020 plasmas [123]. These results provided mechanistic insight to the finding that different variants induce antibody responses that differentially neutralize other variants [42, 211, 43, 187]. For instance, B.1.351-elicited plasmas have a modest decrease in neutralization against D614G or early 2020 viruses, but are substantially less able to neutralize Delta viruses [43, 187]. These results might be explained by differential utilization of antibody germline genes in response to infection with different SARS-CoV-2 variants [210].

9.2 Implications for vaccine development and selection

Influenza virus evolves rapidly on a global scale, requiring frequent updates of the vaccine to incorporate strains that are predicted to be circulating at high frequency in the coming flu season [47]. Current evidence suggests that SARS-CoV-2 evolution will eventually erode vaccine-elicited immunity such that a vaccine update will be required. The data that we have generated with our antibody- and plasma-escape maps can help to interpret the antigenic consequences of SARS-CoV-2 evolution and can be useful in the prioritization of which mutations and variants need to be further investigated using lower-throughput methods such as neutralization assays using pseudotyped virus or live virus system.

Neutralizing antibodies are a correlate of protection against SARS-CoV-2 infection [6, 195, 155]. Because most neutralizing antibodies target the spike RBD [229, 282, 74, 120],

RBD-only vaccine candidates have also been pursued [216, 342, 304, 54] and may be reasonable vaccine targets. But, there has been concern that immunity elicited by an RBD-only vaccine would be more susceptible to erosion by viral evolution because of the smaller number of potential epitopes contained within the RBD compared to the full spike protein. I established that vaccination with the mRNA-1273 vaccine elicits neutralizing antibodies that are even more focused towards the RBD than the infection-elicited response [121]. Moreover, it is presumed that the majority of the residual neutralizing activity that is not directed against the RBD is directed towards the NTD. But all SARS-CoV-2 viral variants of interest or concern that have emerged since late 2020 have contained mutations in the NTD, often in the "antigenic supersite" [201, 200, 199, 48, 283, 202, 207, 337, 154]. From this persepctive, an RBD-only vaccine may be a reasonable immunogen. Indeed, RBD nanoparticle vaccine candidates have been shown to induce broadly protective antibody responses [304, 305].

9.3 Future directions

SARS-CoV-2 continues to evolve, new variants are emerging, and individuals are acquiring increasingly complex infection and vaccination histories. Thus, there are many open questions regarding the antigenic consequences of viral evolution that can be addressed with the methods I have described in the preceding chapters.

9.3.1 Prospective identification of mutations that may define an antigenically drifted descendant of the Delta variant

In my work on the immune response to the B.1.351 variant, we found that this variant induces antibody responses with different immunodominance hierarchies than early SARS-CoV-2 viral isolates. Such changes in immunodominance can have important consequences, as they can contribute to individuals with different exposure histories having different susceptibilities to viral mutants [128, 329]. Although the changes in immunodominance we have observed

here are relatively modest, they could become larger as the virus continues to evolve and different individuals accumulate increasingly disparate exposure histories through infection and vaccination [107, 358].

The Delta (B.1.617.2) variant emerged in mid-2021 [239] and has reached near-fixation in the United States and in many countries throughout the world [138]. The Delta variant has multiple mutations in spike (T19R, E156G, del157–158, L452R, T478K, D614G, P681R, and D950N), and is moderately less susceptible to neutralization by convalescent and vaccine-elicited plasmas [187, 233, 201, 209]. While we cannot be sure that the Delta variant will fix, it seems likely at the moment. If Delta does fix, then the next major variant will be a descendant of Delta. Indeed, sublineages of Delta, sometimes referred to as "Delta+" lineages, have already been described [240]. As variants in the Delta background emerge, their potential for further erosion of population immunity will need to be assessed. Thus, I am extending the methods we have developed to map mutations that reduce binding of polyclonal plasmas to the Delta RBD.

To this end, I am making a deep mutational scanning library in the Delta background and measuring the effects of mutations on ACE2 binding, RBD expression, and antibody binding. I detail our interest in characterizing the antibody response further in the next section.

9.3.2 The influence of SARS-CoV-2 variant infection and vaccination on the specificity antibody response

As SARS-CoV-2 continues to evolve, individuals will gain increasingly complex and potentially disparate immune histories, shaped by serial vaccinations and infections. As SARS-CoV-2 continues to circulate for years or even decades, the effects of immune imprinting or original antigenic sin [102, 316, 35] may start to interact with variant-specific immunodominance hierarchies.

In my ongoing work, I will examine the specificities of the antibody response from individuals who have been vaccinated against the Wuhan-Hu-1 strain, infected with the Delta variant, or vaccinated and then infected with the Delta variant. This will begin to address the role of immune imprinting in the specificity of the immune response. Will individuals who were first vaccinated against the Wuhan-Hu-1 strain and then infected with the Delta variant (which may have a different immunodominance hierarchy) have responses that are more Wuhan-Hu-1-like, or more Delta-like? Preliminary evidence from a SARS-CoV-2 B.1.351 variant booster mRNA vaccine trial suggest that immune imprinting may lead to boosting of the primary antibody response rather than the development of a *de novo* variant response [58].

One way to potentially circumvent the challenges of immune imprinting is to ensure that individual's first antigenic exposure is to a "universal" antigen that elicits a response that is protective against many antigenically distinct variants. While it is too late to rewrite individuals' first antigenic exposures, it might be possible to vaccinate future generations with such a "universal" antigen. The development of a universal flu vaccine has been an NIH/NIAID priority for several years [92]. In the coming years, there may also be increasing interest in the development of a universal SARS-CoV-2 that would be protective against many SARS-CoV-2 variants, or even a universal sarbecovirus vaccine that could protect against pre-emergent viruses of pandemic potential.

9.3.3 Examining the effects of combinations of mutations to the SARS-CoV-2 RBD to define the complete immunodominance hierarchy induced by infection or vaccination

A final direction that I am pursuing is understanding the effects of combinations of mutations on antibody binding. Our work thus far only examines the effects of single mutants, as mutations typically fix in a stepwise manner [276], and processes that require multiple mutations tend to occur more slowly in viral evolution [104]. However, as the SARS-CoV-2

virus continues to circulate in humans, multiple mutations will eventually fix, and so the examination of the antigenic effects of multiple mutations is an important area for future work.

Indeed, SARS-CoV-2 viral variants have emerged with multiple mutations to the RBD, including the Beta, Gamma, Delta, Kappa, Lambda, Mu, and Omicron variants [240, 285, 94]. There is evidence that some mutations in the RBD can act relatively additively to affect neutralization [313, 51, 193, 299], there is additional evidence that some mutations such as K417N and L452R may act in a synergistic manner [201], whereas other mutations that are in partially overlapping epitopes, such as L452R and E484Q may have less-than-additive effects [99]. Thus, a better mechanistic understanding of the combinatorial effects of mutations on antibody binding and neutralization are needed [90].

It is not possible to assay the effects of all possible combinations of mutations to the RBD. Even though the RBD is only 201 amino acids in length, there are over 16 million possible combinations of double mutants, far beyond the feasible scale of our experimental system. But, we can reduce the complexity of a combinatorial library by focusing on sites that are known to be of antigenic importance, based on our prior mapping of escape mutations for monoclonal antibodies and polyclonal sera. We can even further reduce the complexity by making only specific mutations at each of these key antigenic sites, rather than all 19 possible mutations, allowing us to assay higher-order mutations. For instance, the complexity of a library with only 1 mutation at each of 15 sites would be approximately 30,000 variants, only 10-fold higher than our current library diversity size.

9.4 Final thoughts

My dissertation work has specifically leveraged a yeast-displayed deep mutational scanning system to measure the effects of mutations to the SARS-CoV-2 spike RBD on its function and antigenicity. This work has been remarkably useful in the context of the COVID-

19 pandemic and has helped to anticipate the ACE2 binding and antigenic phenotypes of new variants as they have emerged. More broadly, however, future studies may apply high throughput approaches to study other SARS-CoV-2 proteins or domains, such as the spike protein, N-terminal domain, or protease, as well as other viruses. Comprehensive genotype-to-phenotype maps for other viruses may similarly aid in the interpretation of viral surveillance of human endemic viruses and viruses of pandemic potential.

BIBLIOGRAPHY

[1] Novavax COVID-19 vaccine demonstrates 89.3% efficacy in UK phase 3 trial. https://ir.novavax.com/news-releases/news-release-details/novavax-covid-19-vaccine-demonstrates-893-efficacy-uk-phase-3. Accessed: 2021-4-8.

[2] Pfizer and BioNTech confirm high efficacy and no serious safety concerns through up to six months following second dose in updated topline analysis of landmark COVID-19 vaccine study. https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-confirm-high-efficacy-and-no-serious. Accessed: 2021-4-8.

[3] Regeneron's casirivimab and imdevimab antibody cocktail for COVID-19 is first combination therapy to receive FDA emergency use authorization. https://investor.regeneron.com/news-releases/news-release-details/regenerons-regen-cov2-first-antibody-cocktail-covid-19-receive/. Accessed: 2021-10-19.

[4] A study of LY3819253 (LY-CoV555) and LY3832479 (LY-CoV016) in preventing SARS-CoV-2 infection and COVID-19 in nursing home residents and staff - full text view - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/show/NCT04497987. Accessed: 2021-10-19.

[5] Rhys M Adams, Thierry Mora, Aleksandra M Walczak, and Justin B Kinney. Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife*, 5, 30 December 2016.

[6] Amin Addetia, Katharine H D Crawford, Adam Dingens, Haiying Zhu, Pavitra Roychoudhury, Meei-Li Huang, Keith R Jerome, Jesse D Bloom, and Alexander L Greninger. Neutralizing antibodies correlate with protection from SARS-CoV-2 in humans during a fishery vessel outbreak with a high attack rate. *Journal of clinical microbiology*, 58(11), 21 October 2020.

[7] Wafaa B Alsoussi, Jackson S Turner, James B Case, Haiyan Zhao, Aaron J Schmitz, Julian Q Zhou, Rita E Chen, Tingting Lei, Amena A Rizk, Katherine M McIntire, Emma S Winkler, Julie M Fox, Natasha M Kafai, Larissa B Thackray, Ahmed O Hassan, Fatima Amanat, Florian Krammer, Corey T Watson, Steven H Kleinstein,

Daved H Fremont, Michael S Diamond, and Ali H Ellebedy. A potently neutralizing antibody protects mice against SARS-CoV-2 infection. *Journal of immunology*, 205(4):915–922, 15 August 2020.

[8] Fatima Amanat, Mahima Thapa, Tinting Lei, Shaza M Sayed Ahmed, Daniel C Adelsberg, Juan Manuel Carreno, Shirin Strohmeier, Aaron J Schmitz, Sarah Zafar, Julian Q Zhou, Willemijn Rijnink, Hala Alshammary, Nicholas Borcherding, Ana Gonzalez Reiche, Komal Srivastava, Emilia Mia Sordillo, Harm van Bakel, Jackson S Turner, Goran Bajic, Viviana Simon, Ali H Ellebedy, Florian Krammer, and The Personalized Virology Initiative. The plasmablast response to SARS-CoV-2 mRNA vaccination is dominated by non-neutralizing antibodies that target both the NTD and the RBD. 9 March 2021.

[9] Alberto A Amarilla, Julian D J Sng, Rhys Parry, Joshua M Deerain, James R Potter, Yin Xiang Setoh, Daniel J Rawle, Thuy T Le, Naphak Modhiran, Xiaohui Wang, Nias Y G Peng, Francisco J Torres, Alyssa Pyke, Jessica J Harrison, Morgan E Freney, Benjamin Liang, Christopher L D McMillan, Stacey T M Cheung, Darwin J Da Costa Guevara, Joshua M Hardy, Mark Bettington, David A Muller, Fasséli Coulibaly, Frederick Moore, Roy A Hall, Paul R Young, Jason M Mackenzie, Jody Hobson-Peters, Andreas Suhrbier, Daniel Watterson, and Alexander A Khromykh. A versatile reverse genetics platform for SARS-CoV-2 and other positive-strand RNA viruses. *Nature communications*, 12(1):3431, 8 June 2021.

[10] El-Ad David Amir, Kara L Davis, Michelle D Tadmor, Erin F Simonds, Jacob H Levine, Sean C Bendall, Daniel K Shenfeld, Smita Krishnaswamy, Garry P Nolan, and Dana Pe'er. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature biotechnology*, 31(6):545–552, June 2013.

[11] Emanuele Andreano, Giulia Piccini, Danilo Licastro, Lorenzo Casalino, Nicole V Johnson, Ida Paciello, Simeone Dal Monego, Elisa Pantano, Noemi Manganaro, Alessandro Manenti, Rachele Manna, Elisa Casa, Inesa Hyseni, Linda Benincasa, Emanuele Montomoli, Rommie E Amaro, Jason S McLellan, and Rino Rappuoli. SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma. *Proceedings of the National Academy of Sciences of the United States of America*, 118(36), 7 September 2021.

[12] Davide Angeletti, James S Gibbs, Matthew Angel, Ivan Kosik, Heather D Hickman, Gregory M Frank, Suman R Das, Adam K Wheatley, Madhu Prabhakaran, David J Leggat, Adrian B McDermott, and Jonathan W Yewdell. Defining B cell immunodominance to viruses. *Nature immunology*, 18(4):456–463, April 2017.

[13] Davide Angeletti and Jonathan W Yewdell. Understanding and manipulating viral immunity: Antibody immunodominance enters center stage. *Trends in immunology*, 39(7):549–561, July 2018.

[14] Medini K Annavajhala, Hiroshi Mohri, Pengfei Wang, Manoj Nair, Jason E Zucker, Zizhang Sheng, Angela Gomez-Simmonds, Anne L Kelley, Maya Tagliavia, Yaoxing Huang, Trevor Bedford, David D Ho, and Anne-Catrin Uhlemann. Emergence and expansion of SARS-CoV-2 b.1.526 after identification in new york. *Nature*, 24 August 2021.

[15] Yinon M Bar-On, Yair Goldberg, Micha Mandel, Omri Bodenheimer, Laurence Freedman, Nir Kalkstein, Barak Mizrahi, Sharon Alroy-Preis, Nachman Ash, Ron Milo, and Amit Huppert. Protection of BNT162b2 vaccine booster against covid-19 in israel. *The New England journal of medicine*, 385(15):1393–1400, 7 October 2021.

[16] Christopher O Barnes, Claudia A Jette, Morgan E Abernathy, Kim-Marie A Dam, Shannon R Esswein, Harry B Gristick, Andrey G Malyutin, Naima G Sharaf, Kathryn E Huey-Tubman, Yu E Lee, Davide F Robbiani, Michel C Nussenzweig, Anthony P West, Jr, and Pamela J Bjorkman. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature*, 588(7839):682–687, December 2020.

[17] Christopher O Barnes, Anthony P West, Jr, Kathryn E Huey-Tubman, Magnus A G Hoffmann, Naima G Sharaf, Pauline R Hoffman, Nicholas Koranda, Harry B Gristick, Christian Gaebler, Frauke Muecksch, Julio C Cetrulo Lorenzi, Shlomo Finkin, Thomas Hägglöf, Arlene Hurley, Katrina G Millard, Yiska Weisblum, Fabian Schmidt, Theodora Hatziioannou, Paul D Bieniasz, Marina Caskey, Davide F Robbiani, Michel C Nussenzweig, and Pamela J Bjorkman. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *Cell*, 182(4):828–842.e16, 20 August 2020.

[18] Alina Baum, Dharani Ajithdoss, Richard Copin, Anbo Zhou, Kathryn Lanza, Nicole Negron, Min Ni, Yi Wei, Kusha Mohammadi, Bret Musser, Gurinder S Atwal, Adelekan Oyejide, Yenny Goez-Gazi, John Dutton, Elizabeth Clemmons, Hilary M Staples, Carmen Bartley, Benjamin Klaffke, Kendra Alfson, Michal Gazi, Olga Gonzalez, Edward Dick, Jr, Ricardo Carrion, Jr, Laurent Pessaint, Maciel Porto, Anthony Cook, Renita Brown, Vaneesha Ali, Jack Greenhouse, Tammy Taylor, Hanne Andersen, Mark G Lewis, Neil Stahl, Andrew J Murphy, George D Yancopoulos, and Christos A Kyratsous. REGN-COV2 antibodies prevent and treat SARS-CoV-2 infection in rhesus macaques and hamsters. *Science*, 370(6520):1110–1115, 27 November 2020.

[19] Alina Baum, Benjamin O Fulton, Elzbieta Wloga, Richard Copin, Kristen E Pascal, Vincenzo Russo, Stephanie Giordano, Kathryn Lanza, Nicole Negron, Min Ni, Yi Wei, Gurinder S Atwal, Andrew J Murphy, Neil Stahl, George D Yancopoulos, and Christos A Kyratsous. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science*, 369(6506):1014–1018, 21 August 2020.

[20] Etienne Becht, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel W H Kwok, Lai Guan Ng, Florent Ginhoux, and Evan W Newell. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature biotechnology*, 3 December 2018.

[21] Michelle M Becker, Rachel L Graham, Eric F Donaldson, Barry Rockx, Amy C Sims, Timothy Sheahan, Raymond J Pickles, Davide Corti, Robert E Johnston, Ralph S Baric, and Mark R Denison. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 105(50):19944–19949, 16 December 2008.

[22] Trevor Bedford, Alexander L Greninger, Pavitra Roychoudhury, Lea M Starita, Michael Famulare, Meei-Li Huang, Arun Nalla, Gregory Pepper, Adam Reinhardt, Hong Xie, Lasata Shrestha, Truong N Nguyen, Amanda Adler, Elisabeth Brandstetter, Shari Cho, Danielle Giroux, Peter D Han, Kairsten Fay, Chris D Frazar, Misja Ilcisin, Kirsten Lacombe, Jover Lee, Anahita Kiavand, Matthew Richardson, Thomas R Sibley, Melissa Truong, Caitlin R Wolf, Deborah A Nickerson, Mark J Rieder, Janet A Englund, Seattle Flu Study Investigators, James Hadfield, Emma B Hodcroft, John Huddleston, Louise H Moncla, Nicola F Müller, Richard A Neher, Xianding Deng, Wei Gu, Scot Federman, Charles Chiu, Jeffrey S Duchin, Romesh Gautom, Geoff Melly, Brian Hiatt, Philip Dykema, Scott Lindquist, Krista Queen, Ying Tao, Anna Uehara, Suxiang Tong, Duncan MacCannell, Gregory L Armstrong, Geoffrey S Baird, Helen Y Chu, Jay Shendure, and Keith R Jerome. Cryptic transmission of SARS-CoV-2 in washington state. *Science*, 370(6516):571–575, 30 October 2020.

[23] Trevor Bedford, Marc A Suchard, Philippe Lemey, Gytis Dudas, Victoria Gregory, Alan J Hay, John W McCauley, Colin A Russell, Derek J Smith, and Andrew Rambaut. Integrating influenza antigenic dynamics with molecular evolution. *eLife*, 3:e01914, 4 February 2014.

[24] Sandrine Belouzard, Victor C Chu, and Gary R Whittaker. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proceedings of the National Academy of Sciences of the United States of America*, 106(14):5871–5876, 7 April 2009.

[25] M J Birrer, S Udem, S Nathenson, and B R Bloom. Antigenic variants of measles virus. *Nature*, 293(5827):67–69, 3 September 1981.

[26] Jesse D Bloom. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Molecular biology and evolution*, 31(8):1956–1978, August 2014.

[27] Jesse D Bloom, Lizhi Ian Gong, and David Baltimore. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science*, 328(5983):1272–1275, 4 June 2010.

[28] E T Boder and K D Wittrup. Yeast surface display for screening combinatorial polypeptide libraries. *Nature biotechnology*, 15(6):553–557, June 1997.

[29] A A Bogan and K S Thorn. Anatomy of hot spots in protein interfaces. *Journal of molecular biology*, 280(1):1–9, 3 July 1998.

[30] Meagan Bolles, Eric Donaldson, and Ralph Baric. SARS-CoV and emergent coronaviruses: viral determinants of interspecies transmission. *Current opinion in virology*, 1(6):624–634, December 2011.

[31] Maciej F Boni, Philippe Lemey, Xiaowei Jiang, Tommy Tsan-Yuk Lam, Blair W Perry, Todd A Castoe, Andrew Rambaut, and David L Robertson. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature microbiology*, 5(11):1408–1417, November 2020.

[32] Berend Jan Bosch, Ruurd van der Zee, Cornelis A M de Haan, and Peter J M Rottier. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *Journal of virology*, 77(16):8801–8811, August 2003.

[33] Philip J M Brouwer, Tom G Caniels, Karlijn van der Straten, Jonne L Snitselaar, Yoann Aldon, Sandhya Bangaru, Jonathan L Torres, Nisreen M A Okba, Mathieu Claireaux, Gius Kerster, Arthur E H Bentlage, Marlies M van Haaren, Denise Guerra, Judith A Burger, Edith E Schermer, Kirsten D Verheul, Niels van der Velde, Alex van der Kooi, Jelle van Schooten, Mariëlle J van Breemen, Tom P L Bijl, Kwinten Sliepen, Aafke Aartse, Ronald Derking, Ilja Bontjer, Neeltje A Kootstra, W Joost Wiersinga, Gestur Vidarsson, Bart L Haagmans, Andrew B Ward, Godelieve J de Bree, Rogier W Sanders, and Marit J van Gils. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science*, 369(6504):643–650, 7 August 2020.

[34] Jonathan C Brown, Daniel H Goldhill, Jie Zhou, Thomas P Peacock, Rebecca Frise, Niluka Goonawardane, Laury Baillon, Ruthiran Kugathasan, Andreia L Pinto, Paul F McKay, Jack Hassard, Maya Moshe, Aran Singanayagam, Thomas Burgoyne, the AT-ACCC Investigators, PHE Virology Consortium, and Wendy S Barclay. Increased transmission of SARS-CoV-2 lineage b.1.1.7 (VOC 2020212/01) is not accounted for by a replicative advantage in primary airway cells or antibody escape. 1 March 2021.

[35] Alicia P Budd, Lauren Beacham, Catherine B Smith, Rebecca J Garten, Carrie Reed, Krista Kniss, Desiree Mustaquim, Farida B Ahmad, Charisse N Cummings, Shikha Garg, Min Z Levine, Alicia M Fry, and Lynnette Brammer. Birth cohort effects in influenza surveillance data: Evidence that first influenza infection affects later Influenza-Associated illness. *The Journal of infectious diseases*, 220(5):820–829, 31 July 2019.

[36] Yongfei Cai, Jun Zhang, Tianshu Xiao, Hanqin Peng, Sarah M Sterling, Richard M Walsh, Jr, Shaun Rawson, Sophia Rits-Volloch, and Bing Chen. Distinct conformational states of SARS-CoV-2 spike protein. *Science*, 369(6511):1586–1592, 25 September 2020.

[37] Heather M Callaway, Kathrin Welsch, Wendy Weichert, Andrew B Allison, Susan L Hafenstein, Kai Huang, Sho Iketani, and Colin R Parrish. Complex and dynamic interactions between parvovirus capsids, transferrin receptors, and antibodies control cell infection and host range. *Journal of virology*, 92(13), 1 July 2018.

[38] Yunlong Cao, Bin Su, Xianghua Guo, Wenjie Sun, Yongqiang Deng, Linlin Bao, Qinyu Zhu, Xu Zhang, Yinghui Zheng, Chenyang Geng, Xiaoran Chai, Runsheng He, Xiaofeng Li, Qi Lv, Hua Zhu, Wei Deng, Yanfeng Xu, Yanjun Wang, Luxin Qiao, Yafang Tan, Liyang Song, Guopeng Wang, Xiaoxia Du, Ning Gao, Jiangning Liu, Junyu Xiao, Xiao-Dong Su, Zongmin Du, Yingmei Feng, Chuan Qin, Chengfeng Qin, Ronghua Jin, and X Sunney Xie. Potent neutralizing antibodies against SARS-CoV-2 identified by High-Throughput Single-Cell sequencing of convalescent patients' B cells. *Cell*, 182(1):73–84.e16, 9 July 2020.

[39] Loic Carrique, Helen M E Duyvesteyn, Tomas Malinauskas, Yuguang Zhao, Jingshan Ren, Daming Zhou, Thomas S Walter, Julika Radecke, Jiandong Huo, Reinis R Ruza, Pranav N M Shah, Elizabeth E Fry, and David I Stuart. The SARS-CoV-2 spike harbours a lipid binding pocket which modulates stability of the prefusion trimer. *bioRxiv*, page 2020.08.13.249177, 13 August 2020.

[40] James Brett Case, Paul W Rothlauf, Rita E Chen, Zhuoming Liu, Haiyan Zhao, Arthur S Kim, Louis-Marie Bloyet, Qiru Zeng, Stephen Tahan, Lindsay Droit, Ma Xe-

nia G Ilagan, Michael A Tartell, Gaya Amarasinghe, Jeffrey P Henderson, Shane Miersch, Mart Ustav, Sachdev Sidhu, Herbert W Virgin, David Wang, Siyuan Ding, Davide Corti, Elitza S Theel, Daved H Fremont, Michael S Diamond, and Sean P J Whelan. Neutralizing antibody and soluble ACE2 inhibition of a Replication-Competent VSV-SARS-CoV-2 and a clinical isolate of SARS-CoV-2. *Cell host & microbe*, 3 July 2020.

[41] Marina Caskey, Florian Klein, Julio C C Lorenzi, Michael S Seaman, Anthony P West, Jr, Noreen Buckley, Gisela Kremer, Lilian Nogueira, Malte Braunschweig, Johannes F Scheid, Joshua A Horwitz, Irina Shimeliovich, Sivan Ben-Avraham, Maggi Witmer-Pack, Martin Platten, Clara Lehmann, Leah A Burke, Thomas Hawthorne, Robert J Gorelick, Bruce D Walker, Tibor Keler, Roy M Gulick, Gerd Fätkenheuer, Sarah J Schlesinger, and Michel C Nussenzweig. Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature*, 522(7557):487–491, 25 June 2015.

[42] Sandile Cele, Inbal Gazy, Laurelle Jackson, Shi-Hsia Hwa, Houriiyah Tegally, Gila Lustig, Jennifer Giandhari, Sureshnee Pillay, Eduan Wilkinson, Yeshnee Naidoo, Farina Karim, Yashica Ganga, Khadija Khan, Mallory Bernstein, Alejandro B Balazs, Bernadett I Gosnell, Willem Hanekom, Mahomed-Yunus S Moosa, Network for Genomic Surveillance in South Africa, COMMIT-KZN Team, Richard J Lessells, Tulio de Oliveira, and Alex Sigal. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature*, 593(7857):142–146, May 2021.

[43] Sandile Cele, Farina Karim, Gila Lustig, San Emmanuel James, Tandile Hermanus, Eduan Wilkinson, Jumari Snyman, Mallory Bernstein, Khadija Khan, Shi-Hsia Hwa, Houriiyah Tegally, Sasha W Tilles, Jennifer Giandhari, Ntombifuthi Mthabela, Matilda Mazibuko, Yashica Ganga, Bernadett I Gosnell, Salim Abdool Karim, Willem Hanekom, Wesley C Van Voorhis, Thumbi Ndung'u, Richard J Lessells, Penny L Moore, Mahomed-Yunus S Moosa, Tulio de Oliveira, Alex Sigal, and COMMIT-KZN Team. Divergence of delta and beta variants and SARS-CoV-2 evolved in prolonged infection into distinct serological phenotypes. 23 September 2021.

[44] Centers for Disease Control and Prevention. Administration overview for johnson & johnson's janssen COVID-19 vaccine. `https://www.cdc.gov/vaccines/covid-19/info-by-product/janssen/index.html`, 10 August 2021. Accessed: 2021-10-22.

[45] Centers for Disease Control and Prevention. Information about the Pfizer-BioNTech COVID-19 vaccine. `https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines/Pfizer-BioNTech.html`, 20 October 2021. Accessed: 2021-10-22.

[46] Centers for Disease Control and Prevention. Moderna COVID-19 vaccine. `https://www.cdc.gov/vaccines/covid-19/info-by-product/moderna/index.html`, 23 August 2021. Accessed: 2021-10-22.

[47] Centers for Disease Control and Prevention. Selecting viruses for the seasonal influenza vaccine. `https://www.cdc.gov/flu/prevent/vaccine-selection.htm`, 31 August 2021. Accessed: 2021-10-24.

[48] Gabriele Cerutti, Yicheng Guo, Tongqing Zhou, Jason Gorman, Myungjin Lee, Micah Rapp, Eswar R Reddem, Jian Yu, Fabiana Bahna, Jude Bimela, Yaoxing Huang, Phinikoula S Katsamba, Lihong Liu, Manoj S Nair, Reda Rawi, Adam S Olia, Pengfei Wang, Baoshan Zhang, Gwo-Yu Chuang, David D Ho, Zizhang Sheng, Peter D Kwong, and Lawrence Shapiro. Potent SARS-CoV-2 neutralizing antibodies directed against spike n-terminal domain target a single supersite. *Cell host & microbe*, 29(5):819–833.e7, 12 May 2021.

[49] Ginger Chao, Wai L Lau, Benjamin J Hackel, Stephen L Sazinsky, Shaun M Lippow, and K Dane Wittrup. Isolating and engineering human antibodies using yeast surface display. *Nature protocols*, 1(2):755–768, 2006.

[50] Albert Tian Chen, Kevin Altschuler, Shing Hei Zhan, Yujia Alina Chan, and Benjamin E Deverman. COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by locations and dates of interest. *eLife*, 10, 23 February 2021.

[51] Rita E Chen, Xianwen Zhang, James Brett Case, Emma S Winkler, Yang Liu, Laura A VanBlargan, Jianying Liu, John M Errico, Xuping Xie, Naveenchandra Suryadevara, Pavlo Gilchuk, Seth J Zost, Stephen Tahan, Lindsay Droit, Jackson S Turner, Wooseob Kim, Aaron J Schmitz, Mahima Thapa, David Wang, Adrianus C M Boon, Rachel M Presti, Jane A O'Halloran, Alfred H J Kim, Parakkal Deepak, Dora Pinto, Daved H Fremont, James E Crowe, Jr, Davide Corti, Herbert W Virgin, Ali H Ellebedy, Pei-Yong Shi, and Michael S Diamond. Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nature medicine*, 27(4):717–726, April 2021.

[52] Shifu Chen, Yanqing Zhou, Yaru Chen, and Jia Gu. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17):i884–i890, 1 September 2018.

[53] Wen-Hsiang Chen, Lanying Du, Shivali M Chag, Cuiqing Ma, Nancy Tricoche, Xinrong Tao, Christopher A Seid, Elissa M Hudspeth, Sara Lustigman, Chien-Te K Tseng,

Maria Elena Bottazzi, Peter J Hotez, Bin Zhan, and Shibo Jiang. Yeast-expressed recombinant protein of the receptor-binding domain in SARS-CoV spike protein with deglycosylated forms as a SARS vaccine candidate. *Human vaccines & immunotherapeutics*, 10(3):648–658, 2014.

[54] Wen-Hsiang Chen, Xinrong Tao, Anurodh Shankar Agrawal, Abdullah Algaissi, Bi-Hung Peng, Jeroen Pollet, Ulrich Strych, Maria Elena Bottazzi, Peter J Hotez, Sara Lustigman, Lanying Du, Shibo Jiang, and Chien-Te K Tseng. Yeast-expressed SARS-CoV recombinant receptor-binding domain (RBD219-N1) formulated with aluminum hydroxide induces protective immunity and reduces immune enhancement. *Vaccine*, 38(47):7533–7541, 3 November 2020.

[55] Xuemin Chen, Christina A Rostad, Larry J Anderson, He-Ying Sun, Stacey A Lapp, Kathy Stephens, Laila Hussaini, Theda Gibson, Nadine Rouphael, and Evan J Anderson. The development and kinetics of functional antibody-dependent cell-mediated cytotoxicity (ADCC) to SARS-CoV-2 spike protein. *Virology*, 559:1–9, July 2021.

[56] Xiangyang Chi, Renhong Yan, Jun Zhang, Guanying Zhang, Yuanyuan Zhang, Meng Hao, Zhe Zhang, Pengfei Fan, Yunzhu Dong, Yilong Yang, Zhengshan Chen, Yingying Guo, Jinlong Zhang, Yaning Li, Xiaohong Song, Yi Chen, Lu Xia, Ling Fu, Lihua Hou, Junjie Xu, Changming Yu, Jianmin Li, Qiang Zhou, and Wei Chen. A neutralizing human antibody binds to the n-terminal domain of the spike protein of SARS-CoV-2. *Science*, 369(6504):650–655, 7 August 2020.

[57] Alice Cho, Frauke Muecksch, Dennis Schaefer-Babajew, Zijun Wang, Shlomo Finkin, Christian Gaebler, Victor Ramos, Melissa Cipolla, Marianna Agudelo, Eva Bednarski, Justin DaSilva, Irina Shimeliovich, Juan Dizon, Mridushi Daga, Katrina Millard, Martina Turroja, Fabian Schmidt, Fengwen Zhang, Tarek Ben Tanfous, Mila Jankovic, Thiago Y Oliveria, Anna Gazumyan, Marina Caskey, Paul D Bieniasz, Theodora Hatziioannou, and Michel C Nussenzweig. Antibody evolution after SARS-CoV-2 mRNA vaccination. 29 July 2021.

[58] Angela Choi, Matthew Koch, Kai Wu, Laurence Chu, Lingzhi Ma, Anna Hill, Naveen Nunna, Wenmei Huang, Judy Oestreicher, Tonya Colpitts, Hamilton Bennett, Holly Legault, Yamuna Paila, Biliana Nestorova, Baoyu Ding, David Montefiori, Rolando Pajon, Jacqueline M Miller, Brett Leav, Andrea Carfi, Roderick McPhee, and Darin K Edwards. Safety and immunogenicity of SARS-CoV-2 variant mRNA vaccine boosters in healthy adults: an interim analysis. *Nature medicine*, 15 September 2021.

[59] Bina Choi, Manish C Choudhary, James Regan, Jeffrey A Sparks, Robert F Padera, Xueting Qiu, Isaac H Solomon, Hsiao-Hsuan Kuo, Julie Boucau, Kathryn Bowman, U Das Adhikari, Marisa L Winkler, Alisa A Mueller, Tiffany Y-T Hsu, Michaël Desjardins, Lindsey R Baden, Brian T Chan, Bruce D Walker, Mathias Lichterfeld, Manfred Brigl, Douglas S Kwon, Sanjat Kanjilal, Eugene T Richardson, A Helena Jonsson, Galit Alter, Amy K Barczak, William P Hanage, Xu G Yu, Gaurav D Gaiha, Michael S Seaman, Manuela Cernadas, and Jonathan Z Li. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *The New England journal of medicine*, 383(23):2291–2293, 3 December 2020.

[60] Michael D Christian, Susan M Poutanen, Mona R Loutfy, Matthew P Muller, and Donald E Low. Severe acute respiratory syndrome. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 38(10):1420–1427, 15 May 2004.

[61] T Clackson and J A Wells. A hot spot of binding energy in a hormone-receptor interface. *Science*, 267(5196):383–386, 20 January 1995.

[62] Dami A Collier, Anna De Marco, Isabella A T M Ferreira, Bo Meng, Rawlings P Datir, Alexandra C Walls, Steven A Kemp, Jessica Bassi, Dora Pinto, Chiara Silacci-Fregni, Siro Bianchi, M Alejandra Tortorici, John Bowen, Katja Culap, Stefano Jaconi, Elisabetta Cameroni, Gyorgy Snell, Matteo S Pizzuto, Alessandra Franzetti Pellanda, Christian Garzoni, Agostino Riva, CITIID-NIHR BioResource COVID-19 Collaboration, Anne Elmer, Nathalie Kingston, Barbara Graves, Laura E McCoy, Kenneth G C Smith, John R Bradley, Nigel Temperton, Lourdes Ceron-Gutierrez, Gabriela Barcenas-Morales, COVID-19 Genomics UK (COG-UK) Consortium, William Harvey, Herbert W Virgin, Antonio Lanzavecchia, Luca Piccoli, Rainer Doffinger, Mark Wills, David Veesler, Davide Corti, and Ravindra K Gupta. Sensitivity of SARS-CoV-2 b.1.1.7 to mRNA vaccine-elicited antibodies. *Nature*, 593(7857):136–141, May 2021.

[63] Kizzmekia S Corbett, Darin K Edwards, Sarah R Leist, Olubukola M Abiona, Seyhan Boyoglu-Barnum, Rebecca A Gillespie, Sunny Himansu, Alexandra Schäfer, Cynthia T Ziwawo, Anthony T DiPiazza, Kenneth H Dinnon, Sayda M Elbashir, Christine A Shaw, Angela Woods, Ethan J Fritch, David R Martinez, Kevin W Bock, Mahnaz Minai, Bianca M Nagata, Geoffrey B Hutchinson, Kai Wu, Carole Henry, Kapil Bahl, Dario Garcia-Dominguez, Lingzhi Ma, Isabella Renzi, Wing-Pui Kong, Stephen D Schmidt, Lingshu Wang, Yi Zhang, Emily Phung, Lauren A Chang, Rebecca J Loomis, Nedim Emil Altaras, Elisabeth Narayanan, Mihir Metkar, Vlad Presnyak, Cuiping Liu, Mark K Louder, Wei Shi, Kwanyee Leung, Eun Sung Yang, Ande West, Kendra L Gully, Laura J Stevens, Nianshuang Wang, Daniel Wrapp, Nicole A Doria-Rose,

Guillaume Stewart-Jones, Hamilton Bennett, Gabriela S Alvarado, Martha C Nason, Tracy J Ruckwardt, Jason S McLellan, Mark R Denison, James D Chappell, Ian N Moore, Kaitlyn M Morabito, John R Mascola, Ralph S Baric, Andrea Carfi, and Barney S Graham. SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature*, 586(7830):567–571, October 2020.

[64] Davide Corti, Lisa A Purcell, Gyorgy Snell, and David Veesler. Tackling COVID-19 with neutralizing monoclonal antibodies. *Cell*, 184(12):3086–3108, 10 June 2021.

[65] Katharine H D Crawford and Jesse D Bloom. alignparse: A python package for parsing complex features from high-throughput long-read sequencing. *Journal of open source software*, 4(44), 11 December 2019.

[66] Katharine H D Crawford, Adam S Dingens, Rachel Eguia, Caitlin R Wolf, Naomi Wilcox, Jennifer K Logue, Kiel Shuey, Amanda M Casto, Brooke Fiala, Samuel Wrenn, Deleah Pettie, Neil P King, Alexander L Greninger, Helen Y Chu, and Jesse D Bloom. Dynamics of neutralizing antibody titers in the months after SARS-CoV-2 infection. *The Journal of infectious diseases*, 30 September 2020.

[67] Katharine H D Crawford, Rachel Eguia, Adam S Dingens, Andrea N Loes, Keara D Malone, Caitlin R Wolf, Helen Y Chu, M Alejandra Tortorici, David Veesler, Michael Murphy, Deleah Pettie, Neil P King, Alejandro B Balazs, and Jesse D Bloom. Protocol and reagents for pseudotyping lentiviral particles with SARS-CoV-2 spike protein for neutralization assays. *Viruses*, 12(5), 6 May 2020.

[68] B C Cunningham and J A Wells. Comparison of a structural and a functional epitope. *Journal of molecular biology*, 234(3):554–563, 5 December 1993.

[69] Lianpan Dai and George F Gao. Viral targets for vaccines against COVID-19. *Nature reviews. Immunology*, 21(2):73–82, February 2021.

[70] W Dall'Acqua, E R Goldman, W Lin, C Teng, D Tsuchiya, H Li, X Ysern, B C Braden, Y Li, S J Smith-Gill, and R A Mariuzza. A mutational analysis of binding interactions in an antigen-antibody protein-protein complex. *Biochemistry*, 37(22):7981–7991, 2 June 1998.

[71] Nicholas G Davies, Sam Abbott, Rosanna C Barnard, Christopher I Jarvis, Adam J Kucharski, James D Munday, Carl A B Pearson, Timothy W Russell, Damien C Tully, Alex D Washburne, Tom Wenseleers, Amy Gimma, William Waites, Kerry L M Wong, Kevin van Zandvoort, Justin D Silverman, CMMID COVID-19 Working

Group, COVID-19 Genomics UK (COG-UK) Consortium, Karla Diaz-Ordaz, Ruth Keogh, Rosalind M Eggo, Sebastian Funk, Mark Jit, Katherine E Atkins, and W John Edmunds. Estimated transmissibility and impact of SARS-CoV-2 lineage b.1.1.7 in england. *Science*, 372(6538), 9 April 2021.

[72] Tulio de Oliveira, Silvia Lutucuta, John Nkengasong, Joana Morais, Joana Paula Paixão, Zoraima Neto, Pedro Afonso, Julio Miranda, Kumbelembe David, Luzia Inglês, Amilton Pereira Agostinho Paulo Raisa Rivas Carralero, Helga Reis Freitas, Franco Mufinda, Sofonias Kifle Tessema, Houriiyah Tegally, Emmanuel James San, Eduan Wilkinson, Jennifer Giandhari, Sureshnee Pillay, Marta Giovanetti, Yeshnee Naidoo, Aris Katzourakis, Mahan Ghafari, Lavanya Singh, Derek Tshiabuila, Darren Martin, and Richard J Lessells. A novel variant of interest of SARS-CoV-2 with multiple spike mutations detected through travel surveillance in africa. 4 April 2021.

[73] Antony M Dean and Joseph W Thornton. Mechanistic approaches to the study of evolution: the functional synthesis. *Nature reviews. Genetics*, 8(9):675–688, September 2007.

[74] Wanwisa Dejnirattisai, Daming Zhou, Helen M Ginn, Helen M E Duyvesteyn, Piyada Supasa, James Brett Case, Yuguang Zhao, Thomas S Walter, Alexander J Mentzer, Chang Liu, Beibei Wang, Guido C Paesen, Jose Slon-Campos, César López-Camacho, Natasha M Kafai, Adam L Bailey, Rita E Chen, Baoling Ying, Craig Thompson, Jai Bolton, Alex Fyfe, Sunetra Gupta, Tiong Kit Tan, Javier Gilbert-Jaramillo, William James, Michael Knight, Miles W Carroll, Donal Skelly, Christina Dold, Yanchun Peng, Robert Levin, Tao Dong, Andrew J Pollard, Julian C Knight, Paul Klenerman, Nigel Temperton, David R Hall, Mark A Williams, Neil G Paterson, Felicity K R Bertram, C Alistair Siebert, Daniel K Clare, Andrew Howe, Julika Radecke, Yun Song, Alain R Townsend, Kuan-Ying A Huang, Elizabeth E Fry, Juthathip Mongkolsapaya, Michael S Diamond, Jingshan Ren, David I Stuart, and Gavin R Screaton. The antigenic anatomy of SARS-CoV-2 receptor binding domain. *Cell*, 184(8):2183–2200.e22, 15 April 2021.

[75] Marie Laure Delignette-Muller and Christophe Dutang. fitdistrplus: AnRPackage for fitting distributions. *Journal of statistical software*, 64(4), 2015.

[76] Ann Demogines, Michael Farzan, and Sara L Sawyer. Evidence for ACE2-utilizing coronaviruses (CoVs) related to severe acute respiratory syndrome CoV in bats. *Journal of virology*, 86(11):6350–6353, June 2012.

[77] Xianding Deng, Miguel A Garcia-Knight, Mir M Khalid, Venice Servellita, Candace Wang, Mary Kate Morris, Alicia Sotomayor-González, Dustin R Glasner, Kevin R

Reyes, Amelia S Gliwa, Nikitha P Reddy, Claudia Sanchez San Martin, Scot Federman, Jing Cheng, Joanna Balcerek, Jordan Taylor, Jessica A Streithorst, Steve Miller, Bharath Sreekumar, Pei-Yi Chen, Ursula Schulze-Gahmen, Taha Y Taha, Jennifer M Hayashi, Camille R Simoneau, G Renuka Kumar, Sarah McMahon, Peter V Lidsky, Yinghong Xiao, Peera Hemarajata, Nicole M Green, Alex Espinosa, Chantha Kath, Monica Haw, John Bell, Jill K Hacker, Carl Hanson, Debra A Wadford, Carlos Anaya, Donna Ferguson, Phillip A Frankino, Haridha Shivram, Liana F Lareau, Stacia K Wyman, Melanie Ott, Raul Andino, and Charles Y Chiu. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. *Cell*, 20 April 2021.

[78] Xianding Deng, Wei Gu, Scot Federman, Louis du Plessis, Oliver G Pybus, Nuno R Faria, Candace Wang, Guixia Yu, Brian Bushnell, Chao-Yang Pan, Hugo Guevara, Alicia Sotomayor-Gonzalez, Kelsey Zorn, Allan Gopez, Venice Servellita, Elaine Hsu, Steve Miller, Trevor Bedford, Alexander L Greninger, Pavitra Roychoudhury, Lea M Starita, Michael Famulare, Helen Y Chu, Jay Shendure, Keith R Jerome, Catie Anderson, Karthik Gangavarapu, Mark Zeller, Emily Spencer, Kristian G Andersen, Duncan MacCannell, Clinton R Paden, Yan Li, Jing Zhang, Suxiang Tong, Gregory Armstrong, Scott Morrow, Matthew Willis, Bela T Matyas, Sundari Mase, Olivia Kasirye, Maggie Park, Godfred Masinde, Curtis Chan, Alexander T Yu, Shua J Chai, Elsa Villarino, Brandon Bonin, Debra A Wadford, and Charles Y Chiu. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into northern california. *Science*, 369(6503):582–587, 31 July 2020.

[79] Mark R Denison, Rachel L Graham, Eric F Donaldson, Lance D Eckerle, and Ralph S Baric. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA biology*, 8(2):270–279, March 2011.

[80] M Eugenia Dieterle, Denise Haslwanter, Robert H Bortz, 3rd, Ariel S Wirchnianski, Gorka Lasso, Olivia Vergnolle, Shawn A Abbasi, J Maximilian Fels, Ethan Laudermilch, Catalina Florez, Amanda Mengotto, Duncan Kimmel, Ryan J Malonis, George Georgiev, Jose Quiroz, Jason Barnhill, Liise-Anne Pirofski, Johanna P Daily, John M Dye, Jonathan R Lai, Andrew S Herbert, Kartik Chandran, and Rohit K Jangra. A Replication-Competent vesicular stomatitis virus for studies of SARS-CoV-2 Spike-Mediated cell entry and its inhibition. *Cell host & microbe*, 28(3):486–496.e6, 9 September 2020.

[81] Adam S Dingens, Dana Arenz, Haidyn Weight, Julie Overbaugh, and Jesse D Bloom. An antigenic atlas of HIV-1 escape from broadly neutralizing antibodies distinguishes functional and structural epitopes. *Immunity*, 50(2):520–532.e3, 19 February 2019.

[82] Adam S Dingens, Katharine H D Crawford, Amanda Adler, Sarah L Steele, Kirsten Lacombe, Rachel Eguia, Fatima Amanat, Alexandra C Walls, Caitlin R Wolf, Michael Murphy, Deleah Pettie, Lauren Carter, Xuan Qin, Neil P King, David Veesler, Florian Krammer, Jane A Dickerson, Helen Y Chu, Janet A Englund, and Jesse D Bloom. Serological identification of SARS-CoV-2 infections among children visiting a hospital during the initial seattle outbreak. *Nature communications*, 11(1):4378, 1 September 2020.

[83] Jinhui Dong, Seth J Zost, Allison J Greaney, Tyler N Starr, Adam S Dingens, Elaine C Chen, Rita E Chen, James Brett Case, Rachel E Sutton, Pavlo Gilchuk, Jessica Rodriguez, Erica Armstrong, Christopher Gainza, Rachel S Nargi, Elad Binshtein, Xuping Xie, Xianwen Zhan, Pei-Yong Shi, James Logue, Stuart Weston, Marisa E McGrath, Matthew B Frieman, Tyler Brady, Kevin Tuffy, Helen Bright, Yueh-Ming Loo, Patrick McTamney, Mark Esser, Robert H Carnahan, Michael S Diamond, Jesse D Bloom, and James E Crowe. Genetic and structural basis for recognition of SARS-CoV-2 spike protein by a two-antibody cocktail. *bioRxiv*, page 2021.01.27.428529, 28 January 2021.

[84] Michael B Doud and Jesse D Bloom. Accurate measurement of the effects of all Amino-Acid mutations on influenza hemagglutinin. *Viruses*, 8(6), 3 June 2016.

[85] Michael B Doud, Scott E Hensley, and Jesse D Bloom. Complete mapping of viral escape from neutralizing antibodies. *PLoS pathogens*, 13(3):e1006271, March 2017.

[86] Michael B Doud, Juhye M Lee, and Jesse D Bloom. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. *Nature communications*, 9(1):1386, 11 April 2018.

[87] Arthur W D Edridge, Joanna Kaczorowska, Alexis C R Hoste, Margreet Bakker, Michelle Klein, Katherine Loens, Maarten F Jebbink, Amy Matser, Cormac M Kinsella, Paloma Rueda, Margareta Ieven, Herman Goossens, Maria Prins, Patricia Sastre, Martin Deijs, and Lia van der Hoek. Seasonal coronavirus protective immunity is short-lasting. *Nature medicine*, 14 September 2020.

[88] Dirk Eggink, Peter H Goff, and Peter Palese. Guiding the immune response against influenza virus hemagglutinin toward the conserved stalk domain by hyperglycosylation of the globular head domain. *Journal of virology*, 88(1):699–704, January 2014.

[89] Rachel T Eguia, Katharine H D Crawford, Terry Stevens-Ayers, Laurel Kelnhofer-Millevolte, Alexander L Greninger, Janet A Englund, Michael J Boeckh, and Jesse D

Bloom. A human coronavirus evolves antigenically to escape antibody immunity. *PLoS pathogens*, 17(4):e1009453, 8 April 2021.

[90] Tal Einav and Jesse D Bloom. When two are better than one: Modeling the mechanisms of antibody mixtures. *PLoS computational biology*, 16(5):e1007830, May 2020.

[91] Stefan Elbe and Gemma Buckland-Merrett. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global challenges (Hoboken, NJ)*, 1(1):33–46, January 2017.

[92] Emily J Erbelding, Diane J Post, Erik J Stemmy, Paul C Roberts, Alison Deckhut Augustine, Stacy Ferguson, Catharine I Paules, Barney S Graham, and Anthony S Fauci. A universal influenza vaccine: The strategic plan for the national institute of allergy and infectious diseases. *The Journal of infectious diseases*, 218(3):347–354, 2 July 2018.

[93] B Fane, R Villafane, A Mitraki, and J King. Identification of global suppressors for temperature-sensitive folding mutations of the P22 tailspike protein. *The Journal of biological chemistry*, 266(18):11640–11648, 25 June 1991.

[94] Nuno R Faria, Thomas A Mellan, Charles Whittaker, Ingra M Claro, Darlan da S Candido, Swapnil Mishra, Myuki A E Crispim, Flavia C S Sales, Iwona Hawryluk, John T McCrone, Ruben J G Hulswit, Lucas A M Franco, Mariana S Ramundo, Jaqueline G de Jesus, Pamela S Andrade, Thais M Coletti, Giulia M Ferreira, Camila A M Silva, Erika R Manuli, Rafael H M Pereira, Pedro S Peixoto, Moritz U G Kraemer, Nelson Gaburo, Jr, Cecilia da C Camilo, Henrique Hoeltgebaum, William M Souza, Esmenia C Rocha, Leandro M de Souza, Mariana C de Pinho, Leonardo J T Araujo, Frederico S V Malta, Aline B de Lima, Joice do P Silva, Danielle A G Zauli, Alessandro C de S Ferreira, Ricardo P Schnekenberg, Daniel J Laydon, Patrick G T Walker, Hannah M Schlüter, Ana L P Dos Santos, Maria S Vidal, Valentina S Del Caro, Rosinaldo M F Filho, Helem M Dos Santos, Renato S Aguiar, José L Proença-Modena, Bruce Nelson, James A Hay, Mélodie Monod, Xenia Miscouridou, Helen Coupland, Raphael Sonabend, Michaela Vollmer, Axel Gandy, Carlos A Prete, Jr, Vitor H Nascimento, Marc A Suchard, Thomas A Bowden, Sergei L K Pond, Chieh-Hsi Wu, Oliver Ratmann, Neil M Ferguson, Christopher Dye, Nick J Loman, Philippe Lemey, Andrew Rambaut, Nelson A Fraiji, Maria do P S S Carvalho, Oliver G Pybus, Seth Flaxman, Samir Bhatt, and Ester C Sabino. Genomics and epidemiology of the p.1 SARS-CoV-2 lineage in manaus, brazil. *Science*, 372(6544):815–821, 21 May 2021.

[95] Nikhil Faulkner, Kevin W Ng, Mary Y Wu, Ruth Harvey, Marios Margaritis, Stavroula Paraskevopoulou, Catherine Houlihan, Saira Hussain, Maria Greco, William Bolland, Scott Warchal, Judith Heaney, Hannah Rickman, Moria Spyer, Daniel Frampton, Matthew Byott, Tulio de Oliveira, Alex Sigal, Svend Kjaer, Charles Swanton, Sonia Gandhi, Rupert Beale, Steve J Gamblin, John W McCauley, Rodney Stuart Daniels, Michael Howell, David Bauer, Eleni Nastouli, and George Kassiotis. Reduced antibody cross-reactivity following infection with b.1.1.7 than with parental SARS-CoV-2 strains. *eLife*, 10, 29 July 2021.

[96] Joseph R Fauver, Mary E Petrone, Emma B Hodcroft, Kayoko Shioda, Hanna Y Ehrlich, Alexander G Watts, Chantal B F Vogels, Anderson F Brito, Tara Alpert, Anthony Muyombwe, Jafar Razeq, Randy Downing, Nagarjuna R Cheemarla, Anne L Wyllie, Chaney C Kalinich, Isabel M Ott, Joshua Quick, Nicholas J Loman, Karla M Neugebauer, Alexander L Greninger, Keith R Jerome, Pavitra Roychoudhury, Hong Xie, Lasata Shrestha, Meei-Li Huang, Virginia E Pitzer, Akiko Iwasaki, Saad B Omer, Kamran Khan, Isaac I Bogoch, Richard A Martinello, Ellen F Foxman, Marie L Landry, Richard A Neher, Albert I Ko, and Nathan D Grubaugh. Coast-to-Coast spread of SARS-CoV-2 during the early epidemic in the united states. *Cell*, 181(5):990–996.e5, 28 May 2020.

[97] Alison F Feder, Soo-Yon Rhee, Susan P Holmes, Robert W Shafer, Dmitri A Petrov, and Pleuni S Pennings. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. *eLife*, 5, 15 February 2016.

[98] Jared Feldman, Julia Bals, Clara G Altomare, Kerri St Denis, Evan C Lam, Blake M Hauser, Larance Ronsard, Maya Sangesland, Thalia Bracamonte Moreno, Vintus Okonkwo, Nathania Hartojo, Alejandro B Balazs, Goran Bajic, Daniel Lingwood, and Aaron G Schmidt. Naive human B cells engage the receptor binding domain of SARS-CoV-2, variants of concern, and related sarbecoviruses. *Science Immunology*, 0(0):eabl5842, 2021.

[99] Isabella A T M Ferreira, Steven A Kemp, Rawlings Datir, Akatsuki Saito, Bo Meng, Partha Rakshit, Akifumi Takaori-Kondo, Yusuke Kosugi, Keiya Uriu, Izumi Kimura, Kotaro Shirakawa, Adam Abdullahi, Anurag Agarwal, Seiya Ozono, Kenzo Tokunaga, Kei Sato, Ravindra K Gupta, CITIID-NIHR BioResource COVID-19 Collaboration, Indian SARS-CoV-2 Genomics Consortium, and Genotype to Phenotype Japan (G2P-Japan) Consortium. SARS-CoV-2 b.1.617 mutations L452R and E484Q are not synergistic for antibody evasion. *The Journal of infectious diseases*, 224(6):989–994, 17 September 2021.

[100] Pedro M Folegatti, Katie J Ewer, Parvinder K Aley, Brian Angus, Stephan Becker, Sandra Belij-Rammerstorfer, Duncan Bellamy, Sagida Bibi, Mustapha Bittaye, Elizabeth A Clutterbuck, Christina Dold, Saul N Faust, Adam Finn, Amy L Flaxman, Bassam Hallis, Paul Heath, Daniel Jenkin, Rajeka Lazarus, Rebecca Makinson, Angela M Minassian, Katrina M Pollock, Maheshi Ramasamy, Hannah Robinson, Matthew Snape, Richard Tarrant, Merryn Voysey, Catherine Green, Alexander D Douglas, Adrian V S Hill, Teresa Lambe, Sarah C Gilbert, Andrew J Pollard, and Oxford COVID Vaccine Trial Group. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet*, 396(10249):467–478, 15 August 2020.

[101] Douglas M Fowler and Stanley Fields. Deep mutational scanning: a new style of protein science. *Nature methods*, 11(8):801–807, August 2014.

[102] Thomas Francis. On the doctrine of original antigenic sin. *Proceedings of the American Philosophical Society*, 104(6):572–578, 1960.

[103] Hannah K Frank, David Enard, and Scott D Boyd. Exceptional diversity and selection pressure on SARS-CoV and SARS-CoV-2 host receptor in bats compared to other mammals. *bioRxiv : the preprint server for biology*, 20 April 2020.

[104] Thomas C Friedrich, Christopher A Frye, Levi J Yant, David H O'Connor, Nancy A Kriewaldt, Meghan Benson, Lara Vojnov, Elizabeth J Dodds, Candice Cullen, Richard Rudersdorf, Austin L Hughes, Nancy Wilson, and David I Watkins. Extraepitopic compensatory substitutions partially restore fitness to simian immunodeficiency virus variants that escape from an immunodominant cytotoxic-t-lymphocyte response. *Journal of virology*, 78(5):2581–2585, March 2004.

[105] Benjamin O Fulton, David Sachs, Shannon M Beaty, Sohui T Won, Benhur Lee, Peter Palese, and Nicholas S Heaton. Mutational analysis of measles virus suggests constraints on antigenic variation of the glycoproteins. *Cell reports*, 11(9):1331–1338, 9 June 2015.

[106] Christian Gaebler, Zijun Wang, Julio C C Lorenzi, Frauke Muecksch, Shlomo Finkin, Minami Tokuyama, Alice Cho, Mila Jankovic, Dennis Schaefer-Babajew, Thiago Y Oliveira, Melissa Cipolla, Charlotte Viant, Christopher O Barnes, Yaron Bram, Gaëlle Breton, Thomas Hägglöf, Pilar Mendoza, Arlene Hurley, Martina Turroja, Kristie Gordon, Katrina G Millard, Victor Ramos, Fabian Schmidt, Yiska Weisblum, Divya Jha, Michael Tankelevich, Gustavo Martinez-Delgado, Jim Yee, Roshni Patel, Juan Dizon, Cecille Unson-O'Brien, Irina Shimeliovich, Davide F Robbiani, Zhen Zhao,

Anna Gazumyan, Robert E Schwartz, Theodora Hatziioannou, Pamela J Bjorkman, Saurabh Mehandru, Paul D Bieniasz, Marina Caskey, and Michel C Nussenzweig. Evolution of antibody immunity to SARS-CoV-2. *Nature*, 591(7851):639–644, March 2021.

[107] Alain Gagnon, Enrique Acosta, Stacey Hallman, Robert Bourbeau, Lisa Y Dillon, Nadine Ouellette, David J D Earn, D Ann Herring, Kris Inwood, Joaquin Madrenas, and Matthew S Miller. Pandemic paradox: Early life H2N2 pandemic influenza infection enhanced susceptibility to death during the 2009 H1N1 pandemic. *mBio*, 9(1), 16 January 2018.

[108] S Annie Gai and K Dane Wittrup. Yeast surface display for protein engineering and characterization. *Current opinion in structural biology*, 17(4):467–473, August 2007.

[109] Tiziano Gaiotto and Simon E Hufton. Cross-Neutralising nanobodies bind to a conserved pocket in the hemagglutinin stem region identified using yeast display and deep mutational scanning. *PloS one*, 11(10):e0164296, 14 October 2016.

[110] Wilfredo F Garcia-Beltran, Evan C Lam, Kerri St Denis, Adam D Nitido, Zeidy H Garcia, Blake M Hauser, Jared Feldman, Maia N Pavlovic, David J Gregory, Mark C Poznansky, Alex Sigal, Aaron G Schmidt, A John Iafrate, Vivek Naranbhai, and Alejandro B Balazs. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*, 184(9):2372–2383.e9, 29 April 2021.

[111] Meghan E Garrett, Jared Galloway, Helen Y Chu, Hannah L Itell, Caitlin I Stoddard, Caitlin R Wolf, Jennifer K Logue, Dylan McDonald, Haidyn Weight, Frederick A Matsen, 4th, and Julie Overbaugh. High-resolution profiling of pathways of escape for SARS-CoV-2 spike-binding antibodies. *Cell*, 184(11):2927–2938.e11, 27 May 2021.

[112] Meghan E Garrett, Jared G Galloway, Caitlin Wolf, Jennifer K Logue, Nicholas Franko, Helen Y Chu, Frederick A Matsen, and Julie Overbaugh. Comprehensive characterization of the antibody responses to SARS-CoV-2 spike protein after infection and/or vaccination. *bioRxiv : the preprint server for biology*, 5 October 2021.

[113] Meghan E Garrett, Hannah L Itell, Katharine H D Crawford, Ryan Basom, Jesse D Bloom, and Julie Overbaugh. Phage-DMS: A comprehensive method for fine mapping of antibody epitopes. *iScience*, 23(10):101622, 23 October 2020.

[114] Daryl Geers, Marc C Shamier, Susanne Bogers, Gerco den Hartog, Lennert Gommers, Nella N Nieuwkoop, Katharina S Schmitz, Laurine C Rijsbergen, Jolieke A T van

Osch, Emma Dijkhuizen, Gaby Smits, Anouskha Comvalius, Djenolan van Mourik, Tom G Caniels, Marit J van Gils, Rogier W Sanders, Bas B Oude Munnink, Richard Molenkamp, Herbert J de Jager, Bart L Haagmans, Rik L de Swart, Marion P G Koopmans, Robert S van Binnendijk, Rory D de Vries, and Corine H GeurtsvanKessel. SARS-CoV-2 variants of concern partially escape humoral but not t-cell responses in COVID-19 convalescent donors and vaccinees. *Science immunology*, 6(59), 25 May 2021.

[115] R Daniel Gietz and Robert H Schiestl. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature protocols*, 2(1):38–41, 2007.

[116] Pavlo Gilchuk, Robin G Bombardi, Jesse H Erasmus, Qing Tan, Rachel Nargi, Cinque Soto, Peter Abbink, Todd J Suscovich, Lorellin A Durnell, Amit Khandhar, Jacob Archer, Jenny Liang, Mallorie E Fouch, Edgar Davidson, Benjamin J Doranz, Taylor Jones, Elise Larson, Stacey Ertel, Brian Granger, Jasmine Fuerte-Stone, Vicky Roy, Thomas Broge, Thomas C Linnekin, Caitlyn H Linde, Matthew J Gorman, Joseph Nkolola, Galit Alter, Steven G Reed, Dan H Barouch, Michael S Diamond, James E Crowe, Jr, Neal Van Hoeven, Larissa B Thackray, and Robert H Carnahan. Integrated pipeline for the accelerated discovery of antiviral antibody therapeutics. *Nature biomedical engineering*, 3 August 2020.

[117] Pavlo Gilchuk, Charles D Murin, Jacob C Milligan, Robert W Cross, Chad E Mire, Philipp A Ilinykh, Kai Huang, Natalia Kuzmina, Pilar X Altman, Sean Hui, Bronwyn M Gunn, Aubrey L Bryan, Edgar Davidson, Benjamin J Doranz, Hannah L Turner, Tanwee Alkutkar, Robin Flinko, Chiara Orlandi, Robert Carnahan, Rachel Nargi, Robin G Bombardi, Megan E Vodzak, Sheng Li, Adaora Okoli, Morris Ibeawuchi, Benjamin Ohiaeri, George K Lewis, Galit Alter, Alexander Bukreyev, Erica Ollmann Saphire, Thomas W Geisbert, Andrew B Ward, and James E Crowe, Jr. Analysis of a therapeutic antibody cocktail reveals determinants for cooperative and broad ebolavirus neutralization. *Immunity*, 52(2):388–403.e12, 18 February 2020.

[118] Lizhi Ian Gong, Marc A Suchard, and Jesse D Bloom. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife*, 2:e00631, 14 May 2013.

[119] Barry J Grant, Ana P C Rodrigues, Karim M ElSawy, J Andrew McCammon, and Leo S D Caves. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, 22(21):2695–2696, 1 November 2006.

[120] Allison J Greaney, Andrea N Loes, Katharine H D Crawford, Tyler N Starr, Keara D Malone, Helen Y Chu, and Jesse D Bloom. Comprehensive mapping of mutations in

the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell host & microbe*, 29(3):463–476.e6, 10 March 2021.

[121] Allison J Greaney, Andrea N Loes, Lauren E Gentles, Katharine H D Crawford, Tyler N Starr, Keara D Malone, Helen Y Chu, and Jesse D Bloom. Antibodies elicited by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from SARS-CoV-2 infection. *Science translational medicine*, 13(600), 30 June 2021.

[122] Allison J Greaney, Tyler N Starr, Christopher O Barnes, Yiska Weisblum, Fabian Schmidt, Marina Caskey, Christian Gaebler, Alice Cho, Marianna Agudelo, Shlomo Finkin, Zijun Wang, Daniel Poston, Frauke Muecksch, Theodora Hatziioannou, Paul D Bieniasz, Davide F Robbiani, Michel C Nussenzweig, Pamela J Bjorkman, and Jesse D Bloom. Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nature communications*, 12(1):4196, 7 July 2021.

[123] Allison J Greaney, Tyler N Starr, Rachel T Eguia, Andrea N Loes, Khadija Khan, Farina Karim, Sandile Cele, John E Bowen, Jennifer K Logue, Davide Corti, David Veesler, Helen Y Chu, Alex Sigal, and Jesse D Bloom. A SARS-CoV-2 variant elicits an antibody response with a shifted immunodominance hierarchy. 13 October 2021.

[124] Allison J Greaney, Tyler N Starr, Pavlo Gilchuk, Seth J Zost, Elad Binshtein, Andrea N Loes, Sarah K Hilton, John Huddleston, Rachel Eguia, Katharine H D Crawford, Adam S Dingens, Rachel S Nargi, Rachel E Sutton, Naveenchandra Suryadevara, Paul W Rothlauf, Zhuoming Liu, Sean P J Whelan, Robert H Carnahan, James E Crowe, Jr, and Jesse D Bloom. Complete mapping of mutations to the SARS-CoV-2 spike Receptor-Binding domain that escape antibody recognition. *Cell host & microbe*, 29(1):44–57.e9, 13 January 2021.

[125] Hugh K Haddox, Adam S Dingens, and Jesse D Bloom. Experimental estimation of the effects of all Amino-Acid mutations to HIV's envelope protein on viral replication in cell culture. *PLoS pathogens*, 12(12):e1006114, December 2016.

[126] James Hadfield, Colin Megill, Sidney M Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A Neher. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23):4121–4123, 1 December 2018.

[127] Stephen R Hamilton, Piotr Bobrowicz, Beata Bobrowicz, Robert C Davidson, Huijuan Li, Teresa Mitchell, Juergen H Nett, Sebastian Rausch, Terrance A Stadheim, Harry

Wischnewski, Stefan Wildt, and Tillman U Gerngross. Production of complex human glycoproteins in yeast. *Science*, 301(5637):1244–1246, 29 August 2003.

[128] Kathy Hancock, Vic Veguilla, Xiuhua Lu, Weimin Zhong, Eboneé N Butler, Hong Sun, Feng Liu, Libo Dong, Joshua R DeVos, Paul M Gargiullo, T Lynnette Brammer, Nancy J Cox, Terrence M Tumpey, and Jacqueline M Katz. Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. *The New England journal of medicine*, 361(20):1945–1952, 12 November 2009.

[129] Johanna Hansen, Alina Baum, Kristen E Pascal, Vincenzo Russo, Stephanie Giordano, Elzbieta Wloga, Benjamin O Fulton, Ying Yan, Katrina Koon, Krunal Patel, Kyung Min Chung, Aynur Hermann, Erica Ullman, Jonathan Cruz, Ashique Rafique, Tammy Huang, Jeanette Fairhurst, Christen Libertiny, Marine Malbec, Wen-Yi Lee, Richard Welsh, Glen Farr, Seth Pennington, Dipali Deshpande, Jemmie Cheng, Anke Watty, Pascal Bouffard, Robert Babb, Natasha Levenkova, Calvin Chen, Bojie Zhang, Annabel Romero Hernandez, Kei Saotome, Yi Zhou, Matthew Franklin, Sumathi Sivapalasingam, David Chien Lye, Stuart Weston, James Logue, Robert Haupt, Matthew Frieman, Gang Chen, William Olson, Andrew J Murphy, Neil Stahl, George D Yancopoulos, and Christos A Kyratsous. Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science*, 369(6506):1010–1014, 21 August 2020.

[130] Michael J Harms and Joseph W Thornton. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nature reviews. Genetics*, 14(8):559–571, August 2013.

[131] Ahmed O Hassan, Friederike Feldmann, Haiyan Zhao, David T Curiel, Atsushi Okumura, Tsing-Lee Tang-Huau, James Brett Case, Kimberly Meade-White, Julie Callison, Jamie Lovaglio, Patrick W Hanley, Dana P Scott, Daved H Fremont, Heinz Feldmann, and Michael S Diamond. A single intranasal dose of chimpanzee adenovirus-vectored vaccine protects against SARS-CoV-2 infection in rhesus macaques. 26 January 2021.

[132] Biao He, Yuzhen Zhang, Lin Xu, Weihong Yang, Fanli Yang, Yun Feng, Lele Xia, Jihua Zhou, Weibin Zhen, Ye Feng, Huancheng Guo, Hailin Zhang, and Changchun Tu. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in china. *Journal of virology*, 88(12):7070–7082, June 2014.

[133] Wan-Ting He, Meng Yuan, Sean Callaghan, Rami Musharrafieh, Ge Song, Murillo Silva, Nathan Beutler, Wilma Lee, Peter Yong, Jonathan Torres, Mariane Melo, Panpan Zhou, Fangzhu Zhao, Xueyong Zhu, Linghang Peng, Deli Huang, Fabio Anzanello, James Ricketts, Mara Parren, Elijah Garcia, Melissa Ferguson, William Rinaldi, Stephen A Rawlings, David Nemazee, Davey M Smith, Bryan Briney, Yana Safonova, Thomas F Rogers, Shane Crotty, Darrell J Irvine, Andrew B Ward, Ian A Wilson, Dennis R Burton, and Raiees Andrabi. Broadly neutralizing antibodies to SARS-related viruses can be readily induced in rhesus macaques. 6 July 2021.

[134] J O Hendley, H B Fishburne, and J M Gwaltney, Jr. Coronavirus infections in working adults. eight-year study with 229 E and OC 43. *The American review of respiratory disease*, 105(5):805–811, May 1972.

[135] Scott E Hensley, Suman R Das, Adam L Bailey, Loren M Schmidt, Heather D Hickman, Akila Jayaraman, Karthik Viswanathan, Rahul Raman, Ram Sasisekharan, Jack R Bennink, and Jonathan W Yewdell. Hemagglutinin receptor binding avidity drives influenza a virus antigenic drift. *Science*, 326(5953):734–736, 30 October 2009.

[136] Joseph B Hiatt, Rupali P Patwardhan, Emily H Turner, Choli Lee, and Jay Shendure. Parallel, tag-directed assembly of locally derived short sequence reads. *Nature methods*, 7(2):119–122, February 2010.

[137] Sarah Hilton, John Huddleston$, Allison Black, Khrystyna North, Adam Dingens, Trevor Bedford, and Jesse Bloom. dms-view: Interactive visualization tool for deep mutational scanning data. *Journal of open source software*, 5(52):2353, 17 August 2020.

[138] Emma B Hodcroft. CoVariants: SARS-CoV-2 mutations and variants of interest. `https://covariants.org/`, 2021. Accessed: 2021-9-7.

[139] Markus Hoffmann, Hannah Kleine-Weber, Simon Schroeder, Nadine Krüger, Tanja Herrler, Sandra Erichsen, Tobias S Schiergens, Georg Herrler, Nai-Huei Wu, Andreas Nitsche, Marcel A Müller, Christian Drosten, and Stefan Pöhlmann. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181(2):271–280.e8, 16 April 2020.

[140] Yixuan J Hou, Kenichi Okuda, Caitlin E Edwards, David R Martinez, Takanori Asakura, Kenneth H Dinnon, 3rd, Takafumi Kato, Rhianna E Lee, Boyd L Yount, Teresa M Mascenik, Gang Chen, Kenneth N Olivier, Andrew Ghio, Longping V Tse, Sarah R Leist, Lisa E Gralinski, Alexandra Schäfer, Hong Dang, Rodney Gilmore,

Satoko Nakano, Ling Sun, M Leslie Fulcher, Alessandra Livraghi-Butrico, Nathan I Nicely, Mark Cameron, Cheryl Cameron, David J Kelvin, Aravinda de Silva, David M Margolis, Alena Markmann, Luther Bartelt, Ross Zumwalt, Fernando J Martinez, Steven P Salvatore, Alain Borczuk, Purushothama R Tata, Vishwaraj Sontake, Adam Kimple, Ilona Jaspers, Wanda K O'Neal, Scott H Randell, Richard C Boucher, and Ralph S Baric. SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. *Cell*, 182(2):429–446.e14, 23 July 2020.

[141] Ben Hu, Lei-Ping Zeng, Xing-Lou Yang, Xing-Yi Ge, Wei Zhang, Bei Li, Jia-Zheng Xie, Xu-Rui Shen, Yun-Zhi Zhang, Ning Wang, Dong-Sheng Luo, Xiao-Shuang Zheng, Mei-Niang Wang, Peter Daszak, Lin-Fa Wang, Jie Cui, and Zheng-Li Shi. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS pathogens*, 13(11):e1006698, November 2017.

[142] Kuan-Ying A Huang, Tiong Kit Tan, Ting-Hua Chen, Chung-Guei Huang, Ruth Harvey, Saira Hussain, Cheng-Pin Chen, Adam Harding, Javier Gilbert-Jaramillo, Xu Liu, Michael Knight, Lisa Schimanski, Shin-Ru Shih, Yi-Chun Lin, Chien-Yu Cheng, Shu-Hsing Cheng, Yhu-Chering Huang, Tzou-Yien Lin, Jia-Tsrong Jan, Che Ma, William James, Rodney S Daniels, John W McCauley, Pramila Rijal, and Alain R Townsend. Breadth and function of antibody response to acute SARS-CoV-2 infection in humans. *PLoS pathogens*, 17(2):e1009352, 26 February 2021.

[143] Jiandong Huo, Yuguang Zhao, Jingshan Ren, Daming Zhou, Helen M E Duyvesteyn, Helen M Ginn, Loic Carrique, Tomas Malinauskas, Reinis R Ruza, Pranav N M Shah, Tiong Kit Tan, Pramila Rijal, Naomi Coombes, Kevin R Bewley, Julia A Tree, Julika Radecke, Neil G Paterson, Piyada Supasa, Juthathip Mongkolsapaya, Gavin R Screaton, Miles Carroll, Alain Townsend, Elizabeth E Fry, Raymond J Owens, and David I Stuart. Neutralization of SARS-CoV-2 by destruction of the prefusion spike. *Cell host & microbe*, 28(3):445–454.e6, 9 September 2020.

[144] William C Hwang, Yaqiong Lin, Eugenio Santelli, Jianhua Sui, Lukasz Jaroszewski, Boguslaw Stec, Michael Farzan, Wayne A Marasco, and Robert C Liddington. Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. *The Journal of biological chemistry*, 281(45):34610–34616, 10 November 2006.

[145] F Javier Ibarrondo, Jennifer A Fulcher, David Goodman-Meza, Julie Elliott, Christian Hofmann, Mary A Hausner, Kathie G Ferbas, Nicole H Tobin, Grace M Aldrovandi, and Otto O Yang. Rapid decay of Anti–SARS-CoV-2 antibodies in persons with mild

covid-19. *The New England journal of medicine*, 383(11):1085–1087, 10 September 2020.

[146] Benjamin Israelow, Tianyang Mao, Jonathan Klein, Eric Song, Bridget Menasche, Saad B Omer, and Akiko Iwasaki. Adaptive immune determinants of viral clearance and protection in mouse models of SARS-CoV-2. *Science immunology*, 6(64):eabl4509, 15 October 2021.

[147] Lisa A Jackson, Evan J Anderson, Nadine G Rouphael, Paul C Roberts, Mamodikoe Makhene, Rhea N Coler, Michele P McCullough, James D Chappell, Mark R Denison, Laura J Stevens, Andrea J Pruijssers, Adrian McDermott, Britta Flach, Nicole A Doria-Rose, Kizzmekia S Corbett, Kaitlyn M Morabito, Sijy O'Dell, Stephen D Schmidt, Phillip A Swanson, 2nd, Marcelino Padilla, John R Mascola, Kathleen M Neuzil, Hamilton Bennett, Wellington Sun, Etza Peters, Mat Makowski, Jim Albert, Kaitlyn Cross, Wendy Buchanan, Rhonda Pikaart-Tautges, Julie E Ledgerwood, Barney S Graham, John H Beigel, and mRNA-1273 Study Group. An mRNA vaccine against SARS-CoV-2 - preliminary report. *The New England journal of medicine*, 383(20):1920–1931, 12 November 2020.

[148] L Jin, B M Fendly, and J A Wells. High resolution functional analysis of antibody-antigen interactions. *Journal of molecular biology*, 226(3):851–865, 5 August 1992.

[149] Bin Ju, Qi Zhang, Jiwan Ge, Ruoke Wang, Jing Sun, Xiangyang Ge, Jiazhen Yu, Sisi Shan, Bing Zhou, Shuo Song, Xian Tang, Jinfang Yu, Jun Lan, Jing Yuan, Haiyan Wang, Juanjuan Zhao, Shuye Zhang, Youchun Wang, Xuanling Shi, Lei Liu, Jincun Zhao, Xinquan Wang, Zheng Zhang, and Linqi Zhang. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature*, 584(7819):115–119, August 2020.

[150] Boris Julg, Po-Ting Liu, Kshitij Wagh, William M Fischer, Peter Abbink, Noe B Mercado, James B Whitney, Joseph P Nkolola, Katherine McMahan, Lawrence J Tartaglia, Erica N Borducchi, Shreeya Khatiwada, Megha Kamath, Jake A LeSuer, Michael S Seaman, Stephen D Schmidt, John R Mascola, Dennis R Burton, Bette T Korber, and Dan H Barouch. Protection against a mixed SHIV challenge by a broadly neutralizing antibody cocktail. *Science translational medicine*, 9(408), 20 September 2017.

[151] K Katoh and D M Standley. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability, 2013.

[152] Zunlong Ke, Joaquin Oton, Kun Qu, Mirko Cortese, Vojtech Zila, Lesley McKeane, Takanori Nakane, Jasenko Zivanov, Christopher J Neufeldt, Berati Cerikan, John M

Lu, Julia Peukes, Xiaoli Xiong, Hans-Georg Kräusslich, Sjors H W Scheres, Ralf Bartenschlager, and John A G Briggs. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*, 588(7838):498–502, December 2020.

[153] Cheryl Keech, Gary Albert, Iksung Cho, Andreana Robertson, Patricia Reed, Susan Neal, Joyce S Plested, Mingzhu Zhu, Shane Cloney-Clark, Haixia Zhou, Gale Smith, Nita Patel, Matthew B Frieman, Robert E Haupt, James Logue, Marisa McGrath, Stuart Weston, Pedro A Piedra, Chinar Desai, Kathleen Callahan, Maggie Lewis, Patricia Price-Abbott, Neil Formica, Vivek Shinde, Louis Fries, Jason D Lickliter, Paul Griffin, Bethanie Wilkinson, and Gregory M Glenn. Phase 1-2 trial of a SARS-CoV-2 recombinant spike protein nanoparticle vaccine. *The New England journal of medicine*, 383(24):2320–2332, 10 December 2020.

[154] Steven Kemp, William Harvey, Rawlings Datir, Dami Collier, Isabella Ferreira, Alessandro Carabelii, David L Robertson, and Ravindra K Gupta. Recurrent emergence and transmission of a SARS-CoV-2 spike deletion ΔH69/V70. *bioRxiv*, page 2020.12.14.422555, 21 December 2020.

[155] David S Khoury, Deborah Cromer, Arnold Reynaldi, Timothy E Schlub, Adam K Wheatley, Jennifer A Juno, Kanta Subbarao, Stephen J Kent, James A Triccas, and Miles P Davenport. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nature medicine*, 27(7):1205–1211, July 2021.

[156] Michael Kidd, Alex Richter, Angus Best, Jeremy Mirza, Benita Percival, Megan Mayhew, Oliver Megram, Fiona Ashford, Thomas White, Emma Moles-Garcia, Liam Crawford, Andrew Bosworth, Tim Plant, and Alan McNally. S-variant SARS-CoV-2 is associated with significantly higher viral loads in samples tested by ThermoFisher TaqPath RT-QPCR. *medRxiv*, 27 December 2020.

[157] Robert N Kirchdoerfer, Christopher A Cottrell, Nianshuang Wang, Jesper Pallesen, Hadi M Yassine, Hannah L Turner, Kizzmekia S Corbett, Barney S Graham, Jason S McLellan, and Andrew B Ward. Pre-fusion structure of a human coronavirus spike protein. *Nature*, 531(7592):118–121, 3 March 2016.

[158] Kathryn E Kistler and Trevor Bedford. Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229E. *eLife*, 10, 19 January 2021.

[159] Kathryn E Kistler, John Huddleston, and Trevor Bedford. Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. *bioRxiv : the preprint server for biology*, 14 September 2021.

[160] Florian Klein, Lilian Nogueira, Yoshiaki Nishimura, Ganesh Phad, Anthony P West, Jr, Ariel Halper-Stromberg, Joshua A Horwitz, Anna Gazumyan, Cassie Liu, Thomas R Eisenreich, Clara Lehmann, Gerd Fätkenheuer, Constance Williams, Masashi Shingai, Malcolm A Martin, Pamela J Bjorkman, Michael S Seaman, Susan Zolla-Pazner, Gunilla B Karlsson Hedestam, and Michel C Nussenzweig. Enhanced HIV-1 immunotherapy by commonly arising antibodies that target virus escape variants. *The Journal of experimental medicine*, 211(12):2361–2372, 17 November 2014.

[161] Bette Korber, Will M Fischer, Sandrasegaram Gnanakaran, Hyejin Yoon, James Theiler, Werner Abfalterer, Nick Hengartner, Elena E Giorgi, Tanmoy Bhattacharya, Brian Foley, Kathryn M Hastie, Matthew D Parker, David G Partridge, Cariad M Evans, Timothy M Freeman, Thushan I de Silva, Sheffield COVID-19 Genomics Group, Charlene McDanal, Lautaro G Perez, Haili Tang, Alex Moon-Walker, Sean P Whelan, Celia C LaBranche, Erica O Saphire, and David C Montefiori. Tracking changes in SARS-CoV-2 spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell*, 182(4):812–827.e19, 20 August 2020.

[162] Johannes Köster and Sven Rahmann. Snakemake–a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522, 1 October 2012.

[163] J M Kowalski, R N Parekh, J Mao, and K D Wittrup. Protein folding stability can determine the efficiency of escape from endoplasmic reticulum quality control. *The Journal of biological chemistry*, 273(31):19453–19458, 31 July 1998.

[164] J M Kowalski, R N Parekh, and K D Wittrup. Secretion efficiency in saccharomyces cerevisiae of bovine pancreatic trypsin inhibitor mutants lacking disulfide bonds is correlated with thermodynamic stability. *Biochemistry*, 37(5):1264–1273, 3 February 1998.

[165] Florian Krammer. SARS-CoV-2 vaccines in development. *Nature*, 586(7830):516–527, October 2020.

[166] Christoph Kreer, Matthias Zehner, Timm Weber, Meryem S Ercanoglu, Lutz Gieselmann, Cornelius Rohde, Sandro Halwe, Michael Korenkov, Philipp Schommers, Kanika Vanshylla, Veronica Di Cristanziano, Hanna Janicki, Reinhild Brinker, Artem Ashurov, Verena Krähling, Alexandra Kupke, Hadas Cohen-Dvashi, Manuel Koch, Jan Mathis

Eckert, Simone Lederer, Nico Pfeifer, Timo Wolf, Maria J G T Vehreschild, Clemens Wendtner, Ron Diskin, Henning Gruell, Stephan Becker, and Florian Klein. Longitudinal isolation of potent Near-Germline SARS-CoV-2-Neutralizing antibodies from COVID-19 patients. *Cell*, 182(4):843–854.e12, 20 August 2020.

[167] Jeffrey R Kugelman, Johanny Kugelman-Tonos, Jason T Ladner, James Pettit, Carolyn M Keeton, Elyse R Nagle, Karla Y Garcia, Jeffrey W Froude, Ana I Kuehne, Jens H Kuhn, Sina Bavari, Larry Zeitlin, John M Dye, Gene G Olinger, Mariano Sanchez-Lockhart, and Gustavo F Palacios. Emergence of ebola virus escape variants in infected nonhuman primates treated with the MB-003 antibody cocktail. *Cell reports*, 12(12):2111–2120, 29 September 2015.

[168] Sanjeev Kumar, Anmol Chandele, and Amit Sharma. Current status of therapeutic monoclonal antibodies against SARS-CoV-2. *PLoS pathogens*, 17(9):e1009885, September 2021.

[169] Jasmin S Kutter, Monique I Spronken, Pieter L Fraaij, Ron Am Fouchier, and Sander Herfst. Transmission routes of respiratory viruses among humans. *Current opinion in virology*, 28:142–151, February 2018.

[170] Alona Kuzmina, Yara Khalaila, Olga Voloshin, Ayelet Keren-Naus, Liora Boehm-Cohen, Yael Raviv, Yonat Shemer-Avni, Elli Rosenberg, and Ran Taube. SARS-CoV-2 spike variants exhibit differential infectivity and neutralization resistance to convalescent or post-vaccination sera. *Cell host & microbe*, 29(4):522–528.e2, 14 April 2021.

[171] Tommy Tsan-Yuk Lam, Na Jia, Ya-Wei Zhang, Marcus Ho-Hin Shum, Jia-Fu Jiang, Hua-Chen Zhu, Yi-Gang Tong, Yong-Xia Shi, Xue-Bing Ni, Yun-Shi Liao, Wen-Juan Li, Bao-Gui Jiang, Wei Wei, Ting-Ting Yuan, Kui Zheng, Xiao-Ming Cui, Jie Li, Guang-Qian Pei, Xin Qiang, William Yiu-Man Cheung, Lian-Feng Li, Fang-Fang Sun, Si Qin, Ji-Cheng Huang, Gabriel M Leung, Edward C Holmes, Yan-Ling Hu, Yi Guan, and Wu-Chun Cao. Identifying SARS-CoV-2-related coronaviruses in malayan pangolins. *Nature*, 583(7815):282–285, July 2020.

[172] Jun Lan, Jiwan Ge, Jinfang Yu, Sisi Shan, Huan Zhou, Shilong Fan, Qi Zhang, Xuanling Shi, Qisheng Wang, Linqi Zhang, and Xinquan Wang. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807):215–220, May 2020.

[173] Yifei Lang, Wentao Li, Zeshi Li, Danielle Koerhuis, Arthur C S van den Burg, Erik Rozemuller, Berend-Jan Bosch, Frank J M van Kuppeveld, Geert-Jan Boons, Eric G

Huizinga, Hilde M van der Schaar, and Raoul J de Groot. Coronavirus hemagglutinin-esterase and spike proteins coevolve for functional balance and optimal virion avidity. *Proceedings of the National Academy of Sciences of the United States of America*, 117(41):25759–25770, 13 October 2020.

[174] Matthew T Laurie, Jamin Liu, Sara Sunshine, James Peng, Douglas Black, Anthea M Mitchell, Sabrina A Mann, Genay Pilarowski, Kelsey C Zorn, Luis Rubio, Sara Bravo, Carina Marquez, Maya Petersen, Diane Havlir, and Joseph DeRisi. Exposures to different SARS-CoV-2 spike variants elicit neutralizing antibody responses with differential specificity towards established and emerging strains. 12 September 2021.

[175] Nina Le Bert, Anthony T Tan, Kamini Kunasegaran, Christine Y L Tham, Morteza Hafezi, Adeline Chia, Melissa Hui Yen Chng, Meiyin Lin, Nicole Tan, Martin Linster, Wan Ni Chia, Mark I-Cheng Chen, Lin-Fa Wang, Eng Eong Ooi, Shirin Kalimuddin, Paul Anantharajah Tambyah, Jenny Guek-Hong Low, Yee-Joo Tan, and Antonio Bertoletti. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature*, 584(7821):457–462, August 2020.

[176] Juhye M Lee, Rachel Eguia, Seth J Zost, Saket Choudhary, Patrick C Wilson, Trevor Bedford, Terry Stevens-Ayers, Michael Boeckh, Aeron C Hurt, Seema S Lakdawala, Scott E Hensley, and Jesse D Bloom. Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *eLife*, 8, 27 August 2019.

[177] Juhye M Lee, John Huddleston, Michael B Doud, Kathryn A Hooper, Nicholas C Wu, Trevor Bedford, and Jesse D Bloom. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proceedings of the National Academy of Sciences of the United States of America*, 115(35):E8276–E8285, 28 August 2018.

[178] Florian A Lempp, Leah Soriaga, Martin Montiel-Ruiz, Fabio Benigni, Julia Noack, Young-Jun Park, Siro Bianchi, Alexandra C Walls, John E Bowen, Jiayi Zhou, Hannah Kaiser, Maria Agostini, Marcel Meury, Exequiel Dellota, Stefano Jaconi, Elisabetta Cameroni, Herbert W Virgin, Antonio Lanzavecchia, David Veesler, Lisa Purcell, Amalio Telenti, and Davide Corti. Membrane lectins enhance SARS-CoV-2 infection and influence the neutralizing activity of different classes of antibodies. 4 April 2021.

[179] Michael Letko, Andrea Marzi, and Vincent Munster. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature microbiology*, 5(4):562–569, April 2020.

[180] Fang Li, Wenhui Li, Michael Farzan, and Stephen C Harrison. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*, 309(5742):1864–1868, 16 September 2005.

[181] Heng Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 16 March 2013.

[182] Qianqian Li, Jiajing Wu, Jianhui Nie, Li Zhang, Huan Hao, Shuo Liu, Chenyan Zhao, Qi Zhang, Huan Liu, Lingling Nie, Haiyang Qin, Meng Wang, Qiong Lu, Xiaoyu Li, Qiyu Sun, Junkai Liu, Linqi Zhang, Xuguang Li, Weijin Huang, and Youchun Wang. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell*, 182(5):1284–1294.e9, 3 September 2020.

[183] Wenhui Li, Michael J Moore, Natalya Vasilieva, Jianhua Sui, Swee Kee Wong, Michael A Berne, Mohan Somasundaran, John L Sullivan, Katherine Luzuriaga, Thomas C Greenough, Hyeryun Choe, and Michael Farzan. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426(6965):450–454, 27 November 2003.

[184] Wenhui Li, Chengsheng Zhang, Jianhua Sui, Jens H Kuhn, Michael J Moore, Shiwen Luo, Swee-Kee Wong, I-Chueh Huang, Keming Xu, Natalya Vasilieva, Akikazu Murakami, Yaqing He, Wayne A Marasco, Yi Guan, Hyeryun Choe, and Michael Farzan. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *The EMBO journal*, 24(8):1634–1643, 20 April 2005.

[185] Wentao Li, Ruben J G Hulswit, Ivy Widjaja, V Stalin Raj, Ryan McBride, Wenjie Peng, W Widagdo, M Alejandra Tortorici, Brenda van Dieren, Yifei Lang, Jan W M van Lent, James C Paulson, Cornelis A M de Haan, Raoul J de Groot, Frank J M van Kuppeveld, Bart L Haagmans, and Berend-Jan Bosch. Identification of sialic acid-binding function for the middle east respiratory syndrome coronavirus spike glycoprotein. *Proceedings of the National Academy of Sciences of the United States of America*, 114(40):E8508–E8517, 3 October 2017.

[186] C C Linnemann, Jr. Measles vaccine: immunity, reinfection and revaccination. *American journal of epidemiology*, 97(6):365–371, June 1973.

[187] Chang Liu, Helen M Ginn, Wanwisa Dejnirattisai, Piyada Supasa, Beibei Wang, Aekkachai Tuekprakhon, Rungtiwa Nutalai, Daming Zhou, Alexander J Mentzer, Yuguang Zhao, Helen M E Duyvesteyn, César López-Camacho, Jose Slon-Campos, Thomas S Walter, Donal Skelly, Sile Ann Johnson, Thomas G Ritter, Chris Mason,

Sue Ann Costa Clemens, Felipe Gomes Naveca, Valdinete Nascimento, Fernanda Nascimento, Cristiano Fernandes da Costa, Paola Cristina Resende, Alex Pauvolid-Correa, Marilda M Siqueira, Christina Dold, Nigel Temperton, Tao Dong, Andrew J Pollard, Julian C Knight, Derrick Crook, Teresa Lambe, Elizabeth Clutterbuck, Sagida Bibi, Amy Flaxman, Mustapha Bittaye, Sandra Belij-Rammerstorfer, Sarah C Gilbert, Tariq Malik, Miles W Carroll, Paul Klenerman, Eleanor Barnes, Susanna J Dunachie, Vicky Baillie, Natali Serafin, Zanele Ditse, Kelly Da Silva, Neil G Paterson, Mark A Williams, David R Hall, Shabir Madhi, Marta C Nunes, Philip Goulder, Elizabeth E Fry, Juthathip Mongkolsapaya, Jingshan Ren, David I Stuart, and Gavin R Screaton. Reduced neutralization of SARS-CoV-2 b.1.617 by vaccine and convalescent serum. *Cell*, 184(16):4220–4236.e13, 5 August 2021.

[188] Lihong Liu, Pengfei Wang, Manoj S Nair, Jian Yu, Micah Rapp, Qian Wang, Yang Luo, Jasper F-W Chan, Vincent Sahi, Amir Figueroa, Xinzheng V Guo, Gabriele Cerutti, Jude Bimela, Jason Gorman, Tongqing Zhou, Zhiwei Chen, Kwok-Yung Yuen, Peter D Kwong, Joseph G Sodroski, Michael T Yin, Zizhang Sheng, Yaoxing Huang, Lawrence Shapiro, and David D Ho. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature*, 584(7821):450–456, August 2020.

[189] Zhuoming Liu, Laura A VanBlargan, Louis-Marie Bloyet, Paul W Rothlauf, Rita E Chen, Spencer Stumpf, Haiyan Zhao, John M Errico, Elitza S Theel, Mariel J Liebeskind, Brynn Alford, William J Buchser, Ali H Ellebedy, Daved H Fremont, Michael S Diamond, and Sean P J Whelan. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell host & microbe*, page 2020.11.06.372037, 27 January 2021.

[190] Andrea N Loes, Lauren E Gentles, Allison J Greaney, Katharine H D Crawford, and Jesse D Bloom. Attenuated influenza virions expressing the SARS-CoV-2 Receptor-Binding domain induce neutralizing antibodies in mice. *Viruses*, 12(9), 5 September 2020.

[191] Derrick Louz, Hans E Bergmans, Birgit P Loos, and Rob C Hoeben. Animal models in virus research: their utility and limitations. *Critical reviews in microbiology*, 39(4):325–361, November 2013.

[192] Guangwen Lu, Yawei Hu, Qihui Wang, Jianxun Qi, Feng Gao, Yan Li, Yanfang Zhang, Wei Zhang, Yuan Yuan, Jinku Bao, Buchang Zhang, Yi Shi, Jinghua Yan, and George F Gao. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature*, 500(7461):227–231, 8 August 2013.

[193] Carolina Lucas, Chantal B F Vogels, Inci Yildirim, Jessica E Rothman, Peiwen Lu, Valter Monteiro, Jeff R Gehlhausen, Melissa Campbell, Julio Silva, Alexandra Tabachnikova, Mario A Peña-Hernandez, M Catherine Muenker, Mallery I Breban, Joseph R Fauver, Subhasis Mohanty, Jiefang Huang, Yale SARS-CoV-2 Genomic Surveillance Initiative, Albert C Shaw, Albert I Ko, Saad B Omer, Nathan D Grubaugh, and Akiko Iwasaki. Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity. *Nature*, 11 October 2021.

[194] Carolina Lucas, Chantal B F Vogels, Inci Yildirim, Jessica E Rothman, Peiwen Lu, Valter Monteiro, Jeff R Gelhausen, Melissa Campbell, Julio Silva, Alexandra Tabachikova, M Catherine Muenker, Mallery I Breban, Joseph R Fauver, Subhasis Mohanty, Jiefang Huang, Claire Pearson, Anthony Muyombwe, Randy Downing, Jafar Razeq, Mary Petrone, Isabel Ott, Anne Watkins, Chaney Kalinich, Tara Alpert, Anderson Brito, Rebecca Earnest, Steven Murphy, Caleb Neal, Eva Laszlo, Ahmad Altajar, Irina Tikhonova, Christopher Castaldi, Shrikant Mane, Kaya Bilguvar, Nicholas Kerantzas, David Ferguson, Wade Schulz, Marie Landry, David Peaper, Albert C Shaw, Albert I Ko, Saad B Omer, Nathan D Grubaugh, Akiko Iwasaki, and Yale SARS-CoV-2 Genomic Surveillance Initiative. Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity in uninfected and previously infected individuals. 18 July 2021.

[195] Sheila F Lumley, Denise O'Donnell, Nicole E Stoesser, Philippa C Matthews, Alison Howarth, Stephanie B Hatch, Brian D Marsden, Stuart Cox, Tim James, Fiona Warren, Liam J Peck, Thomas G Ritter, Zoe de Toledo, Laura Warren, David Axten, Richard J Cornall, E Yvonne Jones, David I Stuart, Gavin Screaton, Daniel Ebner, Sarah Hoosdally, Meera Chand, Derrick W Crook, Anne-Marie O'Donnell, Christopher P Conlon, Koen B Pouwels, A Sarah Walker, Tim E A Peto, Susan Hopkins, Timothy M Walker, Katie Jeffery, David W Eyre, and Oxford University Hospitals Staff Testing Group. Antibody status and incidence of SARS-CoV-2 infection in health care workers. *The New England journal of medicine*, 23 December 2020.

[196] Oscar A MacLean, Spyros Lytras, Steven Weaver, Joshua B Singer, Maciej F Boni, Philippe Lemey, Sergei L Kosakovsky Pond, and David L Robertson. Natural selection in the evolution of SARS-CoV-2 in bats, not humans, created a highly capable human pathogen. *bioRxiv : the preprint server for biology*, 30 July 2020.

[197] Ikenna G Madu, Shoshannah L Roth, Sandrine Belouzard, and Gary R Whittaker. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *Journal of virology*, 83(15):7411–7421, August 2009.

[198] Kenneth A Matreyek, Lea M Starita, Jason J Stephany, Beth Martin, Melissa A Chiasson, Vanessa E Gray, Martin Kircher, Arineh Khechaduri, Jennifer N Dines, Ronald J Hause, Smita Bhatia, William E Evans, Mary V Relling, Wenjian Yang, Jay Shendure, and Douglas M Fowler. Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nature genetics*, 50(6):874–882, June 2018.

[199] Matthew McCallum, Jessica Bassi, Anna De Marco, Alex Chen, Alexandra C Walls, Julia Di Iulio, M Alejandra Tortorici, Mary-Jane Navarro, Chiara Silacci-Fregni, Christian Saliba, Kaitlin R Sprouse, Maria Agostini, Dora Pinto, Katja Culap, Siro Bianchi, Stefano Jaconi, Elisabetta Cameroni, John E Bowen, Sasha W Tilles, Matteo Samuele Pizzuto, Sonja Bernasconi Guastalla, Giovanni Bona, Alessandra Franzetti Pellanda, Christian Garzoni, Wesley C Van Voorhis, Laura E Rosen, Gyorgy Snell, Amalio Telenti, Herbert W Virgin, Luca Piccoli, Davide Corti, and David Veesler. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science*, 373(6555):648–654, 6 August 2021.

[200] Matthew McCallum, Anna De Marco, Florian A Lempp, M Alejandra Tortorici, Dora Pinto, Alexandra C Walls, Martina Beltramello, Alex Chen, Zhuoming Liu, Fabrizia Zatta, Samantha Zepeda, Julia di Iulio, John E Bowen, Martin Montiel-Ruiz, Jiayi Zhou, Laura E Rosen, Siro Bianchi, Barbara Guarino, Chiara Silacci Fregni, Rana Abdelnabi, Shi-Yan Caroline Foo, Paul W Rothlauf, Louis-Marie Bloyet, Fabio Benigni, Elisabetta Cameroni, Johan Neyts, Agostino Riva, Gyorgy Snell, Amalio Telenti, Sean P J Whelan, Herbert W Virgin, Davide Corti, Matteo Samuele Pizzuto, and David Veesler. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell*, 184(9):2332–2347.e16, 29 April 2021.

[201] Matthew McCallum, Alexandra C Walls, Kaitlin R Sprouse, John E Bowen, Laura Rosen, Ha V Dang, Anna deMarco, Nicholas Franko, Sasha W Tilles, Jennifer Logue, Marcos C Miranda, Margaret Ahlrichs, Lauren Carter, Gyorgy Snell, Matteo Samuele Pizzuto, Helen Y Chu, Wesley C Van Voorhis, Davide Corti, and David Veesler. Molecular basis of immune evasion by the delta and kappa SARS-CoV-2 variants. 11 August 2021.

[202] Kevin R McCarthy, Linda J Rennick, Sham Nambulli, Lindsey R Robinson-McCarthy, William G Bain, Ghady Haidar, and W Paul Duprex. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science*, 371(6534):1139–1142, 12 March 2021.

[203] Katherine McMahan, Jingyou Yu, Noe B Mercado, Carolin Loos, Lisa H Tostanoski, Abishek Chandrashekar, Jinyan Liu, Lauren Peter, Caroline Atyeo, Alex Zhu, Esther A

Bondzie, Gabriel Dagotto, Makda S Gebre, Catherine Jacob-Dolan, Zhenfeng Li, Felix Nampanya, Shivani Patel, Laurent Pessaint, Alex Van Ry, Kelvin Blade, Jake Yalley-Ogunro, Mehtap Cabus, Renita Brown, Anthony Cook, Elyse Teow, Hanne Andersen, Mark G Lewis, Douglas A Lauffenburger, Galit Alter, and Dan H Barouch. Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature*, 590(7847):630–634, February 2021.

[204] Meagan McMahon, Shirin Strohmeier, Madhusudan Rajendran, Christina Capuano, Ali H Ellebedy, Patrick C Wilson, and Florian Krammer. Correctly folded - but not necessarily functional - influenza virus neuraminidase is required to induce protective antibody responses in mice. *Vaccine*, 15 September 2020.

[205] Vineet D Menachery, Boyd L Yount, Jr, Kari Debbink, Sudhakar Agnihothram, Lisa E Gralinski, Jessica A Plante, Rachel L Graham, Trevor Scobey, Xing-Yi Ge, Eric F Donaldson, Scott H Randell, Antonio Lanzavecchia, Wayne A Marasco, Zhengli-Li Shi, and Ralph S Baric. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature medicine*, 21(12):1508–1513, December 2015.

[206] Vineet D Menachery, Boyd L Yount, Jr, Amy C Sims, Kari Debbink, Sudhakar S Agnihothram, Lisa E Gralinski, Rachel L Graham, Trevor Scobey, Jessica A Plante, Scott R Royal, Jesica Swanstrom, Timothy P Sheahan, Raymond J Pickles, Davide Corti, Scott H Randell, Antonio Lanzavecchia, Wayne A Marasco, and Ralph S Baric. SARS-like WIV1-CoV poised for human emergence. *Proceedings of the National Academy of Sciences of the United States of America*, 113(11):3048–3053, 15 March 2016.

[207] Bo Meng, Steven A Kemp, Guido Papa, Rawlings Datir, Isabella A T M Ferreira, Sara Marelli, William T Harvey, Spyros Lytras, Ahmed Mohamed, Giulia Gallo, Nazia Thakur, Dami A Collier, Petra Mlcochova, COVID-19 Genomics UK (COG-UK) Consortium, Lidia M Duncan, Alessandro M Carabelli, Julia C Kenyon, Andrew M Lever, Anna De Marco, Christian Saliba, Katja Culap, Elisabetta Cameroni, Nicholas J Matheson, Luca Piccoli, Davide Corti, Leo C James, David L Robertson, Dalan Bailey, and Ravindra K Gupta. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the alpha variant b.1.1.7. *Cell reports*, 35(13):109292, 29 June 2021.

[208] Jean Kaoru Millet and Gary R Whittaker. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus research*, 202:120–134, 16 April 2015.

[209] Petra Mlcochova, Steven Kemp, Mahesh Shanker Dhar, Guido Papa, Bo Meng, Isabella A T M Ferreira, Rawlings Datir, Dami A Collier, Anna Albecka, Sujeet Singh, Rajesh Pandey, Jonathan Brown, Jie Zhou, Niluka Goonawardane, Swapnil Mishra,

Charles Whittaker, Thomas Mellan, Robin Marwal, Meena Datta, Shantanu Sengupta, Kalaiarasan Ponnusamy, Venkatraman Srinivasan Radhakrishnan, Adam Abdullahi, Oscar Charles, Partha Chattopadhyay, Priti Devi, Daniela Caputo, Tom Peacock, Dr Chand Wattal, Neeraj Goel, Ambrish Satwik, Raju Vaishya, Meenakshi Agarwal, Indian SARS-CoV-2 Genomics Consortium (INSACOG), Genotype to Phenotype Japan (G2P-Japan) Consortium, CITIID-NIHR BioResource COVID-19 Collaboration, Antranik Mavousian, Joo Hyeon Lee, Jessica Bassi, Chiara Silacci-Fegni, Christian Saliba, Dora Pinto, Takashi Irie, Isao Yoshida, William L Hamilton, Kei Sato, Samir Bhatt, Seth Flaxman, Leo C James, Davide Corti, Luca Piccoli, Wendy S Barclay, Partha Rakshit, Anurag Agrawal, and Ravindra K Gupta. SARS-CoV-2 b.1.617.2 delta variant replication and immune evasion. *Nature*, 6 September 2021.

[210] S Momsen Reincke, Meng Yuan, Hans-Christian Kornau, Victor M Corman, Scott van Hoof, Elisa Sánchez-Sendin, Melanie Ramberger, Wenli Yu, Yuanzi Hua, Henry Tien, Marie Luisa Schmidt, Tatjana Schwarz, Lara Maria Jeworowski, Sarah E Brandl, Helle Foverskov Rasmussen, Marie A Homeyer, Laura Stöffler, Martin Barner, Désirée Kunkel, Shufan Huo, Johannes Horler, Niels von Wardenburg, Inge Kroidl, Tabea M Eser, Andreas Wieser, Christof Geldmacher, Michael Hoelscher, Hannes Gänzer, Günter Weiss, Dietmar Schmitz, Christian Drosten, Harald Prüss, Ian A Wilson, and Jakob Kreye. SARS-CoV-2 beta variant infection elicits potent lineage-specific and cross-reactive antibodies. 30 September 2021.

[211] Thandeka Moyo-Gwete, Mashudu Madzivhandila, Zanele Makhado, Frances Ayres, Donald Mhlanga, Brent Oosthuysen, Bronwen E Lambson, Prudence Kgagudi, Houriiyah Tegally, Arash Iranzadeh, Deelan Doolabh, Lynn Tyers, Lionel R Chinhoyi, Mathilda Mennen, Sango Skelem, Gert Marais, Constantinos K Wibmer, Jinal N Bhiman, Veronica Ueckermann, Theresa Rossouw, Michael Boswell, Tulio de Oliveira, Carolyn Williamson, Wendy A Burgers, Ntobeko Ntusi, Lynn Morris, and Penny L Moore. Cross-Reactive neutralizing antibody responses elicited by SARS-CoV-2 501Y.V2 (b.1.351). *The New England journal of medicine*, (NEJMc2104192):2021.03.06.434193, 7 April 2021.

[212] Frauke Muecksch, Yiska Weisblum, Christopher Barnes, Fabian Schmidt, Dennis Schaefer-Babajew, Julio Lorenzi, Andrew Flyak, Andrew DeLaitsch, Kathryn Huey-Tubman, Shurong Hou, Celia Schiffer, Christian Gaebler, Zijun Wang, Justin Da Silva, Daniel Poston, Shlomo Finkin, Alice Cho, Melissa Cipolla, Thiago Oliveira, Katrina Millard, Victor Ramos, Anna Gazumyan, Magdalena Rutkowska, Marina Caskey, Michel Nussenzweig, Pamela Bjorkman, Theodora Hatziioannou, and Paul Bieniasz. Development of potency, breadth and resilience to viral escape mutations in SARS-CoV-2 neutralizing antibodies. *bioRxiv*, page 2021.03.07.434227, 8 March 2021.

[213] Alexander Muik, Ann-Kathrin Wallisch, Bianca Saenger, Kena A Swanson, Julia Muehl, Wei Chen, Hui Cai, Ritu Sarkar, Oezlem Tuereci, Philip R Dormitzer, and Ugur Sahin. Neutralization of SARS-CoV-2 lineage b.1.1.7 pseudovirus by BNT162b2 vaccine-elicited human sera. 19 January 2021.

[214] Julia L Mullen, Ginger Tsueng, Alaa Abdel Latif, Manar Alkuzweny, Marco Cano, Emily Haag, Jerry Zhou, Mark Zeller, Emory Hufbauer, Nate Matteson, Kristian G Andersen, Chunlei Wu, Andrew I Su, Karthik Gangavarapu, Laura D Hughes, and Center for Viral Systems Biology. outbreak.info. `https://outbreak.info/`, 2020. Accessed: 2021-10-24.

[215] Nicola F Müller, Cassia Wagner, Chris D Frazar, Pavitra Roychoudhury, Jover Lee, Louise H Moncla, Benjamin Pelle, Matthew Richardson, Erica Ryke, Hong Xie, Lasata Shrestha, Amin Addetia, Victoria M Rachleff, Nicole A P Lieberman, Meei-Li Huang, Romesh Gautom, Geoff Melly, Brian Hiatt, Philip Dykema, Amanda Adler, Elisabeth Brandstetter, Peter D Han, Kairsten Fay, Misja Ilcisin, Kirsten Lacombe, Thomas R Sibley, Melissa Truong, Caitlin R Wolf, Michael Boeckh, Janet A Englund, Michael Famulare, Barry R Lutz, Mark J Rieder, Matthew Thompson, Jeffrey S Duchin, Lea M Starita, Helen Y Chu, Jay Shendure, Keith R Jerome, Scott Lindquist, Alexander L Greninger, Deborah A Nickerson, and Trevor Bedford. Viral genomes reveal patterns of the SARS-CoV-2 outbreak in washington state. *Science translational medicine*, 13(595), 26 May 2021.

[216] Mark J Mulligan, Kirsten E Lyke, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, Kathleen Neuzil, Vanessa Raabe, Ruth Bailey, Kena A Swanson, Ping Li, Kenneth Koury, Warren Kalina, David Cooper, Camila Fontes-Garfias, Pei-Yong Shi, Özlem Türeci, Kristin R Tompkins, Edward E Walsh, Robert Frenck, Ann R Falsey, Philip R Dormitzer, William C Gruber, Uğur Şahin, and Kathrin U Jansen. Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature*, 586(7830):589–593, October 2020.

[217] Miguel Ángel Muñoz-Alía, Rebecca A Nace, Lianwen Zhang, and Stephen J Russell. Serotypic evolution of measles virus is constrained by multiple co-dominant B cell epitopes on its surface glycoproteins. *Cell reports. Medicine*, 2(4):100225, 20 April 2021.

[218] Sandra C A Nielsen, Fan Yang, Katherine J L Jackson, Ramona A Hoh, Katharina Röltgen, Grace H Jean, Bryan A Stevens, Ji-Yeun Lee, Arjun Rustagi, Angela J Rogers, Abigail E Powell, Molly Hunter, Javaria Najeeb, Ana R Otrelo-Cardoso,

Kathryn E Yost, Bence Daniel, Kari C Nadeau, Howard Y Chang, Ansuman T Satpathy, Theodore S Jardetzky, Peter S Kim, Taia T Wang, Benjamin A Pinsky, Catherine A Blish, and Scott D Boyd. Human B cell clonal expansion and convergent antibody responses to SARS-CoV-2. *Cell host & microbe*, 3 September 2020.

[219] Eri Nobusawa, Katsumi Omagari, Setsuko Nakajima, and Katsuhisa Nakajima. Reactivity of human convalescent sera with influenza virus hemagglutinin protein mutants at antigenic site a. *Microbiology and immunology*, 56(2):99–106, February 2012.

[220] Jakub Otwinowski, David M McCandlish, and Joshua B Plotkin. Inferring the shape of global epistasis. *Proceedings of the National Academy of Sciences of the United States of America*, 115(32):E7550–E7558, 7 August 2018.

[221] Bas B Oude Munnink, Reina S Sikkema, David F Nieuwenhuijse, Robert Jan Molenaar, Emmanuelle Munger, Richard Molenkamp, Arco van der Spek, Paulien Tolsma, Ariene Rietveld, Miranda Brouwer, Noortje Bouwmeester-Vincken, Frank Harders, Renate Hakze-van der Honing, Marjolein C A Wegdam-Blans, Ruth J Bouwstra, Corine GeurtsvanKessel, Annemiek A van der Eijk, Francisca C Velkers, Lidwien A M Smit, Arjan Stegeman, Wim H M van der Poel, and Marion P G Koopmans. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science*, 371(6525):172–177, 8 January 2021.

[222] Bas B Oude Munnink, Nathalie Worp, David F Nieuwenhuijse, Reina S Sikkema, Bart Haagmans, Ron A M Fouchier, and Marion Koopmans. The next phase of SARS-CoV-2 surveillance: real-time molecular epidemiology. *Nature medicine*, 27(9):1518–1524, September 2021.

[223] Our World in Data. Cumulative confirmed COVID-19 cases. `https://ourworldindata.org/grapher/cumulative-covid-cases-region`. Accessed: 2021-10-22.

[224] John E Pak, Chetna Sharon, Malathy Satkunarajah, Thierry C Auperin, Cheryl M Cameron, David J Kelvin, Jayaraman Seetharaman, Alan Cochrane, Francis A Plummer, Jody D Berry, and James M Rini. Structural insights into immune recognition of the severe acute respiratory syndrome coronavirus S protein receptor binding domain. *Journal of molecular biology*, 388(4):815–823, 15 May 2009.

[225] PANUM and PL. Observation made during the epidemic of measles on the faroe islands in the year 1846. *Med Classics*, 3:839–886, 1939.

[226] Norbert Pardi, Michael J Hogan, Frederick W Porter, and Drew Weissman. mRNA vaccines - a new era in vaccinology. *Nature reviews. Drug discovery*, 17(4):261–279, April 2018.

[227] Young-Jun Park, Alexandra C Walls, Zhaoqian Wang, Maximillian M Sauer, Wentao Li, M Alejandra Tortorici, Berend-Jan Bosch, Frank DiMaio, and David Veesler. Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. *Nature structural & molecular biology*, 26(12):1151–1157, December 2019.

[228] Neil Peterman and Erel Levine. Sort-seq under the hood: implications of design choices on large-scale characterization of sequence-function relations. *BMC genomics*, 17:206, 9 March 2016.

[229] Luca Piccoli, Young-Jun Park, M Alejandra Tortorici, Nadine Czudnochowski, Alexandra C Walls, Martina Beltramello, Chiara Silacci-Fregni, Dora Pinto, Laura E Rosen, John E Bowen, Oliver J Acton, Stefano Jaconi, Barbara Guarino, Andrea Minola, Fabrizia Zatta, Nicole Sprugasci, Jessica Bassi, Alessia Peter, Anna De Marco, Jay C Nix, Federico Mele, Sandra Jovic, Blanca Fernandez Rodriguez, Sneha V Gupta, Feng Jin, Giovanni Piumatti, Giorgia Lo Presti, Alessandra Franzetti Pellanda, Maira Biggiogero, Maciej Tarkowski, Matteo S Pizzuto, Elisabetta Cameroni, Colin Havenar-Daughton, Megan Smithey, David Hong, Valentino Lepori, Emiliano Albanese, Alessandro Ceschi, Enos Bernasconi, Luigia Elzi, Paolo Ferrari, Christian Garzoni, Agostino Riva, Gyorgy Snell, Federica Sallusto, Katja Fink, Herbert W Virgin, Antonio Lanzavecchia, Davide Corti, and David Veesler. Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike Receptor-Binding domain by Structure-Guided High-Resolution serology. *Cell*, 183(4):1024–1042.e21, 12 November 2020.

[230] Dora Pinto, Young-Jun Park, Martina Beltramello, Alexandra C Walls, M Alejandra Tortorici, Siro Bianchi, Stefano Jaconi, Katja Culap, Fabrizia Zatta, Anna De Marco, Alessia Peter, Barbara Guarino, Roberto Spreafico, Elisabetta Cameroni, James Brett Case, Rita E Chen, Colin Havenar-Daughton, Gyorgy Snell, Amalio Telenti, Herbert W Virgin, Antonio Lanzavecchia, Michael S Diamond, Katja Fink, David Veesler, and Davide Corti. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature*, 583(7815):290–295, July 2020.

[231] Dora Pinto, Maximilian M Sauer, Nadine Czudnochowski, Jun Siong Low, M Alejandra Tortorici, Michael P Housley, Julia Noack, Alexandra C Walls, John E Bowen, Barbara Guarino, Laura E Rosen, Julia di Iulio, Josipa Jerak, Hannah Kaiser, Saiful Islam, Stefano Jaconi, Nicole Sprugasci, Katja Culap, Rana Abdelnabi, Caroline

Foo, Lotte Coelmont, Istvan Bartha, Siro Bianchi, Chiara Silacci-Fregni, Jessica Bassi, Roberta Marzi, Eneida Vetti, Antonino Cassotta, Alessandro Ceschi, Paolo Ferrari, Pietro E Cippà, Olivier Giannini, Samuele Ceruti, Christian Garzoni, Agostino Riva, Fabio Benigni, Elisabetta Cameroni, Luca Piccoli, Matteo S Pizzuto, Megan Smithey, David Hong, Amalio Telenti, Florian A Lempp, Johan Neyts, Colin Havenar-Daughton, Antonio Lanzavecchia, Federica Sallusto, Gyorgy Snell, Herbert W Virgin, Martina Beltramello, Davide Corti, and David Veesler. Broad betacoronavirus neutralization by a stem helix-specific human antibody. *Science*, 3 August 2021.

[232] Delphine Planas, Timothée Bruel, Ludivine Grzelak, Florence Guivel-Benhassine, Isabelle Staropoli, Françoise Porrot, Cyril Planchais, Julian Buchrieser, Maaran Michael Rajah, Elodie Bishop, Mélanie Albert, Flora Donati, Matthieu Prot, Sylvie Behillil, Vincent Enouf, Marianne Maquart, Mounira Smati-Lafarge, Emmanuelle Varon, Frédérique Schortgen, Layla Yahyaoui, Maria Gonzalez, Jérôme De Sèze, Hélène Péré, David Veyer, Aymeric Sève, Etienne Simon-Lorière, Samira Fafi-Kremer, Karl Stefic, Hugo Mouquet, Laurent Hocqueloux, Sylvie van der Werf, Thierry Prazuck, and Olivier Schwartz. Sensitivity of infectious SARS-CoV-2 b.1.1.7 and b.1.351 variants to neutralizing antibodies. *Nature medicine*, 27(5):917–924, May 2021.

[233] Delphine Planas, David Veyer, Artem Baidaliuk, Isabelle Staropoli, Florence Guivel-Benhassine, Maaran Michael Rajah, Cyril Planchais, Françoise Porrot, Nicolas Robillard, Julien Puech, Matthieu Prot, Floriane Gallais, Pierre Gantner, Aurélie Velay, Julien Le Guen, Najiby Kassis-Chikhani, Dhiaeddine Edriss, Laurent Belec, Aymeric Seve, Laura Courtellemont, Hélène Péré, Laurent Hocqueloux, Samira Fafi-Kremer, Thierry Prazuck, Hugo Mouquet, Timothée Bruel, Etienne Simon-Lorière, Felix A Rey, and Olivier Schwartz. Reduced sensitivity of SARS-CoV-2 variant delta to antibody neutralization. *Nature*, 596(7871):276–280, August 2021.

[234] Lyubov Popova, Kenneth Smith, Ann H West, Patrick C Wilson, Judith A James, Linda F Thompson, and Gillian M Air. Immunodominance of antigenic site B over site a of hemagglutinin of recent H3N2 influenza viruses. *PloS one*, 7(7):e41895, 25 July 2012.

[235] A R Poteete, D Rennell, S E Bouvier, and L W Hardy. Alteration of T4 lysozyme structure by second-site reversion of deleterious mutations. *Protein science: a publication of the Protein Society*, 6(11):2418–2425, November 1997.

[236] Ponraj Prabakaran, Jianhua Gan, Yang Feng, Zhongyu Zhu, Vidita Choudhry, Xiaodong Xiao, Xinhua Ji, and Dimiter S Dimitrov. Structure of severe acute respiratory

syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *The Journal of biological chemistry*, 281(23):15829–15836, 9 June 2006.

[237] Public Health England. *Investigation of novel SARS-COV-2 variant: Variant of Concern 202012/01*. December 2020.

[238] Public Health England. *SARS-CoV-2 variants of concern and variants under investigation in England*. 22 April 2021.

[239] Public Health England. SARS-CoV-2 variants of concern and variants under investigation in england: technical briefing. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993879/Variants_of_Concern_VOC_Technical_Briefing_15.pdf`, 11 June 2021. Accessed: 2021-10-24.

[240] Public Health England. SARS-CoV-2 variants of concern and variants under investigation in england: technical briefing. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1028113/Technical_Briefing_26.pdf`, 22 October 2021. Accessed: 2021-10-24.

[241] Arjun Puranik, Patrick J Lenehan, John C O'Horo, Michiel J M Niesen, Abinash Virk, Melanie D Swift, Walter Kremers, A J Venkatakrishnan, Joel E Gordon, Holly L Geyer, Leigh Lewis Speicher, Venky Soundararajan, and Andrew D Badley. Durability analysis of the highly effective BNT162b2 vaccine against COVID-19. 7 September 2021.

[242] Oliver G Pybus, Andrew Rambaut, Robert Belshaw, Robert P Freckleton, Alexei J Drummond, and Edward C Holmes. Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Molecular biology and evolution*, 24(3):845–852, March 2007.

[243] V Stalin Raj, Huihui Mou, Saskia L Smits, Dick H W Dekkers, Marcel A Müller, Ronald Dijkman, Doreen Muth, Jeroen A A Demmers, Ali Zaki, Ron A M Fouchier, Volker Thiel, Christian Drosten, Peter J M Rottier, Albert D M E Osterhaus, Berend Jan Bosch, and Bart L Haagmans. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*, 495(7440):251–254, 14 March 2013.

[244] A Rambaut, N Loman, O Pybus, W Barclay, J Barrett, A Carabelli, T Connor, T Peacock, D L Robertson, E Volz, and COVID-19 Genomics Consortium UK (CoG-UK). *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations*. Virological, 18 December 2020.

[245] Micah Rapp, Yicheng Guo, Eswar R Reddem, Jian Yu, Lihong Liu, Pengfei Wang, Gabriele Cerutti, Phinikoula Katsamba, Jude S Bimela, Fabiana A Bahna, Seetha M Mannepalli, Baoshan Zhang, Peter D Kwong, Yaoxing Huang, David D Ho, Lawrence Shapiro, and Zizhang Sheng. Modular basis for potent SARS-CoV-2 neutralization by a prevalent VH1-2-derived antibody class. *Cell reports*, 35(1):108950, 6 April 2021.

[246] Matthew I J Raybould, Aleksandr Kovaltsuk, Claire Marks, and Charlotte M Deane. CoV-AbDab: the coronavirus antibody database. *Bioinformatics*, 37(5):734–735, 5 May 2021.

[247] S E Reed. The behaviour of recent isolates of human respiratory coronavirus in vitro and in volunteers: evidence of heterogeneity among 229e-related strains. *Journal of medical virology*, 13(2):179–192, 1984.

[248] Chloe Rees-Spear, Luke Muir, Sarah A Griffith, Judith Heaney, Yoann Aldon, Jonne Snitselaar, Peter Thomas, Carl Graham, Jeffrey Seow, Nayung Lee, Annachiara Rosa, Chloe Roustan, Catherine F Houlihan, Rogier W Sanders, Ravindra K Gupta, Peter Cherepanov, Hans Stauss, Eleni Nastouli, Katie J Doores, Marit J van Gils, and Laura E McCoy. The impact of spike mutations on SARS-CoV-2 neutralization. 19 January 2021.

[249] Wuze Ren, Xiuxia Qu, Wendong Li, Zhenggang Han, Meng Yu, Peng Zhou, Shu-Yi Zhang, Lin-Fa Wang, Hongkui Deng, and Zhengli Shi. Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. *Journal of virology*, 82(4):1899–1907, February 2008.

[250] Alex Renn, Ying Fu, Xin Hu, Matthew D Hall, and Anton Simeonov. Fruitful neutralizing antibody pipeline brings hope to defeat SARS-Cov-2. *Trends in pharmacological sciences*, 41(11):815–829, November 2020.

[251] Davide F Robbiani, Christian Gaebler, Frauke Muecksch, Julio C C Lorenzi, Zijun Wang, Alice Cho, Marianna Agudelo, Christopher O Barnes, Anna Gazumyan, Shlomo Finkin, Thomas Hägglöf, Thiago Y Oliveira, Charlotte Viant, Arlene Hurley, Hans-Heinrich Hoffmann, Katrina G Millard, Rhonda G Kost, Melissa Cipolla, Kristie Gordon, Filippo Bianchini, Spencer T Chen, Victor Ramos, Roshni Patel, Juan Dizon, Irina Shimeliovich, Pilar Mendoza, Harald Hartweger, Lilian Nogueira, Maggi Pack, Jill Horowitz, Fabian Schmidt, Yiska Weisblum, Eleftherios Michailidis, Alison W Ashbrook, Eric Waltari, John E Pak, Kathryn E Huey-Tubman, Nicholas Koranda, Pauline R Hoffman, Anthony P West, Jr, Charles M Rice, Theodora Hatziioannou, Pamela J Bjorkman, Paul D Bieniasz, Marina Caskey, and Michel C Nussenzweig.

Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature*, 584(7821):437–442, August 2020.

[252] Barry Rockx, Eric Donaldson, Matthew Frieman, Timothy Sheahan, Davide Corti, Antonio Lanzavecchia, and Ralph S Baric. Escape from human monoclonal antibody neutralization affects in vitro and in vivo fitness of severe acute respiratory syndrome coronavirus. *The Journal of infectious diseases*, 201(6):946–955, 15 March 2010.

[253] Lauren B Rodda, Jason Netland, Laila Shehata, Kurt B Pruner, Peter A Morawski, Christopher D Thouvenel, Kennidy K Takehara, Julie Eggenberger, Emily A Hemann, Hayley R Waterman, Mitchell L Fahning, Yu Chen, Malika Hale, Jennifer Rathe, Caleb Stokes, Samuel Wrenn, Brooke Fiala, Lauren Carter, Jessica A Hamerman, Neil P King, Michael Gale, Jr, Daniel J Campbell, David J Rawlings, and Marion Pepper. Functional SARS-CoV-2-Specific immune memory persists after mild COVID-19. *Cell*, 23 November 2020.

[254] Katharina Roeltgen, Sandra C A Nielsen, Prabhu S Arunachalam, Fan Yang, Ramona A Hoh, Oliver F Wirz, Alexandra S Lee, Fei Gao, Vamsee Mallajosyula, Chunfeng Li, Emily Haraguchi, Massa J Shoura, James L Wilbur, Jacob N Wohlstadter, Mark M Davis, Benjamin A Pinsky, George B Sigal, Bali Pulendran, Kari C Nadeau, and Scott D Boyd. mRNA vaccination compared to infection elicits an IgG-predominant response with greater SARS-CoV-2 specificity and similar decrease in variant spike recognition. *medRxiv*, page 2021.04.05.21254952, 7 April 2021.

[255] Thomas F Rogers, Fangzhu Zhao, Deli Huang, Nathan Beutler, Alison Burns, Wan-Ting He, Oliver Limbo, Chloe Smith, Ge Song, Jordan Woehl, Linlin Yang, Robert K Abbott, Sean Callaghan, Elijah Garcia, Jonathan Hurtado, Mara Parren, Linghang Peng, Sydney Ramirez, James Ricketts, Michael J Ricciardi, Stephen A Rawlings, Nicholas C Wu, Meng Yuan, Davey M Smith, David Nemazee, John R Teijaro, James E Voss, Ian A Wilson, Raiees Andrabi, Bryan Briney, Elise Landais, Devin Sok, Joseph G Jardine, and Dennis R Burton. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science*, 369(6506):956–963, 21 August 2020.

[256] Colin A Russell, Peter M Kasson, Ruben O Donis, Steven Riley, John Dunbar, Andrew Rambaut, Jason Asher, Stephen Burke, C Todd Davis, Rebecca J Garten, Sandrasegaram Gnanakaran, Simon I Hay, Sander Herfst, Nicola S Lewis, James O Lloyd-Smith, Catherine A Macken, Sebastian Maurer-Stroh, Elizabeth Neuhaus, Colin R Parrish, Kim M Pepin, Samuel S Shepard, David L Smith, David L Suarez, Susan C

Trock, Marc-Alain Widdowson, Dylan B George, Marc Lipsitch, and Jesse D Bloom. Improving pandemic influenza risk assessment. *eLife*, 3:e03883, 16 October 2014.

[257] Jerald Sadoff, Mathieu Le Gars, Georgi Shukarev, Dirk Heerwegh, Carla Truyers, Anne M de Groot, Jeroen Stoop, Sarah Tete, Wim Van Damme, Isabel Leroux-Roels, Pieter-Jan Berghmans, Murray Kimmel, Pierre Van Damme, Jan de Hoon, William Smith, Kathryn E Stephenson, Stephen C De Rosa, Kristen W Cohen, M Juliana McElrath, Emmanuel Cormier, Gert Scheper, Dan H Barouch, Jenny Hendriks, Frank Struyf, Macaya Douoguih, Johan Van Hoof, and Hanneke Schuitemaker. Interim results of a phase 1-2a trial of Ad26.COV2.S covid-19 vaccine. *The New England journal of medicine*, 384(19):1824–1835, 13 May 2021.

[258] Matthew R Sandbulte, Kim B Westgeest, Jin Gao, Xiyan Xu, Alexander I Klimov, Colin A Russell, David F Burke, Derek J Smith, Ron A M Fouchier, and Maryna C Eichelberger. Discordant antigenic drift of neuraminidase and hemagglutinin in H1N1 and H3N2 influenza viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51):20748–20753, 20 December 2011.

[259] Rafael Sanjuán, Miguel R Nebot, Nicola Chirico, Louis M Mansky, and Robert Belshaw. Viral mutation rates. *Journal of virology*, 84(19):9733–9748, October 2010.

[260] K Sato, T Morishita, E Nobusawa, K Tonegawa, K Sakae, S Nakajima, and K Nakajima. Amino-acid change on the antigenic region B1 of H3 haemagglutinin may be a trigger for the emergence of drift strain of influenza a virus. *Epidemiology and infection*, 132(3):399–406, June 2004.

[261] Alexandra Schäfer, Frauke Muecksch, Julio C C Lorenzi, Sarah R Leist, Melissa Cipolla, Stylianos Bournazos, Fabian Schmidt, Rachel M Maison, Anna Gazumyan, David R Martinez, Ralph S Baric, Davide F Robbiani, Theodora Hatziioannou, Jeffrey V Ravetch, Paul D Bieniasz, Richard A Bowen, Michel C Nussenzweig, and Timothy P Sheahan. Antibody potency, effector function, and combinations in protection and therapy for SARS-CoV-2 infection in vivo. *The Journal of experimental medicine*, 218(3), 1 March 2021.

[262] Aaron G Schmidt, Matthew D Therkelsen, Shaun Stewart, Thomas B Kepler, Hua-Xin Liao, M Anthony Moody, Barton F Haynes, and Stephen C Harrison. Viral receptor-binding site antibodies with diverse germline origins. *Cell*, 161(5):1026–1034, 21 May 2015.

[263] O W Schmidt, I D Allan, M K Cooney, H M Foy, and J P Fox. Rises in titers of antibody to human coronaviruses OC43 and 229E in seattle families during 1975-1979. *American journal of epidemiology*, 123(5):862–868, May 1986.

[264] Philipp Schommers, Henning Gruell, Morgan E Abernathy, My-Kim Tran, Adam S Dingens, Harry B Gristick, Christopher O Barnes, Till Schoofs, Maike Schlotz, Kanika Vanshylla, Christoph Kreer, Daniela Weiland, Udo Holtick, Christof Scheid, Markus M Valter, Marit J van Gils, Rogier W Sanders, Jörg J Vehreschild, Oliver A Cornely, Clara Lehmann, Gerd Fätkenheuer, Michael S Seaman, Jesse D Bloom, Pamela J Bjorkman, and Florian Klein. Restriction of HIV-1 escape by a highly broad and potent neutralizing antibody. *Cell*, 180(3):471–489.e22, 6 February 2020.

[265] Till Schoofs, Christopher O Barnes, Nina Suh-Toma, Jovana Golijanin, Philipp Schommers, Henning Gruell, Anthony P West, Jr, Franziska Bach, Yu Erica Lee, Lilian Nogueira, Ivelin S Georgiev, Robert T Bailer, Julie Czartoski, John R Mascola, Michael S Seaman, M Juliana McElrath, Nicole A Doria-Rose, Florian Klein, Michel C Nussenzweig, and Pamela J Bjorkman. Broad and potent neutralizing antibodies recognize the silent face of the HIV envelope. *Immunity*, 50(6):1513–1529.e9, 18 June 2019.

[266] Takuya Sekine, André Perez-Potti, Olga Rivera-Ballesteros, Kristoffer Strålin, Jean-Baptiste Gorin, Annika Olsson, Sian Llewellyn-Lacey, Habiba Kamal, Gordana Bogdanovic, Sandra Muschiol, David J Wullimann, Tobias Kammann, Johanna Emgård, Tiphaine Parrot, Elin Folkesson, Karolinska COVID-19 Study Group, Olav Rooyackers, Lars I Eriksson, Jan-Inge Henter, Anders Sönnerborg, Tobias Allander, Jan Albert, Morten Nielsen, Jonas Klingström, Sara Gredmark-Russ, Niklas K Björkström, Johan K Sandberg, David A Price, Hans-Gustaf Ljunggren, Soo Aleman, and Marcus Buggert. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell*, 183(1):158–168.e14, 1 October 2020.

[267] Emilie Seydoux, Leah J Homad, Anna J MacCamy, K Rachael Parks, Nicholas K Hurlburt, Madeleine F Jennewein, Nicholas R Akins, Andrew B Stuart, Yu-Hsin Wan, Junli Feng, Rachael E Whaley, Suruchi Singh, Michael Boeckh, Kristen W Cohen, M Juliana McElrath, Janet A Englund, Helen Y Chu, Marie Pancera, Andrew T McGuire, and Leonidas Stamatatos. Analysis of a SARS-CoV-2-Infected individual reveals development of potent neutralizing antibodies with limited somatic mutation. *Immunity*, 53(1):98–105.e5, 14 July 2020.

[268] Jian Shang, Gang Ye, Ke Shi, Yushun Wan, Chuming Luo, Hideki Aihara, Qibin Geng,

Ashley Auerbach, and Fang Li. Structural basis of receptor recognition by SARS-CoV-2. *Nature*, 581(7807):221–224, May 2020.

[269] Timothy Sheahan, Barry Rockx, Eric Donaldson, Davide Corti, and Ralph Baric. Pathways of cross-species transmission of synthetically reconstructed zoonotic severe acute respiratory syndrome coronavirus. *Journal of virology*, 82(17):8721–8732, September 2008.

[270] Xiaoying Shen, Haili Tang, Charlene McDanal, Kshitij Wagh, William Fischer, James Theiler, Hyejin Yoon, Dapeng Li, Barton F Haynes, Kevin O Sanders, Sandrasegaram Gnanakaran, Nick Hengartner, Rolando Pajon, Gale Smith, Gregory M Glenn, Bette Korber, and David C Montefiori. SARS-CoV-2 variant b.1.1.7 is susceptible to neutralizing antibodies elicited by ancestral spike vaccines. *Cell host & microbe*, 29(4):529–539.e3, 14 April 2021.

[271] Rui Shi, Chao Shan, Xiaomin Duan, Zhihai Chen, Peipei Liu, Jinwen Song, Tao Song, Xiaoshan Bi, Chao Han, Lianao Wu, Ge Gao, Xue Hu, Yanan Zhang, Zhou Tong, Weijin Huang, William Jun Liu, Guizhen Wu, Bo Zhang, Lan Wang, Jianxun Qi, Hui Feng, Fu-Sheng Wang, Qihui Wang, George Fu Gao, Zhiming Yuan, and Jinghua Yan. A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature*, 584(7819):120–124, August 2020.

[272] Ellen Shrock, Eric Fujimura, Tomasz Kula, Richard T Timms, I-Hsiu Lee, Yumei Leng, Matthew L Robinson, Brandon M Sie, Mamie Z Li, Yuezhou Chen, Jennifer Logue, Adam Zuiani, Denise McCulloch, Felipe J N Lelis, Stephanie Henson, Daniel R Monaco, Meghan Travers, Shaghayegh Habibi, William A Clarke, Patrizio Caturegli, Oliver Laeyendecker, Alicja Piechocka-Trocha, Jon Li, Ashok Khatri, Helen Y Chu, MGH COVID-19 Collection & Processing Team, Alexandra-Chloé Villani, Kyle Kays, Marcia B Goldberg, Nir Hacohen, Michael R Filbin, Xu G Yu, Bruce D Walker, Duane R Wesemann, H Benjamin Larman, James A Lederer, and Stephen J Elledge. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science*, 29 September 2020.

[273] Eric A F Simões, Eduardo Forleo-Neto, Gregory P Geba, Mohamed Kamal, Feng Yang, Helen Cicirello, Matthew R Houghton, Ronald Rideman, Qiong Zhao, Sarah L Benvin, Alicia Hawes, Erin D Fuller, Elzbieta Wloga, Jose M Novoa Pizarro, Flor M Munoz, Scott A Rush, Jason S McLellan, Leah Lipsich, Neil Stahl, George D Yancopoulos, David M Weinreich, Christos A Kyratsous, and Sumathi Sivapalasingam. Suptavumab for the prevention of medically attended respiratory syncytial virus infection in preterm

infants. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 8 September 2020.

[274] Derek J Smith, Alan S Lapedes, Jan C de Jong, Theo M Bestebroer, Guus F Rimmelzwaan, Albert D M E Osterhaus, and Ron A M Fouchier. Mapping the antigenic and genetic evolution of influenza virus. *Science*, 305(5682):371–376, 16 July 2004.

[275] G P Smith. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228(4705):1315–1317, 14 June 1985.

[276] J M Smith. Natural selection and the concept of a protein space. *Nature*, 225(5232):563–564, 7 February 1970.

[277] Misha Soskine and Dan S Tawfik. Mutational effects and the evolution of new protein functions. *Nature reviews. Genetics*, 11(8):572–582, August 2010.

[278] Tyler N Starr, Allison J Greaney, Amin Addetia, William W Hannon, Manish C Choudhary, Adam S Dingens, Jonathan Z Li, and Jesse D Bloom. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science*, 371(6531):850–854, 19 February 2021.

[279] Tyler N Starr, Allison J Greaney, Adam S Dingens, and Jesse D Bloom. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell reports. Medicine*, 2(4):100255, 20 April 2021.

[280] Tyler N Starr, Allison J Greaney, Sarah K Hilton, Daniel Ellis, Katharine H D Crawford, Adam S Dingens, Mary Jane Navarro, John E Bowen, M Alejandra Tortorici, Alexandra C Walls, Neil P King, David Veesler, and Jesse D Bloom. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell*, 182(5):1295–1310.e20, 3 September 2020.

[281] Tyler N Starr, Samantha K Zepeda, Alexandra C Walls, Allison J Greaney, David Veesler, and Jesse D Bloom. ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. 19 July 2021.

[282] Tara L Steffen, E Taylor Stone, Mariah Hassert, Elizabeth Geerling, Brian T Grimberg, Ana M Espino, Petraleigh Pantoja, Consuelo Climent, Daniel F Hoft, Sarah L George, Carlos A Sariol, Amelia K Pinto, and James D Brien. The receptor binding domain of SARS-CoV-2 spike is the key target of neutralizing antibody in human polyclonal sera. *bioRxiv*, page 2020.08.21.261727, 22 August 2020.

[283] Naveenchandra Suryadevara, Swathi Shrihari, Pavlo Gilchuk, Laura A VanBlargan, Elad Binshtein, Seth J Zost, Rachel S Nargi, Rachel E Sutton, Emma S Winkler, Elaine C Chen, Mallorie E Fouch, Edgar Davidson, Benjamin J Doranz, Rita E Chen, Pei-Yong Shi, Robert H Carnahan, Larissa B Thackray, Michael S Diamond, and James E Crowe, Jr. Neutralizing and protective human monoclonal antibodies recognizing the n-terminal domain of the SARS-CoV-2 spike protein. *Cell*, 184(9):2316–2331.e15, 29 April 2021.

[284] Anthony T Tan, Martin Linster, Chee Wah Tan, Nina Le Bert, Wan Ni Chia, Kamini Kunasegaran, Yan Zhuang, Christine Y L Tham, Adeline Chia, Gavin J D Smith, Barnaby Young, Shirin Kalimuddin, Jenny G H Low, David Lye, Lin-Fa Wang, and Antonio Bertoletti. Early induction of functional SARS-CoV-2-specific T cells associates with rapid viral clearance and mild disease in COVID-19 patients. *Cell reports*, 34(6):108728, 9 February 2021.

[285] Houriiyah Tegally, Eduan Wilkinson, Marta Giovanetti, Arash Iranzadeh, Vagner Fonseca, Jennifer Giandhari, Deelan Doolabh, Sureshnee Pillay, Emmanuel James San, Nokukhanya Msomi, Koleka Mlisana, Anne von Gottberg, Sibongile Walaza, Mushal Allam, Arshad Ismail, Thabo Mohale, Allison J Glass, Susan Engelbrecht, Gert Van Zyl, Wolfgang Preiser, Francesco Petruccione, Alex Sigal, Diana Hardie, Gert Marais, Nei-Yuan Hsiao, Stephen Korsman, Mary-Ann Davies, Lynn Tyers, Innocent Mudau, Denis York, Caroline Maslo, Dominique Goedhals, Shareef Abrahams, Oluwakemi Laguda-Akingba, Arghavan Alisoltani-Dehkordi, Adam Godzik, Constantinos Kurt Wibmer, Bryan Trevor Sewell, José Lourenço, Luiz Carlos Junior Alcantara, Sergei L Kosakovsky Pond, Steven Weaver, Darren Martin, Richard J Lessells, Jinal N Bhiman, Carolyn Williamson, and Tulio de Oliveira. Detection of a SARS-CoV-2 variant of concern in south africa. *Nature*, 592(7854):438–443, April 2021.

[286] Jan ter Meulen, Edward N van den Brink, Leo L M Poon, Wilfred E Marissen, Cynthia S W Leung, Freek Cox, Chung Y Cheung, Arjen Q Bakker, Johannes A Bogaards, Els van Deventer, Wolfgang Preiser, Hans Wilhelm Doerr, Vincent T Chow, John de Kruif, Joseph S M Peiris, and Jaap Goudsmit. Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. *PLoS medicine*, 3(7):e237, July 2006.

[287] Emma C Thomson, Laura E Rosen, James G Shepherd, Roberto Spreafico, Ana da Silva Filipe, Jason A Wojcechowskyj, Chris Davis, Luca Piccoli, David J Pascall, Josh Dillen, Spyros Lytras, Nadine Czudnochowski, Rajiv Shah, Marcel Meury, Natasha Jesudason, Anna De Marco, Kathy Li, Jessica Bassi, Aine O'Toole, Dora

Pinto, Rachel M Colquhoun, Katja Culap, Ben Jackson, Fabrizia Zatta, Andrew Rambaut, Stefano Jaconi, Vattipally B Sreenu, Jay Nix, Ruth F Jarrett, Martina Beltramello, Kyriaki Nomikou, Matteo Pizzuto, Lily Tong, Elisabetta Cameroni, Natasha Johnson, Arthur Wickenhagen, Alessandro Ceschi, Daniel Mair, Paolo Ferrari, Katherine Smollett, Federica Sallusto, Stephen Carmichael, Christian Garzoni, Jenna Nichols, Massimo Galli, Joseph Hughes, Agostino Riva, Antonia Ho, Malcolm G Semple, Peter J M Openshaw, J Kenneth Baillie, The ISARIC4C Investigators, the COVID-19 Genomics UK (COG-UK) consortium, Suzannah J Rihn, Samantha J Lycett, Herbert W Virgin, Amalio Telenti, Davide Corti, David L Robertson, and Gyorgy Snell. The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity. *bioRxiv*, page 2020.11.04.355842, 5 November 2020.

[288] Bargavi Thyagarajan and Jesse D Bloom. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife*, 3, 8 July 2014.

[289] Fang Tian, Bei Tong, Liang Sun, Shengchao Shi, Bin Zheng, Zibin Wang, Xianchi Dong, and Peng Zheng. N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. *eLife*, 10, 20 August 2021.

[290] Xiaolong Tian, Cheng Li, Ailing Huang, Shuai Xia, Sicong Lu, Zhengli Shi, Lu Lu, Shibo Jiang, Zhenlin Yang, Yanling Wu, and Tianlei Ying. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerging microbes & infections*, 9(1):382–385, 17 February 2020.

[291] Christine Toelzer, Kapil Gupta, Sathish K N Yadav, Ufuk Borucu, Andrew D Davidson, Maia Kavanagh Williamson, Deborah K Shoemark, Frederic Garzoni, Oskar Staufer, Rachel Milligan, Julien Capin, Adrian J Mulholland, Joachim Spatz, Daniel Fitzgerald, Imre Berger, and Christiane Schaffitzel. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science*, 370(6517):725–730, 6 November 2020.

[292] Suxiang Tong, Christina Conrardy, Susan Ruone, Ivan V Kuzmin, Xiling Guo, Ying Tao, Michael Niezgoda, Lia Haynes, Bernard Agwanda, Robert F Breiman, Larry J Anderson, and Charles E Rupprecht. Detection of novel SARS-like and other coronaviruses in bats from kenya. *Emerging infectious diseases*, 15(3):482–485, March 2009.

[293] M Alejandra Tortorici, Martina Beltramello, Florian A Lempp, Dora Pinto, Ha V Dang, Laura E Rosen, Matthew McCallum, John Bowen, Andrea Minola, Stefano Jaconi, Fabrizia Zatta, Anna De Marco, Barbara Guarino, Siro Bianchi, Elvin J Lauron, Heather Tucker, Jiayi Zhou, Alessia Peter, Colin Havenar-Daughton, Jason A

Wojcechowskyj, James Brett Case, Rita E Chen, Hannah Kaiser, Martin Montiel-Ruiz, Marcel Meury, Nadine Czudnochowski, Roberto Spreafico, Josh Dillen, Cindy Ng, Nicole Sprugasci, Katja Culap, Fabio Benigni, Rana Abdelnabi, Shi-Yan Caroline Foo, Michael A Schmid, Elisabetta Cameroni, Agostino Riva, Arianna Gabrieli, Massimo Galli, Matteo S Pizzuto, Johan Neyts, Michael S Diamond, Herbert W Virgin, Gyorgy Snell, Davide Corti, Katja Fink, and David Veesler. Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science*, 370(6519):950–957, 20 November 2020.

[294] M Alejandra Tortorici and David Veesler. Chapter four - structural insights into coronavirus entry. In Félix A Rey, editor, *Advances in Virus Research*, volume 105, pages 93–116. Academic Press, 1 January 2019.

[295] M Alejandra Tortorici, Alexandra C Walls, Yifei Lang, Chunyan Wang, Zeshi Li, Danielle Koerhuis, Geert-Jan Boons, Berend-Jan Bosch, Félix A Rey, Raoul J de Groot, and David Veesler. Structural basis for human coronavirus attachment to sialic acid receptors. *Nature structural & molecular biology*, 26(6):481–489, June 2019.

[296] Alexandra Trkola, Herbert Kuster, Peter Rusert, Beda Joos, Marek Fischer, Christine Leemann, Amapola Manrique, Michael Huber, Manuela Rehr, Annette Oxenius, Rainer Weber, Gabriela Stiegler, Brigitta Vcelar, Hermann Katinger, Leonardo Aceto, and Huldrych F Günthard. Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies. *Nature medicine*, 11(6):615–622, June 2005.

[297] For Yue Tso, Salum J Lidenge, Lisa K Poppe, Phoebe B Peña, Sara R Privatt, Sydney J Bennett, John R Ngowi, Julius Mwaiselage, Michael Belshan, Jacob A Siedlik, Morgan A Raine, Juan B Ochoa, Julia Garcia-Diaz, Bobby Nossaman, Lyndsey Buckner, W Mark Roberts, Matthew J Dean, Augusto C Ochoa, John T West, and Charles Wood. Presence of antibody-dependent cellular cytotoxicity (ADCC) against SARS-CoV-2 in COVID-19 plasma. *PloS one*, 16(3):e0247640, 4 March 2021.

[298] Jackson S Turner, Wooseob Kim, Elizaveta Kalaidina, Charles W Goss, Adriana M Rauseo, Aaron J Schmitz, Lena Hansen, Alem Haile, Michael K Klebert, Iskra Pusic, Jane A O'Halloran, Rachel M Presti, and Ali H Ellebedy. SARS-CoV-2 infection induces long-lived bone marrow plasma cells in humans. *Nature*, 595(7867):421–425, July 2021.

[299] Keiya Uriu, Izumi Kimura, Kotaro Shirakawa, Akifumi Takaori-Kondo, Taka-Aki Nakada, Atsushi Kaneda, So Nakagawa, Kei Sato, and Genotype to Phenotype Japan

(G2P-Japan) Consortium. Neutralization of the SARS-CoV-2 mu variant by convalescent and vaccine serum. *The New England journal of medicine*, 3 November 2021.

[300] Lucy van Dorp, Mislav Acman, Damien Richard, Liam P Shaw, Charlotte E Ford, Louise Ormond, Christopher J Owen, Juanita Pang, Cedric C S Tan, Florencia A T Boshier, Arturo Torres Ortiz, and François Balloux. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 83:104351, September 2020.

[301] R Vlasak, W Luytjes, W Spaan, and P Palese. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 85(12):4526–4529, June 1988.

[302] Carolina M Voloch, Ronaldo da Silva Francisco, Jr, Luiz G P de Almeida, Cynthia C Cardoso, Otavio J Brustolini, Alexandra L Gerber, Ana Paula de C Guimarães, Diana Mariani, Raissa Mirella da Costa, Orlando C Ferreira, Jr, Covid19-UFRJ Workgroup, LNCC Workgroup, Adriana Cony Cavalcanti, Thiago Silva Frauches, Claudia Maria Braga de Mello, Isabela de Carvalho Leitão, Rafael Mello Galliez, Débora Souza Faffe, Terezinha M P P Castiñeiras, Amilcar Tanuri, and Ana Tereza R de Vasconcelos. Genomic characterization of a novel SARS-CoV-2 lineage from rio de janeiro, brazil. *Journal of virology*, 1 March 2021.

[303] William N Voss, Yixuan J Hou, Nicole V Johnson, George Delidakis, Jin Eyun Kim, Kamyab Javanmardi, Andrew P Horton, Foteini Bartzoka, Chelsea J Paresi, Yuri Tanno, Chia-Wei Chou, Shawn A Abbasi, Whitney Pickens, Katia George, Daniel R Boutz, Dalton M Towers, Jonathan R McDaniel, Daniel Billick, Jule Goike, Lori Rowe, Dhwani Batra, Jan Pohl, Justin Lee, Shivaprakash Gangappa, Suryaprakash Sambhara, Michelle Gadush, Nianshuang Wang, Maria D Person, Brent L Iverson, Jimmy D Gollihar, John Dye, Andrew Herbert, Ilya J Finkelstein, Ralph S Baric, Jason S McLellan, George Georgiou, Jason J Lavinder, and Gregory C Ippolito. Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes. *Science*, 372(6546):1108–1112, 4 May 2021.

[304] Alexandra C Walls, Brooke Fiala, Alexandra Schäfer, Samuel Wrenn, Minh N Pham, Michael Murphy, Longping V Tse, Laila Shehata, Megan A O'Connor, Chengbo Chen, Mary Jane Navarro, Marcos C Miranda, Deleah Pettie, Rashmi Ravichandran, John C Kraft, Cassandra Ogohara, Anne Palser, Sara Chalk, E-Chiang Lee, Kathryn Guerriero, Elizabeth Kepl, Cameron M Chow, Claire Sydeman, Edgar A Hodge, Brieann

Brown, Jim T Fuller, Kenneth H Dinnon, 3rd, Lisa E Gralinski, Sarah R Leist, Kendra L Gully, Thomas B Lewis, Miklos Guttman, Helen Y Chu, Kelly K Lee, Deborah H Fuller, Ralph S Baric, Paul Kellam, Lauren Carter, Marion Pepper, Timothy P Sheahan, David Veesler, and Neil P King. Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. *Cell*, 183(5):1367–1382.e17, 25 November 2020.

[305] Alexandra C Walls, Marcos C Miranda, Alexandra Schäfer, Minh N Pham, Allison Greaney, Prabhu S Arunachalam, Mary-Jane Navarro, M Alejandra Tortorici, Kenneth Rogers, Megan A O'Connor, Lisa Shirreff, Douglas E Ferrell, John Bowen, Natalie Brunette, Elizabeth Kepl, Samantha K Zepeda, Tyler Starr, Ching-Lin Hsieh, Brooke Fiala, Samuel Wrenn, Deleah Pettie, Claire Sydeman, Kaitlin R Sprouse, Max Johnson, Alyssa Blackstone, Rashmi Ravichandran, Cassandra Ogohara, Lauren Carter, Sasha W Tilles, Rino Rappuoli, Sarah R Leist, David R Martinez, Matthew Clark, Roland Tisch, Derek T O'Hagan, Robbert Van Der Most, Wesley C Van Voorhis, Davide Corti, Jason S McLellan, Harry Kleanthous, Timothy P Sheahan, Kelly D Smith, Deborah H Fuller, Francois Villinger, Jesse Bloom, Bali Pulendran, Ralph S Baric, Neil P King, and David Veesler. Elicitation of broadly protective sarbecovirus immunity by receptor-binding domain nanoparticle vaccines. *Cell*, 184(21):5432–5447.e16, 14 October 2021.

[306] Alexandra C Walls, Young-Jun Park, M Alejandra Tortorici, Abigail Wall, Andrew T McGuire, and David Veesler. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, 181(2):281–292.e6, 16 April 2020.

[307] Alexandra C Walls, M Alejandra Tortorici, Berend-Jan Bosch, Brandon Frenz, Peter J M Rottier, Frank DiMaio, Félix A Rey, and David Veesler. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature*, 531(7592):114–117, 3 March 2016.

[308] Alexandra C Walls, M Alejandra Tortorici, Joost Snijder, Xiaoli Xiong, Berend-Jan Bosch, Felix A Rey, and David Veesler. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proceedings of the National Academy of Sciences of the United States of America*, 114(42):11157–11162, 17 October 2017.

[309] Alexandra C Walls, Xiaoli Xiong, Young-Jun Park, M Alejandra Tortorici, Joost Snijder, Joel Quispe, Elisabetta Cameroni, Robin Gopal, Mian Dai, Antonio Lanzavecchia,

Maria Zambon, Félix A Rey, Davide Corti, and David Veesler. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell*, 183(6):1732, 10 December 2020.

[310] Edward E Walsh, Robert W Frenck, Jr, Ann R Falsey, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, Kathleen Neuzil, Mark J Mulligan, Ruth Bailey, Kena A Swanson, Ping Li, Kenneth Koury, Warren Kalina, David Cooper, Camila Fontes-Garfias, Pei-Yong Shi, Özlem Türeci, Kristin R Tompkins, Kirsten E Lyke, Vanessa Raabe, Philip R Dormitzer, Kathrin U Jansen, Uğur Şahin, and William C Gruber. Safety and immunogenicity of two RNA-Based covid-19 vaccine candidates. *The New England journal of medicine*, 383(25):2439–2450, 17 December 2020.

[311] Yushun Wan, Jian Shang, Rachel Graham, Ralph S Baric, and Fang Li. Receptor recognition by the novel coronavirus from wuhan: an analysis based on Decade-Long structural studies of SARS coronavirus. *Journal of virology*, 94(7), 17 March 2020.

[312] Lingshu Wang, Tongqing Zhou, Yi Zhang, Eun Sung Yang, Chaim A Schramm, Wei Shi, Amarendra Pegu, Olamide K Oloniniyi, Amy R Henry, Samuel Darko, Sandeep R Narpala, Christian Hatcher, David R Martinez, Yaroslav Tsybovsky, Emily Phung, Olubukola M Abiona, Avan Antia, Evan M Cale, Lauren A Chang, Misook Choe, Kizzmekia S Corbett, Rachel L Davis, Anthony T DiPiazza, Ingelise J Gordon, Sabrina Helmold Hait, Tandile Hermanus, Prudence Kgagudi, Farida Laboune, Kwanyee Leung, Tracy Liu, Rosemarie D Mason, Alexandra F Nazzari, Laura Novik, Sarah O'Connell, Sijy O'Dell, Adam S Olia, Stephen D Schmidt, Tyler Stephens, Christopher D Stringham, Chloe Adrienna Talana, I-Ting Teng, Danielle A Wagner, Alicia T Widge, Baoshan Zhang, Mario Roederer, Julie E Ledgerwood, Tracy J Ruckwardt, Martin R Gaudinski, Penny L Moore, Nicole A Doria-Rose, Ralph S Baric, Barney S Graham, Adrian B McDermott, Daniel C Douek, Peter D Kwong, John R Mascola, Nancy J Sullivan, and John Misasi. Ultrapotent antibodies against diverse and highly transmissible SARS-CoV-2 variants. *Science*, 373(6556):eabh1766, 13 August 2021.

[313] Pengfei Wang, Manoj S Nair, Lihong Liu, Sho Iketani, Yang Luo, Yicheng Guo, Maple Wang, Jian Yu, Baoshan Zhang, Peter D Kwong, Barney S Graham, John R Mascola, Jennifer Y Chang, Michael T Yin, Magdalena Sobieszczyk, Christos A Kyratsous, Lawrence Shapiro, Zizhang Sheng, Yaoxing Huang, and David D Ho. Antibody resistance of SARS-CoV-2 variants b.1.351 and b.1.1.7. *Nature*, 593(7857):130–135, May 2021.

[314] Zijun Wang, Frauke Muecksch, Dennis Schaefer-Babajew, Shlomo Finkin, Charlotte Viant, Christian Gaebler, Hans-Heinrich Hoffmann, Christopher O Barnes, Melissa Cipolla, Victor Ramos, Thiago Y Oliveira, Alice Cho, Fabian Schmidt, Justin Da Silva, Eva Bednarski, Lauren Aguado, Jim Yee, Mridushi Daga, Martina Turroja, Katrina G Millard, Mila Jankovic, Anna Gazumyan, Zhen Zhao, Charles M Rice, Paul D Bieniasz, Marina Caskey, Theodora Hatziioannou, and Michel C Nussenzweig. Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature*, 595(7867):426–431, July 2021.

[315] Zijun Wang, Fabian Schmidt, Yiska Weisblum, Frauke Muecksch, Christopher O Barnes, Shlomo Finkin, Dennis Schaefer-Babajew, Melissa Cipolla, Christian Gaebler, Jenna A Lieberman, Thiago Y Oliveira, Zhi Yang, Morgan E Abernathy, Kathryn E Huey-Tubman, Arlene Hurley, Martina Turroja, Kamille A West, Kristie Gordon, Katrina G Millard, Victor Ramos, Justin Da Silva, Jianliang Xu, Robert A Colbert, Roshni Patel, Juan Dizon, Cecille Unson-O'Brien, Irina Shimeliovich, Anna Gazumyan, Marina Caskey, Pamela J Bjorkman, Rafael Casellas, Theodora Hatziioannou, Paul D Bieniasz, and Michel C Nussenzweig. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature*, 592(7855):616–622, 10 February 2021.

[316] R G Webster and Ann Arbor. Disquisitions on original antigenic sin. i. evidence in man the. In *Journal of Experimental Medicine*. Citeseer, 1966.

[317] Anna Z Wec, Zachary A Bornholdt, Shihua He, Andrew S Herbert, Eileen Goodwin, Ariel S Wirchnianski, Bronwyn M Gunn, Zirui Zhang, Wenjun Zhu, Guodong Liu, Dafna M Abelson, Crystal L Moyer, Rohit K Jangra, Rebekah M James, Russell R Bakken, Natasha Bohorova, Ognian Bohorov, Do H Kim, Michael H Pauly, Jesus Velasco, Robert H Bortz, 3rd, Kevin J Whaley, Tracey Goldstein, Simon J Anthony, Galit Alter, Laura M Walker, John M Dye, Larry Zeitlin, Xiangguo Qiu, and Kartik Chandran. Development of a human antibody cocktail that deploys multiple functions to confer Pan-Ebolavirus protection. *Cell host & microbe*, 25(1):39–48.e5, 9 January 2019.

[318] Anna Z Wec, Daniel Wrapp, Andrew S Herbert, Daniel P Maurer, Denise Haslwanter, Mrunal Sakharkar, Rohit K Jangra, M Eugenia Dieterle, Asparouh Lilov, Deli Huang, Longping V Tse, Nicole V Johnson, Ching-Lin Hsieh, Nianshuang Wang, Juergen H Nett, Elizabeth Champney, Irina Burnina, Michael Brown, Shu Lin, Melanie Sinclair, Carl Johnson, Sarat Pudi, Robert Bortz, 3rd, Ariel S Wirchnianski, Ethan Laudermilch, Catalina Florez, J Maximilian Fels, Cecilia M O'Brien, Barney S Graham, David Nemazee, Dennis R Burton, Ralph S Baric, James E Voss, Kartik Chandran, John M

Dye, Jason S McLellan, and Laura M Walker. Broad neutralization of SARS-related viruses by human monoclonal antibodies. *Science*, 369(6504):731–736, 7 August 2020.

[319] Payton A Weidenbacher and Peter S Kim. Protect, modify, deprotect (PMD): A strategy for creating vaccines to elicit antibodies targeting a specific epitope. *Proceedings of the National Academy of Sciences of the United States of America*, 116(20):9947–9952, 14 May 2019.

[320] Yiska Weisblum, Fabian Schmidt, Fengwen Zhang, Justin DaSilva, Daniel Poston, Julio Cc Lorenzi, Frauke Muecksch, Magdalena Rutkowska, Hans-Heinrich Hoffmann, Eleftherios Michailidis, Christian Gaebler, Marianna Agudelo, Alice Cho, Zijun Wang, Anna Gazumyan, Melissa Cipolla, Larry Luchsinger, Christopher D Hillyer, Marina Caskey, Davide F Robbiani, Charles M Rice, Michel C Nussenzweig, Theodora Hatziioannou, and Paul D Bieniasz. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife*, 9, 28 October 2020.

[321] Drew Weissman, Mohamad-Gabriel Alameh, Thushan de Silva, Paul Collini, Hailey Hornsby, Rebecca Brown, Celia C LaBranche, Robert J Edwards, Laura Sutherland, Sampa Santra, Katayoun Mansouri, Sophie Gobeil, Charlene McDanal, Norbert Pardi, Nick Hengartner, Paulo J C Lin, Ying Tam, Pamela A Shaw, Mark G Lewis, Carsten Boesler, Uğur Şahin, Priyamvada Acharya, Barton F Haynes, Bette Korber, and David C Montefiori. D614G spike mutation increases SARS CoV-2 susceptibility to neutralization. 24 July 2020.

[322] Alane E Wentz and Eric V Shusta. A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Applied and environmental microbiology*, 73(4):1189–1198, February 2007.

[323] Anthony P West, Joel O Wertheim, Jade C Wang, Tetyana I Vasylyeva, Jennifer L Havens, Moinuddin A Chowdhury, Edimarlyn Gonzalez, Courtney E Fang, Steve S Di Lonardo, Scott Hughes, Jennifer L Rakeman, Henry H Lee, Christopher O Barnes, Priyanthi N P Gnanapragasam, Zhi Yang, Christian Gaebler, Marina Caskey, Michel C Nussenzweig, Jennifer R Keeffe, and Pamela J Bjorkman. Detection and characterization of the SARS-CoV-2 lineage b.1.526 in new york. *bioRxiv : the preprint server for biology*, page 2021.02.14.431043, 22 April 2021.

[324] Constantinos Kurt Wibmer, Frances Ayres, Tandile Hermanus, Mashudu Madzivhandila, Prudence Kgagudi, Brent Oosthuysen, Bronwen E Lambson, Tulio de Oliveira, Marion Vermeulen, Karin van der Berg, Theresa Rossouw, Michael Boswell, Veronica Ueckermann, Susan Meiring, Anne von Gottberg, Cheryl Cohen,

Lynn Morris, Jinal N Bhiman, and Penny L Moore. SARS-CoV-2 501Y.V2 escapes neutralization by south african COVID-19 donor plasma. *Nature medicine*, 27(4):622–625, April 2021.

[325] Eduan Wilkinson, Marta Giovanetti, Houriiyah Tegally, James E San, Richard Lessells, Diego Cuadros, Darren P Martin, David A Rasmussen, Abdel-Rahman N Zekri, Abdoul K Sangare, Abdoul-Salam Ouedraogo, Abdul K Sesay, Abechi Priscilla, Adedotun-Sulaiman Kemi, Adewunmi M Olubusuyi, Adeyemi O O Oluwapelumi, Adnène Hammami, Adrienne A Amuri, Ahmad Sayed, Ahmed E O Ouma, Aida Elargoubi, Nnennaya A Ajayi, Ajogbasile F Victoria, Akano Kazeem, Akpede George, Alexander J Trotter, Ali A Yahaya, Alpha K Keita, Amadou Diallo, Amadou Kone, Amal Souissi, Amel Chtourou, Ana V Gutierrez, Andrew J Page, Anika Vinze, Arash Iranzadeh, Arnold Lambisia, Arshad Ismail, Audu Rosemary, Augustina Sylverken, Ayoade Femi, Azeddine Ibrahimi, Baba Marycelin, Bamidele S Oderinde, Bankole Bolajoko, Beatrice Dhaala, Belinda L Herring, Berthe-Marie Njanpop-Lafourcade, Bronwyn Kleinhans, Bronwyn McInnis, Bryan Tegomoh, Cara Brook, Catherine B Pratt, Cathrine Scheepers, Chantal G Akoua-Koffi, Charles N Agoti, Christophe Peyrefitte, Claudia Daubenberger, Collins M Morang'a, D James Nokes, Daniel G Amoako, Daniel L Bugembe, Danny Park, David Baker, Deelan Doolabh, Deogratius Ssemwanga, Derek Tshiabuila, Diarra Bassirou, Dominic S Y Amuzu, Dominique Goedhals, Donwilliams O Omuoyo, Dorcas Maruapula, Ebenezer Foster-Nyarko, Eddy K Lusamaki, Edgar Simulundu, Edidah M Ong'era, Edith N Ngabana, Edwin Shumba, Elmostafa El Fahime, Emmanuel Lokilo, Enatha Mukantwari, Eromon Philomena, Essia Belarbi, Etienne Simon-Loriere, Etilé A Anoh, Fabian Leendertz, Faida Ajili, Fakayode O Enoch, Fares Wasfi, Fatma Abdelmoula, Fausta S Mosha, Faustinos T Takawira, Fawzi Derrar, Feriel Bouzid, Folarin Onikepe, Fowotade Adeola, Francisca M Muyembe, Frank Tanser, Fred A Dratibi, Gabriel K Mbunsu, Gaetan Thilliez, Gemma L Kay, George Githinji, Gert van Zyl, Gordon A Awandare, Grit Schubert, Gugu P Maphalala, Hafaliana C Ranaivoson, Hajar Lemriss, Happi Anise, Haruka Abe, Hela H Karray, Hellen Nansumba, Hesham A Elgahzaly, Hlanai Gumbo, Ibtihel Smeti, Ikhlas B Ayed, Ikponmwosa Odia, Ilhem Boutiba Ben Boubaker, Imed Gaaloul, Inbal Gazy, Innocent Mudau, Isaac Ssewanyana, Iyaloo Konstantinus, Jean B Lekana-Douk, Jean-Claude C Makangara, Jean-Jacques M Tamfum, Jean-Michel Heraud, Jeffrey G Shaffer, Jennifer Giandhari, Jingjing Li, Jiro Yasuda, Joana Q Mends, Jocelyn Kiconco, John M Morobe, John O Gyapong, Johnson C Okolie, John T Kayiwa, Johnathan A Edwards, Jones Gyamfi, Jouali Farah, Joweria Nakaseegu, Joyce M Ngoi, Joyce Namulondo, Julia C Andeko, Julius J Lutwama, Justin O'Grady, Katherine Siddle, Kayode T Adeyemi, Kefentse A Tumedi, Khadija M Said, Kim Hae-Young, Kwabena O Duedu, Lahcen Belyamani, Lamia Fki-Berrajah, Lavanya

Singh, Leonardo de O. Martins, Lynn Tyers, Magalutcheemee Ramuth, Maha Mastouri, Mahjoub Aouni, Mahmoud el Hefnawi, Maitshwarelo I Matsheka, Malebogo Kebabonye, Mamadou Diop, Manel Turki, Marietou Paye, Martin M Nyaga, Mathabo Mareka, Matoke-Muhia Damaris, Maureen W Mburu, Maximillian Mpina, Mba Nwando, Michael Owusu, Michael R Wiley, Mirabeau T Youtchou, Mitoha O Ayekaba, Mohamed Abouelhoda, Mohamed G Seadawy, Mohamed K Khalifa, Mooko Sekhele, Mouna Ouadghiri, Moussa M Diagne, Mulenga Mwenda, Mushal Allam, My V T Phan, Nabil Abid, Nadia Touil, Nadine Rujeni, Najla Kharrat, Nalia Ismael, Ndongo Dia, Nedio Mabunda, Nei-Yuan Hsiao, Nelson B Silochi, Ngoy Nsenga, Nicksy Gumede, Nicola Mulder, Nnaemeka Ndodo, Norosoa H Razanajatovo, Nosamiefan Iguosadolo, Oguzie Judith, Ojide C Kingsley, Okogbenin Sylvanus, Okokhere Peter, Oladiji Femi, Olawoye Idowu, Olumade Testimony, Omoruyi E Chukwuma, Onwe E Ogah, Chika K Onwuamah, Oshomah Cyril, Ousmane Faye, Oyewale Tomori, Pascale Ondoa, Patrice Combe, Patrick Semanda, Paul E Oluniyi, Paulo Arnaldo, Peter K Quashie, Philippe Dussart, Phillip A Bester, Placide K Mbala, Reuben Ayivor-Djanie, Richard Njouom, Richard O Phillips, Richmond Gorman, Robert A Kingsley, Rosina A A Carr, Saâd El Kabbaj, Saba Gargouri, Saber Masmoudi, Safietou Sankhe, Salako B Lawal, Samar Kassim, Sameh Trabelsi, Samar Metha, Sami Kammoun, Sanaâ Lemriss, Sara H A Agwa, Sébastien Calvignac-Spencer, Stephen F Schaffner, Seydou Doumbia, Sheila M Mandanda, Sherihane Aryeetey, Shymaa S Ahmed, Siham Elhamoumi, Soafy Andriamandimby, Sobajo Tope, Sonia Lekana-Douki, Sophie Prosolek, Soumeya Ouangraoua, Steve A Mundeke, Steven Rudder, Sumir Panji, Sureshnee Pillay, Susan Engelbrecht, Susan Nabadda, Sylvie Behillil, Sylvie L Budiaki, Sylvie van der Werf, Tapfumanei Mashe, Tarik Aanniz, Thabo Mohale, Thanh Le-Viet, Tobias Schindler, Ugochukwu J Anyaneji, Ugwu Chinedu, Upasana Ramphal, Uwanibe Jessica, Uwem George, Vagner Fonseca, Vincent Enouf, Vivianne Gorova, Wael H Roshdy, William K Ampofo, Wolfgang Preiser, Wonderful T Choga, Yaw Bediako, Yeshnee Naidoo, Yvan Butera, Zaydah R de Laurent, Amadou A Sall, Ahmed Rebai, Anne von Gottberg, Bourema Kouriba, Carolyn Williamson, Daniel J Bridges, Ihekweazu Chikwe, Jinal N Bhiman, Madisa Mine, Matthew Cotten, Sikhulile Moyo, Simani Gaseitsiwe, Ngonda Saasa, Pardis C Sabeti, Pontiano Kaleebu, Yenew K Tebeje, Sofonias K Tessema, Christian Happi, John Nkengasong, and Tulio de Oliveira. A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in africa. *Science*, 374(6566):423–431, 2021.

[326] Emma S Winkler, Pavlo Gilchuk, Jinsheng Yu, Adam L Bailey, Rita E Chen, Zhenlu Chong, Seth J Zost, Hyesun Jang, Ying Huang, James D Allen, James Brett Case, Rachel E Sutton, Robert H Carnahan, Tamarand L Darling, Adrianus C M Boon, Matthias Mack, Richard D Head, Ted M Ross, James E Crowe, Jr, and Michael S Diamond. Human neutralizing antibodies against SARS-CoV-2 require intact fc effector

functions for optimal therapeutic protection. *Cell*, 184(7):1804–1820.e16, 1 April 2021.

[327] Alan H M Wong, Aidan C A Tomlinson, Dongxia Zhou, Malathy Satkunarajah, Kevin Chen, Chetna Sharon, Marc Desforges, Pierre J Talbot, and James M Rini. Receptor-binding loops in alphacoronavirus adaptation and evolution. *Nature communications*, 8(1):1735, 23 November 2017.

[328] World Health Organization. COVID-19 vaccine tracker and landscape. `https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines`. Accessed: 2021-10-22.

[329] Michael Worobey, Guan-Zhu Han, and Andrew Rambaut. Genesis and pathogenesis of the 1918 pandemic H1N1 influenza a virus. *Proceedings of the National Academy of Sciences of the United States of America*, 111(22):8107–8112, 3 June 2014.

[330] Michael Worobey, Jonathan Pekar, Brendan B Larsen, Martha I Nelson, Verity Hill, Jeffrey B Joy, Andrew Rambaut, Marc A Suchard, Joel O Wertheim, and Philippe Lemey. The emergence of SARS-CoV-2 in europe and north america. *Science*, 370(6516):564–570, 30 October 2020.

[331] Daniel Wrapp, Dorien De Vlieger, Kizzmekia S Corbett, Gretel M Torres, Nianshuang Wang, Wander Van Breedam, Kenny Roose, Loes van Schie, VIB-CMB COVID-19 Response Team, Markus Hoffmann, Stefan Pöhlmann, Barney S Graham, Nico Callewaert, Bert Schepens, Xavier Saelens, and Jason S McLellan. Structural basis for potent neutralization of betacoronaviruses by Single-Domain camelid antibodies. *Cell*, 181(5):1004–1015.e15, 28 May 2020.

[332] Daniel Wrapp, Nianshuang Wang, Kizzmekia S Corbett, Jory A Goldsmith, Ching-Lin Hsieh, Olubukola Abiona, Barney S Graham, and Jason S McLellan. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483):1260–1263, 13 March 2020.

[333] Antoni G Wrobel, Donald J Benton, Pengqi Xu, Chloë Roustan, Stephen R Martin, Peter B Rosenthal, John J Skehel, and Steven J Gamblin. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nature structural & molecular biology*, 27(8):763–767, August 2020.

[334] Fan Wu, Su Zhao, Bin Yu, Yan-Mei Chen, Wen Wang, Zhi-Gang Song, Yi Hu, Zhao-Wu Tao, Jun-Hua Tian, Yuan-Yuan Pei, Ming-Li Yuan, Yu-Ling Zhang, Fa-Hui Dai, Yi Liu, Qi-Min Wang, Jiao-Jiao Zheng, Lin Xu, Edward C Holmes, and Yong-Zhen

Zhang. A new coronavirus associated with human respiratory disease in china. *Nature*, 579(7798):265–269, March 2020.

[335] Nicholas C Wu, Meng Yuan, Hejun Liu, Chang-Chun D Lee, Xueyong Zhu, Sandhya Bangaru, Jonathan L Torres, Tom G Caniels, Philip J M Brouwer, Marit J van Gils, Rogier W Sanders, Andrew B Ward, and Ian A Wilson. An alternative binding mode of IGHV3-53 antibodies to the SARS-CoV-2 receptor binding domain. *Cell reports*, 33(3):108274, 20 October 2020.

[336] Yan Wu, Feiran Wang, Chenguang Shen, Weiyu Peng, Delin Li, Cheng Zhao, Zhaohui Li, Shihua Li, Yuhai Bi, Yang Yang, Yuhuan Gong, Haixia Xiao, Zheng Fan, Shuguang Tan, Guizhen Wu, Wenjie Tan, Xuancheng Lu, Changfa Fan, Qihui Wang, Yingxia Liu, Chen Zhang, Jianxun Qi, George Fu Gao, Feng Gao, and Lei Liu. A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science*, 368(6496):1274–1278, 2020.

[337] Xuping Xie, Yang Liu, Jianying Liu, Xianwen Zhang, Jing Zou, Camila R Fontes-Garfias, Hongjie Xia, Kena A Swanson, Mark Cutler, David Cooper, Vineet D Menachery, Scott C Weaver, Philip R Dormitzer, and Pei-Yong Shi. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nature medicine*, 27(4):620–621, April 2021.

[338] Xuping Xie, Kumari G Lokugamage, Xianwen Zhang, Michelle N Vu, Antonio E Muruato, Vineet D Menachery, and Pei-Yong Shi. Engineering SARS-CoV-2 using a reverse genetic system. *Nature protocols*, 16(3):1761–1784, March 2021.

[339] Katherine S Xue and Jesse D Bloom. Linking influenza virus evolution within and between human hosts. *Virus evolution*, 6(1):veaa010, January 2020.

[340] Katherine S Xue, Terry Stevens-Ayers, Angela P Campbell, Janet A Englund, Steven A Pergam, Michael Boeckh, and Jesse D Bloom. Parallel evolution of influenza across multiple spatiotemporal scales. *eLife*, 6, 27 June 2017.

[341] Renhong Yan, Yuanyuan Zhang, Yaning Li, Lu Xia, Yingying Guo, and Qiang Zhou. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, 367(6485):1444–1448, 27 March 2020.

[342] Jingyun Yang, Wei Wang, Zimin Chen, Shuaiyao Lu, Fanli Yang, Zhenfei Bi, Linlin Bao, Fei Mo, Xue Li, Yong Huang, Weiqi Hong, Yun Yang, Yuan Zhao, Fei Ye, Sheng Lin, Wei Deng, Hua Chen, Hong Lei, Ziqi Zhang, Min Luo, Hong Gao, Yue Zheng,

Yanqiu Gong, Xiaohua Jiang, Yanfeng Xu, Qi Lv, Dan Li, Manni Wang, Fengdi Li, Shunyi Wang, Guanpeng Wang, Pin Yu, Yajin Qu, Li Yang, Hongxin Deng, Aiping Tong, Jiong Li, Zhenling Wang, Jinliang Yang, Guobo Shen, Zhiwei Zhao, Yuhua Li, Jingwen Luo, Hongqi Liu, Wenhai Yu, Mengli Yang, Jingwen Xu, Junbin Wang, Haiyan Li, Haixuan Wang, Dexuan Kuang, Panpan Lin, Zhengtao Hu, Wei Guo, Wei Cheng, Yanlin He, Xiangrong Song, Chong Chen, Zhihong Xue, Shaohua Yao, Lu Chen, Xuelei Ma, Siyuan Chen, Maling Gou, Weijin Huang, Youchun Wang, Changfa Fan, Zhixin Tian, Ming Shi, Fu-Sheng Wang, Lunzhi Dai, Min Wu, Gen Li, Guangyu Wang, Yong Peng, Zhiyong Qian, Canhua Huang, Johnson Yiu-Nam Lau, Zhenglin Yang, Yuquan Wei, Xiaobo Cen, Xiaozhong Peng, Chuan Qin, Kang Zhang, Guangwen Lu, and Xiawei Wei. A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature*, 586(7830):572–577, October 2020.

[343] Xing-Lou Yang, Ben Hu, Bo Wang, Mei-Niang Wang, Qian Zhang, Wei Zhang, Li-Jun Wu, Xing-Yi Ge, Yun-Zhi Zhang, Peter Daszak, Lin-Fa Wang, and Zheng-Li Shi. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. *Journal of virology*, 90(6):3253–3256, 30 December 2015.

[344] Yuanling Yu, Meiyu Wang, Xiaoai Zhang, Shufen Li, Qingbin Lu, Haolong Zeng, Hongyan Hou, Hao Li, Mengyi Zhang, Fei Jiang, Jiajing Wu, Ruxia Ding, Zehua Zhou, Min Liu, Weixue Si, Tao Zhu, Hangwen Li, Jie Ma, Yuanyuan Gu, Guangbiao She, Xiaokun Li, Yulan Zhang, Ke Peng, Weijin Huang, Wei Liu, and Youchun Wang. Antibody-dependent cellular cytotoxicity response to SARS-CoV-2 in COVID-19 patients. *Signal transduction and targeted therapy*, 6(1):346, 24 September 2021.

[345] Meng Yuan, Deli Huang, Chang-Chun D Lee, Nicholas C Wu, Abigail M Jackson, Xueyong Zhu, Hejun Liu, Linghang Peng, Marit J van Gils, Rogier W Sanders, Dennis R Burton, S Momsen Reincke, Harald Prüss, Jakob Kreye, David Nemazee, Andrew B Ward, and Ian A Wilson. Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science*, 20 May 2021.

[346] Meng Yuan, Hejun Liu, Nicholas C Wu, Chang-Chun D Lee, Xueyong Zhu, Fangzhu Zhao, Deli Huang, Wenli Yu, Yuanzi Hua, Henry Tien, Thomas F Rogers, Elise Landais, Devin Sok, Joseph G Jardine, Dennis R Burton, and Ian A Wilson. Structural basis of a shared antibody response to SARS-CoV-2. *Science*, 369(6507):1119–1123, 28 August 2020.

[347] Meng Yuan, Nicholas C Wu, Xueyong Zhu, Chang-Chun D Lee, Ray T Y So, Huibin Lv, Chris K P Mok, and Ian A Wilson. A highly conserved cryptic epitope in the

receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science*, 368(6491):630–633, 8 May 2020.

[348] Ali M Zaki, Sander van Boheemen, Theo M Bestebroer, Albert D M E Osterhaus, and Ron A M Fouchier. Isolation of a novel coronavirus from a man with pneumonia in saudi arabia. *The New England journal of medicine*, 367(19):1814–1820, 8 November 2012.

[349] Tian-Hao Zhang, Lei Dai, John P Barton, Yushen Du, Yuxiang Tan, Wenwen Pang, Arup K Chakraborty, James O Lloyd-Smith, and Ren Sun. Predominance of positive epistasis among drug resistance-associated mutations in HIV-1 protease. *PLoS genetics*, 16(10):e1009009, October 2020.

[350] Wenjuan Zhang, Brian D Davis, Stephanie S Chen, Jorge M Sincuir Martinez, Jasmine T Plummer, and Eric Vail. Emergence of a novel SARS-CoV-2 variant in southern california. *JAMA: the journal of the American Medical Association*, 325(13):1324–1326, 6 April 2021.

[351] Daming Zhou, Wanwisa Dejnirattisai, Piyada Supasa, Chang Liu, Alexander J Mentzer, Helen M Ginn, Yuguang Zhao, Helen M E Duyvesteyn, Aekkachai Tuekprakhon, Rungtiwa Nutalai, Beibei Wang, Guido C Paesen, Cesar Lopez-Camacho, Jose Slon-Campos, Bassam Hallis, Naomi Coombes, Kevin Bewley, Sue Charlton, Thomas S Walter, Donal Skelly, Sheila F Lumley, Christina Dold, Robert Levin, Tao Dong, Andrew J Pollard, Julian C Knight, Derrick Crook, Teresa Lambe, Elizabeth Clutterbuck, Sagida Bibi, Amy Flaxman, Mustapha Bittaye, Sandra Belij-Rammerstorfer, Sarah Gilbert, William James, Miles W Carroll, Paul Klenerman, Eleanor Barnes, Susanna J Dunachie, Elizabeth E Fry, Juthathip Mongkolsapaya, Jingshan Ren, David I Stuart, and Gavin R Screaton. Evidence of escape of SARS-CoV-2 variant b.1.351 from natural and vaccine-induced sera. *Cell*, 23 February 2021.

[352] Hong Zhou, Xing Chen, Tao Hu, Juan Li, Hao Song, Yanran Liu, Peihan Wang, Di Liu, Jing Yang, Edward C Holmes, Alice C Hughes, Yuhai Bi, and Weifeng Shi. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Current biology: CB*, 30(19):3896, 5 October 2020.

[353] Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, Hui-Dong Chen, Jing Chen, Yun Luo, Hua Guo, Ren-Di Jiang, Mei-Qin Liu, Ying Chen, Xu-Rui Shen, Xi Wang, Xiao-Shuang Zheng, Kai Zhao, Quan-Jiao Chen, Fei Deng, Lin-Lin Liu, Bing Yan, Fa-Xian Zhan, Yan-Yi Wang, Geng-Fu Xiao, and Zheng-Li Shi. A pneumonia outbreak associated

with a new coronavirus of probable bat origin. *Nature*, 579(7798):270–273, March 2020.

[354] Tongqing Zhou, Yaroslav Tsybovsky, Jason Gorman, Micah Rapp, Gabriele Cerutti, Gwo-Yu Chuang, Phinikoula S Katsamba, Jared M Sampson, Arne Schön, Jude Bimela, Jeffrey C Boyington, Alexandra Nazzari, Adam S Olia, Wei Shi, Mallika Sastry, Tyler Stephens, Jonathan Stuckey, I-Ting Teng, Pengfei Wang, Shuishu Wang, Baoshan Zhang, Richard A Friesner, David D Ho, John R Mascola, Lawrence Shapiro, and Peter D Kwong. Cryo-EM structures of SARS-CoV-2 spike without and with ACE2 reveal a pH-Dependent switch to mediate endosomal positioning of Receptor-Binding domains. *Cell host & microbe*, 28(6):867–879.e5, 9 December 2020.

[355] Na Zhu, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, Baoying Huang, Weifeng Shi, Roujian Lu, Peihua Niu, Faxian Zhan, Xuejun Ma, Dayan Wang, Wenbo Xu, Guizhen Wu, George F Gao, Wenjie Tan, and China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in china, 2019. *The New England journal of medicine*, 382(8):727–733, 20 February 2020.

[356] Seth J Zost, Pavlo Gilchuk, James Brett Case, Elad Binshtein, Rita E Chen, Joseph P Nkolola, Alexandra Schäfer, Joseph X Reidy, Andrew Trivette, Rachel S Nargi, Rachel E Sutton, Naveenchandra Suryadevara, David R Martinez, Lauren E Williamson, Elaine C Chen, Taylor Jones, Samuel Day, Luke Myers, Ahmed O Hassan, Natasha M Kafai, Emma S Winkler, Julie M Fox, Swathi Shrihari, Benjamin K Mueller, Jens Meiler, Abishek Chandrashekar, Noe B Mercado, James J Steinhardt, Kuishu Ren, Yueh-Ming Loo, Nicole L Kallewaard, Broc T McCune, Shamus P Keeler, Michael J Holtzman, Dan H Barouch, Lisa E Gralinski, Ralph S Baric, Larissa B Thackray, Michael S Diamond, Robert H Carnahan, and James E Crowe, Jr. Potently neutralizing and protective human antibodies against SARS-CoV-2. *Nature*, 584(7821):443–449, August 2020.

[357] Seth J Zost, Pavlo Gilchuk, Rita E Chen, James Brett Case, Joseph X Reidy, Andrew Trivette, Rachel S Nargi, Rachel E Sutton, Naveenchandra Suryadevara, Elaine C Chen, Elad Binshtein, Swathi Shrihari, Mario Ostrowski, Helen Y Chu, Jonathan E Didier, Keith W MacRenaris, Taylor Jones, Samuel Day, Luke Myers, F Eun-Hyung Lee, Doan C Nguyen, Ignacio Sanz, David R Martinez, Paul W Rothlauf, Louis-Marie Bloyet, Sean P J Whelan, Ralph S Baric, Larissa B Thackray, Michael S Diamond, Robert H Carnahan, and James E Crowe, Jr. Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nature medicine*, 26(9):1422–1427, September 2020.

[358] Seth J Zost, Nicholas C Wu, Scott E Hensley, and Ian A Wilson. Immunodominance and antigenic variation of influenza virus hemagglutinin: Implications for design of universal vaccine immunogens. *The Journal of infectious diseases*, 219(Suppl_1):S38–S45, 8 April 2019.