

# Missense variants in human ACE2 modify binding to SARS-CoV-2 Spike

Stuart A. MacGowan<sup>1\*</sup>, Michael I. Barton<sup>2\*</sup>, Mikhail Kutuzov<sup>2</sup>, Omer Dushek<sup>2</sup>, P. Anton van der Merwe<sup>2+</sup> and Geoffrey J. Barton<sup>1+</sup>

<sup>1</sup>*Division of Computational Biology, School of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH, Scotland, UK*

<sup>2</sup>*Sir William Dunn School of Pathology, University of Oxford, Oxford, Oxfordshire, United Kingdom*

\* These authors contributed equally to this work.

+ Corresponding author email addresses: [anton.vandermerwe@path.ox.ac.uk](mailto:anton.vandermerwe@path.ox.ac.uk), [g.j.barton@dundee.ac.uk](mailto:g.j.barton@dundee.ac.uk)

## Abstract

SARS-CoV-2 infection begins with the interaction of the SARS-CoV-2 Spike (Spike) and human angiotensin-converting enzyme 2 (ACE2). To explore whether population variants in ACE2 might influence Spike binding and hence infection, we selected 10 ACE2 variants based on affinity predictions and prevalence in gnomAD and measured their affinities for Spike receptor binding domain through surface plasmon resonance (SPR). We discovered variants that enhance and reduce binding, including two variants with distinct population distributions that enhanced affinity for Spike. ACE2 p.Ser19Pro ( $\Delta\Delta G = 0.59 \pm 0.08 \text{ kcal mol}^{-1}$ ) is often seen in the gnomAD African cohort (AF = 0.003) whilst p.Lys26Arg ( $\Delta\Delta G = 0.26 \pm 0.09 \text{ kcal mol}^{-1}$ ) is predominant in the Ashkenazi Jewish (AF = 0.01) and European non-Finnish (AF = 0.006) cohorts. Carriers of these alleles may be more susceptible to infection or severe disease and these variants may influence the global epidemiology of Covid-19. We also identified three rare ACE2 variants that strongly inhibited (p.Glu37Lys,  $\Delta\Delta G = -1.33 \pm 0.15 \text{ kcal mol}^{-1}$  and p.Gly352Val, predicted  $\Delta\Delta G = -1.17 \text{ kcal mol}^{-1}$ ) or abolished (p.Asp355Asn) Spike binding. These variants may confer resistance to infection. Finally, we calibrated the mCSM-PPI2  $\Delta\Delta G$  prediction algorithm against our SPR data, give new predictions for all possible ACE2 missense variants at the Spike interface and estimate the overall burden of ACE2 variants on Covid-19 phenotypes.

## 1 Introduction

The COVID-19 pandemic is one of the greatest global health challenges of modern times. Although the disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is usually cleared following mild symptoms, it can progress to serious illness and death<sup>1</sup>. Besides the clear risks associated with age and comorbidities<sup>2,3</sup>, there could be a genetic component that predisposes some individuals to worse outcomes<sup>4</sup>. Genetic association studies have already identified several loci involved in Covid-19 risk<sup>5</sup>. Identifying further genetic factors of COVID-19 susceptibility has implications for clinical decision making and epidemic dynamics. Genetic variation may constitute hidden risk factors and, in

some cases, explain why otherwise healthy individuals in low-risk groups experience severe disease. The identification of specific genetic variants that influence the severity and progression of COVID-19 presents the opportunity for predictive diagnostics, early intervention and personalised treatments whilst the population distribution of such variants could contribute to population specific risk.

Human angiotensin-converting enzyme 2 (ACE2) is the host cell receptor that SARS-CoV-2 exploits to infect human cells<sup>1,6</sup>. As this is the same receptor used by the SARS coronavirus (SARS-CoV) that caused the SARS outbreak in 2002, the detailed body of knowledge built around SARS-CoV infection is relevant to understanding SARS-CoV-2<sup>1,6,7</sup>. The spike glycoprotein (Spike) is the coronavirus entity that recognises and binds host ACE2. Both SARS coronavirus Spikes include an S1 domain that contains ACE2 recognition elements and an S2 domain that is responsible for membrane fusion<sup>6</sup>. Spike is primed for cell fusion by cleavage with host furin<sup>8</sup> and TMPRSS2<sup>6</sup>, in SARS-CoV cleavage by TMPRSS2 is thought to be promoted upon formation of the ACE2 Spike complex<sup>9</sup>. The S1 receptor binding domains (RBDs) from both SARS-CoV<sup>10</sup> and SARS-CoV-2<sup>11</sup> have been co-crystallised with human ACE2. The RBDs from both viruses are similar in overall architecture and interface with roughly the same surface on ACE2. Differences are apparent in the so-called receptor binding motif, which is the region of the RBD responsible for host range and specificity of coronaviruses<sup>10-12</sup>. The binding affinity of Spike and ACE2 is known to be correlated to the infectivity of SARS-CoV and is determined by the complementarity of the interfaces<sup>10,12</sup>. However, despite its essential role in infection, risk variants in ACE2 have not been conclusively identified in genetic association studies.

Missense variants located in protein-protein interaction interfaces can affect altered binding characteristics<sup>13,14</sup> and in the context of virus-host interactions, have been shown to effect susceptibility<sup>15</sup>. This indicates the potential of missense variants in ACE2 to alter Spike binding and therefore influence a key step in SARS-CoV-2 infection. This is also suggested by the fact that the host range of coronaviruses is partly determined by the complementarity of the Spike receptor binding motif and the target hosts' ACE2 sequence<sup>10,12</sup>. A few studies<sup>4,16-21</sup> have addressed this question and gave rise to conflicting conclusions regarding the effects of specific variants on the interaction affinity and their relevance to the pandemic generally. The strongest of these used data from a published deep mutagenesis binding screen<sup>22</sup> to assess the effect of ACE2 population variants on Spike affinity and confirmed the effects of five key variants with further biochemical assays<sup>17</sup>. In our own previous contribution<sup>23</sup>, we employed the mCSM-PPI2 protein-protein interaction affinity prediction algorithm<sup>13</sup> to assess the effects of ACE2 variants on the binding of SARS-CoV-2 Spike and predicted that three reported ACE2 variants would strongly inhibit or abolish binding (p.Asp355Asn, p.Glu37Lys and p.Gly352Val). Confidence was given to these predictions by the performance of mCSM-PPI2<sup>13</sup> in comparison to binding data for 26 ACE2 mutants in complex with SARS-CoV Spike RBD<sup>12</sup>. Here we report experimental binding affinities of 10 ACE2 variants for isolated SARS-CoV-2 Spike RBD determined via surface plasmon resonance (SPR). These results give better insight into the effect of ACE2 variants on SARS-CoV-2 Spike binding, revealing additional variants that enhance Spike binding and also test the quality of our predictions. The SPR data also allowed us to recalibrate the mCSM-PPI2 predictions to provide more accurate estimates of the effect of interface variants that we did not test experimentally.

## 2 Results

### 2.1 ACE2 variant affinities for SARS-CoV-2 Spike

Figure 1 highlights the mutated residues on the structure of ACE2<sup>11</sup> in complex with SARS-CoV-2 Spike. We determined the binding affinity of 10 ACE2 mutants for isolated SARS-CoV-2 Spike RBD via surface plasmon resonance (SPR) to identify variants that may influence an individual's response to infection. Nine of these mutants were selected from the 241 ACE2 missense variants reported in gnomAD<sup>24</sup> on the basis of our previous computational predictions<sup>23</sup> and their reported prevalence, whilst the tenth mutant (p.Thr27Arg) was predicted to enhance Spike binding more than any other possible mutation at the interface.

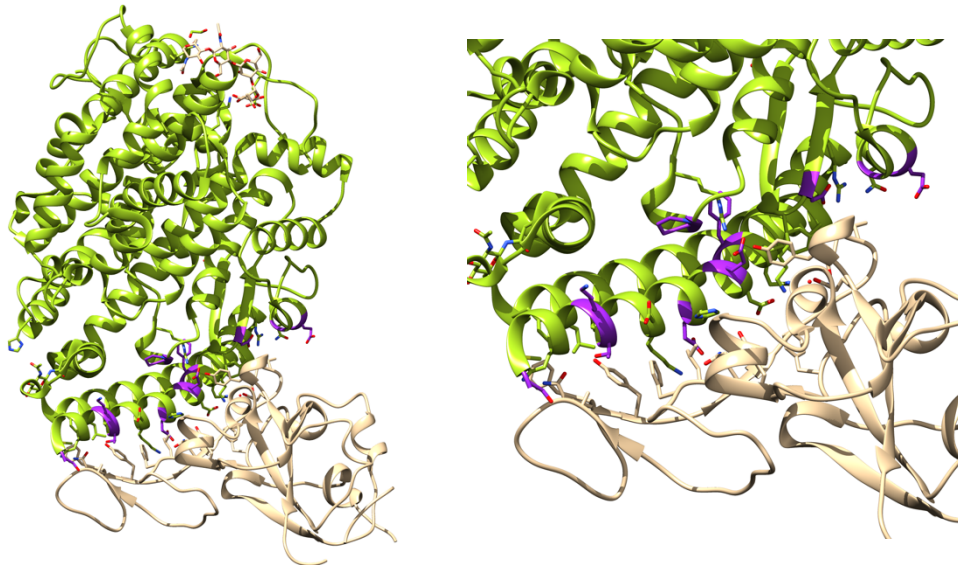


Figure 1. Binding affinity determination of ACE2 variants with SARS-CoV-2 Spike. Left: ACE2 (green) in complex with Spike RBD (tan) complex from biological assembly 1 derived from PDB ID: 6vw1<sup>11</sup>. The positions that were mutated in this work are highlighted magenta. Right: The ACE2 Spike interface. Figure generated with Jalview<sup>25</sup> and UCSF Chimera<sup>26</sup>.

Table 1 presents experimentally determined  $\Delta\Delta G$  and mCSM-PPI2<sup>13</sup> predictions for the 10 ACE2 mutants together with predicted data for a further three variants, alongside variant population frequencies and RBD interacting residues. Our SPR data were collected in two batches. The first batch comprised four variants, two were predicted to strongly reduce or abolish Spike binding (p.Glu37Lys and p.Asp355Asn) and two predicted to enhance Spike binding (p.Gly326Glu and Thr27Arg)<sup>23</sup>. The SPR measurements showed strongly reduced binding for p.Glu37Lys and the total abolition of binding for p.Asp355Asn (within the concentration range assayed), in agreement with the predictions. In contrast, p.Gly326Glu and Thr27Arg, which had been predicted to enhance binding, displayed decreased and slightly decreased binding, in disagreement with the predictions. These discrepancies motivated a second set of SPR measurements that included six of the most prevalent ACE2 variants close to the Spike binding site. Surprisingly, the two most common ACE2 variants tested bound SARS-CoV-2 Spike more strongly than reference ACE2. These were p.Lys26Arg ( $\Delta\Delta G = 0.26 \pm 0.09$ ), which has the highest allele frequency of ACE2 variants near the Spike interface, and p.Ser19Pro ( $\Delta\Delta G = 0.59 \pm 0.09$ ) that has the second highest frequency. Two other variants in this series also increased Spike binding (p.Phe40Leu and p.Pro389His) despite being over 8 Å from the closest Spike residue. Finally, p.Glu329Gly may cause a slight reduction in binding ( $\Delta\Delta G = -0.09 \pm 0.09$ ) whilst p.Glu35Lys had an inhibitory effect ( $\Delta\Delta G = -0.36 \pm 0.09$ ). These results show that ACE2 variants can both enhance and inhibit

## Spike binding, properties that may reasonably be associated with susceptibility and resistance phenotypes to SARS-CoV-2 infection.

*Table 1. Surface plasmon resonance derived  $\Delta\Delta G$  and mCSM-PPI2<sup>13</sup> predictions for ACE2 mutants, including gnomAD<sup>24</sup> missense variants, at or near the ACE2 Spike interface. See Supplementary Table 1 for mCSM-PPI2 predictions for all gnomAD variants in the ACE2 ectodomain.*

Mutation	Distance to Spike	Spike residues	mCSM-PPI2 $\Delta\Delta G$	Recalibrated mCSM-PPI2 $\Delta\Delta G$	SPR $\Delta\Delta G$	Max. prevalence (1 sf)
p.Ser19Pro	2.6	A475,G476,S477	-0.2	0.2	$0.59 \pm 0.09$	0.003 (AFR)
p.Lys26Arg	6.0	(F456)	0.0	0.4	$0.26 \pm 0.09$	0.006 (NFE)
p.Thr27Ala	3.7	A475,N487,F456,Y473,Y489	-0.6	-0.4	-	0.00007 (AMR)
p.Thr27Arg <sup>a</sup>	3.7	“ “	1.4	NA	$-0.11 \pm 0.10$	NA
p.Glu35Lys	2.9	Q493	-0.5	-0.3	$-0.36 \pm 0.09$	0.0001 (EAS)
p.Glu37Lys	3.2	Y505	-1.2	-1.3	$-1.33 \pm 0.18$	0.0003 (FIN)
p.Phe40Leu	8.6	(Y449,Q498)	-0.3	0.0	$0.11 \pm 0.09$	<sup>a</sup> 0.0001 (AMR) <sup>b</sup> 0.0002 (AFR)
p.Met82Ile	3.5	F486	-0.3	0.1	-	0.0003 (AFR)
p.Gly326Glu	5.5	(V503,N506)	1.0	NA	$-0.65 \pm 0.14$	0.0001 (AFR)
p.Glu329Gly	4.1	R439	-0.4	-0.1	$-0.09 \pm 0.09$	0.0001 (NFE)
p.Gly352Val	5.4	(Y505)	-1.1	-1.2	-	0.00005 (NFE)
p.Asp355Asn	3.5	G502,T500	-1.3	-1.5	$(< -3.16 \pm 0.14)^b$	0.00003 (NFE)
p.Pro389His	8.1	(Y505)	-0.1	0.4	$0.27 \pm 0.09$	0.0002 (AMR)

Column legend – Distance to Spike: The minimum distance of the wild-type residue to the SARS-CoV-2 Spike as resolved in PDB 6vw1<sup>11</sup>. Spike residues: Spike residues within 5 angstrom of mutant site (or closest if no residues are in this range). mCSM-PPI2  $\Delta\Delta G$ : The predicted  $\Delta\Delta G$  in kcal mol<sup>-1</sup> for the missense mutation calculated by mCSM-PPI2<sup>13</sup> with PDB 6vw1 as the model structure. Recalibrated mCSM-PPI2  $\Delta\Delta G$ : adjusted mCSM-PPI2  $\Delta\Delta G$  following recalibration with SPR data (§2.4). SPR  $\Delta\Delta G$ :  $\Delta\Delta G$  determined by SPR assay. Max. prevalence: The allele frequency for the gnomAD continental population with the highest frequency (see Table 4 for all population frequencies and definitions of all gnomAD cohort abbreviations). a. p.Thr27Arg was not reported in gnomAD and was selected as it had the highest predicted increased  $\Delta\Delta G$  of any possible ACE2 mutation at the Spike interface. b. No binding was observed for ACE2 p.Asp355Asn, this value corresponds to the calculated maximum affinity that is consistent with this observation.

## 2.2 Structural features of affinity modifying variants

Figure 2 and Figure 3 illustrate the mCSM-PPI2 provided models of p.Ser19Pro and p.Lys26Arg. p.Lys26Arg was not predicted to make any new well-defined contacts with Spike residues but it is predicted to adopt a conformation that extends toward Spike residues Y473 and F456, coming within 7.8 Å and 6.1 Å of these residues' aromatic rings, respectively, slightly beyond typical amino-aromatic interaction distances<sup>27</sup> but potentially favourable when dynamics are considered. Similarly, p.Ser19Pro also does not introduce any new Spike contacts according to the mCSM-PPI2 model and so the enhanced affinity is difficult to explain but it has been suggested that the Pro mutant stabilises the helix to favour Spike interaction<sup>22</sup>. ACE2 p.Ser19Pro is of further interest because of its proximity to Spike Ser477, which has mutated to Asn in circulating SARS-CoV-2 strains and these ACE2 and RBD variants have been found to interact<sup>28</sup>. The mCSM-PPI2 structural models of the p.Asp355Asn, p.Glu37Lys and p.Gly352Val were discussed in detail in our previous work<sup>23</sup>. In summary, p.Asp355Asn was predicted to introduce a number of steric clashes at the interface, whilst p.Glu37Lys abolished an H-bond between ACE2 Glu37 and RBD Tyr505.

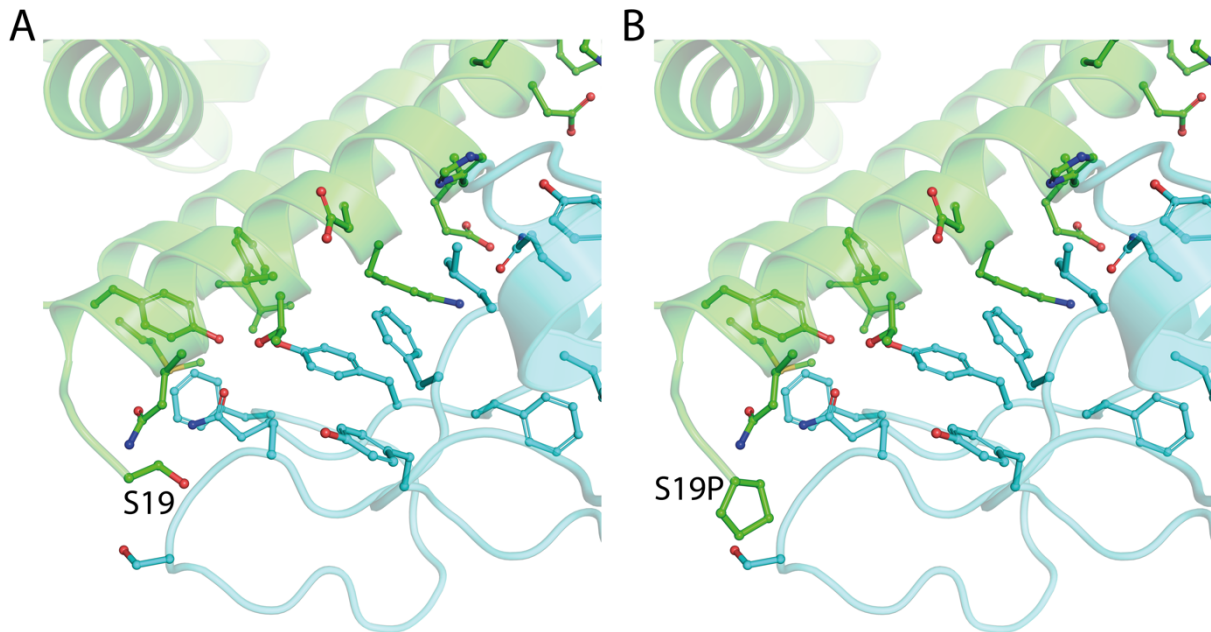


Figure 2. The structure of ACE2 (green) gnomAD<sup>24</sup> missense variant p.Ser19Pro that enhances Spike (light blue) binding affinity. A. The environment of ACE2 Ser19 from PDB ID: 6vw1<sup>11</sup>. B. Model of ACE2 p.Ser19Pro in complex with Spike. The mutant structure was modelled onto 6vw1 with mCSM-PPI<sup>13</sup>. Figure created with PyMol<sup>29</sup>.

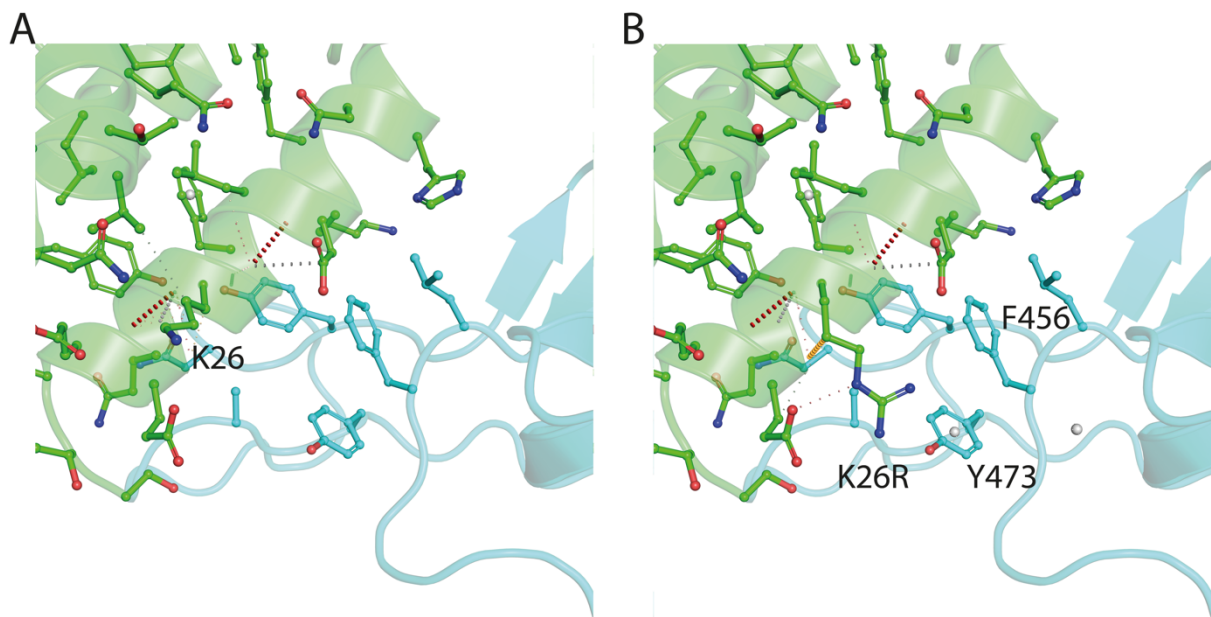


Figure 3. The structure of ACE2 (green) gnomAD<sup>24</sup> missense variant p.Lys26Arg that enhances Spike (light blue) binding affinity. A. The environment of ACE2 Lys26 from PDB ID: 6vw1<sup>11</sup>. B. Model of ACE2 p.Lys26Arg in complex with Spike. The mutant structure was modelled onto 6vw1 with mCSM-PPI<sup>13</sup>. Figure created with PyMol<sup>29</sup>.

### 2.3 Enhanced Spike binding ACE2 variants are relatively common in European (p.Lys26Arg) and African/African-American (p.Ser19Pro) populations

The enhanced binding by p.Lys26Arg and p.Ser19Pro is particularly interesting since this may effect carrier susceptibility or vulnerability toward SARS-CoV-2 infection and they have relatively high prevalence in the gnomAD<sup>24</sup> populations. p.Lys26Arg is the most common missense variant in the ACE2 ectodomain (Total allele count = 797, Total allele frequency = 0.004) and is predominant in the Ashkenazi Jewish cohort (ASJ AF = 0.01) and the European (non-Finnish) population (NFE, AF = 0.006). Amongst NFE sub-populations, it is most

prevalent in north-western Europeans (AF = 0.007) and least prevalent in southern Europeans (AF = 0.003) and Estonians (AF = 0.003). p.Lys26Arg was also observed within this frequency range in Latino/Admixed American (AMR, AF = 0.003) samples. The variant is less frequent in Finnish (FIN, AF = 0.0005), African/African-American (AFR, AF = 0.001), South Asian (0.001) and, especially, East Asian (0.00001) samples. Interestingly, gnomAD reports a second variant at this site, p.Lys26Glu, suggesting that the position is especially tolerant to mutation. p.Ser19Pro is the next most common ACE2 missense variant in proximity to the Spike binding site (AC = 64, AF = 0.0003) and has the highest positive  $\Delta\Delta G$  of all those tested ( $\Delta\Delta G = 0.59 \pm 0.03$ ). This variant is practically unique to the African/African-American gnomAD population (AFR, AC = 63, AF = 0.003). The only other observation of this variant in gnomAD is in a single heterozygote in the labelled “Other” cohort. As a result of these distinct population distributions, it is possible that these variants could contribute to some of the observed epidemiological<sup>30</sup> variation between populations and ethnic groups.

Given the prevalence of these variants we checked for their occurrence in recent GWAS studies on Covid-19 related phenotypes. Table 2 presents GWAS association results for ACE2 p.Lys26Arg from the Covid-19 Host Genetics Initiative (HGI)<sup>31</sup>. These data show some consistency with the hypothesis that p.Lys26Arg contributes additional risk for more severe Covid outcomes, whilst not affecting the likelihood of infection. In the very severe respiratory confirmed Covid vs. population contrast (study A2), a non-statistically significant increased risk was reported ( $\beta = 0.38 \pm 0.24$ ,  $p = 0.12$ ,  $p\text{-het} = 0.74$ ). Other Covid HGI data summaries for contrasts testing for alleles associated with the risk of SARS-CoV-2 infection suggest the variant does not play a role in infection acquisition. Covid vs. lab/self-reported negative ( $\beta = -0.18 \pm 0.14$ ,  $p = 0.21$ ,  $p\text{-het} = 0.92$ ); Covid vs. population ( $\beta = -0.03 \pm 0.11$ ,  $p = 0.81$ ,  $p\text{-het} = 0.65$ ) and predicted Covid from self-reported symptoms vs. predicted or self-reported non-Covid ( $\beta = -0.10 \pm 0.24$ ,  $p = 0.67$ ,  $p\text{-het} = 0.08$ ). Although none of these tests achieved genome wide significance, it would be worthwhile to reassess the data after controlling for other loci with the greatest effect sizes, sex or other appropriate stratifications.

Table 2: Covid Host Genetics Initiative summaries for p.Lys26Arg (rs4646116).

Covid HGI study and description	Beta	SE	p	p-het	meta N	meta AF	N Cases	N Controls
A2 - very severe respiratory confirmed covid vs. population	<b>0.38</b>	0.24	0.12	0.74	35,222	0.003	2,972	284,472
B1 - hospitalized covid vs. not hospitalized covid			Variant not reported				1,389	5,879
B2 - hospitalized covid vs. population			Variant not reported				6,492	1,012,809
C1 - covid vs. lab/self-reported negative	<b>-0.18</b>	0.14	0.21	0.92	64,950	0.004	11,181	116,456
C2 - covid vs. population	-0.03	0.11	0.81	0.65	972,525	0.005	17,607	1,345,334
D1 - predicted covid from self-reported symptoms vs. predicted or self-reported non-covid	-0.10	0.24	0.67	0.08	10,655	0.006	1,777	18,895

#### 2.4 Recalibrated mCSM-PPI2 $\Delta\Delta G$ predictions

Our SPR data provide accurate readouts of the effect of ACE2 variants on Spike binding, but we could not feasibly carry out experiments for all possible ACE2 mutations at the interface. For variants we have not studied, predictions and other high-throughput datasets can be useful, so long as their applicability and limitations are well-understood. In our previous

work<sup>23</sup>, we calibrated predictions against relative binding data for ACE2/SARS-CoV Spike, we now improve this by recalibrating the mCSM-PPI2 predictions with our ACE2/SARS-CoV-2 Spike SPR dataset.

Figure 4 compares experimental and mCSM-PPI2<sup>13</sup> predicted  $\Delta\Delta G$ . The eight mutants with predicted  $\Delta\Delta G < 0$  kcal mol<sup>-1</sup> are linearly correlated with the experimental  $\Delta\Delta G$  ( $R^2 = 0.91$ ,  $p = 0.0006$ ), whilst the two mutants with predicted  $\Delta\Delta G > 1$  kcal mol<sup>-1</sup> (p.Gly326Glu and Thr27Arg) do not follow this behaviour. The absence of detectable binding for p.Asp355Asn and the strongly reduced binding observed for p.Glu37Lys, support our original determination that predicted  $\Delta\Delta G < -1$  kcal mol<sup>-1</sup> was a reliable indicator of inhibitory variants. The difficulty mCSM-PPI2 has with affinity enhancing variants is not altogether surprising since our original calibration showed poorer performance for these variants<sup>23</sup>, which may be because there were fewer experimentally determined affinity enhancing variants in the algorithm's training data. Structurally, although the mCSM-PPI2 model of p.Gly326Glu predicted new contacts with Spike<sup>23</sup>, it is possible that reduced torsional flexibility at mutant Glu326 has other effects that were not recognised (a similar argument applies for p.Ser19Pro but here the error is small). The subtler affinity changes measured for the five variants with predicted  $\Delta\Delta G$  between  $-0.5$  and  $0.0$  kcal mol<sup>-1</sup>, validate our original ambiguity towards those predictions, but the strong linear correlation within this narrow range is encouraging. Taken together, these observations suggest that predictions within the range of p.Glu36Lys ( $\Delta\Delta G^{\text{pred}} = -1.2$  kcal mol<sup>-1</sup>) and p.Lys26Arg ( $\Delta\Delta G^{\text{pred}} = 0.0$  kcal mol<sup>-1</sup>) may be rescaled to provide a better estimate of actual  $\Delta\Delta G$ . Recalibrated  $\Delta\Delta G$  yields correct predictions for all eight mutations with negative predicted  $\Delta\Delta G$  and measurable  $\Delta\Delta G$  from SPR (Table 1). This lends additional confidence in our prediction that ACE2 p.Gly352Val strongly reduces Spike binding (in agreement with published deep mutagenesis experiments<sup>22</sup>).

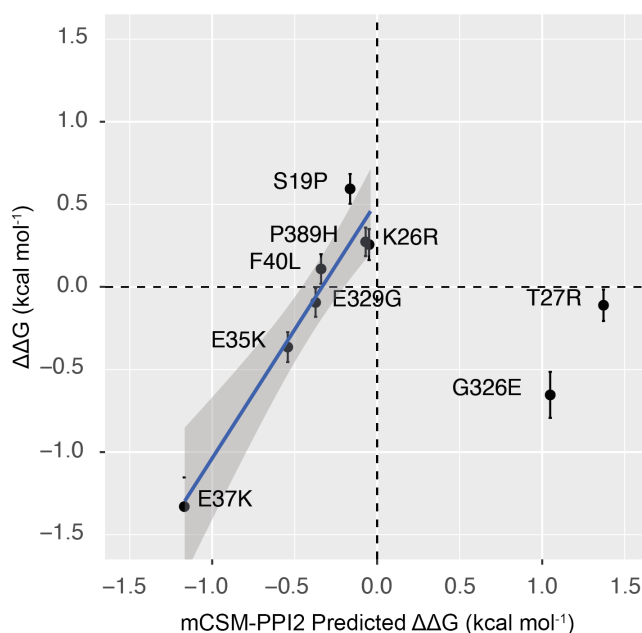


Figure 4. Comparison of SPR  $\Delta\Delta G$  with the mCSM-PPI2 prediction for 9 ACE2 mutations. The linear regression was fit to the eight ACE2 mutants with predicted  $\Delta\Delta G < 0$  kcal mol<sup>-1</sup> (slope = 1.56, intercept = 0.52,  $R^2 = 0.91$ ,  $p = 0.0006$ ). The shaded area indicates the 95 % confidence interval for predicted SPR  $\Delta\Delta G$  from the linear model. Figure generated with R ggplot2.

## 2.5 Predicted burden of rare ACE2 variants with Spike affinity phenotypes

Existing human variation datasets are well-powered to detect common variation (1KG was estimated to detect >99% SNPs with MAF >1%<sup>32</sup> and gnomAD is substantially larger) in the sampled populations but they are far from comprehensive with respect to rare variation<sup>24</sup>. Rare variants in ACE2 that influence Spike binding could have implications for the epidemiology of COVID-19 in addition to the consequences for affected individuals. If there were 10 such variants with an allele frequency of 1 in 50,000, their collective occurrence might be as high as 1 in 5,000 (discounting linkage) and when this is considered alongside the possibility that a high proportion of the global population will be exposed to SARS-CoV-2 it becomes clear that such effects should be investigated. These variants could even be present at significant frequencies in populations missing or underrepresented in gnomAD.

Figure 5 illustrates the distribution of recalibrated mCSM-PPI2  $\Delta\Delta G$  predictions ( $\Delta\Delta G^{\text{recal}}$ ) for all 475 possible ACE2 mutations at 25 ACE2 residues close to the Spike interface and the subset that are accessible via a single base change of the ACE2 coding sequence (these are more likely to be present in human populations than those requiring multiple substitutions). Most of these mutations are predicted to have only a slight effect on Spike binding, but there is a secondary mode below  $-1.0 \text{ kcal mol}^{-1}$  with 126 mutations predicted to lead to strongly reduced binding. Fewer variants received high positive  $\Delta\Delta G^{\text{recal}}$  scores: 17 had  $\Delta\Delta G^{\text{recal}} > 1.0 \text{ kcal mol}^{-1}$ , a further 44 had  $\Delta\Delta G^{\text{recal}} > 0.5 \text{ kcal mol}^{-1}$  and an additional 70 had  $\Delta\Delta G^{\text{recal}} > 0.2 \text{ kcal mol}^{-1}$  (i.e.,  $\Delta\Delta G^{\text{recal}} > \text{p.Lys26Arg}$ ). A similar pattern is observed for the 151 mutants corresponding to 172 single nucleotide variants of the ACE2 coding sequence (Figure 5B). These results suggest that random novel ACE2 missense variants at these loci can inhibit or enhance Spike binding, but are slightly more likely to be inhibitory, meaning that diversity at these positions might typically be beneficial and provide some resistance to infection. Notably, these recalibrated predictions do not display the same bias toward negative  $\Delta\Delta G$  that the raw mCSM-PPI2  $\Delta\Delta G$  predictions did<sup>23</sup> (Supplementary Figure 1), which may be another indicator of the improved quality of these recalibrated predictions.

Even though most of these variants are not reported in gnomAD, they may still occur within the populations represented, especially if they occur at frequencies that are poorly detected. With this in mind, we calculated allele frequencies for these mutations that are compatible with their absence from gnomAD (see Methods) in order to gain a better idea of how widespread the effects of these variants might be.

Table 3 presents estimated allele frequencies for novel variants that are predicted to inhibit and enhance Spike binding at varying  $\Delta\Delta G^{\text{recal}}$  thresholds. Upper bounds for their joint frequencies assuming that they occur at lower frequencies than singleton variants suggest frequency bounds that range 1 in 1,000 to 1 in 2,000 variants per allele for inhibitory variants, and 1 in 1,500 to around 1 in 5,000 for enhancer variants. A second approach, which takes account of the empirical detection of rare variants in ACE2, yields frequencies that span from 1 in 6,250 to 1 in 12,195 for inhibitory variants, and 1 in 8,333 to 1 in 37,037 for enhancer variants. These estimates show that novel variants in ACE2 with any weak Spike affinity phenotype could plausibly be as common as 1 in 3,571 alleles (calculated as the sum of the highest inhibitor and enhancer prevalence's), but the strongest affinity phenotypes are more likely to occur in frequency ranges akin to rare genetic diseases. It should be remembered that these values are calculated to be compatible with their absence



from the gnomAD dataset, but it remains possible that some or all of these variants do not exist at all or that they are very common in one or more populations not represented in gnomAD.

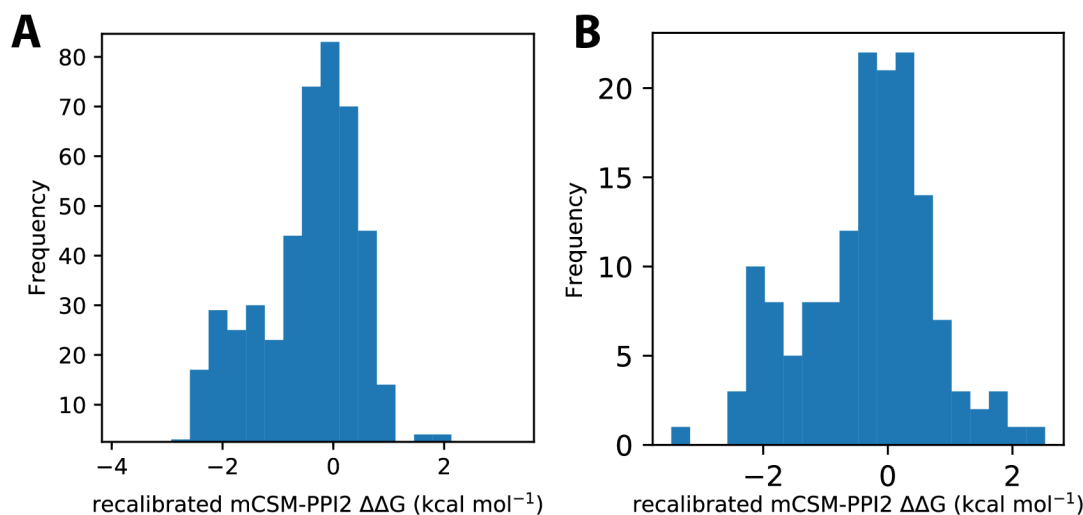


Figure 5. A. Distribution of mCSM-PPI2<sup>23</sup> predicted  $\Delta\Delta G$  from *in silico* saturation mutagenesis of the ACE2-S interface in PDB 6vw1<sup>11</sup>. A. predicted  $\Delta\Delta G$  for 475 mutations across 25 sites on ACE2 corresponding to the 23 residues within 5 Å of SARS-CoV-2 S plus Gly326 and Gly352. B. predicted  $\Delta\Delta G$  for the subset of 151 mutations across these sites that are accessible via a single base change of the ACE2 coding sequence.

Table 3. Estimated allele frequencies of potential novel variants in ACE2 with predicted Spike binding phenotypes. The estimate is calculated from the observed occurrence of rare variants in ACE2 in gnomAD<sup>24</sup> ( $6.1 \times 10^{-6}$ ) and the proportion of SNPs at the 25 residues that are predicted to modify Spike affinity at different thresholds (see Methods).

$\Delta\Delta G^{\text{recal.}}$ (kcal mol <sup>-1</sup> )	N missense	N SNPs	Joint singleton frequency (NFE)	Joint estimated frequency (NFE)	Heterozygotes / 100K	Hemizygotes / 100K
<u>Predicted Spike inhibitory variants</u>						
< -1.0	38	41	0.0005	$8.2 \cdot 10^{-5}$	16.5	8.2
< -0.5	55	59	0.0007	0.00012	23.8	11.9
< -0.2	77	81	0.001	0.00016	33.0	16.5
<u>Predicted Spike enhancer variants</u>						
> 1.0	11	14	0.0002	$2.7 \cdot 10^{-5}$	5.5	2.7
> 0.5	24	30	0.0004	$6.0 \cdot 10^{-5}$	11.9	6.0
> 0.2	48	58	0.0007	0.00012	23.8	11.9

### 3 Discussion and conclusion

#### 3.1 Impact of ACE2 affinity variants on SARS-CoV-2 infection

How far can affinity modulating variants influence SARS-CoV-2 infection? For inhibitory variants, host range specificity and mutagenesis studies<sup>12</sup> highlight how some receptor mutations can provide complete protection from infection. We identified one variant (p.Asp355Asn) that abolished Spike binding altogether within our detection limits and a few others that reduced binding to varying extents. Since these variants were observed in population samples, some individuals carry ACE2 variants that could confer complete resistance to infection. Variants that reduce but do not eliminate binding may confer a degree of resistance proportional to the affinity reduction, or alternatively, there may be an affinity threshold that toggles cellular permissivity as in other enveloped viruses<sup>33</sup>. Detailed infectivity experiments with relevant ACE2 variants and multiple cell types are necessary to answer this question definitively.

It is less clear what effect affinity enhancing variants have on virulence, but there are indications that carriers may be at greater risk of infection and severe disease. Virus attachment proteins in enveloped viruses require a minimum receptor affinity that is proportional to the receptor surface density on the target cell to enable membrane fusion<sup>33</sup>. If this applies to SARS-CoV-2, ACE2 variants that enhance Spike binding could increase viral spreading in carriers due to increased cellular tropism. Greater viral spreading is associated with clinical deterioration<sup>34</sup> and could cause increased infectiousness, akin to the enhanced transmissibility of SARS-CoV-2 vs. SARS-CoV, which is associated with increased viral loads in the upper respiratory tract<sup>6,35</sup> and also correlates with enhanced receptor affinity<sup>11</sup>. The importance of tuned Spike-receptor affinities in SARS-CoV-2 may be enhanced because of the relatively low surface density of Spike on SARS-CoV-2 virions, which may lead to a weak avidity effect<sup>36</sup>. Anecdotal evidence is also provided by Spike RBD mutations, such as N501Y, that enhance ACE2 affinity<sup>37</sup> and are associated with increased transmissibility<sup>38</sup>. Again, infectivity studies are needed but here the prevalence of p.Lys26Arg and p.Ser19Pro may allow the observation of an effect in future Covid-19 GWAS studies.

Some infectivity data for SARS-CoV-2 towards cells expressing ACE2 variants are available, including variants p.Ser19Pro and p.Lys26Arg<sup>16</sup>. Surprisingly, SARS-CoV-2 pseudotypes showed slightly decreased infectivity towards HEK293T cells expressing ACE2 p.Lys26Arg, and no difference compared to reference ACE2 was found for the susceptibility of cells expressing ACE2 p.Ser19Pro<sup>16</sup>. Although these results suggest that affinity is not directly proportional to infectivity in a single cell type, they do not exclude the possibility that these variants allow increased cellular tropism as discussed above. The same study also reported no effect on Spike affinity caused by these variants<sup>16</sup>, which is probably due to the low sensitivity of the cell based binding assay employed, since enhanced binding was also found in a different cell based assay<sup>22</sup>. Indeed, we are especially confident in the accuracy of our  $\Delta\Delta G$  measurements since our results are consistent with published deep mutagenesis binding data<sup>22</sup> (Supplementary Figure 2). In addition, another recent report agreed with our findings that the common ACE2 variants p.Ser19Pro and p.Lys26Arg enhanced affinity for Spike whilst p.Glu37Lys inhibits binding<sup>17</sup>.

Other considerations relevant to the effect of ACE2 Spike affinity variants on carriers include ACE2 carboxypeptidase activity<sup>39</sup>, which may be modulated by the specific binding affinity, the effect of hemizygoty and sex differences in Covid-19 outcomes, and the interplay of affinity variants and ACE2 expression levels. The importance of ACE2 expression was mentioned earlier but it is worth highlighting that it is known to partly determine the cellular specificity of SARS-CoV<sup>40</sup> and was explored as a potential factor in COVID-19 susceptibility and severity, including the interaction between ACE2 variants and ACE2 stimulating drugs<sup>41</sup>. Also, heterozygotes express a proportion of ACE2 alleles whilst hemizygoty carry only a single ACE2 allele so that ACE2 Spike affinity variants ought to always show greater penetrance in hemizygoty, for better or worse. In contrast and since ACE2 escapes complete X-inactivation<sup>20</sup>, heterozygotes have the advantage that the more resistant ACE2 allele could become dominant in infected tissues due to selection over cellular infection cycles, gradually increasing the prevalence of the more Spike resistant ACE2 allele. Besides the possible benefit against infection, there could be other implications

for heterozygotes depending on the persistence X-inactivation bias and the nature of any hitchhiking alleles.

### 3.2 Prevalence of ACE2-Spike affinity genotypes

Table 4 presents detailed allele frequencies from gnomAD<sup>24</sup> for the ACE2 variants investigated in this work. The most prevalent ACE2 variant mediated Covid-19 phenotypes are likely to arise from the two relatively common variants found to enhance Spike binding, p.Ser19Pro and p.Lys26Arg. These variants were found in 3 in 1,000 individuals in the gnomAD African/African-American (AFR) samples and 7 in 1,000 North Western non-Finnish European samples (NW-NFE), respectively. p.Lys26Arg was also observed in other populations, including 1 in 1,000 AFR samples, adding further burden to this cohort. The high affinity displayed by ACE2 p.Ser19Pro is concerning in context with the reported disproportionate impact of Covid-19 on black and minority ethnic groups<sup>30</sup>. Further research to identify the impact of these variants on SARS-CoV-2 pathogenesis should be prioritised with urgency and existing and future GWAS should investigate these variants more closely with appropriate stratifications.

Table 4: Detailed prevalence data from gnomAD<sup>24</sup> for ACE2 variants reported in Table 1.

Variant	$\Delta\Delta G$ (kcal mol <sup>-1</sup> )	AFR	AMR	ASJ	EAS	FIN	NFE	SAS	OTH
p.Ser19Pro	0.59	0.003	-	-	-	-	-	-	0.0002
p.Lys26Arg	0.31	0.001	0.003	0.01	0.00007	0.0005	0.006	0.001	0.003
p.Thr27Ala	(-0.4)	-	0.00007	-	-	-	-	-	-
p.Glu35Lys	-0.29	-	-	-	0.0001	-	0.00001	-	-
p.Glu37Lys	-1.25	0.0001	-	-	-	0.0003	-	-	-
p.Phe40Leu	0.14	<sup>a</sup> 0.0002	<sup>b</sup> 0.0001	-	-	-	-	-	-
p.Met82Ile	(0.1)	0.0003	-	-	-	-	-	-	-
p.Gly326Glu	-0.67	0.0001	-	-	-	-	-	-	-
p.Glu329Gly	-0.06	-	-	-	-	-	0.00007	-	0.0002
p.Gly352Val	(-1.2)	-	-	-	-	-	0.00001	-	-
p.Asp355Asn	No binding	-	-	-	-	-	0.00003	-	-
p.Pro389His	0.12	-	0.0002	-	-	-	0.00002	-	-
Total frequency		0.0047	0.00317	0.01	0.00017	0.0008	0.00614	0.001	0.0034
Enhancers ( $\geq 0.1$ kcal mol <sup>-1</sup> )		0.0045	0.0031	0.01	0.00007	0.0005	0.0062	0.001	0.0032
Inhibitors ( $\leq -0.1$ kcal mol <sup>-1</sup> )		0.0002	0.00007	-	0.0001	0.0003	0.00005	-	-

AFR: African/African-American, AMR: Latino/Admixed American, ASJ: Ashkenazi Jewish, EAS: East Asian, FIN: Finnish, NFE: non-Finnish European, SAS: South Asian, OTH: Other

The remaining gnomAD variants predicted to effect Spike binding are all very rare. The two other variants in our SPR series that also enhanced Spike binding have a joint frequency of around 3 in 10,000 in Latino/Admixed American (AMR) samples whilst amongst Spike inhibitory variants, p.Glu37Lys is the most prevalent occurring in 3 in 10,000 in Finnish samples and one additional African/African American sample. The other strongly inhibitory variants p.Gly352Val (predicted) and p.Asp355Asn are both doubletons observed only in non-Finnish Europeans, corresponding to an extremely low allele frequency of 3 in 100,000. Notably, the inhibitory variant p.Glu35Lys is the only variant predominant in East Asian samples

Our analysis of the effects of all possible missense mutations at the ACE2 Spike interface predicts many other potential novel variants that effect the Spike interaction, but even though a relatively high proportion of possible variants are predicted to effect Spike binding (up to 62%), their collective frequency could be as little as 2.8 in 10,000. These frequencies are comparable to those that might be expected for variants that cause rare diseases.

Although these data suggest that ACE2 alleles with large positive or negative  $\Delta\Delta G$  are extremely rare it is important to note that allele frequencies show significant variation even between large variation datasets<sup>42</sup>. Moreover, it is known that local population allele frequencies can vary substantially from those reported in public datasets<sup>43</sup> and it is therefore possible that some populations have risk or protective mutations at higher frequencies. It is also worth highlighting that the substantially increased Spike affinities of the two most common variants tested may indicate the presence of a past selective effect occurring twice independently, which may increase the possibility of other higher frequency variants amongst populations not well represented by gnomAD.

These frequencies might be useful for modellers to assess the role of rare ACE2 variants with Spike binding phenotypes on the epidemiology of Covid-19. For example, whether rare ACE2 variants that inhibit Spike binding contribute to the prevalence of asymptomatic carriers or if rare ACE2 variants with enhanced Spike binding can account for instances of particularly severe disease in the absence of typical comorbidities.

### 3.3 Conclusion

In summary, we report the binding affinities of 10 ACE2 variants with SARS-CoV-2 Spike RBD. We found that ACE2 p.Ser19Pro and p.Lys26Arg, two of the most common ACE2 missense variants in gnomAD<sup>24</sup>, substantially increase the affinity and are therefore possible risk factors for COVID-19. These variants have distinct distributions across the gnomAD cohorts and could therefore contribute population-specific risk. Two additional rare variants were identified that also enhanced Spike binding, although to a lesser extent. We confirmed our previous predictions that ACE2 p.Glu37Lys and p.Asp355Asn strongly inhibit Spike binding and are therefore potentially protective against COVID-19. We also identified two further rare variants that inhibited binding, including p.Gly326Glu, which we had previously incorrectly predicted would enhance binding. The SPR affinity data were used to recalibrate the mCSM-PPI<sup>13</sup> algorithm to provide improved affinity predictions for all possible ACE2 missense variants that interact with Spike and we estimated the prevalence of novel Spike affinity variants in ACE2. A key point in these burden assessments is the separate consideration of variants predicted to inhibit and enhance Spike affinity since these may have distinct phenotypes. Overall, p.Ser19Pro and p.Lys26Arg are still expected to have the most widespread effects, with the joint prevalence of the rare affinity variants substantially lower, but in all cases the possibility of higher prevalence in local or underrepresented populations remains and the penetrance of each variant will be a key factor. This work has applications in helping to prioritise experimental work into SARS-CoV-2 Spike human ACE2 recognition; developing genetic diagnostic risk profiling for COVID-19 susceptibility and severity, improving detection and interpretation in future COVID-19 genetic association studies and in understanding SARS-CoV-2 Spike evolution and host adaptation.

## Methods

### ACE2 and RBD constructs

The ACE2 construct was kindly provided by Ray Owens at the Oxford Protein Production Facility-UK. The RBD construct was kindly provided by Quentin Sattentau at the Sir William Dunn School of Pathology. ACE2 point mutations were added using Agilent QuikChange II XL

Site-Directed Mutagenesis Kit following the manufactures instructions. The primers were designed using the Agilent QuikChange primer design web program.

### HEK293F suspension cell culture

Cells were grown in FreeStyle™ 293 Expression Medium (12338018) in a 37 °C incubator with 8% CO<sub>2</sub> on a shaking platform at 130 rpm. Cells were passaged every 2-3 days with the suspension volume always kept below 33.3% of the total flask capacity. The cell density was kept between 0.5 and 2 million per ml.

### Transfection of HEK293F suspension cells

Cells were counted to check cell viability was above 95% and the density adjusted to 1.0 million per ml. For 100 ml transfection, 100 µl FreeStyle™ MAX Reagent (16447100) was mixed with 2 ml Opti-MEM (51985034) for 5 minutes. During this incubation 100 µg of expression plasmid was mixed with 2 ml Opti-MEM. For in situ biotinylation of ACE2 90 µg of expression plasmid was mixed with 10 µg of expression plasmid encoding the BirA enzyme. The DNA was then mixed with the MAX Reagent and incubated for 25 minutes before being added to the cell culture. For ACE2 biotinylation biotin was added to the cell culture at a final concentration of 50 µM. The culture was left for 5 days for protein expression to take place.

### Protein purification from HEK293F suspension cell supernatant

Cells were harvested by centrifugation and the supernatant collected and filtered through a 0.22 µm filter. Imidazole was added to a final concentration of 10 mM and PMSF added to a final concentration of 1 mM. 1 ml of Ni-NTA Agarose (30310) was added per 100 ml of supernatant and the mix was left on a rolling platform at 4 °C overnight. The supernatant mix was poured through a gravity flow column to collect the Ni-NTA Agarose. The Ni-NTA Agarose was washed 3 times with 25 ml of wash buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl and 20 mM imidazole at pH 8). The protein was eluted from the Ni-NTA Agarose with elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl and 250 mM imidazole at pH 8). The protein was concentrated, and buffer exchanged into size exclusion buffer (25 mM NaH<sub>2</sub>PO<sub>4</sub>, 150 mM NaCl at pH 7.5) using a protein concentrator with a 10,000 molecular weight cut-off. The protein was concentrated down to less than 500 µl before loading onto a Superdex 200 10/300 GL size exclusion column. Fractions corresponding to the desired peak were pooled and frozen at -80 °C. Samples from all observed peaks were analysed on an SDS-PAGE gel (Supplementary Figure 3).

### Surface plasmon resonance (SPR)

SARS-CoV-2 receptor binding domain binding to human extracellular ACE2 were analysed on a Biacore T200 instrument (GE Healthcare Life Sciences) at 37°C and a flow rate of 30 µl/min. Running buffer was HBS-EP (BR100669). Streptavidin was coupled to a CM5 sensor chip (29149603) using an amine coupling kit (BR100050) to near saturation, typically 10000-12000 response units (RU). Biotinylated ACE2 WT and variants were injected into the experimental flow cells (FC2–FC4) for different lengths of time to produce desired immobilisation levels (600–700 RU). FC1 was used as a reference and contained streptavidin only. Excess streptavidin was blocked with two 40 s injections of 250 µM biotin (Avidity). Before RBD injections, the chip surface was conditioned with 8 injections of the running

buffer. A dilution series of RBD was then injected simultaneously in all FCs. Buffer was injected after every 2 or 3 RBD injections. The lowest RBD concentration was injected at the beginning and at the end of each dilution series to ensure reproducibility. The length of all injections was 30 s, and dissociation was monitored from 180-300 s. Binding measured in FC1 was subtracted from the other three FCs. Additionally, all binding and dissociation data were double referenced using the closest buffer injections<sup>44</sup>. In all experiments, an ACE2-specific antibody (NOVUS Biologicals, AC384) was injected at 5 µg/ml for 10 minutes with the disassociation monitored for 10 minutes (Supplementary Figure 4). Only ACE2 T27R did not bind AC384 as expected but since this mutant displays RBD binding comparable to WT ACE2, this most likely indicates direct inhibition of AC384 binding rather than the presence of unfolded protein.

### SPR data fitting

Double referenced binding data was plotted and fit with GraphPad Prism (Supplementary Figure 5). To find the equilibrium  $K_D$  (dissociation constant) the association phase was fit with a One-phase association model and the plateau binding measurements were extracted and plotted against the corresponding concentration of RBD. This plot was then fit with a One-site specific binding model below and the value for the equilibrium  $K_D$  extracted.

$$Y = \frac{B_{\max} * X}{(K_D + X)}$$

To convert  $K_D$  values to  $\Delta G$  the equation below was used. Where R is the gas constant measured in cal mol<sup>-1</sup> K<sup>-1</sup> and T is the temperature measured in K.

$$\Delta G = R * T * \ln K_D$$

A  $\Delta\Delta G$  value could then be found for each mutant by subtracting the  $\Delta G$  of the WT from the  $\Delta G$  of each mutant.

For ACE2 variant D355N binding was too poor to fit accurately therefore, an estimate for the lower limit for the  $K_D$  was calculated using the formula below. Where the “Maximum [RBD]” is the highest concentration of RBD flown over the surface, the “Binding at  $K_D$  (WT)” is the RU value at the equilibrium  $K_D$  for the WT protein on the same chip and the “Binding at maximum [RBD] (variant)” is the maximum RU value for the RBD binding D355N at the highest concentration. The estimated  $K_D$  lower limit could then be converted into  $\Delta G$  and then  $\Delta\Delta G$  using the same method above.

$$\text{Estimated } K_D \text{ lower limit} = \text{Maximum [RBD]} * \frac{\text{Binding at } K_D \text{ (WT)}}{\text{Binding at maximum [RBD] (D355N)}}$$

### Integration of structure, variant and mutagenesis data

The pyDRSASP suite<sup>45</sup> was used to integrate 3D structure, population variant and mutagenesis assay data for analysis. Population variants from gnomAD v2<sup>24</sup> were mapped to ACE2 with VarAlign<sup>45</sup>. Residue mappings were derived from the Ensembl VEP annotations present in the gnomAD VCF. In addition, we manually checked the gnomAD multi nucleotide polymorphisms (MNPs) data file and found no records for ACE2. The structure of chimeric

SARS-CoV-2 Spike receptor binding domain in complex with human ACE2 (PDB ID: 6vw1)<sup>11</sup> was downloaded from PDB. Residue-residue contacts were calculated with ARPEGGIO<sup>46</sup>. These operations were run with our ProIntVar<sup>45</sup> Python package, which processes all these data into conveniently accessible Pandas DataFrames. ACE2 Spike interface residues were defined as those with any interprotein interatomic contact (Supplementary Table 2).

### Prediction of missense variant effects on Spike – ACE2 interaction

The mCSM-PPI2<sup>13</sup> web server was used to predict the effect of mutations on the SARS-CoV-2 Spike-ACE2 interface topology and binding affinity with the structure PDB ID: 6vw1<sup>11</sup> according to our previous protocol<sup>23</sup>.

### Recalibrated mCSM-PPI2 predictions with SPR tested variants

The SPR determined  $\Delta\Delta G$  (Kd-plateau) were regressed against the mCSM-PPI2 prediction, restricting the regression to the variants with negative predicted  $\Delta\Delta G$ , with the *lm* function in R<sup>47</sup>. Recalibrated scores were calculated with the *predict.lm* function and applied to ACE2 variants within 10 angstroms of Spike.

### Enumerating possible ACE2 missense SNPs

The ACE2 gene (ENSG00000130234) was retrieved from Ensembl in Jalview<sup>25</sup>. Two identical CDS transcripts (ENST00000427411 and ENST00000252519) were found with the Get Cross-References command corresponding to ACE2 full-length proteins (ENSP00000389326 and ENSP00000252519). These correspond to the UniProt ACE2 sequence Q9BYF1. The CDS was saved in Fasta format and this was parsed in Python with Biopython. The CDS was broken into codons and all possible single base changes were enumerated and translated using the standard genetic code. This provided the set of amino acids accessible to each residue via a single base change.

### Estimated Prevalence of Novel Rare Variants in ACE2 with Spike Affinity Phenotypes

#### Upper bound from gnomAD singleton frequency (Minimum reportable frequency)

Upper bounds for the total prevalence of potential Spike affinity variants were calculated based on the conservative assumption that novel variants must occur at lower frequencies than the minimum reportable variant frequency (i.e., minimum singleton frequency) in gnomAD. The theoretical minimum reportable frequency is the allele frequency of a singleton variant at a site where all samples have been effectively called. For alleles on the X chromosome, the proportion of XX and XY samples is important since the number of alleles sequenced is  $2N_{female} + N_{male}$ . Practically, many reported singleton frequencies are greater than this theoretical minimum because at a given loci not all samples have sufficient sequence data quality to be called. Therefore, the minimum reportable frequency varies by genomic position, as well as population. Since gnomAD reports the allele number (AN) only for variant sites, we used the maximum observed allele number in ACE2. For example, the maximum allele number in ACE2 corresponding to non-Finnish European samples is 80,119 (AN\_NFE = 80,119). This yields a minimum observed variant frequency in ACE2 amongst non-Finnish Europeans in gnomAD (v2 exomes) of  $1/80,119 = 1.2 \times 10^{-5}$ , or 2.5 variants per 200,000 alleles. The total prevalence is then found by multiplying this value by the number of SNPs being considered.

### Empirical detection rate of rare ACE2 variants and the affinity active ratio based estimate

Our second approach to estimate plausible frequencies of novel variants ( $P$ ) was to project the empirical detection rate of rare ACE2 variants in gnomAD ( $k$ ) onto the sites that correspond to the Spike interface ( $n$ ), and then adjust this by the proportion of variants that are predicted to effect Spike binding ( $\alpha$ ) so that,

$$P = \alpha kn$$

For example, there are 1,181 variant alleles in the 56,885 non-Finnish European cohort (AN\_NFE = 80,119) arising from variants with AF < 0.01 in the 2,415 nucleotide ACE2 coding sequence (n.b. at this AF threshold, 90 % are missense). This corresponds to a variation rate  $k = 6.1 \times 10^{-6}$  variants per nucleotide per allele. Projecting rate  $k$  onto the 25 ACE2 sites considered ( $n = 75$  nucleotides) we find  $kn = 4.6 \times 10^{-4}$  variants per allele, or 91.6 variants per 200,000 alleles. Note that this should be a conservative estimate of variability at these sites, since surface residues tend to be more variable than the core or other functional regions of the protein<sup>48</sup>, unless the site is under specific selection. The proportion of SNPs that are defined to have an affinity phenotype is dependent on the  $\Delta\Delta G^{\text{recal}}$  threshold. For example, there are 38 substitutions corresponding to 41 single nucleotide variants out of the a possible 225 that are predicted to inhibit binding with  $\Delta\Delta G^{\text{recal}} < -1.0 \text{ kcal mol}^{-1}$  (Table 3) so that  $\alpha = 41 / 225 = 0.18$ . Altogether this gives  $P = 8.2 \times 10^{-5}$  for variants with  $\Delta\Delta G^{\text{recal}} < -1.0 \text{ kcal mol}^{-1}$ . This calculation could be improved (e.g., to account for missense/synonymous ratios in  $\alpha$ ) but in its current form is suitable to provide estimates that indicate plausible orders of magnitude of these variants' prevalence as intended.

### Software

Jalview 2.11<sup>25</sup> was used for interactive sequence data retrieval, sequence analysis, structure data analysis and figure generation. UCSF Chimera<sup>26</sup> and PyMol<sup>29</sup> were used for structure analysis and figure generation.

The pyDRSASP<sup>45</sup> packages, comprising ProteoFAV, ProIntVar and VarAlign were used for data retrieval and analysis. Biopython was used to process sequence data.

Data analyses were coded in R and Python Jupyter Notebooks. Numpy, Pandas and Scipy were used for data analysis. Matplotlib, Seaborn and ggplot2 were used to plot data.

### Code availability

All code and analysis notebooks used in this study are available from the Barton Group public GitHub repository at <https://github.com/bartongroup/covid19-ace2-variants>.

### Data availability

This study employed several public datasets, which are available from their original sources and all necessary identifiers and accessions are provided in the methods. Data that was compiled and calculated as part of this study are available from the BioStudies database (<https://www.ebi.ac.uk/biostudies>) under accession S-BSST649.



## References

- 1 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).
- 2 Verity, R. *et al.* Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*, doi:10.1016/S1473-3099(20)30243-7 (2020).
- 3 Guan, W. J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*, doi:10.1056/NEJMoa2002032 (2020).
- 4 Cao, Y. *et al.* Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell Discov* **6**, 11, doi:10.1038/s41421-020-0147-1 (2020).
- 5 Païro-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92-98, doi:10.1038/s41586-020-03065-y (2021).
- 6 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, doi:10.1016/j.cell.2020.02.052 (2020).
- 7 Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *Journal of Virology* **94**, e00127-00120, doi:10.1128/JVI.00127-20 (2020).
- 8 Johnson, B. A. *et al.* Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **591**, 293-299, doi:10.1038/s41586-021-03237-4 (2021).
- 9 Heurich, A. *et al.* TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. *J Virol* **88**, 1293-1307, doi:10.1128/JVI.02202-13 (2014).
- 10 Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864-1868, doi:10.1126/science.1116480 (2005).
- 11 Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature*, doi:10.1038/s41586-020-2179-y (2020).
- 12 Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* **24**, 1634-1643, doi:10.1038/sj.emboj.7600640 (2005).
- 13 Rodrigues, C. H. M., Myung, Y., Pires, D. E. V. & Ascher, D. B. mCSM-PPI2: predicting the effects of mutations on protein-protein interactions. *Nucleic Acids Res* **47**, W338-W344, doi:10.1093/nar/gkz383 (2019).
- 14 David, A. & Sternberg, M. J. The Contribution of Missense Mutations in Core and Rim Residues of Protein-Protein Interfaces to Human Disease. *J Mol Biol* **427**, 2886-2898, doi:10.1016/j.jmb.2015.07.004 (2015).
- 15 Lindesmith, L. *et al.* Human susceptibility and resistance to Norwalk virus infection. *Nat Med* **9**, 548-553, doi:10.1038/nm860 (2003).
- 16 Hashizume, M., Gonzalez, G., Ono, C., Takashima, A. & Iwasaki, M. Population-Specific ACE2 Single-Nucleotide Polymorphisms Have Limited Impact on SARS-CoV-2 Infectivity In Vitro. *Viruses* **13**, doi:10.3390/v13010067 (2021).
- 17 Suryamohan, K. *et al.* Human ACE2 receptor polymorphisms and altered susceptibility to SARS-CoV-2. *Commun Biol* **4**, 475, doi:10.1038/s42003-021-02030-3 (2021).

- 18 Othman, H. *et al.* Interaction of the spike protein RBD from SARS-CoV-2 with ACE2: Similarity with SARS-CoV, hot-spot analysis and effect of the receptor polymorphism. *Biochem Biophys Res Commun* **527**, 702-708, doi:10.1016/j.bbrc.2020.05.028 (2020).
- 19 Hussain, M. *et al.* Structural Variations in Human ACE2 may Influence its Binding with SARS-CoV-2 Spike Protein. *Journal of Medical Virology* **n/a**, doi:10.1002/jmv.25832 (2020).
- 20 Asselta, R., Paraboschi, E. M., Mantovani, A. & Duga, S. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. *Aging (Albany NY)* **12**, 10087-10098, doi:10.18632/aging.103415 (2020).
- 21 Benetti, E. *et al.* ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. *Eur J Hum Genet* **28**, 1602-1614, doi:10.1038/s41431-020-0691-z (2020).
- 22 Chan, K. K. *et al.* Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* **369**, 1261-1265, doi:10.1126/science.abc0870 (2020).
- 23 MacGowan, S. A. & Barton, G. J. Missense variants in ACE2 are predicted to encourage and inhibit interaction with SARS-CoV-2 Spike and contribute to genetic risk in COVID-19. *bioRxiv*, 2020.2005.2003.074781, doi:10.1101/2020.05.03.074781 (2020).
- 24 Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210, doi:10.1101/531210 (2019).
- 25 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191, doi:10.1093/bioinformatics/btp033 (2009).
- 26 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
- 27 Burley, S. K. & Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett* **203**, 139-143, doi:10.1016/0014-5793(86)80730-x (1986).
- 28 Barton, M. I. *et al.* Effects of common mutations in the SARS-CoV-2 Spike RBD domain and its ligand the human ACE2 receptor on binding affinity and kinetics. *bioRxiv*, 2021.2005.2018.444646, doi:10.1101/2021.05.18.444646 (2021).
- 29 Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 2.3.5. (2015).
- 30 Patel, P., Hiam, L., Sowemimo, A., Devakumar, D. & McKee, M. Ethnicity and covid-19. *BMJ* **369**, m2282, doi:10.1136/bmj.m2282 (2020).
- 31 The, C.-H. G. I. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *European Journal of Human Genetics* **28**, 715-718, doi:10.1038/s41431-020-0636-6 (2020).
- 32 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
- 33 Hasegawa, K. *et al.* Affinity thresholds for membrane fusion triggering by viral glycoproteins. *J Virol* **81**, 13149-13157, doi:10.1128/JVI.01415-07 (2007).
- 34 Cao, W. & Li, T. COVID-19: towards understanding of pathogenesis. *Cell Res*, 1-3, doi:10.1038/s41422-020-0327-4 (2020).
- 35 Wolfel, R. *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature*, doi:10.1038/s41586-020-2196-x (2020).

- 36 Ke, Z. *et al.* Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **588**, 498-502, doi:10.1038/s41586-020-2665-2 (2020).
- 37 Supasa, P. *et al.* Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell*, doi:10.1016/j.cell.2021.02.033 (2021).
- 38 Leung, K., Shum, M. H., Leung, G. M., Lam, T. T. & Wu, J. T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill* **26**, doi:10.2807/1560-7917.ES.2020.26.1.2002106 (2021).
- 39 Lu, J. & Sun, P. D. High affinity binding of SARS-CoV-2 spike protein enhances ACE2 carboxypeptidase activity. *J Biol Chem* **295**, 18579-18588, doi:10.1074/jbc.RA120.015303 (2020).
- 40 Hofmann, H. & Pohlmann, S. Cellular entry of the SARS coronavirus. *Trends Microbiol* **12**, 466-472, doi:10.1016/j.tim.2004.08.008 (2004).
- 41 Fang, L., Karakiulakis, G. & Roth, M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir Med* **8**, e21, doi:10.1016/S2213-2600(20)30116-8 (2020).
- 42 Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291, doi:10.1038/nature19057 (2016).
- 43 Rentoft, M. *et al.* A geographically matched control population efficiently limits the number of candidate disease-causing variants in an unbiased whole-genome analysis. *PLoS One* **14**, e0213350, doi:10.1371/journal.pone.0213350 (2019).
- 44 Myszka, D. G. Improving biosensor analysis. *J Mol Recognit* **12**, 279-284, doi:10.1002/(SICI)1099-1352(199909/10)12:5<279::AID-JMR473>3.0.CO;2-3 (1999).
- 45 MacGowan, S. A. *et al.* The Dundee Resource for Sequence Analysis and Structure Prediction. *Protein Sci* **29**, 277-297, doi:10.1002/pro.3783 (2020).
- 46 Jubb, H. C. *et al.* Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J Mol Biol* **429**, 365-371, doi:10.1016/j.jmb.2016.12.004 (2017).
- 47 Team, R. C. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* (2017).
- 48 MacGowan, S. A. *et al.* Human Missense Variation is Constrained by Domain Structure and Highlights Functional and Pathogenic Residues. *bioRxiv*, 127050, doi:10.1101/127050 (2017).

## Acknowledgements

We thank Johannes Pettmann for help with protein expression and Anna Huhn for help with data analysis. We thank Dr Jim Procter for alerting us to the release of the structure of ACE2 in complex with SARS-CoV-2 S and his initial observations regarding variants at the ACE2-S interface. We thank the Dundee Research Computing team for supporting our IT infrastructure and remote working. This work was supported by Biotechnology and Biological Sciences Research Council Grants (BB/J019364/1 and BB/R014752/1) and Wellcome Trust Biomedical Resources Grant (101651/Z/13/Z). OD is supported by a Wellcome Trust Senior Fellowship in Basic Biomedical Sciences (207537/Z/17/Z828).

## Author contributions

SM performed the computational and population genetic analyses, compiled and analysed data, and wrote the first draft manuscript. MB conducted all wet lab experiments and compiled and analysed the associated data. GJB and SM conceived the research. All authors interpreted results and edited and reviewed the final manuscript.

## Author information

GJB and SM: Division of Computational Biology, School of Life Sciences, University of Dundee, Scotland, United Kingdom

MB, MK, OD and AvdW: Sir William Dunn School of Pathology, University of Oxford, Oxford, Oxfordshire, United Kingdom

## Conflict of interest statement

AvdW declares ownership of shares in BioNTech SE.

## Extended Data

Supplementary Table 1. mCSM-PPI2 predictions for all *gnomAD* ACE2 missense variants in residues resolved in PDB 6vw1.

ID	Mutation	Distance to Interface	mCSM-PPI2 $\Delta\Delta G$	Allele Count	AC Male	AC Popn. Max	Popn. Max
rs73635825	p.Ser19Pro	2.6	-0.2	40	6	39	AFR
-	p.Ile21Thr	7.3	0.3	1	1	1	AMR
rs778030746	p.Ile21Val	7.3	-0.4	2	2	2	NFE
rs756231991	p.Glu23Lys	6.2	-0.4	1	0	1	NFE
rs4646116	p.Lys26Arg	6.0	0.0	704	258	85	ASJ
-	p.Lys26Glu	6.0	0.1	1	1	1	NFE
rs781255386	p.Thr27Ala	3.7	-0.6	2	0	2	AMR
-	p.Glu35Lys	2.9	-0.5	3	1	2	EAS
rs146676783	p.Glu37Lys	3.2	-1.2	6	3	5	FIN
-	p.Phe40Leu	8.6	-0.3	3	0	3	AMR
-	p.Ser43Arg	8.5	-0.1	1	0	1	SAS
-	p.Tyr50Phe	13.0	-0.3	1	0	1	NFE
rs760159085	p.Asn51Asp	14.2	0.1	2	1	2	SAS
rs775273812	p.Thr55Ala	15.3	0.0	1	1	1	NFE
rs771621249	p.Asn58Lys	12.0	-0.3	2	0	2	SAS
-	p.Asn58His	12.0	-0.1	2	1	2	NFE
rs759162332	p.Gln60Arg	15.5	0.1	2	0	2	NFE
-	p.Met62Val	13.5	0.0	1	1	1	NFE
-	p.Asn64Lys	13.2	0.1	1	1	1	AFR
rs755691167	p.Lys68Glu	7.3	0.1	2	2	2	SAS
-	p.Phe72Val	8.7	-0.4	1	1	1	EAS
-	p.Ser77Phe	9.3	-0.1	1	1	1	NFE
rs766996587	p.Met82Ile	3.5	-0.3	0	0	0	
rs766996587	p.Met82Ile	3.5	-0.3	2	0	2	AFR
rs759134032	p.Pro84Thr	7.1	-0.1	1	0	1	SAS
rs746808776	p.Gln86Arg	13.1	0.2	2	1	2	NFE
rs763395248	p.Thr92Ile	15.1	-0.1	2	2	2	NFE
-	p.Gln102Pro	14.8	-0.2	2	0	1	NFE
rs143158922	p.Asn103His	14.8	0.0	2	0	2	AFR
rs139773121	p.Val107Ala	19.8	-0.1	2	1	2	AFR
rs201900069	p.Arg115Gln	30.4	0.0	30	15	9	EAS
rs751227277	p.Ser128Thr	28.6	-0.1	1	0	1	SAS
rs766124365	p.Pro138Ala	45.4	-0.1	1	1	1	NFE
-	p.Cys141Tyr	38.5	0.0	1	0	1	AFR
-	p.Asn154Lys	38.4	0.0	2	0	1	AMR
rs746034076	p.Asn159Ser	49.4	0.1	2	1	1	AMR
-	p.Ala164Ser	42.6	0.3	1	0	1	AFR
rs769593006	p.Glu166Gln	45.4	-0.5	1	0	1	NFE
rs748076875	p.Glu171Val	41.6	-0.1	1	1	1	NFE
rs754511501	p.Gly173Ser	40.6	-0.4	4	2	1	SAS
rs779651019	p.Pro178Leu	40.2	-0.1	1	0	1	NFE
rs758142853	p.Val184Ala	33.1	-0.1	8	1	8	EAS
rs750052167	p.Leu186Ser	28.7	-0.1	1	0	1	AMR
-	p.Met190Thr	23.6	0.1	1	0	1	SAS
rs765733397	p.Ala191Pro	24.7	0.0	1	0	1	NFE
rs762219565	p.Ala193Glu	21.1	0.2	3	0	3	AMR
rs764661406	p.His195Asn	19.1	-0.1	1	1	1	SAS
rs764661406	p.His195Tyr	19.1	0.0	1	0	1	NFE
-	p.Glu197Gly	28.0	0.0	2	0	1	ASJ
-	p.Asp198Asn	29.0	-0.7	1	0	1	NFE
rs750145841	p.Tyr199Cys	28.1	-0.4	4	3	1	AMR
-	p.Tyr199His	28.1	-0.2	1	0	1	SAS
rs779790336	p.Arg204Thr	28.9	-0.2	1	0	1	SAS
rs142443432	p.Asp206Gly	22.4	0.0	52	23	50	NFE
rs754237613	p.Tyr207Cys	26.2	-0.4	1	0	1	NFE
-	p.Val209Ile	25.6	0.0	1	1	1	NFE
rs148771870	p.Gly211Arg	23.9	-0.3	243	91	165	NFE
rs761269389	p.Asp216Tyr	27.3	0.2	1	0	1	NFE
rs753164828	p.Asp216Glu	27.3	-0.1	2	1	2	NFE
rs372272603	p.Arg219Cys	24.9	-0.2	58	17	52	NFE
rs759590772	p.Arg219His	24.9	0.0	18	11	17	SAS
rs774621083	p.Gly220Ser	32.6	-0.1	3	0	1	AFR
-	p.Thr229Ile	37.5	-0.1	1	0	1	NFE

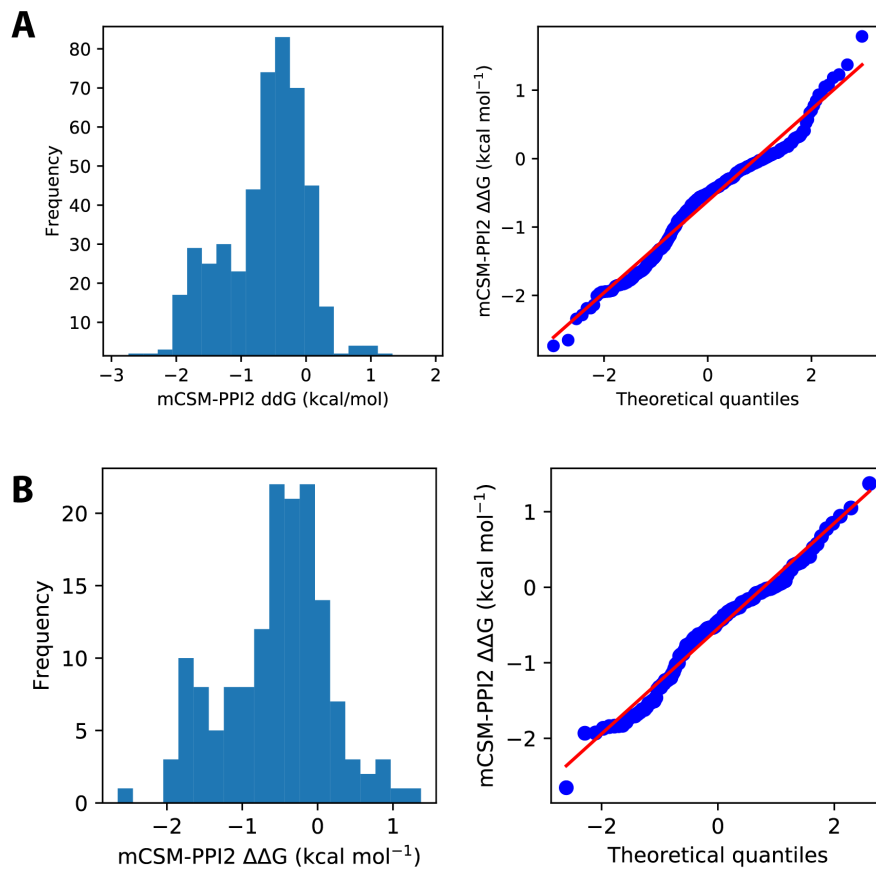
ID	Mutation	Distance to Interface	mCSM-PPI2 $\Delta\Delta G$	Allele Count	AC Male	AC Popn. Max	Popn. Max
-	p.His239Gln	49.7	0.0	1	0	1	NFE
-	p.Ala242Val	50.9	0.4	3	1	3	NFE
rs771769548	p.Tyr252Cys	47.9	-0.3	2	0	2	NFE
rs745514718	p.Ser257Asn	56.7	-0.3	1	0	1	EAS
-	p.Ile259Thr	57.2	-0.1	1	0	1	AFR
rs200745906	p.Pro263Ser	47.8	0.1	9	0	9	NFE
-	p.Gly268Cys	38.9	-0.7	1	0	1	ASJ
-	p.Asp269Glu	36.9	-0.4	0	0	0	
rs766319182	p.Met270Val	37.7	-0.1	4	2	4	NFE
-	p.Ser280Tyr	43.8	0.0	1	1	1	NFE
rs763264372	p.Gln287Arg	49.6	0.0	1	0	1	NFE
-	p.Gln287Lys	49.6	-0.1	1	0	1	AMR
rs763994205	p.Asn290His	39.9	0.0	2	1	1	SAS
rs756358940	p.Ile291Lys	36.3	0.0	3	2	3	NFE
-	p.Asp292Asn	35.0	-0.5	1	0	1	NFE
rs776226831	p.Asp295Gly	34.8	-0.1	7	1	6	NFE
rs772143709	p.Met297Leu	27.6	-0.3	1	0	1	AFR
rs745744395	p.Met297Ile	27.6	-0.2	1	1	1	NFE
rs772143709	p.Met297Val	27.6	-0.1	0	0	0	
rs773936807	p.Gln300Arg	29.5	-0.1	2	0	2	NFE
rs749750821	p.Asp303Asn	22.0	-0.8	2	0	2	SAS
rs766518362	p.Phe308Leu	17.0	-0.5	1	0	1	NFE
rs780574871	p.Glu312Lys	12.5	-0.4	3	1	2	AFR
rs759579097	p.Gly326Glu	5.5	1.0	1	0	1	AFR
rs143936283	p.Glu329Gly	4.1	-0.4	5	0	4	NFE
rs185525294	p.Met332Leu	11.2	-0.2	3	2	3	AFR
rs762357444	p.Gly337Arg	20.2	0.0	1	1	1	NFE
-	p.Asn338Ser	22.9	0.0	3	1	2	AMR
rs772710779	p.Val339Gly	20.0	-0.2	1	1	1	SAS
rs138390800	p.Lys341Arg	18.1	0.0	53	10	46	AFR
-	p.Pro346Ser	18.9	0.0	1	1	1	NFE
rs370610075	p.Gly352Val	5.4	-1.1	2	0	2	NFE
-	p.Asp355Asn	3.5	-1.3	2	0	2	NFE
-	p.Met360Leu	16.0	-0.1	1	1	1	NFE
rs758568640	p.Met366Thr	28.4	-0.1	2	0	2	NFE
-	p.Asp368Asn	23.6	-0.4	1	0	1	AMR
-	p.His374Arg	20.1	0.0	1	0	1	AMR
-	p.Glu375Asp	16.6	-0.2	3	1	3	AMR
rs767462182	p.Gly377Glu	17.3	-0.5	1	1	1	NFE
rs142984500	p.His378Arg	15.3	0.1	15	4	15	NFE
rs751572714	p.Gln388Leu	6.6	-0.1	4	3	4	AMR
rs762890235	p.Pro389His	8.1	-0.1	7	2	4	AMR
-	p.Asn397Asp	20.9	-0.1	2	0	2	NFE
rs772619843	p.Glu398Lys	22.5	-0.4	2	1	2	NFE
-	p.Gly405Glu	22.7	-0.1	1	1	1	NFE
-	p.Lys419Thr	33.1	0.0	1	1	1	EAS
-	p.Ile421Thr	26.4	-0.2	1	1	1	NFE
-	p.Pro426Ala	38.8	0.0	1	0	0	
-	p.Asp427Tyr	41.7	0.1	2	1	2	AFR
-	p.Gln429Lys	44.4	0.0	1	1	1	SAS
-	p.Asn437His	42.2	0.1	0	0	0	
rs764772589	p.Thr445Met	34.1	-0.1	1	1	1	SAS
rs776328956	p.Val447Phe	38.5	0.9	10	1	4	FIN
rs763655186	p.Gly448Glu	39.7	-0.6	1	1	1	EAS
rs760321012	p.Leu450Val	35.9	-0.2	1	0	1	NFE
rs775744448	p.Met455Ile	39.9	-0.1	1	1	1	SAS
-	p.Trp461Arg	34.1	-0.3	1	0	1	AMR
rs772445388	p.Val463Ile	37.7	-0.5	1	1	1	SAS
rs774978137	p.Gly466Trp	37.7	0.0	1	0	1	EAS
-	p.Glu467Lys	40.4	-0.1	4	1	1	AMR
rs191860450	p.Ile468Val	42.6	-0.5	142	40	138	EAS
-	p.Lys481Asn	42.4	-0.1	0	0	0	
rs748359955	p.Arg482Gln	49.3	-0.1	2	1	2	NFE
rs779752560	p.Glu483Asp	49.2	0.1	7	0	7	AMR
rs200973492	p.Val488Ala	49.7	0.0	1	0	1	AFR
rs750415079	p.Val491Met	50.7	-0.4	1	1	1	NFE
rs765152220	p.Asp494Val	49.7	-0.2	8	3	4	SAS
rs140473595	p.Ala501Thr	36.2	0.5	2	0	2	NFE
-	p.Phe504Leu	26.3	-0.6	1	0	1	NFE

ID	Mutation	Distance to Interface	mCSM-PPI2 $\Delta\Delta G$	Allele Count	AC Male	AC Popn. Max	Popn. Max
rs760281053	p.Phe504Ile	26.3	-0.3	2	0	2	NFE
-	p.His505Arg	28.6	0.0	1	0	1	EAS
rs775181355	p.Val506Ala	33.0	0.1	1	1	1	NFE
rs779199005	p.Tyr510His	23.2	-0.1	0	0	0	
rs868506911	p.Arg514Gln	23.4	-0.1	0	0	0	
-	p.Arg518Thr	27.3	0.4	1	1	1	SAS
-	p.Thr519Ile	30.3	0.0	1	1	1	NFE
rs759200680	p.Tyr521His	26.7	-0.3	1	0	1	NFE
rs763593286	p.Ala532Thr	28.5	0.1	10	5	10	SAS
rs773746588	p.Lys534Arg	32.6	-0.1	1	1	1	NFE
-	p.Pro538Leu	38.7	0.1	1	0	1	SAS
-	p.Lys541Ile	35.8	-0.1	1	0	1	AFR
rs769821600	p.Ile544Asn	25.7	-0.2	1	0	1	NFE
rs756905974	p.Asn546Ser	25.3	-0.1	1	0	1	SAS
rs761944150	p.Asn546Asp	25.3	0.0	4	1	3	AFR
rs373025684	p.Ser547Cys	24.3	-0.2	43	19	36	NFE
rs373025684	p.Ser547Phe	24.3	0.0	1	0	1	EAS
-	p.Arg559Ser	11.7	-0.1	1	0	1	AMR
rs375352455	p.Ser563Leu	15.2	0.1	1	0	1	NFE
rs750955307	p.Pro565Ser	21.6	0.0	1	1	1	NFE
-	p.Thr567Ala	25.4	0.0	1	0	1	NFE
-	p.Val573Ala	25.4	0.1	2	1	2	EAS
-	p.Val574Ile	28.4	-0.2	1	1	1	SAS
-	p.Gly575Arg	30.3	-0.1	1	0	1	NFE
rs372924787	p.Arg582Ser	38.7	0.0	1	0	1	AFR
rs150172355	p.Arg582Lys	38.7	0.0	2	0	2	NFE
-	p.Leu585Pro	39.8	-0.7	1	0	1	EAS
rs768685219	p.Asn586Tyr	40.9	0.0	1	0	1	FIN
rs760929786	p.Phe588Ser	38.6	-0.4	1	0	1	AMR
-	p.Glu589Gly	43.9	-0.1	1	0	1	FIN
rs140857723	p.Thr593Asn	49.6	-0.1	1	0	1	AFR
rs148036434	p.Leu595Val	47.7	-0.2	2	1	2	NFE
rs145437639	p.Asp597Glu	54.8	0.0	13	4	11	AFR
-	p.Thr608Ile	53.9	0.0	1	1	1	SAS
rs747988885	p.Asp609Asn	57.4	0.0	2	1	1	AFR
rs201715513	p.Ala614Ser	60.5	0.0	35	17	7	FIN

Supplementary Table 2. Structural description of the ACE2 – Spike-protein interface.

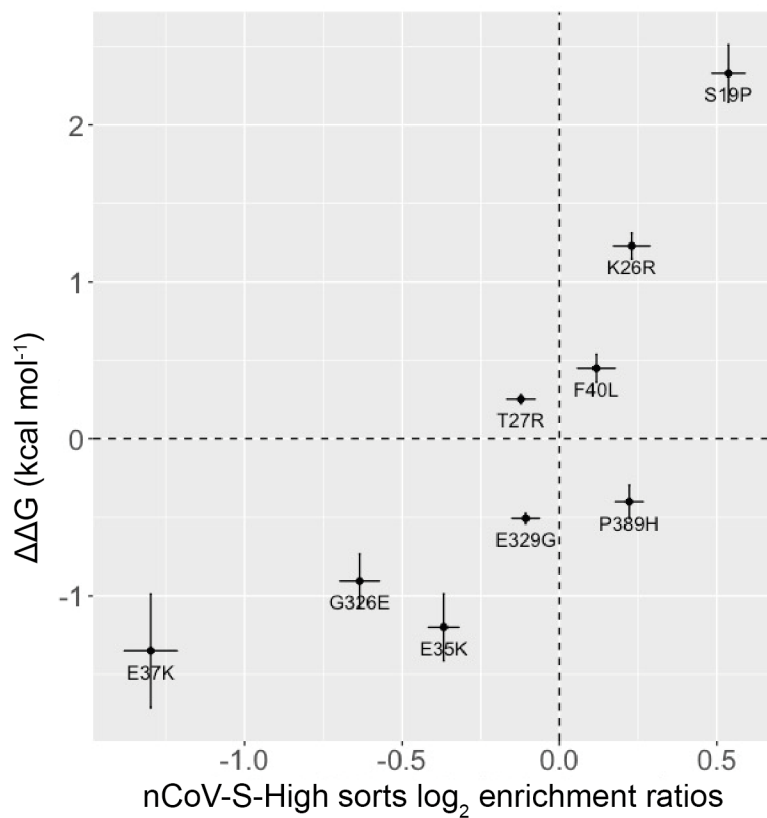
Residue	Contact Number	Min. VdW Distance	Spike counterparts	ARPEGGIO Interaction Types
S19	3	0.0	A475, G476, S477	Polar-Bond, VdW-Bond, Clash
Q24	6	0.0	A475, N487, G476, F486, S477, Y489	H-Bond, Clash
T27	5	0.4	A475, N487, F456, Y473, Y489	VdW-Proximal
F28	1	0.2	Y489	H-Bond, Polar-Bond
D30	2	0.7	L455, F456	Hydrophobic-Bond
K31	6	0.0	Q493, L455, L492, F456, F490, Y489	Hydrophobic-Bond, Clash
H34	3	0.0	Q493, L455, Y453	Aromatic-Bond, H-Bond, Hydrophobic-Bond, Polar-Bond, VdW-Bond, Clash
E35	1	0.0	Q493	H-Bond, Polar-Bond, Clash
E37	1	0.1	Y505	H-Bond, Hydrophobic-Bond, Polar-Bond
D38	4	0.0	Q498, G496, HOH701, Y449	H-Bond, Polar-Bond, VdW-Bond, Clash
Y41	4	0.0	N501, Q498, HOH701, T500	H-Bond, Polar-Bond, VdW-Bond, Clash
Q42	4	0.0	Q498, HOH701, T446, Y449	VdW-Bond, Clash
L45	2	0.4	Q498, T500	VdW-Proximal
L79	2	0.1	G485, F486	Hydrophobic-Bond
M82	1	0.1	F486	Hydrophobic-Bond, VdW-Bond
Y83	3	0.0	N487, F486, Y489	Aromatic-Bond, H-Bond, Hydrophobic-Bond, Polar-Bond, VdW-Bond, Clash
Q325	2	1.4	R439, V503	VdW-Proximal
E329	1	1.0	R439	VdW-Proximal
N330	1	0.3	T500	VdW-Proximal
K353	7	0.0	N501, Q498, G496, G502, HOH701, F497, Y505	H-Bond, Hydrophobic-Bond, Polar-Bond, VdW-Bond, Clash
G354	3	0.2	N501, G502, Y505	Polar-Bond
D355	2	0.2	G502, T500	VdW-Proximal
R357	1	0.2	T500	VdW-Proximal
R393	1	0.5	Y505	VdW-Proximal

Residue	Contact Number	Min. VdW Distance	Spike counterparts	ARPEGGIO Interaction Types
N90-glycan (BMA703)	2	0.1	R408, T415	H-Bond, Polar-Bond, VdW-Bond



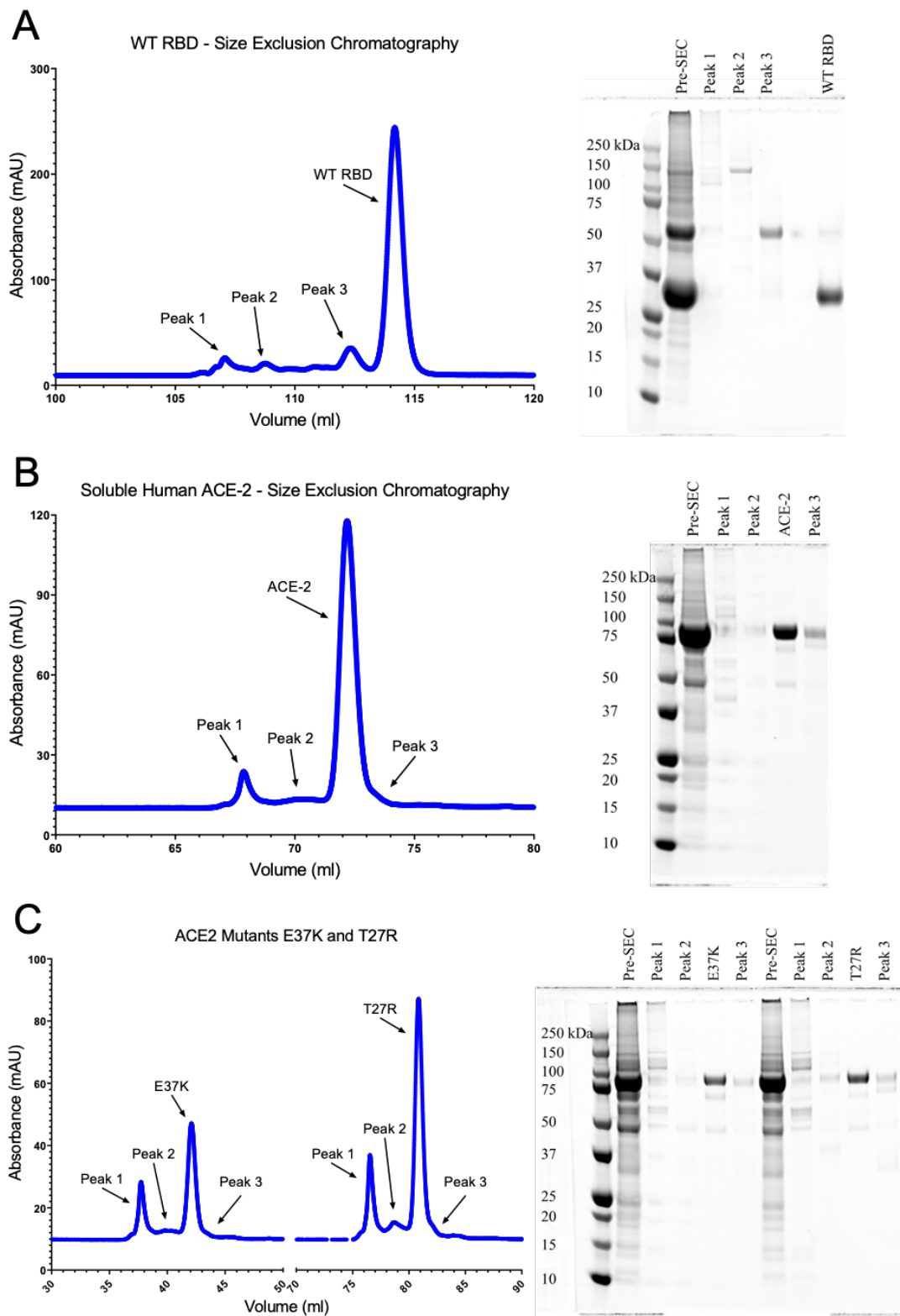
Supplementary Figure 1. A. Distribution of mCSM-PPI2<sup>23</sup> predicted  $\Delta\Delta G$  from *in silico* saturation mutagenesis of the ACE2-S interface in PDB 6vw1<sup>11</sup>. A. predicted  $\Delta\Delta G$  for 475 mutations across 25 sites on ACE2 corresponding to the 23 residues within 5 Å of SARS-CoV-2 S plus Gly326 and Gly352. B. predicted  $\Delta\Delta G$  for the subset of 151 mutations across these sites that are accessible via a single base change of the ACE2 coding sequence.



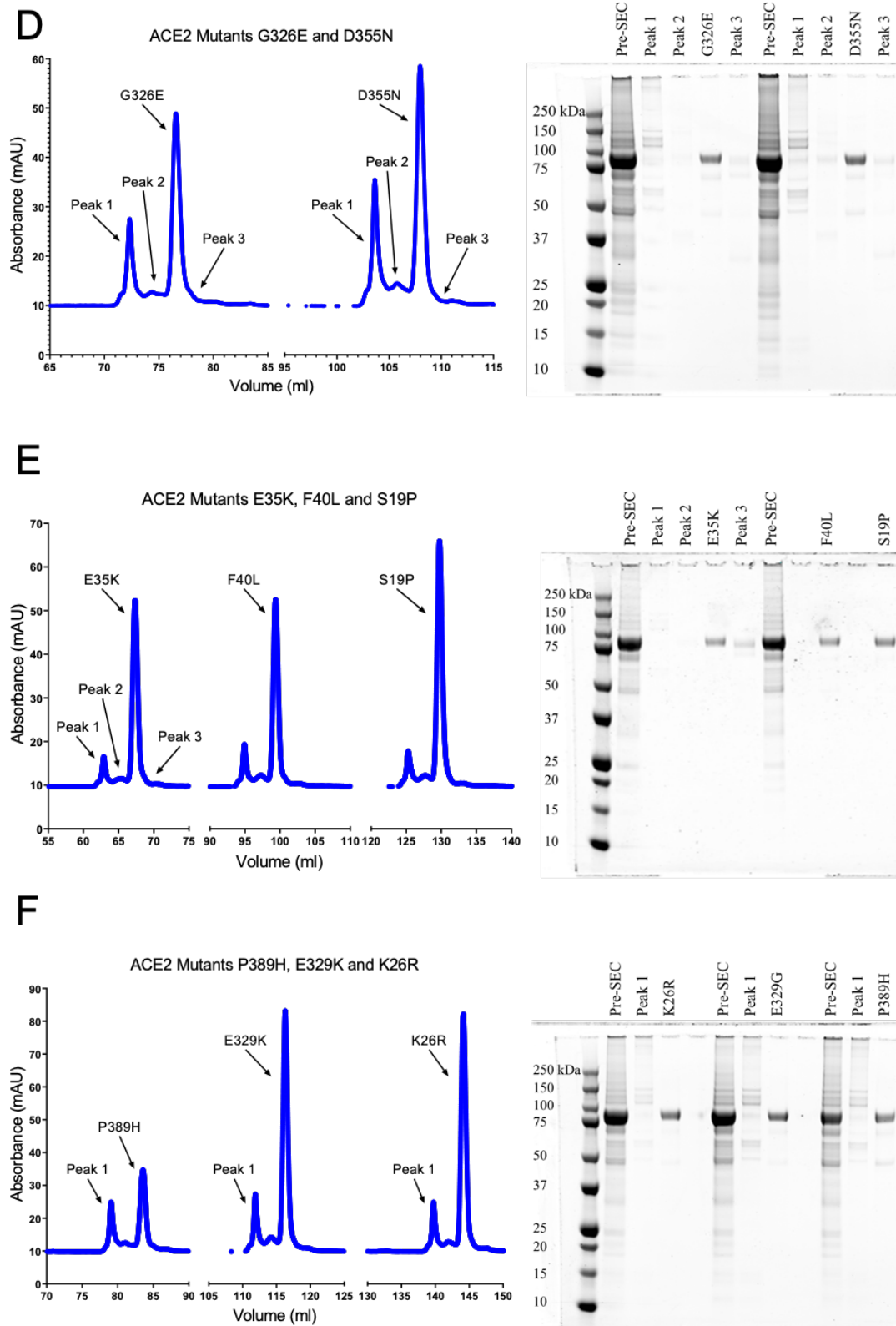


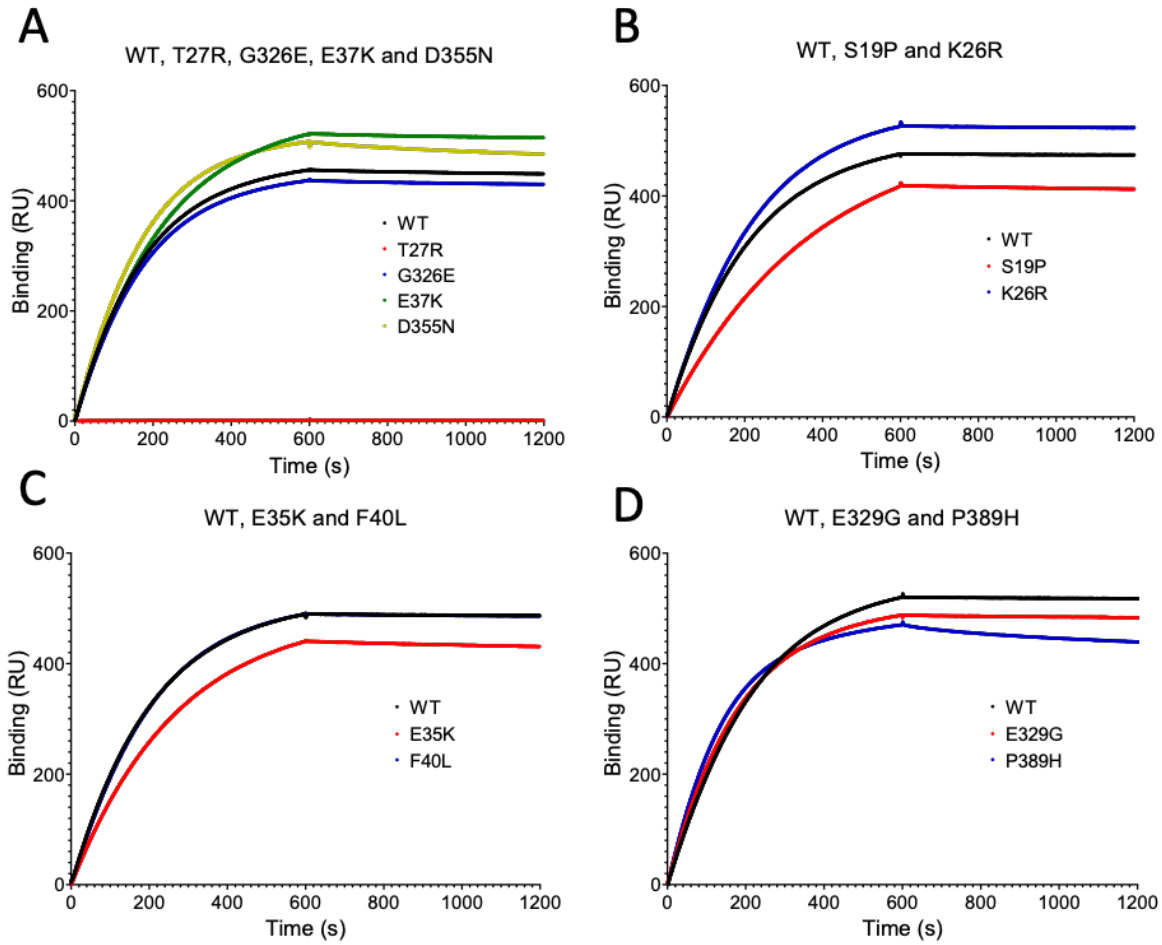
Supplementary Figure 2. Comparison of SPR  $\Delta\Delta G$  with the  $\log_2$  enrichment ratios from the nCoV-S-High sorts from Procko and co-workers<sup>22</sup>. Figure generated with R ggplot2.

Supplementary Figure 3. Protein purification size exclusion chromatography and corresponding SDS-PAGE of labelled peaks. A. WT RBD, B. WT ACE2, (C) ACE2 mutants E37K and T27R, (continued next page) D. ACE2 mutants G326E and D355N, E. ACE2 mutants E35K, F40L and S19P, F. ACE2 mutants P329H, E329K and K26R. Each panel shows the *l*.

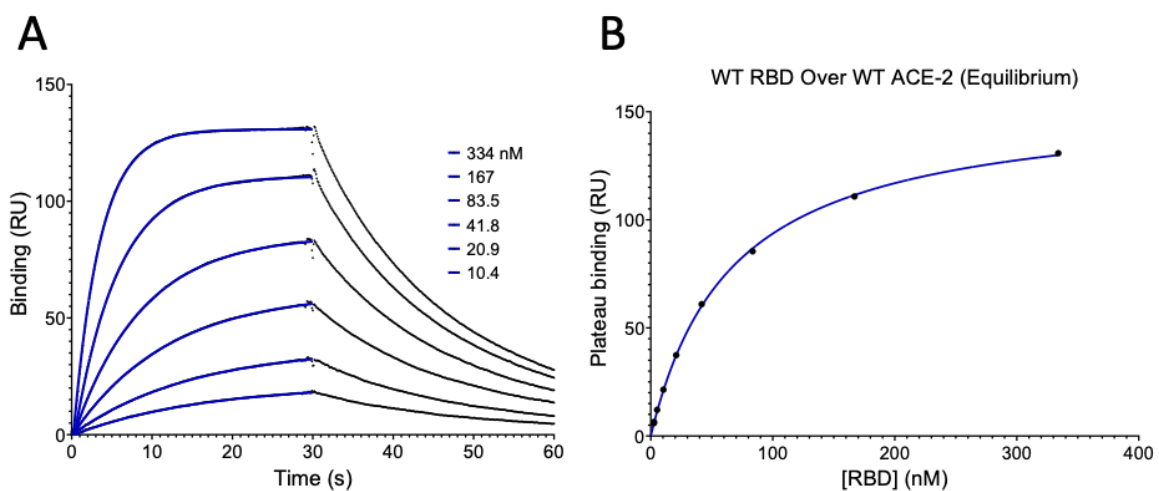


Supplementary Figure 1 (continued)





Supplementary Figure 4. Anti ACE2 Antibody (NOVUS AC384-NBP2-80038) binding to WT and 10 ACE2 variants. A. Anti ACE2 binding to WT and four ACE2 variants: T27R, G326E, E37K and D355N. B. Binding to WT, S19P and K26R. C. Binding to WT, E35K and F40L. D. Binding to WT, E329G and P389H. Anti ACE2 was injected for 600 s association and buffer for a further 600 s for dissociation.



Supplementary Figure 5. WT RBD binding WT ACE2 plus equilibrium  $K_D$ . A. Representative binding trace for WT RBD binding immobilised WT ACE2. Injections start from lowest to highest concentration of RBD and fits are shown in blue. B. Plateau binding values from A are plotted against the concentration of RBD injected and the equilibrium  $K_D$  is extracted from the fit shown in blue.