Title

Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity

Journal

Cell, 184(5)

ISSN

0092-8674

Authors

Thomson, Emma C

Rosen, Laura E

Shepherd, James G

et al.

Publication Date

2021-03-01

DOI

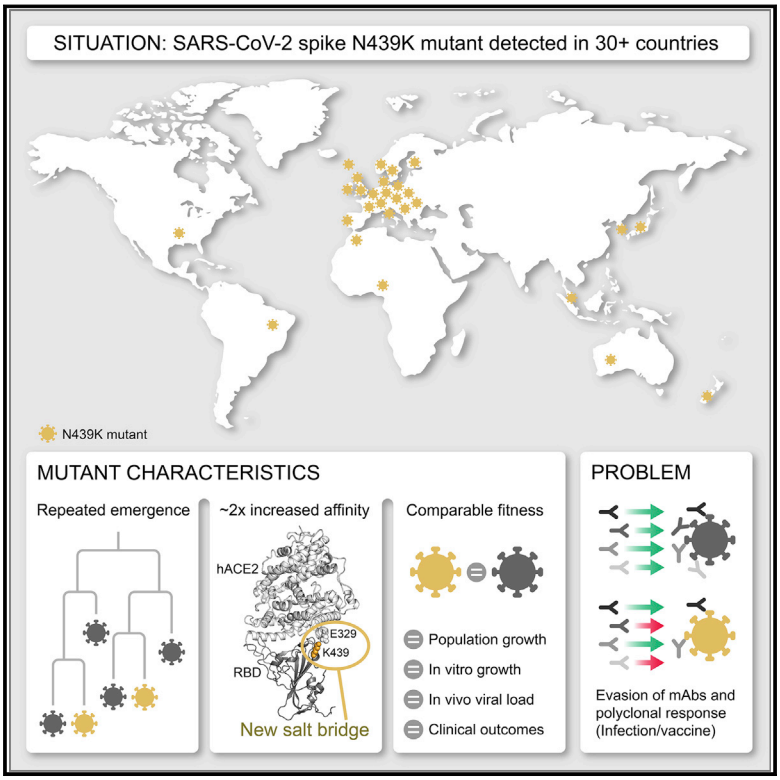
10.1016/j.cell.2021.01.037

Peer reviewed

Cell

Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity

Graphical Abstract



Authors

Emma C. Thomson, Laura E. Rosen, James G. Shepherd, ..., Davide Corti, David L. Robertson, Gyorgy Snell

Correspondence

david.l.robertson@glasgow.ac.uk (D.L.R.),
gsnell@vir.bio (G.S.)

In Brief

Epidemiological, clinical, molecular, and structural characterization of the N439K mutation in the SARS-CoV-2 spike receptor binding motif demonstrates that it results in similar viral fitness compared to wild-type while conferring resistance against some neutralizing monoclonal antibodies and reducing the activity of some polyclonal antibody responses.

Highlights

- The receptor-binding motif (RBM) is a highly variable region of SARS-CoV-2 spike
- RBM mutation N439K has emerged independently in multiple lineages
- N439K increases spike affinity for hACE2; viral fitness and disease are unchanged
- N439K confers resistance to several mAbs and escapes some polyclonal responses

Thomson et al., 2021, Cell *184*, 1171–1187
March 4, 2021
https://doi.org/10.1016/j.cell.2021.01.037

CellPress

Cell

CellPress
OPEN ACCESS

Article

Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity

Emma C. Thomson,[1,2,29] Laura E. Rosen,[3,29] James G. Shepherd,[1,29] Roberto Spreafico,[3,29] Ana da Silva Filipe,[1] Jason A. Wojcechowskyj,[3] Chris Davis,[1] Luca Piccoli,[4] David J. Pascall,[5] Josh Dillen,[3] Spyros Lytras,[1] Nadine Czudnochowski,[3] Rajiv Shah,[1] Marcel Meury,[3] Natasha Jesudason,[1] Anna De Marco,[4] Kathy Li,[1] Jessica Bassi,[4] Aine O'Toole,[6] Dora Pinto,[4] Rachel M. Colquhoun,[6] Katja Culap,[4] Ben Jackson,[6] Fabrizia Zatta,[4] Andrew Rambaut,[6] Stefano Jaconi,[4] Vattipally B. Sreenu,[1] Jay Nix,[7] Ivy Zhang,[8,9] Ruth F. Jarrett,[1] William G. Glass,[8] Martina Beltramello,[4] Kyriaki Nomikou,[1] Matteo Pizzuto,[4] Lily Tong,[1] Elisabetta Cameroni,[4] Tristan I. Croll,[10] Natasha Johnson,[1] Julia Di Iulio,[3] Arthur Wickenhagen,[1] Alessandro Ceschi,[11,12,13] Aoife M. Harbison,[14] Daniel Mair,[1] Paolo Ferrari,[15,16] Katherine Smollett,[1] Federica Sallusto,[17,18] Stephen Carmichael,[1] Christian Garzoni,[19] Jenna Nichols,[1] Massimo Galli,[20] Joseph Hughes,[1]

(Author list continued on next page)

[1]MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow G61 1QH, UK
[2]Department of Clinical Research, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK
[3]Vir Biotechnology, San Francisco, CA 94158, USA
[4]Humabs Biomed SA, a subsidiary of Vir Biotechnology, 6500 Bellinzona, Switzerland
[5]Institute of Biodiversity, Animal Health and Comparative Medicine, Boyd Orr Centre for Population and Ecosystem Health, University of Glasgow, Glasgow G61 1QH, UK
[6]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, UK
[7]Molecular Biology Consortium, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[8]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[9]Tri-Institutional PhD Program in Computational Biology and Medicine, Weill Cornell Graduate School of Medical Sciences, New York, NY 10065, USA
[10]Cambridge Institute for Medical Research, Department of Haematology, University of Cambridge, Cambridge CB2 0XY, UK
[11]Faculty of Biomedical Sciences, Università della Svizzera italiana, 6900 Lugano, Switzerland
[12]Division of Clinical Pharmacology and Toxicology, Institute of Pharmacological Sciences of Southern Switzerland, Ente Ospedaliero Cantonale, 6900 Lugano, Switzerland
[13]Department of Clinical Pharmacology and Toxicology, University Hospital Zurich, 8091 Zurich, Switzerland
[14]Department of Chemistry and Hamilton Institute, Maynooth University, Maynooth, Ireland
[15]Department of Nephrology, Ospedale Civico Lugano, Ente Ospedaliero Cantonale, 6900 Lugano, Switzerland

(Affiliations continued on next page)

SUMMARY

SARS-CoV-2 can mutate and evade immunity, with consequences for efficacy of emerging vaccines and antibody therapeutics. Here, we demonstrate that the immunodominant SARS-CoV-2 spike (S) receptor binding motif (RBM) is a highly variable region of S and provide epidemiological, clinical, and molecular characterization of a prevalent, sentinel RBM mutation, N439K. We demonstrate N439K S protein has enhanced binding affinity to the hACE2 receptor, and N439K viruses have similar *in vitro* replication fitness and cause infections with similar clinical outcomes as compared to wild type. We show the N439K mutation confers resistance against several neutralizing monoclonal antibodies, including one authorized for emergency use by the US Food and Drug Administration (FDA), and reduces the activity of some polyclonal sera from persons recovered from infection. Immune evasion mutations that maintain virulence and fitness such as N439K can emerge within SARS-CoV-2 S, highlighting the need for ongoing molecular surveillance to guide development and usage of vaccines and therapeutics.

INTRODUCTION

SARS-CoV-2, the cause of COVID-19, emerged in late 2019 and expanded globally, resulting in over 82 million confirmed cases as of the end of 2020. Molecular epidemiology studies across the world have generated over 330,000 viral genomic sequences, shared with unprecedented speed via the GISAID Initiative (https://gisaid.org). These data are essential for

Agostino Riva,[20] Antonia Ho,[1] Marco Schiuma,[20] Malcolm G. Semple,[21,22] Peter J.M. Openshaw,[23] Elisa Fadda,[14] J. Kenneth Baillie,[24,25] John D. Chodera,[8] The ISARIC4C Investigators,[26] the COVID-19 Genomics UK (COG-UK) Consortium,[27] Suzannah J. Rihn[1] Samantha J. Lycett,[24] Herbert W. Virgin,[3,28] Amalio Telenti,[3] Davide Corti,[4] David L. Robertson,[1,*] and Gyorgy Snell[3,30,*]

[16]Prince of Wales Hospital Clinical School, University of New South Wales, Sydney, NSW 2052, Australia
[17]Institute for Research in Biomedicine, Università della Svizzera italiana, 6500 Bellinzona, Switzerland
[18]ETH Institute of Microbiology, ETH Zurich, 8093 Zürich, Switzerland
[19]Clinic of Internal Medicine and Infectious Diseases, Clinica Luganese Moncucco, 6900 Lugano, Switzerland
[20]III Division of Infectious Diseases, ASST Fatebenefratelli Sacco, Luigi Sacco Hospital, 20157 Milan, Italy
[21]NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences, Faculty of Health and Life Sciences, University of Liverpool, Liverpool L69 7BE, UK
[22]Respiratory Medicine, Alder Hey Children's Hospital, Liverpool L12 2AP, UK
[23]National Heart and Lung Institute, Imperial College London, London SW3 6LY, UK
[24]The Roslin Institute, University of Edinburgh, Edinburgh EH25 9RG, UK
[25]Intensive Care Unit, Royal Infirmary Edinburgh, Edinburgh EH16 4SA, UK
[26]ISARIC4C Investigators
[27]https://www.cogconsortium.uk
[28]Washington University School of Medicine, Saint Louis, MO 63110, USA
[29]These authors contributed equally
[30]Lead contact
*Correspondence: david.l.robertson@glasgow.ac.uk (D.L.R.), gsnell@vir.bio (G.S.)
https://doi.org/10.1016/j.cell.2021.01.037

monitoring virus transmission and spread (Meredith et al., 2020). Of special interest is the evolution of the SARS-CoV-2 surface protein, spike (S), which is responsible for viral entry via its interaction with the human angiotensin-converting enzyme 2 (hACE2) receptor on host cells. The S protein is the target of neutralizing antibodies generated by infection (Jiang et al., 2020) or vaccination (Folegatti et al., 2020; Jackson et al., 2020; Keech et al., 2020) as well as monoclonal antibody (mAb) drugs currently in clinical trials and/or approved for Emergency Use Authorization (EUA) by the US Food and Drug Administration (FDA) (Chen et al., 2021; Hansen et al., 2020; Jones et al., 2020; Pinto et al., 2020).

A SARS-CoV-2 S amino acid change, D614G, is now dominant in most places around the globe (Korber et al., 2020). Studies *in vitro* indicate that this mutation confers greater infectivity while molecular epidemiology correlates it with an increase in transmissibility with no evidence to date for increased virulence (Hou et al., 2020; Hu et al., 2020; Korber et al., 2020; Volz et al., 2021; Zhang et al., 2020). Amino acid 614 is located outside the receptor binding domain (RBD) of S, the domain targeted by 90% of neutralizing antibody activity in serum of SARS-CoV-2 survivors (Piccoli et al., 2020). Initial studies suggest that D614G viruses exhibit increased sensitivity to neutralizing antibodies, likely due to the effect of the mutation on the molecular dynamics of the S protein (Hou et al., 2020; Weissman et al., 2021; Yurkovetskiy et al., 2020). Therefore, this now dominant variant is unlikely to jeopardize natural or vaccine-derived antibody-mediated immunity generated in response to D614 S protein.

The low numbers of novel mutations reaching high frequency in sequenced SARS-CoV-2 genomes relates to the moderate intrinsic error rate of SARS-CoV-2 RNA replication (Li et al., 2020c; Robson et al., 2020). Nevertheless, the increasing number of infected individuals and the large reservoir of hosts susceptible to infection increase the likelihood that novel variants that impact vaccine and therapeutic development will emerge and spread by chance. Moreover, the full impact of immune selection, which can drive variant selection, has not yet influenced the pandemic, because herd immunity has not been attained. As population immunity increases and vaccines are deployed at scale, this will very likely change. The potential for circulating viral variants to derail promising vaccine or antibody-based prophylactics or treatments, even in the absence of selective pressure from the drug or vaccine, is demonstrated by the failure of a phase III clinical trial of a mAb targeting the respiratory syncytial virus (Simões et al., 2020) and the need for new influenza vaccines on a yearly basis. It is therefore critical to understand whether and how SARS-CoV-2 may evolve to evade antibody-dependent immunity.

Here, we examine the immunodominant SARS-CoV-2 receptor binding motif (RBM), the primary target of the neutralizing Ab response within the RBD (Piccoli et al., 2020), and find it to be a highly variable region of the S protein in circulating viruses. To understand the implications of this structural plasticity, which could allow the RBD to accommodate amino acids changes that could contribute to immune evasion, we defined the clinical and epidemiological impact, molecular features, and immune response to the RBM mutation N439K. This amino acid replacement has arisen independently multiple times, and in two cases formed lineages of more than 500 sequences. As of January 6, 2021, it was observed in 34 countries and was the second most commonly observed RBD mutation worldwide, and the sixth most common S mutation. We find that the N439K mutation results in enhanced RBD affinity for hACE2, it is associated with a similar clinical spectrum of disease and slightly higher viral loads *in vivo* compared to viruses with the wild-type (WT) N439 residue, and it results in immune escape from polyclonal sera from a proportion of recovered individuals and some neutralizing mAbs.

N439K provides a sentinel example of immune escape, indicating that RBM variants must be evaluated when considering vaccines and the therapeutic or prophylactic use of mAbs.

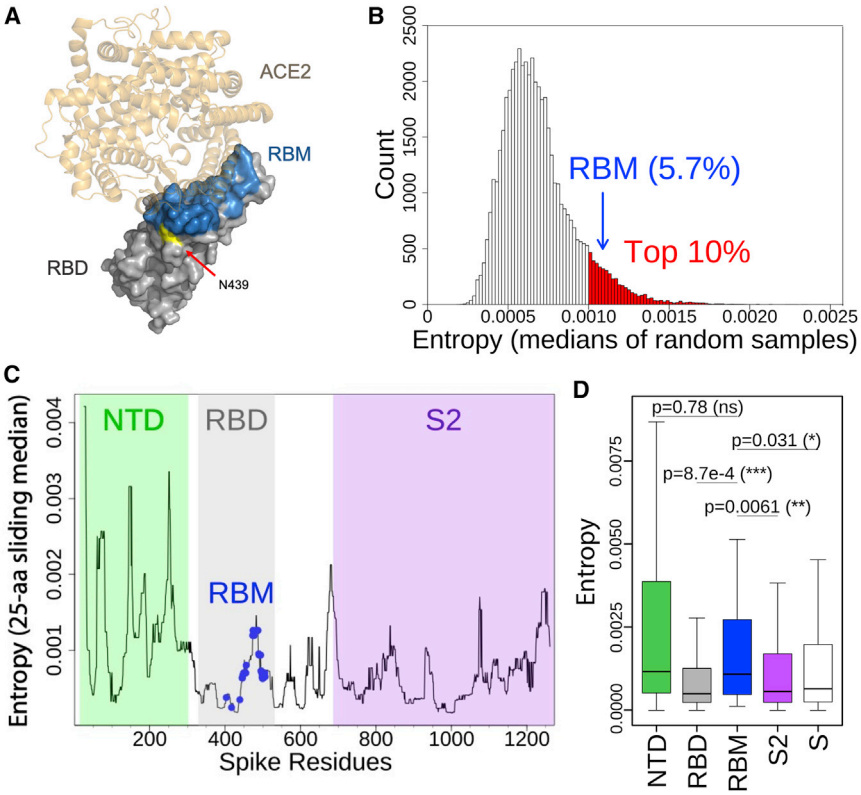


Figure 1. The RBM exhibits significant natural diversity in circulating SARS-CoV-2 viruses

SARS-CoV-2 variants (retrieved from CoV-GLUE) are based on 209,239 high-quality sequences downloaded from GISAID on November 30, 2020.

(A) Structure of the SARS-CoV-2 RBD-hACE2 complex (PDB: 6M0J) highlighting the RBM (blue) and residue N439 (yellow).

(B) Thirty-four residues (the size of the RBM) were randomly sampled without replacement 50,000 times from the mature S protein (excluding the RBM). Median entropies were computed for each draw. The resulting 50,000 median entropies were used to build the entropy distribution of residues other than the RBM. The top 10% medians are highlighted in red. The median entropy of RBM residues was compared with the non-RBM entropy distribution to determine the variability of the RBM relative to non-RBM residues. To allow for a fair comparison, sampling was performed without enforcing residue contiguity, as the RBM is not contiguous in sequence space. Therefore, in any given sample, residues are unlikely to share any functional relationship.

(C) Per-residue entropies of the mature S protein were smoothed by plotting medians of a 25-aa center-aligned sliding window. Smoothing allows visualizing local peaks of variability. The RBM residues and the NTD, RBD, and S2 domains are highlighted. Due to the non-contiguous nature of the RBM in sequence space, the sliding window median at RBM residues is diluted by neighboring non-RBM residues.

(D) Boxplot of per-residue entropies in four S domains (or full mature S protein). The lower and upper hinges correspond to the first and third quartiles. The lower/upper whiskers extend from the hinge to the smallest/largest value no further than 1.5 times the inter-quartile range. Outliers beyond the end of the whiskers are not plotted but are retained for statistical testing. Pairwise comparisons by Mann-Whitney U tests. p value thresholds are 0.05 (*), 0.01 (**) and 0.001 (***); ns, not significant.

See also Figures S1 and S2.

Long-term control of the pandemic with vaccines will require systematic monitoring of immune escape variants and may require new vaccine preparations that address the variants circulating globally.

RESULTS

The RBM is a variable region of the SARS-CoV-2 S protein

Competing pressures influence the evolution of the S RBM. First, the RBM mediates viral entry (Shang et al., 2020; Walls et al., 2020; Wrapp et al., 2020b) and therefore must maintain sufficient affinity to engage the entry receptor hACE2. Second, it is a major target of neutralizing antibodies (Piccoli et al., 2020; Robbiani et al., 2020; Rogers et al., 2020; Wec et al., 2020) and so would be a primary location for the emergence of immune escape mutations. We set out to understand these competing pressures by evaluating the landscape of RBM sequence divergence observed in circulating SARS-CoV-2 variants and in other viruses of the *Sarbecovirus* lineage.

We used re-refined published X-ray structures of SARS-CoV and SARS-CoV-2 RBD:hACE2 complexes (Lan et al., 2020; Li et al., 2005) to define the RBM residues using a 6 Å distance cutoff (Figures 1A–1D and S1A, 2). We evaluated SARS-CoV-2 genomic sequences deposited in GISAID as of November 30th, 2020 and observed a high number of variants occurring in the RBM. To understand how the variability of the RBM compares to the variability of the entire RBD and the whole S protein, we evaluated well-defined S protein domains: within S1, the N-terminal domain (NTD) and the RBD (further split into RBM and non-RBM), and the S2 domain. Analysis of entropy, which estimates sequence variability at a given position in a protein alignment, identified the RBM as a highly variable region of the RBD and of the entire S protein (Figures 1B–1D), with a median entropy within the top 10% of equivalently sized sets of randomly sampled residues (Figure 1B). This result is confirmed by an analysis of sequence variability that is not weighted by total counts of each variant, thereby capturing the diversity of circulating variants with mitigated bias toward oversampled variants (Figure S1A).

To understand constraints on RBM variability, we evaluated the published deep mutational scanning (DMS) dataset of the RBD (Starr et al., 2020b) and compared it to sequences of circulating viruses. The DMS data define the effect of each possible single amino acid change on both expression of the RBD and its capacity to bind hACE2. For each position in the RBM, we

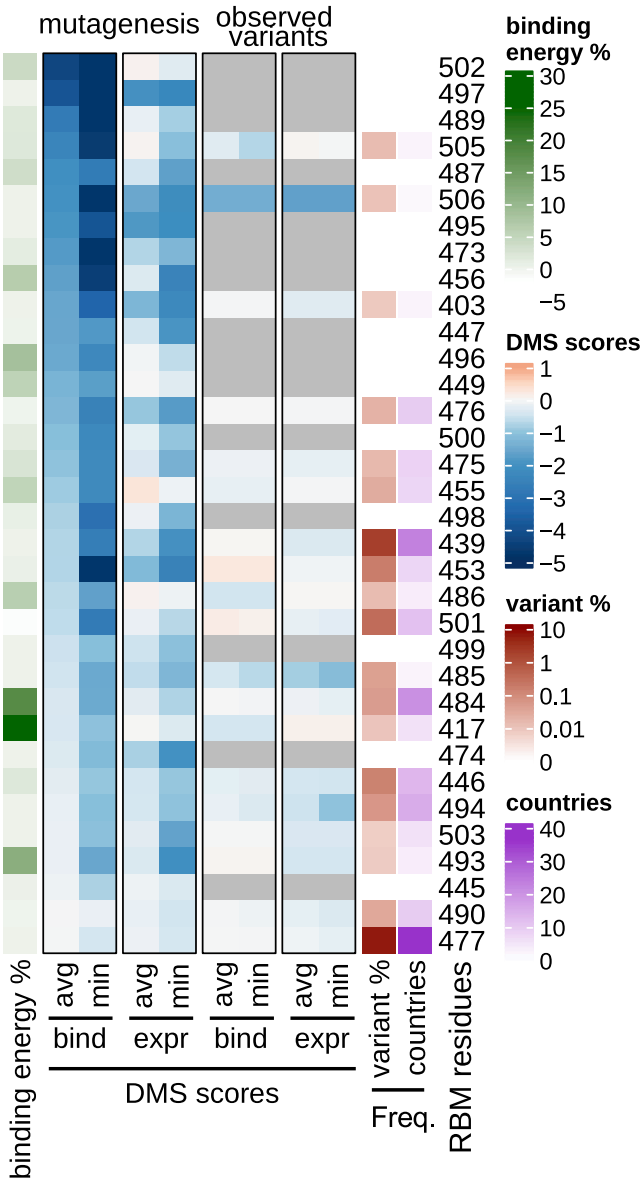


Figure 2. RBM functional constraints compared to RBM natural diversity

Each residue in the RBM is annotated by several metrics, depicted as a heatmap. DMS scores: outlined in black boxes (center) are summaries of hACE2 binding and RBD expression deep mutational scanning (DMS) experimental results (Starr et al., 2020b). DMS score is the binding or expression fold change of a variant over WT on a log10 scale (red indicating improvement and blue indicating loss as compared to WT). In the "mutagenesis" columns, DMS results are given for each residue as either the minimum (most disruptive variant) or the average score across all possible variants of a residue, except for the reference residue and the stop codon. In the "observed variants" columns, minimum and average scores are computed only across variants that have been observed in GISAID (same set of sequences as used for Figure 1). When no natural variants have been observed, cells are gray. Data were sorted on the leftmost DMS column. Frequency: each RBM position is annotated with the frequency of non-reference amino acids in deposited sequences (darker red indicating higher frequency; at least 1 supporting sequence per 25,000 deposited sequences is required to call a variant). The number of countries in which variants have been observed is also annotated (darker purple indicating more countries). Binding energy: a re-refined SARS-CoV-2 RBD:hACE2 complex X-ray structure (PDB: 6M0J) was used to determine the approximate, decomposed binding free energy associated with each RBM residue. Results for each RBM residue are expressed as a percentage of the total binding interface interaction energy (darker green indicating stronger contribution to the binding energy).

See also Figures S1 and S2.

compared the DMS results for all amino acid replacements at that position versus only changes that have been observed in circulating SARS-CoV-2 variants (Figure 2). A subset of residues shows the largest loss of hACE2 binding on mutation (top ~1/3 of RBM residues in Figure 2) and, as would be expected, few natural occurrences of mutations at these residues have been observed to be circulating. However, for the majority of the RBM (bottom ~2/3 of RBM residues in Figure 2), variation in circulating virus sequences confirms the tolerance to mutation predicted by the DMS data.

To further assess the ability of the RBM to accommodate mutations without disrupting hACE2 binding, we examined the structural dynamics and energetics of the RBM:hACE2 binding interface. We performed an approximate, residue-level decomposition of binding free energy based on the RBD:hACE2 complex X-ray structure (green in Figure 2) as well as molecular dynamics simulations of the complex, resulting in ~118 μs of aggregate simulation data (Figure S1B). Consistent with expectation, the two residues with the highest variant frequency (S477 and N439) contribute weakly to the binding energy (Figure 2). Surprisingly, the two RBM residues with the strongest interactions with hACE2 based on the X-ray structure (K417 and E484, dark green in Figure 2) were not highly conserved (variant % in red, Figure 2), with ~10-fold more variants for E484. This could be explained by results from the molecular dynamics simulation: K417 formed close interactions with hACE2 70% of the simulation time, while E484 only 3% of the time (Figure S1B). The low percent for E484 is also consistent with the non-conservative amino acid replacements observed for circulating variants (e.g., the most common E484 substitution is currently E484K), with a positively charged lysine substituting for the negatively charged glutamate. Overall, these results demonstrate that the RBM has a high degree of structural plasticity whereby it is able to accommodate amino acid changes without disrupting hACE2 binding.

Evolutionary analysis of the *Sarbecovirus* subgenus provides further support for RBM plasticity (Boni et al., 2020; Li et al., 2020b; Rambaut et al., 2020). The SARS-CoV RBM is highly divergent from the SARS-CoV-2 RBM (Figures S2A and S2B) while maintaining hACE2 binding affinity. Additionally, there are many sequence changes in the RBM across a panel of related coronaviruses from animal isolates (Figures S2A and S2B; Table S1). To determine the ability of members of the *Sarbecovirus* lineage to bind hACE2, we produced nine recombinant RBD proteins corresponding to seven animal isolates, SARS-CoV-2, and SARS-CoV, and evaluated their binding to recombinant hACE2 (Figure S2C). We found that three of the RBDs from animal isolates showed strong affinity for hACE2: GD Pangolin, which has a highly similar RBM to SARS-CoV-2, GX Pangolin, which has a more divergent RBM, and Bat CoV WIV1 which is highly

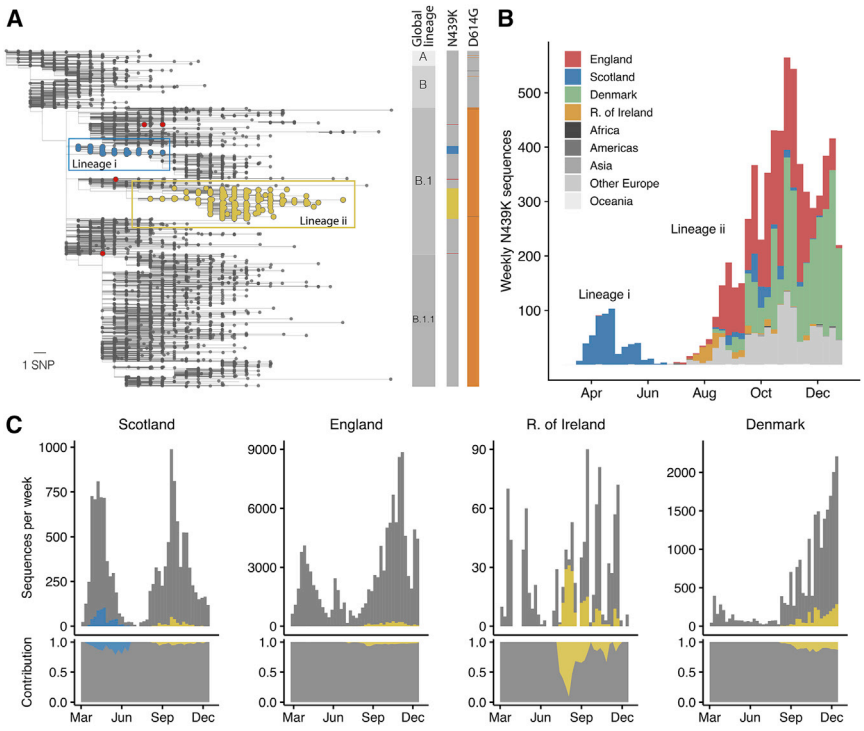


Figure 3. The N439K RBM mutation has arisen independently multiple times, twice forming significant lineages

(A) Phylogenetic tree (de-duplicated and down-sampled) showing the relationship among representative global SARS-CoV-2 variants, with N439K variants highlighted in color. Two significant N439K lineages, one in Scotland (>500 sequences, blue circles) and one in 32 countries (>6,000 sequences, yellow circles) were detected as of January 6, 2021. The N439K mutation has also emerged independently on at least seven occasions (red circles show four of these) bringing the total country count to 34. Vertical bars indicate global lineage, the presence of N439K (same colors as tree), D614G (orange) or D614N (dark gray). The scale bar corresponds to a single nucleotide polymorphism (SNP).

(B) Frequency of N439K variants relative to sampling time and their geographical area of occurrence (see key): Africa (Morocco, Nigeria), Americas (Brazil, USA), Asia (Japan, Singapore, South Korea), the European countries Denmark, England, Republic of Ireland and Scotland and other European countries (Belgium, Bosnia-Herzegovina, Croatia, Czech Republic, Faroe Islands, Finland, France, Germany, Hungary, Italy, Luxembourg, Netherlands, Northern Ireland, Norway, Poland, Romania, Slovakia, Sweden, Switzerland, Wales), and Oceania (Australia, New Zealand). The prominent light gray bars correspond to other European countries. See Table S2 for total numbers for each country.

(C) Frequency of the two N439K lineages (same colors as A) over time relative to all sequences for that country (gray) and their normalized contributions (lower panels) in Scotland, England, Republic of Ireland, and Denmark.

See also Figure S3.

divergent (Figures S2A and S2B). This further indicates that the RBM is structurally plastic, retaining binding with hACE2 as a receptor despite changes to sequence. Given this plasticity, we next considered whether an RBM variant can lead to immune evasion while retaining virulence.

Phylogenetic analysis of the prevalent SARS-CoV-2 RBM mutation N439K

N439K is a prevalent RBM mutation (the second most common mutation in the RBD through the end of 2020) which was first sampled in March 2020 in Scotland from lineage B.1 (Rambaut et al., 2020) on the background of D614G. Using phylogenetic analysis, we determined that the earliest reported N439K sequences represented a single SARS-CoV-2 lineage (Figure 3A) that increased in frequency to 542 sequences in Scotland by June 20, 2020 (~10% of the available Scottish viral genome sequences for this time period). Subsequently, numbers of N439K and all other variants decreased in Scotland concurrent with control of the pandemic after initiation of stringent public health measures, with this specific N439K lineage (designated here as lineage i) not being detected since June 2020 (Figures 3B and 3C). However, the N439K mutation appears in >6,000 additional sequences in the GISAID database as of January 6, 2021. Our analysis demonstrates that the majority of these sequences represent a second, independent lineage (designated lineage ii) which was first sampled in Romania on May 13, 2020, then Norway on June 23, 2020, and is now detected to be circulating in 32 countries (Figures 3A–3C). N439K lineages i and ii have recently received the lineage designations B.1.141 and B.1.258, respectively (Rambaut et al., 2020). We also observe at least seven instances of the N439K mutation that have arisen independently of these two large lineages, including again in the United States in at least four linked infections, and in Brazil and Nigeria where no lineage ii/B.1.258 has been observed, resulting in a total of 34 countries where N439K has been detected to date (Figures 3A and 3B).

Sequence counts are heavily influenced by sampling frequency, which varies widely between countries, and N439K as a percentage of total sequences appears low: as of January 6, 2021, there have been 6,868 N439K observations in GISAID, 2% out of ~290,000 SARS-CoV-2 genome sequences for the 34 countries where this mutation has been detected (Table S2). Nevertheless, when comparing the percentage of N439K sequences over time in countries with sufficient data, the proportion can be significant: ~10% in Scotland from March to June 2020 and ~10% in Denmark from August to December 2020, both countries with high sequencing rates, and ~13% in Ireland from July to December 2020, where regional coverage is reasonable, but the sequencing rate is lower (Figure 3C). Importantly, on the scale of a pandemic, small proportions correspond to large numbers of infections. If the proportion of N439K sequences in each country predicts what proportion of its confirmed infections are associated with N439K variants, then N439K variants correspond to ~764,000 of the confirmed SARS-CoV-2 infections as

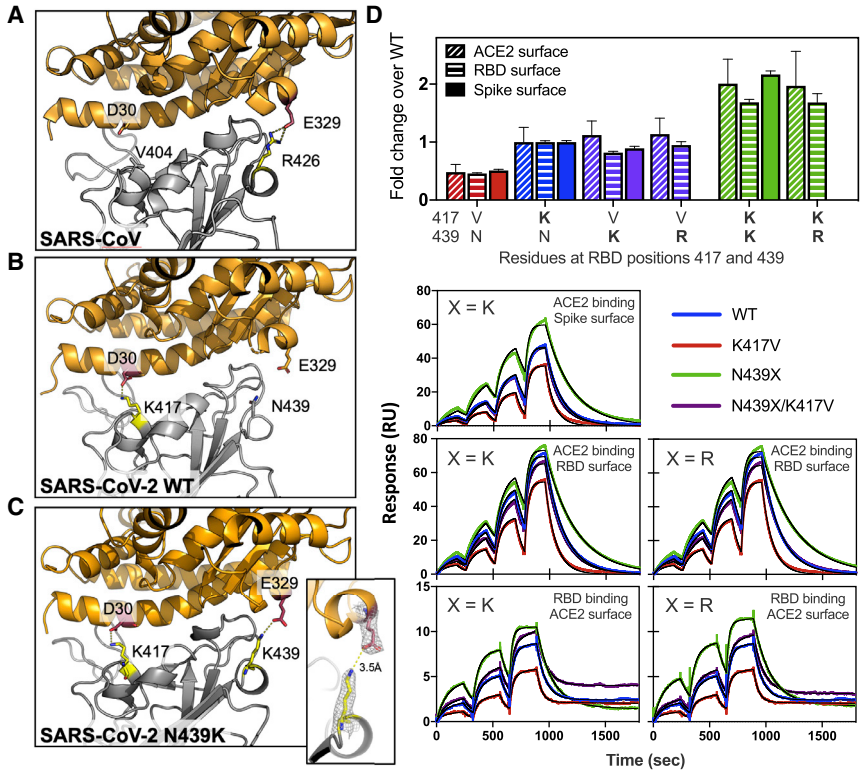


Figure 4. N439K creates a new RBD:hACE2 salt bridge and enhances RBD:hACE2 affinity

(A–C) X-ray structures of the SARS-CoV (A), SARS-CoV-2 WT (B), and SARS-CoV-2 N439K (C) RBD in complex with hACE2 (based on 2AJF, 6M0J, and current work, respectively). Select interface residues are shown as sticks. hACE2 is shown in orange and RBD in gray. The inset in (C) shows the 2Fo-Fc electron density contoured at 1σ for the K439-E329 salt bridge.

(D) Binding affinity of RBD and Spike variants for hACE2 measured by surface plasmon resonance. Monomeric hACE2 is injected successively at 11, 33, 100, and 300 nM onto surface-captured spike extracellular domain (ECD) or RBD; alternately, RBD is injected successively at 3.1, 12.5, and 50 nM onto surface-captured hACE2. All spike ECD contain the D614G mutation. Bar graph: affinity measurements (averages of 3–4 replicates) expressed as a fold change relative to WT binding within each experiment format, where >1 indicates improved binding (smaller K_D) relative to WT. WT K_D values measured as: 95 ± 1.6 nM (Spike surface), 63 ± 1.0 nM (RBD surface), 19 ± 3.3 nM (hACE2 surface); errors are SEM.

See also Table S3.

of January 6, 2021 (Table S2). If detected cases represent 5%–33% of true infections, as has been estimated for the United States (Wu et al., 2020b), then a very rough approximation of the actual cumulative number of N439K-associated infections would be in the range of 2–15 million.

Overall, the spread of N439K to at least 34 countries is concerning, as is its repeated independent emergence. At the nucleotide level, all N439K variants to date have arisen from the same mutation: a C-to-A transversion in the third codon position. Interestingly, 4,209 of sequences in lineage ii/B.1.258 also carry the S 69-70 deletion that has occurred independently multiple times in the pandemic and most notably with the Y453F amino acid replacement associated with mink infections (Oude Munnink et al., 2021). In both cases the 69-70 deletion mutation has arisen subsequent to the RBM mutation and then been retained in all subsequent variants. This deletion has also been recently reported to provide an escape for NTD-specific neutralizing antibodies (McCarthy et al., 2021). Very recently, this deletion has also been observed to co-occur with another RBM mutation, N501Y (Volz et al., 2021).

Because there is concern that mutations with high prevalence may have increased virus transmissibility, we next evaluated whether any difference could be detected in the rate of spread of the N439K lineages as compared to other lineages. Because Scotland has a high sampling frequency for its population size (Table S2), it is possible to calculate a growth rate for N439K lineage i based on a comparison with other Scottish lineages (see STAR methods and http://sars2.cvr.gla.ac.uk/RiseFallScotCOVID/). We find that while the N439K/D614G lineage is one of the largest to emerge in Scotland, its growth rate is similar to the median N439/D614 or N439/D614G WT growth rates, with no evidence for a faster growth conferred by the N439K mutation (Figure S3A).

N439K RBD forms a new interaction with hACE2 and has enhanced hACE2 affinity

In addition to its frequency and repeated emergence, the N439K mutation stood out from other circulating RBM mutations as having a plausible mechanism for maintenance of viral fitness. The equivalent position to N439K in the SARS-CoV RBM is also a positively charged amino acid (R426), which forms a salt bridge with hACE2 (Li et al., 2005) (Figure 4A). We therefore hypothesized that the N439K SARS-CoV-2 variant may form a similar salt bridge at the RBD-hACE2 interface (RBD N439K:hACE2 E329) (Figure 4B). We determined the X-ray structure of the N439K RBD in complex with hACE2 at 2.8 Å resolution and observed that this new interaction does indeed form (Figure 4C; Table S3). Because salt bridges can be strong non-covalent bonds, and therefore the N439K mutation plausibly adds a strong interaction at the binding interface, we hypothesized that the N439K variant has enhanced binding for hACE2.

To test this hypothesis, we used surface plasmon resonance (SPR) to evaluate binding of recombinant N439K S or RBD protein to recombinant hACE2. We also evaluated the N439R and K417V variants, each of which are found in SARS-CoV at these positions, and the latter of which would remove a salt bridge at the RBD:hACE2 interface. Across multiple assay formats, we found that the N439K and N439R variants exhibited an ~2-fold enhanced binding affinity for hACE2 as compared to the original N439 variant (termed herein WT) (Figure 4D). The magnitude of this enhancement was paralleled by an ~2-fold loss of binding

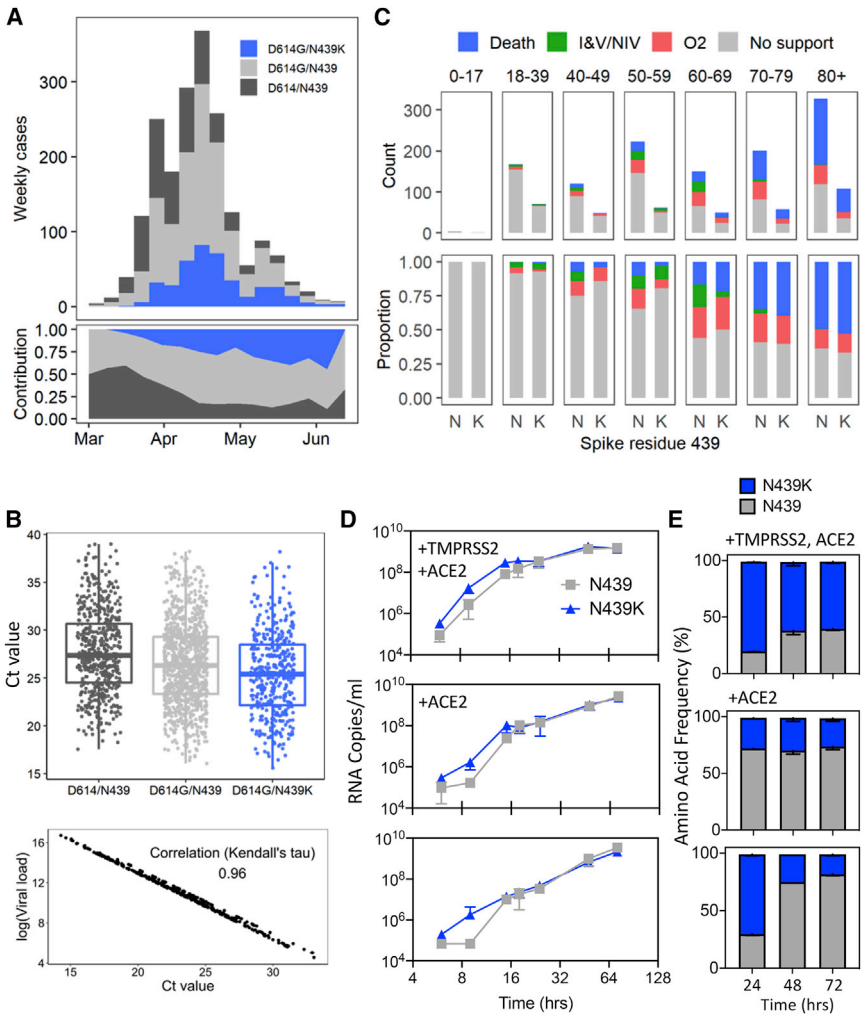


Figure 5. Clinical outcomes and virological evaluation of N439K lineage i indicate maintenance of fitness relative to WT virus

(A) Epidemiological growth of the N439/D614, N439/D614G, or N439K/D614G virus in the National Health Service (NHS) Greater Glasgow and Clyde (GGC) Health Board area relative to sampling time in epidemiological (epi) weeks (top) and their relative contributions (bottom) for 1,918 patients whose diagnostic samples were sequenced.

(B) Top: real-time PCR data for N439/D614, N439/D614G, and N439K/D614G groups, same patient population as in (A). The N439K genotype was associated with marginally lower Ct values than the N439 genotype (posterior mean Ct value difference between N439K/D614G and N439/D614G: −0.65, 95% CI: −1.22, −0.07). Bottom: correlation between Ct and quantitative viral load.

(C) Severity of disease within NHS GGC for a subset of 1,591 patients. Ordinal scale scored by requirement for supplementary oxygen: (1) no respiratory support, (2) supplemental oxygen, (3) invasive or non-invasive ventilation or oxygen delivered by high-flow nasal cannula, and (4) death. Ordinal regression analysis indicated that the N439K viral genotype was associated with similar clinical outcomes compared to the N439 genotype (posterior mean of N439K/D614G genotype effect: 0.06, 95% CI: −1.21, 1.33).

(D) Growth curves for GLA1 (N439/D614G) or GLA2 (N439K/D614G) virus isolates in Vero E6 cells with ACE2 and TMPRSS2 overexpression (+TMPRSS2 +ACE2), ACE2 overexpression (+ACE2), or no overexpression. Error bars are SD from three replicates.

(E) Competition of GLA1 and GLA2 virus isolates for growth in Vero E6 cells with ACE2 and TMPRSS2 overexpression (+TMPRSS2 +ACE2), ACE2 overexpression (+ACE2), or no overexpression, after inoculation at a matched MOI. Quantification of each virus was performed by tracking the frequency of N439K within the spike gene using metagenomic NGS. Error bars are SD from three replicates.

See also Figure S3 and Tables S4–S6.

affinity for the K417V variant relative to WT. Our data are in line with the DMS results (Starr et al., 2020b), which show a 2-fold loss of binding for K417V and no change for N439K/R, as the two assays are inherently different and the DMS data are much higher-throughput but lower sensitivity. We also tested the effect of the N439K/R and K417V mutations in combination. These double mutants swap one salt bridge at the hACE2 binding interface at RBD position 417 for one at position 439; we found they had an hACE2 affinity similar to the WT (Figure 4D).

Overall, these data indicate that acquisition of the N439K mutation enhances hACE2 binding, which could have implications *in vivo* in the context of infection and transmission. At a minimum, we found no evidence for any decreased success of N439K lineage i relative to other lineages present in Scotland at the same time (Figure S3A). The enhanced affinity could compensate for other mutations that would otherwise decrease binding (e.g., K417V), further highlighting the plasticity of the RBM and the need for surveillance.

N439K SARS-CoV-2 maintains fitness and virulence

The enhanced hACE2 affinity conferred by the N439K mutation, its geographical emergence as independent lineages, as well as its prevalence among circulating viral isolates is consistent with no effect on viral fitness. We set out to directly examine N439K impact on viral fitness by evaluating clinical data and outcomes associated with virus carrying the N439K mutation versus WT N439, as well as by direct *in vitro* viral growth and competition. Clinical data including age, gender, date of diagnosis, hospitalization status, and mortality were collected prospectively, and sequencing was carried out in real time, as part of the Scottish strategy for COVID-19 surveillance.

We used qPCR to evaluate viral load (as measured by cycle threshold [Ct]) in 1,918 Scottish patients whose positive samples had been sequenced (Figures 5A and 5B). Variants were either N439K/D614G (n = 406), N439/D614G (n = 978), or ancestral (N439/D614) (n = 534). Our analysis found strong evidence that the N439K/D614G genotype was associated with marginally lower

Ct than the N439/D614G genotype, even after controlling for confounders: age, sex, viral co-ancestry, and epidemic stage (mean Ct value difference between N439K/D614G and N439/D614G: −0.65, 95% confidence interval [CI]: −1.22, −0.07) (Figure 5B; Table S4). Assuming the PCR was 95% efficient, then a mean Ct difference of 0.65 would represent an RNA copy number increase of 1.54-fold in N439K/D614G relative to N439/D614G. Because Ct measurements were from multiple locations in Scotland, a subanalysis of viral load using RNA standards was carried out with available samples. This analysis showed a near-complete correlation with Ct values (Figure 5B). D614G has previously been associated with higher viral loads/lower Ct values (Korber et al., 2020; Lorenzo-Redondo et al., 2020; Mueller et al., 2020; Volz et al., 2021); although our data suggest a similar trend in a naive analysis, when controlling for confounders (given above), we could not detect this effect (Table S4).

Clinical outcomes were also obtained for a subset of these patients (n = 1,591), who were scored for severity of disease based on oxygen requirement: (1) no respiratory support, (2) supplemental oxygen, (3) invasive or non-invasive ventilation or high flow nasal cannula, or (4) death (Figures 5C and S3B). The requirement for oxygen therapy or ventilation was collected retrospectively. Variant counts for the clinical outcome analysis were double mutant (N439K/D614G, n = 399), D614G mutants (with N439 WT, n = 735), or ancestral genotype (N439/D614, n = 457). Our ordinal regression indicated that the N439K/D614G viral genotype was associated with similar clinical outcomes compared to D614G or ancestral genotypes (posterior mean of N439K/D614G genotype effect: 0.06, 95% CI: −1.21, 1.33) (Table S5). All other results from the severity analysis were qualitatively similar to a previous analysis of the D614G mutation (Volz et al., 2021). These clinical outcome data indicate that the N439K virus is neither linked to an attenuated phenotype nor linked to increased severity.

We next experimentally tested growth of two representative SARS-CoV-2 isolates, GLA1 (N439) and GLA2 (N439K), both with the D614G background (Table S6). Culture was carried out for 72 h in Vero E6 cells with either hACE2 and TMPRSS2 overexpression, hACE2 overexpression, or no overexpression. There was no significant difference between the growth of these isolates after inoculation at multiplicities of infection (MOIs) of 0.005 and 0.01. The N439K variant replicated slightly faster initially after inoculation (Figure 5D). These experimental data indicate that the N439K mutation does not exhibit positive or negative effects on viral growth. To further assess fitness for replication in cultured cells, we carried out a cross-competition assay using inoculation of cells at a matched MOI followed by quantitation of N439 and N439K by metagenomic sequencing over time (Figure 5E). N439K demonstrated similar fitness as the WT N439 variant, with a slight fitness advantage for N439K in cells expressing TMPRSS2. Collectively, these results indicate that the N439K mutation results in viral fitness that is similar or possibly slightly improved relative to the WT N439 virus. These results may relate to the improved hACE2 affinity measured for the N439K RBD in the SPR binding assays, or could relate to additional mechanisms, such as changes to S density on the viral particle surface or changes to the conformational dynamics of the S protein.

The N439K mutation promotes evasion of antibody-mediated immunity

Having established that the N439K mutation has no detectable effect on virus replication, we sought to test whether it promotes evasion of antibody-mediated immunity by evaluating recognition of N439K RBD by mAbs and by polyclonal immune serum from 442 recovered individuals, including six donors who were infected by the SARS-CoV-2 N439K variant. 6.8% of the tested sera showed a >2-fold reduction in binding to N439K RBD as compared to WT (Figures 6A, 6B, and S4; Data S1). In some individuals, the >2-fold reduction diminished the RBD ED_{50} response below 30 (Figure 6A; Data S1), a threshold previously determined to be a cutoff for specific binding (Piccoli et al., 2020). Thus, the response to the RBD can be significantly influenced by the N439K mutation in a number of individuals infected by WT SARS-CoV-2. The majority of serum samples for which there was a loss of binding were those that had overall lower Ab titers against WT RBD. The sera from the six individuals known to have recovered from infection with SARS-CoV-2 N439K virus all showed <2-fold change in binding levels to WT RBD as compared to N439K RBD (Figures 6A, 6B, and S4). This may reflect a true variant-specific response or that differential binding could not be measured due to the limited number of samples analyzed.

To understand our results at the level of individual antibodies, we evaluated a panel of 140 mAbs isolated from individuals recovered from SARS-CoV-2 infection early in the pandemic (likely N439 WT virus), which are a representative sample of the RBD-targeting mAbs generated after infection (Piccoli et al., 2020; Tortorici et al., 2020). We also evaluated mAbs REGN10933, REGN10987, LY-CoV555, and S309 (the parent of VIR-7831) which are either clinical stage or approved for EUA (Baum et al., 2020; Chen et al., 2021; Hansen et al., 2020; Pinto et al., 2020). 16.7% of these mAbs demonstrated a >2-fold reduction of RBD binding in response to the N439K mutation (Figures 6C, 6D, and S5; Data S1). For comparison, we also evaluated the K417V and N439K/K417V mutations. A similar percentage, 9.7% for K417V and 14.6% for N439K/K417V, lost >2-fold binding to these variants (Figures 6C, 6D, and S5; Data S1). Of note, some mAbs demonstrated a larger loss of binding to the double mutant as compared to either single mutant (Figures 6C, 6D; and S5; Data S1). The reduced binding of mAbs to these RBD mutants was also confirmed by bio-layer interferometry analysis (Figures 6E and S6). The mAb panel was evaluated by RBD-binding competition experiments with hACE2 as well as with three structurally characterized antibodies defining distinct epitopes on the RBD: S304/site II, S309/site IV, and S2H14/site I, the latter significantly overlapping with the RBM (Piccoli et al., 2020). The majority of the panel were site I, hACE2-blocking mAbs; the mAbs with sensitivity to N439K were enriched for site I mAbs with moderate or weak/no hACE2 blockade, consistent with the positioning of N439K at the edge of the RBM (Figures 1A and 6F; Data S1).

To define the potential biological importance of these mutations for evasion of antibody-mediated neutralization, we tested mAbs against pseudoviruses expressing S variants N439K, K417V, and N439K/K417V (Figures 7A–7C and S7; Data S1). Neutralization of pseudoviruses containing these mutations

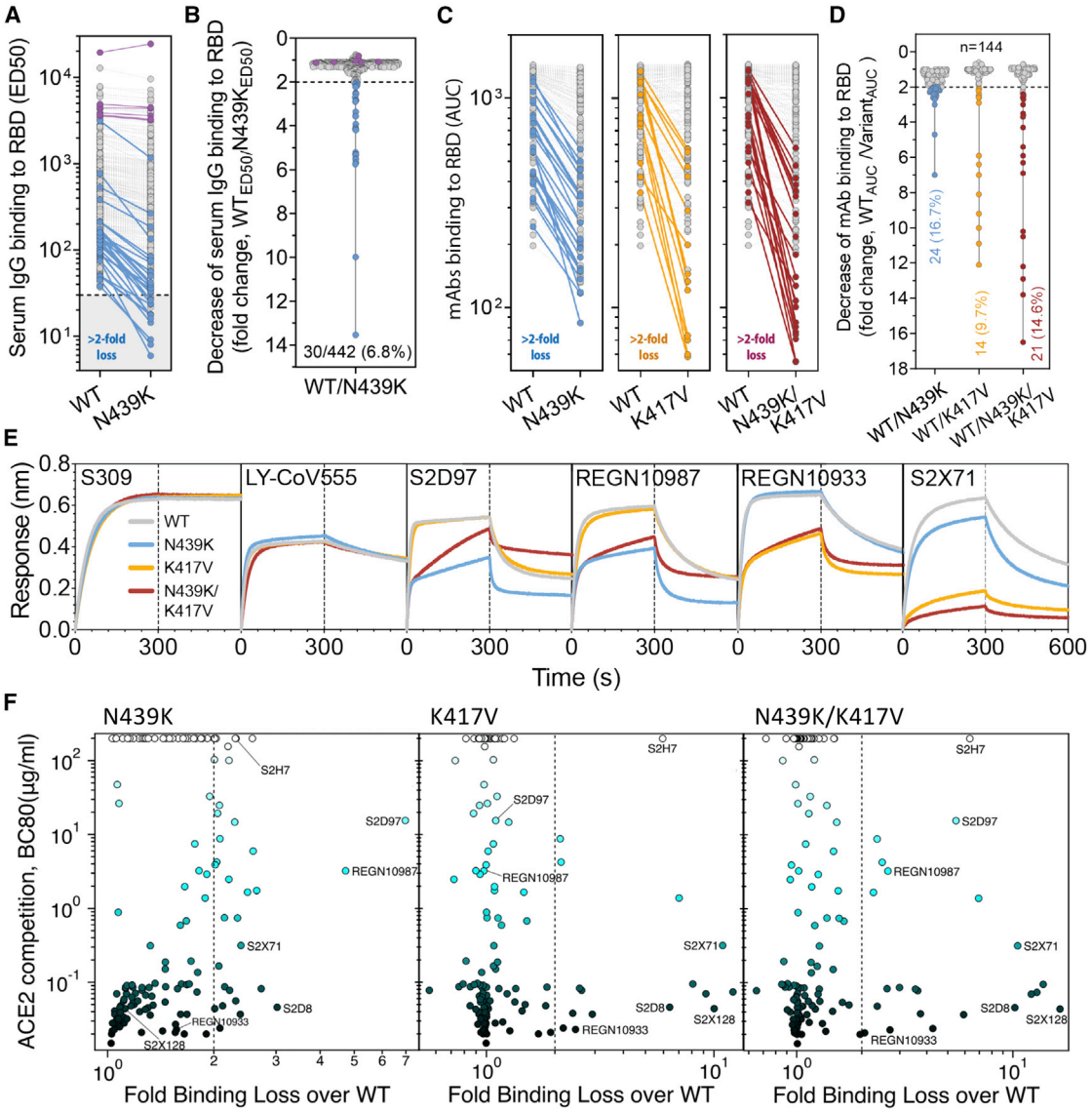


Figure 6. RBM variants exhibit escape from monoclonal antibodies and sera binding

(A and B) Binding of serum and plasma samples from 442 SARS-CoV-2 infected individuals against WT and N439K RBD plotted as (A) ELISA ED_{50} for each RBD (cut-off for positive binding to WT set at 30) and (B) fold change relative to WT. Data shown are the average of two independent replicates (source data given in Data S1). Blue dots indicate sera with at least 2-fold loss of binding to the N439K RBD variant as compared to WT in both replicates. Purple dots indicate sera from individuals infected with SARS-CoV-2 N439K variant.

(C and D) Binding of 140 mAbs from SARS-CoV-2 infected individuals and four clinical-stage or EUA-approved mAbs against WT, N439K, K417V, and N439K/K417V RBD, plotted as (C) ELISA AUC for each RBD and (D) fold change relative to WT. Data shown are the average of two independent replicates (source data given in Data S1). For all, the colored dots indicate mAbs demonstrating at least 2-fold loss of binding to the variant RBD as compared to WT (counted if the average of both replicates is at least 2-fold and each individual replicate is at least 1.7-fold).

(E) Kinetics of binding to RBD variants by Octet of six representative mAbs (representative of n = 2 independent experiments).

(F) Distribution of the 144 mAbs based on binding to RBD variants (expressed as fold-change over WT) and hACE2 competition (expressed as the mAb concentration blocking 80% of hACE2 binding, BC80, also indicated as a blue gradient; source data in Data S1). Higher BC80 values (lighter blue) correspond to less hACE2 competition, with mAbs indicated at the top of the panels (white) showing no competition at all.

See also Figures S4, S5, and S6.

was significantly diminished for certain mAbs, including some that are currently in use in patients under EAU. As predicted by its non-RBM epitope (Pinto et al., 2020), S309 was capable of neutralizing each of these variants. We also evaluated a cross-reactive camelid nanobody, VHH-72, which has enhanced potency for SARS-CoV as compared to SARS-CoV-2, predicted to be partially due to a contact with R426 in SARS-CoV RBD, the same position as 439 in SARS-CoV-2 RBD (Wrapp et al.,

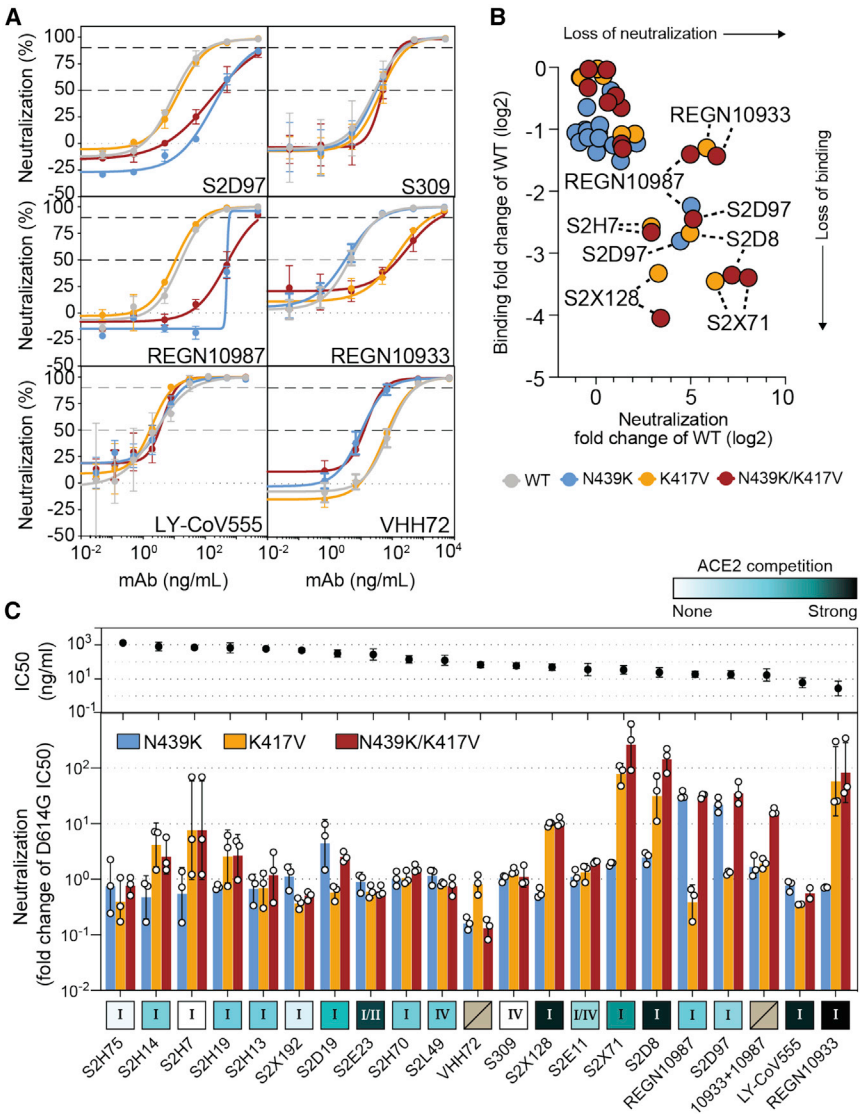


Figure 7. Neutralization of four RBM variants by a panel of antibodies and a nanobody

(A) Neutralization of four VSV-pseudovirus variants by six of the mAbs tested. Data shown are representative of n = 3 biological replicates, bars = SD of technical duplicate (Data S1).

(B) Correlation of ELISA-binding fold change and neutralization fold change for each variant relative to WT.

(C) Top: neutralization IC_{50} of the D614G virus determined as the geometric mean of three biological replicates. Bottom: neutralization results for all mAbs tested, expressed as a fold-change relative to D614G (all variants are in the background of D614G) (Data S1). The individual values of the three replicates are shown as open circles, their geometric mean as colored bars and the geometric SD as error bars. Each antibody is annotated according to its hACE2 competition (as shown in Figure 6F) as well as its epitope (site I, II, or IV) (Data S1). Gray boxes with a slash indicate not tested for hACE2 competition or epitope analysis.

See also Figure S7.

2020a). Consistent with this prediction, VHH-72 showed enhanced potency against N439K SARS-CoV-2 pseudovirus compared to WT N439 (Figures 7A and 7C), highlighting the possibility that a single mutation can impact antibody efficacy positively as well as negatively. Sensitivity of a few neutralizing mAbs to mutations at positions 417 and 439 have also been reported in other studies (Baum et al., 2020; Gaebler et al., 2021; Greaney et al., 2021; Li et al., 2020a; Starr et al., 2020a; Weisblum et al., 2020), although combinations of mutations have typically not been evaluated. Overall, our results demonstrate that mutations compatible with equivalent viral fitness to WT can result in immune evasion from both monoclonal and polyclonal antibody responses.

DISCUSSION

Here, we describe an example of a circulating RBM mutation, N439K, which can evade antibody-mediated immunity without losing fitness relative to WT. The success of variants with the N439K mutation is evidenced by their repeated emergence by convergent evolution on at least nine occasions, spread to 34 countries as of January 2021, significant representation in sampled genome sequences (indicative of high infection rates), the fact that the N439K RBD retains a high-affinity interaction with the hACE2 receptor, and efficient replication of N439K virus in cultured cells. Additionally, we observed no evidence for change in disease severity in a large cohort of individuals infected with N439K virus as compared to WT N439 virus, although we acknowledge some limitations in the data collection, including variations in testing guidelines and availability of testing during the course of the study (da Silva Filipe et al., 2021).

The success of the N439K mutation is consistent with our findings that the RBM is a highly variable region of S. It demonstrates the ability of SARS-CoV-2 to accommodate mutations at the RBM while retaining efficient hACE2 binding. This ability could have emerged by chance or in response to immune pressure from neutralizing Ab responses in viral hosts. There is precedent for the most immunogenic region of a viral surface protein to be highly divergent despite harboring the receptor binding site; for example, the immunogenic globular head domain of the influenza virus hemagglutinin surface protein, which contains the sialic acid receptor binding site, evolves faster than the stalk region (Doud et al., 2018; Kirkpatrick et al., 2018). The ability to readily accommodate mutations in the RBM indicates a high likelihood that potentially immune-evading SARS-CoV-2 variants will continue to emerge, with implications for reinfection,

vaccines, and both monoclonal and polyclonal antibody therapeutics.

A few other circulating RBM mutations have become prominent since N439K first emerged. S477N appeared in the sequence databases in March 2020 but did not become the most prevalent RBD mutation until the summer (as of January 2021, it has >19,000 counts). Consistent with the high prevalence, position 477 is the RBM position where mutations are predicted to be the most well-tolerated with respect to hACE2 binding (Figure 2). Studies across multiple mAb panels have not found this mutation to be conferring resistance (Gaebler et al., 2021; Greaney et al., 2021; Tortorici et al., 2020; Weisblum et al., 2020). In contrast, mutations at position 484 in the RBM have been reported to confer resistance to many mAbs across multiple studies (Baum et al., 2020; Gaebler et al., 2021; Greaney et al., 2021; Tortorici et al., 2020; Weisblum et al., 2020) and also appeared in a persistent (>150 days) infection in an immunocompromised individual who did not respond to treatment with a two-mAb cocktail (Choi et al., 2020). The variant count at position 484 has been steadily increasing (>500 as of January 2021), and the possibility for no fitness consequences for these variants is notable (Figures 2 and S1B) (Starr et al., 2020b), as is a recent study identifying this position as particularly important for escape from polyclonal serum antibodies (Greaney et al., 2021a) and the appearance of this mutation in a new, fast-growing viral lineage (Tegally et al., 2020). The Y453F mutation has become noteworthy recently for its association with virus circulating in mink farms and its transmission back to humans (Oude Munnink et al., 2021) and the DMS measurement indicating it confers significantly increased hACE2 binding (Starr et al., 2020b). To date, we know of only one example of published immune escape documented for Y453F (Baum et al., 2020), but more examples may arise as this new mutation is investigated further. Last, the N501Y mutation has gained notice in the final weeks of 2020 for its association with high rates of infection (Tegally et al., 2020; Volz et al., 2021), although further research is needed to determine the impact on immune escape.

SARS-CoV-2 appears to be evolving relatively slowly consistent with its low mutation rate, the highly susceptible human population, and its generalist nature (Conceicao et al., 2020) and, at present, evidence indicates it will be controllable by vaccines based on early SARS-CoV-2 genome sequences (Dearlove et al., 2020). Additionally, for the majority of our tested sera and mAbs, a single amino acid change in the RBM was not sufficient to confer resistance. Nevertheless, our data indicate that individuals with a mild antibody response to vaccination or first infection could be at risk from a virus carrying a mutation in the RBM. Furthermore, considering the high level of structural plasticity of the RBM demonstrated in the present study, there could be many combinations of RBM mutations, including some requiring compensatory changes, that are compatible with high viral fitness. Some of these combinations will contribute to efficient immune escape. For example, our data show that N439K can compensate for a mutation (K417V) that otherwise decreases receptor binding affinity (Figure 4D) and that several mAbs were more sensitive to these mutations in combination versus individually (Figure 6D; Data S1). This particular combination of mutations is plausibly compatible with maintained viral fitness as it parallels SARS-CoV RBM:hACE2 interactions (salt bridge at SARS-CoV RBD position R426 and no salt bridge at V404) (Figure 4A). Current SARS-CoV-2 mutations have arisen in the absence of pressure from significant population immunity. However, as immunity to the WT virus becomes more widespread, immune escape mutations can be expected to increasingly circulate. In the final weeks of 2020, SARS-CoV-2 variants carrying multiple mutations in the S protein, in both the RBM and Domain A, have been observed (Volz et al., 2021) including one variant carrying three simultaneous RBM mutations (K417N, E484K, and N501Y) (Tegally et al., 2020). Accumulation of multiple changes may increase the risk of immune escape from vaccines that are based on early SARS-CoV-2 sequences.

Mutations in the RBM will also impact the prophylactic or therapeutic use of mAbs. In our profile of immune escape from the N439K variant, we observed resistance to a mAb which is part of a two-mAb cocktail that recently received EAU. The promise of using cocktails of mAbs is that they should significantly lower the likelihood of drug-induced selection of resistant viruses (Baum et al., 2020). However, if circulating viral variants already carry resistant mutations to one antibody in the cocktail, this could cause the cocktail to be reduced to a monotherapy. Additionally, we observed that two mutations together (N439K/K417V) conferred resistance *in vitro* to the two-mAb cocktail (Figure 7C).

Two approaches will be critical for minimizing the impact of mAb escape mutations. One is to develop mAbs with epitopes that are highly resistant to viral escape. This may include epitopes outside of the RBM and/or epitopes that are cross-reactive across SARS-CoV and SARS-CoV-2, indicating conserved epitopes with a low tolerance for mutation (Garrett Rappazzo et al., 2021; Pinto et al., 2020; Wec et al., 2020; Wrapp et al., 2020a). A comparison of epitopes of RBM-targeting mAbs with the most conserved regions of the RBM (Figure 2) may also identify RBM mAbs with a higher barrier to escape. The second will be to screen patients, likely at the population level, for the presence of potential resistance variants prior to drug administration. The availability of multiple different mAb therapeutics in the clinic could provide the opportunity to tailor the choice of therapeutics to local circulating variants.

In general, given that access to therapeutic mAbs is expanding, and as more people develop immune responses to the WT virus via infection or vaccination, monitoring the evolution of SARS-CoV-2 for escape mutants will be critical. Although we only report on evasion of antibody-mediated immunity here, it would be surprising if similar changes are not observed that confer evasion of T cell immunity and innate immunity.

Limitations of study

This study presents the finding that the RBM is a highly variable region of the SARS-CoV-2 S protein, and we provide a thorough characterization of the N439K RBM amino acid replacement, and the ability of this mutation to confer immune evasion without attenuating (or enhancing) fitness or disease. When this study was initiated in June 2020, the general consensus was that the slow rate of evolution of SARS-CoV-2 would result in no immediate threat to vaccines or therapies. N439K was the first RBM amino acid replacement, relative to the ancestral SARS-CoV-2

variant used in vaccine preparations, to increase to high frequency and so can be viewed as a sentinel mutation for SARS-CoV-2 antigenic drift. Since initial submission to Cell in late October 2020, the emergence of multiple highly transmissible variants carrying other RBM mutations of significance has brought the study of immune evasion variants to the forefront of SARS-CoV-2 research. Future retrospective studies will confirm whether these new RBM mutations fall into the same category as N439K: mutations that do not attenuate viral fitness or disease but cause immune evasion. Our results on the plasticity of the RBM also anticipated the emergence of the RBM mutations present in the SARS-CoV-2 variants of concern.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
  - Sample donors
- METHOD DETAILS
  - Structural analysis
  - RBM variability across SARS-CoV-2 sequences
  - Evaluation of deep mutational scanning (DMS) data
  - Molecular dynamics simulations
  - System solvation and parametrization
  - System equilibration
  - Folding@home simulations
  - Simulation analysis
  - RBM variability across Sarbecoviruses
  - Recombinant glycoprotein production
  - Crystallization, data collection, structure determination, and analysis
  - Binding measurements using surface plasmon resonance (SPR)
  - hACE2 binding measurements using bio-layer interferometry (BLI)
  - Epidemiological and genome surveillance
  - Phylogenetic and phylodynamic analysis
  - Evaluation of clinical samples
  - qPCR of clinical samples
  - Viral growth curve
  - Competition assay
  - Ab discovery and recombinant expression
  - Enzyme-linked immunosorbent assay (ELISA)
  - Antibody binding measurements using bio-layer interferometry (BLI)
  - Blockade of RBD binding to ACE2
  - RBD epitope mapping (blockade of binding assay)
  - VSV pseudovirus generation
  - Pseudovirus neutralization
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.01.037.

CONSORTIA

The members of the COG-UK consortia are Thomas R. Connor, Nicholas J. Loman, Samuel C. Robson, Tanya Golubchik, M. Estee Torok, William L. Hamilton, David Bonsall, Ali R. Awan, Sally Corden, Ian Goodfellow, Darren L. Smith, Martin D. Curran, Surendra Parmar, James G. Shepherd, Matthew D. Parker, Catherine Moore, Derek J. Fairley, Matthew W. Loose, Joanne Watkins, Matthew Bull, Sam Nicholls, David M. Aanensen, Sharon Glaysher, Matthew Bashton, Nicole Pacchiarini, Anthony P. Underwood, Thushan I. de Silva, Dennis Wang, Monique Andersson, Anoop J. Chauhan, Mariateresa de Cesare, Catherine Ludden, Tabitha W. Mahungu, Rebecca Dewar, Martin P. McHugh, Natasha G. Jesudason, Kathy K. Li, Rajiv N. Shah, Yusri Taha, Kate E. Templeton, Simon Cottrell, Justin O'Grady, Andrew Rambaut, Colin P. Smith, Matthew T.G. Holden, Emma C. Thomson, Samuel Moses, Meera Chand, Chrystala Constantinidou, Alistair C. Darby, Julian A. Hiscox, Steve Paterson, Meera Unnikrishnan, Andrew J. Page, Erik M. Volz, Charlotte J. Houldcroft, Aminu S. Jahun, James P. McKenna, Luke W. Meredith, Andrew Nelson, Sarojini Pandey, Gregory R. Young, Anna Price, Sara Rey, Sunando Roy, Ben Temperton, Matthew Wyles, Stefan Rooke, Sharif Shaaban, Helen Adams, Yann Bourgeois, Katie F. Loveson, Áine O'Toole, Richard Stark, Ewan M. Harrison, David Heyburn, Sharon J. Peacock, David Buck, Michaela John, Dorota Jamrozy, Joshua Quick, Rahul Batra, Katherine L. Bellis, Beth Blane, Sophia T. Girgis, Angie Green, Anita Justice, Mark Kristiansen, Rachel J. Williams, Radoslaw Poplawski, Garry P. Scarlett, John A. Todd, Christophe Fraser, Judith Breuer, Sergi Castellano, Stephen L. Michell, Dimitris Gramatopoulos, Jonathan Edgeworth, Gemma L. Kay, Ana da Silva Filipe, Aaron R. Jeffries, Sascha Ott, Oliver Pybus, David L. Robertson, David A. Simpson, Chris Williams, Cressida Auckland, John Boyes, Samir Dervisevic, Sian Ellard, Sonia Goncalves, Emma J. Meader, Peter Muir, Husam Osman, Reenesh Prakash, Venkat Sivaprakasam, Ian B. Vipond, Jane A.H. Masoli, Nabil-Fareed Alikhan, Matthew Carlile, Noel Craine, Sam T. Haldenby, Nadine Holmes, Ronan A. Lyons, Christopher Moore, Malorie Perry, Ben Warne, Thomas Williams, Lisa Berry, Andrew Bosworth, Julianne Rose Brown, Sharon Campbell, Anna Casey, Gemma Clark, Jennifer Collins, Alison Cox, Thomas Davis, Gary Eltringham, Cariad Evans, Clive Graham, Fenella Halstead, Kathryn Ann Harris, Christopher Holmes, Stephanie Hutchings, Miren Iturriza-Gomara, Kate Johnson, Katie Jones, Alexander J. Keeley, Bridget A. Knight, Cherian Koshy, Steven Liggett, Hannah Lowe, Anita O. Lucaci, Jessica Lynch, Patrick C McClure, Nathan Moore, Matilde Mori, David G. Partridge, Pinglawathee Madona, Hannah M. Pymont, Paul Anthony Randell, Mohammad Raza, Felicity Ryan, Robert Shaw, Tim J. Sloan, Emma Swindells, Alexander Adams, Hibo Asad, Alec Birchley, Tony Thomas Brooks, Giselda Bucca, Ethan Butcher, Sarah L. Caddy, Laura G. Caller, Yasmin Chaudhry, Jason Coombes, Michelle Cronin, Patricia L. Dyal, Johnathan M. Evans, Laia Fina, Bree Gatica-Wilcox, Iliana Georgana, Lauren Gilbert, Lee Graham, Danielle C. Groves, Grant Hall, Ember Hilvers, Myra Hosmillo, Hannah Jones, Sophie Jones, Fahad A. Khokhar, Sara Kumziene-Summerhayes, George MacIntyre-Cockett, Rocio T. Martinez Nunez, Caoimhe McKerr, Claire McMurray, Richard Myers, Yasmin Nicole Panchbhaya, Malte L. Pinckert, Amy Plimmer, Joanne Stockton, Sarah Taylor, Alicia Thornton, Amy Trebes, Alexander J. Trotter, Helena Jane Tutill, Charlotte A. Williams, Anna Yakovleva, Wen C. Yew, Mohammad T. Alam, Laura Baxter, Olivia Boyd, Fabricia F. Nascimento, Timothy M. Freeman, Lily Geidelberg, Joseph Hughes, David Jorgensen, Benjamin B. Lindsey, Richard J. Orton, Manon Ragonnet-Cronin, Joel Southgate, Sreenu Vattipally, Igor Starinskij, Joshua B. Singer, Khalil Abudahab, Leonardo de Oliveira Martins, Thanh Le-Viet, Mirko Menegazzo, Ben E.W. Taylor, Corin A. Yeats, Sophie Palmer, Carol M. Churcher, Alisha Davies, Elen De Lacy, Fatima Downing, Sue Edward, Nikki Smith, Frances Bolt, Alex Alderton, Matt Berriman, Ian G. Charles, Nicholas Cortes, Tanya Curran, John Danesh, Sahar Eldirdiri, Ngozi Elumogo, Andrew Hattersley, Alison Holmes, Robin Howe, Rachel Jones, Anita Kenyon, Robert A. Kingsley, Dominic Kwiatkowski, Cordelia Langford, Jenifer Mason, Alison E. Mather, Lizzie Meadows, Sian Morgan, James Price, Trevor I. Robinson, Giri

Shankar, John Wain, Mark A. Webber, Declan T. Bradley, Michael R. Chapman, Derrick Crooke, David Eyre, Martyn Guest, Huw Gulliver, Sarah Hoosdally, Christine Kitchen, Ian Merrick, Siddharth Mookerjee, Robert Munn, Timothy Peto, Will Potter, Dheeraj K Sethi, Wendy Smith, Luke B. Snell, Rachael Stanley, Claire Stuart, Elizabeth Wastenge, Erwan Acheson, Safiah Afifi, Elias Allara, Roberto Amato, Adrienn Angyal, Elihu Aranday-Cortes, Cristina Ariani, Jordan Ashworth, Stephen Attwood, Alp Aydin, David J. Baker, Carlos E. Balcazar, Angela Beckett, Robert Beer, Gilberto Betancor, Emma Betteridge, David Bibby, Daniel Bradshaw, Catherine Bresner, Hannah E. Bridgewater, Alice Broos, Rebecca Brown, Paul E. Brown, Kirstyn Brunker, Stephen N. Carmichael, Jeffrey K.J. Cheng, Rachel Colquhoun, Gavin Dabrera, Johnny Debebe, Eleanor Drury, Louis du Plessis, Richard Eccles, Nicholas Ellaby, Audrey Farbos, Ben Farr, Jacqueline Findlay, Chloe L. Fisher, Leysa Marie Forrest, Sarah Francois, Lucy R. Frost, William Fuller, Eileen Gallagher, Michael D. Gallagher, Matthew Gemmell, Rachel A.J. Gilroy, Scott Goodwin, Luke R. Green, Richard Gregory, Natalie Groves, James W. Harrison, Hassan Hartman, Andrew R. Hesketh, Verity Hill, Jonathan Hubb, Margaret Hughes, David K. Jackson, Ben Jackson, Keith James, Natasha Johnson, Ian Johnston, Jon-Paul Keatley, Moritz Kraemer, Angie Lackenby, Mara Lawniczak, David Lee, Rich Livett, Stephanie Lo, Daniel Mair, Joshua Maksimovic, Nikos Manesis, Robin Manley, Carmen Manso, Angela Marchbank, Inigo Martincorena, Tamyo Mbisa, Kathryn McCluggage, J.T. McCrone, Shahjahan Miah, Michelle L. Michelsen, Mari Morgan, Gaia Nebbia, Charlotte Nelson, Jenna Nichols, Paola Niola, Kyriaki Nomikou, Steve Palmer, Naomi Park, Yasmin A. Parr, Paul J. Parsons, Vineet Patel, Minal Patel, Clare Pearson, Steven Platt, Christoph Puethe, Mike Quail, Jayna Raghwani, Lucille Rainbow, Shavanthi Rajatileka, Mary Ramsay, Paola C. Resende Silva, Steven Rudder, Chris Ruis, Christine M. Sambles, Fei Sang, Ulf Schaefer, Emily Scher, Carol Scott, Lesley Shirley, Adrian W. Signell, John Sillitoe, Christen Smith, Katherine L. Smollett, Karla Spellman, Thomas D. Stanton, David J. Studholme, Grace Taylor-Joyce, Ana P. Tedim, Thomas Thompson, Nicholas M. Thomson, Scott Thurston, Lily Tong, Gerry Tonkin-Hill, Rachel M. Tucker, Edith E. Vamos, Tetyana Vasylyeva, Joanna Warwick-Dugdale, Danni Weldon, Mark Whitehead, David Williams, Kathleen A. Williamson, Harry D. Wilson, Trudy Workman, Muhammad Yasir, Xiaoyu Yu, Alex Zarebski, Evelien M. Adriaenssens, Shazaad S.Y. Ahmad, Adela Alcolea-Medina, John Allan, Patawee Asamaphan, Laura Atkinson, Paul Baker, Jonathan Ball, Edward Barton, Mathew A. Beale, Charlotte Beaver, Andrew Beggs, Andrew Bell, Duncan J Berger, Louise Berry, Claire M. Bewshea, Kelly Bicknell, Paul Bird, Chloe Bishop, Tim Boswell, Cassie Breen, Sarah K. Buddenborg, Shirelle Burton-Fanning, Vicki Chalker, Joseph G. Chappell, Themoula Charalampous, Claire Cormie, Nick Cortes, Lindsay J. Coupland, Angela Cowell, Rose K. Davidson, Joana Dias, Maria Diaz, Thomas Dibling, Matthew J. Dorman, Nichola Duckworth, Scott Elliott, Sarah Essex, Karlie Fallon, Theresa Feltwell, Vicki M Fleming, Sally Forrest, Luke Foulser, Maria V. Garcia-Casado, Artemis Gavriil, Ryan P. George, Laura Gifford, Harmeet K. Gill, Jane Greenaway, Luke Griffith, Ana Victoria Gutierrez, Antony D. Hale, Tanzina Haque, Katherine L. Harper, Ian Harrison, Judith Heaney, Thomas Helmer, Ellen E. Higginson, Richard Hopes, Hannah C. Howson-Wells, Adam D. Hunter, Robert Impey, Dianne Irish-Tavares, David A. Jackson, Kathryn A. Jackson, Amelia Joseph, Leanne Kane, Sally Kay, Leanne M. Kermack, Manjinder Khakh, Stephen P. Kidd, Anastasia Kolyva, Jack C.D. Lee, Laura Letchford, Nick Levene, Lisa J. Levett, Michelle M. Lister, Allyson Lloyd, Joshua Loh, Louissa R. Macfarlane-Smith, Nicholas W. Machin, Mailis Maes, Samantha McGuigan, Liz McMinn, Lamia Mestek-Boukhibar, Zoltan Molnar, Lynn Monaghan, Catrin Moore, Plamena Naydenova, Alexandra S. Neaverson, Rachel Nelson, Marc O. Niebel, Elaine O'Toole, Debra Padgett, Gaurang Patel, Brendan A.I. Payne, Liam Prestwood, Veena Raviprakash, Nicola Reynolds, Alex Richter, Esther Robinson, Hazel A. Rogers, Aileen Rowan, Garren Scott, Divya Shah, Nicola Sheriff, Graciela Sluga, Emily Souster, Michael Spencer-Chapman, Sushmita Sridhar, Tracey Swingler, Julian Tang, Graham P. Taylor, Theocharis Tsoleridis, Lance Turtle, Sarah Walsh, Michelle Wantoch, Joanne Watts, Sheila Waugh, Sam Weeks, Rebecca Williams, Iona Willingham, Emma L. Wise, Victoria Wright, Sarah Wyllie, Jamie Young, Amy Gaskin, Will Rowe, Igor Siveroni, and Robert Johnson. See Data S3 for affiliation information. The members of the ISARIC4C consortia are Consortium Lead Investigator: J. Kenneth Baillie; Chief Investigator: Malcolm G. Semple; Co-Lead Investigator: Peter J.M. Openshaw; ISARIC Clinical Coordinator: Gail Carson; Co-Investigators: Beatrice Alex, Benjamin Bach, Wendy S. Barclay, Debby Bogaert, Meera Chand, Graham S. Cooke, Annemarie B. Docherty, Jake Dunning, Ana da Silva Filipe, Tom Fletcher, Christopher A. Green, Ewen M. Harrison, Julian A. Hiscox, Antonia Ying Wai Ho, Peter W. Horby, Samreen Ijaz, Saye Khoo, Paul Klenerman, Andrew Law, Wei Shen Lim, Alexander J. Mentzer, Laura Merson, Alison M. Meynert, Mahdad Noursadeghi, Shona C. Moore, Massimo Palmarini, William A. Paxton, Georgios Pollakis, Nicholas Price, Andrew Rambaut, David L. Robertson, Clark D. Russell, Vanessa Sancho-Shimizu, Janet T. Scott, Thushan de Silva, Louise Sigfrid, Tom Solomon, Shiranee Sriskandan, David Stuart, Charlotte Summers, Richard S. Tedder, Emma C. Thomson, A.A. Roger Thompson, Ryan S. Thwaites, Lance C.W. Turtle, and Maria Zambon; Project Managers: Hayley Hardwick, Chloe Donohue, Ruth Lyons, Fiona Griffiths, and Wilna Oosthuyzen; Data Analysts: Lisa Norman, Riinu Pius, Tom M. Drake, Cameron J. Fairfield, Stephen Knight, Kenneth A. Mclean, Derek Murphy, and Catherine A. Shaw; Data and Information System Managers: Jo Dalton, James Lee, Daniel Plotkin, Michelle Girvan, Egle Saviciute, Stephanie Roberts, Janet Harrison, Laura Marsh, Marie Connor, Sophie Halpin, Clare Jackson, and Carrol Gamble; Data Integration and Presentation: Gary Leeming, Andrew Law, Murray Wham, Sara Clohisey, Ross Hendry, and James Scott-Brown; Material Management: William Greenhalf, Victoria Shaw, and Sarah McDonald; Patient Engagement: Seán Keating; Outbreak Laboratory Staff and Volunteers: Katie A. Ahmed, Jane A. Armstrong, Milton Ashworth, Innocent G. Asiimwe, Siddharth Bakshi, Samantha L. Barlow, Laura Booth, Benjamin Brennan, Katie Bullock, Benjamin W.A. Catterall, Jordan J. Clark, Emily A. Clarke, Sarah Cole, Louise Cooper, Helen Cox, Christopher Davis, Oslem Dincarslan, Chris Dunn, Philip Dyer, Angela Elliott, Anthony Evans, Lorna Finch, Lewis W.S. Fisher, Terry Foster, Isabel Garcia-Dorival, Willliam Greenhalf, Philip Gunning, Catherine Hartley, Antonia Ho, Rebecca L. Jensen, Christopher B. Jones, Trevor R. Jones, Shadia Khandaker, Katharine King, Robyn T. Kiy, Chrysa Koukorava, Annette Lake, Suzannah Lant, Diane Latawiec, L. Lavelle-Langham, Daniella Lefteri, Lauren Lett, Lucia A. Livoti, Maria Mancini, Sarah McDonald, Laurence McEvoy, John McLauchlan, Soeren Metelmann, Nahida S. Miah, Joanna Middleton, Joyce Mitchell, Shona C. Moore, Ellen G. Murphy, Rebekah Penrice-Randal, Jack Pilgrim, Tessa Prince, Will Reynolds, P. Matthew Ridley, Debby Sales, Victoria E. Shaw, Rebecca K. Shears, Benjamin Small, Krishanthi S. Subramaniam, Agnieska Szemiel, Aislynn Taggart, Jolanta Tanianis-Hughes, Jordan Thomas, Erwan Trochu, Libby van Tonder, Eve Wilcock, and J. Eunice Zhang; Local Principal Investigators: Kayode Adeniji, Daniel Agranoff, Ken Agwuh, Dhiraj Ail, Ana Alegria, Brian Angus, Abdul Ashish, Dougal Atkinson, Shahedal Bari, Gavin Barlow, Stella Barnass, Nicholas Barrett, Christopher Bassford, David Baxter, Michael Beadsworth, Jolanta Bernatoniene, John Berridge, Nicola Best, Pieter Bothma, David Brealey, Robin Brittain-Long, Naomi Bulteel, Tom Burden, Andrew Burtenshaw, Vikki Caruth, David Chadwick, Duncan Chambler, Nigel Chee, Jenny Child, Srikanth Chukkambotla, Tom Clark, Paul Collini, Catherine Cosgrove, Jason Cupitt, Maria-Teresa Cutino-Moguel, Paul Dark, Chris Dawson, Samir Dervisevic, Phil Donnison, Sam Douthwaite, Ingrid DuRand, Ahilanadan Dushianthan, Tristan Dyer, Cariad Evans, Chi Eziefula, Chrisopher Fegan, Adam Finn, Duncan Fullerton, Sanjeev Garg, Sanjeev Garg, Atul Garg, Effrossyni Gkrania-Klotsas, Jo Godden, Arthur Goldsmith, Clive Graham, Elaine Hardy, Stuart Hartshorn, Daniel Harvey, Peter Havalda, Daniel B. Hawcutt, Maria Hobrok, Luke Hodgson, Anil Hormis, Michael Jacobs, Susan Jain, Paul Jennings, Agilan Kaliappan, Vidya Kasipandian, Stephen Kegg, Michael Kelsey, Jason Kendall, Caroline Kerrison, Ian Kerslake, Oliver Koch, Gouri Koduri, George Koshy, Shondipon Laha, Steven Laird, Susan Larkin, Tamas Leiner, Patrick Lillie, James Limb, Vanessa Linnett, Jeff Little, Michael MacMahon, Emily MacNaughton, Ravish Mankregod, Huw Masson, Elijah Matovu, Katherine McCullough, Ruth McEwen, Manjula Meda, Gary Mills, Jane Minton, Mariyam Mirfenderesky, Kavya Mohandas, Quen Mok, James Moon, Elinoor Moore, Patrick Morgan, Craig Morris, Katherine Mortimore, Samuel Moses, Mbiye Mpenge, Rohinton Mulla, Michael Murphy, Megan Nagel, Thapas Nagarajan, Mark Nelson, Igor Otahal, Mark Pais, Selva Panchatsharam, Hassan Paraiso, Brij Patel, Natalie Pattison, Justin Pepperell, Mark Peters, Mandeep Phull, Stefania Pintus, Jagtur Singh Pooni, Frank Post, David Price, Rachel Prout, Nikolas Rae, Henrik Reschreiter, Tim Reynolds, Neil Richardson, Mark Roberts, Devender Roberts, Alistair Rose, Guy Rousseau, Brendan Ryan, Taranprit Saluja,

Aarti Shah, Prad Shanmuga, Anil Sharma, Anna Shawcross, Jeremy Sizer, Manu Shankar-Hari, Richard Smith, Catherine Snelson, Nick Spittle, Nikki Staines, Tom Stambach, Richard Stewart, Pradeep Subudhi, Tamas Szakmany, Kate Tatham, Jo Thomas, Chris Thompson, Robert Thompson, Ascanio Tridente, Darell Tupper-Carey, Mary Twagira, Andrew Ustianowski, Nick Vallotton, Lisa Vincent-Smith, Shico Visuvanathan, Alan Vuylsteke, Sam Waddy, Rachel Wake, Andrew Walden, Ingeborg Welters, Tony Whitehouse, Paul Whittaker, Ashley Whittington, Meme Wijesinghe, Martin Williams, Lawrence Wilson, Sarah Wilson, Stephen Winchester, Martin Wiselka, Adam Wolverson, Daniel G. Wooton, Andrew Workman, Bryan Yates, and Peter Young.

ACKNOWLEDGMENTS

We thank all Scottish NHS virology laboratories who provided samples for sequencing and Scott Arkison for HPC maintenance. We thank Chiara Silacci-Fregni from Humabs BioMed, Sandra Jovic, Blanca Fernandez Rodriguez, and Federico Mele, from the Institute for Research in Biomedicine in Bellinzona, and Tatiana Terrot from Ente Ospedaliero Cantonale in Lugano for the help in collecting serum samples. We thank Cindy Ng for help with protein production, Jesse Bloom for helpful comments, and Siro Bianchi for help with the graphical abstract. We gratefully acknowledge the authors originating and submitting laboratories of the sequences from GISAID (https://www.gisaid.org). Our genome sequence acknowledgments can be found in Data S2. Molecular Biology Consortium beamline 4.2.2 of the Advanced Light Source, a DOE Office of Science User Facility under Contract DE-AC02-05CH11231, is supported in part by the ALS-ENABLE program funded by the NIH, National Institute of General Medical Sciences (P30 GM124169-01). The ISARIC WHO CCP-UK study protocol is available at http://isaric4c.net/protocols; study registry https://www.isrctn.com/ISRCTN66726260. This work uses data provided by patients and collected by the NHS as part of their care and support #DataSavesLives. We are grateful to the 2,648 frontline NHS clinical and research staff and volunteer medical students who collected the data in challenging circumstances and the generosity of the participants and their families for their individual contributions in these difficult times. We also acknowledge the support of Jeremy J. Farrar and Nahoko Shindo. We are grateful to Gregory R. Bowman at Washington University in St. Louis, Joseph Coffland at Cauldron Development, Peter K. Eastman at Stanford University for developing and maintaining the Folding@home infrastructure, Amazon Web Services for critical Folding@home infrastructure support, and the many Folding@home volunteers who contributed their computational resources to this project (FAH Project 17311). We are also grateful to Rafal P. Wiewiora at Memorial Sloan Kettering Cancer Center and Redesign Science and Sukrit Singh at Washington University in St. Louis for their guidance on setting up and managing simulations on Folding@home. For funding, we thank the Medical Research Council (MRC) (MC UU 1201412), Wellcome Trust (Collaborators Award 206298/Z/17/Z-ARTIC Network), and Chief Scientist Office (Project COV/EDI/20/11). COG-UK is supported by funding from the MRC part of UK Research & Innovation (UKRI), the NIH Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute. We also thank NIHR (CO-CIN-01), MRC (MC_PC_19059), the NIHR Health Protection Research Unit (HPRU) in Emerging and Zoonotic Infections at University of Liverpool in partnership with Public Health England (PHE) and in collaboration with Liverpool School of Tropical Medicine and the University of Oxford (200907), NIHR HPRU in Respiratory Infections at Imperial College London with PHE (200927), Wellcome Trust and Department for International Development (DID; 215091/Z/18/Z), the Bill and Melinda Gates Foundation (OPP1209135), Liverpool Experimental Cancer Medicine Centre (C18616/A25153), NIHR Biomedical Research Centre at Imperial College London (IS-BRC-1215-20013), EU Platform for European Preparedness Against (Re-)emerging Epidemics (PREPARE; FP7 project 602525), and NIHR Clinical Research Network for providing infrastructure support for this research. P.J.M.O. is supported by an NIHR senior investigator award (201385). The views expressed are those of the authors and not necessarily those of the Department of Health and Social Care, DID, NIHR, MRC, Wellcome Trust, or PHE. J.D.C. acknowledges support from NIH (P30 CA008748, R01 GM121505, and R01 GM132386), NSF (CHI-1904822), and the Sloan Kettering Institute. F.S. is supported by the Henry Krenter Foundation. F.S. and the Institute for Research in Biomedicine are supported by the Helmut Horten Foundation. I.Z. acknowledges support from the Tri-Institutional PhD Program in Computational Biology and Medicine, Vir Biotechnology, the Molecular Sciences Software Institute, and the Sloan Kettering Institute. W.G.G. acknowledges support from Bayer and the Sloan Kettering Institute. T.I.C acknowledges support from Randy Read's Wellcome Trust (209407/Z/17/Z).

AUTHOR CONTRIBUTIONS

Conceived research, E.C.T., R. Spreafico, H.W.V., A.T., D.C., D.L.R., and G.S.; designed experiments, E.C.T., A.T., D.C., and G.S.; donors' recruitment and sample collection for serological analysis and mAbs isolation, A.C., P.F., F.S., C.G., M.G., A. Riva, A.H., M.S., M.G.S., P.J.M.O., and J.K.B.; isolation of mAbs, D.P., K.C., F.Z., M.B., M.P., and E.C.; expression and purification of proteins: J.D., N.C., M.M., and S.J.; crystallization, X-ray structure determination and refinement, J.D., N.C., J. Nix, T.I.C., and G.S.; preparation and analysis of molecular dynamics simulations, I.Z., W.G.G., T.I.C., A.M.H., E.F., and J.D.C.; binding and neutralization assays, L.E.R., J.A.W., L.P., A.D.M., J.B., and S.J.; NGS sequencing and analysis, E.C.T., A.S.F., J.H., V.B.S., K.N., L.T., N. Johnson, D.M., K.S., S.C., and J. Nichols; phylogenetic and epidemiological analysis, S.L., J.H., R.M.C., B.J., A. Rambaut, A.O., S.J.L., and D.L.R.; cross-competition and growth assays of primary isolates, A.S.F., C.D., A.W., and S.J.R.; collected and analyzed clinical data, E.C.T., J.G.S., D.J.P., R. Shah, N. Jesudason, K.L., M.G.S., P.J.M.O., and J.K.B.; performed and analyzed real-time qPCR assays, R. Shah, N. Jesudason, and R.F.J.; analyzed data, E.C.T., L.E.R., J.G.S., R. Spreafico, J.A.W., L.P., D.J.P., A. Rambaut, J. Nix, J.D.I., S.J.L., D.L.R., and G.S.; wrote the manuscript, E.C.T., L.E.R., H.W.V., A.T., D.C., D.L.R., and G.S.; edited the manuscript, R. Spreafico, L.P., A.S.F., J.A.W., D.J.P., J.D., N.C., E.C., J.G.S., I.Z., W.G.G., and J.D.C.; supervised the project, E.C.T., D.L.R., G.S.

DECLARATION OF INTERESTS

L.E.R., R. Spreafico, J.A.W., L.P., J.D., N.C., M.M., A.D.M., J.B., D.P., K.C., F.Z., S.J., M.B., M.P., E.C., J.D.I., H.W.V., A.T., D.C., and G.S. are or were employees of Vir Biotechnology and may hold shares in Vir Biotechnology. C.G. is an external scientific advisor for Humabs BioMed SA. J. Nix and T.I.C. are consultants with Vir Biotechnology. M.G.S. declares interest in Integrum Scientific, Greensboro, NC, outside the scope of this work. J.D.C. is a current member of the Scientific Advisory Board of OpenEye Scientific Software and is a scientific consultant to Foresite Labs. The Chodera laboratory (I.Z., W.G.G., and J.D.C.) receives or has received funding from multiple sources, including the NIH, the National Science Foundation, the Parker Institute for Cancer Immunotherapy, Relay Therapeutics, Entasis Therapeutics, Silicon Therapeutics, EMD Serono (Merck KGaA), AstraZeneca, Vir Biotechnology, XtalPi, the Molecular Sciences Software Institute, the Starr Cancer Consortium, the Open Force Field Consortium, Cycle for Survival, a Louis V. Gerstner Young Investigator Award, and the Sloan Kettering Institute. A complete funding history for the Chodera lab can be found at https://www.choderalab.org/funding. The other authors declare no competing interests.

Received: October 30, 2020
Revised: December 12, 2020
Accepted: January 22, 2021
Published: January 28, 2021

REFERENCES

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

Boni, M.F., Lemey, P., Jiang, X., Lam, T.T., Perry, B.W., Castoe, T.A., Rambaut, A., and Robertson, D.L. (2020). Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat. Microbiol. *5*, 1408–1417.

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the R package brms. R J. *10*, 395–411.

Casalino, L., Gaieb, Z., Goldsmith, J.A., Hjorth, C.K., Dommer, A.C., Harbison, A.M., Fogarty, C.A., Barros, E.P., Taylor, B.C., McLellan, J.S., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent. Sci. *6*, 1722–1734.

Case, D.A., Cerutti, D.S., Cheatham, T.E., III, and Darden, T.A. (2017). AMBER 2017 (University of California, San Francisco). https://ambermd.org/doc12/Amber17.pdf.

Chen, P., Nirula, A., Heller, B., Gottlieb, R.L., Boscia, J., Morris, J., Huhn, G., Cardona, J., Mocherla, B., Stosor, V., et al. (2021). SARS-CoV-2 Neutralizing Antibody LY-CoV555 in Outpatients with Covid-19. N. Engl. J. Med. *384*, 229–237.

Choi, B., Choudhary, M.C., Regan, J., Sparks, J.A., Padera, R.F., Qiu, X., Solomon, I.H., Kuo, H.H., Boucau, J., Bowman, K., et al. (2020). Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. N. Engl. J. Med. *383*, 2291–2293.

Conceicao, C., Thakur, N., Human, S., Kelly, J.T., Logan, L., Bialy, D., Bhat, S., Stevenson-Leggett, P., Zagrajek, A.K., Hollinghurst, P., et al. (2020). The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. PLoS Biol. *18*, e3001016.

Corti, D., Voss, J., Gamblin, S.J., Codoni, G., Macagno, A., Jarrossay, D., Vachieri, S.G., Pinna, D., Minola, A., Vanzetta, F., et al. (2011). A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. Science *333*, 850–856.

COVID-19 Genomics UK (COG-UK) Consortium (2020). An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe *1*, e99–e100.

Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr. D Struct. Biol. *74*, 519–530.

da Silva Filipe, A., Shepherd, J.G., Williams, T., Hughes, J., Aranday-Cortes, E., Asamaphan, P., Ashraf, S., Balcazar, C., Brunker, K., Campbell, A., et al.; COVID-19 Genomics UK (COG-UK) Consortium (2021). Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland Europe into Scotland. Nat. Microbiol. *6*, 112–122.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. J. Chem. Physiol. *98*, 10089.

De Maio, N., Walker, C., Borges, R., Weilguny, L., Slodkowicz, G., and Goldman, N. (2020). Issues with SARS-CoV-2 sequencing data. nCoV-2019 Genomic Epidemiology. https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473.

Dearlove, B., Lewitus, E., Bai, H., Li, Y., Reeves, D.B., Joyce, M.G., Scott, P.T., Amare, M.F., Vasan, S., Michael, N.L., et al. (2020). A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. Proc. Natl. Acad. Sci. USA *117*, 23652–23662.

Doud, M.B., Lee, J.M., and Bloom, J.D. (2018). How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. Nat. Commun. *9*, 1386.

Eastman, P., Swails, J., Chodera, J.D., McGibbon, R.T., Zhao, Y., Beauchamp, K.A., Wang, L.P., Simmonett, A.C., Harrigan, M.P., Stern, C.D., et al. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput. Biol. *13*, e1005659.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Folegatti, P.M., Ewer, K.J., Aley, P.K., Angus, B., Becker, S., Belij-Rammerstorfer, S., Bellamy, D., Bibi, S., Bittaye, M., Clutterbuck, E.A., et al.; Oxford COVID Vaccine Trial Group (2020). Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. Lancet *396*, 467–478.

Gaebler, C., Wang, Z., Lorenzi, J.C.C., Muecksch, F., Finkin, S., Tokuyama, M., Ladinsky, M., Cho, A., Jankovic, M., Schaefer-Babajew, D., et al. (2021). Evolution of Antibody Immunity to SARS-CoV-2. Nature. https://doi.org/10.1038/s41586-021-03207-w.

Garrett Rappazzo, C., Tse, L.V., Kaku, C.I., Wrapp, D., Sakharkar, M., Huang, D., Deveau, L.M., Yockachonis, T.J., Herbert, A.S., Battles, M.B., et al. (2021). Broad and potent activity against SARS-like viruses by an engineered human monoclonal antibody. Science. https://doi.org/10.1126/science.abf4830.

Giroglou, T., Cinatl, J., Jr., Rabenau, H., Drosten, C., Schwalbe, H., Doerr, H.W., and von Laer, D. (2004). Retroviral vectors pseudotyped with severe acute respiratory syndrome coronavirus S protein. J. Virol. *78*, 9007–9015.

Gowers, R.J., Linke, M., Barnoud, J., and Reddy, T.J.E. (2016). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference, 98–105.

Greaney, A.J., Loes, A.N., Crawford, K.H.D., Starr, T.N., Malone, K.D., Helen, Y., Chu, H.Y., and Bloom, J.D. (2021a). Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. Cell Host & Microbe. https://doi.org/10.1016/j.chom.2021.02.003.

Greaney, A.J., Starr, T.N., Gilchuk, P., Zost, S.J., Binshtein, E., Loes, A.N., Hilton, S.K., Huddleston, J., Eguia, R., Crawford, K.H.D., et al. (2021). Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. Cell Host & Microbe. https://doi.org/10.1016/j.chom.2020.11.007.

Grolemund, G., and Wickham, H. (2011). Dates and Times Made Easy with lubridate. J. Stat. Softw. *40.3*, 1–25.

Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul, L.M., Brackney, D.E., Grewal, S., et al. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. *20*, 8.

Hansen, J., Baum, A., Pascal, K.E., Russo, V., Giordano, S., Wloga, E., Fulton, B.O., Yan, Y., Koon, K., Patel, K., et al. (2020). Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. Science *369*, 1010–1014.

Harbison, A.M., Brosnan, L.P., Fenlon, K., and Fadda, E. (2019). Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. Glycobiology *29*, 94–103.

Hou, Y.J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K.H., 3rd, Leist, S.R., Schäfer, A., Nakajima, N., Takahashi, K., et al. (2020). SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science *370*, 1464–1468.

Hu, J., He, C.-L., Gao, Q.-Z., Zhang, G.-J., Cao, X.-X., Long, Q.-X., Deng, H.-J., Huang, L.-Y., Chen, J., Wang, K., et al. (2020). D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity. bioRxiv. https://doi.org/10.1101/2020.06.20.161323.

Jackson, L.A., Anderson, E.J., Rouphael, N.G., Roberts, P.C., Makhene, M., Coler, R.N., McCullough, M.P., Chappell, J.D., Denison, M.R., Stevens, L.J., et al. (2020). An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. N. Engl. J. Med. *383*, 1920–1931.

Jiang, S., Hillyer, C., and Du, L. (2020). Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses. Trends Immunol. *41*, 355–359.

Jones, B.E., Brown-Augsburger, P.L., Corbett, K.S., Westendorf, K., Davies, J., Cujec, T.P., Wiethoff, C.M., Blackbourne, J.L., Heinz, B.A., Foster, D., et al. (2020). LY-CoV555, a rapidly isolated potent neutralizing antibody, provides protection in a non-human primate model of SARS-CoV-2 infection. bioRxiv. https://doi.org/10.1101/2020.09.30.318972.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., and Impey, R.W. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Keech, C., Albert, G., Cho, I., Robertson, A., Reed, P., Neal, S., Plested, J.S., Zhu, M., Cloney-Clark, S., Zhou, H., et al. (2020). Phase 1-2 Trial of a SARS-CoV-2 Recombinant Spike Protein Nanoparticle Vaccine. N. Engl. J. Med. *383*, 2320–2332.

Kirkpatrick, E., Qiu, X., Wilson, P.C., Bahl, J., and Krammer, F. (2018). The influenza virus hemagglutinin head evolves faster than the stalk domain. Sci. Rep. *8*, 10432.

Kirschner, K.N., Yongye, A.B., Tschampel, S.M., González-Outeiriño, J., Daniels, C.R., Foley, B.L., and Woods, R.J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J. Comput. Chem. *29*, 622–655.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., et al. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In Positioning and Power in Academic Publishing: Players, Agents and Agendas, F. Loizides and B. Scmidt, eds. (IOS Press), pp. 87–90.

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

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., and Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature *581*, 215–220.

Leimkuhler, B., and Matthews, C. (2013). Robust and efficient configurational molecular sampling via Langevin dynamics. J. Chem. Phys. *138*, 174102.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, F., Li, W., Farzan, M., and Harrison, S.C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science *309*, 1864–1868.

Li, P., Song, L.F., and Merz, K.M., Jr. (2015). Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. J. Chem. Theory Comput. *11*, 1645–1657.

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

Li, X., Giorgi, E.E., Marichannegowda, M.H., Foley, B., Xiao, C., Kong, X.P., Chen, Y., Gnanakaran, S., Korber, B., and Gao, F. (2020b). Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci. Adv. *6*, eabb9153.

Li, X., Wang, W., Zhao, X., Zai, J., Zhao, Q., Li, Y., and Chaillon, A. (2020c). Transmission dynamics and evolutionary history of 2019-nCoV. J. Med. Virol. *92*, 501–511.

Liu, Y., Gelman, A., and Zheng, T. (2015). Simulation-efficient shortest probability intervals. Stat. Comput. *25*, 809–819.

Lorenzo-Redondo, R., Nam, H.H., Roberts, S.C., Simons, L.M., Jennings, L.J., Qi, C., Achenbach, C.J., Hauser, A.R., Ison, M.G., Hultquist, J.F., and Ozer, E.A. (2020). A clade of SARS-CoV-2 viruses associated with lower viral loads in patient upper airways. EBioMedicine *62*, 103112.

Lycett, S.J., Hughes, J., McHugh, M.P., da Silva Felipe, A., Dewar, R., Lu, L., Doherty, T., Shepherd, A., Inward, R., Rossi, G., et al. (2021). Epidemic waves of COVID-19 in Scotland: a genomic perspective on the impact of the introduction and relaxation of lockdown on SARS-CoV-2. medRxiv. https://doi.org/10.1101/2021.01.08.20248677.

Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. *11*, 3696–3713.

McCarthy, K.R., Rennick, L.J., Nambulli, S., Robinson-McCarthy, L.R., Bain, W.G., Haidar, G., and Duprex, P. (2021). Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science. https://doi.org/10.1126/science.abf6950.

McCoy, A.L., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernández, C.X., Schwantes, C.R., Wang, L.P., Lane, T.J., and Pande, V.S. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys. J. *109*, 1528–1532.

Meredith, L.W., Hamilton, W.L., Warne, B., Houldcroft, C.J., Hosmillo, M., Jahun, A.S., Curran, M.D., Parmar, S., Caller, L.G., Caddy, S.L., et al. (2020). Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. Lancet Infect. Dis. *20*, P1263–P1271.

Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., and Beckstein, O. (2011). MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. *32*, 2319–2327.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534.

Mueller, N.F., Wagner, C., Frazar, C.D., Roychoudhury, P., Lee, J., Moncla, L.H., Pelle, B., Richardson, M., Ryke, E., Xie, H., et al. (2020). Viral genomes reveal patterns of the SARS-CoV-2 outbreak in Washington State. medRxiv. https://doi.org/10.1101/2020.09.30.20204230.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. *67*, 355–367.

Oude Munnink, B.B., Sikkema, R.S., Nieuwenhuijse, D.F., Molenaar, R.J., Munger, E., Molenkamp, R., van der Spek, A., Tolsma, P., Rietveld, A., Brouwer, M., et al. (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science *371*, 172–177.

Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.

Perez, F., and Granger, B.E. (2007). IPython: A System for Interactive Scientific Computing. Comput. Sci. Eng. *9*, 21–29.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82.

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

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

Rambaut, A., Holmes, E.C., O'Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., and Pybus, O.G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. *5*, 1403–1407.

Riblett, A.M., Blomen, V.A., Jae, L.T., Altamura, L.A., Doms, R.W., Brummelkamp, T.R., and Wojcechowskyj, J.A. (2015). A Haploid Genetic Screen Identifies Heparan Sulfate Proteoglycans Supporting Rift Valley Fever Virus Infection. J. Virol. *90*, 1414–1423.

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

Robson, F., Khan, K.S., Le, T.K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P., and Ng, W.L. (2020). Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. Mol. Cell *79*, 710–727.

Rodrigues, J.P.G.L.M., Teixeira, J.M.C., Trellet, M., and Bonvin, A.M.J.J. (2018). pdb-tools: a swiss army knife for molecular structures. F1000Res. *7*, 1961.

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

Shajahan, A., Archer-Hartmann, S., Supekar, N.T., Gleinich, A.S., Heiss, C., and Azadi, P. (2020). Comprehensive characterization of N- and O- glycosylation of SARS-CoV-2 human receptor angiotensin converting enzyme 2. Glycobiology. Published online October 29, 2020. https://doi.org/10.1093/glycob/cwaa101.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. Nature *581*, 221–224.

Shirts, M., and Pande, V.S. (2000). COMPUTING: Screen Savers of the World Unite!. Science *290*, 1903–1904.

Simões, E.A.F., Forleo-Neto, E., Geba, G.P., Kamal, M., Yang, F., Cicirello, H., Houghton, M.R., Rideman, R., Zhao, Q., Benvin, S.L., et al. (2020). Suptavumab for the Prevention of Medically Attended Respiratory Syncytial Virus Infection in Preterm Infants. Clin. Infect. Dis., ciaa951.

Starr, T.N., Greaney, A.J., Addetia, A., Hannon, W.W., Choudhary, M.C., Dingens, A.S., Li, J.Z., and Bloom, J.D. (2020a). Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. bioRxiv. https://science.sciencemag.org/content/early/2021/01/22/science.abf9302. https://doi.org/10.1101/2020.11.30.405472.

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

Takada, A., Robison, C., Goto, H., Sanchez, A., Murti, K.G., Whitt, M.A., and Kawaoka, Y. (1997). A system for functional analysis of Ebola virus glycoprotein. Proc. Natl. Acad. Sci. USA *94*, 14764–14769.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., et al. (2020). 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. https://doi.org/10.1101/2020.12.21.20248640.

Tortorici, M.A., Beltramello, M., Lempp, F.A., Pinto, D., Dang, H.V., Rosen, L.E., McCallum, M., Bowen, J., Minola, A., Jaconi, S., et al. (2020). Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. Science *370*, 950–957.

Volz, E.M., and Frost, S.D.W. (2017). Scalable relaxed clock phylogenetic dating. Virus Evol. *3*, vex025.

Volz, E.M., Hill, V., McCrone, J.T., Price, A., Jorgensen, D., O'Toole, A., Southgate, A., Johnson, R., Jackson, B., Nascimento, F.F., et al. (2021). Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. Cell. https://doi.org/10.1016/j.cell.2020.11.020.

Volz, E., Mishra, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., Hinsley, W.R., Laydon, D.J., Dabrera, G., O'Toole, A., et al. (2021). Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. medRxiv. https://doi.org/10.1101/2020.12.30.20249034.

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

Watanabe, Y., Allen, J.D., Wrapp, D., McLellan, J.S., and Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. Science *369*, 330–333.

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

Weisblum, Y., Schmidt, F., Zhang, F., DaSilva, J., Poston, D., Lorenzi, J.C., Muecksch, F., Rutkowska, M., Hoffmann, H.H., Michailidis, E., et al. (2020). Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. eLife *9*, e61312.

Weissman, D., Alameh, M.G., de Silva, T., Collini, P., Hornsby, H., Brown, R., LaBranche, C.C., Edwards, R.J., Sutherland, L., Santra, S., et al. (2021). D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. Cell Host Microbe *29*, 23–31.e4.

Wrapp, D., De Vlieger, D., Corbett, K.S., Torres, G.M., Wang, N., Van Breedam, W., Roose, K., van Schie, L., Hoffmann, M., Pöhlmann, S., et al.; VIB-CMB COVID-19 Response Team (2020a). Structural Basis for Potent Neutralization of Betacoronaviruses by Single-Domain Camelid Antibodies. Cell *181*, 1004–1015.e15.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020b). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science *367*, 1260–1263.

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., et al. (2020a). A new coronavirus associated with human respiratory disease in China. Nature *579*, 265–269.

Wu, S.L., Mertens, A.N., Crider, Y.S., Nguyen, A., Pokpongkiat, N.N., Djajadi, S., Seth, A., Hsiang, M.S., Colford, J.M., Jr., Reingold, A., et al. (2020b). Substantial underestimation of SARS-CoV-2 infection in the United States. Nat. Commun. *11*, 4507.

Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T.P., Wang, Y., Baum, A., Diehl, W.E., Dauphin, A., Carbone, C., et al. (2020). Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell *183*, 739–751.e8.

Zhang, L., Jackson, C.B., Mou, H., Ojha, A., Peng, H., Quinlan, B.D., Rangarajan, E.S., Pan, A., Vanderheiden, A., Suthar, M.S., et al. (2020). SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. Nature Communications. https://doi.org/10.1038/s41467-020-19808-4.

Zhao, P., Praissman, J.L., Grant, O.C., Cai, Y., Xiao, T., Rosenbalm, K.E., Aoki, K., Kellman, B.P., Bridger, R., Barouch, D.H., et al. (2020). Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. Cell Host Microbe *28*, 586–601.e6.

Zimmerman, M.I., Porter, J.R., Ward, M.D., Singh, S., Vithani, N., Meller, A., Mallimadugula, U.L., Kuhn, C.E., Borowsky, J.H., Wiewiora, R.P., et al. (2020). SARS-CoV-2 Simulations Go Exascale to Capture Spike Opening and Reveal Cryptic Pockets Across the Proteome. bioRxiv. https://doi.org/10.1101/2020.06.27.175430.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
S304, S309 IgG and Fab fragments	Pinto et al., 2020	PDB: 7JX3
S2H13, S2H14, S2A4, S2X35 IgG	Piccoli et al., 2020	PDB: 7JV2, 7JXC, 7JXD, 7JXE
S2E12, S2M11 IgG	Tortorici et al., 2020	PDB: 7K3Q, 7K43
LY-CoV555 IgG	Eli Lilly and Company	N/A
REGN10933, REGN10987 IgG	Hansen et al., 2020	PDB: 6XDG
VHH-72	Wrapp et al., 2020a	PDB: 6WAQ
Panel of Human IgG	This study	N/A
Goat Anti-Human IgG-AP	Southern Biotech	Cat. No. 2040-04; RRID:AB_2795643
Goat F(ab')2 Anti-Mouse IgG(H+L), Human ads-AP	Southern Biotech	Cat. No. 1030-04; RRID:AB_2794293
Anti-Avi-tag Antibody, pAb, Rabbit	GenScript	Cat. No. A00674; RRID:AB_915553
Bacterial strains and pseudotype viruses		
VSV-G-glycoprotein-pseudotyped virus	Kerafast	Cat. No. EH1020-PM
VSV-SARS-CoV-2 S-glycoprotein-pseudotyped virus	This study	N/A
Biological samples		
Serum and plasma of SARS-CoV-2 infected individuals	Piccoli et al., 2020	N/A
Serum from SARS-CoV-2 N439K infected individuals	ISARIC4C https://isaric4c.net/	N/A
Nasopharyngeal swabs from SARS-CoV-2 infected individuals	West of Scotland Specialist Virology Centre	N/A
Sputum from SARS-CoV-2 infected individuals	West of Scotland Specialist Virology Centre	N/A
Chemicals and recombinant proteins		
PEI MAX	Polysciences	Cat. No. POL24765-1
TransIT-Lenti	Mirus	Cat. No. 6600
4-Nitrophenyl phosphate disodium salt hexahydrate (pNPP)	Sigma-Aldrich	Cat. No. N2765-100TAB
Blocker Casein (1%) in PBS	Thermo Fisher Scientific	Cat. No. 37528
Tween 20	Sigma Aldrich	Cat. No. 93773
Bovine Serum Albumine	Sigma	Cat. No. 3059
hACE2, mFc tag	ATUM	N/A
BioLock - Biotin Blocking Solution	IBA GmbH	2-0205-050
PNGase F	New England Biolabs	P0704L
EndoH	New England Biolabs	P0702L
Thrombin	Sigma-Aldrich	T1063-250UN
RBD mouse Fc-tagged	Sino Biological	Cat. No. 40592-V05H
Streptavidin-alkaline phosphatase conjugated	Jackson ImmunoResearch	Cat. No. 016-050-084
Cell lines		
ExpiCHO-S	Thermo Fisher Scientific	Cat. No. A29127
Expi293F	Thermo Fisher Scientific	Cat. No. A14527
Lenti-X 293T cells	Takara	Cat. No. 632180
Vero E6 cells	ATCC	Cat. No. CRL-1586

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Commercial assays, kits, and products		
HiTrap Protein A columns (HiTrap Mab select PrismA)	Cytiva	Cat. No. 17-5498-54P
Strep-Tactin XT Superflow high capacity cartridge	IBA GmbH	Cat. No. 2-4026-001
HisTALON Superflow Cartridges, 5 mL	Takara Bio	Cat. No. 635683
HisTALON Superflow Cartridges, 1 mL	Takara Bio	Cat. No. 635650
Superdex 200 Increase 10/300 GL	Cytiva	Cat. No. 28-9909-44
Superose 6 Increase 10/300 GL	Cytiva	Cat. No. 29-0915-96
StrepTrap HP column, 1 mL	Cytiva	Cat. No. 28-9075-46
ExpiFectamine 293 Transfection Kit	Thermo Fisher Scientific	Cat. No. A14524
ExpiFectamine CHO Transfection Kit	Thermo Fisher Scientific	Cat. No. A29129
ExpiFectamine CHO Transfection Kit	Thermo Fisher Scientific	Cat. No. A29130
HiPrep 26/10 desalting columns	Cytiva	Cat. No. 17-5087-02
CaptureSelect C-tag Affinity Matrix	Thermo Fisher Scientific	Cat. No. 2943072010
Zeba Spin Desalting columns, 7 K MWCO, 5 ml	Thermo Fisher Scientific	Cat. No. 89892
Bio-Glo	Promega	Cat. No. G7940
Biosensor Protein A	FortéBio	Cat. No. 733-2137
MiSeq Reagent v2 500 cycle kit	Illumina	Cat. No. MS-102-2003
NextSeq 500/550 High Output Kit v2.5 (300 Cycles)	Illumina	Cat. No. 20024908
DMEM GlutaMAX	Thermo Fisher Scientific	Cat. No. 10566016
Fetal Bovine Serum (FBS)	Thermo Fisher Scientific	Cat. No. A4766801
DNaseI	Thermo Fisher Scientific	Cat. No. AM2222
Agencourt RNA Clean AMPure XP Beads	Beckman Coulter	Cat. No. A63987
Qubit dsDNA HS Kit	Thermo Fisher Scientific	Cat No. Q32854
SuperScript III	Thermo Fisher Scientific	Cat No. 18080044
NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module	New England Biolabs	Cat. No. E6111L
Kapa HyperPrep kit	Roche	Cat. No. KK8504
Kapa LTP Library Preparation Kit for Illumina Platforms	Roche	Cat. No. KK8232
NEBNext Multiplex Oligos for Illumina 96 Unique Dual Index Primer Pairs	New England Biolabs	Cat. No. E6442S
High Sensitivity D5000 Screentape	Agilent	Cat. No. 5067-5592
NEB Luna Universal Probe One-Step RT-qPCR Kit	New England Biolabs	Cat No. E3006E
2019-nCoV_N1 assay RT-qPCR assay	Integrated DNA Technologies	Cat No. 10006713
HBS-N buffer	Cytiva	Cat. No. BR100369
HBS-EP+ buffer	Cytiva	Cat. No. BR100669
Series S Sensor Chip CM5	Cytiva	Cat. No. BR100530
Series S Sensor Chip C1	Cytiva	Cat. No. BR100535
Mouse antibody capture kit	Cytiva	Cat. No. BR100838
Twin-Strep-Tag Capture Kit	IBA GmbH	Cat. No. 2-4370-000
EZ-Link NHS-PEG Solid-Phase Biotinylation Kit - Mini-Spin Columns	Thermo Fisher Scientific	Cat. No. 21450
Spectraplate-384 with high protein binding	Perkin Elmer	Cat. No. CUSG83093
Nunc-Immuno plates - 96-well plate, MaxiSorp	Sigma-Aldrich Chemie GmbH	Cat. No. M9410-1CS
Deposited data		
SARS-CoV-2 RBD N439K/ hACE2/ S304 Fab/ S309 Fab X-ray structure	This study	PDB: 7L0N

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SARS-CoV-2 RBD/hACE2 Molecular Dynamics trajectories	This study	https://covid.molssi.org//simulations/#foldinghome-simulations-of-the-sars-cov-2-spike-rbd-bound-to-human-ace2
Recombinant DNA		
Human antibody expression vectors (IgG1)	This study	N/A
Plasmids encoding SARS-CoV-2 2P S ectodomain variants (amino acids 14-1211) with C-terminal AviTag-8xHis-C-tag	This study (ATUM Bio)	GenBank: NC_045512.2
Plasmids encoding SARS-CoV-2 RBD WT or variants (amino acids 328-531) with C-terminal thrombin cleavage site-TwinStrep-8xHis-tag	This study	GenBank: NC_045512.2
Plasmid encoding SARS-CoV-2 RBD WT (amino acids 328-531) with N-terminal 'ETGT' and C-terminal GS linker-Strep-8xHis-tag	Piccoli et al., 2020	GenBank: NC_045512.2
Plasmid encoding SARS-CoV-2 RBD N439K (amino acids 328-531) with N-terminal 'ETGT' and C-terminal 8xHis-tag	This study	GenBank: NC_045512.2
Plasmids encoding Sarbecovirus RBDs with C-terminal thrombin cleavage site-TwinStrep-8xHis-tag	This study	See Data S1
Plasmid encoding human ACE2 receptor (amino acids 19-615) plus C-terminal thrombin cleavage site-TwinStrep-10xHis-GGG-tag	Piccoli et al., 2020	UniProt: Q9BYF1
Plasmids encoding SARS-CoV-2 D19 Spike variants	This study	N/A
Software and algorithms		
Prism	GraphPad	Version 8
UNICORN	Cytiva	Versions 7.3 and 7.5
Biacore T200 Evaluation Software	Cytiva	Version 3.1
SDS software	Thermo Fisher Scientific	Version 2.3
Refmac5	Murshudov et al., 2011	Version 5.8.0258
Coot	Emsley et al., 2010	Version 0.9
XDS	Kabsch, 2010	N/A
MOE	Chemical Computing Group	Version 2019.0102
BioEdit	Tom Hall http://bioedit.software.informer.com/7.0	Version 7.0.5.3
R: A Language and Environment for Statistical Computing	R Foundation for Statistical Computing	Version 4.0.3
Skygrowth	https://github.com/mrc-ide/skygrowth	N/A
SPln	Liu et al., 2015	Version 1.1
IQ-TREE 2	Minh et al., 2020	Version 2.0.6
lubridate	https://github.com/tidyverse/lubridate	Version 1.7.4
ape	Paradis and Schliep, 2019	Version 5.3
brms	Bürkner, 2018	Version 2.13.5
drc	https://cran.r-project.org/web/packages/drc/drc.pdf	Version 3.0-1
entropy	https://cran.r-project.org/web/packages/entropy/	Version 1.2.1
RcppRoll	https://cran.r-project.org/web/packages/RcppRoll/index.html	Version 0.3.0

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MinKNOW	Oxford Nanopore technologies	Version 19.12.6
Porechop	https://github.com/rrwick/Porechop	Version 0.2.4
Guppy basecaller	Oxford Nanopore technologies	Version 3.2.10
Nanopolish	https://github.com/jts/nanopolish	Version 0.11.3
trim_galore	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/	Version 0.6.5
BWA	Li, 2013	Version 0.7.5
iVar	Grubaugh et al., 2019	Version 1.2.2
Minimap2	Li, 2018	Version 2.17
Baltic Python library	https://github.com/evogytis/baltic	N/A
Artic sequencing bioinformatic pipeline	Artic network https://artic.network/ncov-2019	N/A
Miniconda	Anaconda http://www.anaconda.com	Anaconda Version 2-2.4.0 Miniconda Version 4.9.0
Folding@home	Shirts and Pande, 2000; Zimmerman et al., 2020	N/A
IPython	Perez and Granger, 2007	Version 7.14.0
Jupyter Notebook	Kluyver et al., 2016	Version 6.1.5
MDAnalysis	Michaud-Agrawal et al., 2011; Gowers et al., 2016	Version 1.0.0
NumPy	https://numpy.org	Version 1.19.1
OpenMM	Eastman et al., 2017	Version 7.4.2
OpenMMTools	https://github.com/choderalab/openmmtools	Version 0.20.0
PyMOL	Schrödinger	Version 2.3.2
ISOLDE	Croll, 2018	Version 1.0.1
ChimeraX	Pettersen et al., 2021	Version 1.0
AmberTools	Case et al., 2017	Version 17.0
pdb-tools	Rodrigues et al., 2018	Version 2.0.5
MDTraj	McGibbon et al., 2015	Version 1.9.4
Pandas	https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf	Version 1.0.5
Custom code, molecular dynamics set up and processing	This paper	https://github.com/choderalab/rbd-ace2-contact-analysis
Custom code, evaluation of clinical samples	This paper	https://github.com/dpascall/SARS-CoV-2-mutation-analysis
Instruments		
ÄKTA Xpress FPLC	Cytiva	N/A
ÄKTA Pure 25	Cytiva	N/A
Synergy H1 Hybrid Multi-Mode plate reader	Biotek	N/A
EL406 washer/dispenser BSL2 M	Biotek	N/A
Biacore T200	Cytiva	N/A
Octet Red96	Pall FortéBio	N/A
7500 Fast Real-Time PCR System	Applied Biosystems	N/A
Illumina MiSeq	Illumina	SY-410-1003
Illumina'sNextSeq550	Illumina	SY-415-1002

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Flowcell R9.4.1	Oxford Nanopore technologies	FLO-MIN106D
Envision multimode plate reader	PerkinElmer	2105

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Gyorgy Snell (gsnell@vir.bio).

Materials availability

Materials generated in this study will be made available on request, but we may require a completed materials transfer agreement.

Data and code availability

Datasets generated during this study are included in the article or are available from the corresponding authors on request. The X-ray structure data and model has been deposited with accession code PDB: 7L0N. The code used to set up, run, and analyze the molecular dynamics simulations is available at: https://github.com/choderalab/rbd-ace2-contact-analysis. Raw and processed molecular dynamics trajectory data are available at the MolSSI COVID-19 Molecular Structure and Therapeutics Hub: https://covid.molssi.org//simulations/#foldinghome-simulations-of-the-sars-cov-2-spike-rbd-bound-to-human-ace2. Code for evaluation of clinical samples is available from GitHub: https://github.com/dpascall/SARS-CoV-2-mutation-analysis.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines

Cell lines were obtained from ATCC (Vero E6) or Thermo Fisher Scientific (Expi293F, ExpiCHO-S). Expi293F and ExpiCHO-S cells were maintained in Expi293 Expression Medium and ExpiCHO- Expression Medium (Thermo Fisher Scientific), respectively.

Sample donors

Samples from 442 SARS-CoV-2 infected individuals were obtained from the Ticino healthcare workers cohort (Switzerland), described previously (Piccoli et al., 2020), and under study protocols approved by the local Institutional Review Board (Canton Ticino Ethics Committee, Switzerland). All donors provided written informed consent for the use of blood and blood components (such as PBMCs, sera or plasma). In the Ticino region of Switzerland and during the time period of collection (February-March 2020) no N439K SARS-CoV-2 isolates were reported.

Samples from six N439K variant infected individuals were obtained from the ISARIC4C consortium (https://isaric4c.net/). Ethical approval was given by the South Central-Oxford C Research Ethics Committee in England (reference 13/SC/0149), and by the Scotland A Research Ethics Committee (reference 20/SS/0028). The study was registered at https://www.isrctn.com/ISRCTN66726260.

Residual nucleic acid extracts derived from the nose-throat swabs of 1918 SARS-CoV-2 positive individuals whose diagnostic samples were submitted to the West of Scotland specialist virology center between 3rd March and 30th June 2020 were sequenced as part of the COG-UK consortium under study protocols approved by the relevant national biorepositories (16/WS/0207NHS and 10/S1402/33) (consortiumcontact@cogconsortium.uk, 2020).

METHOD DETAILS

Structural analysis

RBM residues were determined based on the RBD:hACE2 complex crystal structures 2AJF for SARS-CoV (Li et al., 2005) and 6M0J for SARS-CoV-2 (Lan et al., 2020). The 2AJF structure was obtained from the PDB-REDO server (https://pdb-redo.eu) and was subsequently prepared in the molecular modeling software MOE (v2019.0102, https://www.chemcomp.com) using the structure preparation, protonation and energy minimization steps with default settings. RBD residues within 6.0 Å distance of any hACE2 atoms (determined using MOE) were determined for each of the two copies of the complex in the asymmetric unit, and then were combined to obtain the definition of the RBM used in this work (Figure 2). 6M0J was obtained from the Coronavirus Structural Task Force server (https://github.com/thorn-lab/coronavirus_structural_task_force) and was further refined (using Refmac5 v5.8.0258), manually fitted (using Coot v0.9) and prepared (using MOE, as described above) in multiple iterative cycles. The final structure was analyzed for RBD-hACE2 contact residues with a 6.0 Å cutoff to obtain the RBM (using MOE). The final list of RBM residues (Figure 1C) was arrived at by combining the SARS-CoV and SARS-CoV-2 results.

Using MOE, the pairwise binding energy (the sum of van der Waals, ionic, and hydrogen-bond interactions) between each residue in SARS-CoV-2 RBD and each residue in hACE2, and the total binding energy for all interactions, was determined at cutoff distances 3.0 Å, 3.5 Å, 4.0 Å, 4.5 Å, 5.0 Å, 5.5 Å, 6.0 Å, 6.5 Å and 7.0 Å. The percentage of the total binding energy for each interacting RBD residue was calculated for each distance cutoff and was then averaged over all cutoffs. The resulting values are shown in green in Figure 1C.

RBM variability across SARS-CoV-2 sequences

Using CoV-GLUE-reported variants (http://cov-glue.cvr.gla.ac.uk/, downloaded from GISAID on November 30th 2020, n = 209,239) the Shannon's entropy (natural log units, as implemented in the R package entropy) was computed at each residue of the mature (excluding signal peptide) spike protein. Then, entropy was aggregated by domain, in sliding windows (using the R package RcppRoll), or in bins of randomly sampled residues, as detailed in figure legends. Due to the non-normal distribution of variant frequencies, the median rather than the mean was used as the aggregation metric.

As an alternative to entropy, we also quantitated variability by counting the number of variants passing an increasing threshold of supporting sequences. Unlike entropy, this metric only uses variant frequency for thresholding a digital (presence/absence) variant call, hence it is less affected by sampling/deposition bias.

Evaluation of deep mutational scanning (DMS) data

The DMS dataset was retrieved from Starr et al. (2020b). Variant-level DMS scores were aggregated by residue by taking the minimum (most disruptive variant) or the average score across all variants of that residue, except for the reference amino acid and stop codons. Alternatively, minimum and average scores were computed only across variants that have been observed as naturally occurring. Data were represented as a heatmap annotated with: frequency of variants from CoV-GLUE (at least 1 supporting sequence per 25,000 deposited sequences was required to call a variant); number of countries in which a variant was observed; and percentage of total binding energy computed from an X-ray crystal structure (*cf.* structural analysis methods section).

Molecular dynamics simulations

Structure preparation

The RBD:hACE2 complex was constructed from individual RBD (PDB: 6m0j, Chain E) and hACE2 (PDB: 1r42, Chain A) monomers aligned to the full RBD:hACE2 structure (PDB: 6m0j). The 1r42 structure was used for hACE2 because 1) 1r42 is higher resolution (2.20 Å, whereas 6m0j is 2.45 Å) and 2) the electron density map of 1r42 clearly reveals N-acetylglucosamine (NAG) orientation at each glycosylated asparagine residue, providing a reliable building block on which to construct more complex glycan structures. These complex glycans were constructed at each NAG due to earlier work suggesting their role in mediating RBD:hACE2 binding (Zhao et al., 2020), as it is in the spike proteins' intrinsic RBD dynamics (Casalino et al., 2020).

In order to start from the most reliable structural models, we obtained 6m0j and 1r42 from the Coronavirus Structural Taskforce (CST) database, which contains refined structural models based on careful examination of the electron density. In the RBD of the refined 6m0j structure, amino acid rotamers and peptide bonds were flipped to increase Ramachandran favorability, decrease rotamer outliers, reduce clashes, and improve fit to density. A more detailed summary of the 6m0j refinement details is available at: https://github.com/thorn-lab/coronavirus_structural_task_force/blob/master/pdb/surface_glycoprotein/SARS-CoV-2/6m0j/isolde/notes.txt. The 1r42 refined structure differs from the PDB-deposited structure in that it includes the missing C-terminal domain of hACE2 (copied from the 6m17 PDB structure). A more detailed summary of the 1r42 refinement details is available at: https://github.com/thorn-lab/coronavirus_structural_task_force/blob/master/pdb/human_interaction_partners/ACE2/1r42/isolde/notes.txt.

The resulting RBD and hACE2 monomers were then aligned in PyMOL 2.3.2 (Schrödinger, LLC) to the CST 6m0j structure to create an initial RBD:hACE2 complex. The overall root-mean-square deviation (RMSD) was 0.426 Å and the interface RMSD was 0.405 Å, where RMSD was computed for all atoms and the interface residues were defined as all residues within 4 Å of the other binding partner.

Next, the full glycosylation patterns for hACE2 and RBD glycans were determined from Shajahan et al. (2020) and Watanabe et al. (2020). For the constructed RBD:hACE2 complex, these included sites: N53, N90, N103, N322, N432, N546, and N690 on hACE2 and N343 on the RBD. The glycan structures used for each site (FA2, FA26G1, FA2, FA2, FA2G2, A2, FA2, FA2G2, respectively) correspond to the most stable conformers obtained from multi microsecond MD simulations of cumulative sampling (Harbison et al., 2019). Base NAG residues at the reducing end of each glycan structure were aligned to the corresponding NAG stub in the RBD:hACE2 model in PyMOL 2.3.2 (Schrödinger, LLC) and any resulting clashes were refined in ISOLDE (Croll, 2018). Full details of the glycosylation patterns / structures used and full workflow are available at: https://github.com/choderalab/rbd-ace2-contact-analysis.

System solvation and parametrization

The refined glycosylated RBD:hACE2 complex was prepared for simulation using the AmberTools17 tleap suite (Case et al., 2017). All relevant disulfide bridges were specified as well as covalent connectivity within each glycan structure. The glycosylated protein was parameterized with the Amber ff14SB (Maier et al., 2015) and GLYCAM_06j-1 (Kirschner et al., 2008) force fields. The system was solvated using the TIP3P rigid water model (Jorgensen et al., 1983) in a cubic box with 1.5 nm solvent padding on all sides. The solvated system was then minimally neutralized with 0.15 M NaCl using the Li/Merz ion parameters of monovalent ions for the TIP3P water model (12-6 normal usage set) (Li et al., 2015). Full details and tleap scripts can be found at: https://github.com/choderalab/rbd-ace2-contact-analysis.

System equilibration

The system was energy-minimized with an energy tolerance of 10 kJ $mol^{-1}$and equilibrated using the OpenMM 7.4.2 (Eastman et al., 2017) Langevin integrator for 300 ns in the NPT (p = 1 atm, T = 310 K) ensemble with a timestep of 4.0 femtoseconds, a collision rate of 1.0 picoseconds $^{-1}$, and a constraint tolerance of 1 × 10^{-5}. Hydrogen atom masses were set to 4.0 amu by transferring mass from connected heavy atoms, bonds to hydrogen were constrained, and center of mass motion was not removed. Pressure was controlled by a molecular-scaling Monte Carlo barostat with an update interval of 25 steps. Non-bonded interactions were treated with the Particle Mesh Ewald method (Darden et al., 1993) using a real-space cutoff of 1.0 nm and the OpenMM (Eastman et al., 2017) default relative error tolerance of 0.0005, with grid spacing selected automatically. For improved stability, the structure was then equilibrated using the OpenMMTools 0.20.0 BAOAB Langevin integrator (Leimkuhler and Matthews, 2013) for 10 ns using all of the same simulation parameters described above. This simulation was subsequently packaged to seed for production simulation on Folding@home (Shirts and Pande, 2000; Zimmerman et al., 2020). Default parameters were used unless noted otherwise. Further details of the equilibration protocol are available at: https://github.com/choderalab/rbd-ace2-contact-analysis

Folding@home simulations

The equilibrated structure was then used to initiate parallel distributed MD simulations on Folding@home (Shirts and Pande, 2000; Zimmerman et al., 2020). Simulations were run with OpenMM 7.4.2 (Eastman et al., 2017), Folding@home core22 0.0.13). Production simulations used the same Langevin integrator as the NPT equilibration described above. In total, 2000 independent MD simulations were generated on Folding@home. Conformational snapshots (frames) were stored at an interval of 0.5 ns/frame for subsequent analysis. The resulting final dataset contained 2000 trajectories, 183.8 μs of aggregate simulation time, and 367610 frames. This amount of simulation time corresponds to approximately 13.7 GPU-years on an NVIDIA GeForce GTX 1080Ti. This trajectory dataset with solvent is available at the MolSSI COVID-19 Molecular Structure and Therapeutics Hub: https://covid.molssi.org//simulations/#foldinghome-simulations-of-the-sars-cov-2-spike-rbd-bound-to-human-ace2.

Simulation analysis

The longest 1000 trajectories were chosen for analysis, ranging from 90 ns to 230 ns in length, which represent an aggregate simulation time of 118.7 μs. Each frame in a trajectory was aligned using MDAnalysis (Gowers et al., 2016; Michaud-Agrawal et al., 2011) to the equilibrated structure. This was to ensure no crossing of periodic boundary conditions during calculation of distances between residue pairs. The distances between residue pairs (K417-D30, E484-K31, Q493-E35, Q493-K31, G496-K353, G502-K353, Y449-D38, Y449-Q42, K31-E35) were calculated once every 5 frames (2.5 ns) using MDAnalysis after discarding the first 100 frames (50 ns) of each trajectory to ensure relaxation away from the initial seed conformation. Distance was defined as the minimum distance between sidechain heavy atoms for a given residue pair. Further details of the analysis pipeline are available at: https://github.com/choderalab/rbd-ace2-contact-analysis

RBM variability across Sarbecoviruses

A pairwise comparison of *Sarbecovirus* RBD sequences (see Table S1) to SARS-CoV-2 RBD was performed by calculating percent identity over a window size of 30 amino acids at each RBD position. For the site-specific entropy plot across the RBD alignment of SARS-CoV-2 and 68 related viruses, entropy for each position *I* (H(*I*)) was calculated using Shannon's entropy formula with a natural log as implemented in Bioedit ($H(I) = -\Sigma f(a,I)\ln(f(a,I))$; $f(a,I)$ being the frequency of amino acid *a* at position *I*).

Recombinant glycoprotein production

Prefusion-stabilized SARS-CoV-2 spike protein variants (residues 14-1211), containing the 2P and Furin cleavage site mutations (Walls et al., 2020) with a mu-phosphatase signal peptide and a C-terminal Avi-8xHis-C-tag (ATUM Bio) were expressed in Expi293F cells at 37°C and 8% CO_2. Transfections were performed using the ExpiFectamine 293 Transfection Kit (Thermo Fisher Scientific). Cell culture supernatant was collected after four days and purified over a 5 mL C-tag affinity matrix (Thermo Fisher Scientific). Elution fractions were concentrated and injected on a Superose 6 Increase 10/300 GL column (Cytiva) with 1x PBS pH 7.4 as running buffer.

SARS-CoV-2 RBD WT (with N-terminal signal peptide and C-terminal thrombin cleavage site-TwinStrep-8xHis-tag) and variants were expressed in Expi293F cells at 37°C and 8% CO_2. Transfections were performed using the ExpiFectamine 293 Transfection Kit (Thermo Fisher Scientific). Cell culture supernatant was collected three days after transfection and supplemented with 10x PBS to a final concentration of 2.5x PBS (342.5 mM NaCl, 6.75 mM KCl and 29.75 mM phosphates), or 3.2x for RBD N439R. SARS-CoV-2 RBDs were purified using 1 or 5 mL HisTALON superflow cartridges (Takara Bio) and subsequently buffer exchanged into 1x HBS-N buffer (Cytiva) or PBS using Zeba Spin Desalting or HiPrep 26/10 desalting columns.

RBDs from other sarbecoviruses and SARS-CoV-2 RBD WT (with N-terminal signal peptide and 'ETGT', and C-terminal GS linker-Strep-8xHis-tag) were expressed in Expi293F cells at 37°C and 8% CO_2. Cells were transfected using PEI MAX (Polysciences) at a DNA:PEI ratio of 1:3.75. Transfected cells were supplemented three days after transfection with 3 g/L glucose (Bioconcept) and 5 g/L soy hydrolysate (Sigma-Aldrich Chemie GmbH). Cell culture supernatant (423 mL) was collected seven days after transfection and supplemented with 47 mL 10x binding buffer (1 M Tris-HCl, 1.5 M NaCl, 20 mM EDTA, pH 8.0) and 25 mL BioLock (IBA GmbH) and

incubated on ice for 30 min. Proteins were purified using a 5 mL Strep-Tactin XT Superflow high capacity cartridge (IBA GmbH) followed by buffer exchange to PBS using HiPrep 26/10 desalting columns (Cytiva).

For S binding measurements, recombinant ACE2 (residues 19-615 from Uniprot Q9BYF1 with a C-terminal thrombin cleavage site-TwinStrep-10xHis-GGG-tag, and N-terminal signal peptide) was expressed in Expi293F cells at 37°C and 8% CO_2. Transfections were performed using the ExpiFectamine 293 Transfection Kit (Thermo Fisher Scientific). Cell culture supernatant was collected seven or eight days after transfection and supplemented to a final concentration of 80 mM Tris-HCl pH 8.0, 100 mM NaCl, and then incubated with BioLock (IBA GmbH) solution. After filtration through a 0.22 μm filter, ACE2 was purified using a 1 mL StrepTrap HP column (Cytiva) followed by isolation of the monomeric ACE2 by size exclusion chromatography using a Superdex 200 Increase 10/300 GL column (Cytiva) pre-equilibrated in PBS.

For crystallography, the same hACE2 construct as above was expressed in ExpiCHO-S cells at 37°C and 8% CO_2 with kifunensine added to 10 μM. Transfections were performed using the ExpiFectamine CHO transfection kit (Thermo Fisher Scientific). Cell culture supernatant was collected six days after transfection and supplemented to a final concentration of 80 mM Tris-HCl pH 8.0, 100 mM NaCl, and then incubated with BioLock (IBA GmbH) solution for one hour. hACE2 was purified using a 1 mL StrepTrap HP column (Cytiva).

For SPR binding measurements with surface-captured RBD, recombinant hACE2 (residues 19-615 from Uniprot Q9BYF1 with a C-terminal AviTag-10xHis-GGG-tag, and N-terminal signal peptide) was expressed in HEK293.sus using standard methods (ATUM Bio). Protein was purified via Ni Sepharose resin followed by isolation of the monomeric ACE2 by size exclusion chromatography using a Superdex 200 Increase 10/300 GL column pre-equilibrated with PBS.

For binding measurements with surface-captured hACE2, recombinant hACE2 (residues 18-615 with a C-terminal GS-IgG2a-Mm-Fc tag, and N-terminal signal peptide) was stably transfected in CHO-K1 GS knock-down cell line (ATUM Bio). Protein was purified via protein A and buffer exchanged into PBS.

Crystallization, data collection, structure determination, and analysis

The SARS-CoV-2 RBD N439K-hACE2 complex was formed together with two Fab fragments (S304 and S309) to aid in crystallization. Prior to forming the SARS-CoV-2 RBD N439K-ACE2-S304-S309 complex, recombinant hACE2 protein was digested using EndoH (New England Biolabs) and thrombin (Sigma-Aldrich). Recombinant SARS-CoV-2 RBD N439K was digested with PNGase F (New England Biolabs) and thrombin (Sigma-Aldrich). RBD was mixed with a 1.3-fold molar excess of deglycosylated hACE2, S304 Fab, and S309 Fab. The complex was purified on a Superdex 200 10/300 GL column pre-equilibrated with 20 mM Tris-HCl pH 7.5, 150 mM NaCl. Crystals of the SARS-CoV-2 RBD N439K-hACE2-S304-S309 complex were obtained at 20°C by sitting drop vapor diffusion. A total of 200 nL of the complex at 6 mg/mL were mixed with 200 nL mother liquor solution containing 0.1 M ammonium sulfate, 20% v/v ethylene glycol, 10% w/v PEG 8000, and 0.1 M bicine/tris pH 8.5.

Data were collected at the Molecular Biology Consortium beamline 4.2.2 at the Advanced Light Source synchrotron facility in Berkeley, CA. Datasets from two crystals were individually processed and then merged with the XDS software package (Kabsch, 2010) for a final dataset of 2.78 Å in space group $P2_1$. The RBD N439K-hACE2-S304-S309 complex structure was solved by molecular replacement using phaser (McCoy et al., 2007) from starting models consisting of RBD-S304-S309 (PDB: 7JX3) and hACE2 (PDB: 6m0j). Several subsequent rounds of model building and refinement were performed using Coot (Emsley et al., 2010), ISOLDE (Croll, 2018), Refmac5 (Murshudov et al., 2011), and MOE (https://www.chemcomp.com), to arrive at a final model for the quarternary complex.

Binding measurements using surface plasmon resonance (SPR)

SPR binding measurements were performed using a Biacore T200 instrument. S protein was surface captured via anti-AviTag pAb (Genscript) covalently immobilized on a CM5 chip, RBD protein was surface captured via StrepTactin XT (Twin-Strep-Tag Capture Kit, IBA GmbH) covalently immobilized on a CM5 chip, and ACE2-mFc was surface captured via covalent immobilization of the Cytiva Mouse antibody capture kit on a C1 chip. Running buffer was HBS-EP+ pH 7.4 (Cytiva) and all measurements were performed at 25°C. All experiments were performed as single-cycle kinetics, with a 3-fold dilution series of monomeric hACE2 starting from 300 nM, each concentration injected for 180 s, or a 3-fold dilution series of RBD starting from 50 nM, each concentration injected for 240 s. All data were double reference-subtracted and fit to a binding model using Biacore Evaluation software. For one representative replicate, capture levels were normalized to WT for visualization. Binding data with hACE2 as analyte were fit to a 1:1 binding model. Binding data with RBD as analyte were fit to a Heterogeneous Ligand binding model, due to an artifactual kinetic phase with very slow dissociation that arises when RBD is an analyte; the lower affinity of the two K_Ds reported by the fit is reported as the K_D of the RBD-ACE2 interaction (the two reported K_Ds are separated by at least two orders of magnitude for all fits). The measured K_D for hACE2 binding to S is likely influenced by conformational dynamics of the RBDs in the context of the prefusion S trimer. Reported K_Ds are an average of 3-4 replicates measured on at least two separate days, with error given as SEM.

Differences between the SPR assay and the published DMS binding assay (Starr et al., 2020b) include using targeted measurements of purified proteins expressed in mammalian cells versus yeast surface display, as well as the use of dimeric hACE2 in the DMS experiment, which incorporates avidity effects into the RBD-hACE2 binding measurements that can mask modest changes in binding affinity.

hACE2 binding measurements using bio-layer interferometry (BLI)

Binding measurements of Sarbecovirus RBDs to hACE2 were performed by diluting RBDs to 8 μg/ml in kinetic buffer (PBS supplemented with 0.05% BSA) and immobilization on Anti-Penta-His Biosensors of an Octet RED96 system (FortéBio). RBD-coated biosensors were incubated for 5 min with a solution containing 5, 1 or 0.2 μg/ml of hACE2. A dissociation step was then performed by incubating the biosensors for 10 min in kinetic buffer. The change in material bound to the biosensors caused a shift in the interference pattern that was recorded in real time and plotted using GraphPad Prism 8 software.

Epidemiological and genome surveillance

A national sequencing collaboration formed at the start of the epidemic in the UK, CoG-UK consortium (COVID-19 Genomics UK (COG-UK) Consortium, 2020) has facilitated the tracking of SARS-CoV-2 sequences across Scotland since the start of the outbreak in February 2020 and real-time monitoring of genetic changes in the Spike gene that might be associated with changes in virulence or transmissibility. Sequencing was carried out using an amplicon-based protocol in real-time at a rate of up to 300 genomes per week. 50% of samples were selected as surveillance samples, representing Scottish health boards proportionately based on population size, while 50% were selected to allow intervention with local issues such as nosocomial infection in hospitals and nursing homes. The N439K mutation was noted to become increasingly prevalent during April 2020. This was noted to be particularly common in the Greater Glasgow & Clyde NHS health board region but spread to adjacent Scottish health boards also.

Sequencing libraries were prepared according to the ARTIC nCoV-2019 described in detail at https://artic.network/ncov-2019. Briefly, PCR amplicons were generated using the nCoV-2019 PrimalSeq sequencing primers using 25-35 cycles of amplification. Generated amplicons were used to prepare either Oxford Nanopore or Illumina sequencing libraries. Oxford Nanopore libraries were prepared as described in the link above and sequenced in a flow cell R9.4.1 (Oxford Nanopore Technologies, Part Number FLO-MIN106D), using MinKNOW version 19.12.6. Raw FAST5 files were basecalled using Guppy version 3.2.10 in high accuracy mode with a minimum quality score of 7. Reads were size filtered, demultiplexed and trimmed with Porechop (https://github.com/rrwick/Porechop), and mapped against reference strain Wuhan-Hu-1 (MN908947). Variants were called using Nanopolish 0.11.3 and accepted if they had a log-likelihood score of greater than 200 and minimum read coverage of 20. For Illumina sequencing, amplicons were used to prepare libraries using the Kapa HyperPrep kit (Roche, Part Number KK8504) and further processed as described in the competition assay sequencing method. Sequencing was carried out on Illumina's MiSeq system (Illumina, Part Number SY-410-1003) using a MiSeq Reagent v2 500 cycle kit (Illumina, Part Number MS-102-2003). Reads were trimmed with trim_galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and mapped with BWA (Li and Durbin, 2009) to the Wuhan-Hu-1 (MN908947) reference sequence, followed by primer trimming and consensus calling with iVar (Grubaugh et al., 2019) and a minimum read coverage of 10.

Phylogenetic and phylodynamic analysis

UK sequences were obtained from the COG-UK consortium (https://www.cogconsortium.uk) and global sequences from the GISAID Initiative (https://www.gisaid.org) on November 23 2020. The sequences were mapped using minimap2 and padded against the Wuhan/WH04/2020 reference. The sequences were downsampled with weights that normalize sequence count per epiweek, maximize the number of countries and lineages represented, and enriching for sequences with the N439K mutation. A maximum-likelihood phylogenetic tree was constructed using IQ-TREE with the the following parameters: -czb -blmin 0.0000000001 -m HKY–runs 5 and all other parameters set to default. The tree was visualized with custom python code using the baltic library (https://github.com/evogytis/baltic).

For the phylodynamic analysis, Scottish "introduction" lineages with ten or more sequences were identified (Lycett et al., 2021), and the skygrowth package in R was used to estimate the effective population size over time (using up to ten time intervals), and the growth rates of the lineages within Scotland (Volz and Frost, 2017). The data used for analysis were sampled between Feb 28, 2020 and Aug 18, 2020. Growth estimates were calculated for the intervals in between the time points that Ne is estimated over, from the TMRCA onward. Lineages with less than ten sequences in total, less than 50% and/or less than five Scottish sequences were excluded.

Evaluation of clinical samples

Clinical samples submitted to the West of Scotland Specialist Virology Centre for SARS-CoV-2 diagnostic rt-PCR testing were selected for sequencing as part of the COVID-19 UK Genomics UK Consortium (COG-UK) project, resulting in 1918 whole genome sequences originating from the NHS Greater Glasgow and Clyde Health Board region. Sequences were linked to electronic patient records and basic metadata including sample date, age, sex, admission to hospital and mortality at 28 days post diagnosis extracted. The electronic patient records of a subset of 1591 patients underwent full case-note review and clinical severity was recorded based on a 4-level ordinal scale: 1. no requirement for respiratory support, 2. treatment with supplemental oxygen via facemask or low-flow nasal cannulae, 3. intubation and ventilation, non-invasive ventilation or oxygen delivery by high flow nasal cannulae devices, 4. death within the 28 days following diagnosis. We modified the WHO ordinal scale to these 4 points as described previously (Volz et al., 2021) to avoid using hospitalisation as a criterion of severity because 1) many patients in nursing homes had severe infection but were not admitted to hospital, and 2) early in the outbreak, all cases were hospitalised irrespective of the severity of their infection.

These data had previously been analyzed to test for an effect of the D614G mutation on the severity of disease (Volz et al., 2021); we extend that analysis here using the same methodology to test for an effect of the N439K mutation. Additionally, we perform a new analysis using a model with the same structure to test for an effect of both the D614G mutation and the D614G/N439K mutation combination on the viral load of infected patients, as measured by cycle threshold (Ct) value. Ct values were generated in different locations, on different platforms and then collected centrally.

In both cases we cannot estimate the marginal effect of the N439K mutation, as we only have the mutation on the 614G genetic background, so the individual effect of N439K cannot be separated from any potential epistatic interactions between the mutations.

Briefly, the structure of the model used previously (Volz et al., 2021) and in the present study is a phylogenetic generalized additive model with mutation being the primary predictor of interest. The model controls for biological sex, age and the number of days since the first reported case in the dataset, with the latter two being included as penalised splines with a maximum of 30 knots. If the patient was part of a cluster of cases, this was included as a random effect, with individuals not part of clusters being assigned their own levels. Correlations driven by the rest of the genome are controlled for by a phylogenetic random effect using a correlation matrix generated under a Brownian motion assumption from a phylogeny estimated in IQ-TREE 2 v. 2.0.6 (Minh et al., 2020) using a HKY + Γ model, masking the positions recommended by (De Maio et al. 2020) as of 22/7/2020 (https://virological.org/t/issues-with-sars-cov-2-sequencing-data/473/13), rooted on the first sequenced SARS-CoV-2 genome (Wu et al., 2020a). The priors for the severity model were those used in the previous analysis of this data. The priors for the model of the viral load were a student-t (mean = 20, scale = 10, degrees of freedom = 3) prior on the model intercept, a Gaussian (mean = 0, standard deviation = 10) prior over the fixed effects, and an exponential (lambda = 0.1) prior over the random effect, penalised spline and residual standard deviations.

There are two key structural differences between the model used previously (Volz et al., 2021) and the model used here. First, mutation is a three level rather than two level factor (D614/N439, D614G/N439 and D614G/N439K) with the ancestral D614/N439 being the reference level. Second, as we are now interested in two mutations, we estimated the phylogeny used to control for the effect of the rest of the genome excluding both the nucleotide position underlying the D614G mutation and the nucleotide position underlying the N439K mutation (in addition to the sites from De Maio et al. mentioned above).

The severity model used a cumulative error structure while the model on the CT values used a Gaussian error structure. In both cases, the models were estimated in brms v. 2.13.5 (Bürkner, 2018) . The presented models had no divergent transitions, Rhat values less than 1.01, and appropriate bulk and tail effective sample sizes for all parameters. Shortest probability intervals were calculated using the R package SPIn v. 1.1 (Liu et al., 2015). Trees were manipulated using ape v. 5.3 (Paradis and Schliep, 2019), and dates were manipulated using lubridate v. 1.7.4 (Grolemund and Wickham, 2011). Analysis code is available at https://github.com/dpascall/SARS-CoV-2-mutation-analysis.

qPCR of clinical samples

All samples were tested in duplicate using the 2019-nCoV_N1 assay RT-qPCR assay (https://www.fda.gov/media/134922/download). Ready-mixed primers and probe were obtained from IDT (Leuven, Belgium). PCR was carried out using NEB Luna Universal Probe One-Step RT-qPCR Kit (New England Biolabs, Herts, UK), primers and probe at 500 nM and 127.5 nM, respectively, and 5 μL of RNA sample in a final volume of 20 μL. No template negative controls were included after every seventh sample. Six ten-fold dilutions of SARS-CoV-2 RNA standards were tested in duplicate in each assay; standards were calibrated using a plasmid containing the N sequence that had been quantified using droplet digital PCR. Thermal cycling was performed on an Applied Biosystems 7500 Fast PCR instrument running SDS software v2.3 (Thermo Fisher Scientific) under the following conditions: 55°C for 10 minutes and 95°C for 1 minute followed by 45 cycles of 95°C for 10 s and 58°C for 1 minute. Assays were repeated if the reaction efficiency was < 90% or the R2 value of the standard curve was $\leq$ 0.998. Where possible, testing of samples was repeated if the %CV of the duplicates was < 10%.

Viral growth curve

Vero E6-hACE2 cells (Vero E6 cells induced to overexpress hACE2) either with or without TMPRSS2 overexpression (S.J.R., S.B., and A.W., unpublished data) were seeded in a 12-well plate and inoculated with an MOI of 0.01 with either the GLA1 (N439/D614G) or GLA2 (N439K/D614G) virus isolates for 1 hr before washing the cells three times in PBS and replacing with DMEM supplemented with 2% FBS. 100 uL of media was removed at each time point, RNA was extracted using the RNAdvance Blood kit (Beckman Coulter), and the presence of SARS-CoV-2 determined using 2019-nCOV-N1 assays (IDT) with an NEB Luna Universal Probe One-Step RT-qPCR Kit. A standard curve was used to determine the copy number present per mL of cell culture media. 100 uL of the fresh media was also tested for the presence of virus, which was undetectable in all wells. Experiment was performed in triplicate, each with an independent preparation of virus inoculum.

Competition assay

Three T25 flasks were seeded with Vero E6-hACE2 or Vero E6-hACE2-TMPRSS2 and inoculated with either single viruses or both GLA1 and GLA2 virus strains at an MOI of 0.01 for 1 hr. The flasks were washed three times with PBS, with 100 uL of the final wash being retained to determine the presence of free virus, before adding 5 mL of fresh DMEM, supplemented with 2% FBS. At 24, 48, and 72 hr, 500 uL of media was removed, which was replaced with 500 uL fresh media. 300 uL was used for RNA extraction using the RNAdvance Blood kit (Beckman Coulter) and NGS analysis of the frequencies of the specific positions within the spike protein. The single virus inoculations showed no alternations in the frequency of the amino acid positions and the final wash showing no

free virus in the supernatant. We used an unbiased metagenomic NGS sequencing pipeline to quantify variation across the whole viral genome on the Illumina NGS Next Seq platform. Briefly, extracted nucleic acid was incubated with DNaseI (Thermo Fisher Scientific, Part Number AM2222), cleaned with RNA Ampure beads (Agencourt RNA Clean AMPure XP Beads, Beckman Coulter, Cat. No. A63987) followed by cDNA synthesis using SuperScript III (Thermo Fisher Scientific, Part Number 18080044) and NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module (New England Biolabs, Part Number E6111L). Samples were further processed using the Kapa LTP Library Preparation Kit for Illumina Platforms (Roche, Part Number KK8232) and indexed with the NEBNext Multiplex Oligos for Illumina 96 Unique Dual Index Primer Pairs (New England Biolabs, Part Number E6442S). Libraries were characterized utilizing using the Qubit dsDNA HS Kit (Thermo Fisher Scientific, Cat No. Q32854) and the sequenced on Illumina's NextSeq 550 System (Illumina, Part Number SY-415-1002), using a NextSeq 500/550 High Output Kit v2.5 (300 Cycles) (Illumina, Cat. No. 20024908) generating aproximately 10 million pairs of reads per sample. Experiment was performed in triplicate, each with an independent preparation of virus inoculum.

Ab discovery and recombinant expression

Human mAbs were isolated from plasma cells or memory B cells of SARS-CoV or SARS-CoV-2 immune donors, as previously described (Corti et al., 2011; Pinto et al., 2020; Tortorici et al., 2020). Recombinant antibodies were expressed in ExpiCHO cells at 37°C and 8% CO_2. Cells were transfected using ExpiFectamine. Transfected cells were supplemented 1 day after transfection with ExpiCHO Feed and ExpiFectamine CHO Enhancer. Cell culture supernatant was collected eight days after transfection and filtered through a 0.2 μm filter. Recombinant antibodies were affinity purified on an ÄKTA xpress FPLC device using 5 mL HiTrap MabSelect PrismA columns followed by buffer exchange to Histidine buffer (20 mM Histidine, 8% sucrose, pH 6) using HiPrep 26/10 desalting columns.

Enzyme-linked immunosorbent assay (ELISA)

A total of 144 human monoclonal antibodies or 442 human sera were tested for binding to RBD WT and mutants, as previously described (Piccoli et al., 2020). Spectraplate-384 plates with high protein binding treatment (custom made from Perkin Elmer) were coated overnight at 4 °C with 0.1 μg/mL (for mAbs) or 5 ug/mL (for sera) SARS-CoV-2 RBD WT, N439K, K417V or N439K/K417V in phosphate-buffered saline (PBS), pH 7.2. Plates were subsequently blocked with Blocker Casein 1% supplemented with 0.05% Tween 20 (Sigma-Aldrich) for 1 h at room temperature. The coated plates were incubated with serial dilutions of the monoclonal antibodies or of the sera for 1 h at room temperature. The plates were then washed with PBS containing 0.05% Tween-20 (PBS-T), and alkaline phosphatase-goat anti-human IgG (Southern Biotech) was added and incubated for 1 h at room temperature. After 3 washing steps with PBS-T, p-NitroPhenyl Phosphate (pNPP, Sigma-Aldrich) substrate was added and incubated for 30 min at room temperature. The absorbance of 405 nm was measured by a microplate reader (Synergy H1 Hybrid Multi-Mode plate reader, Biotek). For mAbs, fitting was performed using the drc R package with a 4-parameter logistic (4PL) model, yielding dose-response curves from which the area under the curve (AUC) between 5 and 500 ng/mL was computed. The AUC allows to capture, in a single metric, shifts in two parameters of the 4PL model: EC_{50} and upper asymptote. For sera, fitting was performed using GraphPrism 8 with a 4PL model from which ED_{50} was calculated. A cutoff of 30 for serum ED_{50} was set based on previously published data (Piccoli et al., 2020). A difference in reactivity to RBD mutants compared to WT was considered when a minimum 2-fold-variation in the AUC or ED_{50} WT/mutant ratios was observed in two independent experiments (for mAbs, fold variation as low as 1.7 was accepted provided that the average of two experiments was greater or equal to 2-fold).

Antibody binding measurements using bio-layer interferometry (BLI)

BLI binding measurement was performed on a selection of human monoclonal antibodies tested by ELISA (16/144). Antibodies were diluted to 2.7 μg/mL in kinetic buffer (PBS supplemented with 0.05% BSA) and immobilized on Protein A Biosensors of an Octet RED96 system (FortéBio). Antibody-coated biosensors were incubated for 5 min with a solution containing 5 μg /mL of SARS-CoV2 RBD WT, N439K, K417V or N439/K417V in kinetic buffer. A dissociation step was then performed by incubating the biosensors for 5 min in kinetic buffer. Change in molecules bound to the biosensors caused a shift in the interference pattern that was recorded in real time and plotted using GraphPad Prism 8 software.

Blockade of RBD binding to ACE2

Blockade of WT RBD binding to hACE2 was performed, as previously described (Piccoli et al., 2020). Unlabeled mAbs were serially diluted, mixed with RBD mouse Fc-tagged antigen (Sino Biological, final concentration 20 ng/mL) and incubated for 30 min at 37°C. The mix was added for 30 min to ELISA 96-well plates (Corning) pre-coated overnight at 4°C with 2 μg/mL hACE2 in PBS. Plates were washed (EL406 washer/dispenser BSL2 M, Biotek) and RBD binding was revealed using a secondary goat anti-mouse IgG (Southern Biotech). After washing, pNPP substrate was added and plates were read at 405 nm (Synergy H1 Hybrid Multi-Mode plate reader, Biotek). The percentage of inhibition was calculated as follow: (1−(OD sample−OD neg ctr)/(OD pos ctr−OD neg ctr)) × 100.

RBD epitope mapping (blockade of binding assay)

RBD epitope mapping of the 144 mAbs was performed through blockade of binding (BOB) assay as previously described (Piccoli et al., 2020). Human mAbs binding to RBD site I (S2H14), site II (S304) and site IV (S309) were biotinylated using the EZ-Link

NHS-PEO solid phase biotinylation kit (Pierce). Labeled mAbs were tested for binding to RBD by ELISA and the optimal concentration of each mAb to achieve 80% maximal binding was determined. Unlabeled mAbs were serially diluted and added to ELISA 96-well plates (Corning) pre-coated overnight at 4°C with 1 μg/mL of RBD mouse Fc-tagged antigen (Sino Biological) in PBS. After 30 min, biotinylated anti-RBD mAbs were added at the concentration achieving 80% maximal binding and the mixture was incubated at room temperature for 20 min. Plates were washed (EL406 washer/dispenser BSL2 M, Biotek) and antibody binding was revealed using alkaline phosphatase-conjugated streptavidin (Jackson ImmunoResearch). After washing, pNPP substrate (Sigma-Aldrich) was added and plates were read at 405 nm (Synergy H1 Hybrid Multi-Mode plate reader, Biotek). The percentage of inhibition was calculated as follow: (1−(OD sample−OD neg ctr)/ (OD pos ctr−OD neg ctr)) × 100.

VSV pseudovirus generation

Replication defective VSV pseudovirus (Takada et al., 1997) expressing SARS-CoV-2 spike protein were generated as previously described (Riblett et al., 2015) with some modifications. Plasmids encoding SARS-CoV-2 spike variants were generated by site-directed mutagenesis of the wild-type plasmid, pcDNA3.1(+)-spike-D19 (Giroglou et al., 2004). Lenti-X 293T cells (Takara, 632180) were seeded in 10-cm dishes at a density of 1e5 cells/cm^2 and the following day transfected with 5 μg of spike expression plasmid with TransIT-Lenti (Mirus, 6600) according to the manufacturer's instructions. One day post-transfection, cells were infected with VSV-luc (VSV-G) (Kerafast, EH1020-PM) for 1 h, rinsed three times with PBS, then incubated for an additional 24 h in complete media at 37°C. The cell supernatant was clarified by centrifugation, filtered (0.45 μm), aliquoted, and frozen at −80°C.

Pseudovirus neutralization

Vero E6 cells (ATCC CRL-1586) were grown in DMEM supplemented with 10% FBS and seeded into clear bottom white 96 well plates (Costar, 3903) at a density of 2e4 cells per well. The next day, mAbs were serially diluted in pre-warmed complete media, mixed at a 1:1 ratio with pseudovirus and incubated for 1 h at 37°C in round bottom polypropylene plates. Media from cells was aspirated and 50 μL of virus-mAb complexes were added to cells and then incubated for 1 h at 37°C. An additional 100 μL of prewarmed complete media was then added on top of complexes and cells incubated for an additional 16-24 h. Conditions were tested in duplicate wells on each plate and at least six wells per plate contained uninfected, untreated cells (mock) and infected, untreated cells ('no mAb control'). Virus-mAb-containing media was then aspirated from cells and 100 uL of a 1:4 dilution of Bio-glo (Promega, G7940) in PBS was added to cells. Plates were incubated for 10 min at room temperature and then were analyzed on the Envision plate reader (PerkinElmer). Relative light units (RLUs) for infected wells were subtracted by the average of RLU values for the mock wells (background subtraction) and then normalized to the average of background subtracted "no mAb control" RLU values within each plate. Percent neutralization was calculated by subtracting from 1 the normalized mAb infection condition. Data were analyzed and visualized with Prism (Version 8.4.3). IC_{50} and IC80 values were calculated from the interpolated value from the log(inhibitor) versus response – variable slope (four parameters) nonlinear regression with an upper constraint of < 100. Each neutralization experiment was conducted on three independent days, i.e., biological replicates, where each biological replicate contains a technical duplicate. IC_{50} values across biological replicates are presented as geometric mean ± geometric standard deviation. The loss or gain of neutralization potency across spike variants was calculated by dividing the variant IC_{50} by the parental (D614G) IC_{50} within each biological replicate, and then visualized as geometric mean ± geometric standard deviation.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification and statistical analyses were performed using GraphPad Prism (v8), R, and Biacore T200 Evaluation software, as described in the Method details.

CellPress
OPEN ACCESS

Supplemental Figures

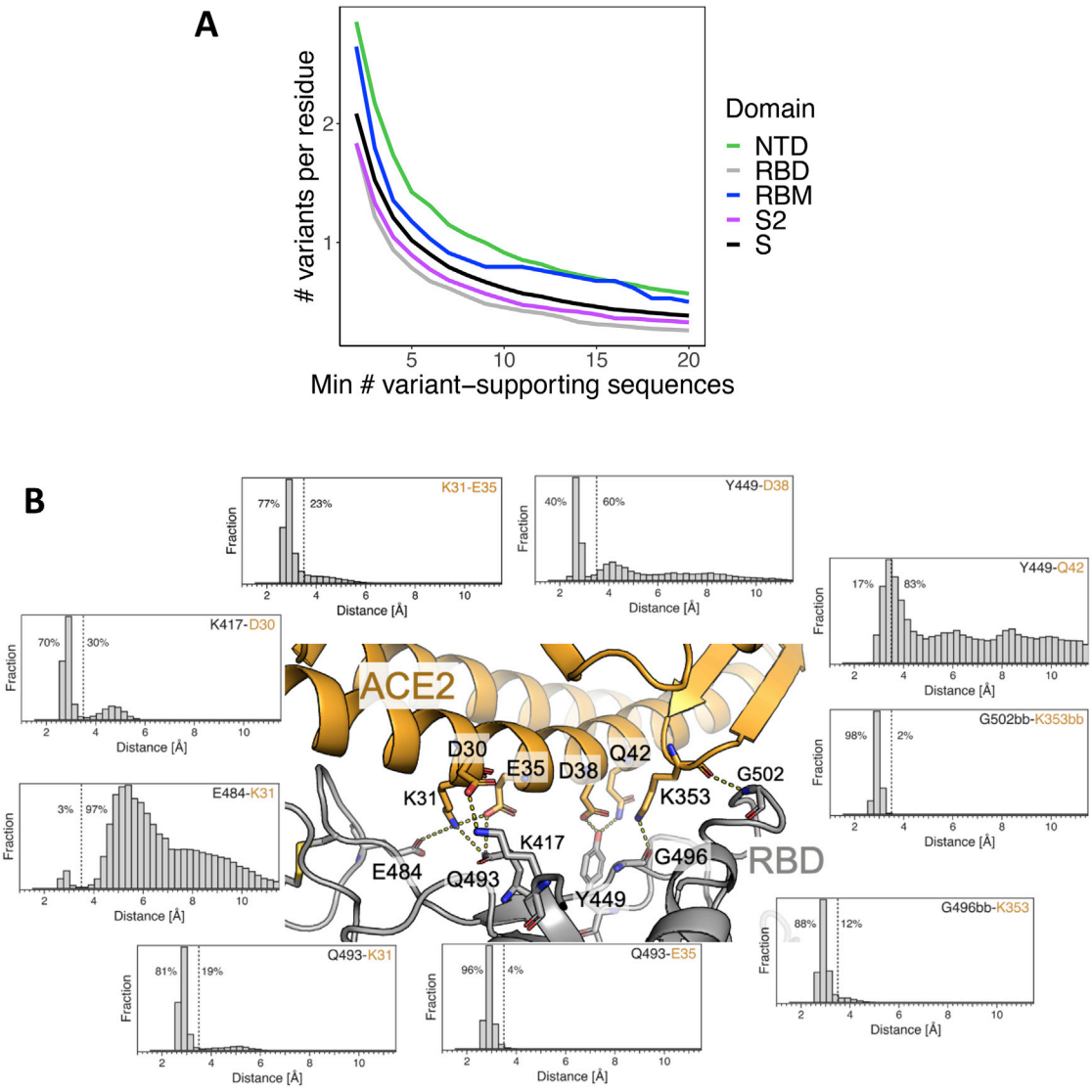
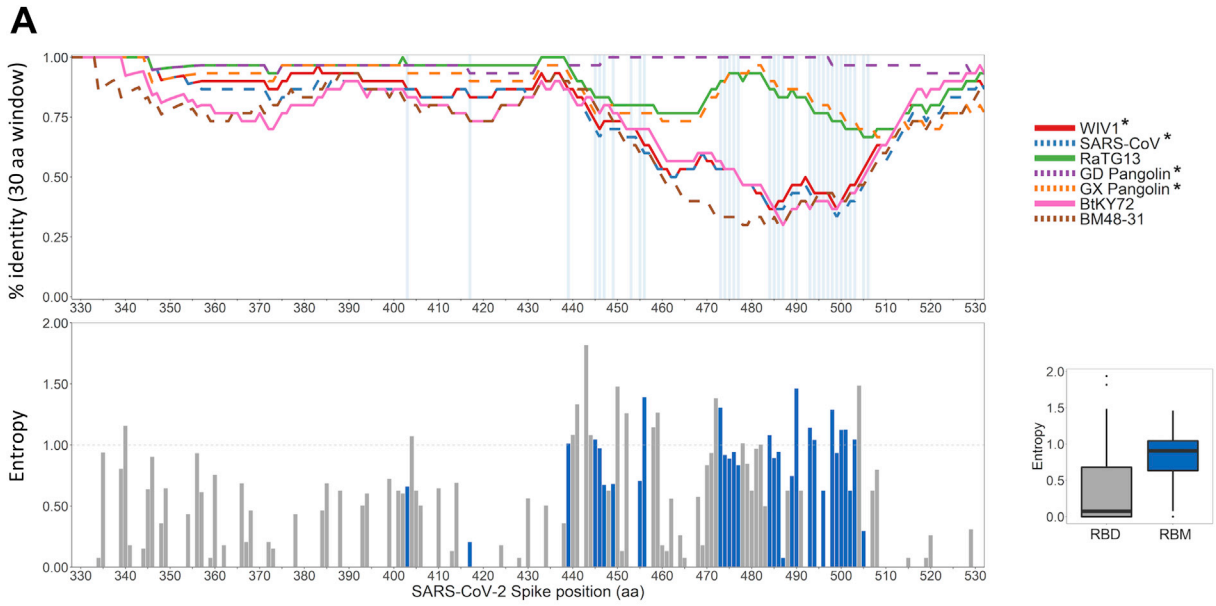
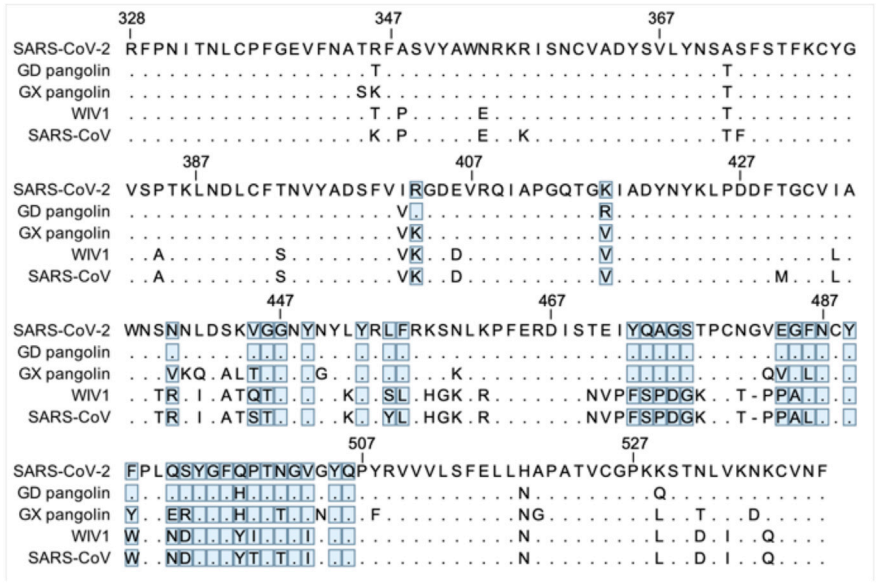


Figure S1. High RBM variability in deposited SARS-CoV-2 sequences is consistent with a dynamic RBD:hACE2 binding interface, related to Figures 1 and 2

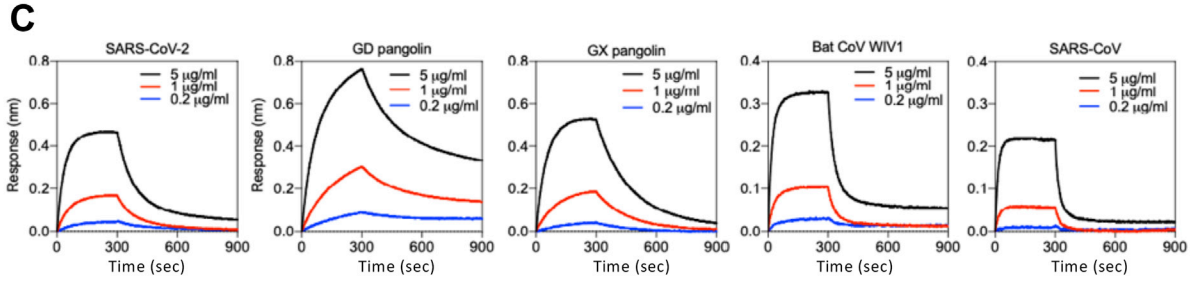
(A) Number of observed variants in four S domains (or full mature S protein) normalized by the total number of residues in each domain, where the number of observed isolates required to call a variant is varied along the x axis.

(B) Distributions of distances observed for RBD (gray):hACE2 (gold) residue pairs: K417-D30, E484-K31, Q493-K31, Q493-E35, G496bb-K353, G502bb-K353bb, Y449-Q42, Y449-D38, K31-E35 (bb = backbone interaction). RBD:ACE2 residue pairs were chosen based on RBM residues with high binding energies as determined by the binding energy % column (green) in Figure 2. Distances were computed every 2.5 ns from 118.7 μs of molecular dynamics simulation data. Dashed lines indicate a distance of 3.5 Å and the percentage of distances below and above 3.5 Å are annotated to the left and right of the lines, respectively.





	RBM	RBD
GD Pangolin	94.1%	96.7%
GX Pangolin	67.7%	86.5%
Bat CoV WIV1	41.2%	76.6%
SARS-CoV	35.3%	74.3%



(legend on next page)

Figure S2. RBDs from bat and pangolin *Sarbecovirus* isolates bind to hACE2 despite RBM divergence, related to Figures 1 and 2

(A) Top – Percent identity to SARS-CoV-2 using a sliding window size of 30 amino acids for seven related *Sarbecoviruses* (see figure key, *: viruses which bind to hACE2) across the RBD region of the Spike protein. Bottom – Site-specific entropy plot across the RBD protein alignment of SARS-CoV-2 and 68 related viruses (Table S1). Sites constituting the RBM are annotated in blue; the x axis refers to absolute positions in the SARS-CoV-2 Spike protein sequence. Right – boxplot of site-specific entropy values for the RBM sites (blue) and the full RBD (gray).

(B) Sequence alignment (left) and identity for RBM and RBD (right) to SARS-CoV-2 of the RBD sequences showing binding to hACE2. RBM residues indicated by blue boxes.

(C) Binding of hACE2 to human, pangolin, and bat *Sarbecovirus* RBDs by BLI. Bat CoV RaTG13, Bat CoVs ZC45, BtKY72 and BGR2008 have also been tested and did not bind hACE2.

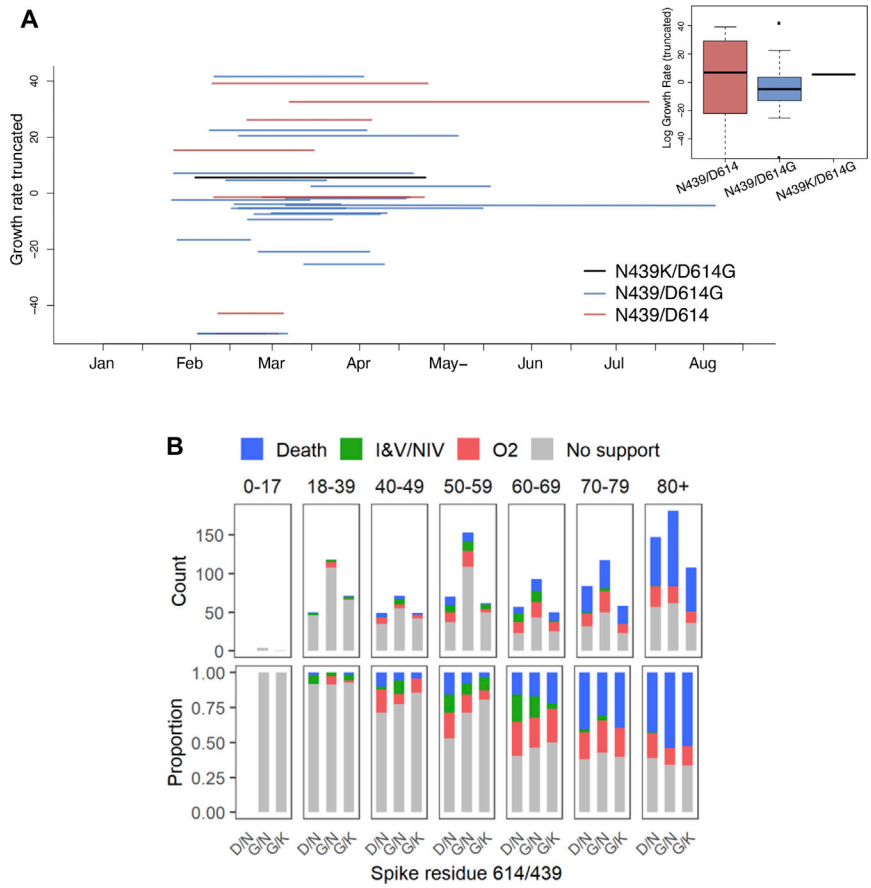


Figure S3. Virological and clinical results stratified by positions 439 and 614, related to Figures 3 and 5

(A) Phylodynamic analysis showing lineage growth rates relative to sampling times for UK lineages in Scotland. Data used for analysis were sampled between Feb 28, 2020 and Aug 18, 2020 (see STAR methods and http://sars2.cvr.gla.ac.uk/RiseFallScotCOVID/). The Scottish N439K lineage i (which co-occurs with D614G) is indicated in black along with whether wild-type N439 lineages are D614 (red) or D614G (blue). The inset shows a boxplot for the distributions of these genotypes. Note, only the growth rates between −50 and 50 are plotted.

(B) Comparison of clinical severity between D614/N439, D614G/N439 and D614G/N439K genotypes by patient age group for 1591 patients whose diagnostic samples were sequenced. Ordinal scale scored by oxygen requirement: 1. No respiratory support, 2: Supplemental oxygen, 3: Invasive or non-invasive ventilation or oxygen delivery by high flow nasal cannulae, 4: Death.

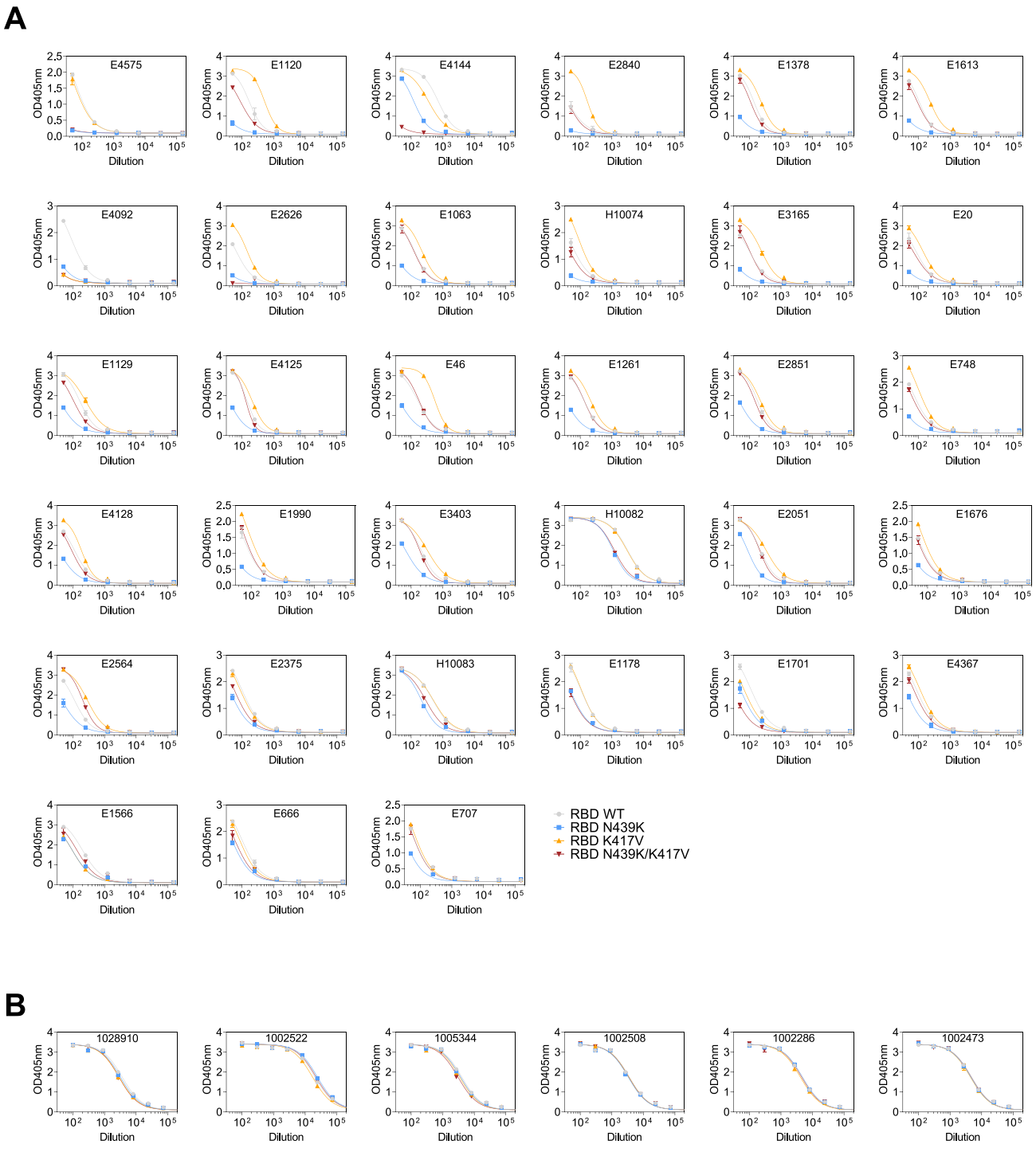


Figure S4. Sera ELISA results, related to Figure 6

ELISA binding of the 33 human sera with a > 2-fold reduction of binding to RBD N439K (A) and of the 6 sera of individuals infected with SARS-CoV-2 N439K viruses (B) to RBD WT (gray), N439K (blue), K417V (yellow) and N439K/K417V (red). Representative of n = 2 independent experiments.

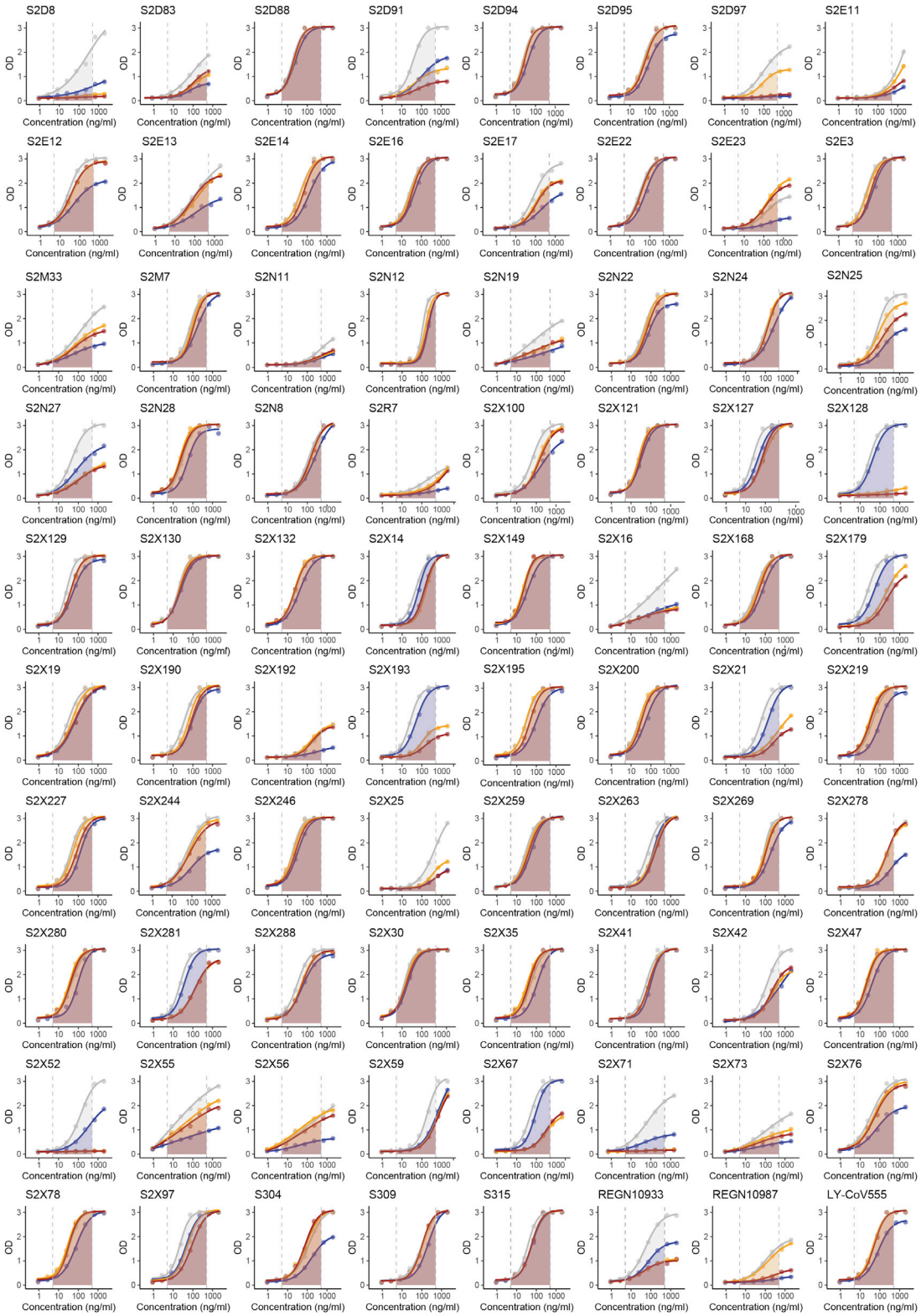


Figure S5. mAb ELISA results, related to Figure 6

ELISA binding of 80 out of the 144 mAbs to RBD WT (gray), N439K (blue), K417V (yellow) and N439K/K417V (red). AUC used for quantification is highlighted between dotted lines. Representative of n = 2 independent experiments. See Data S1 for results of all 144 mAbs.

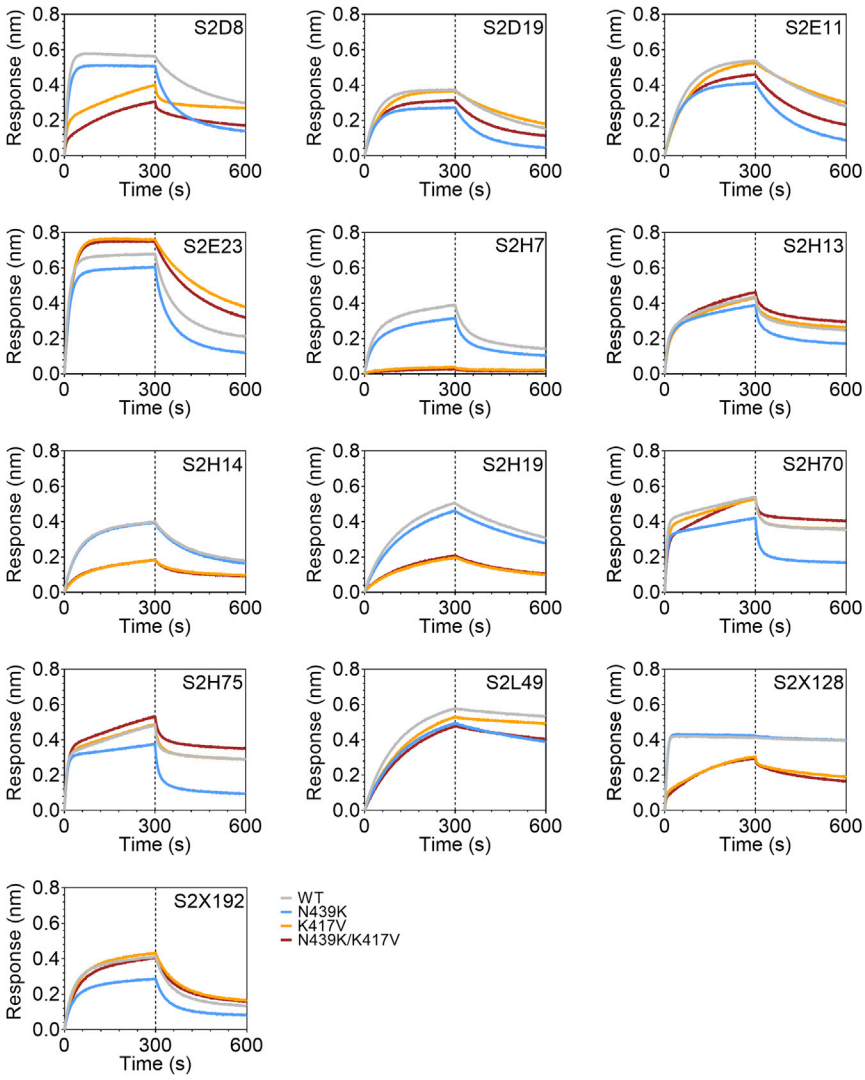


Figure S6. mAb BLI results, related to Figure 6

Binding of 13 selected mAbs to RBD WT (gray), N439K (blue), K417V (yellow) and N439K/K417V (red) as measured by BLI. Representative of n = 2 independent experiments.

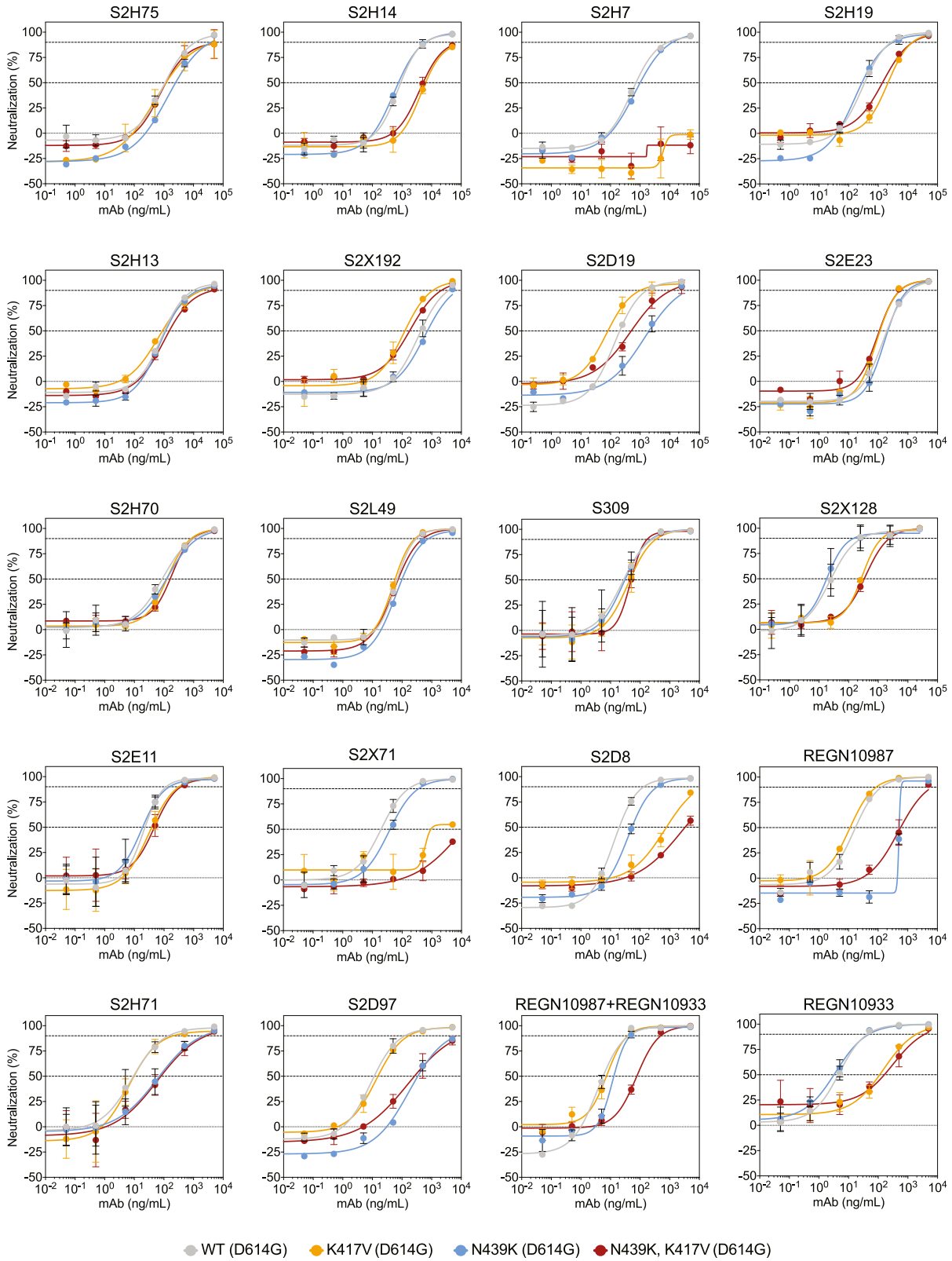


Figure S7. VSV pseudovirus neutralization curves of all mAbs tested, related to Figure 7

Representative of n = 3 biological replicates, bars = SD of technical duplicate.