










# SARS-CoV-2 variants, spike mutations and immune escape

William T. Harvey <sup>1,2,8</sup>, Alessandro M. Carabelli <sup>3,8</sup>, Ben Jackson <sup>4</sup>, Ravindra K. Gupta<sup>5</sup>, Emma C. Thomson<sup>6,7</sup>, Ewan M. Harrison<sup>3,7</sup>, Catherine Ludden<sup>3</sup>, Richard Reeve <sup>1</sup>, Andrew Rambaut <sup>4</sup>, COVID-19 Genomics UK (COG-UK) Consortium\*, Sharon J. Peacock<sup>3</sup> and David L. Robertson <sup>2</sup> 

**Abstract** | Although most mutations in the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome are expected to be either deleterious and swiftly purged or relatively neutral, a small proportion will affect functional properties and may alter infectivity, disease severity or interactions with host immunity. The emergence of SARS-CoV-2 in late 2019 was followed by a period of relative evolutionary stasis lasting about 11 months. Since late 2020, however, SARS-CoV-2 evolution has been characterized by the emergence of sets of mutations, in the context of ‘variants of concern’, that impact virus characteristics, including transmissibility and antigenicity, probably in response to the changing immune profile of the human population. There is emerging evidence of reduced neutralization of some SARS-CoV-2 variants by postvaccination serum; however, a greater understanding of correlates of protection is required to evaluate how this may impact vaccine effectiveness. Nonetheless, manufacturers are preparing platforms for a possible update of vaccine sequences, and it is crucial that surveillance of genetic and antigenic changes in the global virus population is done alongside experiments to elucidate the phenotypic impacts of mutations. In this Review, we summarize the literature on mutations of the SARS-CoV-2 spike protein, the primary antigen, focusing on their impacts on antigenicity and contextualizing them in the protein structure, and discuss them in the context of observed mutation frequencies in global sequence datasets.

## Variants

In the context of viruses, genetically distinct viruses with mutations different from those of other viruses. ‘Variants’ can also refer to the founding virus of a cluster or lineage and is used to refer collectively to the resulting variants that form the lineage. Variants with changed biological characteristics or antigenicity have been termed ‘variants of interest’, ‘variants under investigation’ or ‘variants of concern’ by public health bodies.

<sup>✉</sup>*e-mail*: [david.l.robertson@glasgow.ac.uk](mailto:david.l.robertson@glasgow.ac.uk)  
<https://doi.org/10.1038/s41579-021-00573-0>

As of April 2021, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, accounted for more than 143 million infections and more than three million deaths worldwide<sup>1</sup>. Virus genomic sequences are being generated and shared at an unprecedented rate, with more than one million SARS-CoV-2 sequences available via the [Global Initiative on Sharing All Influenza Data \(GISAID\)](#), permitting near real-time surveillance of the unfolding pandemic<sup>2</sup>. The use of pathogen genomes on this scale to track the spread of the virus internationally, study local outbreaks and inform public health policy signifies a new age in virus genomic investigations<sup>3</sup>. Further to understanding epidemiology, sequencing enables identification of emerging SARS-CoV-2 variants and sets of mutations potentially linked to changes in viral properties.

As highly deleterious mutations are rapidly purged, most mutations observed in genomes sampled from circulating SARS-CoV-2 virions are expected to be either neutral or mildly deleterious. This is because although

high-effect mutations that contribute to virus adaptation and fitness do occur, they tend to be in the minority compared with tolerated low-effect or no-effect ‘neutral’ amino acid changes<sup>4</sup>. A small minority of mutations are expected to impact virus phenotype in a way that confers a fitness advantage, in at least some contexts. Such mutations may alter various aspects of virus biology, such as pathogenicity, infectivity, transmissibility and/or antigenicity. Although care has to be taken not to confound mutations being merely present in growing lineages with mutations that change virus biology<sup>5</sup>, fitness-enhancing mutations were first detected to have arisen within a few months of the evolution of SARS-CoV-2 within the human population. For example, the spike protein amino acid change D614G was noted to be increasing in frequency in April 2020 and to have emerged several times in the global SARS-CoV-2 population, and the coding sequence exhibits a high dN/dS ratio, suggesting positive selection at the codon position 614 (REFS<sup>6,7</sup>). Subsequent studies indicated that D614G confers a moderate advantage for infectivity<sup>8,9</sup>

## Mutations

The substitutions, insertions or deletions of one or more nucleotides in the virus RNA genome. Non-synonymous nucleotide substitutions in protein-coding sequence result in a change in amino acid (referred to as a substitution or replacement), whereas synonymous nucleotide substitutions do not change the amino acid.

## Lineages

Monophyletic clusters of viruses assigned on the basis of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) global phylogenetic tree.

## dN/dS ratio

The ratio of non-synonymous mutations per non-synonymous site (dN) to synonymous mutations per synonymous site (dS), which is used to estimate the balance between neutral mutations, purifying selection and positive selection acting on gene or a specific codon.

## Amino acid substitution

A change in a specific amino acid of a protein. This is caused by non-synonymous mutations. By convention, an amino acid substitution is written in the form N501Y to denote the wild-type amino acid (N (asparagine)) and the substituted amino acid (Y (tyrosine)) at site 501 in the amino acid sequence.

## Monoclonal antibodies

(mAbs). Antibodies made by cloning a unique white blood cell, which usually has monovalent binding affinity for a specific epitope. Virus particles can be saturated with mAbs, and the structure can be solved to determine the antibody footprint or mAbs can be used to select for mutations that escape recognition.

and transmissibility<sup>10</sup>. Several other spike mutations of note have now arisen and are discussed in this Review, with particular focus on mutations affecting antigenicity.

The extent to which mutations affecting the antigenic phenotype of SARS-CoV-2 will enable variants to circumvent immunity conferred by natural infection or vaccination remains to be determined. However, there is growing evidence that mutations that change the antigenic phenotype of SARS-CoV-2 are circulating and affect immune recognition to a degree that requires immediate attention. The spike protein mediates attachment of the virus to host cell-surface receptors and fusion between virus and cell membranes<sup>11</sup> (BOX 1). It is also the principal target of neutralizing antibodies generated following infection by SARS-CoV-2 (REFS<sup>12,13</sup>), and is the SARS-CoV-2 component of both mRNA and adenovirus-based vaccines licensed for use and others awaiting regulatory approval<sup>14</sup>. Consequently, mutations that affect the antigenicity of the spike protein are of particular importance. In this Review, we explore the literature on these mutations and their antigenic consequences, focusing on the spike protein and antibody-mediated immunity, and discuss them in the context of observed mutation frequencies in global sequence datasets.

## Spike mutations receiving early attention

The rate of evolution of SARS-CoV-2 from December 2019 to October 2020 was consistent with the virus acquiring approximately two mutations per month in the global population<sup>15,16</sup>. Although our understanding of the functional consequences of spike mutations is rapidly expanding, much of this knowledge involves the reactive investigation of amino acid changes identified as rapidly increasing in frequency or being associated with unusual epidemiological characteristics. Following the emergence of D614G, an amino acid substitution within the receptor-binding motif (RBM), N439K, was noted as increasing in frequency in Scotland in March 2020. Whereas this first lineage with N439K (designated B.1.141 with the Pango nomenclature system<sup>17</sup>) quickly became extinct, another lineage that independently acquired N439K (B.1.258) emerged and circulated widely in many European countries<sup>18</sup>. N439K is noteworthy as it enhances the binding affinity for the ACE2 receptor and reduces the neutralizing activity of some monoclonal antibodies (mAbs) and polyclonal antibodies present in sera from people who have recovered

from infection<sup>18</sup>. Another RBM amino acid change, Y453F — associated with increased ACE2-binding affinity<sup>19</sup> — received considerable attention following its identification in sequences associated with infections in humans and mink; most notably one lineage identified in Denmark and initially named ‘cluster 5’ (now B.1.1.298)<sup>20</sup>. As of 5 November 2020, 214 humans infected with SARS-CoV-2 related to mink were all carrying the mutation Y453F<sup>21</sup>. The B.1.1.298 lineage also has Δ69–70, an amino-terminal domain (NTD) deletion that has emerged several times across the global SARS-CoV-2 population, including in the second N439K lineage, B.1.258. Δ69–70 is predicted to alter the conformation of an exposed NTD loop and has been reported to be associated with increased infectivity<sup>22</sup>.

Genomic analyses indicate a change in host environment and signatures of increased selective pressures acting upon immunologically important SARS-CoV-2 genes sampled from around November 2020 (REF.<sup>23</sup>). This coincided with the emergence of variants with higher numbers of mutations relative to previous circulating variants. These lineages because of their association with increased transmissibility were named ‘variants of concern’. They are defined by multiple convergent mutations that are hypothesized to have arisen either in the context of chronic infections or in previously infected individuals<sup>24–29</sup>. In addition to understanding the transmissibility and pathogenicity of these emerging variants, it is crucially important to characterize their antigenicity and the level of cross-protection provided by infection by earlier viruses that are genetically and antigenically similar to the virus that first emerged in December 2019 and which is used in all of the current vaccine preparations. Information on how spike mutations affect antigenic profiles can be derived from structural studies, mutations identified in viruses exposed to mAbs or plasma containing polyclonal antibodies, targeted investigations of variants using site-directed mutagenesis and deep mutational scanning (DMS) experiments that systematically investigate the possibility of mutations arising.

## Immunogenic regions of spike

Several studies have probed the antigenicity of the SARS-CoV-2 spike protein by epitope mapping approaches, including solving the structure of the spike protein in complex with the antigen-binding fragment of particular antibodies<sup>13,30–32</sup>. Serological analyses of almost 650 individuals infected with SARS-CoV-2 indicated that ~90% of the plasma or serum neutralizing antibody activity targets the spike receptor-binding domain (RBD)<sup>12</sup>. A relative lack of glycan shielding may contribute to the immunodominance of the RBD<sup>33</sup>. One study reported structural, biophysical and bioinformatics analyses of 15 SARS-CoV-2 RBD-binding neutralizing antibodies<sup>31</sup>. Antibody footprints were generated by structural analyses of the spike residues considering potential hydrogen bonds and van der Waals interactions with a mAb atom that were less than 4.0 Å. Structural analyses allowed the categorization of RBD-binding neutralizing antibodies into four classes (FIG. 1a,b): ACE2-blocking antibodies that bind the spike protein in the open conformation (class 1); ACE2-blocking antibodies

## Author addresses

<sup>1</sup>Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK.

<sup>2</sup>MRC–University of Glasgow Centre for Virus Research, Glasgow, UK.

<sup>3</sup>Department of Medicine, University of Cambridge, Cambridge, UK.

<sup>4</sup>Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK.

<sup>5</sup>Cambridge Institute of Therapeutic Immunology and Infectious Disease, University of Cambridge, Cambridge, UK.

<sup>6</sup>Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, UK.

<sup>7</sup>Wellcome Sanger Institute, Hinxton, UK.

<sup>8</sup>These authors contributed equally: William T. Harvey, Alessandro M. Carabelli.

\*A list of members and their affiliations appears in Supplementary information.

### Box 1 | Spike protein structure and function

As with other coronaviruses, the entry of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) into host cells is mediated by the transmembrane spike glycoprotein, which forms homotrimers on the surface of the virion. The SARS-CoV-2 spike protein is highly glycosylated, with 66 potential N-glycosylation sites per trimer<sup>98,99</sup> (FIG. 4a). The SARS-CoV-2 spike protein is post-translationally cleaved by mammalian furin into two subunits: S1 and S2 (FIG. 4a). The S1 subunit largely consists of the amino-terminal domain and the receptor-binding domain (RBD), and is responsible for binding to the host cell-surface receptor, ACE2, whereas the S2 subunit includes the trimeric core of the protein and is responsible for membrane fusion (FIG. 4b). The presence of a polybasic furin cleavage site at the S1–S2 boundary, which is unique within the subgenus *Sarbecovirus*, is important for infectivity and virulence<sup>100</sup>, with furin cleavage facilitating the conformational change required for receptor binding<sup>50</sup>. The spike protein transiently undergoes conformational changes between a closed conformation and an open conformation in which a hinge-like movement raises the RBD<sup>50</sup>. The residues comprising the receptor-binding motif are revealed on the upright RBD, enabling binding to ACE2, which induces a progressively more open structure until a fully open, three-ACE2-bound structure is formed, initiating S2 unsheddung and membrane fusion<sup>101</sup>.

#### Epitope mapping

Experimental determination of the binding site, or epitope, of an antibody. Approaches include X-ray co-crystallography or cryogenic electron microscopy of an antigen–antibody complex and the mapping of systematic mutations introduced by site-directed mutagenesis.

#### Glycan shielding

The process by which a virus can cloak underlying protein, impeding antibody binding. This is mediated by glycans, bulky sugar molecules that are covalently attached to amino acid side chains of the viral protein.

#### Immunodominance

The phenomenon by which the host immune response against a viral particle is mostly focused on a few antigens and mediated by potentially neutralizing antibodies.

#### Antibody footprints

Amino acid residues of a 3D folded protein that are targeted and contacted by a binding antibody.

#### Glycoprotein

A protein with oligosaccharide chains (glycans) covalently attached to amino acid side chains. Virus surface glycoproteins embedded in the membrane often have a role in interactions with host cells, including receptor binding, and are also commonly targeted by host antibodies.

that bind the RBD in both the open conformation and the closed conformation (class 2); antibodies that do not block ACE2 and bind the RBD in both the open conformation and the closed conformation (class 3); and neutralizing antibodies that bind outside the ACE2 site and only in the open conformation (class 4)<sup>31</sup>. Within the RBD, RBM epitopes overlapping the ACE2 site are immunodominant, whereas other RBD sites generate lower and variable responses in different individuals<sup>12</sup>.

Although the RBD is immunodominant, there is evidence for a substantial role of other spike regions in antigenicity, most notably the NTD<sup>13,30,34</sup>. Early structural characterization of NTD-specific antibodies 4A8 (REF.<sup>32</sup>) and 4–8 (REF.<sup>13</sup>) revealed similar epitope locations towards the upper side of the most prominently protruding area of the NTD. Cryogenic electron microscopy was used to determine the antibody footprint of the neutralizing antibody 4A8, and showed key interactions involving spike residues Y145, H146, K147, K150, W152, R246 and W258 (REF.<sup>32</sup>). Epitope binning of 41 NTD-specific mAbs led to the identification of six antigenic sites, one of which is recognized by all known NTD-specific neutralizing antibodies and has been termed the ‘NTD supersite’, consisting of residues 14–20, 140–158 and 245–264 (REF.<sup>30</sup>) (FIG. 1a,b). The mechanism of neutralization by which NTD-specific antibodies act remains to be fully determined, although it may involve the inhibition of conformational changes or proposed interactions with auxiliary receptors such as DC-SIGN or L-SIGN<sup>32,35</sup>. Relatively little is known of antigenicity in the S2 subunit, with immunogenicity thought to be impeded by extensive glycan shielding<sup>36</sup>, and although both linear and cross-reactive conformational S2 epitopes have been described<sup>37,38</sup>, the biological significance of these is not yet known.

#### Spike RBD mutations and immune escape

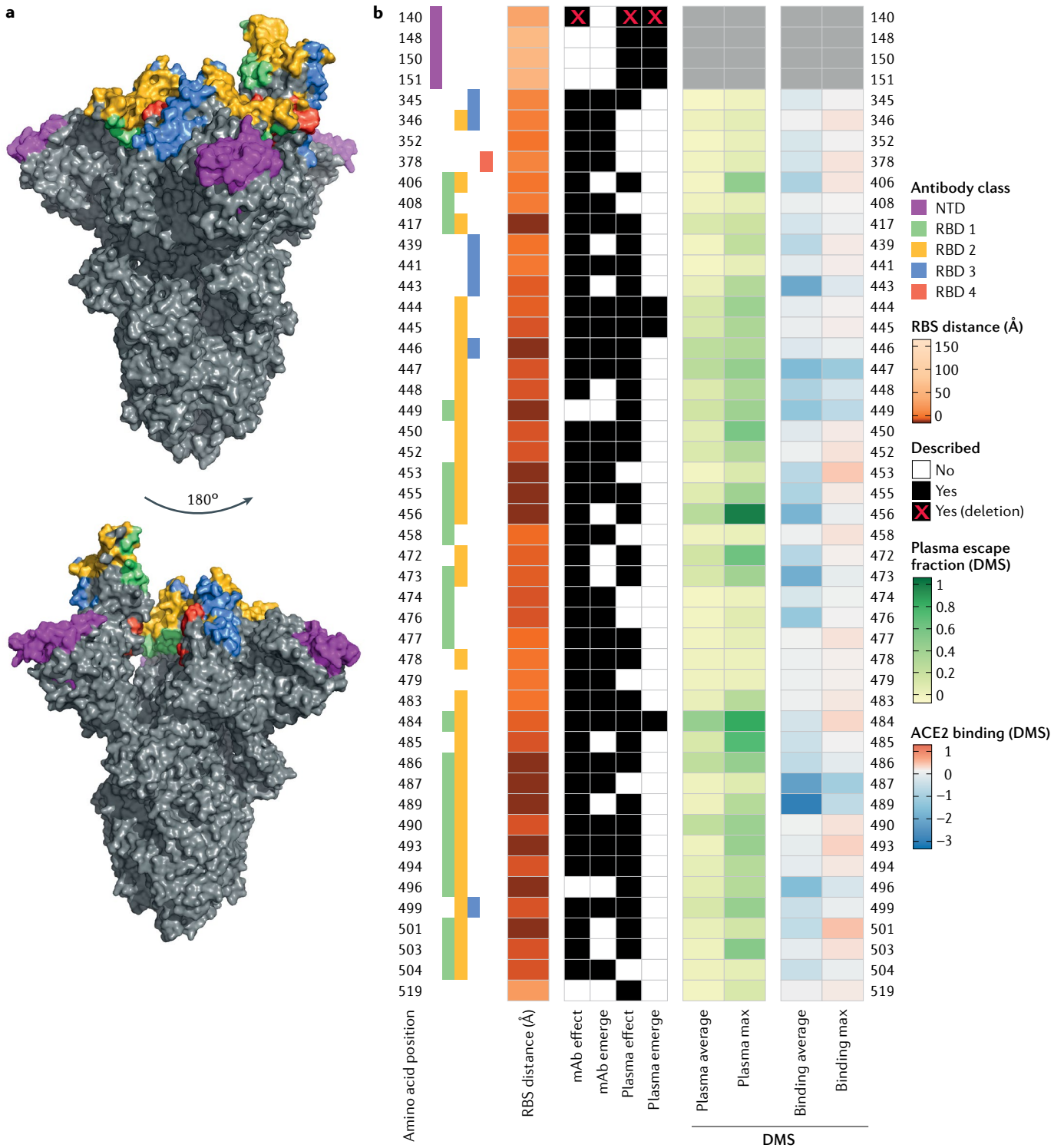
Several studies have contributed to the current understanding of how mutations in the SARS-CoV-2 spike protein affect neutralization. These studies include traditional escape mutation work that identifies mutations that emerge in virus populations exposed to either

mAbs<sup>39</sup> or convalescent plasma containing polyclonal antibodies<sup>40,41</sup>; targeted characterization of particular mutations<sup>18,42</sup>; and wider investigations of either large numbers of circulating variants<sup>13</sup> or all possible amino acid substitutions in the RBD<sup>39,44–46</sup>. For spike residues where mutations have been shown to influence polyclonal antibody recognition, the observation of an effect on either mAbs or plasma is indicated in FIG. 1b. For a smaller number of residues, escape mutations emerging in virus exposed to mAbs or polyclonal plasma have been described (‘mAb emerge’ and ‘plasma emerge’ in FIG. 1b).

In a DMS study, researchers assessed all possible single amino acid variants using a yeast-display system and detected variants that escape either nine neutralizing SARS-CoV-2 mAbs<sup>45</sup> or convalescent plasma from 11 individuals taken at two time points after infection<sup>39</sup> (shades of green in FIG. 1b). The resulting heat maps provide rich data on the antigenic consequence of RBD mutations, with the plasma escape mutations being of particular interest given that they impact neutralization by polyclonal antibodies of the kind SARS-CoV-2 encounters in infections, with significant levels of immunity acquired through prior exposure or vaccination. Although significant interperson and intraperson heterogeneity in the impact of mutations on neutralization by polyclonal serum has been described, the mutations that reduce antibody binding the most occur at a relatively small number of RBD residues, indicating substantial immunodominance within the RBD<sup>39</sup>.

Of all RBD residues for which substitutions affected recognition by convalescent sera, DMS identified E484 as being of principal importance, with amino acid changes to K, Q or P reducing neutralization titres by more than an order of magnitude<sup>39</sup>. E484K has also been identified as an escape mutation that emerges during exposure to mAbs C121 and C144 (REF.<sup>40</sup>) and convalescent plasma<sup>41</sup>, and was the only mutation described in one study as able to reduce the neutralizing ability of a combination of mAbs (REGN10989 and REGN10934) to an unmeasurable level<sup>47</sup>. In an escape mutation study using 19 mAbs, substitutions at E484 emerged more frequently than at any other residue (in response to four mAbs), and each of the four 484 mutants identified (E484A, E484D, E484G and E484K) subsequently conferred resistance to each of four convalescent sera tested<sup>48</sup>. No other mAb-selected escape mutants escaped each of the four sera, although the mutations K444E, G446V, L452R and F490S escaped three of the four sera tested<sup>48</sup>.

Mutations at position 477 of the spike protein (S477G, S477N and S477R) rank prominently among mAb escape mutations identified by one study, and the mutation S477G conferred resistance to two of the four sera tested<sup>48</sup>. However, substitutions at 477 were not identified as being important in DMS with convalescent plasma<sup>39</sup>. The mutation N439K increases affinity for ACE2 (REF.<sup>19</sup>), is predicted to result in an additional salt bridge at the RBM–ACE2 interface and is thought to preferentially reduce the neutralization potential of plasma that already has low neutralizing activity<sup>18</sup>. However, a DMS study<sup>39</sup> did not find that the mutation



N439K significantly alters neutralization by polyclonal antibodies in plasma, in contrast to previous studies that found that N439K reduced neutralization by mAbs and convalescent plasma<sup>18</sup>. One explanation for this inconsistency is that the mechanism of immune escape conferred by N439K is through increased ACE2 affinity rather than by directly affecting antibody epitope recognition and that perhaps the experimental design of the DMS study is less sensitive to detecting immune evasion mutations of this type.

**Spike NTD mutations and immune escape**

In the NTD, most of the evidence for immune evasion focuses on a region centred at a conformational epitope consisting of residues 140–156 (N3 loop) and 246–260 (N5 loop), which includes the epitope of the antibody 4A8<sup>32</sup> (FIG. 1, magenta). In studies that identified the emergence of antibody escape mutations in virus populations exposed to convalescent plasma, mutations were roughly evenly distributed between the RBD and the NTD (FIG. 1b). One study described the emergence of escape

**Epitopes**

The specific parts of an antigen recognized by the immune system: antibodies, B cells or T cells.



◀ **Fig. 1 | Neutralizing antibody classes defined by structural analyses and properties of spike protein residues.** **a** | Amino acid residues of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein are coloured according to the class of the antibody that binds to an epitope. Receptor-binding domain (RBD) antibody classes 1–4 (REF.<sup>31</sup>) are shown: green for class 1 (ACE2-blocking antibodies that bind the spike protein in the open conformation), yellow for class 2 (ACE2-blocking antibodies that bind the RBD in both the open conformation and the closed conformations), blue for class 3 (antibodies that do not block ACE2 and bind the RBD in both the open conformation and the closed conformations) and red for class 4 (neutralizing antibodies that bind outside the ACE2 site and only in the open conformation). When residues belong to epitopes of multiple classes, priority colouring is given to antibodies that block ACE2 and bind the closed spike protein. The amino-terminal domain (NTD) supersite<sup>30</sup> is coloured in magenta. **b** | Aligned heat maps showing properties of amino acid residues where substitutions affect binding by antibodies in polyclonal human blood plasma or emerge as antibody escape mutations. The distance in angstroms to the ACE2-contacting residues that form the receptor-binding site (RBS) is shown in shades of orange; each residue is classified as having evidence for mutations affecting neutralization by either monoclonal antibodies (mAbs)<sup>40,43,47,48</sup> or polyclonal antibodies in plasma from previously infected individuals (convalescent)<sup>39–41,43,48</sup> or vaccinated individuals<sup>59</sup> ('mAb effect' and 'plasma effect', respectively). A subset of these residues has mutations described as emerging upon exposure (co-incubation) to mAbs<sup>40,47,48</sup> or plasma<sup>40,41</sup> in laboratory experiments ('mAb emerge' and 'plasma emerge', respectively). When an observation includes a deletion, this is indicated by a red cross. Shades of green depict the results of deep mutational scanning (DMS) experiments where yeast cells expressing RBD mutants are incubated with multiple samples of human convalescent plasma<sup>39</sup>. The escape fraction (that is, a quantitative measure of the extent to which a mutation reduced polyclonal antibody binding) averaged across all amino acid substitutions at a residue ('plasma average') and the maximally resistant substitution ('plasma max') are indicated. DMS data on ACE2-binding affinity<sup>19</sup> are shown in shades of red or blue representing higher or lower ACE2 affinity, respectively. The mean change in binding affinity averaged across all mutations at each site ('binding average') and alternatively the maximally binding mutant ('binding max') is shown. Scores represent binding constants ( $\Delta\log_{10} K_D$ ) relative to the wild-type reference amino acid.

mutations in viruses exposed to convalescent plasma from two individuals, one of which selected for NTD mutations only (N148S, K150R, K150E, K150T, K150Q and S151P)<sup>40</sup>. This was despite the plasma being a source of the highly potent RBD-targeting mAb C144 (REF.<sup>40</sup>). NTD antibody escape mutations were not observed for the other samples of plasma investigated, and furthermore, the 148–151 mutants exhibited only marginal reductions in sensitivity to the plasma tested, indicating individual immune responses may be differentially affected by mutations of RBD and NTD epitopes<sup>40</sup>.

Deletions in the NTD have been observed repeatedly in the evolution of SARS-CoV-2 and have been described as changing NTD antigenicity<sup>30,41,42</sup>. One study identified four recurrently deleted regions (RDRs) within the NTD and tested five frequently observed deletions within these:  $\Delta 69-70$  (RDR1),  $\Delta 141-144$  and  $\Delta 146$  (RDR2),  $\Delta 210$  (RDR3) and  $\Delta 243-244$  (RDR4)<sup>42</sup>. Of the four RDRs, RDR1, RDR2 and RDR4 correspond to NTD loops N2, N3 and N5, whereas RDR3 falls between N4 and N5 in another accessible loop (FIG. 2a, asterisk). Both RDR2 deletions,  $\Delta 141-144$  and  $\Delta 146$ , and  $\Delta 243-244$  (RDR4) abolished binding of 4A8 (REF.<sup>42</sup>). Further evidence of the role of RDR2 deletions in immune escape was provided by a study that describes the emergence of  $\Delta 140$  in SARS-CoV-2 co-incubated with potentially neutralizing convalescent plasma, causing a fourfold reduction in neutralization titre<sup>41</sup>. This  $\Delta 140$  spike mutant subsequently acquired the E484K mutation, resulting in a further fourfold drop in neutralization titre, and thus a two-residue

change across the NTD and the RBD can drastically evade the polyclonal antibody response. The  $\Delta 140$ +E484K double mutant next acquired an 11-residue insertion in the NTD N5 loop between Y248 and L249, completely abolishing neutralization. This insertion, which also introduced a new glycosylation motif in the vicinity of RDR4, is predicted to alter the structure of the antigenic N3 and N5 NTD loops<sup>41</sup>. This finding further demonstrates the structural plasticity of the NTD and indicates that insertions and the acquisition of additional glycosylation motifs in the NTD are further mechanisms in addition to deletion that lead to immune evasion. Other examples of mutations that impact the epitope–paratope interface indirectly include mutations in the signal peptide region and at cysteine residues 15 and 136, which form a disulfide bond that 'staples' the NTD amino terminus against the galectin-like  $\beta$ -sandwich<sup>30</sup>. Mutations at those sites (for example, C136Y and S12P, which alter the cleavage occurring between residues C15 and V16) have been shown to affect the neutralizing activity of several mAbs, likely disrupting the disulfide bond and therefore dislodging the supersite targeted by several antibodies<sup>30</sup>.

Across the spike protein, some mutations that confer escape to neutralizing mAbs have little impact on serum antibody binding<sup>39,40,44</sup>, possibly because those mAbs are rare in polyclonal sera, targeting subdominant epitopes<sup>12,39,44</sup>. Escape mutations emerging in viruses exposed to convalescent plasma have been identified in both the NTD ( $\Delta F140$ , N148S, K150R, K150E, K150T, K150Q and S151P) and the RBD (K444R, K444N, K444Q, V445E and E484K)<sup>40,41</sup> (FIG. 1b). Notably, mutations emerging under selective pressure from convalescent plasma may be different from those selected by the most frequent mAb isolated from the same plasma<sup>40</sup>. Potentially, observed differences arise because mutations selected by convalescent plasma facilitate escape from multiple mAbs. Fewer data on the antigenic effects of S2 mutations exist, though D769H has been described as conferring decreased susceptibility to neutralizing antibodies<sup>24</sup>. Residue 769 is positioned in a surface-exposed S2 loop, and D769H was found to arise, in linkage with  $\Delta 69-70$ , in an immunocompromised individual treated with convalescent plasma<sup>24</sup>.

### Conformational epitopes in spike

To evaluate potential antigenicity across the spike protein, we analysed the protein using BEpro, a program for the prediction of conformational epitopes based on tertiary structure<sup>49</sup>. This approach calculates a structure-based epitope score, which approximates antibody accessibility for each amino acid position. For each residue, the calculated score accounts for the local protein structure: half-sphere exposure measures and propensity scores each depend on all atoms within 8–16 Å of the target residue, with weighting towards closer atoms. Due to this aggregation, calculated scores are relatively insensitive to the effects of single amino acid substitutions. Scores were calculated for the spike protein in both the closed conformation and the open conformation (FIG. 2). It has been estimated that ~34% of spike proteins are closed and 27% are open (with the remainder in an intermediate form) following furin cleavage<sup>50</sup>.

#### Epitope binning

An approach that uses a competitive immunoassay to sort a library of monoclonal antibodies into discrete groups of antibodies that compete for access to overlapping epitopes.

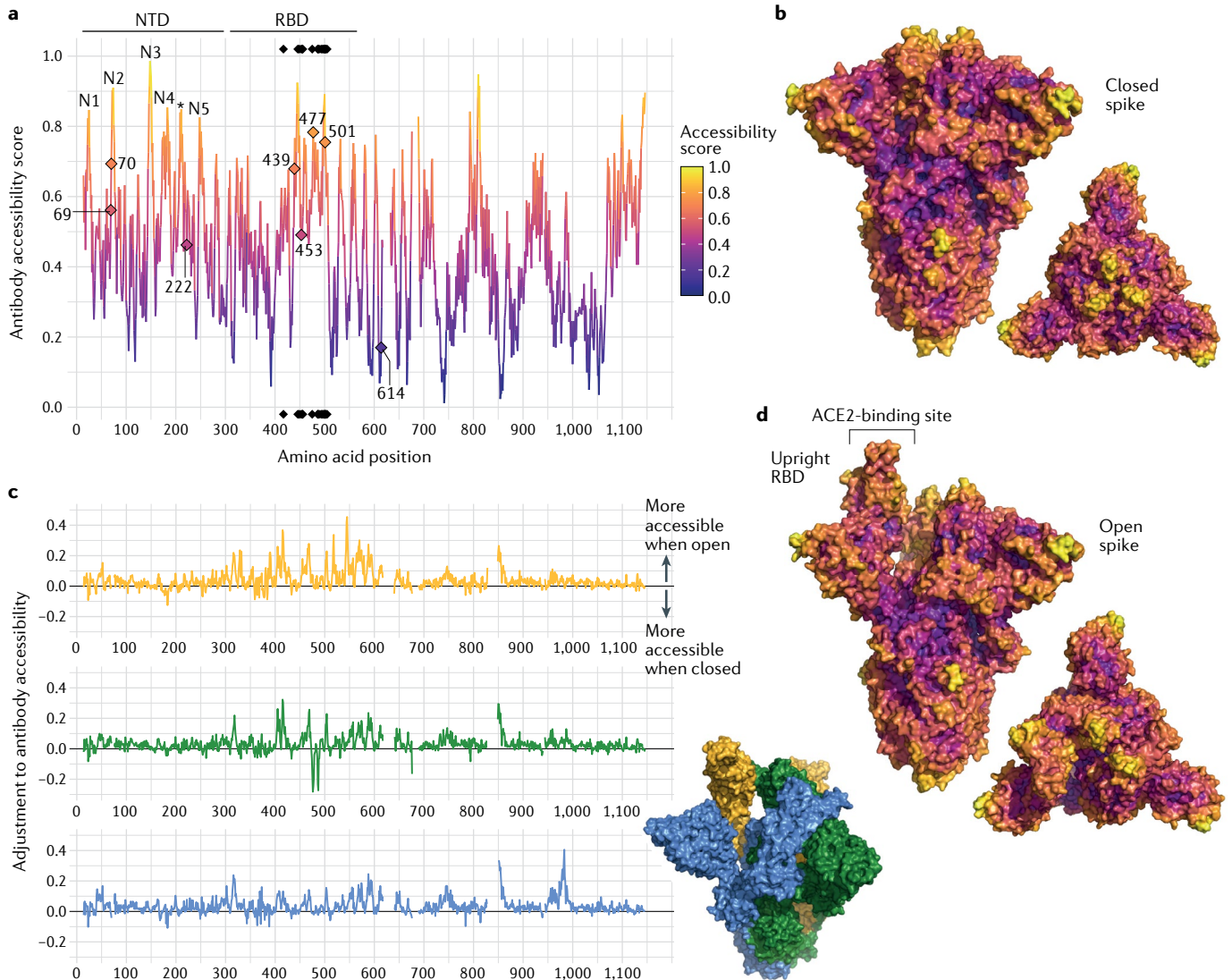
#### Convalescent plasma

Blood serum of a previously infected individual that usually contains a mixture of different antibodies referred to as polyclonal antibodies. Similarly, postvaccination serum includes polyclonal antibodies generated by vaccination.

Scores rescaled between 0 and 1 are plotted for the closed conformation in FIG. 2a and are represented on the structure in FIG. 2b. A limitation of this approach is that it does not account for glycan shielding of residues and likely overestimates scores at the base of the ectodomain for residues closest to the carboxy terminus.

Comparisons with reporting of antibody footprints and the impact of mutations on antigenicity indicate that residues with mutations described as affecting recognition by mAbs or antibodies in convalescent plasma (FIG. 1b) tend to occur at residues with higher structure-based antibody accessibility scores compared with other residues belonging to epitope footprints and residues not implicated in antigenicity (Supplementary

Fig. 1b). Notably, scores for residues with mutations described as affecting plasma antibody recognition are also slightly higher on average compared with those with mutations described as affecting mAbs only. Epitope scores are particularly high for residues with mutations described as emerging during exposure to convalescent plasma<sup>40,41</sup> (Supplementary Fig. 1b). Experimental data on the emergence of mutations under selective pressure from polyclonal antibodies are relatively rare, although these trends for higher scores associated with such mutations indicate that information from structural analysis approaches of this kind can contribute to the ranking of residues at which substitutions are likely to impact the polyclonal antibody response.



**Fig. 2 | Structure-based analysis of conformational epitopes on the spike protein.** **a** | Structure-based antibody accessibility scores for each spike protein ectodomain residue in the closed form were calculated with BEpro<sup>49</sup>. Black diamonds at the top and bottom of the plot indicate the positions of ACE2-contacting residues. Accessible amino-terminal domain (NTD) loops N1–N5 are labelled, and a loop falling between these is indicated with an asterisk. **b** | Two surface colour representations of antibody accessibility scores for the spike protein in the closed conformation according to the colour scheme in part **a**: a trimer axis vertical view (left) and an orthogonal top-down

view along this axis (right). **c** | The extent to which each spike residue becomes more or less accessible when the spike protein is in its open form is shown. For each spike monomer (upright receptor-binding domain (RBD) (yellow), closed RBD clockwise adjacent (green) and closed RBD anticlockwise adjacent (blue)), the difference relative to the score calculated for the closed form (shown in part **a**) is shown. **d** | Two surface colour representations of antibody accessibility scores for the spike protein in the open conformation with a single monomer with an upright RBD are shown: a trimer axis vertical view (left) and an orthogonal top-down view along this axis (right).

Within the RBD, the two areas with high structure-based antibody accessibility scores for the closed spike structure (FIG. 2a, peaks with consecutive residues with scores greater than 0.8) are centred at residues 444–447 and residues 498–500. These areas are represented as yellow patches near the centre of the top-down view of the spike structure in FIG. 2b. FIGURE 2c shows that, in general, residues become more accessible and are likelier to form epitopes when the spike protein is in the open conformation, and this is especially true for the RBD, particularly for the upright RBD (FIG. 2c, yellow). In the open form, residues close to the ACE2-binding site (405, 415, 416, 417 and 468) become much more exposed on both the upright RBD and the clockwise adjacent closed RBD (FIG. 2c, green). The effect of mutations at these positions is likely to be greater for antibodies belonging to RBD class 1. Residues centred at 444–447 and 498–500 maintain high scores on the upright RBD and are joined by residues in areas 413–417 and 458–465. The only RBD residues that become notably less accessible in the open spike structure are residues 476, 477, 478, 586 and 487 of the closed RBD clockwise adjacent to the upright RBD, which become blocked by the upright RBD (FIG. 2c, green). Several RBD-specific antibodies are able to bind only the open spike protein (RBD classes 1 and 4 (REF.<sup>31</sup>)), and interestingly, it has been observed that D614G makes the spike protein more vulnerable to neutralizing antibodies by increasing the tendency for the open conformation to occur<sup>51</sup>.

Within the NTD, the highest-scoring spike residues in the closed form belong to a loop centred at residues 147–150, which each have scores greater than 0.9 (FIG. 2a, yellow patch to the extreme right of the structure viewed from the side in FIG. 2b). This loop, known as the N3 loop, is described as forming key interactions with the neutralizing antibody 4A8 (REF.<sup>32</sup>). One study described the structure of five previously unmodelled, protruding NTD loops, denoting them N1–N5. In addition to N3, high-scoring residues (greater than 0.7) are found at positions 22–26 (N1), 70 (N2), 173–187 (N4), 207–213 (FIG. 2a, asterisk) and 247–253 (N5). Structural analysis indicates NTD-binding antibodies are likely able to bind epitopes when the spike protein is in either the closed conformation or open the conformation (FIG. 2c). Outside the NTD and the RBD, the highest-scoring residues are residues 676 and 689 (which lie on either side of the loop containing the S1–S2 furin cleavage site, which is disordered in both the open conformation and the closed conformation<sup>50</sup>), 793–794, 808–812, 1,099–1,100 and 1,139–1,146 (FIG. 2a). When the spike protein is in the open conformation, increased accessibility results in substantially higher potential epitope scores for S2 residues centred at 850–854, which become more accessible on all three spike monomers (FIG. 2c), and residues 978–984, which become more accessible on the monomer anticlockwise adjacent to the upright RBD monomer (FIG. 2c, blue).

### Structural context of spike mutations

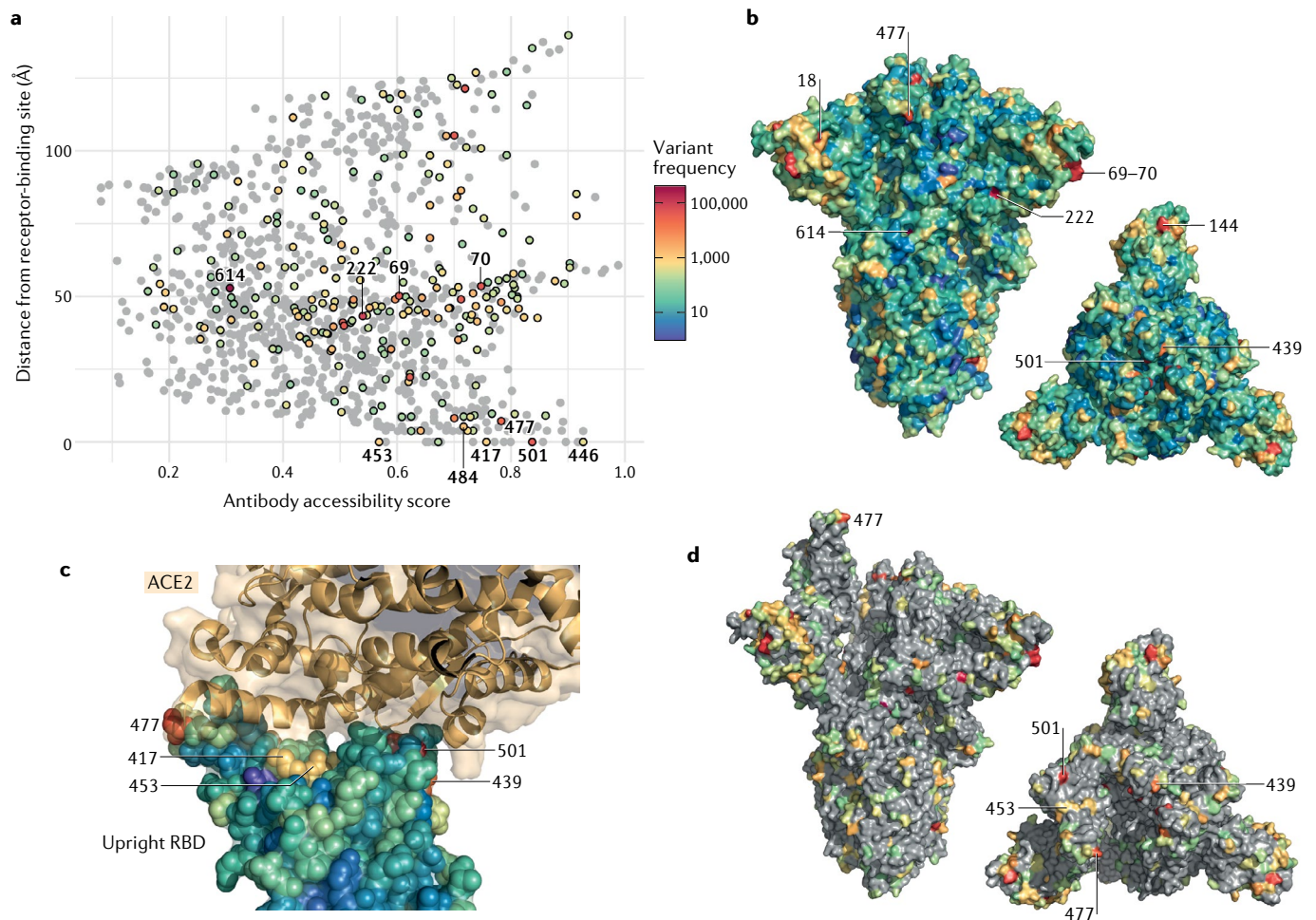
To assess the impact of spike mutations and their immunological role in the global SARS-CoV-2 population, we combined structural analyses with the observed

frequency of mutations in circulating variants (FIG. 3). Globally, the highest number of amino acid variants, mapped against the Wuhan-Hu-1 reference sequence (MN908947), are recorded at amino acid positions 614, 222 and 18 (FIG. 4a) (among 426,623 high-quality sequences retrieved from the GISAID database on 3 February 2021 and processed using CoV-GLUE). Residues at positions 614 and 222 have relatively low antibody access scores and are positioned ~50 Å from the RBS residues when the spike protein is in the open conformation (FIG. 3a,b). As mentioned earlier, there is evidence indicating that D614G confers a moderate advantage for infectivity<sup>8,9</sup> and increases transmissibility<sup>10</sup>. The spike amino acid substitution with the second highest frequency is A222V, which is present in the 20A.EU1 SARS-CoV-2 cluster (also designated lineage B.1.177). This lineage has spread widely in Europe and is reported to have originated in Spain<sup>52</sup>. There is no evidence for a notable impact of A222V on virus phenotype (that is, infectivity and transmissibility), and so its increase in frequency is generally presumed to have been fortuitous rather than a selective advantage. The substitution L18F has occurred ~21 times in the global population<sup>53</sup> and is associated with escape from multiple NTD-binding mAbs<sup>30</sup>.

Among the 5,106 independent substitutions observed in the spike protein (BOX 1), 161 are described as affecting recognition by mAbs or polyclonal antibodies in sera, of which 22 are present in more than 100 sequences. On average, variant frequency is higher at amino acid positions where mutations are described as affecting antibody recognition than at positions with no described substitutions of antigenic importance (Supplementary Fig. 1a), and high levels of amino acid substitutions are observed at some amino acid positions where mutations are described as affecting recognition by antibodies in convalescent plasma, including positions 439 and 484. This indicates that, generally, the amino acid positions at which antibody escape mutations have been detected *in vitro* tolerate mutations at least to some degree *in vivo*.

Within the RBD, the positions at which amino acid substitutions are present at the highest frequency are located close to the RBD–ACE2 interface (FIG. 3). Of the three RBD amino acid substitutions present in several thousand sequences, N439K and N501Y were described earlier, and N501Y is discussed in more detail in the next section in the context of variants of concern. The other substitution, S477N, is estimated to have emerged at least seven times in the global SARS-CoV-2 population and has persisted at a frequency of between 4% and 7% of sequences globally since mid-June 2020 (REF.<sup>53</sup>). One study described multiple mAbs that selected for the emergence of S477N and found this mutant to be resistant to neutralization by the entire panel of RBD-targeting mAbs that were tested. By contrast, when tested with convalescent serum, neutralization of the S477N mutant was similar to that of the wild type<sup>48</sup>. In common with N439K and N501Y, S477N results in increased affinity for the ACE2 receptor, although to a lesser extent<sup>19,54</sup>. As described in BOX 2, substitutions may facilitate immune escape by increasing receptor-binding affinity independently of any effect that they may have on antibody





**Fig. 3 | Structural context of spike amino acid mutations in the global virus population.** Spike amino acid residues are coloured according to the frequency of amino acid substitutions or deletions. Variants (retrieved from CoV-GLUE) are based on 426,623 high-quality sequences downloaded from the Global Initiative on Sharing All Influenza Data (GISAID) database on 3 February 2021. **a** | Points representing each spike amino acid residue are positioned according to the antibody accessibility score and the distance to the nearest residue in the receptor-binding site. Residues with at least 100 sequences possessing a substitution or deletion are coloured according to the frequency scale shown, with the remainder shaded grey.

**b** | Spike protein in closed form with all residues coloured according to the frequency scale shown; a trimer axis vertical view (left) and an orthogonal top-down view along this axis (right) are shown. **c** | A close-up view of the receptor-binding domain (RBD) bound to ACE2 (RCSB Protein Data Bank ID 6M0J<sup>95</sup>), with RBD residues shown as spheres coloured by amino acid variant frequency and ACE2 shown in gold. Amino acid variants are present at high frequency in positions at the RBD–ACE2 interface. **d** | Spike protein in open form with residues where at least 100 sequences possessing a substitution are highlighted; a trimer axis vertical view (left) and an orthogonal top-down view along this axis (right) are shown.

recognition of epitopes; therefore, it is possible that such a mechanism contributes to the impact of S477N on neutralization. Variant frequency is also moderately high at RBD–ACE2 interface amino acid positions 417, 453 and 446. Of these positions, 446 occurs in a location in the spike structure that is predicted to be highly antigenic, and substitutions at this site are described as affecting neutralization by both mAbs and antibodies present in polyclonal serum<sup>39,43,46,48</sup>. Substitutions at amino acid positions 417 and 453 are described in the next section in the context of variants of concern.

**Variants of interest or concern**

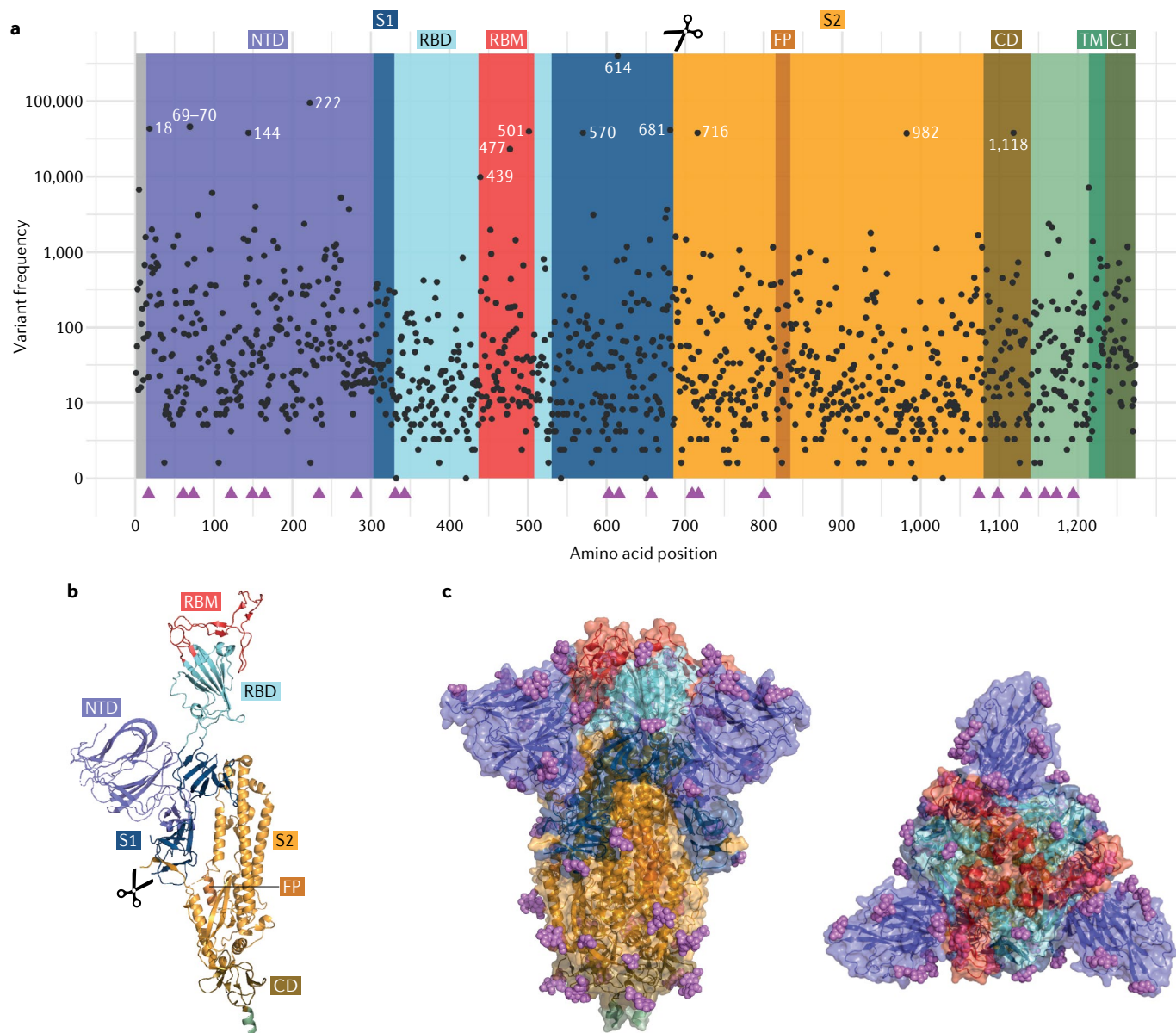
In addition to single mutations of note, more heavily mutated SARS-CoV-2 lineages have emerged. Arguably the first variant of interest defined by the presence of several spike mutations, and referred to as B.1.1.298 (cluster 5),

was detected in Denmark spreading among farmed mink and a small number of people<sup>20</sup>. This lineage is characterized by four amino acid differences, ΔH69–V70, Y453F, I692V and M1229I (FIG. 5). Of these, the Y453F substitution occurs at a residue within the ACE2 footprint and has been shown by DMS to increase ACE2 affinity<sup>19</sup>. In addition, Y453F has been described as reducing neutralization by mAbs<sup>17</sup>. In late 2020 and early 2021, the emergence and sustained transmission of lineages with mutations that affect the characteristics of the virus received much attention, most notably lineages B.1.1.7, B.1.351 and P.1 (also known as 501Y.V1, 501Y.V2 and 501Y.V3, respectively). The locations of the spike mutations in the B.1.1.298, B.1.1.7, B.1.351 and P.1 lineages are annotated in FIG. 5a, and information on the structural context and consequences of mutations for antibody recognition and ACE2 binding are shown in FIG. 5b.



Of the lineages summarized in FIG. 5, several amino acid substitutions are convergent, having arisen independently in different lineages: N501Y, which is present in lineages B.1.1.7, B.1.351 and P.1; E484K, which is present in lineages B.1.351 and P.1 and has been detected as emerging within the B.1.1.7 lineage<sup>55</sup>; and  $\Delta$ H69–V70

in lineages B.1.1.298 and B.1.1.7. Additionally, lineages B.1.351 and P.1 possess alternative amino acid substitutions K417N and K417T, respectively. Further lineages with these mutations have also been identified; for example, an independent emergence of N501Y in the B.1.1.70 lineage, which is largely circulating in Wales.



**Fig. 4 | Spike protein sequence variability and structure. a** | The domain organization of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein showing amino acid sequence variability. The spike protein is synthesized as a 1,273 amino acid polypeptide, and the frequency of amino acid variants, including both substitutions and deletions, at each of the positions is shown. These variants, relative to the Wuhan-Hu-1 reference sequence, were identified with use of CoV-GLUE<sup>96</sup>, which filters out Global Initiative on Sharing All Influenza Data (GISAID) sequences<sup>97</sup> identified as being of low quality or from non-human hosts (sequences retrieved from the GISAID database on 3 February 2021). Among 426,623 genomes after filtering, 5,106 different amino acid replacements or substitutions across 1,267 spike positions were identified, of which 320 at 259 positions were observed in at least 100 sequences. In addition to substitutions, several deletions have been observed,

particularly within the amino-terminal domain (NTD). The most frequently detected NTD deletion is the two-residue deletion at positions 69 and 70 ( $\Delta$ 69–70), present in 45,898 sequences. The S1–S2 boundary is at amino acid position 685. **b** | Spike protein monomer displaying an upright receptor-binding domain (RBD). **c** | Spike protein structure in the closed conformation overlaid with surface representations shown with a trimer axis vertical view (left) and an orthogonal top-down view along this axis (right). Domains are coloured as in part **a**. The RCSB Protein Data Bank IDs for the SARS-CoV-2 spike protein structures are 6ZGG and 6ZGE<sup>50</sup>. The magenta spheres represent glycans, and the magenta triangles represent potential N-linked glycosylation sites. The scissors represent the S1–S2 boundary at amino acid position 685. CD, connecting domain; CT, cytoplasmic tail; FP, fusion peptide; RBM, receptor-binding motif; TM, transmembrane domain.

## Box 2 | Mechanisms of antigenic change

In common with other virus surface glycoproteins responsible for attachment to host cell-surface receptors, such as influenza virus haemagglutinin and the envelope glycoprotein GP120 of HIV, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike glycoprotein is an important target for neutralizing antibodies. There are various distinct mechanisms by which mutations can alter the antigenic properties of a glycoprotein.

**Amino acid substitutions that alter the epitope**

A change in the biophysical properties of an epitope residue directly diminishes antibody binding. For example, the neutralizing antibody 4A8 forms salt bridges with spike protein residues K147 and K150, and therefore substitutions at these residues are likely to inhibit binding. The E484K amino acid substitution has received attention for its effect on monoclonal antibodies and convalescent plasma neutralizing activity. Its position has been described as belonging to the footprint of several antibodies, and a change in charge caused by replacement of a glutamate residue with a lysine residue has the potential to diminish antibody binding.

**Increasing receptor-binding avidity**

Substitutions that individually increase receptor-binding affinity can shift the binding equilibrium between glycoprotein and neutralizing antibodies in favour of a higher-avidity interaction between glycoprotein and the cellular receptor<sup>102</sup>. The spike amino acid substitution N501Y, which increases ACE2-binding affinity<sup>19</sup>, has been described as emerging in individuals treated with convalescent plasma, potentially as a means of immune escape.

**Changes in glycosylation**

Glycans are bulky sugar molecules that may shield epitopes from antibody binding. N-linked glycans are typically prominent in glycan shielding of virus surface glycoprotein epitopes<sup>33</sup>, although O-linked glycans can also contribute<sup>103</sup>. A substitution can introduce an additional N-linked glycosylation motif. The acquisition of epitope-masking glycans during the evolution of human influenza viruses is well described<sup>104</sup>.

**Deletions and insertions**

The deletion or insertion of residues has the potential to alter epitope conformation, diminishing antibody binding. Several deletions in the spike amino-terminal domain (NTD) that affect recognition by neutralizing antibodies have been described<sup>41,42</sup>. In laboratory experiments, a multiresidue insertion in the spike NTD has been described as emerging and contributing to escape from polyclonal antibodies in convalescent plasma<sup>41</sup>.

**Allosteric structural effects**

Similarly to deletions or insertions, an amino acid substitution outside an epitope footprint may affect antibody binding by changing the protein conformation in such a way that an epitope is altered or differently displayed. In the spike NTD, changes to disulfide bonds are thought to reduce binding by multiple monoclonal antibodies through this mechanism<sup>30</sup>.

Residue 501 is at the RBD–ACE2 interface (FIG. 2c), and N501Y has been shown experimentally to result in one of the highest increases in ACE2 affinity conferred by a single RBD mutation<sup>19</sup>. E484 has been identified as an immunodominant spike protein residue, with various substitutions, including E484K, facilitating escape from several mAbs<sup>40,47,48</sup> as well as antibodies in convalescent plasma<sup>39–41,48</sup>. E484K is estimated to have emerged repeatedly in the global SARS-CoV-2 population<sup>53</sup> and has been described in two other lineages originating from lineage B.1.1.28 in addition to lineage P.1 reported to be spreading in the state of Rio de Janeiro in Brazil (lineage P.2)<sup>56</sup> and in the Philippines (lineage P.3)<sup>57</sup>. Whereas K417 is described in the epitopes of RBD class 1 and class 2 antibodies<sup>31</sup>, alterations to K417 tend to affect class 1 antibody binding and are therefore somewhat less important for the polyclonal antibody response to the RBD, which is dominated by class 2 antibody responses, which are more susceptible to substitutions such as E484K<sup>44,58,59</sup>. In addition to their

antigenic effect, both K417N and K417T are expected to moderately decrease ACE2-binding affinity<sup>19</sup> (FIG. 5b). The  $\Delta$ H69–V70 deletion has been identified in variants associated with immune escape in immunocompromised individuals treated with convalescent plasma<sup>24</sup>. Experiments have shown that  $\Delta$ H69–V70 does not reduce neutralization by a panel of convalescent sera; however, it may compensate for infectivity deficits associated with affinity-boosting RBM mutations that may emerge due to immune-mediated selection<sup>22</sup>.

The first genomes belonging to the B.1.1.7 lineage were sequenced in the south of England in September 2020. Lineage B.1.1.7 is defined by the presence of 23 nucleotide mutations across the genome that map to a single branch of the phylogenetic tree<sup>3</sup>. Of these 23 mutations, 14 encode amino acid changes and three are deletions, including six amino acid substitutions in the spike protein (N501Y, A570D, P681H, T716I, S982A and D1118H) and two NTD deletions ( $\Delta$ H69–V70 and  $\Delta$ Y144)<sup>3</sup>. The lineage has been associated with a rapidly increasing proportion of reported SARS-CoV-2 cases, and phylogenetic analyses indicate that this lineage was associated with a growth rate estimated to be 40–70% higher than that of other lineages<sup>60,61</sup>. There is also evidence that this lineage may be associated with a higher viral load<sup>62</sup>. In addition to N501Y, for which there is some evidence that it reduces neutralization by a small proportion of RBD antibodies<sup>63</sup>, there is evidence for an antigenic effect of  $\Delta$ Y144 (FIG. 5b). This deletion is expected to alter the conformation of the N3 NTD loop (amino acid positions 140–156) and has been demonstrated to abolish neutralization by a range of neutralizing antibodies<sup>30</sup>. The B.1.1.7 spike mutations have been shown to diminish neutralization of a higher proportion of NTD-specific neutralizing antibodies (9 of 10; 90%) than RBD-specific neutralizing antibodies (5 of 31; 16%)<sup>63</sup>. Given the immunodominance of the RBD, this could explain the modest reductions in neutralizing activity of convalescent sera against authentic B.1.1.7 or pseudoviruses carrying the B.1.1.7 spike mutations<sup>64,65</sup>. The co-occurrence of  $\Delta$ Y144 and E484K is concerning with respect to the polyclonal antibody response as the N3 loop, which  $\Delta$ Y144 changes, is predicted to be among the most immunogenic regions of the spike protein (FIG. 2a), and amino acid substitutions at position 484 diminish neutralization by a range of RBD-targeting antibodies.

Reports of lineages with N501Y circulating in the UK were followed by reports of another lineage possessing N501Y circulating in South Africa (lineage B.1.351), which has been rapidly expanding in frequency since December 2020 (REF. 66). In addition to N501Y, lineage B.1.351 is defined by the presence of five further spike amino acid substitutions (D80A, D215G, K417N, E484K and A701V) and a deletion in the NTD,  $\Delta$ 242–244. High numbers of B.1.351 viruses also have the spike amino acid substitutions L18F, R246I and D614G. A similar NTD deletion,  $\Delta$ 243–244, abolishes binding by the antibody 4A8 (REF. 42), and L18F and R246I also occur within the NTD supersite and likely affect antibody binding as well<sup>30</sup>. The co-occurrence of K417N, E484K and these NTD substitutions suggests that lineage B.1.351 may overcome the polyclonal antibody response by reducing

**Avidity**

Also referred to as functional affinity, the accumulated binding strength of multiple affinities of individual interactions, such as between a virus receptor-binding site and its cellular receptor.

**IC<sub>50</sub>**

The half-maximal inhibitory concentration, a quantitative measure that indicates how much of an inhibitory substance (for example, postvaccination serum) is required to inhibit a biological process (for example, virus forming plaques or regions of infected cells in culture) by 50%.

neutralization by class 1 and class 2 RBD-specific antibodies and NTD-specific antibodies (FIG. 5b). Data reported in one study showed that nearly half of examined convalescent plasma samples (21 of 44; 48%) had no detectable neutralization activity against the B.1.351 variant<sup>58</sup>. Other experiments with pseudotyped viruses showed that the B.1.351 variant was also resistant to the neutralizing activity of some mAbs (12 of 17; 70%)<sup>67</sup>. Similarly, a study showed that the neutralizing effect of convalescent plasma collected from 14 individuals was strongly reduced against a live (authentic) B.1.351 virus (with IC<sub>50</sub> reduced by 3.2-fold to 41.9-fold relative to the first-wave virus)<sup>68</sup>.

The P.1 lineage, a sublineage of B.1.1.28, was first detected in Brazil<sup>69</sup> and in travellers from Brazil to Japan<sup>70</sup>. Lineage P.1 is characterized by the presence of several amino acid substitutions in the spike protein: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y and T1027I<sup>69</sup>. In addition to substitutions at positions 417, 484 and 501 discussed above, the P.1 lineage has a cluster of substitutions close to the described antigenic regions of the NTD, including L18F, which is known to reduce neutralization by some antibodies<sup>30</sup>. The substitutions T20N and P26S also occur in or near the NTD supersite<sup>30</sup> at residues with high antibody accessibility scores (FIG. 5b), and T20N introduces a potential glycosylation site that could result in glycan shielding (BOX 2) of part of the supersite. The P.1 lineage has also been associated with a reinfection case in Manaus, Brazil<sup>71</sup>, indicating it is contributing to antigenic circumvention of what might have been an otherwise effective immune response. Analyses integrating genomic and mortality data estimate that P.1 may be 1.7 to 2.4-fold more transmissible and that previous infection by non-P.1 SARS-CoV-2 provides 54–79% of the protection against P.1 infection compared with non-P.1 lineages<sup>71</sup>. More details of the frequency and geographic distribution of the P.1 lineage can be found at the Pango lineages website<sup>72</sup>.

Increasingly, lineages possessing independent occurrences of mutations in common with the variants of concern B.1.1.7, B.1.351 and P.1 are being detected, demonstrating convergent evolution. For example, viruses of lineage B.1.525, which has been observed in several countries, albeit at low frequency to date, have NTD deletions ΔH69–V70 and ΔY144 in common with viruses of the B.1.1.7 lineage; E484K in common with the B.1.351 and P.1 lineages; and spike amino acid substitutions Q52R, Q677H and F888L<sup>73</sup>. Repeated amino acid substitutions at position 677 and the independent emergence of Q677H in several lineages in the USA provides strong evidence of adaptation, potentially through an effect of this mutation on the proximal polybasic furin cleavage site, although further experiments are required to determine its impact<sup>74</sup>. Other novel variants have been identified spreading in California and New York, USA (B.1.427 and B.1.429, and B.1.526, respectively). The B.1.427 and B.1.429 variants carry an antigenically noteworthy substitution, L452R<sup>75</sup>, which has been shown to reduce neutralization by several mAbs<sup>43,45,48,59</sup> and convalescent plasma<sup>43</sup>. L452R independently appeared in several other lineages around the globe

between December 2020 and February 2021, indicating that this amino acid substitution is probably the result of viral adaptation due to increasing immunity in the population<sup>75</sup>. L452R is also present in the A.27 lineage associated with a cluster of cases identified on the island of Mayotte<sup>76</sup>. The lineage B.1.526 has been found to carry either S477N or E484K, among other lineage-defining mutations<sup>77,78</sup>, both of which were described as antigenically important above. A new variant carrying E484K currently designated A.VOI.V2 was recently identified in Angola from cases involving travel from Tanzania<sup>79</sup>. This variant carries several amino acid substitutions in the spike protein and three deletions in the NTD, some of which are within the antigenic supersite<sup>79</sup>. Another variant within the A lineage, the prevalence of which is rising in Uganda (A.23.1), shares with the B.1.1.7 lineage a substitution at position 681 within the furin cleavage site (P681R has been found in the A lineage, whereas P681H has been found in the B.1.1.7 lineage), and additionally has the amino acid substitutions R102I, F157L, V367F and Q613H. Q613H is speculated to be important as it occurs at a position neighbouring D614G<sup>30</sup>. Amino acid position 157 has been identified as an epitope residue, with F157A reducing neutralization by the mAb 2489 (REF.<sup>34</sup>).

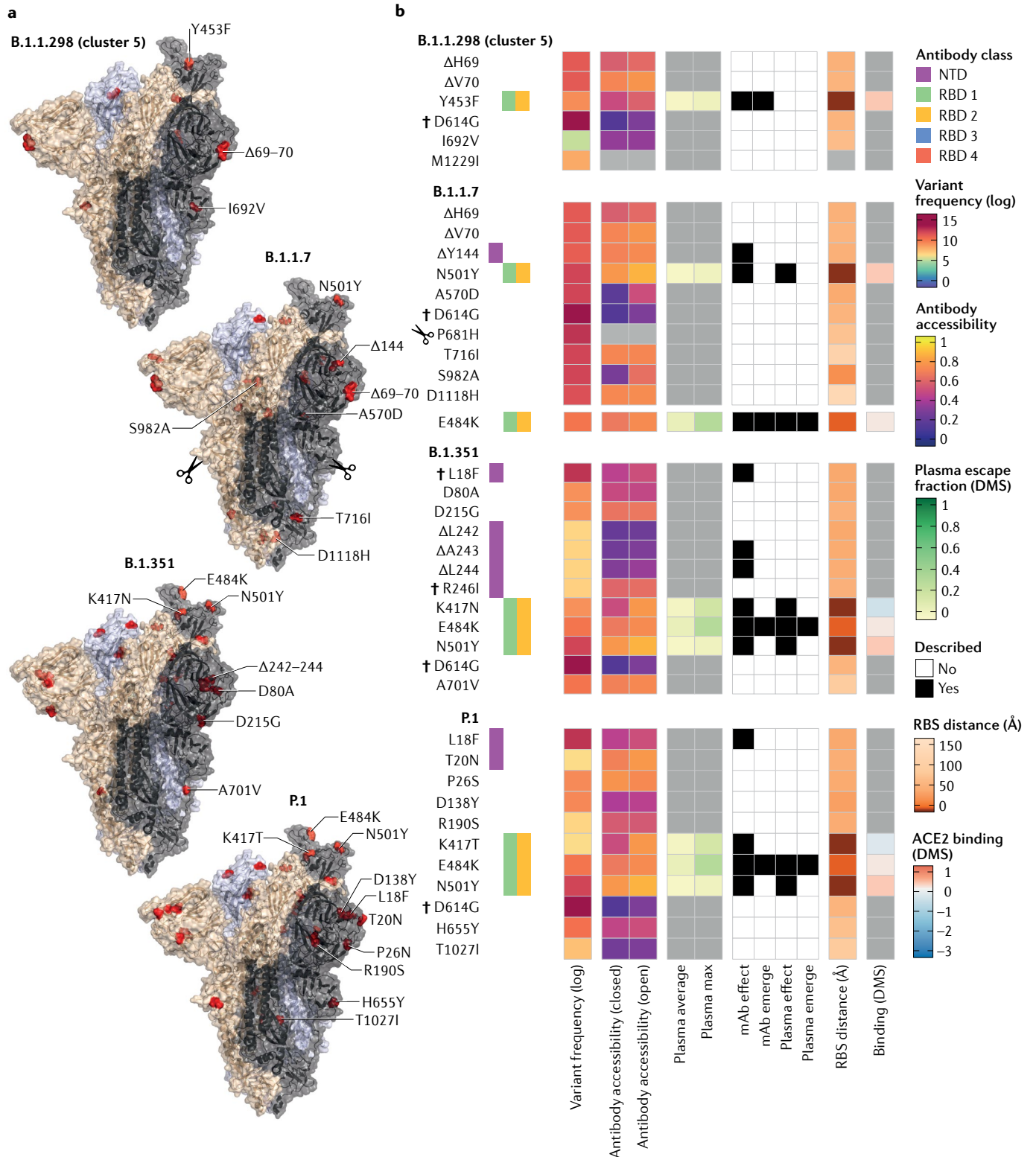
New variants will continue to emerge, and although it is important to understand the phenotypes of emerging variants in terms of infectivity, transmissibility, virulence and antigenicity, it is also important to quantify the phenotypic impacts of specific mutations present in variants, both individually and in combination with other mutations. As new variants with unforeseen combinations of mutations continue to emerge, such insights will allow predictions of virus phenotype. For example, recently detected viruses of lineage B.1.617.1 were anticipated to show altered antigenicity due to the presence of the substitutions L452R and E484Q, which have been described as affecting antibody recognition<sup>39,43,45,48,81</sup>. Moving forwards, the experimental characterization of SARS-CoV-2 spike mutations to date will continue to provide extremely useful information on individual mutations or combinations of mutations that may not yet have been seen in circulating viruses.

### Vaccine efficacy against new variants

To date, vaccines have been licensed and rolled out very successfully in several countries, but the number of individuals vaccinated still represents a small fraction of the global population (Supplementary Table 1). To assess the impacts of mutations on vaccine efficacy, authentic viruses and pseudoviruses possessing particular spike mutations (either individually or in combination) and larger sets of mutations representing variants of concern and other circulating spike mutations have been assessed by neutralization assays with postvaccination sera (Supplementary Table 1). Typically, studies report a fold change in variant virus, or pseudovirus, neutralization relative to wild-type virus (the serum concentration at which 50% neutralization (IC<sub>50</sub>) is achieved with the variant divided by the average IC<sub>50</sub> for the wild-type virus).

Postvaccination sera from a cohort of 20 volunteers immunized with the mRNA vaccine mRNA-1273





(Moderna) or BNT162b2 (Pfizer–BioNTech) showed high binding titres for anti-SARS-CoV-2 spike IgM and IgG with plasma neutralizing activity and relative numbers of RBD-specific antibodies equivalent to those in natural infection<sup>59</sup>. Furthermore, epitope mapping of mAbs isolated from postvaccination sera showed they targeted a range of RBD epitopes similar to those isolated from naturally infected individuals<sup>59</sup>. The plasma

neutralizing activity and the numbers of RBD-specific memory B cells were found to be equivalent to those of plasma from individuals who had recovered from natural SARS-CoV-2 infection<sup>59</sup>. Investigations with pseudoviruses possessing RBD mutations carried by variants of concern demonstrated that the neutralizing activity of plasma from vaccinated individuals showed a small but significant decreases of onefold to threefold against

◀ Fig. 5 | **Amino acid mutations that characterize variants of concern.** **a** | Spike heterotrimer in the open conformation overlaid with the surface representation (RCSB Protein Data Bank ID 6ZGC<sup>30</sup>). The locations of amino acid substitutions and deletions that define variants of concern are highlighted as red spheres. For B.1.1.7, scissors mark the approximate position of substitution P681H within the furin cleavage site, which is absent from the structural model. **b** | Aligned heat maps showing properties of amino acid residues or of the specific amino acid substitution, as appropriate. Epitope residues are coloured to indicate the amino-terminal domain (NTD) or the receptor-binding domain (RBD) class<sup>30</sup>. Structure-based antibody access scores for the spike protein in the closed and open conformations are shown. For RBD residues, the results of deep mutational scanning (DMS) studies show the escape fraction (that is, a quantitative measure of the extent to which a mutation reduced polyclonal antibody binding) for each mutant averaged across plasma ('plasma average') and for the most sensitive plasma ('plasma max')<sup>39</sup>. Each mutation is classified as having evidence for mutations affecting neutralization by either monoclonal antibodies (mAbs) or antibodies in convalescent plasma<sup>39</sup> or vaccinated individuals<sup>59</sup>, and emerging in selection experiments using mAbs<sup>40,47,48</sup> or post-infection serum<sup>40,47,48</sup>. The distance to the ACE2-contacting residues that form the receptor-binding site RBS is shown (for residue 681, this is estimated with use of the nearest residues present in published structures). DMS data on ACE2-binding affinity<sup>19</sup> are shown by aggregation of scores and averaging across each mutant at a residue and alternatively the maximally binding mutant. Scores represent the binding constant ( $\Delta\log_{10} K_D$ ) relative to the wild-type reference amino acid. Mutations that are present in a variant but that are also widespread in the virus population in which a variant emerged, or exhibit high diversity within a lineage, are marked with a dagger.

E484K, N501Y or the K417N + E484K + N501Y triple mutant<sup>59</sup>. Other data indicate that the effect of N501Y alone on neutralization is relatively modest, and other studies using sera from 20 participants in a trial of the BNT162b2 vaccine showed neutralizing titres equivalent to those of pseudoviruses carrying the N501 and Y501 mutations<sup>82</sup>. Other investigations with recombinant viruses carrying N501Y,  $\Delta$ H69–V70 + N501Y + D614G or E484K + N501Y + D614G demonstrated that compared with the Wuhan-Hu-1 reference virus, only E484K + N501Y + D614G resulted in a small and modest reduction in neutralization by postvaccination sera elicited by two BNT162b2 doses, and only modest differences in neutralization were seen compared with the Wuhan-Hu-1 reference virus<sup>83</sup>.

As stated earlier, convalescent plasma from individuals infected with pre-B.1.1.7 viruses (that is, viruses that circulated before the emergence of the B.1.1.7 lineage) shows only a modest reduction in neutralization activity against B.1.1.7 or pseudovirus possessing B.1.1.7 spike mutations<sup>63,78</sup>, and results obtained with postvaccination sera are broadly consistent with this. Pseudoviruses carrying the set of B.1.1.7 spike mutations evaluated with postvaccination serum from individuals who received the BNT162b2 vaccine (two doses)<sup>63,78,84</sup> or mRNA-1273 vaccine (two doses)<sup>63</sup> exhibited only a modest reduction in neutralization titres (less than threefold). However, assays using pseudovirus carrying B.1.1.7 spike mutations and with the addition of E484K, a combination that has been observed in sequencing of circulating isolates, showed larger, more significant drops (6.7-fold) in neutralization with postvaccination sera isolated from individuals who received the BNT162b2 vaccine<sup>85</sup>. In a live-virus neutralization assay, neutralizing titres of ChAdOx1 nCoV-19 (Oxford–AstraZeneca) postvaccination sera were nine times lower than titres against the B.1.1.7 lineage relative to a canonical non-B.1.1.7 lineage (Wuhan-Hu-1 with the S247R spike mutation)<sup>86</sup>.

Similarly, neutralizing activity of sera elicited by the inactivated vaccine Covaxin (Bharat Biotech) against B.1.1.7 viruses was largely preserved<sup>87</sup>. Pseudovirus and live-virus neutralization assays showed that the neutralizing activity of sera from individuals after the two doses of the ChaAdOx1 vaccine against the B.1.351 variant was reduced or abrogated<sup>86</sup>. Postvaccination sera from individuals who received two doses of mRNA-1273 (28 days apart) showed reduced neutralization of the B.1.351 variant (6.4-fold reduction)<sup>88</sup>. By contrast, neutralizing activity of sera elicited by the inactivated vaccine BBIBP-CorV (Sinopharm) against the authentic virus B.1.351 showed only a slight reduction (less than twofold)<sup>89</sup>.

Comparison of the differing extents to which variants affect neutralization by postvaccination serum is complicated by the different methods used in various studies. However, one study tested eight SARS-CoV-2 variants of interest or concern, including B.1.1.298, B.1.1.7 and P.1, as well as three B.1.351 variants, distinguished by their combination of NTD mutations, representing sequence diversity in circulating viruses of this lineage. Pseudoviruses were generated by the same system and tested with postvaccination sera from individuals who received two doses of either the BNT162b2 vaccine ( $n = 30$ ) or the mRNA-1273 vaccine ( $n = 35$ )<sup>90</sup>. Compared with wild type, pseudoviruses with D614G or the mutations defining lineages B.1.1.7, B.1.1.298 and B.1.429 each showed non-statistically significant decreases in neutralization<sup>90</sup>. Lineages P.1 and P.2 each showed significant decreases, with both BNT162b2 (6.7-fold and 5.8-fold, respectively) and mRNA-1273 (4.5-fold and 2.9-fold, respectively) postvaccination sera<sup>90</sup>. The three B.1.351 variants investigated, representing the majority of deposited B.1.351 sequences, showed much larger decreases in neutralization activity, ranging from 34-fold to 42-fold (BNT162b2) and from 19.2-fold to 27.7-fold (mRNA-1273). Taken together, these data indicate that E484K is the primary determinant of the decreases in neutralization titres, which distinguish P.1, P.2 and the three B.1.351 variants from the other pseudoviruses tested. In addition to E484K, further mutations that are shared by each of the three B.1.351 variants, but are not possessed by the P.1 and P.2 lineages, are D80A,  $\Delta$ 242–244, K417N (though K417T is present in P.1) and A701V.

To complement the experimental data provided by neutralization assays, there is emerging evidence from clinical trials on the impact of variants on vaccine efficacy. Early indications suggest that these are broadly consistent with the laboratory results, with the B.1.351 variant showing greater signs of vaccine escape. The ChAdOx1 nCoV-19 vaccine showed clinical efficacy against the B.1.1.7 variant but failed to provide protection against mild to moderate disease caused by the B.1.351 variant, with vaccine efficacy against the variant estimated at 10.4% (95% confidence interval –76.8 to 54.8)<sup>85,86,91</sup>. Preliminary data from clinical trials reported that the NVX-CoV2373 (Novavax) protein-based vaccine provides 95.6% efficacy against the wild-type virus and that this is moderately lower for the B.1.1.7 variant (85.6%) and is further reduced for the B.1.351 variant (60.0%)<sup>91</sup>. Similarly, the single-dose vaccine JNJ-78436735 (Johnson & Johnson/Janssen) has been

shown to provide 72% protection against moderate to severe SARS-CoV-2 infections in the USA, but the proportion significantly decreased to 57% in South Africa (at a time when the B.1.351 variant was widespread)<sup>92</sup>. These data indicate that NVX-CoV2373 and JNJ-78436735 are clinically efficacious against the B.1.1.7 variant and variants circulating in the USA, and are consistent in that the B.1.351 variant is associated with a larger reduction in vaccine efficacy.

In addition to evaluation of vaccine efficacy against SARS-CoV-2 variants and mutations, the effects of mutations on some mAbs used as therapeutics have been described (Supplementary Table 2). Single mAb treatment can exert a selective pressure that potentially increases the possibility of mutational escape of the targeted antigen. The risk is likely to be reduced with the use of cocktails of two or more mAbs targeting non-overlapping epitopes. REGN-COV2 (Regeneron) (included in the RECOVERY trial in the UK) and AZD7742 (AstraZeneca) are two examples of mAbs cocktails that have been developed<sup>93</sup>. Importantly, some mutations in the RBM have already been identified in variants which are circulating in the UK (for example, N439K, T478I and V483I) and are likely to impact antigenicity.

### Conclusions

There is now clear evidence of the changing antigenicity of the SARS-CoV-2 spike protein and of the amino acid changes that affect antibody neutralization. Spike amino acid substitutions and deletions that impact neutralizing antibodies are present at significant frequencies in the global virus population, and there is emerging evidence of variants exhibiting resistance to antibody-mediated immunity elicited by vaccines. Greater understanding of the correlates of immune protection is required to provide a context for the results of studies reporting changes in neutralization. A comprehensive understanding of the consequences of spike mutations for antigenicity will encompass both T cell-mediated immunity and non-spike epitopes recognized by antibodies. To monitor vaccine efficacy and to better understand the implications of antigenic variation for vaccine effectiveness, it will be important to collect information on vaccine status and viral genome sequence data from individuals infected with SARS-CoV-2. More generally, a broader understanding of the phenotypic impacts of mutations across the SARS-CoV-2 genome and their consequences for variant fitness will help elucidate drivers of transmission and evolutionary success.

Recent studies have shown the potential selective pressure exerted by convalescent plasma and mAb treatments on SARS-CoV-2 evolution in immunocompromised individuals<sup>24–26</sup>. Such circumstances, involving long-term virus shedders, may have contributed to the sporadic emergence of the more heavily mutated variants

(for example, seen in the B.1.1.7 and B.1.351 lineages). Given that therapeutics (vaccines and antibody-based therapies) target mainly the SARS-CoV-2 spike protein, the selection pressures that favour the emergence of new variants carrying immune escape mutations generated in chronic infections<sup>24–26</sup> will be similar to those selecting for mutations that allow reinfections within the wider population<sup>27–29</sup>. Therefore, sequencing of viruses associated with prolonged infections will provide useful information on mutations that could contribute to increased transmissibility or escape from vaccine-mediated immunity.

The collective data on the effect of mutations on vaccines and convalescent serum efficacy show that the polyclonal antibody response is focused on a few immunodominant regions, indicating the high probability of future mutation-mediated escape from host immunity. As antigenically different variants are continuing to emerge, it will become necessary to routinely collect serum samples from vaccinated individuals and from individuals who have been infected with circulating variants of known sequence. Cross-reactive immunity between circulating lineages can be assessed by measuring the ability of sera to neutralize panels of circulating viruses. The systematic surveillance of antigenic SARS-CoV-2 variants will be enhanced by the establishment of a network similar to the WHO-coordinated Global Influenza Surveillance and Response System (GISRS), a collaborative global effort responsible for tracking the antigenic evolution of human influenza viruses and making recommendations on vaccine composition. Modelling approaches to predict the evolutionary trajectories of emerging variants based on an understanding of the phenotypic effects of mutations will assist this process, as is the case for influenza virus<sup>94</sup>.

Prediction of the mutational pathways by which a virus such as SARS-CoV-2 will evolve is extremely challenging. Nonetheless, there is a rapidly expanding knowledge base regarding the effect of SARS-CoV-2 spike mutations on antigenicity and other aspects of virus biology. The integration of these data and emerging SARS-CoV-2 sequences has the potential to facilitate the automated detection of potential variants of concern at low frequency (that is, before they are spreading widely). Tracking the emergence of these viruses flagged as potential antigenically significant variants will help to guide the implementation of targeted control measures and further laboratory characterization. An important part of this process will be the preparation of updated vaccines tailored to emerging antigenic variants that are maximally cross-reactive against all circulating variants. All of these processes will benefit from close international collaboration and the rapid and open sharing of data.

Published online: 01 June 2021

1. WHO. Coronavirus (COVID-19) Dashboard. <https://covid19.who.int/> (2021).
2. Meredith, L. W. et al. 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**, 1263–1272 (2020).
3. Rambaut, A. et al. *Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel set of Spike Mutations*. <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563> (2020).
4. Frost, S. D. W., Magalis, B. R. & Kosakovsky Pond, S. L. Neutral theory and rapidly evolving viral pathogens. *Mol. Biol. Evol.* **35**, 1348–1354 (2018).
5. MacLean, O. A., Orton, R. J., Singer, J. B. & Robertson, D. L. No evidence for distinct types in the evolution of SARS-CoV-2. *Virus Evol.* <https://doi.org/10.1093/ve/veaa034> (2020).



6. Zhan, X.-Y. et al. Molecular evolution of SARS-CoV-2 structural genes: evidence of positive selection in spike glycoprotein. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.06.25.170688> (2020).
7. Korber, B. et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812–827 e819 (2020).
8. Hou, Y. J. et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464 (2020).
9. Yurkovetskiy, L. et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell* **183**, 739–751 e738 (2020).
10. Volz, E. et al. 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> (2020).
11. Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).
12. Piccoli, L. et al. Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* **183**, 1024–1042 e1021 (2020).
13. Liu, L. et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450–456 (2020).
14. Dai, L. & Gao, G. F. Viral targets for vaccines against COVID-19. *Nat. Rev. Immunol.* **21**, 73–82 (2021).
15. Duchene, S. et al. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol.* <https://doi.org/10.1093/ve/vaa061> (2020).
16. Worobey, M. et al. The emergence of SARS-CoV-2 in Europe and North America. *Science* **370**, 564–570 (2020).
17. Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
18. Thomson, E. C. et al. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **184**, 1171–1187 e1120 (2021).
19. Starr, T. N. et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310 e1220 (2020).
20. Fonager J. et al. Working paper on SARS-CoV-2 spike mutations arising in Danish mink, their spread to humans and neutralization data. [https://files.ssi.dk/Mink-cluster-5-short-report\\_AFO2](https://files.ssi.dk/Mink-cluster-5-short-report_AFO2) (2020).
21. ECDC. *Detection of new SARS-CoV-2 Variants Related to Mink*. <https://www.ecdc.europa.eu/sites/default/files/documents/RRA-SARS-CoV-2-in-mink-12-nov-2020.pdf> (2020).
22. Kemp, S. et al. Recurrent emergence and transmission of a SARS-CoV-2 spike deletion ΔH69/N70. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.12.14.422555> (2020).
23. Martin, D. P. et al. The emergence and ongoing convergent evolution of the N501Y lineages coincides with a major global shift in the SARS-CoV-2 selective landscape. Preprint at *medRxiv* <https://doi.org/10.1101/2021.02.23.21252268> (2021).
24. Kemp, S. A. et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* <https://doi.org/10.1038/s41586-021-03291-y> (2021).
25. Choi, B. et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N. Eng. J. Med.* **383**, 2291–2293 (2020).
26. Avanzato, V. A. et al. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. *Cell* **183**, 1901–1912 e1909 (2020).
27. Naveca, F. et al. SARS-CoV-2 Reinfection by the New Variant of Concern (VOC) P.1 in Amazonas, Brazil. <https://virological.org/t/sars-cov-2-reinfection-by-the-new-variant-of-concern-voc-p-1-in-amazonas-brazil/596> (2021).
28. Nonaka, C. K. V. et al. *Genomic Evidence of a SARS-CoV-2 Reinfection Case with E484K Spike Mutation in Brazil*. <https://www.preprints.org/manuscript/202101.0132/v1> (2021).
29. Resende, P. C. et al. *Spike E484K Mutation in the First SARS-CoV-2 Reinfection Case Confirmed in Brazil, 2020*. <https://virological.org/t/spike-e484k-mutation-in-the-first-sars-cov-2-reinfection-case-confirmed-in-brazil-2020/584> (2021).
30. McCallum, M. et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* <https://doi.org/10.1016/j.cell.2021.03.028> (2021).
31. Barnes, C. O. et al. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020).
32. Chi, X. et al. A neutralizing human antibody binds to the N-terminal domain of the spike protein of SARS-CoV-2. *Science* **369**, 650 (2020).
33. Watanabe, Y. et al. Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat. Commun.* **11**, 2688 (2020).
34. Suryadevara, N. et al. Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell* <https://doi.org/10.1016/j.cell.2021.03.029> (2021).
35. Soh, W. T. et al. The N-terminal domain of spike glycoprotein mediates SARS-CoV-2 infection by associating with L-SIGN and DC-SIGN. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.11.05.369264> (2020).
36. Casalino, L. et al. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. *ACS Cent. Sci.* **6**, 1722–1734 (2020).
37. Shrock, E. et al. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* **370**, eabd4250 (2020).
38. Zheng, Z. et al. Monoclonal antibodies for the S2 subunit of spike of SARS-CoV-1 cross-react with the newly-emerged SARS-CoV-2. *Eurosurveillance* **25**, 2000291 (2020).
39. Greaney, A. J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476 e466 (2021).
40. Weisblum, Y. et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* <https://doi.org/10.7554/eLife.61312> (2020).
41. Andreano, E. et al. SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.12.28.424451> (2020).
42. McCarthy, K. R. et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* **371**, 1139–1142 (2021).
43. Li, Q. et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* **182**, 1284–1294 e1289 (2020).
44. Greaney, A. J. et al. Mutational escape from the polyclonal antibody response to SARS-CoV-2 infection is largely shaped by a single class of antibodies. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.17.435863> (2021).
45. Greaney, A. J. et al. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **29**, 44–57 e49 (2021).
46. Starr, T. N. et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850 (2021).
47. Baum, A. et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* <https://doi.org/10.1126/science.abd0831> (2020).
48. Liu, Z. et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* **29**, 477–488 e474 (2021).
49. Sweredoski, M. J. & Baldi, P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics* **24**, 1459–1460 (2008).
50. Wrobel, A. G. et al. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* **27**, 763–767 (2020).
51. Weissman, D. et al. D614G spike mutation increases SARS-CoV-2 susceptibility to neutralization. *Cell Host Microbe* **29**, 23–31 e24 (2021).
52. Hodcroft, E. B. et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. Preprint at *medRxiv* <https://doi.org/10.1101/2020.10.25.20219063> (2020).
53. Nextstrain. *Genomic epidemiology of novel coronavirus - Global subsampling*. [https://nextstrain.org/ncov/global?c=gt-S\\_477&gmax=24271&gmin=22482](https://nextstrain.org/ncov/global?c=gt-S_477&gmax=24271&gmin=22482) (2021).
54. Zahradnik, J. et al. SARS-CoV-2 RBD in vitro evolution follows contagious mutation spread, yet generates an able infection inhibitor. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.01.06.425392> (2021).
55. Wise, J. Covid-19: the E484K mutation and the risks it poses. *Br. Med. J.* **372**, n359 (2021).
56. Voloch, C. M. et al. Genomic characterization of a novel SARS-CoV-2 lineage from rio de Janeiro, Brazil. *J. Virol.* **95**, e00119-21 (2021).
57. Tablizo, F. A. et al. Genome sequencing and analysis of an emergent SARS-CoV-2 variant characterized by multiple spike protein mutations detected from the Central Visayas Region of the Philippines. Preprint at *medRxiv* <https://doi.org/10.1101/2021.03.03.21252812> (2021).
58. Wibmer, C. K. et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* **27**, 622–625 (2021).
59. Wang, Z. et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **592**, 616–622 (2021).
60. ECDC. *Risk Related to Spread of New SARS-CoV-2 Variants of Concern in the EU/EEA*. <https://www.ecdc.europa.eu/sites/default/files/documents/COVID-19-risk-related-to-spread-of-new-sars-cov-2-variants-EU-EEA.pdf> (2020).
61. Volz, E. et al. Transmission of SARS-CoV-2 lineage B.1.1.7 in England: insights from linking epidemiological and genetic data. Preprint at *medRxiv* <https://doi.org/10.1101/2020.12.30.20249034> (2021).
62. Kidd, M. et al. S-variant SARS-CoV-2 is associated with significantly higher viral load in samples tested by TaqPath polymerase chain reaction. *J. Infect. Dis.* <https://doi.org/10.1093/infdis/jiab082> (2021).
63. Collier, D. A. et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* <https://doi.org/10.1038/s41586-021-03412-7> (2021).
64. Rees-Spear, C. et al. The impact of spike mutations on SARS-CoV-2 neutralization. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.01.15.426849> (2021).
65. Hu, J. et al. Emerging SARS-CoV-2 variants reduce neutralization sensitivity to convalescent sera and monoclonal antibodies. *Cell Mol. Immunol.* **18**, 1061–1063 (2021).
66. Tegally, H. et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. Preprint at *medRxiv* <https://doi.org/10.1101/2020.12.21.20248640> (2020).
67. Li, Q. et al. No higher infectivity but immune escape of SARS-CoV-2 501Y.V2 variants. *Cell* <https://doi.org/10.1016/j.cell.2021.02.042> (2021).
68. Cele, S. et al. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature* <https://doi.org/10.1038/s41586-021-03471-w> (2021).
69. Faria, N. R. et al. *Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in Manaus: Preliminary Findings*. <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586> (2021).
70. Naveca, F. et al. *Phylogenetic Relationship of SARS-CoV-2 Sequences from Amazonas with Emerging Brazilian Variants Harboring Mutations E484K and N501Y in the Spike Protein*. <https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585> (2021).
71. Faria, N. R. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815–821 (2021).
72. Rambaut, A. et al. *Global Report Investigating Novel Coronavirus Haplotypes*. [https://cov-lineages.org/global\\_report.html](https://cov-lineages.org/global_report.html) (2020).
73. O'Toole, A. *Proposal for New Lineage within B.1 #4: B.1.525, cov-lineages/pango-designation*. <https://github.com/cov-lineages/pango-designation/issues/4> (2021).
74. Hodcroft, E. B. et al. Emergence in late 2020 of multiple lineages of SARS-CoV-2 spike protein variants affecting amino acid position 677. Preprint at *medRxiv* <https://doi.org/10.1101/2021.02.12.21251658> (2021).
75. Tchesnokova, V. et al. Acquisition of the L452R mutation in the ACE2-binding interface of spike protein triggers recent massive expansion of SARS-CoV-2 variants. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.02.22.432189> (2021).
76. Wagner, C., Hodcroft, E., Bell, S. M., Neher, R. & Bedford, T. *Resurgence of SARS-CoV-2 19B Clade Corresponds with Possible Convergent Evolution*. <https://virological.org/t/resurgence-of-sars-cov-2-19b-clade-corresponds-with-possible-convergent-evolution/620> (2021).

77. Annavajhala, M. K. et al. A novel SARS-CoV-2 variant of concern, B.1.526, identified in New York. Preprint at *medRxiv* <https://doi.org/10.1101/2021.02.23.21252259> (2021).
78. West, A. P., Barnes, C. O., Yang, Z. & Bjorkman, P. J. SARS-CoV-2 lineage B.1.526 emerging in the New York region detected by software utility created to query the spike mutational landscape. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.02.14.431043> (2021).
79. de Oliveira, T. et al. A Novel Variant of Interest of SARS-CoV-2 with Multiple Spike Mutations Detected Through Travel Surveillance in Africa. <https://www.krisp.org.za/publications.php?pubid=330> (2021).
80. Bugembe, D. L. et al. A SARS-CoV-2 lineage A variant (A.23.1) with altered spike has emerged and is dominating the current Uganda epidemic. Preprint at *medRxiv* <https://doi.org/10.1101/2021.02.08.21251393> (2021).
81. Cherian, S. et al. Convergent evolution of SARS-CoV-2 spike mutations, L452R, E484Q and P681R, in the second wave of COVID-19 in Maharashtra, India. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.04.22.440932> (2021).
82. Xie, X. et al. Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.01.07.425740> (2021).
83. Xie, X. et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K, and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* <https://doi.org/10.1038/s41591-021-01270-4> (2021).
84. Faulkner, N. et al. Reduced antibody cross-reactivity following infection with B.1.1.7 than with parental SARS-CoV-2 strains. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.01.433314> (2021).
85. Emary, K. R. W. et al. Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/01 (B.1.1.7): an exploratory analysis of a randomised controlled trial. *Lancet* [https://doi.org/10.1016/S0140-6736\(21\)00628-0](https://doi.org/10.1016/S0140-6736(21)00628-0) (2021).
86. Madhi, S. A. et al. Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa2102214> (2021).
87. Sapkal, G. N. et al. Neutralization of UK-variant VUI-202012/01 with COVAXIN vaccinated human serum. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.01.26.426986> (2021).
88. Wu, K. et al. mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.01.25.427948> (2021).
89. Huang, B. et al. Neutralization of SARS-CoV-2 VOC 501Y.V2 by human antisera elicited by both inactivated BBIBP-CorV and recombinant dimeric RBD ZF2001 vaccines. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.02.01.429069> (2021).
90. Garcia-Beltran, W. F. et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* <https://doi.org/10.1016/j.cell.2021.03.013> (2021).
91. Mahase, E. Covid-19: Novavax vaccine efficacy is 86% against UK variant and 60% against South African variant. *Br. Med. J.* **372**, n296 (2021).
92. Mahase, E. Covid-19: Where are we on vaccines and variants? *Br. Med. J.* **372**, n597 (2021).
93. Matthews, D. B. A cocktail of antibodies for COVID-19 therapy. *Nat. Rev. Immunol.* **20**, 591 (2020).
94. Morris, D. H. et al. Predictive modeling of influenza shows the promise of applied evolutionary biology. *Trends Microbiol.* **26**, 102–118 (2018).
95. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
96. Singer, J., Gifford, R., Cotten, M. & Robertson, D. L. *CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation*. <http://cov-glue.cvr.gla.ac.uk/#/home> (2020).
97. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance* **22**, 30494 (2017).
98. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330 (2020).
99. Kumar, S., Maurya, V. K., Prasad, A. K., Bhatt, M. L. B. & Saxena, S. K. Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). *Virusdisease* **31**, 13–21 (2020).
100. Hoffmann, M., Kleine-Weber, H. & Pohlmann, S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol. Cell* **78**, 779–784 e775 (2020).
101. Benton, D. J. et al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **588**, 327–330 (2020).
102. Hensley, S. E. et al. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science* **326**, 734–736 (2009).
103. Silver, Z. A. et al. Discovery of O-linked carbohydrate on HIV-1 envelope and its role in shielding against one category of broadly neutralizing antibodies. *Cell Rep.* **30**, 1862–1869.e1864 (2020).
104. Das, S. R. et al. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. *Proc. Natl Acad. Sci. USA* **108**, E1417 (2011).

#### Acknowledgements

The authors thank all of the researchers who have shared genome data openly via the Global Initiative on Sharing All Influenza Data (GISAID). The COVID-19 Genomics UK

(COG-UK) Consortium is supported by funding from the UK Medical Research Council (MRC), part of UK Research and Innovation, the UK National Institute of Health Research and Genome Research Limited, operating as the Wellcome Sanger Institute. W.T.H. is funded by the MRC (MR/R024758/1). R.R. is funded by the UK Biotechnology and Biological Sciences Research Council (BB/R012679/1). D.L.R. and E.C.T. are funded by the MRC (MC\_UU\_12014/12) and acknowledge the support of the G2P-UK National Virology Consortium (MR/W005611/1) funded by UK Research and Innovation. D.L.R. also acknowledges support of the Wellcome Trust (220977/Z/20/Z). A.R. acknowledges the support of the Wellcome Trust (Collaborators Award 206298/Z/17/Z — ARTIC network) and the European Research Council (grant agreement no. 725422 — ReservoirDOCS).

#### Author contributions

W.T.H., A.M.C., B.J., R.K.G., E.C.T., E.M.H., C.L., A.R. and D.L.R. researched data for the article. W.T.H., A.M.C., A.R., S.J.P. and D.L.R. contributed substantially to discussion of the content. W.T.H., A.M.C. and D.L.R. wrote the article. W.T.H., A.M.C., R.K.G., E.C.T., R.R., S.J.P. and D.L.R. reviewed and/or edited the manuscript before submission.

#### Competing interests

The authors declare no competing interests.

#### Peer review information

*Nature Reviews Microbiology* thanks Y. Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

#### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1038/s41579-021-00573-0>.

#### RELATED LINKS

COG-UK Mutation Explorer: <http://sars2.cvr.gla.ac.uk/cog-uk/>  
 COVID-19 Genomics UK (COG-UK) Consortium: <https://www.cogconsortium.uk>  
 CoV-GLUE — A Web Application for Tracking SARS-CoV-2 Genomic Variation: <http://cov-glue.cvr.gla.ac.uk/>  
 Global Initiative on Sharing All Influenza Data (GISAID): <https://www.gisaid.org>  
 Global Report Investigating Novel Coronavirus Haplotypes: [https://cov-lineages.org/global\\_report.html](https://cov-lineages.org/global_report.html)

© Springer Nature Limited 2021