bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Naive human B cells engage the receptor binding domain of SARS-CoV-2, variants of concern, and related sarbecoviruses

Jared Feldman[1†], Julia Bals[1†], Clara G. Altomare[2], Kerri St. Denis[1], Evan C. Lam[1], Blake M. Hauser[1], Larance Ronsard[1], Maya Sangesland[1], Thalia Bracamonte Moreno[1], Vintus Okonkwo[1], Nathania Hartojo[1] Alejandro B. Balazs[1], Goran Bajic[2], Daniel Lingwood[1*] and Aaron G. Schmidt[1,3*]

[1]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, 02139, USA

[2] Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029

[3]Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA

*Correspondence to: Email: dlingwood@mgh.harvard.edu (D.L.); aschmidt@crystal.harvard.edu (A.G.S)

†These authors contributed equally to this work.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Exposure to a pathogen elicits an adaptive immune response aimed to control and eradicate. Interrogating the abundance and specificity of the naive B cell repertoire contributes to understanding how to potentially elicit protective responses. Here, we isolated naive B cells from 8 seronegative human donors targeting the SARS-CoV-2 receptor-binding domain (RBD). Single B cell analysis showed diverse gene usage with no restricted complementarity determining region lengths. We show that recombinant antibodies engage SARS-CoV-2 RBD, circulating variants, and pre-emergent coronaviruses. Representative antibodies signal in a B cell activation assay and can be affinity matured through directed evolution. Structural analysis of a naive antibody in complex with spike shows a conserved mode of recognition shared with infection-induced antibodies. Lastly, both naive and affinity-matured antibodies can neutralize SARS-CoV-2. Understanding the naive repertoire may inform potential responses recognizing variants or emerging coronaviruses enabling the development of pan-coronavirus vaccines aimed at engaging germline responses.

One Sentence Summary: Isolation of antibody germline precursors targeting the receptor binding domain of coronaviruses.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MAIN TEXT

Initial exposure to viral antigens by natural infection or vaccination primes an immune response and often establishes an immune memory which can prevent or control future infections. The naive repertoire contains potential B cell receptor (BCR) rearrangements capable of recognizing these antigens, often the surface-exposed glycoproteins. An early step in generating humoral immunity involves activation of these naive B cells through recognition of a cognate antigen (*1*) which in turn can lead to affinity maturation through somatic hypermutation (SHM) and subsequent differentiation (*2*). The initial engagement of the naive repertoire begins this cascade and often coincides with the eventual generation of a protective or neutralizing antibody response (*3, 4*).

For SARS-CoV-2, the etiological agent of COVID-19, the development of a neutralizing antibody response after primary infection or vaccination is associated with protection against reinfection in non-human primates (*5-9*). In humans, the presence of neutralizing antibodies can predict disease severity and survival after primary SARS-CoV-2 infection (*10*) or vaccination (*11*) and correlates with protection from symptomatic secondary infection (*12, 13*). Further, the two arms of humoral immune memory, long-lived bone marrow plasma cells (*14*) and circulating memory B cells (*15-19*), were induced by natural infection in humans and may persist for at least 8 months after primary infection providing potentially durable long-term protection. Comparable levels of neutralizing antibody titers were present in convalescent COVID-19 subjects and vaccine recipients (*20-22*) further supporting the role of adaptive immune responses in helping to control and prevent disease severity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Both infection- and vaccine-elicited antibodies target the major envelope glycoprotein, spike, present on the virion surface (*23*). A substantial component of the neutralizing response engages the receptor binding domain (RBD) (*24-29*) and does so by directly blocking interactions with the viral receptor ACE2 (*30-35*). Isolated RBD-directed monoclonal antibodies derive from diverse heavy- and light-chain variable gene segments suggesting that multiple biochemical solutions for developing RBD-directed antibodies are encoded within the human B-cell repertoire (*24, 26, 29, 36*). Potential immunogenicity of this antigenic site is based on the human naive B cell repertoire, and the overall frequency of naive BCRs that have some level of intrinsic affinity to stimulate their elicitation (*37-40*). However, antigen-specificity of naive B cells is largely undefined.

Traditional approaches for studying antigen-specific naive B cells include bioinformatic mining of available BCR datasets and inference of likely germline precursors by "germline-reverting" mature BCR sequences, which can be limited by the availability of heavy and light chain paired sequence data and unreliable CDR3 (complementarity-determining region 3) loop approximation, respectively. Here, we address this limitation by characterizing human naive B cells specific for the SARS-CoV-2 RBD directly from the peripheral blood of seronegative donors to understand their relative abundance, intrinsic affinity, and potential for activation. Furthermore, we asked whether the SARS-CoV-2 specific naive repertoire could also engage related circulating variants of concern and pre-pandemic CoVs. We find that SARS-CoV-2 RBD-specific naive B cells were of unrestricted gene usage and several isolated B cells had affinity for circulating SARS-CoV-2 variants and related CoV-RBDs. We determined the structure of a representative naive antibody that binds the SARS-CoV-2 RBD with a mode of recognition similar to a multi-donor class of antibodies prevalent in human responses to SARS-CoV-2 infection (*41*). Further, we improved the

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

affinity for two representative naive antibodies to RBD and showed that the starting naive specificity dictated the breadth of evolved clones to circulating variants. The analysis of the human naive antigen-specific B cell repertoire for the SARS-CoV-2 RBD and its capacity to recognize related variants and emerging CoVs may inform the rational design of epitope-focused immunogens for next generation vaccines.

Isolated SARS-CoV-2-specific naive B cells are genetically diverse

To measure the reactivity of naive human B cells specific for the SARS-CoV-2 RBD we adapted an *ex vivo* B cell profiling approach used previously to study epitope-specific naive precursors targeting neutralizing sites on HIV (*42-44*) and influenza virus surface glycoproteins (*37*). We first designed a SARS-CoV-2 RBD construct that positions two glycans at residues 475 and 501 to selectively block binding to ACE2 and the receptor-binding motif (RBM)-directed antibody, B38 (**fig. S1**) (*45*). Using this "ΔRBM" probe, in addition to wildtype SARS-CoV-2 spike, and RBD probes, we isolated naive ($CD19^+/IgD^+/IgG^-$) B cells specific to the RBD and, more finely, the RBM from the peripheral blood of 8 SARS-CoV-2 seronegative human donors (**Fig. 1A** and **fig. S1E**). We defined RBM-specificity as B cells that bound to fluorescently labeled spike and RBD, but not the ΔRBM probe (**fig. S2A**). Although rare, all 8 donors had detectable populations of RBM-specific naive B cells (**fig. S2B**). The median frequency of RBM-specific B cells among total and naive B cells was 0.0021% and 0.0023%, respectively (**Fig. 1B**). Within spike-reactive, naive cells, the median frequency of RBM-specific B cells was 3.2% (**Fig. 1C**); this potentially suggests that a large proportion of spike epitopes targeted by naive responses reside outside of the RBD. The majority of IgD^+ RBM-specific B cells were $CD27^-$ (mean frequency ~97%), in agreement with the naive B cell phenotype (**fig. S2C**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To understand in more detail the properties of this naive repertoire, we obtained 163 paired heavy- and light-chain antibody sequences from 5 of the 8 donors **(Fig. 1D** and **Table S1)**. Sequence analysis showed that all clones were unique with diverse gene usage for both heavy and light chains and minimal gene pairing preferences (**Table S1**). These data reflect the polyclonal gene usage observed in RBD-specific memory B cells sequenced from COVID-19 convalescent individuals (*26, 29, 36*) and vaccine recipients (*23*), suggesting that a diverse pool of antibody precursors can be activated upon antigen exposure. In comparing this naive repertoire to gene usage distribution from non-SARS-CoV-2-specific repertoires (*46*), we observed an increase in mean repertoire frequency of ~20% for IGHV3-9 in 4 out of 5 sequenced donors (**fig. S3A**). Notably, this enrichment of IGHV3-9 was also observed in isolated memory B cells from convalescent individuals (*47*) and vaccine recipients (*23*), as well as in expanded IgG^+ B cells sequenced from a cohort of COVID-19 subjects during acute infection (*36*). These expanded clones detected shortly after symptom onset displayed low levels of SHM (*36*), suggesting potential IGHV3-9 usage in an early extrafollicular response in which naive B cells differentiate into short-lived plasma cells (*48*). Additionally, IGHV3-53 and 3-30 gene segments, over-represented in RBD-specific antibodies isolated from convalescent subjects (*27, 35, 49*), were recovered from three sequenced donors (13 total clones; ~8.0% of total). The amino acid length of heavy and light chain third complementarity-determining regions (CDR3) ranged from 8 to 27 (average length ~16) for HCDR3 and 4 to 13 (average length ~10) for LCDR3 (**Fig. 1E**). These lengths are normally distributed relative to both unselected human repertoires (*46, 50*) and RBD-specific memory B cell repertoires (*23, 26, 27, 29*); this is in contrast to antibody precursors targeting the influenza and HIV receptor binding sites which have strict requirements for length

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(*51*) or gene usage (*52, 53*). These data suggest that overall HCDR3 length does not restrict precursor frequency and there appears no inherent bias for CDR3 length conferring RBM-specificity. The majority of obtained sequences were at germline in both the variable heavy (V_H) and light (V_L) chains. However, despite sorting B cells with a naive phenotype, some sequences were recovered that deviated from germline. Specifically, the V_H ranged from 91.4 to 100% identity to germline, with a median of 99.7%; the V_L ranged from 93.6 to 100%, with a median of 99.3% (**Fig. 1F, fig. 2B, C**).

Naive antibodies engage SARS-CoV-2 RBD with high affinity

To obtain affinities of the isolated naive antibodies, we cloned and recombinantly expressed 44 IgGs selected to reflect the polyclonal RBD-specific repertoire with representatives from diverse variable region gene segments (**Table S1**). Additionally, we ensured diversity in terms of HCDR3 length, kappa and lambda usage, as well as representation from all 5 donors. By ELISA, we identified IgGs with detectable binding to SARS-CoV-2 RBD; we summarize these results for all antibodies (**Fig. 2A**) and parsed by donor (**fig. S3D**). Across 5 donors, 36 (~81%) bound to monomeric SARS-CoV-2 RBD (**Fig. 2A**) with EC_{50} values ranging from 3.3 to 410 nM and a mean of 62 nM (**Fig. 2A** and **fig. S3E**). These antibodies included 32 unique variable heavy and light chain pairings (**Table S1**). Of the binding population, there is no apparent predisposition for HCDR3 length or light chain pairing (**Fig. 2 C, D**). We further defined the epitopic region of these IgGs using the ΔRBM construct and the individual glycan variants, Δ501 and Δ475, both of which independently block ACE2 cell-surface binding but are on opposite sides of the RBM (**fig. S1E, F**). 11 IgGs had no detectable ΔRBM binding (e.g., ab079, ab119), while 21 IgGs had reduced ELISA binding relative to wild-type RBD, reflected in the reduced ΔRBM median EC_{50} values

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(**fig. S3E**). We also identified examples of antibodies sensitive to only Δ475 (e.g., ab185) and only Δ501 (e.g., ab007) (**Fig. 2A** and **fig. S3E).**

To obtain binding kinetics independent of avidity effects from bivalent IgGs, 12 antibodies were selected for expression as Fabs to determine monovalent binding affinity (K_{DS}) by biolayer interferometry (BLI). Using monomeric RBD as the analyte, 10 of the 12 Fabs had detectable binding with K_{DS} ranging from ~6.5 to ~75 μM; the other two remaining Fabs (ab177, ab185), gave unreliable affinity measurements (i.e., >100 μM) (**fig. S4**). Notably, all Fabs had characteristically fast off rates (k_{off}). This observation is consistent for germline B cells where fast off-rates are compensated by avidity due to overall BCR surface density (*54*); subsequent affinity gains via SHM often result in slowing of the off-rate and is a canonical mechanism of improved antigen binding (*55-57*).

Naive antibodies engage SARS-CoV-2 variants of concern

The emergence of SARS-CoV-2 variants with mutations in RBD has raised significant concern that antigenic evolution will impair recognition of RBD-directed antibodies elicited by prior infection and vaccination with an antigenically distinct SARS-CoV-2 variant (*58-61*). We therefore asked whether these naive antibodies, isolated using wild-type SARS-CoV-2 RBD, could recognize circulating viral variants, B.1.1.7 (mutations N501Y) (*62*) and B.1.351 (mutations K417N/E484K/N501Y) (*63*); the former has now become the most prevalent circulating variant in the US and many other countries (*64*). we find that 89% of the antibodies with wild-type RBD affinity also bound to the B.1.1.7 variant with a comparable mean affinity of 68.9 nM (**Fig. 2D, F**). For B.1.351, a concerning variant prevalent in South Africa (*64*), 62% of the wild-type SARS-

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CoV-2 RBD binding IgGs also bound to the B.1.351 variant, many of which displayed reduced ELISA binding relative to wild-type RBD with a mean affinity of 262 nM (**Fig. 2E, F**). A more pronounced reduction in cross-reactivity to the B.1.351 variant may be predictive of reduced sera binding and neutralization titers from convalescent individuals and vaccine-recipients (*22, 58, 65*).

Naive antibodies engage pre-emerging CoVs

We next tested the cross-reactivity of these naive antibodies to related sarbecovirus RBDs, which also use ACE2 as a host receptor (*66*). Our panel included the previously circulating SARS-CoV RBD and representative preemergent bat CoV RBDs from WIV1 (*67*), RaTG13 (*68*), and SHC014 (*69*). These RBDs share 73 to 90% paired-sequence identity with the highest degree of amino acid conservation in residues outside of the RBM (*70*). 13 antibodies cross-reacted with at least one additional RBD in our panel, with decreasing affinity for RBDs with more divergent amino acid sequence identity (**Fig. 2A, G**). Notably, ab017, ab072, ab109, and ab114 had broad reactivity to all tested sarbecovirus RBDs, suggesting binding to highly conserved epitopes. Of these cross-reactive antibodies, ab017 and ab114, derive from the same IGHV3-33 and IGVL2-14 paring but were isolated from different donors, suggesting a shared or public clonotype.

Naive antibodies are not polyreactive and do not engage seasonal coronaviruses.

Prior studies have shown that germline antibodies are more likely to display polyreactivity relative to affinity-matured antibodies with higher levels of SHM from mature B cell compartments (*71-74*). We therefore tested the polyreactivity of all 44 naive antibodies using three common autoantigens, double-stranded DNA (dsDNA), *Escherichia coli* lipopolysaccharide (LPS), and human insulin in ELISA (**Fig. 2A**) We observed no polyreactivity of any naive antibody, including

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

those that are broadly reactive. Furthermore, none of the naive antibodies bound RBDs from the human seasonal betacoronaviruses (hCoVs), OC43 and HKU1 (**Fig. 2A**), which share 22 and 19% paired-sequence identity to SARS-CoV-2 RBD, respectively. Together, these results suggest that the isolated naive B cells encode BCRs with specificity to sarbecoviruses.

In vitro reconstitution of naive B cell activation

Physiological interactions between a naive BCR and cognate antigen occurs at the B cell surface. Naive BCRs are displayed as a bivalent membrane-bound IgM and multivalent antigen binding can initiate intracellular signaling resulting in an activated B cell with the capacity to differentiate to antibody secreting plasma cells or memory cells (*75*). To determine whether the isolated RBD-specific naive BCRs have the capacity to be activated, we generated stable Ramos B cell lines expressing ab090 or ab072 as cell-surface BCRs and measured their activation by monitoring calcium flux *in vitro* (*76*). These antibodies were selected to represent divergent germline gene usage and specificities: 1) ab090 (IGHV1-2/IGKV3-15) bound SARS-CoV-2 and variant B.1.1.7 RBDs, but not variant B.1.351 and WIV1 RBDs (**Fig. 3A**); and 2) ab072 (IGHV3-23/IGLV2-14) had broad reactivity to all RBDs (**Fig. 3B**). To assess BCR activation, we generated ferritin-based nanoparticles (NPs) for multivalent RBD display using SpyTag-SpyCatcher (*70*, *77*, *78*); these RBD NPs included SARS-CoV-2, B.1.1.7, B.1.351 and WIV RBDs. We found that ab090 expressing Ramos B cells were only activated by SARS-CoV-2 RBD and variant B.1.1.7 RBD NPs (**Fig. 3C**), while ab072 Ramos B cells were activated by all RBD-NPs (**Fig. 3D**). Notably, these data parallel the observed recombinant binding specificity of each antibody. Importantly, neither ab090 nor ab072 Ramos B cell lines were activated by influenza hemagglutinin NPs, suggesting that this activation is sarbecovirus RBD-specific (**Fig. 3C, D**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ab090 engages the SARS-CoV-2 RBM

To further characterize the epitope specificity of a representative naive antibody, we determined the structure of ab090 in complex with SARS-CoV-2 spike (S) by electron cryomicroscopy (cryo-EM). A ~6.7-Å structure showed one Fab bound to an RBD in the "up" conformation (**Fig. 4A, B and fig. S5)**. Based on this modest resolution structure, we make the following general descriptions of the antibody-antigen interface. The interaction between ab090 and the RBD is mediated primarily by the antibody heavy chain, with the germline encoded HCDR1, HCDR2, and the framework 3 DE-loop centered over the RBM epitope (**Fig. 4B**). The ab090 light chain is oriented distal to the RBD and does not appear to substantially contribute to the paratope (**Fig. 4B**). IGHV1-2 antibodies represent a prevalent antibody class in human responses to SARS-CoV-2 infection, many of which display high neutralization potency (*41, 79*). ab090 shares a V_H-centric mode of contact and angle of approach similar to members of this class of infection-elicited antibodies (**Fig. 4D**), despite varying HCDR3 lengths and diverse light chain pairings (**Fig. 4D**) (*41, 79*). Additionally, members of the IGHV1-2 antibody class contain relatively few SHMs (**fig. S5B**). We note that many of the infection-elicited IGHV1-2 RBD-specific memory B cells derive from the IGHV1-2*02 allele, while ab090 is encoded by the IGHV1-2*06 allelic variant (**fig. S5B**). The IGHV1-2*06 allele is represented by a single nucleotide polymorphism encoding an arginine rather than a tryptophan at position 50 (*80*) (**fig. S5B**). Notably, a potent neutralizing antibody, H4, derives from the same *06 allele (*34*). In conjunction with the structure, we biochemically defined the sensitivity of ab090 to variant B.1.351 by testing the binding to individual mutations. Binding affinity was detected to SARS-CoV-2 RBDs with either N501Y or K417N mutations, but not to E484K alone (**Fig. 4C**). Based on the structure, the E484K mutation, is grossly positioned

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proximal to the CDRH2 loop (**Fig. 4C**), which has a germline-encoded motif critical for IGHV1-2 antibody binding to RBD (**fig. S5B**) (*41*). Indeed, infection-elicited IGHV1-2 antibodies are susceptible to escape by E484K alone, which disrupts a CDRH2 hydrogen binding network (*81*). Together, the cryo-EM structure and binding data suggest that ab090 represents a precursor of a class of RBM directed SARS-CoV-2 neutralizing antibodies. More generally, structural characterization of germline antibody complexes has been limited to hapten antigens (*82*), simple peptides (*83*) and to protein antigens bearing engineered affinity to inferred germline sequences(*84*). We present, to our knowledge, the first structure of a naturally occurring naive human antibody bound to non-engineered viral protein.

***in vitro* affinity-matured naive antibodies retain intrinsic specificity**

After initial antigen recognition and subsequent activation, naive B cells can undergo successive rounds of somatic hypermutation within the germinal center (GC) that ultimately result in higher affinity antibodies for the cognate antigen. To determine how somatic hypermutation might influence overall affinity and specificity, we used yeast surface display to *in vitro* mature ab072 and ab090. We randomly mutagenized the single chain variable fragment (scFv) variable heavy and light chain regions to generate ab072 and ab090 variant display libraries (*85*). After two rounds of selections using SARS-CoV-2 RBD, we enriched the ab072 and ab090 libraries for improved binding over their respective parental clones (**Fig. 5A, D and fig. S6A**). We also observed increased binding to B.1.351 for the ab072 library but not for ab090; notably this corresponded with the respective specificity of the parent clones (**Fig. 5A, D**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We next isolated and sequenced individual clones from the enriched libraries. For ab090, we observed a dominant mutation, R72H, in the FRWH3 region present in 60% of sequenced clones (**fig. S6B**). Notably, multiple mutations at position 72 conferred a ~3- to 5-fold improvement in monovalent affinity relative to parental ab090 for wild-type and B.1.1.7 RBDs, with no detectable B.1.351 binding for affinity matured progeny (**Fig. 5B, C**). We observed no mutations within the light chain which appears to be consistent with the V_H-centric binding mode in the cryo-EM structure (**Fig. 4**). For the broadly reactive ab072, isolated clones had mutations in both the V_H and V_L ; ~35% of the sequenced clones had mutation S31P in the HCDR1 (**fig. S6B, C**). There was 3- to ~5-fold improvement in monovalent affinity of ab072 progeny relative to parent for SARS-CoV-2, B.1.1.7 and B.1.351 RBDs (**Fig. 5E, F)**. Collectively, these data identify potential mutations that can improve affinity while retaining initial parental antigen specificity.

SARS-CoV-2 pseudovirus neutralization by naive and affinity-matured Abs

We next used a SARS-CoV-2 pseudovirus assay (*10*) to ask whether any of the isolated naive antibodies and affinity matured clones were capable of blocking transduction of target cells. We found that of the 36 RBD-binding antibodies tested in this assay, 5 had detectable levels of neutralization (~14%) (**Fig. 6A**). These antibodies, obtained from multiple donors, have no commonality with respect to their gene usages and HCDR3 lengths (**Fig. 6B**). While these naive antibodies were not as potent as B38, isolated from a memory B cell (*34*), the observation, nevertheless, that the naive repertoire has antibodies that neutralize is noteworthy.

To determine whether improved affinity correlated with enhanced neutralization potency, we evaluated the affinity matured progeny of ab090 in a SARS-CoV-2 pseudovirus neutralization

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

assay (**Fig. 6B**). We find that all three ab090 progeny that had higher affinity for SARS-CoV-2 RBD also had increased neutralization potency. ab090_A08 bearing the R72H mutation had the highest affinity gain and was the most potent neutralizer with a K_D of 1.7μM and an IC50 of 0.37 μg/ml, respectively. Notably, ab090 progeny had IC50 values similar to other IGHV1-2 memory B cells isolated from convalescent donors (*41*); this increase in potency is conferred through minimal somatic hypermutation.

DISCUSSION

The development of a protective humoral immune response upon infection or vaccination relies on the recruitment, activation, and maturation of antigen-specific naive B cells. However, the specificity of the naive B cell repertoire remains largely undefined. Here, we showed that coronavirus-specific naive B cells are present across distinct seronegative donors, are of unrestricted gene usage and when recombinantly expressed as IgGs, have affinity for SARS-CoV-2 RBD, circulating variants of concern, and at least four related coronaviruses. These data suggest that RBD-specific precursors are likely present across a large fraction of individual human naive repertories, consistent with longitudinal studies of SARS-CoV-2 infected individuals in which most convalescent individuals seroconverted with detectable RBD serum antibodies and neutralization titers (*17, 86, 87*). The naive B cells characterized here engage epitopes across the RBM with a range of angles of approach as defined by our glycan variant probes and cross-reactivity profiles; this is also consistent with infection and vaccine elicited, RBD-specific repertoire characterized by epitope-mapping, deep mutational scanning and structural analyses (*30, 32, 88*). Having naive BCRs recognizing distinct or partially overlapping epitopes across the

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RBM may be advantageous for eliciting a polyclonal response more able to recognize variants of concern.

The presence of broadly reactive naive B cells inherently capable of recognizing sarbecovirus RBDs and circulating variants suggests that these precursors could be vaccine-amplified. Recent work showed that uninfected individuals have pre-existing SARS-CoV-2 S-reactive serum antibodies (*89-91*) and memory B cells (*28, 92*) which cross-react with hCoVs and can be boosted upon SARS-CoV-2 infection. These cross-reactive antibodies appear to be specific to the S2 domain and are predominantly IgG or IgA. Notably, this observation contrasts the cross-reactive B cells described here that engage the RBD, have no reactivity to hCoV and are IgG^{-} naive B cells suggesting that they are distinct from previously described S-reactive pre-existing antibodies.

Data suggests that the competitive success of a naive B cell within a GC is influenced by precursor frequencies and antigen affinities (*40*). However, the biologically relevant affinities necessary for activation remain unclear—indeed several studies suggest that B cell activation and affinity maturation is not restricted by immeasurably low affinity BCR interactions (*93-95*). Recently, two studies involving naive precursors of receptor-binding site (RBS) directed HIV-1 broadly neutralizing antibodies (bnAbs) contributed to our understanding of these parameters (*38, 39*). Using an *in vivo* murine adoptive transfer model, these RBS-directed precursors were recruited into a GC reaction at a precursor frequency of ~1:10,000 and a monovalent antigen affinity of 14μM (*39*). For comparison, here we defined the SARS CoV-2 RBM-specific naive precursor frequency as 1:41,000 by flow cytometric gating (**fig. S2**) with monovalent affinities ranging from 6.5 to >100μM. These data suggest that these isolated naive B cells, especially those with

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

demonstrable monomeric affinity, could be readily elicited upon antigen exposure. However, longitudinal studies tracking antigen specific naive B cells pre- and post-exposure are required to determine the fate (i.e., plasma cell, memory, or germinal center B cell compartments) of potential precursors and define relevant naive affinities for elicitation by SARS-CoV-2.

Through biochemical and structural analyses, we characterized a naive antibody, ab090, which resembles a commonly elicited class of potent neutralizing antibodies utilizing the IGHV1-2 gene (*41*). This class of antibodies share restricted binding specificity for wild-type SARS-CoV RBD (the vaccine strain) and the prevalent B.1.1.7 variant. This recombinant binding pattern also paralleled the reconstituted *in vitro* B cell activation dynamics of ab090 in the highly avid assay with the capacity to detect immeasurably low affinity interactions (*54*). *In vitro* affinity maturation of ab090 against corresponded to a single H-FR3 mutation, which improved monovalent affinity ~5-fold to wild-type SARS-CoV-2 and B.1.1.7 RBDs relative to parent and pseudovirus neutralization to IC50 values less than 1µg/ml. This observation is consistent with the low levels of SHM within IGHV1-2 neutralizing antibodies (*41*) and with reports of other potent RBD-directed neutralizing antibodies with a limited level of somatic hypermutation (*24, 26, 29, 36, 96*). Further, a recent study monitoring RBD-specific memory B cell evolution up to 12 months after SARS-CoV-2 infection revealed examples of affinity matured clones with increased neutralizing breadth over time against circulating RBD variants (*97*). While *in vitro* affinity gains and neutralization potency are generally correlated (*98*), we note that affinity does not necessarily correlate to neutralization potency for all SARS-CoV-2 RBD targeting antibodies, where fine epitope specificity appears to be most relevant (*28, 99*).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Probing and characterizing the human naive B cell antigen-specific repertoire can identify precursors for vaccine or infection-specific naive B cells and expand our understanding of basic B cell biology. Germline-endowed specificity for neutralizing antibody targets on the RBD may also contribute to the strong clinical efficacy observed for the current SARS-CoV-2 vaccines (*100, 101*). Furthermore, understanding the naive B cell repertoire to potential pandemic coronaviruses may reveal commonalties in antigen-specific precursors, enabling the development of pan-coronavirus vaccines aimed at engaging broadly protective germline responses.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

METHODS

Donor Samples

PBMCs were isolated from blood donors obtained from the MGH blood donor center (8 donors total). Prior to donating blood, subjects were required to sign a donor attestation/consent statement, as per hospital requirements, stating ''I give permission for my blood to be used for transfusion to patients or for research''. The gender and age are not recorded, however eligible donors are of at least 16 years old and weigh a minimum of 110lbs. All experiments were conducted with MGH Institutional Biosafety Committee approval (MGH protocol 2014B000035). Isolated PBMCs were used for B cell enrichment and single cell sorting (described below); plasma was aliquoted and stored at -80 ºC until further use. Additionally, the control convalescent sera used for ELISA was obtained under the approved Partners Institutional Review Board (protocol 2020P000895) for use of patient samples for the development and validation of SARS-CoV-2 diagnostic tests (*10*).

Expression and purification of recombinant CoV Antigens

Plasmids encoding the receptor binding domains (RBDs) were designed based on GenBank sequences MN975262.1 (SARS-CoV-2), ABD72970.1 (SARS-CoV), AGZ48828.1 (WIV-1), MN996532.2 (RaTG13), QJE50589.1 (SHC014), AAT98580.1 (HKU1), and AAT84362 (OC43). Constructs were codon optimized and synthesized by IDT. QuikChange Mutagenesis (Agilent) was used to insert glycosylation sites at SARS-CoV-2 RBD residues 501 and/or 475 as well as for RBD variant mutations, B.1.351 (K417N/E484K/N501Y) and B.1.1.7 (N501Y). SARS-CoV-2 spike contained a C-terminal foldon trimerization domain and HRV 3C-cleavable 6xHis and 2xStrep II tags (*102*). All proteins were transiently expressed in Expi293F cells (ThermoFisher). 5 to 7 days post-transfection, supernatants were harvested by centrifugation and further purified

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

using immobilized metal affinity chromatography (IMAC) with cobalt-TALON resin (Takara) followed by Superdex 200 Increase 10/300 GL size exclusion column (GE Healthcare).

Expression and purification IgGs and Fabs

IgG and Fab genes for the heavy- and light-chain variable domains were synthesized and codon optimized by IDT and subcloned into pVRC protein expression vectors and sequence confirmed (Genewiz). Fabs and IgGs were similarly expressed and purified as described above for RBDs. IgGs were buffer exchanged into PBS while Fabs were concentrated and further purified by Superdex 200 Increase 10/300 GL size exclusion column.

ELISA

Both sera and monoclonal antibody reactivity to CoV antigens were assayed by ELISA. Briefly, 96-well plates (Corning) were coated with 5 µg/ml of monomeric RBDs in PBS at 100µl/well and incubated overnight at 4ºC. Plates were blocked with 1% BSA in PBS containing 1% Tween-20 (PBS-T) for 1hr at room temperature (RT). Blocking solution was discarded and 4-fold serial dilutions of human plasma (1:20 starting dilution) or isolated monoclonal antibodies (150 µg/ml starting concentration) in PBS were added to wells and incubated for 1hr at RT. Plates were then washed three times with PBS-T. Secondary, anti-human IgG-HRP (Abcam), was added to each well at 1:20,000 dilution in PBS-T and incubated for 1hr at RT. Plates were then washed three times with PBS-T and developed with 1-Step ABTS substrate (ThermoFisher) per manufacturer recommendations. Absorbance was measured using a plate reader at 405nm. EC_{50} values were determined for monoclonal antibodies by non-linear regression (sigmoidal) using GraphPad Prism

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

8.4.3 software. ELISAs against OC43 and HKU1 RBDs were done at a single IgG concentration (150 µg/ml) in replicate. Positive binding was defined by an $OD_{405} \geq 0.30$.

For polyreactivity ELISAs against human insulin (MilliporeSigma) and dsDNA (Calf Thymus DNA; Invitrogen), plates were coated with 2µg/ml and 50µg/ml, respectively, in PBS at 100µl/well and incubated overnight at 4ºC. Plates were then blocked and incubated with IgGs as described above for CoV antigens. LPS ELISAs were measured according to a previously described method (*103, 104*). Briefly, plates were coated with 30µg/ml LPS (*Escherichia coli* O55:B5; MilliporeSigma) in carbonate buffer (100mM Na_2CO_2, 20mM EDTA, pH 9.6) at 100µl/well for 3hrs at 37ºC, washed three times with water, and air-dried overnight at RT. Coated plates were blocked with 200µl/well of HS buffer (50mM HEPES, 0.15mM NaCl, pH 7.4) plus 10mg/ml. Plates were incubated with IgGs diluted in HS buffer containing 1mg/ml BSA for 3hrs at 37ºC, washed three times with HS buffer, and developed as detailed above for CoV antigens. All polyreacivity ELISAs were performed at a single IgG concentration (15µg/ml) in replicate with positive binding was defined by an $OD_{405} \geq 0.30$.

ACE-2 cell binding assay

ACE-2 expressing 293T cells were incubated with 200 nM of RBD antigen in PBS for 1hr on ice. Cells were resuspended in 50µL of secondary stain containing streptavidin-PE (Invitrogen) at a 1:200 dilution and incubated for 30 min on ice. Cell binding was analyzed by flow cytometry using a Stratedigm S1300Exi Flow Cytometer equipped with a 96 well plate high throughput sampler. Rsulting data were analyzed using FlowJo (10.7.1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Probe Generation

SARS-CoV-2 RBD and ΔRBM constructs were expressed as dimeric murine-Fc (mFc; IgG1) fusion proteins containing a HRV 3C-cleavable C-terminal 8xHis and SBP tags and purified as described above. SBP-tagged RBD- and ΔRBM-mFc dimers were individually mixed with fluorescently labeled streptavidin, SA-BV650 and SA-BV786 (BioLegend), to form RBD-mFc-BV650 and ΔRBM-mFc-BV786 tetramers. SARS-CoV-2 spike with a C-terminal Strep II tag was labeled separately with StrepTactin PE and APC (IBA) to form spike-PE and -APC tetramers, respectively. Both labeling steps were performed for 30 min at 4 ºC prior to sorting.

Single B Cell Sorting

Naive B cells were purified from PBMCs using the MACS Human B Cell isolation kit (Miltenyi Biotec) and incubated with 25nM of each SARS-CoV-2 probe (RBD-mFc-BV650, ΔRBM-mFc-BV786, spike-PE, and spike-APC) for 30 min at 4ºC. Cells were stained with anti-human CD19 (Alexa-700), CD3 (PerCP-Cy5), IgD (PE-Cy7), IgG (BV711), CD27 (BV510), LiveDead Violet (Invitrogen), and Calcien (Invitrogen) for an additional 30 min. RBM-specific naive B cells, defined as $CD19^{+}/CD3^{-}/IgG^{-}/IgD^{+}$/spike PE^{+}/spike $APC^{+}/RBD^{+}/\Delta RBM^{--}$, were single-cell sorted using BD FACS Aria II (BD Biosciences) into 96-well plates containing lysis buffer supplemented with 1% BME. Within the $CD19^{+}/IgG^{-}/IgD^{+}$ gated cells, we also confirmed that 97% of the events were CD27 negative. Plates were stored at -80 ºC for subsequent analysis. Flow cytometry data was analyzed using FlowJo software version 10.7.1.

BCR Sequencing

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

BCR Sequencing was carried out as described previously (*37*). Briefly, whole transcriptome amplification (WTA) was performed on the sorted cell-lysates according to the Smart-Seq2 protocol (*105*). We then amplified heavy and light chain sequences from the WTA products utilizing pools of partially degenerate pools of V region specific primers (Qiagen HotStar Taq Plus). Heavy and light chain amplifications were carried out separately, with each pool containing pooled primers against human IGHV and heavy chain constant region genes, or human IGLV, IGKV, and light chain constant region genes. Cellular barcodes and index adapters (based on Nextera XT Index Adapters, Illumina Inc.) were added using a step-out PCR method. Amplicons were then pooled and sequenced using a 250x250 paired end 8x8 index reads on an Illumina Miseq System. The data were then demultiplexed, heavy and light chain reads were paired, and overlapping sequence reads were obtained (Panda-Seq) (*106*) and aligned against the human IMGT database (*107*).

Interferometry binding experiments

Interferometry experiments were performed using a BLItz instrument (ForteBio). Fabs (0.1 mg/ml) were immobilized on Ni-NTA biosensors. The SARS-CoV-2 RBD analyte was titrated (10µM, 5µM, 2.5µM, and 1µM) to acquire binding affinities; the K_D was obtained through global fit of the titration curves by applying a 1:1 binding isotherm using vendor-supplied software.

Pseudotyped neutralization assay

SARS-CoV-2 neutralization was assessed using lentiviral particles pseudotyped as previously described (*10, 108*). Briefly, lentiviral particles were produced via transient transfection of 293T cells. The titers of viral supernatants were determined via flow cytometry on 293T-ACE2 cells

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(*108*) and via the HIV-1 p24^{CA} antigen capture assay (Leidos Biomedical Research, Inc.). Assays were performed in 384-well plates (Grenier) using a Fluent Automated Workstation (Tecan). IgGs starting at 150 µg/ml, were serially diluted (3-fold) in 20µL followed by addition of 20 µL of pseudovirus containing 250 infectious units and incubated at room temperature for 1 hr. Finally, 10,000 293T-ACE2 cells (*108*) in 20 µL cell media containing 15 µg/ml polybrene were added to each well and incubated at 37 °C for 60-72 hrs. Following transduction, cells were lysed using a previously described assay buffer (*109*) and shaken for 5 min prior to quantitation of luciferase expression using a Spectramax L luminometer (Molecular Devices). Percent neutralization was determined by subtracting background luminescence measured from cells control wells (cells only) from sample wells and dividing by virus control wells (virus and cells only). Data were analyzed using Graphpad Prism.

cryo-EM sample preparation, data collection and processing

SARS-CoV-2 spike HexaPro was incubated with ab090 Fab at 0.6 mg/mL at a molar ratio of 1.5:1 Fab:Spike for 20 minutes at 4°C and two 3 µl aliquots were applied to UltrAuFoil gold R0.6/1 grids and subsequently blotted for 3 seconds at blot force 3 twice, then plunge-frozen in liquid ethane using an FEI Vitrobot Mark IV. Grids were imaged on a Titan Krios microscope operated at 300 kV and equipped with a Gatan K3 Summit direct detector. 10,690 movies were collected in counting mode at 16e^{-}/pix/s at a magnification of 81,000, corresponding to a calibrated pixel size of 1.058 Å. Defocus values were at around -2.00 µm. Micrographs were aligned and dose weighted using Relion's (*110*) implementation of MotionCorr2 (*111*). Contrast transfer function estimation was done in GCTF (*112*). Particles were picked with crYOLO (*113*) with a model trained with 12 manually picked micrographs with particle diameter value of 330Å. Initial processing was

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

performed in Relion. The picked particles were binned to ~ 12Å/pixel and subjected to a 2D classification. Selected particles were then extracted to ~6Å/pixel then subjected to a second round of 2D classification. An initial model was generated on the selected particles at ~6Å/pixel and used as a reference for two rounds of 3D classification; first to select particles containing SARS-CoV-2 spike then to select particles containing both spike and ab090. Selected particles were unbinned then aligned using 3D auto-refine and subjected to a third round of 3D classification to select for a single class with SARS-CoV-2 spike bound with one ab090 Fab. Selected particles were aligned using 3D auto-refine before undergoing CTF refinement and Bayesian polishing. Polished particles were then simultaneously focus-aligned relative to the RBD and ab090 region (Figure S5 A) to aid in model building of this region of interest and imported to cryoSPARC (*114*). Imported particles were aligned using non-uniform refinement and local resolution estimation (Figure S5B). Non-uniform refined maps were then sharpened with DeepEMhancer then used to dock a previously built SARS-CoV-2-spike model (PDB ID 7LQW).

cryo-EM model building

Backbone models were built by docking the variable regions of structurally similar Fabs (PDB ID 2D2P and 6FG1 for heavy and light chains, respectively) and a previously built RBD (6M0J) into the focus refined maps using UCSF Chimera (*115*) variable regions were then mutated and manually built using COOT (*116*). For the remainder of the spike, a previously published model (PDB ID 6VXX) was docked into the full, sharpened map in UCSF Chimera.

RBD nanoparticle production and conjugation

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Monomeric SARS-CoV-2 wild-type, B.1.1.7, B.1.351, and WIV1 RBDs were recombinantly produced and purified as described above with an 8xHis and SpyTag (cite) at the C-terminus. *Helicobacter pylori* ferritin nanoparticles (NP) were expressed separately with N-terminal 8xHis and SpyCatcher tags. SpyTag-SpyCatcher conjugations were performed overnight at 4ºC with a 4-fold molar excess of SpyTag-RBD relative to SpyCatcher-NP. The conjugated RBD-NPs were repurified by size-exclusion chromatography to remove excess RBD-SpyTag.

In vitro BCR triggering

The capacity of RBD-NPs to trigger naive BCR signaling was determined through activation of Ramos cells engineered to display mono-specific IgM BCRs of interest, as previously described (*76*). Briefly, BCRs for ab090 and ab072 were stably expressed in an IgM negative Ramos B cell clone by lentiviral transduction. Five to seven days post transduction, confluent BCR-expressing B cells were FACS sorted on IgM (APC anti-human IgM; BioLegend) and kappa light chain (PE anti-human kappa light chain; BD Biosciences) double positivity using a SH800S Cell Sorter (Sony Biotechnology). Sorted cells were expanded in RPMI (GIBCO) and evaluated for B cell activation by labeling 10 million cells with 0.5µg/ml Fura red dye (Invitrogen) in 2ml of RPMI at 37ºC for 30 min. Cells were then washed and resuspended to 4 million cells/ml in RPMI. BCR triggering was measured in response to the RBD-NPs described above by flow cytometry (LSR II, BD Biosciences) as the ratio of Ca^{2+} bound/unbound states of Fura red. Ratiometric measures for individual B cell lines were normalized to the maximum Ca^{2+} flux as measured by exposure to 10µg/ml ionomycin.

in vitro affinity maturation of ab090 and ab072

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To build yeast display libraries for ab090 and ab072, variable heavy and light chains were reformatted into an scFv and synthesized as gBlocks (IDT). gBlocks were amplified by polymerase chain reaction (PCR) using Q5 polymerase (New England BioLabs) following the manufacturer's protocol. The amplified DNA purified and subsequently mutagenized by error-prone PCR (ePCR) via the GeneMorph II Random Mutagenesis Kit (Agilent Technologies) with a target nucleotide mutation frequency of 0-4.5 mutations/kb. Mutagenized scFv DNA products were combined with the linearized yeast display vector pCHA (*117*) and electroporated into EBY100 grown to mid-log phase in YPD media, where the full plasmid was reassembled by homologous recombination (*85*). The final library size was estimated to be 4 x 10^7.

The scFv libraries and selection outputs were passaged in selective SDCAA media (20 g/L dextrose, 6.7 g/L Yeast Nitrogen base, 5 g/L Bacto casamino acids, 5.4 g/: Na_2HPO_4 and 8.56 g/L $NaH_2PO_4 \cdot H_2O$) at shaking at 30ºC and induced in SGCAA media (same as SDCAA wit 20 g/L galactose instead of dextrose) at 20ºC. The scFv libraries were induced covering at least 10-fold of their respective diversities and subject to three rounds of selection for binding to SBP-tagged SARS-CoV-2 RBD-Fc. Induced yeast libraries were stained for antigen binding (RBD-Fc APC tetramers) and scFv expression (chicken anti-c-myc IgY; Invitrogen). Following two washes in PBSF (1x PBS, 0.1% w/v BSA), yeast was stained with donkey anti-chicken IgY AF488 (Jackson ImmunoResearch). Two gates were drawn for cells with improved RBD binding over parental clones, a more stringent "edge" gate represented ~1% and a "diversity" gate represented ~3-5% of the improved output. Yeast from the final round of selection were resuspended in SDCAA media and plated on SDCAA agar plates for single colony isolation and Sanger sequencing from which IgGs and Fabs were cloned and recombinantly expressed as described above.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We thank members of the Schmidt and Lingwood Labs for helpful discussions, especially Tim Caradonna, Catherine Jacob-Dolan, and Daniel Maurer. We thank Samuel Kazer, James Gatter and Alex Shalek for BCR sequencing advice, Jason McLellan for the SARS-CoV-2 spike plasmid, and Nir Hacohen and Michael Farzan for ACE2 expressing 293T cells. Some of this work was performed at the National Center for CryoEM Access and Training (NCCAT) and the Simons Electron Microscopy Center located at the New York Structural Biology Center, supported by the NIH Common Fund Transformative High Resolution Cryo-Electron Microscopy program (U24 GM129539), and by grants from the Simons Foundation (SF349247) and NY State Assembly.

Funding: We acknowledge support from NIH (R01AI146779, R01AI124378, R01AI137057, R01AI153098, R01AI155447, DP2DA042422, DP2DA040254, T32 AI007245), a Massachusetts Consortium on Pathogenesis Readiness (MassCPR) grant to A.G.S. and a MGH Transformative Scholars Program and Charles H. Hood Foundation to A.B.B.

Author contributions: J.F., J.B., A.B.B., G.B., D.L., A.G.S. designed research; J.F., J.B., C.G.A., K.S.D., E.C.L, B.M.H., L.R., M.S., T.B.M. ,V.O., N.H., performed research; J.F., J.B., C.G.A, G.B., D.L., A.G.S analyzed data; J.F. and A.G.S. wrote the paper. J.F., J.B., C.G.A., B.M.H, A.B.B., G.B., D.L., A.G.S edited and commented on the paper.

Competing interests: Authors declare no competing interests**.**

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data and materials availability: All data are provided in the Supplementary Materials. Requests for material should be addressed to Daniel Lingwood (dlingwood@mgh.harvard.edu) or Aaron G. Schmidt (aschmidt@crystal.harvard.edu). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material. The EM maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession code: EMD-24279.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MAIN TEXT FIGURES

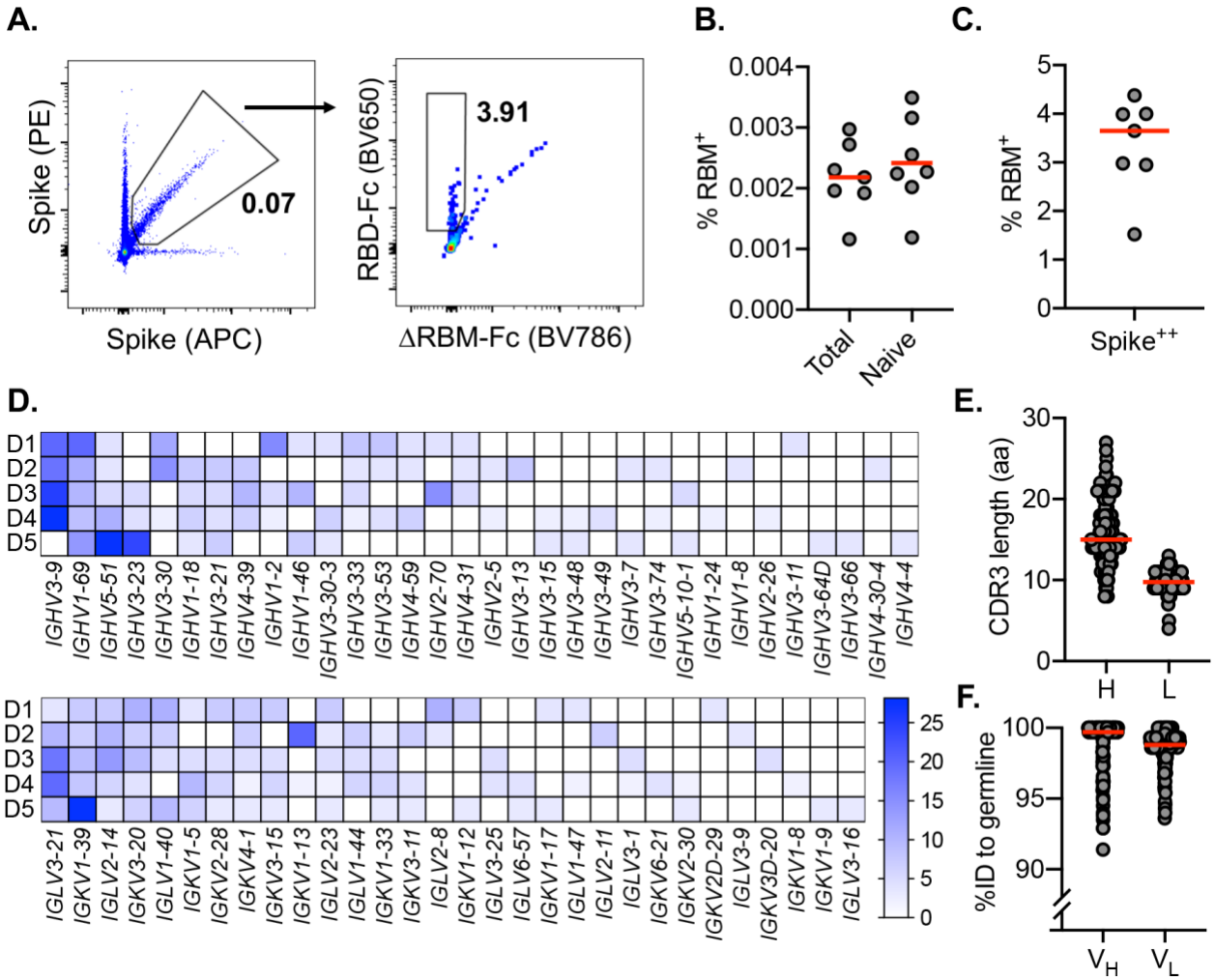


Fig. 1. SARS-CoV-2-specific naive B cells isolation and characterization. **(A)** RBM-specific naive B cells from seronegative human donors were isolated by fluorescence-activated cell sorting gated on $CD19^{+}IgD^{+}IgG^{-}$; representative plot from donors 1 and 2 is shown. ΔRBM is a sorting probe with N-linked glycans at residues 501 and 475. RBM-specific B cell frequency among (**B**) total, naive, and (**C**) spike-positive cells from each donor (n = 8). (**D**) Heat map showing variable-gene usage for all paired B cell sequences. Scale indicates percent of total sequences for each donor separately. (**E**) Heavy (H) and light (L) CDR3 amino acid length distribution determined using IMGT numbering. Red bars indicate median amino acid length. (**F**) Divergence from inferred germline gene sequences. Red bars indicate the median percent values.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

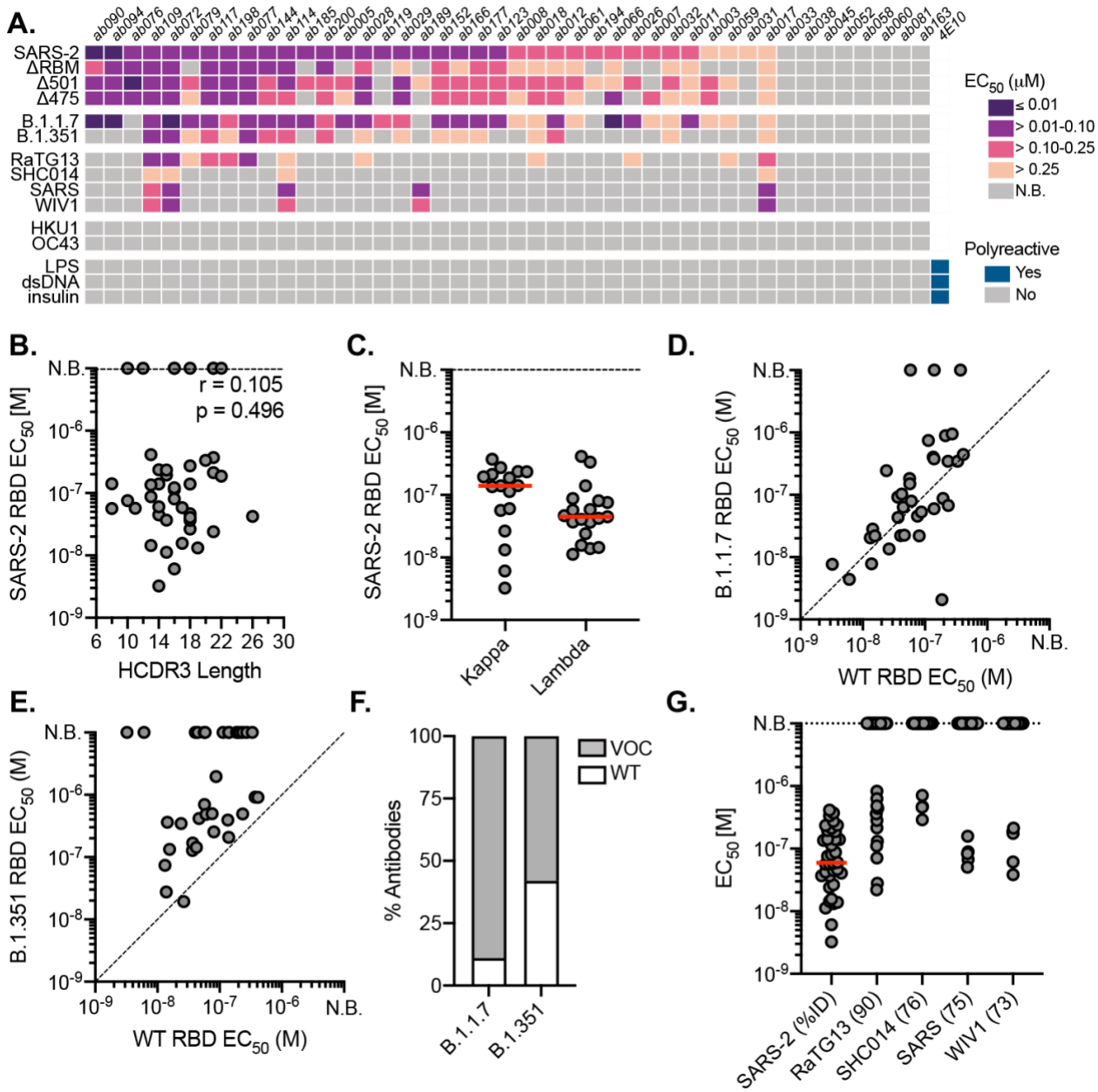


Fig. 2. Binding properties and specificity of isolated naive antibodies. (**A**) ELISA binding heat map of 44 naive IgGs. Binding to wildtype SARS-CoV-2 RBD (SARS-2), ΔRBM, individual RBD glycan variants, circulating variants, related CoVs, hCoVs and polyreactivity antigens. (**B**) Pearson correlation analysis of SARS-CoV-2 RBD affinities and HCDR3 length. (**C**) ELISA EC_{50}s for IgGs with detectable SARS-CoV-2 RBD binding (n = 36) based on kappa or lambda gene usage. Red bars indicate the mean EC_{50} values. (**D**) Wildtype SARS-CoV-2 RBD ELISA EC_{50}s plotted against EC_{50}s for B.1.1.7 RBD (**E**) B.1.351 RBD. (**F**) Proportion of SARS-CoV-2 RBD binders with detectable ELISA affinity for variants of concern (VOC) B.1.1.7 and B.1.351 RBDs. (**G**) ELISA EC_{50} values to related sarbecovirus RBDs displayed in decreasing order of paired-sequence identity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

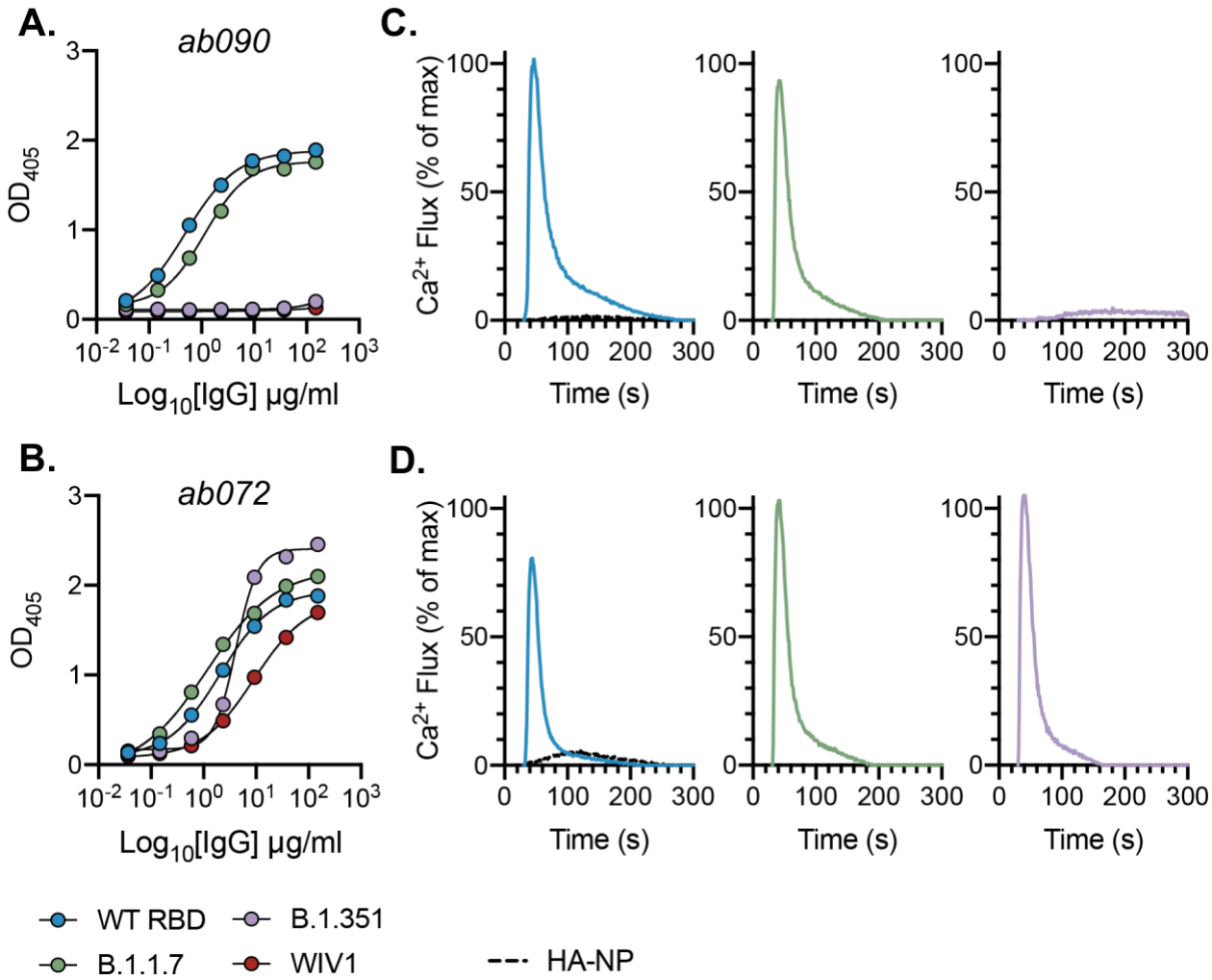
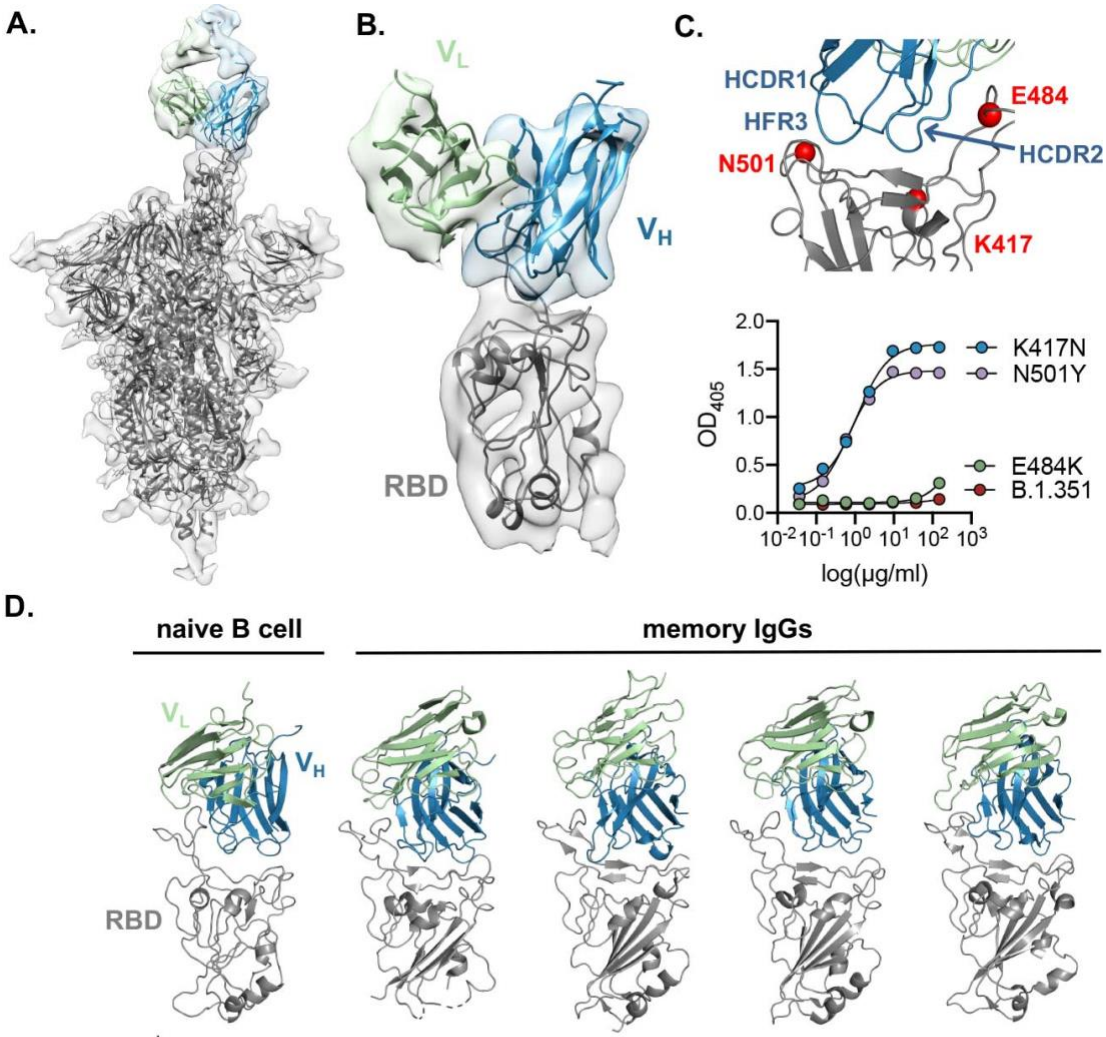


Fig. 3. *In vitro* reconstitution of naive B cell activation. (**A**) ELISA binding reactivity shows restricted specificity of ab090 and (**B**) broad binding of ab072 to wildtype (WT) SARS-CoV-2, B.1.1.7, B.1.351, and WIV1 RBDs. (**C**) BCR activation as measured by calcium flux in a Ramos B cell line expressing ab090 membrane-anchored IgM (mIgM) and (**D**) and ab072 mIgM in response to ferritin nanoparticles (NPs) displaying WT SARS-CoV-2, B.1.1.7, B.1.351, and WIV1 RBDs. Influenza hemagglutinin (HA) NP was used as a negative control.

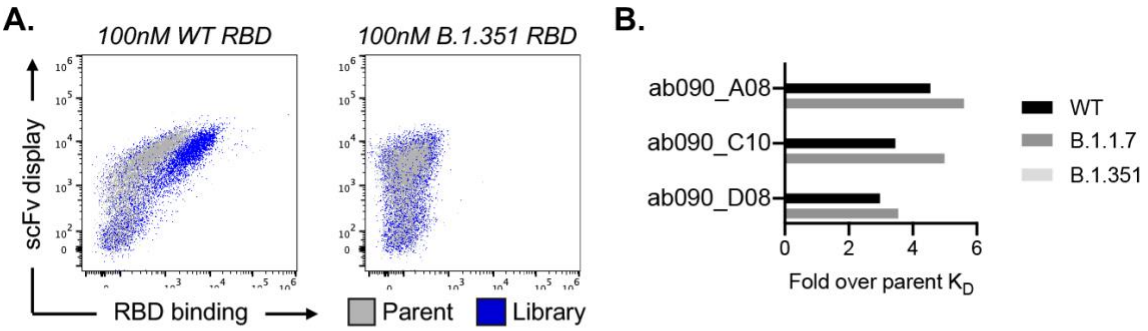
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Ab	ab090	2-15	2-43	2-4	H4
PBD ID	N/A	7L5B	7L56	6XEY	7L58
heavy chain	V_H1-2	V_H1-2	V_H1-2	V_H1-2	V_H1-2
HCDR3 (aa)	14	22	22	15	19
light chain	V_K3-15	V_L2-14	V_L2-14	V_L2-8	V_K2-40
SHM (Δaa)	0	5	2	3	0

Fig. 4. ab090 recognizes the SARS-CoV-2 RBM. (**A**) Cryo-EM structure of the SARS-CoV-2 spike trimer (grey) with ab090 Fab bound to one RBD in the up position. (**B**) ab090 recognizes the SARS-CoV-2 RBM with through a paratope centered on the V_H (blue). (**C**) Close-up view showing the approximate locations of HCDR loops proximal to the RBM epitope and B.1.351 RBD mutations highlighted in red (top). ELISA binding reactivity of ab090 to individual mutations from B.1.351 RBD (bottom). (**D**) ab090 binds to the RBM with a similar mode and angle of approach to IGHV1-2 neutralizing antibodies isolated from memory B cells from convalescent COVID-19 donors. RBDs (grey) are shown in the same relative orientation in each panel with PBD codes and sequence attributes listed in the below.

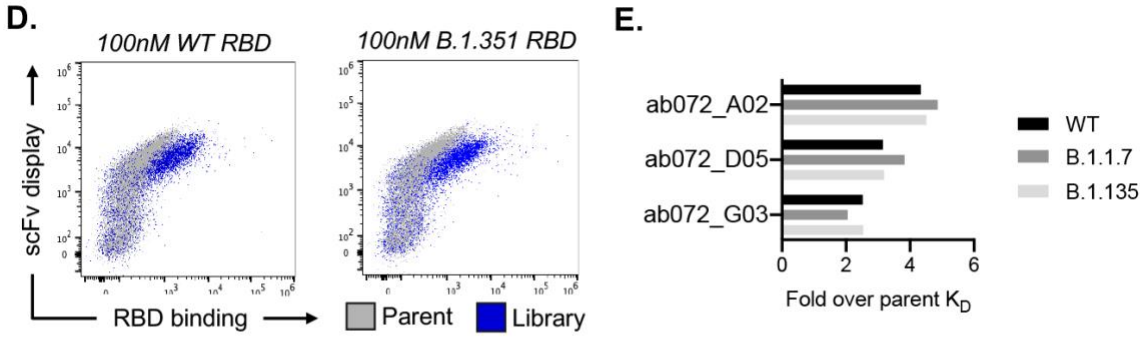
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



C.

ab090
IGHV1-2/IGKV3-15

Clone	Mutations		WT RBD			B.1.1.7			B.1.351		
	V_H	V_L	K_D (μM)	k_a (10^4) ($M^{-1}s^{-1}$)	k_d (10^{-1}) (s^{-1})	K_D (μM)	k_a (10^4) ($M^{-1}s^{-1}$)	K_d (10^{-1}) (s^{-1})	K_D (μM)	k_a (10^4) ($M^{-1}s^{-1}$)	k_d (10^{-1}) (s^{-1})
Parent	---	---	7.7	2.0	1.5	9.2	1.5	1.4	n.b	n.b	n.b
A08	R72H	---	1.7	4.4	0.7	1.7	4.2	0.7	n.b	n.b	n.b
C10	R72C	---	2.2	4.0	0.9	1.9	3.9	0.7	n.b	n.b	n.b
D08	R72S	---	2.6	3.8	1.0	2.6	3.5	0.9	n.b	n.b	n.b



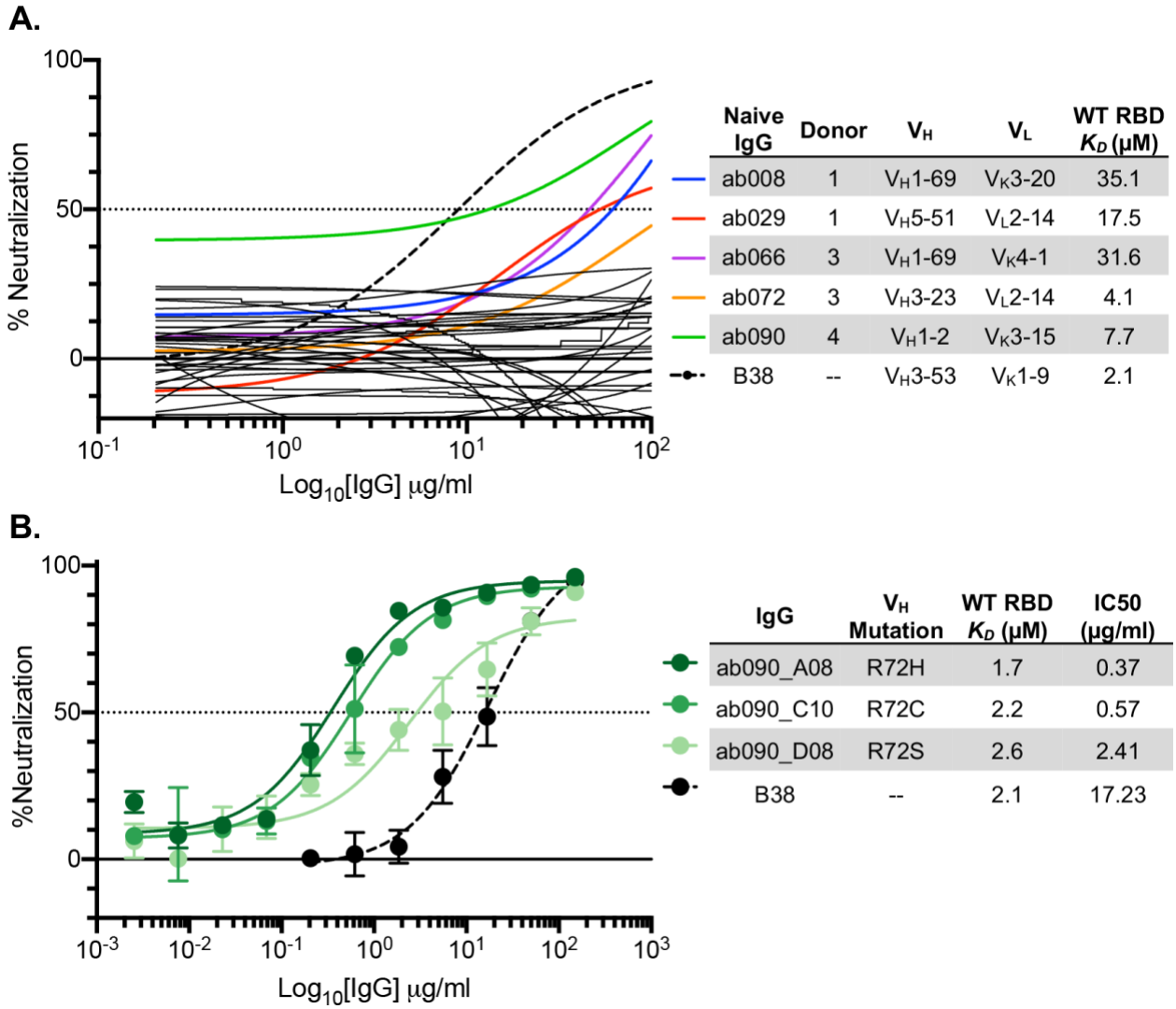
F.

ab072
IGHV3-23/IGLV2-14

Clone	Mutations		WT RBD			B.1.1.7			B.1.351		
	V_H	V_L	K_D (μM)	k_a (10^4) ($M^{-1}s^{-1}$)	k_d (10^{-1}) (s^{-1})	K_D (μM)	k_a (10^4) ($M^{-1}s^{-1}$)	k_d (10^{-1}) (s^{-1})	K_D (μM)	k_a (10^4) ($M^{-1}s^{-1}$)	k_d (10^{-1}) (s^{-1})
Parent	---	---	4.1	2.1	0.9	6.5	1.7	1.1	4.0	2.2	0.9
A02	S31P	---	1.0	3.5	0.3	1.3	3.4	0.5	0.9	3.7	0.3
D05	P99S	N55D	1.3	3.0	0.4	1.7	3.0	0.5	1.3	3.1	0.4
G03	G53D	---	1.6	3.9	0.6	3.2	2.8	0.9	1.6	3.9	0.6

Fig. 5. *in vitro* affinity-matured naive antibodies retain intrinsic specificity. (**A**) Enrichment of ab090 parent (grey) and affinity matured (blue) libraires to 100nM SARS-CoV-2 or B.1.351 RBD using flow cytometry (**B**) Fold enrichment in monovalent K_D over ab090 parent for selected affinity matured progeny. (**C**) Kinetic analysis using biolayer interferometry (BLI) for ab090 parent and progeny Fabs to monomeric WT and variant RBDs. (**D**) Enrichment of ab072 parent (grey) and affinity matured (blue) libraires to 100nM SARS-CoV-2 or B.1.351 RBD using flow cytometry (**E**) Fold enrichment in monovalent K_D over ab072 parent for selected affinity matured progeny. (**F**) Kinetic analysis using biolayer interferometry (BLI) for ab072 parent and progeny Fabs to monomeric WT and variant RBDs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Naive IgG	Donor	V_H	V_L	WT RBD K_D (μM)
ab008	1	V_H1-69	V_K3-20	35.1
ab029	1	V_H5-51	V_L2-14	17.5
ab066	3	V_H1-69	V_K4-1	31.6
ab072	3	V_H3-23	V_L2-14	4.1
ab090	4	V_H1-2	V_K3-15	7.7
B38	--	V_H3-53	V_K1-9	2.1

IgG	V_H Mutation	WT RBD K_D (μM)	IC50 (μg/ml)
ab090_A08	R72H	1.7	0.37
ab090_C10	R72C	2.2	0.57
ab090_D08	R72S	2.6	2.41
B38	--	2.1	17.23

Fig. 6. SARS-CoV-2 pseudovirus neutralization by naive and affinity-matured Abs. (**A**) SARS-CoV-2 pseudovirus neutralization assay for 36 purified IgGs. Curves in color highlighted antibodies with neutralizing activity with donor and monovalent wild-type RBD affinity listed for this subset of antibodies. The neutralizing monoclonal antibody, B38, was used as a positive control. Dashed lines indicate IC_{50} values and data represent means ± SD of two technical replicates. (**B**) SARS-CoV-2 pseudovirus neutralization for select affinity matured progeny from the ab090 lineage with respective mutations relative to ab090 parent sequence, monovalent wild-type RBD affinity, and IC50 listed.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

1. N. E. Harwood, F. D. Batista, Early events in B cell activation. *Annual review of immunology* **28**, 185-210 (2009).
2. G. D. Victora, M. C. Nussenzweig, Germinal centers. *Annual review of immunology* **30**, 429-457 (2012).
3. C. Havenar-Daughton, R. K. Abbott, W. R. Schief, S. Crotty, When designing vaccines, consider the starting material: the human B cell repertoire. *Current opinion in immunology* **53**, 209-216 (2018).
4. F. Sallusto, A. Lanzavecchia, K. Araki, R. Ahmed, From vaccines to memory and back. *Immunity* **33**, 451-463 (2010).
5. W. Deng *et al.*, Primary exposure to SARS-CoV-2 protects against reinfection in rhesus macaques. *Science* **369**, 818-823 (2020).
6. A. Chandrashekar *et al.*, SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science*, (2020).
7. N. B. Mercado *et al.*, Single-shot Ad26 vaccine protects against SARS-CoV-2 in rhesus macaques. *Nature* **586**, 583-588 (2020).
8. Q. Gao *et al.*, Development of an inactivated vaccine candidate for SARS-CoV-2. *Science* **369**, 77-81 (2020).
9. K. S. Corbett *et al.*, Immune Correlates of Protection by mRNA-1273 Immunization against SARS-CoV-2 Infection in Nonhuman Primates. *bioRxiv*, 2021.2004.2020.440647 (2021).
10. W. F. Garcia-Beltran *et al.*, COVID-19 neutralizing antibodies predict disease severity and survival. *Cell*, (2020).
11. D. S. Khoury *et al.*, Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nature Medicine*, (2021).
12. A. Addetia *et al.*, 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**, e02107-02120 (2020).
13. S. F. Lumley *et al.*, Antibodies to SARS-CoV-2 are associated with protection against reinfection. *medRxiv*, 2020.2011.2018.20234369 (2020).
14. J. S. Turner *et al.*, SARS-CoV-2 infection induces long-lived bone marrow plasma cells in humans. *Nature*, 1-8 (2021).
15. A. Sokal *et al.*, Maturation and persistence of the anti-SARS-CoV-2 memory B cell response. *bioRxiv*, 2020.2011.2017.385252 (2020).
16. C. Gaebler *et al.*, Evolution of antibody immunity to SARS-CoV-2. *Nature*, (2021).
17. J. M. Dan *et al.*, Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science*, eabf4063 (2021).
18. L. B. Rodda *et al.*, Functional SARS-CoV-2-specific immune memory persists after mild COVID-19. *Cell*, (2020).
19. M. Sakharkar *et al.*, Prolonged evolution of the human B cell response to SARS-CoV-2 infection. *Science immunology* **6**, (2021).
20. E. J. Anderson *et al.*, Safety and immunogenicity of SARS-CoV-2 mRNA-1273 vaccine in older adults. *New England Journal of Medicine*, (2020).
21. E. E. Walsh *et al.*, Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine Candidates. *New England Journal of Medicine* **383**, 2439-2450 (2020).
22. Z. Wang *et al.*, mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature*, (2021).
23. Z. Wang *et al.*, mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *bioRxiv*, 2021.2001.2015.426911 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24. P. J. M. Brouwer *et al.*, Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643-650 (2020).
25. B. Ju *et al.*, Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **584**, 115-119 (2020).
26. C. Kreer *et al.*, Longitudinal isolation of potent near-germline SARS-CoV-2-neutralizing antibodies from COVID-19 patients. *Cell* **182**, 843-854. e812 (2020).
27. D. F. Robbiani *et al.*, Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437-442 (2020).
28. A. Z. Wec *et al.*, Broad neutralization of SARS-related viruses by human monoclonal antibodies. *Science* **369**, 731-736 (2020).
29. S. J. Zost *et al.*, Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nature Medicine* **26**, 1422-1427 (2020).
30. C. O. Barnes *et al.*, SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature*, (2020).
31. C. O. Barnes *et al.*, Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828-842.e816 (2020).
32. L. Piccoli *et al.*, Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* **183**, 1024-1042.e1021 (2020).
33. R. Shi *et al.*, A human neutralizing antibody targets the receptor binding site of SARS-CoV-2. *Nature*, 1-8 (2020).
34. Y. Wu *et al.*, A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* **368**, 1274-1278 (2020).
35. M. Yuan *et al.*, Structural basis of a shared antibody response to SARS-CoV-2. *Science* **369**, 1119-1123 (2020).
36. S. C. A. Nielsen *et al.*, Human B Cell Clonal Expansion and Convergent Antibody Responses to SARS-CoV-2. *Cell Host & Microbe* **28**, 516-525.e515 (2020).
37. M. Sangesland *et al.*, Germline-encoded affinity for cognate antigen enables vaccine amplification of a human broadly neutralizing response against influenza virus. *Immunity* **51**, 735-749. e738 (2019).
38. P. Dosenovic *et al.*, Anti–HIV-1 B cell responses are dependent on B cell precursor frequency and antigen-binding affinity. *Proceedings of the National Academy of Sciences*, 201803457 (2018).
39. R. K. Abbott *et al.*, Precursor frequency and affinity determine B cell competitive fitness in germinal centers, tested with germline-targeting HIV vaccine immunogens. *Immunity* **48**, 133-146. e136 (2018).
40. A. Amitai *et al.*, Defining and Manipulating B cell Immunodominance Hierarchies to Elicit Broadly Neutralizing Antibody Responses Against Influenza Virus. *Cell Systems* **11**, 573-588. e579 (2020).
41. M. Rapp *et al.*, Modular basis for potent SARS-CoV-2 neutralization by a prevalent VH1-2-derived antibody class. *Cell Reports* **35**, 108950 (2021).
42. C. Havenar-Daughton *et al.*, The human naive B cell repertoire contains distinct subclasses for a germline-targeting HIV-1 vaccine immunogen. *Science translational medicine* **10**, eaat0381 (2018).
43. J. G. Jardine *et al.*, HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. *Science* **351**, 1458-1463 (2016).
44. J. M. Steichen *et al.*, A generalized HIV vaccine design strategy for priming of broadly neutralizing antibody responses. *Science* **366**, (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

45. B. M. Hauser *et al.*, Engineered receptor binding domain immunogens elicit pan-coronavirus neutralizing antibodies. *bioRxiv*, 2020.2012.2007.415216 (2020).
46. B. Briney, A. Inderbitzin, C. Joyce, D. R. Burton, Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* **566**, 393-397 (2019).
47. B. Ju *et al.*, Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **584**, 115-119 (2020).
48. I. C. MacLennan *et al.*, Extrafollicular antibody responses. *Immunological reviews* **194**, 8-18 (2003).
49. Y. Cao *et al.*, Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell* **182**, 73-84.e16 (2020).
50. C. Soto *et al.*, High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* **566**, 398-402 (2019).
51. A. G. Schmidt *et al.*, Viral receptor-binding site antibodies with diverse germline origins. *Cell* **161**, 1026-1034 (2015).
52. T. Zhou *et al.*, Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. *Immunity* **39**, 245-258 (2013).
53. A. P. West, R. Diskin, M. C. Nussenzweig, P. J. Bjorkman, Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proceedings of the National Academy of Sciences* **109**, E2083-E2090 (2012).
54. D. Lingwood *et al.*, Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature* **489**, 566 (2012).
55. A. G. Schmidt *et al.*, Immunogenic stimulus for germline precursors of antibodies that engage the influenza hemagglutinin receptor-binding site. *Cell reports* **13**, 2842-2850 (2015).
56. F. D. Batista, M. S. Neuberger, Affinity dependence of the B cell response to antigen: a threshold, a ceiling, and the importance of off-rate. *Immunity* **8**, 751-759 (1998).
57. J. M. Tas *et al.*, Visualizing antibody affinity maturation in germinal centers. *Science* **351**, 1048-1054 (2016).
58. S. Cele *et al.*, Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature* **593**, 142-146 (2021).
59. R. E. Chen *et al.*, Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. *Nature Medicine* **27**, 717-726 (2021).
60. P. Wang *et al.*, Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130-135 (2021).
61. C. K. Wibmer *et al.*, SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nature Medicine* **27**, 622-625 (2021).
62. N. G. Davies *et al.*, Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021).
63. H. Tegally *et al.*, Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*, 2020.2012.2021.20248640 (2020).
64. E. B. Hodcroft, CoVariants: SARS-CoV-2 Mutations and Variants of Interest. *https://covariants.org/*, (2021).
65. W. F. Garcia-Beltran *et al.*, Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372-2383. e2379 (2021).
66. M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology* **5**, 562-569 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

67. V. D. Menachery *et al.*, SARS-like WIV1-CoV poised for human emergence. *Proceedings of the National Academy of Sciences* **113**, 3048-3053 (2016).
68. A. G. Wrobel *et al.*, SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nature Structural & Molecular Biology* **27**, 763-767 (2020).
69. V. D. Menachery *et al.*, A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature Medicine* **21**, 1508-1513 (2015).
70. A. A. Cohen *et al.*, Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science* **371**, 735-741 (2021).
71. E. Goodwin *et al.*, Infants Infected with Respiratory Syncytial Virus Generate Potent Neutralizing Antibodies that Lack Somatic Hypermutation. *Immunity* **48**, 339-349.e335 (2018).
72. H.-X. Liao *et al.*, Initial antibodies binding to HIV-1 gp41 in acutely infected subjects are polyreactive and highly mutated. *Journal of Experimental Medicine* **208**, 2237-2249 (2011).
73. Z.-H. Zhou, A. G. Tzioufas, A. L. Notkins, Properties and function of polyreactive antibodies and polyreactive antigen-binding B cells. *Journal of Autoimmunity* **29**, 219-228 (2007).
74. L. Shehata *et al.*, Affinity Maturation Enhances Antibody Specificity but Compromises Conformational Stability. *Cell Reports* **28**, 3300-3308.e3304 (2019).
75. N. E. Harwood, F. D. Batista, Early Events in B Cell Activation. *Annual Review of Immunology* **28**, 185-210 (2010).
76. G. C. Weaver *et al.*, In vitro reconstitution of B cell receptor–antigen interactions to evaluate potential vaccine candidates. *Nature protocols* **11**, 193 (2016).
77. B. Zakeri *et al.*, Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proceedings of the National Academy of Sciences* **109**, E690-E697 (2012).
78. B. Zhang *et al.*, A platform incorporating trimeric antigens into self-assembling nanoparticles reveals SARS-CoV-2-spike nanoparticles to elicit substantially higher neutralizing responses than spike alone. *Scientific Reports* **10**, 18149 (2020).
79. M. Yuan, H. Liu, N. C. Wu, I. A. Wilson, Recognition of the SARS-CoV-2 receptor binding domain by neutralizing antibodies. *Biochemical and Biophysical Research Communications* **538**, 192-203 (2021).
80. V. Giudicelli *et al.*, IMGT/LIGM-DB, the IMGT® comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Research* **34**, D781-D784 (2006).
81. M. Yuan *et al.*, Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science*, eabh1139 (2021).
82. D. K. Sethi, A. Agarwal, V. Manivel, K. V. Rao, D. M. Salunke, Differential epitope positioning within the germline antibody paratope enhances promiscuity in the primary immune response. *Immunity* **24**, 429-438 (2006).
83. L. C. James, P. Roversi, D. S. Tawfik, Antibody multispecificity mediated by conformational diversity. *Science* **299**, 1362-1367 (2003).
84. Z. Guo, E. Kraka, D. Cremer, Description of local and global shape properties of protein helices. *J Mol Model* **19**, 2901-2911 (2013).
85. G. Chao *et al.*, Isolating and engineering human antibodies using yeast surface display. *Nature Protocols* **1**, 755-768 (2006).
86. D. Stadlbauer *et al.*, Repeated cross-sectional sero-monitoring of SARS-CoV-2 in New York City. *Nature*, 1-5 (2020).
87. K. Vanshylla *et al.*, Kinetics and correlates of the neutralizing antibody response to SARS-CoV-2 infection in humans. *Cell Host & Microbe* **29**, 917-929.e914 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

88. A. J. Greaney *et al.*, Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host & Microbe*, (2020).
89. E. M. Anderson *et al.*, Seasonal human coronavirus antibodies are boosted upon SARS-CoV-2 infection but not associated with protection. *Cell* **184**, 1858-1864.e1810 (2021).
90. K. W. Ng *et al.*, Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science* **370**, 1339-1343 (2020).
91. E. Shrock *et al.*, Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* **370**, eabd4250 (2020).
92. P. Nguyen-Contant *et al.*, S Protein-Reactive IgG and Memory B Cell Production after Human SARS-CoV-2 Infection Includes Broad Reactivity to the S2 Subunit. *mBio* **11**, e01991-01920 (2020).
93. C. Viant *et al.*, Antibody Affinity Shapes the Choice between Memory and Germinal Center B Cell Fates. *Cell* **183**, 1298-1311. e1211 (2020).
94. R. Di Niro *et al.*, Salmonella infection drives promiscuous B cell activation followed by extrafollicular affinity maturation. *Immunity* **43**, 120-131 (2015).
95. M. Kuraoka *et al.*, Complex antigens drive permissive clonal selection in germinal centers. *Immunity* **44**, 542-552 (2016).
96. S. I. Kim *et al.*, Stereotypic neutralizing VH antibodies against SARS-CoV-2 spike protein receptor binding domain in COVID-19 patients and healthy individuals. *Science Translational Medicine*, (2021).
97. Z. Wang *et al.*, Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature*, (2021).
98. P. W. H. I. Parren, D. R. Burton, in *Advances in Immunology*. (Academic Press, 2001), vol. 77, pp. 195-262.
99. C. G. Rappazzo *et al.*, Broad and potent activity against SARS-like viruses by an engineered human monoclonal antibody. *Science*, eabf4830 (2021).
100. L. R. Baden *et al.*, Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med* **384**, 403-416 (2021).
101. D. M. Skowronski, G. De Serres, Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* **384**, 1576-1577 (2021).
102. D. Wrapp *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).
103. M. Sangesland *et al.*, A Single Human VH-gene Allows for a Broad-Spectrum Antibody Response Targeting Bacterial Lipopolysaccharides in the Blood. *Cell Reports* **32**, 108065 (2020).
104. P. S. Tobias, K. Soldau, R. J. Ulevitch, Identification of a Lipid A Binding Site in the Acute Phase Reactant Lipopolysaccharide Binding Protein. *Journal of Biological Chemistry* **264**, 10867-10871 (1989).
105. J. J. Trombetta *et al.*, Preparation of single-cell RNA-seq libraries for next generation sequencing. *Current protocols in molecular biology* **107**, 4.22. 21-24.22. 17 (2014).
106. A. P. Masella, A. K. Bartram, J. M. Truszkowski, D. G. Brown, J. D. Neufeld, PANDAseq: paired-end assembler for illumina sequences. *BMC bioinformatics* **13**, 1-7 (2012).
107. B. Shi *et al.*, Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST. *Theoretical Biology and Medical Modelling* **11**, 1-11 (2014).
108. M. J. Moore *et al.*, Retroviruses pseudotyped with the severe acute respiratory syndrome coronavirus spike protein efficiently infect cells expressing angiotensin-converting enzyme 2. *Journal of virology* **78**, 10628-10635 (2004).
109. E. Siebring-van Olst *et al.*, Affordable luciferase reporter assay for cell-based high-throughput screening. *Journal of biomolecular screening* **18**, 453-461 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

110. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012).
111. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).
112. K. Zhang, Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12 (2016).
113. T. Wagner *et al.*, SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* **2**, 218 (2019).
114. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
115. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
116. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
117. J. Mata-Fink *et al.*, Rapid Conformational Epitope Mapping of Anti-gp120 Antibodies with a Designed Mutant Panel Displayed on Yeast. *Journal of Molecular Biology* **425**, 444-456 (2013).
118. M. I. J. Raybould, A. Kovaltsuk, C. Marks, C. M. Deane, CoV-AbDab: the coronavirus antibody database. *Bioinformatics* **37**, 734-735 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplemental Information

Naive human B cells engage the receptor binding domain of SARS-CoV-2, variants of concern, and related sarbecoviruses

Jared Feldman[1†], Julia Bals[1†], Clara G. Altomare[2], Kerri St. Denis[1], Evan C. Lam[1], Blake M. Hauser[1], Larance Ronsard[1], Maya Sangesland[1], Thalia Bracamonte Moreno[1], Vintus Okonkwo[1], Nathania Hartojo[1] Alejandro B. Balazs[1], Goran Bajic[2], Daniel Lingwood[1*] and Aaron G. Schmidt[1,3*]

[1]Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, 02139, USA

[2] Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029

[3]Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA

*Correspondence to: Email: dlingwood@mgh.harvard.edu (D.L.); aschmidt@crystal.harvard.edu (A.G.S)

†These authors contributed equally to this work.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

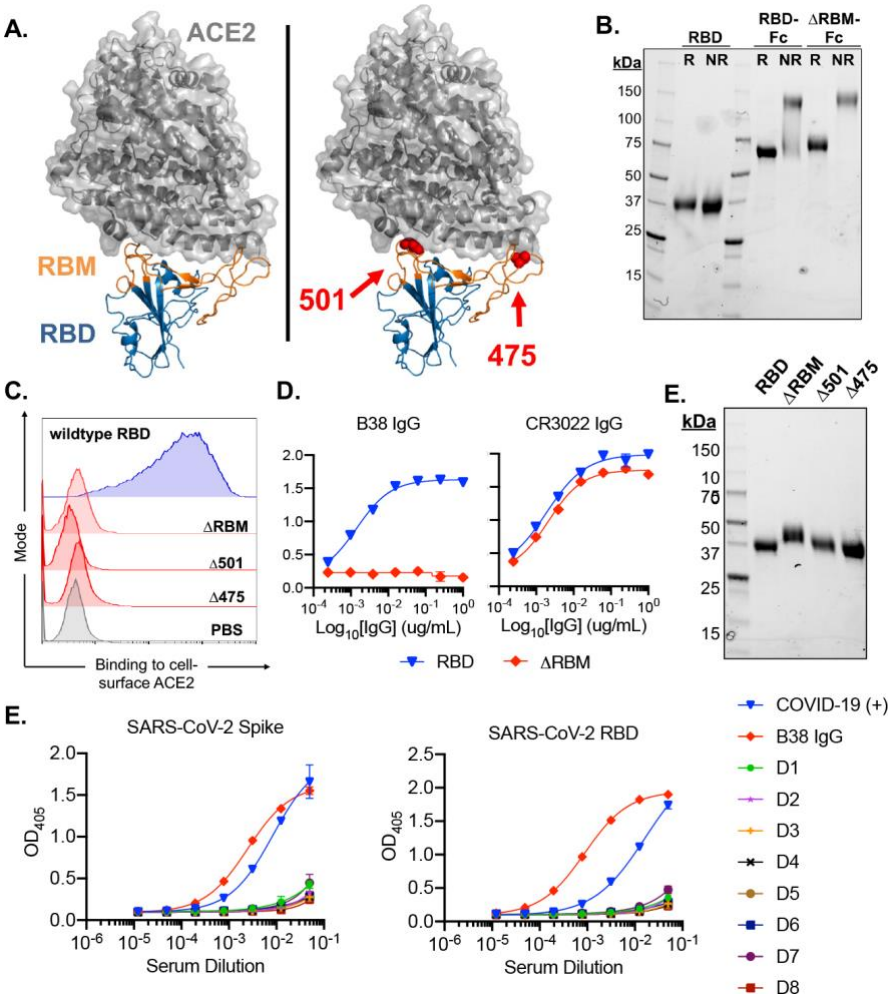


fig. S1. Design and characterization of SARS-CoV-2 antigens and healthy donor sera binding. (A) SARS-CoV-2 RBD in complex with viral receptor, ACE2 shown in blue and grey, respectively (PDB 6M0J). Wild-type RBD with, the receptor binding motif (RBM), shown in orange (left panel). Structural model of the ΔRBM probe designed to abrogate binding to ACE2 (right panel). Putative N-linked glycosylation sites engineered onto the RBM are shown in red spheres at amino acid positions 501 and 475. (**B**) SDS-PAGE gel under reducing (R) and non-reducing (NR) conditions for monomeric RBD, RBD-Fc and ΔRBM-Fc. (**C**) Wildtype RBD, ΔRBM and single glycan variant binding to ACE2-expressing 293T cells by flow cytometry. Wild-type RBD binding shown in blue, glycan variant binding shown in red. Streptavidin-PE was used to detect the relative intensity of antigen binding to cell-surface ACE2. A PBS control (gray) indicates secondary-only staining. (**D**) Control antibody ELISA binding to RBD and ΔRBM antigens. RBM-specific antibody, B38 (left). Non-RBM-specific control antibody, CR3022 (right). (**E**) ΔRBM and Δ501 and Δ475 variants analyzed by SDS-PAGE gel under reducing conditions; wildtype RBD is shown for comparison. (E) SARS-CoV-2 spike (left) and RBD (right) sera ELISA from human subjects 1-8. Sera from a COVID-19 convalescent patient and control antibody, B38, were included as positive controls.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

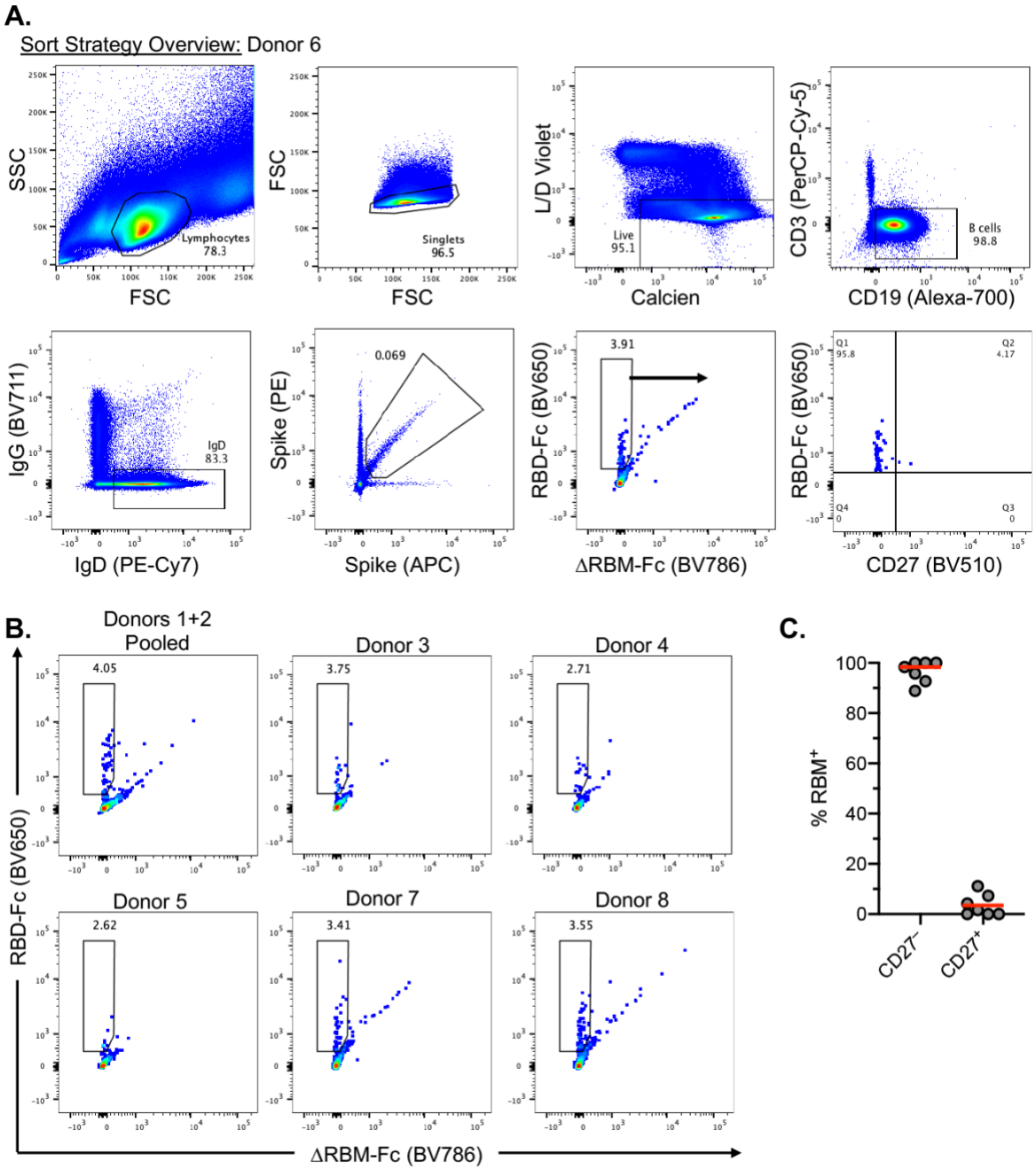


fig. S2. **PBMC flow cytometry analyses.** (**A**) Representative gating strategy used for FACS of PBMCs pooled from donors 1 and 2. Gating was on naive B cells defined by single living lymphocytes that were $CD19^{+}CD3^{-}IgD^{+}IgG^{-}$. Sorted cells were RBM-specific as defined by spike-PE^{+}/spike-APC^{+}/RBD-Fc-$BV650^{+}$/ΔRBM-Fc-BC650-. Sort gate is denoted by the blue arrow. The bottom right plot shows CD27 staining of sorted RBM-specific naive B cells. (**B**) Flow cytometry showing the sort gate and percentage of RBM-specific B cells for the remaining 6 healthy human donors. (**C**) RBM-specific B cell frequency among $CD27^{+}$ and $CD27^{-}$ cells. Each symbol represents a different donor (n = 8).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

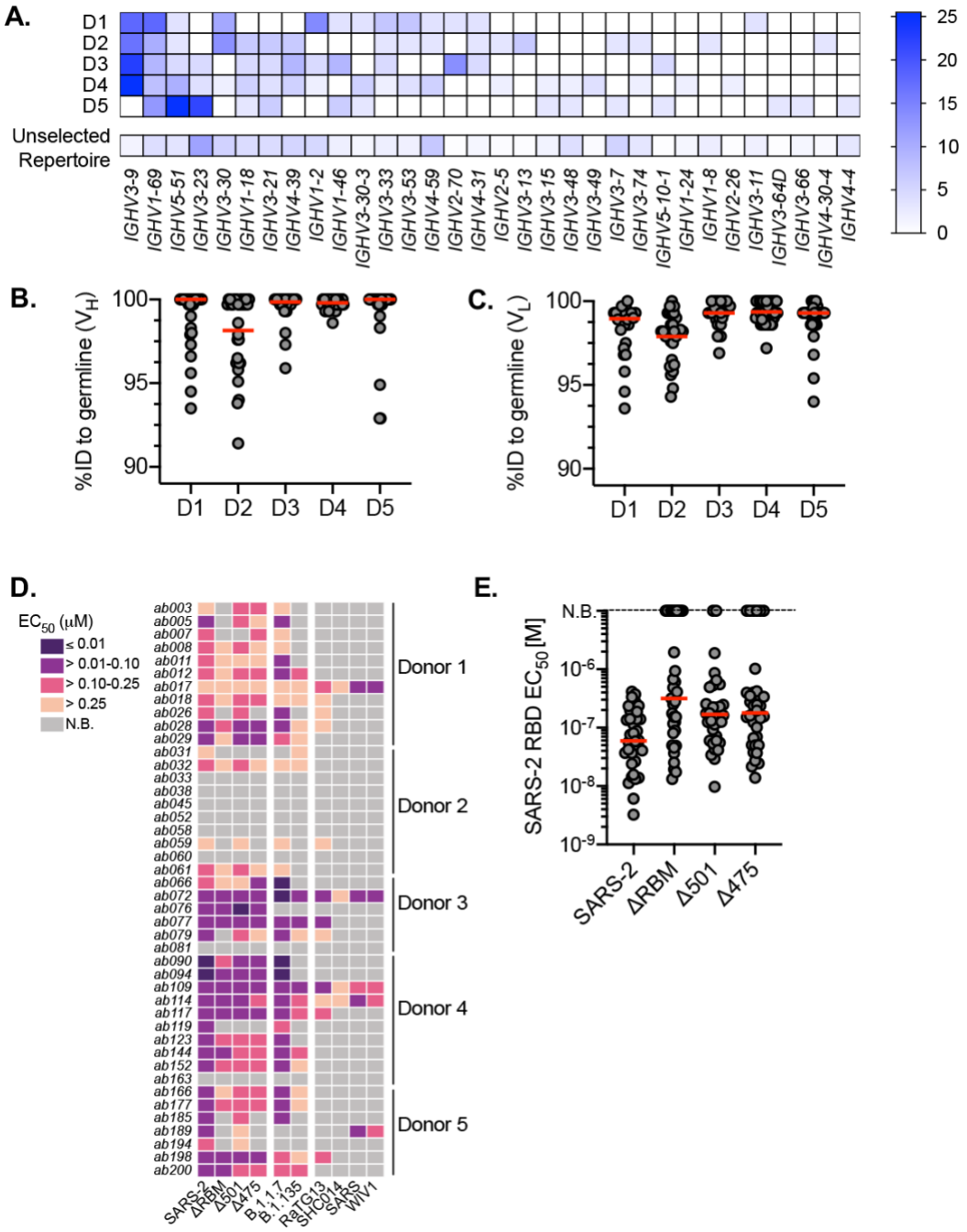
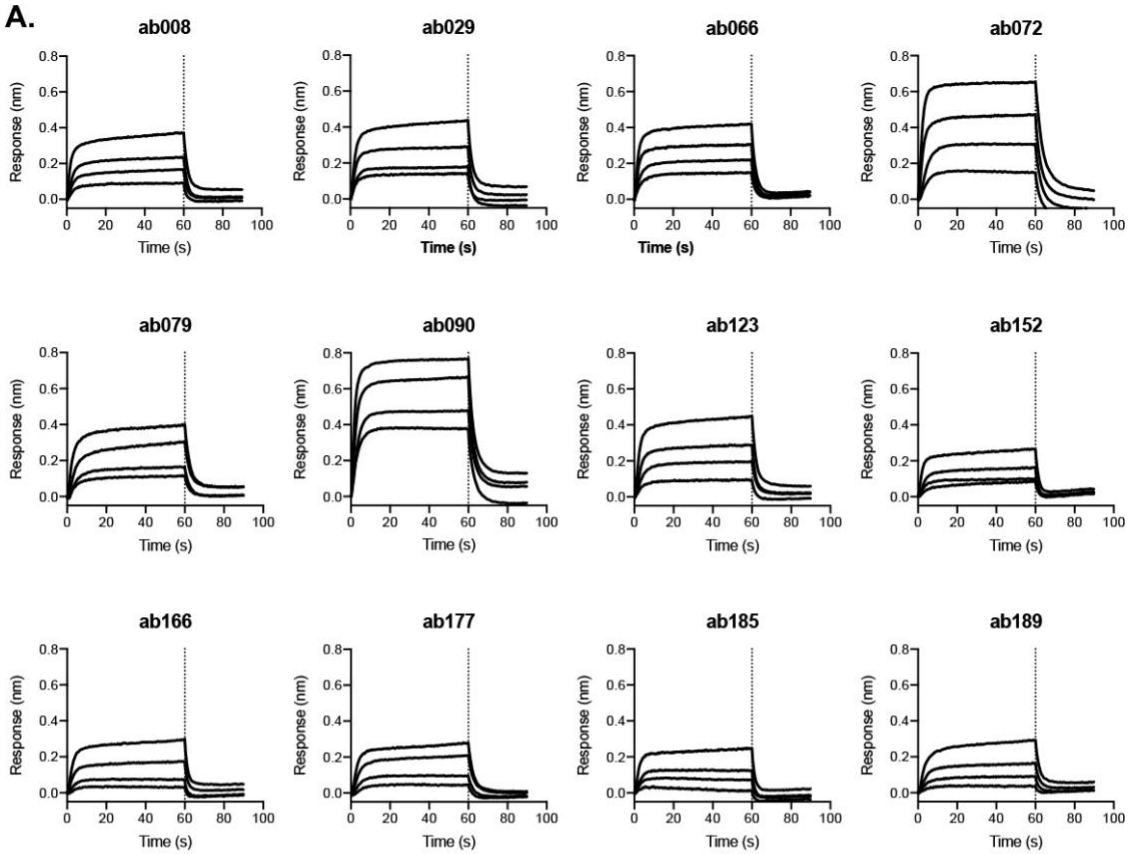


fig. S3. **Repertoire comparison, germline identity, and IgG binding by individual donor**. (**A**) Heatmap showing V_H-gene usage of isolated antibodies derived from donors 1-5. Unselected repertoire gene usage derive from a high-throughput sequencing data set of circulating B cells across 10 human subjects (*46*). Heatmap scale represents percent of total paired sequence from each donor. Divergence from inferred germline gene sequences separated by individual donor for (**B**) V_H and (**C**) V_L. Red bars indicate the median percent values, and each dot represents an individual paired sequence. (**D**) Heatmap showing IgG binding to RBDs (n = 44) sorted by donor. (**E**) ELISA EC_{50} values for IgGs with detectable SARS-CoV-2 RBD binding (n = 36) against RBM glycan probes. Red bars indicate the mean EC_{50} values.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



B.

Naive IgG	IGHV	IGLV	K_D (μM)	K_a (10^3) ($M^{-1}s^{-1}$)	K_{off} (10^{-1}) (s^{-1})
B38	*IGHV3-53*	*IGKV1-9*	*2.1*	*10.21*	*0.21*
ab072	IGHV	IGLV	4.1	20.90	0.87
ab090	IGHV1-2	IGKV3-15	7.7	19.50	1.50
ab029	IGHV5-51	IGLV2-14	17.5	13.91	2.44
ab066	IGHV1-69	IGKV4-1	31.6	9.23	2.92
ab008	IGHV1-69	IGKV3-20	35.1	7.21	2.53
ab123	IGHV3-9	IGLV3-21	40.2	5.86	2.36
ab079	IGHV4-31	IGLV3-21	42.6	4.30	1.83
ab152	IGHV5-51	IGKV2-30	45.1	6.39	2.88
ab166	IGHV1-69	IGLV1-40	55	4.52	2.49
ab189	IGHV3-30-3	IGLV3-21	73.7	2.37	1.75
ab177	IGHV3-21	IGLV2-14	>100	0.06	2.39
ab185	IGHV3-23	IGLV2-23	>100	0.13	3.79

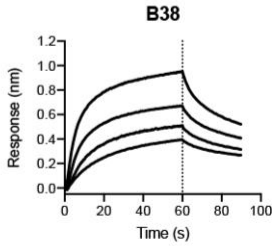
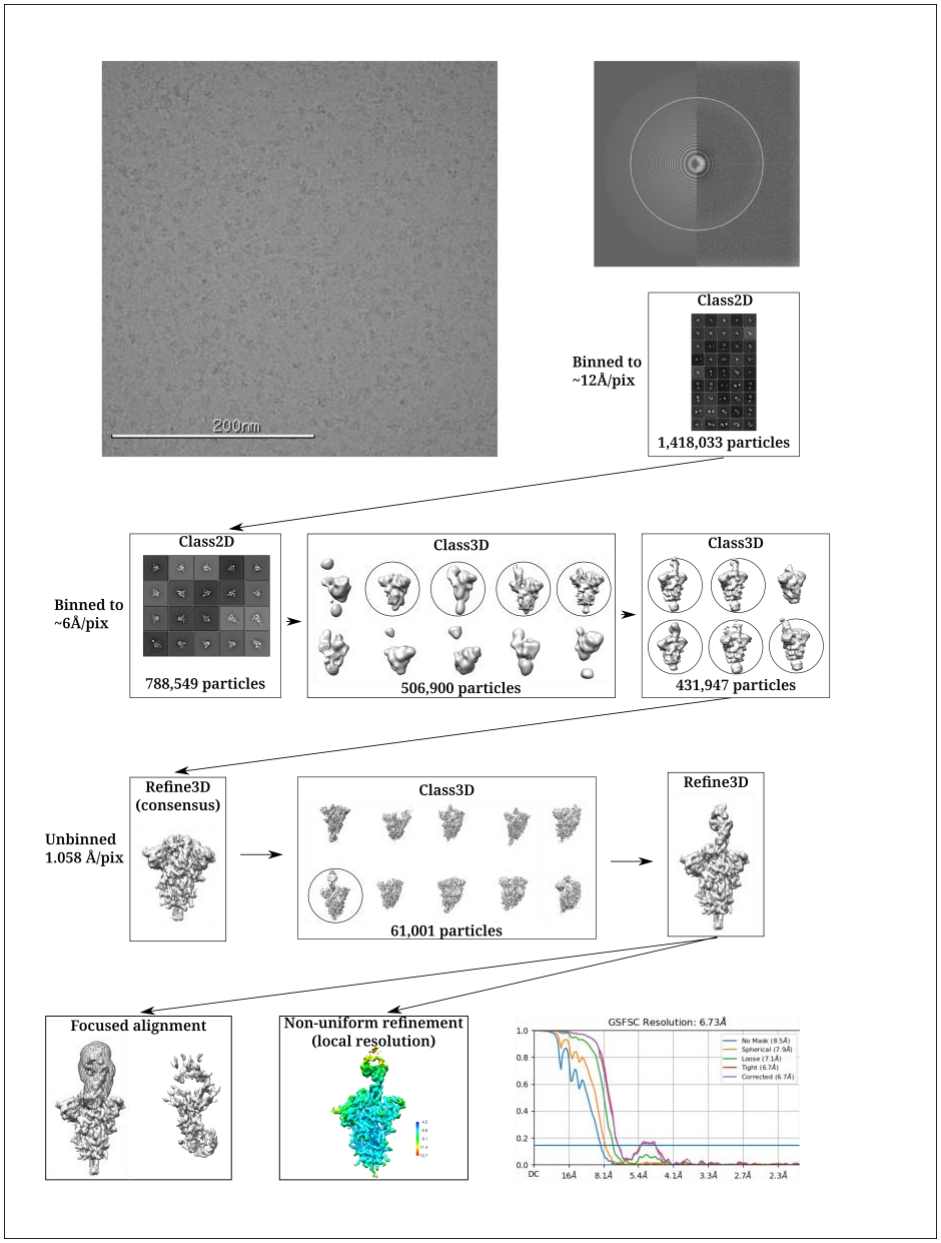


fig. S4. **SARS-CoV-2 RBD-binding kinetics of isolated naive antibodies.** (**A**) Biolayer interferometry (BLI) binding kinetic analysis of titrated SARS-CoV-2 RBD to immobilized Fabs. Dotted line at 60 s denotes the start of the dissociation phase. (**B**) Kinetic and equilibrium constants for binding to RBD calculated from a 1:1 binding model using a global fit to all curves for each Fab using vendor supplied software. B38 Fab is used as a positive control.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A.



bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B.

```
                    10        20        30        40        50        60        70        80        90        100       110       120       130
IGHV1-2*06  QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMHWVRQAPGQGLEWMGRINPNSGGTNYAQKFQGRVTMTRDTSISTAYMELSRLRSDDTAVYYC-------------------------------------
ab090       ........................................................................................................ARHYGDYGEDAFDI---------------WGQGTMVTVSS
2-4         ........................................W.........T.M.....................V.............ARDRSWAVVYYYMDV---------------WGKGTTVTVSS
2-15        ................R.......................W...I.D.........W...........V................ARGGSRCSGGNCYGWAYD-------AFDIWGQGTMITVSS
2-43        ........................................W.......................T......R...........ARGLGVGCSGGNCYLDYY-------YMDVWGKGTTVTVSS
C004        ........................................W...I...............................ASPASRGYSGYDHGYYYY--------MDVWGKGTTVTVSS
C009        ......................M.................W...................................ARDSPFSALGASNDY---------------WGQGTLVTVSS
C121        ........................................W.S.V...............................ARAPLFPTGVLAGDYYYY-------GMDVWGQGTTVTVSS
C127        ........................................W...................................ATAHPRRIQGVFFLGPGV------------WGQGTTVTVSS
C212        ........................V....I..............W.S..............W......M..T...................ARERYFDLGGMDV----------------WGQGTTVTVSS
CC12.4      ........................................W.S...............W........V...........F........ATESWVYGSGSYSSGAFD-----I------WGQGTMVTVSS
CC12.5      E........................IYS...................W.S.D........R...........TT..............M...ARGPRYSGTYFDY----------------WGQGTLVTVSS
CC12.6      ........................I.S...................W.S.D.....................T.G.....G..........ARGPRYSGTYFDY----------------WGQGTLVTVSS
CC12.7      E..................T..T...I.S.........V........W.S.D..A.................T.S.V..TW.K.........ARGPRYSGTYFDF----------------WGQGTLVTVSS
CC12.8      ........................I.S...T...............W.S.D.....................T...V...G..........ARGPRYSGTYFDY----------------WGQGTLVTVSS
CC12.9      E........................I.S...................W.S.D........N...............G............M...ARGPRYSGTYFDY----------------WGQGVLVTVSS
CC12.10     ........................IYS..F................W.S.D..A....T...........TT..H...............ARGPRYSGTHFDY----------------WGQGTLVTVSS
CC12.11     ........................I.S...T...............W.S.D.....................T...V...G..........ARGPRYSGTYFDY----------------WGQGTLVTVSS
CC12.12     ........................IYS...................W.S.D........R...........TT..................ARGPRYSGTYFDY----------------WGQGTLVTVSS
CC12.27     ........................I....................W...........................V................AREMPAAMGYYYYGMDV------------WGQGTTVTVSS
CnC2t1p1_E8 ...........N.............I..N..I............V.W.HSL....S..........L...A..R........G.....L...ARASVSTITDFDY----------------WGQGTLVAVSS
CnC2t1p1_G6 ........................I..N..I............V.W.HSL....S..........L...AP.R.......G.G.....ARASVATITDFDY----------------WGQGTLVAVSS
COV2-2050   ........................D.................W.....R.......................V.......T........ARVVVLGYGRPNNYYDGR------NVWDYWGQGTLVTVSS
COV2-2504   ........................D.................W.....R.......................V.......T........ARVVVLGYGRPNNYYDGR------NVWDYWGQGTLVTVSS
COV2-2632   ........................D.................W.....R.......................V.......T........ARVVVLGYGRPNNYYDGR------NVWDYWGQGTLVTVSS
COVA1-12    E.......T...................P...I...............W......VAKS.............S....V.LDVT---.........ARDLVWATVSGTMDV----------------WGQGTTVTVSS
COVA2-31    E....E..V..............F......QD...............W...S.A.................................S....RKMLTIFGKVNQTMLLISGAKGQWSPSLQWGQGTMVTVSS
COVA2-45    E...LE.............I.........................W...................................ARGDGDYYDSSGYYRPTL-----YNWLDPWGQGTRVTVSS
COVOX-316   ........................................W..........T.............................S.ARDMAFSMVRGSFDY----------------WGQGTLVTVSS
CQTS013     ............................I...............W.D.D....DFV.......................K.........ARGTPPDSVVVPAAIHPY------DWFDPWGQGTLVTVSS
CV05-163    ........................................W...................................AREVMVRGALPPYGMDV-------------WGQGTTVTVSS
CV07-200    ........T.M.............................W.......................R.V................ARGPFYYDNSGTLGGLDV------------WGQGTTVTVSS
CV07-222    ........................D..V...............W...I...............................ARGPYYYDSSGSLGAFDI------------WGQGTMVTVSS
CV07-255    ..............E.........................W.........................V................ARDSRFSYVNGEFDY----------------WGQGILVTVSS
CV07-262    ........................................W.............T.......T.T................ARVGWYDFGTPGDYYYYY-------GMDVWGQGTTVTVSS
CV07-283    ................A...............I.......L...I...........D.......................VRGPFYYDSSGPLGGMDV------------WGQGTTVTVSS
CV17        ........................................W...................................ARVDYGSGSYGWGWFDP-------------WGQGTLVTVSS
CV19        ........................................W...................................AREYYYDSSVYPYYYYAM---------DVWGQGTTVTVSS
CV32        ........................................W.....DV...........................AREARDYYGSGSLDY----------------WGQGTLVTVSS
CV36        ........................................W..........V.....................N..........F.ARDLTTTAGTDYYYGMDV------------WGQGTTVTVSS
DH1151      ........................D...................S......V.................S................ARDLEDFWSGYPPLGYAL---------DVWGQGTTVTVSS
DH1173      ........................................W......A.........W.........N......R..K......F..ARETSFAIFGGGGMDV-------------WGQGTTVTVSS
DH1194      ................M.......D.....................S...................................ARDSSSWRYNWFDP----------------WGQGTLVTVSS
H4          ........................................................................................ARVPYCSSTSCHRDWYF----------DLWGRGTLVTVSS
REGN10923   ........................IN..I...............W...................................AIITIFGVVTWFDP----------------WGQGTLVTVSS
REGN10989   ........................I.....................W.......A..........L......T.V........F........ARGSRYDWNQNNWF-------------DPWGQGTLVTVSS
S2M11       E.......................................W...I.S..S...T........S....T..................ARAAPFYDFWSGYSYFDY------------WGQGTLVTVSS
```

fig. S5. Structural characterization and analysis. (A) Cryo-EM data processing scheme of ab090 Fab bound with SARS-CoV-2 spike. See the Methods section for more details. (**B**) Heavy chain amino acid sequence alignment of ab090 with IGHV1-2 derived antibodies from convalescent COVID-19 patients. Sequences were obtained from CoV-AbDab (*118*) and aligned to the IGHV1-2*06 reference. Residues forming the germline-encoded HCDR1 and HCDR2 motif contacting the SARS-CoV-2 RBD are highlighted in blue. The single nucleotide polymorphism in the *06 allele at position 50 is highlighted red. The site of the dominant mutation from *in vitro* affinity maturation efforts with ab090 is highlighted in green.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.02.429458; this version posted July 9, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

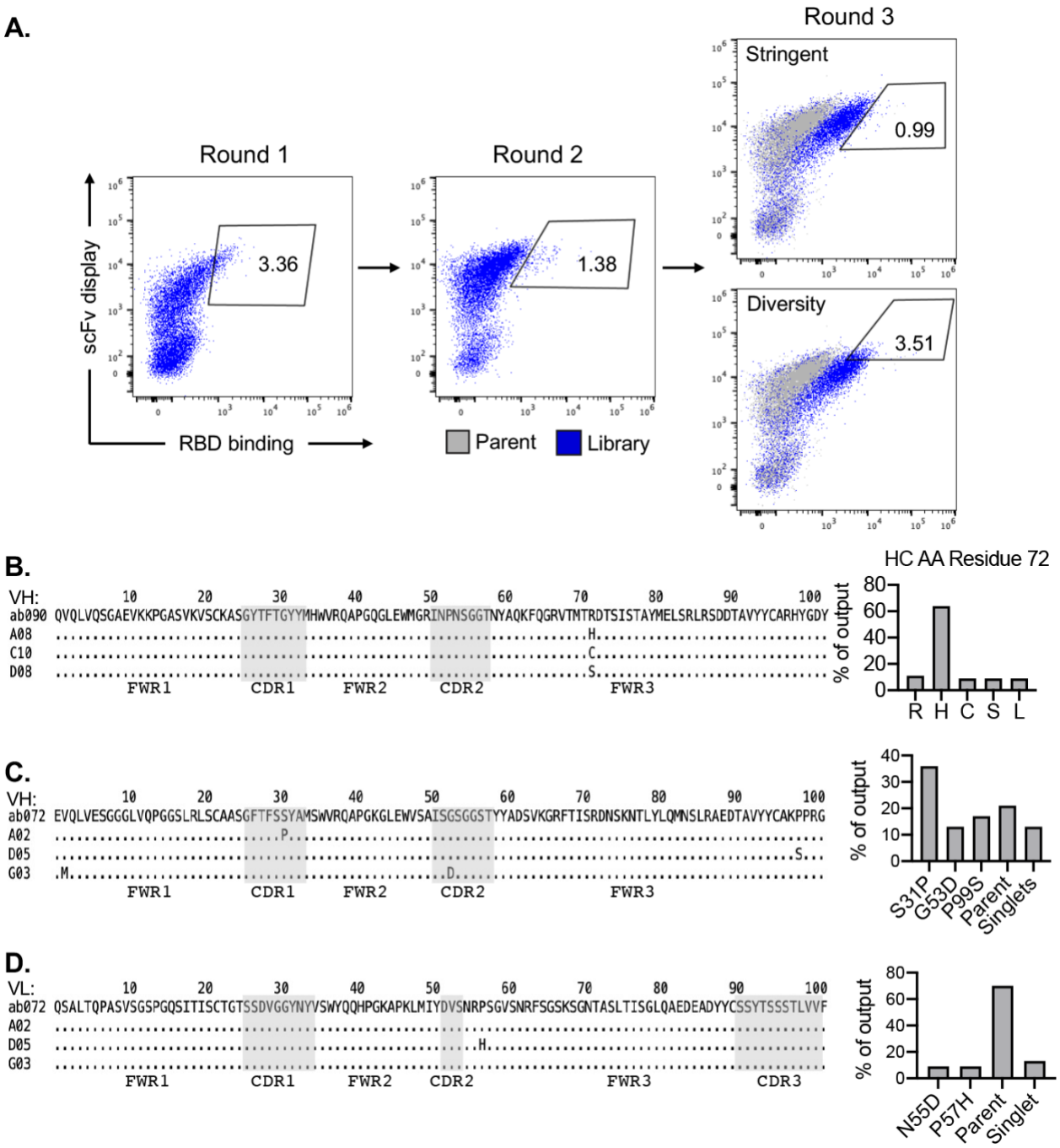


fig. S6. Representative affinity maturation selection strategy and output sequence overview. (**A**) Flow cytometric sorting of diversified single chain variable fragment (scFv) libraries of ab090. Gates represent the yeast population sorted for subsequent selections. After 2 rounds of enrichment for wildtype SARS-CoV-2 binding, a "stringent" and "diversity gate were sorted in round 3 indicating the yeast populations sorted for individual colony isolation and sequencing. Alignment of the V_H sequencing output clones for ab090 (**B**) and ab072 (**C**) with the output frequency of each mutation from a total of 48 single colonies. (**D**) Alignment of the V_L sequencing output clones ab072 with the output frequency of each mutation from a total of 48 single colonies. The V_L output for ab090 was exclusively parent.