

Human ACE2 receptor polymorphisms predict SARS-CoV-2 susceptibility

Eric W. Stawiski^{1*}, Devan Diwanji^{2,3*}, Kushal Suryamohan^{1*}, Ravi Gupta⁴, Frederic A. Fellouse⁵, J. Fah Sathirapongsasuti¹, Jiang Liu⁶, Ying-Ping Jiang⁶, Aakrosh Ratan^{7,8}, Monika Mis¹, Devi Santhosh¹, Sneha Somasekar⁹, Sangeetha Mohan¹⁰, Sameer Phalke¹⁰, Boney Kuriakose¹¹, Aju Antony¹⁰, Jagath R. Junutula⁶, Stephan C. Schuster^{8,12}, Natalia Jura^{2,3#}, Somasekar Seshagiri^{6,13#}

¹Research and Development Department, MedGenome Inc., Foster City, CA 94404, USA;

²Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA

94158, USA; ³Department of Cellular and Molecular Pharmacology University of California San

Francisco, San Francisco, CA 94158, USA; ⁴MedGenome Labs Ltd., 3rd Floor, Narayana

Nethralaya Building, Narayana Health City, #258/A, Bommasandra, Hosur Road, Bangalore,

Karnataka 560099, India; ⁵ModMab Therapeutics, Accelerator for Donnelly Collaboration,

University of Toronto, Toronto, Ontario M5S 1A8, Canada; ⁶ModMab Therapeutics, 348 Hatch

Drive, Foster City, CA 94404, USA; ⁷Center for Public Health Genomics, University of Virginia,

Charlottesville, VA, USA; ⁸GenomeAsia100K Consortium; ⁹Midwestern University, Glendale, AZ

85308; ¹⁰Department of Molecular Biology, SciGenom Labs Pvt Ltd, Kerala 682037, India;

¹¹AgriGenome Labs Private Ltd, Kochi, Kerala 682030, India; ¹²Singapore Centre for

Environmental Life Sciences Engineering, Nanyang Technological University, Singapore,

Singapore; ¹³SciGenom Research Foundation, 3rd Floor, Narayana Nethralaya Building,

Narayana Health City, #258/A, Bommasandra, Hosur Road, Bangalore, Karnataka, 560099,

India.

* co-first author

correspondence: NJ - natalia.Jura@ucsf.edu and SS – sekar@sgrf.org

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of coronavirus disease (COVID-19) that has resulted in a global pandemic. It is a highly contagious positive strand RNA virus and its clinical presentation includes severe to critical respiratory disease that appears to be fatal in ~3-5% of the cases. The viral spike (S) coat protein engages the human angiotensin-converting enzyme2 (ACE2) cell surface protein to invade the host cell. The SARS-CoV-2 S-protein has acquired mutations that increase its affinity to human ACE2 by ~10-15-fold compared to SARS-CoV S-protein, making it highly infectious. In this study, we assessed if ACE2 polymorphisms might alter host susceptibility to SARS-CoV-2 by affecting the ACE2 S-protein interaction. Our comprehensive analysis of several large genomic datasets that included over 290,000 samples representing >400 population groups identified multiple ACE2 protein-altering variants, some of which mapped to the S-protein-interacting ACE2 surface. Using recently reported structural data and a recent S-protein-interacting synthetic mutant map of ACE2, we have identified natural ACE2 variants that are predicted to alter the virus-host interaction and thereby potentially alter host susceptibility. In particular, human ACE2 variants S19P, I21V, E23K, K26R, T27A, N64K, T92I, Q102P and H378R are predicted to increase susceptibility. The T92I variant, part of a consensus NxS/T N-glycosylation motif, confirmed the role of N90 glycosylation in immunity from non-human CoVs. Other ACE2 variants K31R, N33I, H34R, E35K, E37K, D38V, Y50F, N51S, M62V, K68E, F72V, Y83H, G326E, G352V, D355N, Q388L and D509Y are putative protective variants predicted to show decreased binding to SARS-CoV-2 S-protein. Overall, ACE2 variants are rare, consistent with the lack of selection pressure given the recent history of SARS-CoV epidemics, however, are likely to play an important role in altering susceptibility to CoVs.

Introduction

Coronaviruses (CoVs) are widely distributed in nature and pose a serious threat to humans and a range of mammalian hosts causing respiratory, gastrointestinal, and central nervous system diseases (Li, 2016). CoVs are enveloped non-segmented positive-sense single stranded RNA viruses and are classified into α -, β -, γ -, and δ -CoVs (Li, 2016). While α - and β -CoVs infect mammals, the γ - and δ -CoVs generally infect birds (Li, 2016). Previously, α -CoVs HCoV-229E and HCoV-NL63, and β -CoVs HCoV-HKU1 and HCoV-OC43 have been found to infect humans leading to mild symptoms (Graham and Baric, 2010; Li, 2016). However, three β -CoVs, severe acute respiratory syndrome coronavirus (SARS-CoV) in 2003 (Holmes, 2003; Li, 2016), Middle-East respiratory syndrome coronavirus in 2012 (MERS-CoV) (Li, 2016; Zaki et al., 2012), and more recently SARS-CoV-2 in 2019 (Chan et al., 2020; Huang et al., 2020; Zhu et al., 2020), have crossed species barrier to infect humans resulting in respiratory illnesses including pneumonia that is fatal.

SARS-CoV-2 is a novel coronavirus (2019-nCoV) first reported in December 2019 and is the cause of an ongoing global pandemic (Chan et al., 2020; Huang et al., 2020; Zhu et al., 2020). It has infected over 1.2 million people in 181 countries leading to over 69,000 deaths as of April 5th, 2020 (JHU, 2020). SARS-CoV-2 genome sequence analysis revealed that it is closer to bat CoV RaTG13 (96.2% identical) than to SARS-CoV (79.5% identical) responsible for the 2003 epidemic, suggesting that this novel virus originated in bats independently before jumping to humans either directly or through an yet to be determined intermediary host (Guo et al., 2020).

As with SARS-CoV and a related alpha-coronavirus NL63 (HCoV-NL63), SARS-CoV-2 employs the human ACE2 cell surface protein as a receptor to gain entry into cells (Hoffmann et al., 2020; Letko et al., 2020; Lin et al., 2008; Ou et al., 2020; Wan et al., 2020; Zhou et al., 2020a). The virus surface spike glycoprotein (S-protein) constitutes a key determinant of viral host range and contains two domains, S1 and S2, which are separated by a protease cleavage site (Li, 2016). A successful host cell invasion by the virus involves direct binding of the virus S1 receptor binding domain (RBD) to the host ACE2 peptidase extracellular domain (PD), exposing the S1-S2 inter-domain protease site that upon cleavage by host proteases leads to S2-mediated virus-

host cell membrane fusion (Belouzard et al., 2009; Hoffmann et al., 2020; Li, 2016; Li et al., 2005a; Simmons et al., 2005).

The SARS-CoV-2 S-protein is 98% identical to the bat CoV RaTG13 S protein, with the exception of an insertion that is also absent in the SARS-CoV S-protein in the S1/S2 inter-domain protease cleavage site. This difference has been proposed to alter SARS-CoV-2 tropism and enhance its transmissibility (Walls et al., 2020).

Several structural studies involving the SARS-CoV-2 S-protein RBD and ACE2 PD have identified key residues involved in their interaction (Shang et al., 2020; Walls et al., 2020; Wrapp et al., 2020; Yan et al., 2020). The S-protein RBD was reported to bind ACE2 PD with ~10- to 20-fold higher affinity (~15 nM) when compared to the SARS-CoV S-protein RBD (Shang et al., 2020; Wrapp et al., 2020), potentially contributing to high rate of SARS-CoV-2 infection.

As the interactions between the ACE2 receptor and S-protein RBD interface are critical for the cellular entry of the virus, we wondered if there were natural ACE2 variations that decrease or increase its affinity to the S-protein RBD that may protect or render individuals more susceptible to the virus. Consistent with this possibility, a saturation mutagenesis screen of select ACE2 PD residues identified variants that showed enhanced or decreased binding to S-protein (Procko, 2020).

In this study, we have analyzed ACE2 protein-altering variants in a large cohort of human population groups and identified polymorphisms that either likely protect or render individuals more susceptible to the virus. Understanding these changes at the molecular level, combined with the genotype and epidemiological data will allow the elucidation of population risk profiles and also help advance therapeutics such as a rationally designed soluble ACE2 receptor for treatment of COVID-19.

Results

Human ACE2 population polymorphism

SARS-CoV-2 S-protein interacts with the ACE2 PD to enter the human host cells. Analysis of S-protein RBD domain of SARS-CoV-2, SARS-CoV and closely related bat CoV RaTG13 identified changes that have increased the affinity of CoV-2 S1 RBD for human ACE2, which likely contributes to its increased infectivity (Shang et al., 2020;

Wrapp et al., 2020). It is very likely that there are natural variations in ACE2 in human populations, though not under selection, that may increase or decrease its affinity to SARS-CoV-2 S-protein and thereby render individuals more resistant or susceptible to the virus. To investigate this, we assessed ACE2 protein-altering variations from a number of databases including the gnomAD (Karczewski et al., 2019), RotterdamStudy (Ikram et al., 2017), ALSPAC (Fraser et al., 2013) and Asian-specific databases which included GenomeAsia100k (GenomeAsia, 2019), HGDP (Bergstrom et al., 2020), TOMMO-3.5kjpvn2 (Tadaka et al., 2019), and IndiGen (<https://indigen.igib.in/>), and HGDP (Bergstrom et al., 2020) (Supplementary Table 1). We found a total of 298 unique protein altering variants across 256 codons distributed throughout the 805 amino acid long human ACE2 (**Figure 1a, 1b, Supplementary Figure 1, and Supplementary Table 1**). The most frequent variant, N720D (1.6% allele frequency; n=3054, gnomAD), was found in the C-terminal collectrin domain that is not involved in the SARS-CoV-2 S-protein interaction. Overall, we found human ACE2 receptor polymorphisms to be low with a weighted mean Fst (fixation index) value of 0.0168, and the ACE2 PD showed even more reduced variation (Wilcoxon p=0.0656, **Supplementary Figure 2a**, see Methods). Further, we found ACE2 to be highly intolerant of loss of function variants (pLI=0.9977, gnomAD; **Supplementary Figure 2b**, see Methods), though we observed 5 predicted LOF singleton alleles (**Supplementary Table 1**).

Structural studies involving SARS-CoV and, more recently, the SARS-CoV-2 S-protein and its complex with human ACE2 have identified three regions in an ~120 amino acid claw-like exposed outer surface of the human ACE2 (ACE2-claw) that contributes to its binding to the S-protein (Shang et al., 2020; Walls et al., 2020; Wrapp et al., 2020; Yan et al., 2020). The main residues at the interface include S19, Q24, T27, F28, D30, K31, H34, E35, E37, D38, Y41, Q42, L45, L79, M82, Y83, T324, Q325, G326, E329, N330, K353, G354, D355, R357, P389, and R393 (**Figure 1b**). Mutagenesis of four residues in the S-protein-binding interface of rat ACE2 was sufficient to convert rat ACE2 into a human SARS-CoV receptor, further indicating the importance of this region in determining the host range and specificity of CoVs (Li et al., 2005b). Considering these findings, we focused on variants within the human ACE2-claw S-protein RBD-binding interface and identified protein alterations in 44 codons that

resulted in 49 unique variants for a total of 968 allelic variants. This included K26R, the second most frequent human ACE2 protein-altering variant (0.4% allele frequency; allele count=797, gnomAD), S19P, T27A, K31R, N33I, H34R, E35K, E37K, D38V, N51S, N64K, K68E, F72V, T92I, Q102P, G326E, G352V, D355N, H378R, Q388L, and D509Y (**Supplementary Table 2**). These variants are likely to either increase or decrease the binding affinity of ACE2 to the S-protein and thereby alter the ability of the virus to infect the host cell.

A recent mutagenesis screen using a synthetic human ACE2 mutant library identified variants that either increased or decreased its binding to SARS-CoV-2 S-protein (Procko, 2020). Using a sequencing-based enrichment assay, the fold enrichment or depletion of the mutant sequences was measured in this study (Procko, 2020). Mapping the enrichment z-scores from this study (Procko, 2020) to the spectrum of natural ACE2 polymorphisms, we identified several rare ACE2 variants (**Figure 1c**) that likely alter their binding to the SARS-CoV-2 S-protein and thereby protect or render individuals more susceptible to the virus (Supplementary Table 2). The majority of the variants that were predicted to alter the interaction between ACE2 and the virus S-protein were clustered around the N-terminal region of ACE2 that interacts with the S-protein (**Figure 1b**).

Included among the ACE2 polymorphic variants that increase ACE2/S-protein interaction are S19P, I21V, E23K, K26R, K26E, T27A, N64K, T92I, Q102P, M383T and H378R (**Supplementary Table 2 and Supplementary Figure 3**). Among these, the T92I polymorphism stands out in particular because it is part of a NxT/S (where x is any amino acid except proline) consensus N-glycosylation motif (Gavel and von Heijne, 1990) where N90 is the site of N-glycan addition. The ACE2 NxT/S motif, while conserved in 96 out of 296 jawed vertebrate with ACE2 sequence available is absent or altered in several species, including the civet cat (*Paguma larvata*) and several bat species where residue N90 is mutated, a proline is present at position 91 or the T92 is altered to any amino acid except serine (**Figure 1d, Supplementary Figure 4 and Supplementary Table 3**) (Demogines et al., 2012; Gavel and von Heijne, 1990; Li et al., 2005b). These ACE2 variations are expected to abolish glycosylation at N90 (Gavel and von Heijne, 1990). Furthermore, a mutation that altered the NxT/S motif in human

ACE2 to a civet ACE2-like sequence (90-NLTV-93 to DAKI), expected to abolish the N-glycosylation, increased the SARS-CoV infectivity and S-protein binding (**Figure 1d**) (Li et al., 2005b). The T92I mutant we identified showed a strong enrichment in the sequencing-based screen for S-protein binders (Procko, 2020). Considering these observations, we conclude that the T92I mutation increases the ACE2/S-protein binding affinity rendering individuals harboring this mutation more susceptibility to the virus.

Variants that are predicted to reduce the virus S-protein interactions and thereby decrease S/ACE2 binding affinity include K31R, N33I, H34R, E35K, E37K, D38V, Y50F, N51S, K68E, F72V, Y83H, G326E, G352V, D355N and Q388L. Below we discuss the structural basis for the inhibitory effect on ACE2/S-protein binding for this selected set of mutations, as well as for the enhancing effect of the selected polymorphisms that were shown to increase ACE2/S-protein binding in vitro (Procko, 2020).

Structural evaluation of ACE2 polymorphism

To further understand the effect of the polymorphisms on receptor recognition by the SARS-CoV-2 RBD and to confirm our predictions, we structurally modeled the identified ACE2 variants using the recently published cryo-EM and crystal structures of ACE2/SARS-CoV-2 RBD complexes (**Supplementary Table 4**) (Shang et al., 2020; Walls et al., 2020; Wrapp et al., 2020; Yan et al., 2020). We combined our structural analysis predictions with data from the Procko study (Procko, 2020) and classified the polymorphisms into five categories. These included mutations that directly enhanced or disrupted the residues at the ACE2/S-protein binding interface or residues interacting with the N90-linked glycan. We term these mutations as “directly enhancing” or “directly disrupting”. We also identified variants that were “indirectly enhancing” or “indirectly disrupting”. These mutants were found to affect the residues that are at the ACE2/S-protein binding interface but not in direct contact with the CoV-2 RBD residues. Lastly, we found several mutants that are located distal to the ACE2/S-protein binding site and do not mediate direct or indirect contacts with CoV-2 RBD. We classified these variants as “not relevant” in the context of our analysis. The enhancing variants were enriched in the mutagenesis screen (Procko, 2020), while those ACE2 variants predicted to disrupt or weaken the ACE2/S-protein interactions were depleted (Procko, 2020).

Polymorphic variants mapped onto the ACE2 structure remarkably segregate into two distinct clusters at the ACE2/CoV-2 RBD interface (**Figure 2**). Enhancing variants cluster to the ACE2 surface most proximal to the receptor-binding ridge of CoV-2 RBD (**Figure 3a**) whereas the majority of the disrupting variants reside centrally on the two major ACE2 α -helices that substantially contribute to the buried surface area at the interface (**Figure 4a**). Interestingly, the loop conformation in the receptor-binding ridge differs significantly in SARS-CoV-2 from that of SARS-CoV owing to the presence of bulky residues (V483 and E484) in the loop (Shang et al., 2020). This feature allows the CoV-2 loop to extend further towards ACE2 establishing more extensive contacts with the receptor. Hence, natural ACE2 variants in this region could be exploited by the CoV-2 loop, increasing susceptibility to viral infection. In contrast, most interactions that CoV-2 makes with the core of the ACE2 interface are centered on two α -helices and are mostly not unique to CoV-2. They seem to encompass critical binding hotspots, discussed below, and thus centrally located polymorphic variants are more likely to reduce viral recognition.

By far the most frequent variant identified in our data, K26R (~0.4% allele frequency), is predicted to enhance ACE2 affinity for SARS-CoV-2. Structural analysis of this polymorphism shows that K26 establishes polar contacts with the first mannose moiety of the ACE2 N90-linked glycan and likely stabilizes the position of the glycan relative to ACE2 (**Figure 3b**). As discussed above, the N90-linked glycan emerges as an important determinant of CoV-2 infectivity and may diminish ACE2 affinity for the RBD possibly through steric hindrance imposed by branching of the sugar modifications (Demogines et al., 2012). We predict that K26R would abrogate stabilizing polar contacts with N90, impairing coordination of the glycan (**Figure 3b**) and lead to an increase in the affinity of the virus to the ACE2 receptor. At the same time, R26 is now primed to establish backbone and side chain interactions with ACE2 D30 which then is poised to build a salt-bridge with CoV-2 RBD K417 (**Figure 3b**). The net effect of R26 polymorphism would then be the stabilization of core α -helices that increases ACE2 binding affinity to CoV-2 RBD at the cost of glycan rigidity. As discussed above, another enhancing variant, T92I, is structurally predicted to lead to similar effects by directly eliminating the N90-linked glycan.

Whereas the K26R variant stabilizes core ACE2 α -helical interactions, other naturally occurring polymorphic variants appear to locally destabilize or alter the N-terminus of ACE2 helix 1 (α 1) conformation. The T27A mutant (**Figure 3c**) removes side chain-backbone and backbone-backbone interactions between T27 and E30 likely increasing the local dynamics of helix α 1. This would allow the N-terminus of α 1 to bend slightly and accommodate the unique CoV-2 RBD receptor binding-ridge loop that more intimately contacts ACE2 compared to its SARS-CoV counterpart. Another predicted effect of the T27A variant is increased hydrophobicity at the interface, which could contribute to an increase in binding affinity. Similar destabilizing patterns can be inferred for S19P and E23K (**Supplementary Table 4**). Thus, the local α 1 N-terminal helical flexibility along with N90 glycan destabilization may help accommodate the protrusive CoV-2 RBD receptor-binding ridge to form more extensive contacts with ACE2 and facilitate viral entry specifically for this virus.

ACE2 polymorphic variants predicted to confer protection against the virus (**Supplementary Tables 2 and 4**) present compelling mechanistic explanations for how they may offer protection against the virus. The vast majority of disruptive variants map to the core α -helical bundle of ACE2 and to residues known to form contacts with the RBD (**Figure 4a**). There are two key hotspots in the α -helical bundle of the ACE2 interface that are important for CoV-2 RBD binding: K31 and K353 (**Figure 4a, right panel**). To enable interaction with the virus, these charged residues need to be accommodated in a largely hydrophobic environment at the binding interface and hence their neutralization is critical to the binding of coronavirus RBDs to human ACE2 (Li, 2008; Shang et al., 2020; Wu et al., 2012). A recent elegant study (Shang et al., 2020) showed that SARS-CoV-2 S-protein is more effective in neutralization of the lysine hotspots than SARS-CoV due to the presence of Q493 and L455 that stabilize K31, and N501 that stabilizes K353, (**Figure 4A, right panel**). Interestingly, K31R is one of the human ACE2 polymorphisms that we identified (**Supplementary Table 2**). Introduction of an arginine not only maintains the positive charge at position 31 but is also predicted to break an interaction with Q493 in the RBD (**Supplementary Figure 5a**) and destabilize the charge-neutralizing interaction with the virus. Thus, individuals carrying K31R ACE2 variants are predicted to be less prone to SARS-CoV-2 infection. While we

did not identify any polymorphic variants at residue K353, we detected an ACE2 mutation that changes the identity of D38 residue (**Supplementary Table 2**), which forms an electrostatic interaction with K353 (**Supplementary Table 4**). This mutation (D38V) would compromise the neutralizing effect of the K353-D38 interaction at the interface and is predicted to significantly reduce binding affinity between the virus and the host receptor.

Another recurrent polymorphism in ACE2 maps to residue E35 and changes it to a lysine (**Supplementary Table 1**). E35 establishes a critical polar contact with SARS-CoV-2 S-protein residue Q493, which is predicted to be attenuated in the presence of the positively charged lysine (**Figure 4B**). Interestingly, E35 is not conserved between SARS-CoV and SARS-CoV-2 S-proteins (Shang et al., 2020) and hence we predict that it could offer selective protection from the SARS-CoV-2 infection in individuals carrying this variant. Other variants found in our analysis, including H34R (**Figure 4C**) and D38V, similarly result in a loss of interface polar contacts which are predicted to reduce ACE2 affinity for the viral RBD domain. Another interesting polymorphism at position 83 results in Y83H alteration. Residue Y83 underlies a hydrophobic pocket into which F486 from SARS-CoV-2 RBD is inserted (**Supplementary Figure 5b**). This is another unique interaction involving ACE2 and the SARS-CoV-2 RBD F486 that is absent in SARS-CoV RBD where, the equivalent residue is a leucine (Shang et al., 2020). The polymorphism that replaces Y83 with a polar histidine will compromise the hydrophobic character of this unique pocket in addition to removing a polar contact with N487 (**Supplementary Figure 5b**), potentially offering selective protection from the SARS CoV-2 infections.

Discussion

The host-virus evolutionary arms race over time leads to natural selection that alters both the host and the viral proteins allowing both to increase their fitness (Daugherty and Malik, 2012). In this context multiple studies have analyzed and identified the origin, evolution and successful adaption of the SARS coronaviruses as human pathogens (Andersen et al., 2020; Guo et al., 2020). Viral genome sequencing and analysis has identified bats as the most likely natural host of origin for both SARS-

CoV and the recent SARS-CoV-2 (Guo et al., 2020). In particular, several studies have focused on the viral S-protein RBD that interacts with its host ACE2 receptor and identified key changes between the bat CoVs and other suspected intermediary host CoVs found in the civet and pangolin (Andersen et al., 2020; Chen et al., 2020; Shang et al., 2020; Walls et al., 2020; Wrapp et al., 2020; Yan et al., 2020). These studies have identified S-protein changes that have rendered the human cells permissive to the SARS-CoV and SARS-CoV-2 infection (Chen et al., 2020; Shang et al., 2020; Walls et al., 2020; Wrapp et al., 2020; Yan et al., 2020).

Thus far, the role of variations in human ACE2 receptor in susceptibility to both SARS CoVs had not been comprehensively examined. A recent study analyzed a limited ACE2 population variation data set and concluded that these polymorphisms did not confer resistance to the virus (Cao et al., 2020a). In this study, we have examined human ACE2 variation data compiled from multiple data sets and identified polymorphisms that will either likely render individuals more susceptible to the SARS-CoV-2 or protect them from the virus. Using structural predictions based on published protein structures and data from an elegant mutagenesis screen that used deep sequencing to assess enrichment or depletion of S-protein binding ACE2 variants, we classified the variants identified in this study for the effects on susceptibility to SARS-CoV (Procko, 2020; Shang et al., 2020; Walls et al., 2020; Wrapp et al., 2020; Yan et al., 2020). In particular, human ACE2 variants K26R, S16P, T27A, K31R, H34R, E35K, E37K, D38V, N51S, N64K, K68E, F72V, T92I, Q102P, G326E, G352V, D355N, H378R, Q388L, and D509Y are predicted to increase the susceptibility of the individuals carrying these variations. It is interesting to note that the T92I ACE2 variant is part of the consensus NxS/T N-glycosylation motif and is predicted to abolish glycosylation of the conserved N90 residue. Our structural investigation suggests that this mutation will favor improved viral S-protein binding. A previous study showed that the ACE2 N90 renders human cells resistant to Civet CoV (Li et al., 2005b). Recently, N90 and T92 ACE2 mutations were enriched in a screen for CoV-2 S-protein binding (Procko, 2020). Taken together, these observations suggest that N90 and T92 are critical ACE2 residues that confer protection and are CoV host modifiers. We also found variants K31R, E35K, E37K, D38V, N33I, H34R, Q388L and Y83H in ACE2 that are predicted to

show decreased binding to SARS-CoV-2 S-protein and thus protect individuals corresponding to these genotypes.

Overall, we find the ACE2 population variants, that either increase or decrease susceptibility, to be rare, which is consistent with the overall low population ACE2 receptor polymorphisms (mean F_{st} 0.0167). Also, we did not observe significant differences in ACE2 variant allele frequency among population groups. The variant alleles also did not show discernable gender distribution differences, even though ACE2 is a X-linked gene. The SARS-CoV infections and its deadly effects in humans are more recent and thus the pathogenic and protective variants have not been subject to purifying selection and therefore the variants we observe are predictably rare.

The expression levels of ACE2 and its variants in appropriate host tissue may modulate the deleterious effect of the virus. To further understand the importance of the ACE2 variants in susceptibility, it will be important to correlate clinical outcomes with ACE2 genotypes at population scale. The extremes in COVID-19 clinical symptoms reported ranging from asymptomatic infected adult individuals to those that show acute respiratory syndrome leading to death (Cao et al., 2020b; Cascella et al., 2020; Yuen et al., 2020), suggest a role for additional factors, including the role of innate and adaptive immunity, besides ACE2 variants in modifying disease outcomes.

Currently, there are no approved therapeutics for treating or preventing COVID-19 caused by the SARS-CoV-2. Therefore, development of therapeutics to treat patients and mitigate the COVID-19 pandemic is urgently needed (Cascella et al., 2020; Jiang, 2020). Several small molecules and neutralizing antibodies for treatment are in development (Li and De Clercq, 2020; Zhou et al., 2020b). Soluble ACE2 and ACE2-Fc fusion protein have been proposed as decoy SARS-CoV-2 receptor therapeutic (Hofmann et al., 2004; Kruse, 2020; Lei et al., 2020). Soluble ACE2, as a therapy for pulmonary arterial hypertension, has been shown to be safe in early in-human clinical studies (Guignabert et al., 2018; Haschke et al., 2013). A rationally designed, catalytically inactive, human ACE2 that carries one or more of the natural variants predicted to show improved binding to SARS viral S-protein RBD could be safely developed as a soluble protein with or without an Fc domain for treatment of COVID-19.

Such a recombinant ACE2 protein can be engineered to create a pan-CoV neutralizing drug that is broad and can neutralize CoVs that may emerge during future epidemics.

Methods

Identification of ACE2 variation

We queried multiple genomic databases including gnomAD (Karczewski et al., 2019) (<https://gnomad.broadinstitute.org/>), DicoverEHR (Dewey et al., 2016), RotterdamStudy (Ikram et al., 2017), ALSPAC (Fraser et al., 2013) and Asian specific databases which included GenomeAsia100k (GenomeAsia, 2019), HGDP (Bergstrom et al., 2020), TOMMO-3.5kjpnv2 (Tadaka et al., 2019) IndiGen (<https://indigen.igib.in/>) and Other aggregated data for ACE2 protein altering variations in populations groups across the world. The ACE2 genotypes in this study were from over 290,000 samples representing over 400 population groups across the world.

Fst Analysis

To assess genetic variation in the coding region of ACE2, we calculated the fixation index (Fst) from 2,381 unrelated individuals across 26 populations in the 1000 Genomes Project Phase 3 and 57,783 female individuals across eight populations in gnomAD. For 1000 Genome data, we used the Weir and Cockerham (1984) method as implemented in vcftools (Version 0.1.17); the weighted Fst were calculated from 88 variants. For gnomAD (v2.1.1), because we only have access to the allele counts, we used the original formulation by Wright (1969) and reported the weighted mean Fst as described in Bhatia et al. (2013); 277 variants were used. Because Fst values vary based on variants used (Bhatia et al. 2013), we calculated the Fst in a set of randomly selected genes on the same chromosomes matched by the length decile to use for comparison. To assess if variants in the peptidase domain has lower genetic variation, we used the one-sided Wilcoxon rank-sum test to compare 15 variants in the peptidase domain against 50 variants outside. Variants with $F_{st} < 1e-4$ were removed as they were uninformative.

ACE2 ortholog sequence analysis

A total of 295 Human ACE2 orthologs were obtained from NCBI (**Supplementary Table 3** for accession numbers). A snake ACE2 ortholog protein was obtained from the

published Indian cobra genome (Suryamohan et al., 2020). Multiple sequence alignment of residues surrounding the ACE2 NxT/S motif was performed using MCoffee (www.tcoffee.org). Phylogenetic trees were constructed using the PhyML webserver (www.phylogeny.fr).

Structural Analysis

Each identified variant was mapped, modeled, and analyzed in Pymol using the recently deposited crystal structures 6VW1 and 6LZG of human ACE2 bound to either chimeric SARS CoV-2 RBD (6VW1) or complete SARS CoV-2 RBD (6LZG).

Acknowledgements

DD was supported by an F30 fellowship. Grant number: 1F30CA247147. AR and SCS were supported by SCELSE.

Figure Legend

Figure 1. ACE2 polymorphisms. **a.** Pie chart representing protein altering variations in ACE2 by allele count and source. **b.** ACE2 protein domain showing positions with polymorphisms that were predicted to lead to increased (red line) or decreased (blue line) S-protein binding. Recurrent polymorphisms ($n > 1$) that were predicted to not impact S-protein binding are shown in light grey. Residues within the ACE2 PD known to interact with viral S-protein are shown as black vertical lines within the peptidase domain in the ACE2 diagram. **c.** Log base 10 pseudo count adjusted (+1) observed ACE2 allele counts of mutants predicted to impact S-protein binding. Singletons are marked with a ^ and direct S-protein contact residues are underlined. **d.** Multiple sequence alignment of the S-protein interacting ACE2 sequence from indicated species. ACE2 NxT/S glycosylation motif disrupted in dog, rat, palm civet and several bat ACE2 is highlighted in red. ACE2 residues that mediate contact with NL63-CoV, SARS-CoV and SARS-CoV-2 are shown as blue, green and orange bars, respectively.

Figure 2 Identified polymorphisms in human ACE2 mapped to the structure of human ACE2 in complex with the SARS-CoV-2 RBD. Residues in ACE2 showing polymorphic variation in human population were mapped on to the structure of the ACE2/SARS-CoV-2 RBD (PDB: 6VW1) and colored according to their effect on the predicted affinity between human ACE2 and viral spike S-protein systematically as measured by a deep sequence mutagenesis screen (Procko, 2020). Polymorphisms that were predicted to enhance the binding between ACE2 and the S-protein are colored in magenta if they are located directly at the ACE2/SARS-CoV-2 interface or in light pink if they do not participate directly in the interface. Polymorphisms that are predicted to enhance the binding between ACE2 and the S-protein are colored in dark blue if they are located directly at the ACE2/SARS-CoV-2 interface or in light blue if they do not participate directly in the interface interactions.

Figure 3. Structural basis of enhanced interaction between human ACE2 polymorphic variants and SARS-CoV-2 S-protein. (a) Zoomed-in view of the indicated polymorphic variants that enhance the interaction between the human ACE2 and the viral S-protein. The mutation color codes used are same as in Fig 1. (b-c) Analysis of the effect of the most frequent polymorphism, K26R (b) and-T27A (c) polymorphisms. PDB structures shown are 6VW1 (a) and 6LZG (b,c).

Figure 4. Structural basis of decreased interaction between human ACE2 polymorphic variants and SARS-CoV-2 S-protein. (a) Zoomed-in view of the indicated polymorphic variants that destabilize the interaction between human ACE2 and the viral S-protein. The mutation color codes used are same as in Fig 1.. Two hotspot lysine residues on ACE2 (K31 and K353) are highlighted with dot representation in the panel on the right. (b) Analysis of the recurring polymorphism at E35 of ACE2. (c) Analysis of the effect of ACE2 H34 mutation to arginine. PDB structures shown are 6VW1 (a) and 6LZG (b,c).

Supplementary figures legend

Figure S1. A cartoon of ACE2 protein showing protein altering polymorphic variants observed across the entire protein. Allele counts for each polymorphism is shown inside or above each circle. Empty circles indicate singletons.

Figure S2. (a) Fst index of exonic variants of ACE2, calculated from 57,783 female individuals across eight populations in gnomAD. Canonical transcript of ACE2 (ENST00000427411) and two Pfam domains are shown along with the positions of known SARS-CoV-2 contact residues. Peptidase domain harbor variants with lower variation (Wilcox $p=0.0656$). **(b)** ACE2 is highly constrained ($pLI=0.9977$), with the observed-to-expected ratio of the number of pLoF variants of 0.0968, consistent with the constrained genes (highlighted in cyan).

Figure S3. Heatmap showing human ACE2 polymorphism that map to the ACE2-RBD interaction region and the corresponding enrichment/depletion scores from a recent study by Procko 2020 (doi.org/10.1101/2020.03.16.994236)

Figure S4. ACE2 sequence comparison (a). Phylogenetic tree of ACE2 sequences from selected species, **(b)** Multiple sequence alignment of representative primate ACE2 sequences and ACE2 sequences of putative natural and intermediate reservoirs of coronaviruses. Pink boxes highlight species where the canonical NxT/S motif is absent or altered.

Figure S5. Structural basis for the destabilizing effects of K31R and Y83H human polymorphisms on the interactions with SARS-CoV-2. **(a)** Analysis of the polymorphism at K31 of ACE2 shows that its mutation to an arginine breaks an energetically favorable electrostatic interaction with Q493 of SARS-CoV-2 RBD (PDB: 6LZG). **(b)** Analysis of the effect of ACE2 Y83 mutation to histidine shows loss of polar contact with N487 as well as reduced hydrophobic packing with the unique SARS-CoV-2 RBD F486 (PDB: 6LZG).

References

- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine* 17 March 2020.
- Belouzard, S., Chu, V. C., and Whittaker, G. R. (2009). Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc Natl Acad Sci U S A* 106, 5871-5876.
- Bergstrom, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., *et al.* (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367, doi: 10.1126/science.aay5012.
- Cao, Y., Li, L., Feng, Z., Wan, S., Huang, P., Sun, X., Wen, F., Huang, X., Ning, G., and Wang, W. (2020a). Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell Discov* 6, 11.
- Cao, Y., Liu, X., Xiong, L., and Cai, K. (2020b). Imaging and Clinical Features of Patients With 2019 Novel Coronavirus SARS-CoV-2: A systematic review and meta-analysis. *J Med Virol*.
- Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C., and Di Napoli, R. (2020). Features, Evaluation and Treatment Coronavirus (COVID-19). In *StatPearls*, (Treasure Island (FL)).
- Chan, J. F., Yuan, S., Kok, K. H., To, K. K., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C., Poon, R. W., *et al.* (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 395, 514-523.
- Chen, Y., Guo, Y., Pan, Y., and Zhao, Z. J. (2020). Structure analysis of the receptor binding of 2019-nCoV. *Biochem Biophys Res Commun*, doi: 10.1016/j.bbrc.2020.1002.1071.
- Daugherty, M. D., and Malik, H. S. (2012). Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet* 46, 677-700.
- Demogines, A., Farzan, M., and Sawyer, S. L. (2012). Evidence for ACE2-utilizing coronaviruses (CoVs) related to severe acute respiratory syndrome CoV in bats. *J Virol* 86, 6350-6353.
- Dewey, F. E., Murray, M. F., Overton, J. D., Habegger, L., Leader, J. B., Fetterolf, S. N., O'Dushlaine, C., Van Hout, C. V., Staples, J., Gonzaga-Jauregui, C., *et al.* (2016). Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* 354, doi: 10.1126/science.aaf6814.

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., *et al.* (2013). Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* 42, 97-110.

Gavel, Y., and von Heijne, G. (1990). Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng* 3, 433-442.

GenomeAsia, K. C. (2019). The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* 576, 106-111.

Graham, R. L., and Baric, R. S. (2010). Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol* 84, 3134-3146.

Guignabert, C., de Man, F., and Lombes, M. (2018). ACE2 as therapy for pulmonary arterial hypertension: the good outweighs the bad. *Eur Respir J* 51, doi: 10.1183/13993003.13900848-13992018.

Guo, Y. R., Cao, Q. D., Hong, Z. S., Tan, Y. Y., Chen, S. D., Jin, H. J., Tan, K. S., Wang, D. Y., and Yan, Y. (2020). The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. *Mil Med Res* 7, 11.

Haschke, M., Schuster, M., Poglitsch, M., Loibner, H., Salzberg, M., Bruggisser, M., Penninger, J., and Krahenbuhl, S. (2013). Pharmacokinetics and pharmacodynamics of recombinant human angiotensin-converting enzyme 2 in healthy human subjects. *Clin Pharmacokinet* 52, 783-792.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., *et al.* (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, doi.org/10.1016/j.cell.2020.1002.1052.

Hofmann, H., Geier, M., Marzi, A., Krumbiegel, M., Peipp, M., Fey, G. H., Gramberg, T., and Pohlmann, S. (2004). Susceptibility to SARS coronavirus S protein-driven infection correlates with expression of angiotensin converting enzyme 2 and infection can be blocked by soluble receptor. *Biochem Biophys Res Commun* 319, 1216-1221.

Holmes, K. V. (2003). SARS-associated coronavirus. *N Engl J Med* 348, 1948-1951.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., *et al.* (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497-506.

Ikram, M. A., Brusselle, G. G. O., Murad, S. D., van Duijn, C. M., Franco, O. H., Goedegebure, A., Klaver, C. C. W., Nijsten, T. E. C., Peeters, R. P., Stricker, B. H., *et al.* (2017). The Rotterdam Study: 2018 update on objectives, design and main results. *Eur J Epidemiol* 32, 807-850.

JHU (2020). <https://coronavirus.jhu.edu/>.

Jiang, S. (2020). Don't rush to deploy COVID-19 vaccines and drugs without sufficient safety guarantees. *Nature* 579, 321.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., *et al.* (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* doi.org/10.1101/531210.

Kruse, R. L. (2020). Therapeutic strategies in an outbreak scenario to treat the novel coronavirus originating in Wuhan, China. *F1000Res* 9, 72.

Lei, C., Fu, W., Qian, K., Li, T., Zhang, S., Ding, M., and Hu, S. (2020). Potent neutralization of 2019 novel coronavirus by recombinant ACE2-Ig. *bioRxiv* doi.org/10.1101/2020.02.01.929976.

Letko, M., Marzi, A., and Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* 5, 562-569.

Li, F. (2008). Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. *J Virol* 82, 6984-6991.

Li, F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* 3, 237-261.

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

Li, G., and De Clercq, E. (2020). Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat Rev Drug Discov* 19, 149-150.

Li, W., Zhang, C., Sui, J., Kuhn, J. H., Moore, M. J., Luo, S., Wong, S. K., Huang, I. C., Xu, K., Vasilieva, N., *et al.* (2005b). Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* 24, 1634-1643.

Lin, H. X., Feng, Y., Wong, G., Wang, L., Li, B., Zhao, X., Li, Y., Smaill, F., and Zhang, C. (2008). Identification of residues in the receptor-binding domain (RBD) of the spike protein of human coronavirus NL63 that are critical for the RBD-ACE2 receptor interaction. *J Gen Virol* 89, 1015-1024.

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., Guo, L., Guo, R., Chen, T., Hu, J., *et al.* (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* 11, 1620.

Procko, E. (2020). The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2. *bioRxiv* <https://doi.org/10.1101/2020.1103.1116.994236>.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, <https://www.nature.com/articles/s41586-41020-42179-y>.

Simmons, G., Gosalia, D. N., Rennekamp, A. J., Reeves, J. D., Diamond, S. L., and Bates, P. (2005). Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci U S A* 102, 11876-11881.

Suryamohan, K., Krishnankutty, S. P., Guillory, J., Jevit, M., Schroder, M. S., Wu, M., Kuriakose, B., Mathew, O. K., Perumal, R. C., Koludarov, I., *et al.* (2020). The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nat Genet* 52, 106-117.

Tadaka, S., Katsuoka, F., Ueki, M., Kojima, K., Makino, S., Saito, S., Otsuki, A., Gocho, C., Sakurai-Yageta, M., Danjoh, I., *et al.* (2019). 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. *Hum Genome Var* 6, 28.

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veerler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 10.1016/j.cell.2020.1002.1058.

Wan, Y., Shang, J., Graham, R., Baric, R. S., and Li, F. (2020). Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J Virol* 94.

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

Wu, K., Peng, G., Wilken, M., Geraghty, R. J., and Li, F. (2012). Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J Biol Chem* 287, 8904-8911.

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367, 1444-1448.

Yuen, K. S., Ye, Z. W., Fung, S. Y., Chan, C. P., and Jin, D. Y. (2020). SARS-CoV-2 and COVID-19: The most important research questions. *Cell Biosci* 10, 40.

Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., and Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367, 1814-1820.

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., *et al.* (2020a). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270-273.

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., and Cheng, F. (2020b). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov* 6, 14.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., *et al.* (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 382, 727-733.

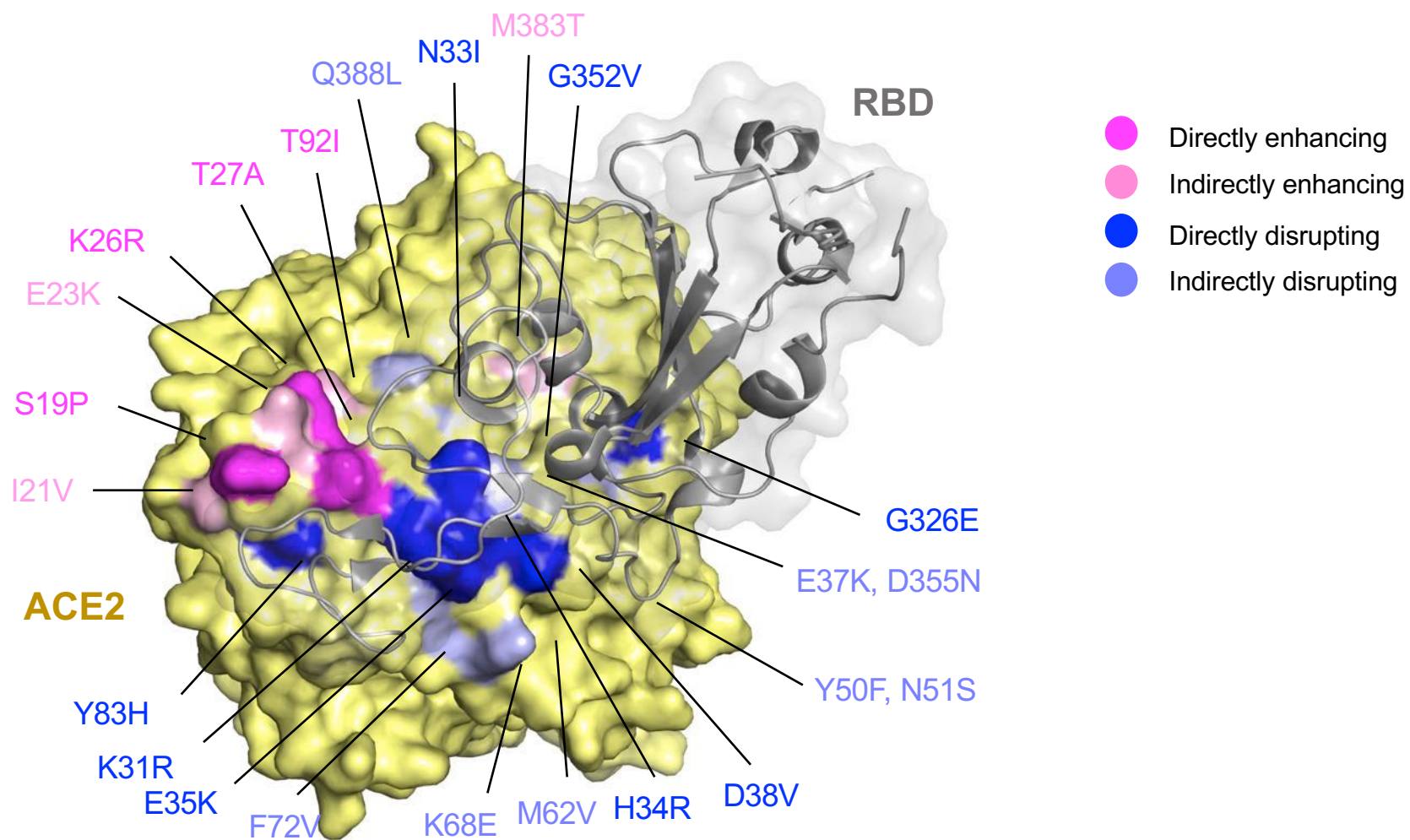


Figure 2

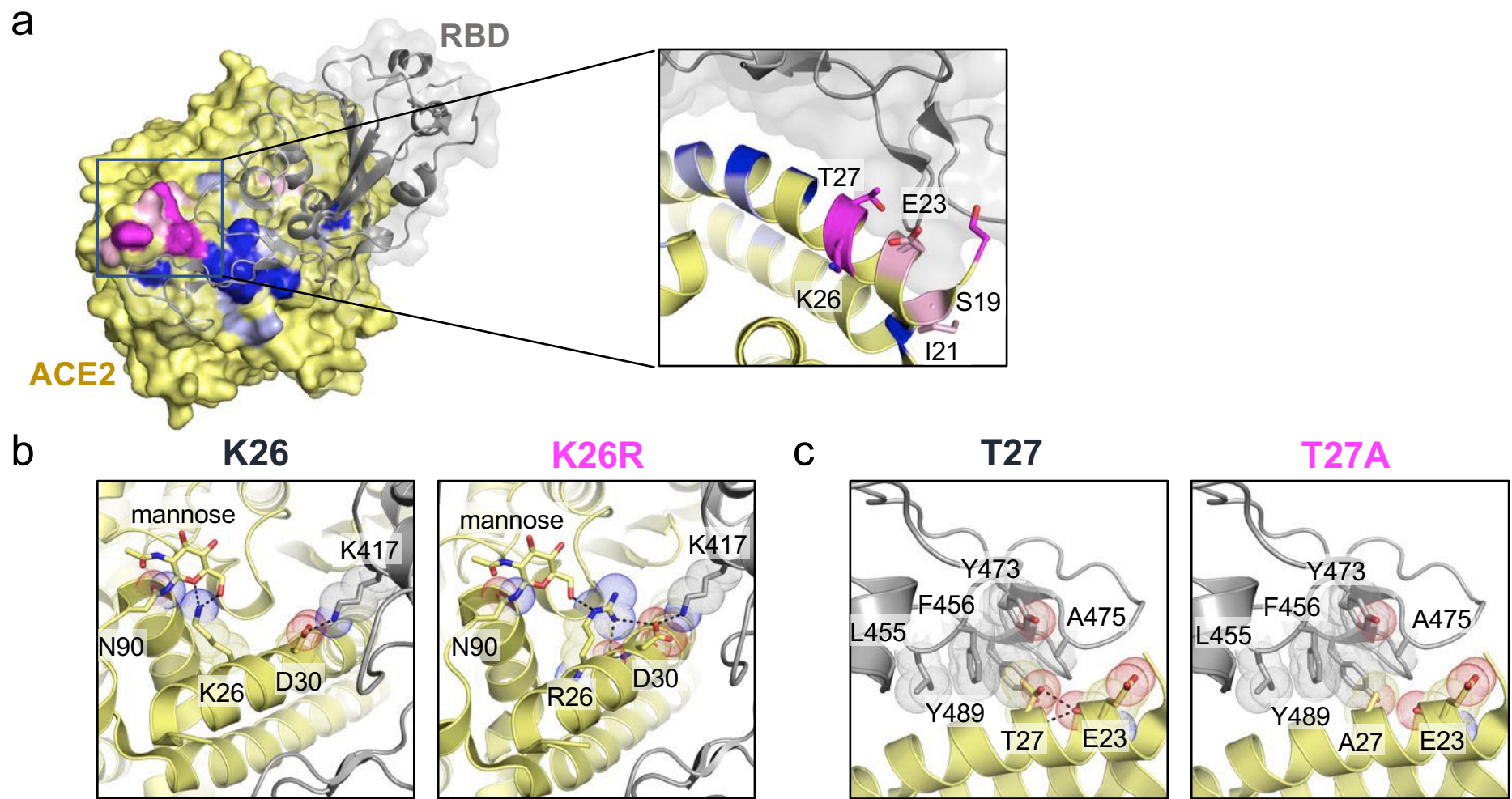


Figure 3

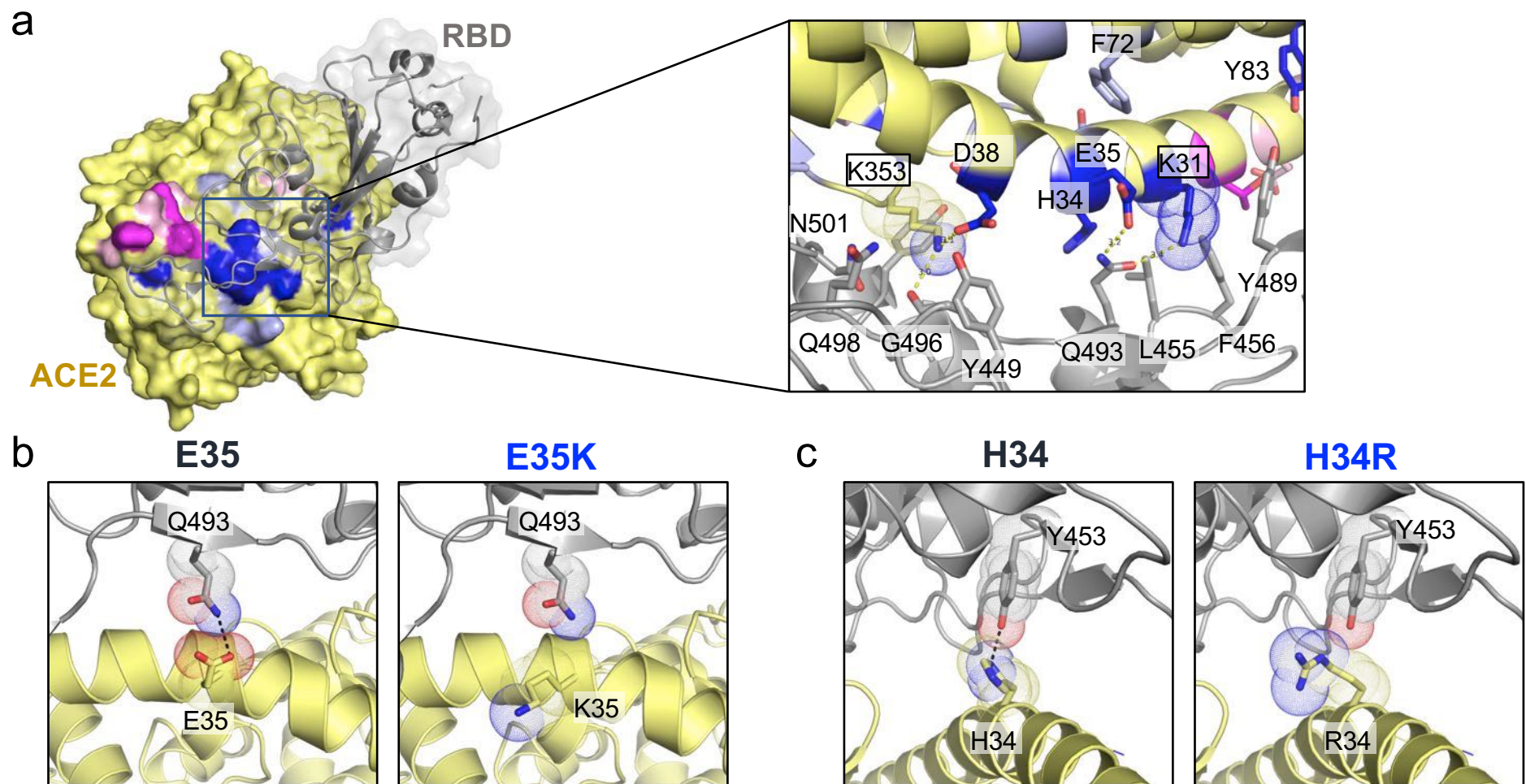
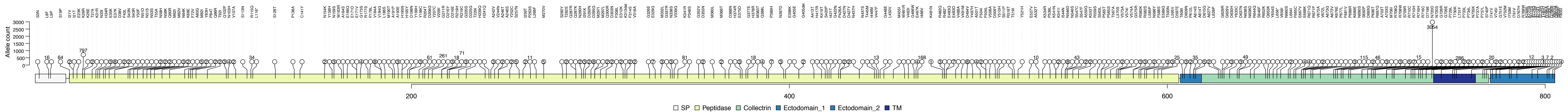
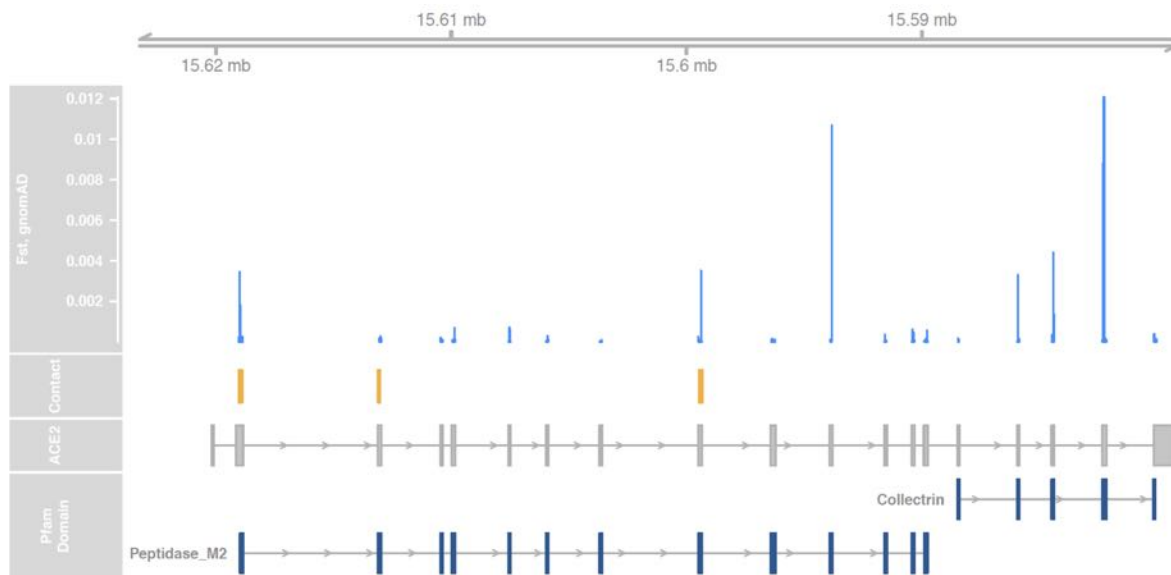
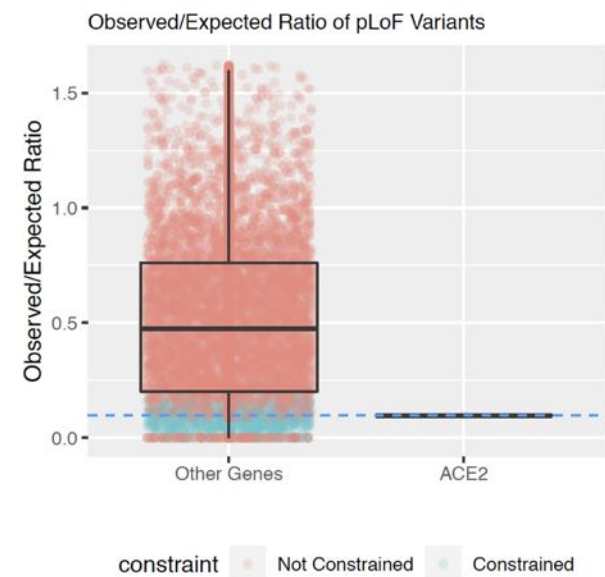
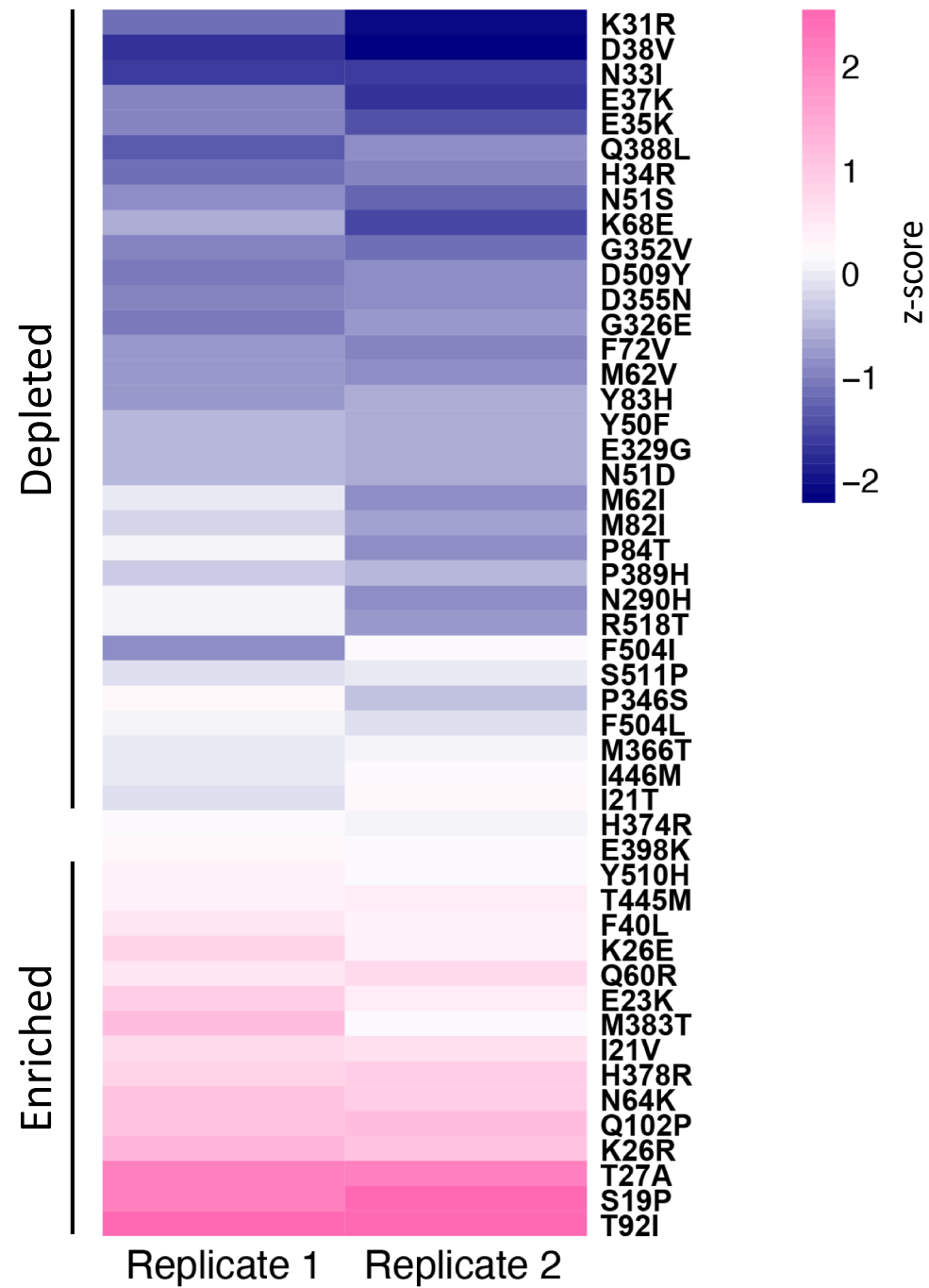


Figure 4



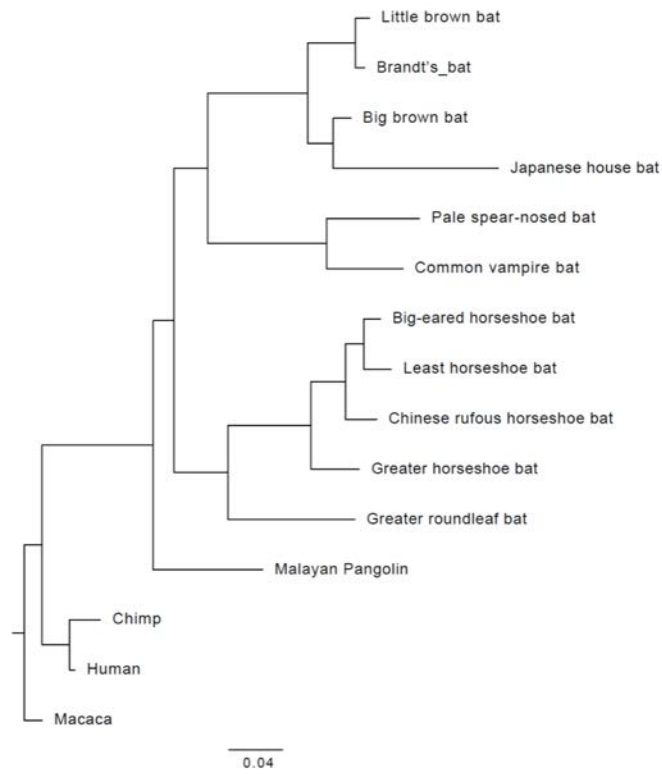
a**b**

Supplementary Figure 2

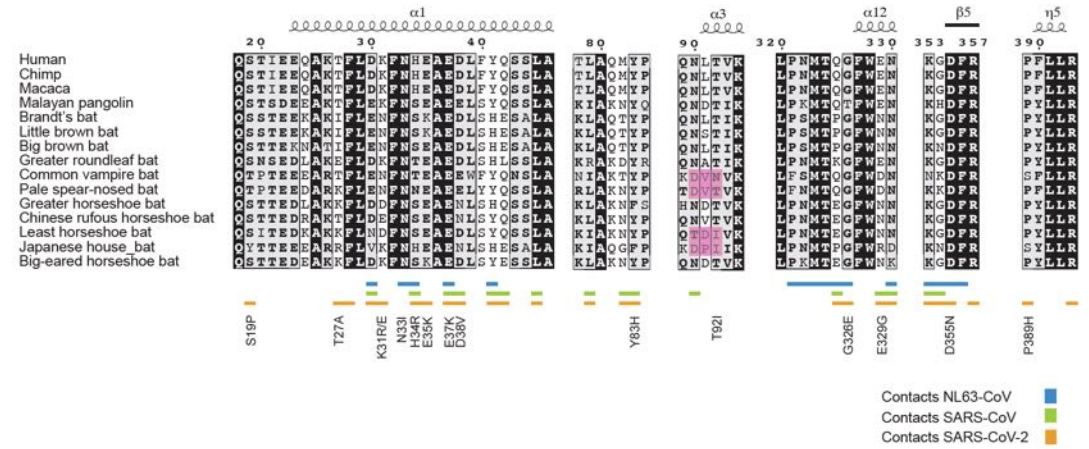


Supplementary Figure 3

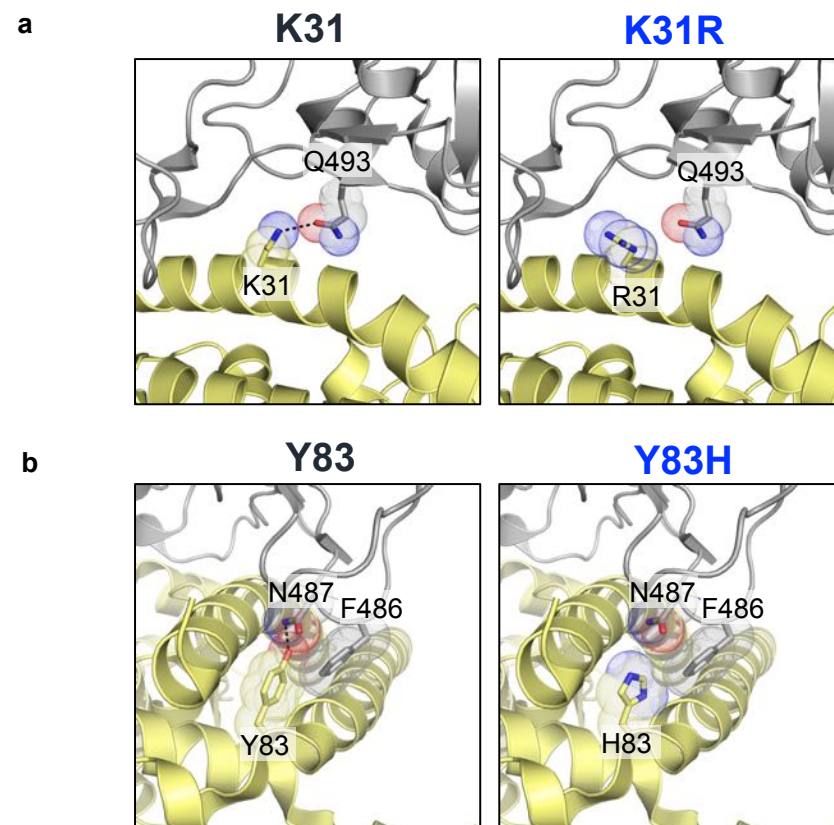
a



b



Supplementary Figure 4



Supplementary Figure 5

Supplementary Tables

Supplementary Table 1 - Human ACE2 protein altering population variants

Supplementary Table 2 - ACE2 polymorphisms predicted to alter SARS-CoV/Cov-2 S-protein binding

Supplementary Table 3 - Conservation of N90 glycosylation motif in annotated jawed vertebrate ACE2 orthologs

Supplementary Table 4 - Structure evaluation findings

	A	B	C	D	E	F	G	H	I
1	Supplementary Table 1 - Human ACE2 protein altering population variants								
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean Procko Score	Contact Residue
3	3	S3N	p.Ser3Asn	1	5.83625E-06	gnomAD	AsianSpecificDB		
4	8	L8F	p.Leu8Phe	16	8.05814E-05	gnomAD	Other		
5	9	L9P	p.Leu9Pro	1	5.62943E-06	gnomAD	Other		
6	19	S19P	p.Ser19Pro	64	0.000312919	gnomAD	Other	2.327818365	
7	21	I21V	p.Ile21Val	2	1.0927E-05	gnomAD	Other	0.697012397	
8	21	I21T	p.Ile21Thr	1	5.46254E-06	gnomAD		0.072108031	
9	23	E23K	p.Glu23Lys	1	5.45777E-06	gnomAD		0.647777027	
10	26	K26R	p.Lys26Arg	797	0.003883296	gnomAD	AsianSpecificDB,UKBio	1.229246696	
11	26	K26E	p.Lys26Glu	1	5.45476E-06	gnomAD		0.573492928	
12	27	T27A	p.Thr27Ala	2	1.09099E-05	gnomAD		2.113424949	Y
13	31	K31R	p.Lys31Arg	1	0	AsianSpecif	AsianSpecificDB	-1.609477184	Y
14	33	N33I	p.Asn33Ile	1	0	AsianSpecif	AsianSpecificDB	-1.568303739	
15	34	H34R	p.His34Arg	1	0	AsianSpecif	AsianSpecificDB	-1.076031319	Y
16	35	E35K	p.Glu35Lys	3	1.636E-05	gnomAD		-1.199253155	Y
17	37	E37K	p.Glu37Lys	8	3.89708E-05	gnomAD	AsianSpecificDB	-1.349986562	Y
18	38	D38V	p.Asp38Val	1	0	Other		-1.963734719	Y
19	40	F40L	p.Phe40Leu	2	3.0862E-05	Gnomad		0.449051719	
20	43	S43N	p.Ser43Asn	2	0	Other			
21	43	S43R	p.Ser43Arg	1	5.46185E-06	gnomAD			
22	50	Y50F	p.Tyr50Phe	1	5.4817E-06	gnomAD		-0.535360358	
23	51	N51D	p.Asn51Asp	1	5.48501E-06	gnomAD		-0.500865398	
24	51	N51S	p.Asn51Ser	1	5.48781E-06	gnomAD		-1.052778389	
25	53	N53S	p.Asn53Ser	1	0	AsianSpecif	AsianSpecificDB		
26	55	T55A	p.Thr55Ala	1	5.52312E-06	gnomAD			
27	58	N58H	p.Asn58His	2	1.11702E-05	gnomAD			
28	58	N58K	p.Asn58Lys	2	1.11867E-05	gnomAD			
29	60	Q60R	p.Gln60Arg	2	1.12455E-05	gnomAD		0.602848439	
30	62	M62I	p.Met62Ile	1	0	AsianSpecif	AsianSpecificDB	-0.433221503	

	A	B	C	D	E	F	G	H	I
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean Procko Score	Contact Residue
31	62	M62V	p.Met62Val	1	5.66646E-06	gnomAD	AsianSpecificDB	-0.782795497	
32	64	N64K	p.Asn64Lys	3	1.4664E-05	gnomAD		1.047209649	
33	68	K68E	p.Lys68Glu	2	1.09457E-05	gnomAD	AsianSpecificDB	-1.052667031	
34	72	F72V	p.Phe72Val	1	5.47411E-06	gnomAD		-0.88673533	
35	80	A80G	p.Ala80Gly	1	0	AsianSpecif	AsianSpecificDB		
36	82	M82I	p.Met82Ile	5	2.44178E-05	gnomAD		-0.428501783	Y
37	83	Y83H	p.Tyr83His	2	0	SAS(1)	AsianSpecificDB	-0.667134906	Y
38	84	P84T	p.Pro84Thr	1	5.4707E-06	gnomAD		-0.419131731	
39	86	Q86R	p.Gln86Arg	2	1.09411E-05	gnomAD	Other		
40	92	T92I	p.Thr92Ile	2	1.09557E-05	gnomAD		2.534286782	
41	102	Q102P	p.Gln102Pro	3	1.47451E-05	gnomAD		1.191328285	
42	103	N103H	p.Asn103His	4	1.96661E-05	gnomAD			
43	107	V107A	p.Val107Ala	2	1.10552E-05	gnomAD	Other		
44	113	S113N	p.Ser113Asn	1	0	Other			
45	115	R115Q	p.Arg115Gln	34	0.000170308	gnomAD	Other		
46	128	S128T	p.Ser128Thr	1	5.73809E-06	gnomAD			
47	138	P138A	p.Pro138Ala	1	5.73283E-06	gnomAD			
48	141	C141Y	p.Cys141Tyr	1	5.8307E-06	gnomAD			
49	154	N154K	p.Asn154Lys	2	1.09536E-05	gnomAD			
50	158	Y158H	p.Tyr158His	1	0	Other			
51	159	N159S	p.Asn159Ser	3	1.63682E-05	gnomAD	AsianSpecificDB,UKBio		
52	163	W163R	p.Trp163Arg	1	5.45557E-06	gnomAD			
53	164	A164S	p.Ala164Ser	1	5.45461E-06	gnomAD			
54	166	E166Q	p.Glu166Gln	1	5.4536E-06	gnomAD			
55	171	E171D	p.Glu171Asp	1	4.55519E-05	gnomAD			
56	171	E171V	p.Glu171Val	1	5.45411E-06	gnomAD			
57	173	G173S	p.Gly173Ser	4	2.18172E-05	gnomAD			
58	177	R177S	p.Arg177Ser	1	0	AsianSpecif	AsianSpecificDB		
59	178	P178L	p.Pro178Leu	1	5.455E-06	gnomAD			
60	184	V184A	p.Val184Ala	8	4.36503E-05	gnomAD	AsianSpecificDB,UKBio		
61	184	V184G	p.Val184Gly	1	4.58358E-05	gnomAD			

	A	B	C	D	E	F	G	H	I
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean Procko Score	Contact Residue
62	186	L186S	p.Leu186Ser	1	5.4615E-06	gnomAD			
63	190	M190T	p.Met190Thr	1	5.46591E-06	gnomAD			
64	191	A191P	p.Ala191Pro	1	5.46819E-06	gnomAD			
65	193	A193E	p.Ala193Glu	3	1.64126E-05	gnomAD			
66	195	H195N	p.His195Asn	1	5.47627E-06	gnomAD	AsianSpecificDB		
67	195	H195Y	p.His195Tyr	1	5.47627E-06	gnomAD			
68	198	D198N	p.Asp198Asn	1	5.88166E-06	gnomAD			
69	199	Y199H	p.Tyr199His	1	0	AsianSpecif	AsianSpecificDB		
70	199	Y199C	p.Tyr199Cys	3	1.68777E-05	gnomAD	AsianSpecificDB		
71	204	R204T	p.Arg204Thr	1	5.48152E-06	gnomAD	AsianSpecificDB		
72	206	D206G	p.Asp206Gly	61	0.000299999	gnomAD	Other		
73	207	Y207C	p.Tyr207Cys	1	5.46866E-06	gnomAD			
74	209	V209I	p.Val209Ile	1	5.46735E-06	gnomAD			
75	211	G211R	p.Gly211Arg	261	0.001279889	gnomAD	AsianSpecificDB,UKBio		
76	216	D216Y	p.Asp216Tyr	1	5.46015E-06	gnomAD	Other		
77	216	D216E	p.Asp216Glu	3	1.47049E-05	gnomAD			
78	219	R219H	p.Arg219His	18	9.83128E-05	gnomAD	AsianSpecificDB,UKBio		
79	219	R219C	p.Arg219Cys	71	0.000348058	gnomAD	Other		
80	220	G220S	p.Gly220Ser	3	1.63894E-05	gnomAD	AsianSpecificDB		
81	225	D225G	p.Asp225Gly	1	0	AsianSpecif	AsianSpecificDB		
82	229	T229I	p.Thr229Ile	1	5.4787E-06	gnomAD			
83	239	H239Q	p.His239Gln	1	5.85967E-06	gnomAD			
84	241	H241Q	p.His241Gln	1	0	Other			
85	242	A242V	p.Ala242Val	3	1.7254E-05	gnomAD			
86	246	A246T	p.Ala246Thr	2	0	AsianSpecif	AsianSpecificDB		
87	246	A246S	p.Ala246Ser	1	5.68311E-06	gnomAD			
88	252	Y252C	p.Tyr252Cys	2	1.12827E-05	gnomAD	AsianSpecificDB		
89	257	S257N	p.Ser257Asn	5	2.57054E-05	gnomAD	AsianSpecificDB		
90	259	I259T	p.Ile259Thr	2	1.03069E-05	gnomAD			
91	263	P263S	p.Pro263Ser	11	5.78415E-05	gnomAD	Other		
92	266	L266F	p.Leu266Phe	1	0	Other			

	A	B	C	D	E	F	G	H	I
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean Procko Score	Contact Residue
93	270	M270V	p.Met270Val	5	2.98299E-05	gnomAD	AsianSpecificDB,UKBio		
94	280	S280Y	p.Ser280Tyr	1	5.65141E-06	gnomAD			
95	282	T282S	p.Thr282Ser	2	0	AsianSpecif	AsianSpecificDB		
96	287	Q287R	p.Gln287Arg	1	5.66046E-06	gnomAD			
97	287	Q287K	p.Gln287Lys	1	5.65384E-06	gnomAD			
98	290	N290H	p.Asn290His	2	1.13323E-05	gnomAD		-0.382086808	
99	291	I291K	p.Ile291Lys	3	1.70594E-05	gnomAD	Other		
100	292	D292N	p.Asp292Asn	2	1.13967E-05	gnomAD			
101	295	D295G	p.Asp295Gly	8	4.6594E-05	gnomAD	Other		
102	297	M297I	p.Met297Ile	1	5.86304E-06	gnomAD			
103	297	M297L	p.Met297Leu	1	5.85144E-06	gnomAD			
104	300	Q300R	p.Gln300Arg	2	1.18694E-05	gnomAD			
105	303	D303N	p.Asp303Asn	2	1.20266E-05	gnomAD	AsianSpecificDB		
106	308	F308L	p.Phe308Leu	1	5.73365E-06	gnomAD			
107	312	E312K	p.Glu312Lys	2	1.13409E-05	gnomAD			
108	314	F314S	p.Phe314Ser	2	0	Other			
109	318	V318A	p.Val318Ala	2	0	AsianSpecif	AsianSpecificDB		
110	326	G326E	p.Gly326Glu	1	5.51675E-06	gnomAD		-0.90546886	Y
111	329	E329G	p.Glu329Gly	7	3.44298E-05	gnomAD	Other	-0.506466064	Y
112	332	M332L	p.Met332Leu	3	1.65432E-05	gnomAD			
113	337	G337R	p.Gly337Arg	1	5.53168E-06	gnomAD			
114	338	N338S	p.Asn338Ser	3	1.65937E-05	gnomAD			
115	339	V339G	p.Val339Gly	1	5.53661E-06	gnomAD			
116	341	K341R	p.Lys341Arg	81	0.000400192	gnomAD	AsianSpecificDB,UKBio		
117	346	P346S	p.Pro346Ser	1	5.59945E-06	gnomAD		-0.066906284	
118	352	G352V	p.Gly352Val	1	5.75218E-06	gnomAD		-1.025319393	
119	355	D355N	p.Asp355Asn	2	1.17433E-05	gnomAD	Other	-0.911052386	
120	360	M360L	p.Met360Leu	1	5.53836E-06	gnomAD			
121	366	M366T	p.Met366Thr	2	1.09845E-05	gnomAD		0.034196848	
122	368	D368N	p.Asp368Asn	1	5.49125E-06	gnomAD	AsianSpecificDB		
123	374	H374R	p.His374Arg	1	5.47528E-06	gnomAD		0.110711661	

	A	B	C	D	E	F	G	H	I
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean Procko Score	Contact Residue
124	375	E375D	p.Glu375Asp	3	1.64213E-05	gnomAD			
125	377	G377E	p.Gly377Glu	1	5.47345E-06	gnomAD			
126	378	H378R	p.His378Arg	18	8.79104E-05	gnomAD	Other	0.83154494	
127	383	M383T	p.Met383Thr	1	0	AsianSpecif	AsianSpecificDB	0.676250072	
128	388	Q388L	p.Gln388Leu	4	2.18607E-05	gnomAD		-1.084467232	
129	389	P389H	p.Pro389His	7	3.82576E-05	gnomAD		-0.400730258	
130	397	N397D	p.Asn397Asp	3	1.46412E-05	gnomAD			
131	398	E398K	p.Glu398Lys	1	5.46224E-06	gnomAD		0.205797456	
132	405	G405E	p.Gly405Glu	1	5.46072E-06	gnomAD			
133	413	A413T	p.Ala413Thr	1	0	Other			
134	417	H417R	p.His417Arg	1	0	AsianSpecif	AsianSpecificDB		
135	419	K419T	p.Lys419Thr	1	5.46263E-06	gnomAD			
136	420	S420P	p.Ser420Pro	1	4.54876E-05	gnomAD			
137	421	I421T	p.Ile421Thr	1	5.46379E-06	gnomAD			
138	426	P426A	p.Pro426Ala	1	5.46866E-06	gnomAD			
139	427	D427N	p.Asp427Asn	2	0	Other			
140	427	D427Y	p.Asp427Tyr	2	1.0948E-05	gnomAD			
141	437	N437H	p.Asn437His	1	0	AsianSpecif	AsianSpecificDB		
142	437	N437S	p.Asn437Ser	1	0	AsianSpecif	AsianSpecificDB		
143	445	T445M	p.Thr445Met	1	5.86937E-06	gnomAD		0.422616407	
144	446	I446M	p.Ile446Met	1	4.53968E-05	gnomAD		0.059171832	
145	447	V447F	p.Val447Phe	13	6.68439E-05	gnomAD			
146	448	G448E	p.Gly448Glu	1	5.77481E-06	gnomAD			
147	450	L450V	p.Leu450Val	1	5.71984E-06	gnomAD			
148	455	M455I	p.Met455Ile	1	5.65163E-06	gnomAD			
149	461	W461R	p.Trp461Arg	1	5.59851E-06	gnomAD			
150	463	V463I	p.Val463Ile	1	5.57479E-06	gnomAD			
151	466	G466W	p.Gly466Trp	1	5.57336E-06	gnomAD			
152	467	E467K	p.Glu467Lys	4	2.24252E-05	gnomAD			
153	468	I468V	p.Ile468Val	168	0.000838441	gnomAD	AsianSpecificDB,UKBio		
154	481	K481N	p.Lys481Asn	1	0	Other			

	A	B	C	D	E	F	G	H	I
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean Procko Score	Contact Residue
155	482	R482Q	p.Arg482Gln	3	1.7609E-05	gnomAD	Other		
156	483	E483Q	p.Glu483Gln	1	0	AsianSpecific	AsianSpecificDB		
157	483	E483D	p.Glu483Asp	7	4.63908E-05	gnomAD			
158	488	V488A	p.Val488Ala	1	6.32495E-06	gnomAD			
159	491	V491M	p.Val491Met	1	6.28828E-06	gnomAD			
160	494	D494V	p.Asp494Val	8	4.9578E-05	gnomAD	Other		
161	497	Y497H	p.Tyr497His	1	0	AsianSpecific	AsianSpecificDB		
162	501	A501T	p.Ala501Thr	4	2.21659E-05	gnomAD	AsianSpecificDB,UKBio		
163	504	F504I	p.Phe504Ile	2	1.26824E-05	gnomAD		-0.33180346	
164	504	F504L	p.Phe504Leu	1	4.6307E-05	gnomAD		-0.035560813	
165	506	V506A	p.Val506Ala	1	6.56052E-06	gnomAD			
166	509	D509Y	p.Asp509Tyr	1	0	Other		-0.942921896	
167	510	Y510H	p.Tyr510His	1	6.86304E-06	gnomAD		0.246017376	
168	511	S511P	p.Ser511Pro	1	0	AsianSpecific	AsianSpecificDB	-0.096523335	
169	518	R518T	p.Arg518Thr	1	5.4904E-06	gnomAD		-0.34082089	
170	519	T519I	p.Thr519Ile	1	5.48276E-06	gnomAD			
171	521	Y521H	p.Tyr521His	1	5.47312E-06	gnomAD			
172	527	E527V	p.Glu527Val	1	0	AsianSpecific	AsianSpecificDB		
173	532	A532T	p.Ala532Thr	10	5.46305E-05	gnomAD	AsianSpecificDB		
174	534	K534R	p.Lys534Arg	1	5.46138E-06	gnomAD			
175	538	P538L	p.Pro538Leu	1	5.46379E-06	gnomAD			
176	541	K541N	p.Lys541Asn	1	0	Other			
177	541	K541I	p.Lys541Ile	2	9.75895E-06	gnomAD			
178	544	I544N	p.Ile544Asn	1	5.47714E-06	gnomAD			
179	546	N546D	p.Asn546Asp	8	3.91654E-05	gnomAD			
180	546	N546S	p.Asn546Ser	1	5.48585E-06	gnomAD			
181	547	S547C	p.Ser547Cys	43	0.00021059	gnomAD			
182	547	S547F	p.Ser547Phe	1	5.49058E-06	gnomAD			
183	550	A550G	p.Ala550Gly	2	0	Other			
184	553	K553T	p.Lys553Thr	2	9.84518E-06	gnomAD			
185	559	R559S	p.Arg559Ser	1	5.6744E-06	gnomAD			

	A	B	C	D	E	F	G	H	I
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean Procko Score	Contact Residue
186	563	S563L	p.Ser563Leu	1	5.6211E-06	gnomAD			
187	565	P565T	p.Pro565Thr	1	0	AsianSpecif	AsianSpecificDB		
188	565	P565S	p.Pro565Ser	1	5.57588E-06	gnomAD			
189	567	T567A	p.Thr567Ala	1	5.55222E-06	gnomAD			
190	570	L570S	p.Leu570Ser	1	4.59876E-05	gnomAD	Other		
191	573	V573A	p.Val573Ala	2	1.09806E-05	gnomAD	Other		
192	574	V574I	p.Val574Ile	1	5.49082E-06	gnomAD	AsianSpecificDB		
193	574	V574A	p.Val574Ala	1	0	AsianSpecif	AsianSpecificDB		
194	575	G575R	p.Gly575Arg	1	5.48799E-06	gnomAD			
195	582	R582K	p.Arg582Lys	3	1.64348E-05	gnomAD	Other		
196	582	R582S	p.Arg582Ser	3	1.46802E-05	gnomAD			
197	585	L585P	p.Leu585Pro	1	5.47948E-06	gnomAD			
198	586	N586Y	p.Asn586Tyr	2	9.78852E-06	gnomAD			
199	588	F588S	p.Phe588Ser	1	5.48252E-06	gnomAD			
200	589	E589G	p.Glu589Gly	1	5.48477E-06	gnomAD			
201	593	T593N	p.Thr593Asn	3	1.47442E-05	gnomAD			
202	595	L595V	p.Leu595Val	3	1.65411E-05	gnomAD			
203	597	D597E	p.Asp597Glu	25	0.000123366	gnomAD	AsianSpecificDB,UKBio		
204	608	T608I	p.Thr608Ile	1	5.67617E-06	gnomAD			
205	609	D609N	p.Asp609Asn	3	1.71863E-05	gnomAD	Other		
206	612	P612L	p.Pro612Leu	2	0	AsianSpecif	AsianSpecificDB		
207	614	A614S	p.Ala614Ser	35	0.00017115	gnomAD	Other		
208	614	A614T	p.Ala614Thr	1	0	Other			
209	615	D615G	p.Asp615Gly	5	2.73928E-05	gnomAD			
210	627	A627V	p.Ala627Val	2	1.0948E-05	gnomAD			
211	628	L628P	p.Leu628Pro	1	0	Other			
212	629	G629R	p.Gly629Arg	1	0	AsianSpecif	AsianSpecificDB		
213	629	G629V	p.Gly629Val	1	5.47411E-06	gnomAD			
214	630	D630H	p.Asp630His	3	1.64245E-05	gnomAD			
215	633	Y633C	p.Tyr633Cys	1	5.93384E-06	gnomAD			
216	637	D637N	p.Asp637Asn	3	0	Other			

	A	B	C	D	E	F	G	H	I
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean Procko Score	Contact Residue
217	638	N638S	p.Asn638Ser	49	0.000253484	gnomAD	AsianSpecificDB,UKBio		
218	638	N638D	p.Asn638Asp	1	0	AsianSpecif	AsianSpecificDB		
219	644	R644Q	p.Arg644Gln	1	0	AsianSpecif	AsianSpecificDB		
220	646	S646F	p.Ser646Phe	1	0	AsianSpecif	AsianSpecificDB		
221	652	R652K	p.Arg652Lys	1	5.75844E-06	gnomAD			
222	653	Q653K	p.Gln653Lys	1	0	Other			
223	654	Y654S	p.Tyr654Ser	1	4.55208E-05	gnomAD			
224	658	V658I	p.Val658Ile	1	5.95575E-06	gnomAD			
225	660	N660S	p.Asn660Ser	1	5.99797E-06	gnomAD			
226	664	L664I	p.Leu664Ile	2	0	Other			
227	665	F665C	p.Phe665Cys	1	0	Other			
228	667	E667V	p.Glu667Val	1	5.52077E-06	gnomAD	Other		
229	668	E668K	p.Glu668Lys	4	2.19972E-05	gnomAD	AsianSpecificDB		
230	671	R671Q	p.Arg671Gln	4	1.96187E-05	gnomAD			
231	671	R671P	p.Arg671Pro	1	4.61553E-05	gnomAD			
232	672	V672A	p.Val672Ala	1	5.48195E-06	gnomAD			
233	672	V672L	p.Val672Leu	2	1.09705E-05	gnomAD			
234	673	A673G	p.Ala673Gly	1	5.47921E-06	gnomAD			
235	673	A673V	p.Ala673Val	1	5.47921E-06	gnomAD			
236	676	K676E	p.Lys676Glu	1	0	AsianSpecif	AsianSpecificDB		
237	677	P677L	p.Pro677Leu	1	5.47525E-06	gnomAD			
238	681	F681V	p.Phe681Val	1	5.47202E-06	gnomAD			
239	688	P688R	p.Pro688Arg	1	5.47282E-06	gnomAD			
240	689	K689E	p.Lys689Glu	3	1.64049E-05	gnomAD			
241	690	N690S	p.Asn690Ser	1	5.47211E-06	gnomAD			
242	692	S692P	p.Ser692Pro	115	0.000561883	gnomAD	AsianSpecificDB,UKBio		
243	693	D693G	p.Asp693Gly	1	5.47217E-06	gnomAD			
244	696	P696T	p.Pro696Thr	2	1.09609E-05	gnomAD	Other		
245	697	R697G	p.Arg697Gly	46	0.000252134	gnomAD	AsianSpecificDB,UKBio		
246	703	A703T	p.Ala703Thr	1	0	AsianSpecif	AsianSpecificDB		
247	703	A703S	p.Ala703Ser	3	1.65621E-05	gnomAD			

	A	B	C	D	E	F	G	H	I
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean Procko Score	Contact Residue
248	706	M706I	p.Met706Ile	1	7.08572E-06	gnomAD			
249	708	R708Q	p.Arg708Gln	1	6.91247E-06	gnomAD	Other		
250	708	R708W	p.Arg708Trp	3	1.80397E-05	gnomAD			
251	709	S709R	p.Ser709Arg	1	6.88108E-06	gnomAD			
252	710	R710C	p.Arg710Cys	5	2.89195E-05	gnomAD	Other		
253	710	R710H	p.Arg710His	7	3.99657E-05	gnomAD			
254	716	R716H	p.Arg716His	15	8.18027E-05	gnomAD	Other		
255	716	R716C	p.Arg716Cys	1	6.19149E-06	gnomAD			
256	719	D719E	p.Asp719Glu	1	6.01525E-06	gnomAD			
257	720	N720D	p.Asn720Asp	3054	0.016215011	gnomAD	AsianSpecificDB,UKBio		
258	720	N720S	p.Asn720Ser	1	6.00402E-06	gnomAD			
259	726	G726R	p.Gly726Arg	2	1.15294E-05	gnomAD			
260	726	G726E	p.Gly726Glu	1	5.76558E-06	gnomAD			
261	729	P729L	p.Pro729Leu	2	1.00837E-05	gnomAD			
262	730	T730K	p.Thr730Lys	1	5.63146E-06	gnomAD			
263	731	L731F	p.Leu731Phe	286	0.001434915	gnomAD	AsianSpecificDB,UKBio		
264	733	P733L	p.Pro733Leu	2	0	AsianSpecif	AsianSpecificDB		
265	734	P734L	p.Pro734Leu	2	1.12724E-05	gnomAD	AsianSpecificDB		
266	735	N735K	p.Asn735Lys	1	5.64127E-06	gnomAD			
267	737	P737A	p.Pro737Ala	1	5.64723E-06	gnomAD	AsianSpecificDB		
268	737	P737L	p.Pro737Leu	2	0	AsianSpecif	AsianSpecificDB		
269	740	S740P	p.Ser740Pro	1	4.59749E-05	gnomAD			
270	741	I741V	p.Ile741Val	20	0.000100345	gnomAD	Other		
271	745	V745I	p.Val745Ile	2	1.01287E-05	gnomAD			
272	751	G751E	p.Gly751Glu	2	1.1659E-05	gnomAD			
273	752	V752M	p.Val752Met	1	0	AsianSpecif	AsianSpecificDB		
274	753	I753M	p.Ile753Met	1	5.83216E-06	gnomAD	AsianSpecificDB		
275	761	I761V	p.Ile761Val	1	6.28943E-06	gnomAD			
276	767	D767H	p.Asp767His	2	1.34282E-05	gnomAD			
277	768	R768W	p.Arg768Trp	2	1.38522E-05	gnomAD			
278	769	K769E	p.Lys769Glu	1	6.92919E-06	gnomAD			

	A	B	C	D	E	F	G	H	I
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean Procko Score	Contact Residue
279	771	K771R	p.Lys771Arg	1	4.53762E-05	gnomAD			
280	772	N772S	p.Asn772Ser	12	6.02101E-05	gnomAD			
281	774	A774G	p.Ala774Gly	1	5.61605E-06	gnomAD			
282	774	A774P	p.Ala774Pro	1	5.61549E-06	gnomAD			
283	774	A774T	p.Ala774Thr	1	5.61549E-06	gnomAD			
284	776	S776R	p.Ser776Arg	1	4.55021E-05	gnomAD			
285	781	Y781H	p.Tyr781His	3	1.4797E-05	gnomAD			
286	782	A782V	p.Ala782Val	11	6.07839E-05	gnomAD	AsianSpecificDB,UKBio		
287	785	D785N	p.Asp785Asn	7	3.86678E-05	gnomAD	AsianSpecificDB,UKBio		
288	793	P793L	p.Pro793Leu	1	5.57324E-06	gnomAD			
289	796	Q796R	p.Gln796Arg	2	1.11324E-05	gnomAD			
290	801	V801G	p.Val801Gly	1	5.61728E-06	gnomAD	AsianSpecificDB,UKBio		
291	802	Q802R	p.Gln802Arg	1	0	AsianSpecif	AsianSpecificDB		
292	804	S804F	p.Ser804Phe	1	5.80124E-06	gnomAD			
293	805	F805I	p.Phe805Ile	4	2.3209E-05	gnomAD			
294	482	R482*	p.Arg482Ter	1	0	Other			
295	244	V244fs	p.Val244LeufsTe	1	0	Other			
296	656	L656*	p.Leu656Ter	1	5.9153E-06	gnomAD	AsianSpecificDB		
297	405	G405del	p.Gly405del	2	1.09215E-05	gnomAD			
298	422	G422fs	p.Gly422ValfsTe	1	5.46472E-06	gnomAD			
299	116	L116*	p.Leu116Ter	1	6.43882E-06	gnomAD			
300	313	K313del	p.Lys313del	1	5.642E-06	gnomAD			

	A	B	C	D	E	F	G	H	I	J	K	L
1	Supplementary Table 2 - ACE2 polymorphisms predicted to alter SARS-CoV/Cov-2 S-protein binding											
2	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean enrichment score (Procko 2020)	Contact Residue	Strctural mechanism (Fig 2)	impact (binding to S-protein)	Suceptibility to SARS-CoV/SRAS-CoV-2 (combined interpretation)
3	38	D38V	p.Asp38Val	1	0	Other		-1.9637347	Y	Direct	Decrease	Decrease
4	31	K31R	p.Lys31Arg	1	0	AsianSpecific	AsianSpecific	-1.6094772	Y	Direct	Decrease	Decrease
5	33	N33I	p.Asn33Ile	1	0	AsianSpecific	AsianSpecific	-1.5683037		Direct	Decrease	Decrease
6	37	E37K	p.Glu37Lys	8	3.897E-05	gnomAD	AsianSpecific	-1.3499866	Y	Indirect	Decrease	Decrease
7	35	E35K	p.Glu35Lys	3	1.636E-05	gnomAD		-1.1992532	Y	Direct	Decrease	Decrease
8	388	Q388L	p.Gln388Leu	4	2.186E-05	gnomAD		-1.0844672		Indirect	Decrease	Decrease
9	34	H34R	p.His34Arg	1	0	AsianSpecific	AsianSpecific	-1.0760313	Y	Direct	Decrease	Decrease
10	51	N51S	p.Asn51Ser	1	5.488E-06	gnomAD		-1.0527784		Indirect	Decrease	Decrease
11	68	K68E	p.Lys68Glu	2	1.095E-05	gnomAD	AsianSpecific	-1.052667		Indirect	Decrease	Decrease
12	352	G352V	p.Gly352Val	1	5.752E-06	gnomAD		-1.0253194		Direct	Decrease	Decrease
13	509	D509Y	p.Asp509Tyr	1	0	Other		-0.9429219		-	-	-
14	355	D355N	p.Asp355Asn	2	1.174E-05	gnomAD	Other	-0.9110524		Indirect	Decrease	Decrease
15	326	G326E	p.Gly326Glu	1	5.517E-06	gnomAD		-0.9054689	Y	Direct	Decrease	Decrease
16	72	F72V	p.Phe72Val	1	5.474E-06	gnomAD		-0.8867353		Indirect	Decrease	Decrease
17	62	M62V	p.Met62Val	1	5.666E-06	gnomAD	AsianSpecific	-0.7827955		Indirect	Decrease	Decrease
18	83	Y83H	p.Tyr83His	2	0	SAS(1)	AsianSpecific	-0.6671349	Y	Direct	Decrease	Decrease
19	50	Y50F	p.Tyr50Phe	1	5.482E-06	gnomAD		-0.5353604		Indirect	Decrease	Decrease
20	329	E329G	p.Glu329Gly	7	3.443E-05	gnomAD	Other	-0.5064661	Y	-	Decrease	Decrease
21	51	N51D	p.Asn51Asp	1	5.485E-06	gnomAD		-0.5008654		-	Decrease	Decrease
22	62	M62I	p.Met62Ile	1	0	AsianSpecific	AsianSpecific	-0.4332215		-	Decrease	Decrease
23	82	M82I	p.Met82Ile	5	2.442E-05	gnomAD		-0.4285018	Y	-	Decrease	Decrease
24	84	P84T	p.Pro84Thr	1	5.471E-06	gnomAD		-0.4191317		-	Decrease	Decrease
25	389	P389H	p.Pro389His	7	3.826E-05	gnomAD		-0.4007303		-	Decrease	Decrease
26	290	N290H	p.Asn290His	2	1.133E-05	gnomAD		-0.3820868		-	-	-
27	518	R518T	p.Arg518Thr	1	5.49E-06	gnomAD		-0.3408209		-	-	-
28	504	F504I	p.Phe504Ile	2	1.268E-05	gnomAD		-0.3318035		-	-	-
29	511	S511P	p.Ser511Pro	1	0	AsianSpecific	AsianSpecific	-0.0965233		-	-	-
30	346	P346S	p.Pro346Ser	1	5.599E-06	gnomAD		-0.0669063		-	-	-
31	504	F504L	p.Phe504Leu	1	4.631E-05	gnomAD		-0.0355608		-	-	-
32	366	M366T	p.Met366Thr	2	1.098E-05	gnomAD		0.0341968		-	-	-

	A	B	C	D	E	F	G	H	I	J	K	L
	Position	Mutation	Protein Consequence	Allele Count	gnomAD Allele Frequency	Primary Source	Additional Sources	Mean enrichment score (Procko 2020)	Contact Residue	Strctural mechanism (Fig 2)	impact (binding to S-protein)	Suceptibility to SARS-CoV/SRAS-CoV-2 (combined interpretation)
2												
33	446	I446M	p.Ile446Met	1	4.54E-05	gnomAD		0.0591718		-	-	-
34	21	I21T	p.Ile21Thr	1	5.463E-06	gnomAD		0.072108		-	-	-
35	374	H374R	p.His374Arg	1	5.475E-06	gnomAD		0.1107117		-	-	-
36	398	E398K	p.Glu398Lys	1	5.462E-06	gnomAD		0.2057975		-	-	-
37	510	Y510H	p.Tyr510His	1	6.863E-06	gnomAD		0.2460174		-	-	-
38	445	T445M	p.Thr445Met	1	5.869E-06	gnomAD		0.4226164		-	-	-
39	40	F40L	p.Phe40Leu	2	3.086E-05	Gnomad		0.4490517		-	Increase	Increased
40	26	K26E	p.Lys26Glu	1	5.455E-06	gnomAD		0.5734929		-	Increase	Increased
41	60	Q60R	p.Gln60Arg	2	1.125E-05	gnomAD		0.6028484		-	Increase	Increased
42	23	E23K	p.Glu23Lys	1	5.458E-06	gnomAD		0.6477777		Indirect	Increase	Increased
43	383	M383T	p.Met383Thr	1	0	AsianSpecific	AsianSpecific	0.6762501		-	Increase	Increased
44	21	I21V	p.Ile21Val	2	1.093E-05	gnomAD	Other	0.6970124		Direct	Increase	Increased
45	378	H378R	p.His378Arg	18	8.791E-05	gnomAD	Other	0.8315449		-	Increase	Increased
46	64	N64K	p.Asn64Lys	3	1.466E-05	gnomAD		1.0472096		-	Increase	Increased
47	102	Q102P	p.Gln102Pro	3	1.475E-05	gnomAD		1.1913283		-	Increase	Increased
48	26	K26R	p.Lys26Arg	797	0.0038833	gnomAD	AsianSpecific	1.2292467		Direct	Increase	Increased
49	27	T27A	p.Thr27Ala	2	1.091E-05	gnomAD		2.1134249	Y	Indirect	Increase	Increased
50	19	S19P	p.Ser19Pro	64	0.0003129	gnomAD	Other	2.3278184		Direct	Increase	Increased
51	92	T92I	p.Thr92Ile	2	1.096E-05	gnomAD		2.5342868		Direct	Increase	Increased

	A	B	C	D	E	F	G	H
1	Supplementary Table 3 - Conservation of N90 glycosylation motif in annotated jawed vertebrate ACE2 orthologs							
2								
3	Common name	Refseq ID	Residue #90	Residue #90	Residue #91	Residue #92	Residue #93	Motif NxT/S present?
4	Human	NP_001358344.1	Q	N	L	T	V	Yes
5	house mouse	NP_081562.2	Q	T	P	I	I	No
6	Norway rat	NP_001012006.1	Q	N	A	T	I	No
7	zebrafish	XP_005169416.1	S	D	P	I	I	No
8	pig	NP_001116542.1	Q	T	L	I	L	No
9	Rhesus monkey	NP_001129168.1	Q	N	L	T	V	Yes
10	cattle	XP_005228485.1	Q	N	L	T	L	Yes
11	dog	NP_001158732.1	Q	D	S	T	V	No
12	rabbit	XP_002719891.1	Q	N	L	T	V	Yes
13	tropical clawed	XP_002938293.2	T	D	P	S	I	No
14	chicken	XP_416822.2	Q	D	A	V	T	No
15	chimpanzee	XP_016798468.1	Q	N	L	T	V	Yes
16	domestic cat	XP_023104564.1	H	N	T	T	V	Yes
17	sheep	XP_011961657.1	Q	N	L	T	L	Yes
18	rainbow trout	XP_021433278.1	S	D	P	L	I	No
19	Atlantic cod	XP_030232530.1	K	D	P	V	V	No
20	giant panda	XP_002930657.1	H	N	S	T	V	Yes
21	Brandt's bat	XP_014399780.1	Q	N	L	T	I	Yes
22	elephant shark	XP_007889845.1	S	D	N	I	I	No
23	domestic ferret	NP_001297119.1	Q	D	P	I	I	No
24	golden hamster	XP_005074266.1	Q	N	L	T	I	Yes
25	naked mole-rat	XP_004866157.1	Q	N	L	T	V	Yes
26	wild yak	XP_005903173.1	Q	N	L	T	L	Yes
27	barramundi per	XP_018539189.1	K	D	Q	E	I	No
28	white-tufted-ea	XP_008987241.1	Q	N	L	T	V	Yes
29	horse	XP_001490241.1	Q	N	L	T	V	Yes
30	greater amberj	XP_022605054.1	K	N	P	E	I	No

	A	B	C	D	E	F	G	H
3	Common name	Refseq ID	Residue #90	Residue #90	Residue #91	Residue #92	Residue #93	Motif NxT/S present?
31	turquoise killifish	XP_015808977.1	K	D	P	E	V	No
32	two-lined caecilian	XP_029459086.1	T	E	P	E	I	No
33	Microcaecilia unicolor	XP_030058174.1	T	D	P	E	T	No
34	Mexican tetraodon	XP_022523929.1	S	D	E	L	V	No
35	coelacanth	XP_005997915.2	T	D	P	H	I	No
36	spotted gar	XP_006639185.1	A	D	K	K	I	No
37	Atlantic herring	XP_031414786.1	N	D	L	E	I	No
38	goldfish	XP_026131313.1	S	D	P	L	I	No
39	channel catfish	XP_017313836.1	S	D	H	E	V	No
40	electric eel	XP_026867211.1	T	D	P	E	I	No
41	northern pike	XP_010884777.1	K	D	P	L	I	No
42	Atlantic salmon	XP_014062928.1	.	.	.	.	.	NA
43	Arctic char	XP_023998967.1	S	V	I	I	D	No
44	mummichog	XP_021178197.1	K	D	P	Q	I	No
45	guppy	XP_008402714.1	S	D	P	V	I	No
46	southern platyfish	XP_005799835.1	N	D	P	V	I	No
47	Nile tilapia	XP_003445853.2	N	D	L	E	I	No
48	Burton's mouthbrooder	XP_005943362.1	N	D	L	E	I	No
49	eastern happygob	XP_026020155.1	N	D	L	E	I	No
50	yellow perch	XP_028441363.1	K	D	P	E	I	No
51	gilthead seabream	XP_030271236.1	K	D	R	E	L	No
52	black rockcod	XP_010790455.1	T	D	A	T	I	No
53	Japanese flounder	XP_019935235.1	K	D	A	K	I	No
54	Green sea turtle	XP_007070561.1	M	D	P	I	V	No
55	Painted turtle	XP_023964517.1	T	D	P	I	V	No
56	American alligator	XP_019350687.1	M	D	P	L	I	No
57	Australian saltwater crocodile	XP_019384826.1	.	D	P	V	I	No
58	mainland tiger	XP_026530754.1	S	N	E	T	I	Yes
59	Pseudonaja texensis	XP_026570054.1	A	N	E	T	I	Yes
60	emu	XP_025976569.1	T	D	D	L	I	No
61	mallard	XP_012949915.2	Q	D	P	L	L	No
62	chimney swift	XP_009992128.1	S	D	A	L	I	No

	A	B	C	D	E	F	G	H
3	Common name	Refseq ID	Residue #90	Residue #90	Residue #91	Residue #92	Residue #93	Motif NxT/S present?
63	rock pigeon	XP_021154486.1	Q	D	D	L	T	No
64	peregrine falco	XP_005231984.2	Q	D	A	L	T	No
65	white-tailed eag	XP_009925641.1	Q	D	D	L	T	No
66	helmeted guine	XP_021240731.1	Q	D	A	V	T	No
67	Ring-necked pl	XP_031451919.1	Q	D	A	A	T	No
68	turkey	XP_019467554.1	Q	D	A	A	T	No
69	Common cana	XP_009087922.1	K	D	D	L	T	No
70	Great Tit	XP_015486815.1	T	D	D	L	T	No
71	Common starlin	XP_014731370.1	T	D	D	L	I	No
72	great cormoran	XP_009509070.1	Q	D	A	L	T	No
73	emperor pengu	XP_009275140.1	Q	D	A	L	T	No
74	Adelie penguin	XP_009323767.1	Q	D	T	L	T	No
75	Anna's hummir	XP_008492997.2	T	D	A	L	I	No
76	platypus	XP_001515597.2	S	D	R	S	L	No
77	Tasmanian dev	XP_031814825.1	S	A	Y	P	I	No
78	nine-banded ar	XP_004449124.1	S	N	L	T	N	Yes
79	western Europe	XP_007538670.1	Q	N	P	T	V	No
80	small Madagasca	XP_004710002.1	T	D	P	I	I	No
81	black flying fox	XP_006911709.1	Q	D	P	I	L	No
82	Egyptian rouse	XP_015974412.1	Q	D	P	E	L	No
83	common vamp	XP_024425698.1	K	D	V	N	V	No
84	Tufted capuchi	XP_032141854.1	Q	N	L	T	V	Yes
85	sooty mangabe	XP_011891198.1	Q	N	L	T	V	Yes
86	crab-eating ma	XP_005593094.1	Q	N	L	T	V	Yes
87	pig-tailed maca	XP_011733505.1	Q	N	L	T	V	Yes
88	olive baboon	XP_021788732.1	Q	N	L	T	V	Yes
89	gelada	XP_025227847.1	Q	N	L	T	V	Yes
90	drill	XP_011850923.1	Q	N	L	T	V	Yes
91	western gorilla	XP_018874749.1	Q	N	L	T	I	Yes
92	pygmy chimpan	XP_008972428.1	Q	N	L	T	V	Yes
93	Sumatran oran	NP_001124604.1	Q	N	L	T	V	Yes
94	red fox	XP_025842512.1	Q	D	S	T	V	No

	A	B	C	D	E	F	G	H
3	Common name	Refseq ID	Residue #90	Residue #90	Residue #91	Residue #92	Residue #93	Motif NxT/S present?
95	leopard	XP_019273508.1	H	N	T	T	V	Yes
96	puma	XP_025790417.1	H	N	T	T	V	Yes
97	California sea lion	XP_027465353.1	Q	D	S	T	V	No
98	Pacific walrus	XP_004415448.1	Q	D	S	T	V	No
99	Weddell seal	XP_030886750.1	.	.	.	.	.	NA
100	harbor seal	XP_032245506.1	Q	D	S	T	V	No
101	long-finned pilot whale	XP_030703991.1	R	N	L	T	L	Yes
102	killer whale	XP_004269705.1	R	N	L	T	L	Yes
103	common bottlenose dolphin	XP_019781177.1	R	N	L	T	L	Yes
104	beluga whale	XP_022418360.1	R	N	L	T	L	Yes
105	sperm whale	XP_023971279.1	Q	N	L	T	L	Yes
106	African savanna elephant	XP_023410960.1	S	S	S	I	I	No
107	ass	XP_014713133.1	Q	N	L	T	V	Yes
108	Przewalski's horse	XP_008542995.1	Q	N	L	T	V	Yes
109	Bactrian camel	XP_010966303.1	Q	N	V	T	L	Yes
110	Arabian camel	XP_010991717.1	Q	N	V	T	L	Yes
111	Odokoileus virginianus	XP_020768965.1	Q	N	L	T	L	Yes
112	zebu cattle	XP_019811719.1	Q	N	L	T	L	Yes
113	goat	NP_001277036.1	Q	N	L	T	L	Yes
114	Malayan pangolin	XP_017505746.1	Q	N	D	T	I	Yes
115	American pika	XP_004597549.2	Q	N	L	T	T	Yes
116	Alpine marmot	XP_015343540.1	Q	N	F	T	L	Yes
117	Arctic ground sloth	XP_026252505.1	Q	N	F	T	L	Yes
118	Ord's kangaroo rat	XP_012887572.1	Q	N	P	I	L	No
119	Chinese hamster	XP_003503283.1	Q	N	L	I	I	No
120	white-footed mouse	XP_028743609.1	P	N	L	I	I	No
121	Ryukyu mouse	XP_021009138.1	Q	T	P	I	I	No
122	shrew mouse	XP_021043935.1	Q	N	P	V	I	No
123	domestic guinea pig	XP_023417808.1	Q	N	L	T	V	Yes
124	degu	XP_023575315.1	Q	N	L	T	V	Yes
125	gray short-tailed opossum	XP_007500935.1	T	N	A	T	V	Yes
126	Chinese soft-shelled turtle	XP_006122891.1	T	N	H	T	V	Yes

	A	B	C	D	E	F	G	H
3	Common name	Refseq ID	Residue #90	Residue #90	Residue #91	Residue #92	Residue #93	Motif NxT/S present?
127	Bolivian squirrel	XP_010334925.1	Q	N	L	T	V	Yes
128	reedfish	XP_028655640.1	S	N	Y	T	I	Yes
129	green anole	XP_008105455.1	N	N	D	T	I	Yes
130	Cape elephant	XP_006892457.1	S	D	P	S	I	No
131	sheepshead m	XP_015226730.1	K	D	L	Q	I	No
132	polar bear	XP_008694637.1	H	N	S	T	V	Yes
133	big brown bat	XP_008153150.1	Q	N	L	T	I	Yes
134	Hawaiian monk	XP_021536480.1	Q	D	S	T	L	No
135	common wombat	XP_027691156.1	S	D	P	Q	I	No
136	Opisthocomus	XP_009938970.1	Q	D	A	L	T	No
137	Northern fulmar	XP_009574896.1	H	D	A	L	T	No
138	Nothoprocta p	XP_025891105.1	K	N	D	L	I	No
139	hybrid cattle	XP_027389727.1	Q	N	L	T	L	Yes
140	alpaca	XP_006212709.1	E	N	V	T	L	Yes
141	gray mouse lemur	XP_020140826.1	Q	N	L	T	I	Yes
142	small-eared galap	XP_003791912.1	Q	N	R	T	V	Yes
143	Indian medaka	XP_024150631.1	K	D	P	E	I	No
144	torafugu	XP_029702274.1	K	N	A	E	I	No
145	Monterrey platy	XP_027871671.1	N	D	P	V	I	No
146	cheetah	XP_026910297.1	H	N	T	T	V	Yes
147	long-tailed chin	XP_013362428.1	Q	N	L	T	V	Yes
148	northern fur seal	XP_025713397.1	Q	D	S	T	V	No
149	Steller sea lion	XP_027970822.1	Q	D	S	T	V	No
150	Western terrestrial	XP_032082934.1	T	N	E	T	I	Yes
151	Thamnophis sierrae	XP_013926936.1	T	N	E	T	I	Yes
152	southern multicolored	XP_031226742.1	K	T	P	I	I	No
153	Dalmatian pelican	XP_009478920.1	Q	D	D	L	T	No
154	ermine	XP_032187677.1	Q	D	P	I	I	No
155	mangrove rivulus	XP_017295385.1	H	D	P	T	V	No
156	meerkat	XP_029786256.1	Q	N	T	T	V	Yes
157	red-throated loach	XP_009816127.1	Q	D	A	L	I	Yes
158	Ma's night monkey	XP_012290105.1	Q	N	L	T	V	Yes

	A	B	C	D	E	F	G	H
3	Common name	Refseq ID	Residue #90	Residue #90	Residue #91	Residue #92	Residue #93	Motif NxT/S present?
159	koala	XP_020863153.1	S	D	P	Q	I	No
160	Chinese alligator	XP_025066628.1	.	D	P	L	I	No
161	narwhal	XP_029095804.1	R	N	L	T	L	Yes
162	European shrew	XP_004612266.1	T	D	P	K	V	No
163	red-bellied piranha	XP_017550079.1	S	D	P	L	I	No
164	thirteen-lined ground sloth	XP_005316051.3	Q	N	F	T	L	Yes
165	Bison bison bison	XP_010833001.1	Q	N	L	T	L	Yes
166	swamp eel	XP_020465646.1	R	D	A	E	I	No
167	white-throated sparrow	XP_005491832.2	T	D	E	L	T	No
168	Oreochromis aeneus	XP_031584810.1	N	D	L	E	I	No
169	Amazon molly	XP_007560208.1	N	D	P	V	I	No
170	sailfin molly	XP_014895313.1	N	D	P	V	I	No
171	Poecilia mexicana	XP_014837025.1	N	D	P	V	I	No
172	medium ground finch	XP_005426221.1	T	D	E	L	T	No
173	killdeer	XP_009887331.1	Q	D	P	L	I	No
174	lesser Egyptian tortoise	XP_004671523.1	Q	N	P	T	I	No
175	bald eagle	XP_010579828.1	Q	D	D	L	T	No
176	Austrofundulus holbrooki	XP_013888928.1	N	D	P	N	I	No
177	brown roatelo	XP_010178703.1	Q	D	P	L	I	No
178	Red-legged seriema	XP_009703695.1	Q	D	A	L	T	No
179	sunbittern	XP_010156467.1	Q	D	A	L	I	No
180	common cuckoo	XP_009563864.1	Q	D	A	L	T	No
181	Barn owl	XP_009969209.1	Q	D	A	L	T	No
182	Cottoperca gobio	XP_029283581.1	R	D	S	T	I	No
183	ballan wrasse	XP_020493627.1	T	I	P	E	I	No
184	rifleman	XP_009082150.1	.	.	.	.	.	NA
185	bar-tailed trogon	XP_009867056.1	G	D	D	L	I	No
186	speckled mouse lemur	XP_010206054.1	.	.	.	.	.	NA
187	carmine bee-eater	XP_008937519.1	Q	N	A	T	T	Yes
188	little brown bat	XP_023609437.1	Q	N	S	T	I	Yes
189	zebra finch	XP_002194303.3	A	D	D	P	T	No
190	collared flycatcher	XP_005037422.1	T	D	D	L	T	No

	A	B	C	D	E	F	G	H
3	Common name	Refseq ID	Residue #90	Residue #90	Residue #91	Residue #92	Residue #93	Motif NxT/S present?
191	green monkey	XP_007989304.1	Q	N	L	T	V	Yes
192	Canada lynx	XP_030160839.1	H	N	T	T	I	Yes
193	black snub-nose	XP_017744069.1	Q	N	L	T	V	Yes
194	golden snub-nose	XP_010364367.2	Q	N	L	T	V	Yes
195		XP_003261132.2	Q	N	L	T	I	Yes
196	flier cichlid	XP_030582139.1	Q	D	L	E	I	No
197	climbing perch	XP_026233431.1	S	D	P	E	I	No
198	Amur tiger	XP_007090142.1	H	N	T	T	V	Yes
199	prairie vole	XP_005358818.1	Q	N	L	L	L	No
200	spiny chromis	XP_022063988.1	T	D	P	E	I	No
201	clown anemonefish	XP_023124156.1	K	D	P	E	I	No
202	American crow	XP_017583883.1	.	.	.	.	.	NA
203	Camarhynchus	XP_030811385.1	T	D	E	L	T	No
204	water buffalo	XP_006041602.1	Q	N	L	T	L	Yes
205	pale spear-nose	XP_028378317.1	T	D	V	T	V	No
206	Pacific white-sided	XP_026951598.1	R	N	L	T	L	Yes
207	yellow-bellied marmoset	XP_027802308.1	Q	N	F	T	L	Yes
208	Japanese quail	XP_015742063.1	Q	D	A	V	T	No
209	white-throated	XP_010217584.1	K	D	D	L	I	No
210	Gharial	XP_019381060.1	.	D	P	L	I	No
211	White-tailed tropicbird	XP_010290019.1	Q	D	A	L	T	No
212	central bearded dragon	XP_020642422.1	S	N	E	T	I	Yes
213	Protobothrops	XP_029140508.1	T	N	E	T	I	Yes
214	zebra mbuna	XP_004543482.1	N	D	L	E	I	No
215	tiger tail seahorse	XP_019742561.1	K	D	P	Q	I	No
216	Asian bonytongue	XP_018584732.1	T	D	P	T	V	No
217	saffron-crested	XP_027544864.1	E	D	N	L	I	No
218	Ursus arctos horreorum	XP_026333865.1	H	N	S	T	V	Yes
219	Downy woodpecker	XP_009909849.1	.	.	.	.	.	NA
220	Yangtze River catfish	XP_007466389.1	Q	N	L	T	L	Yes
221	red-crested turp	XP_009978415.1	Q	D	A	L	I	No
222	Nanorana parkmanni	XP_018418558.1	T	D	E	M	L	No

	A	B	C	D	E	F	G	H
3	Common name	Refseq ID	Residue #90	Residue #90	Residue #91	Residue #92	Residue #93	Motif NxT/S present?
223	crested ibis	XP_009474590.1	Q	D	A	L	T	No
224	large flying fox	XP_011361275.1	Q	D	P	I	L	No
225	star-nosed mole	XP_012585871.1	Q	D	P	I	V	No
226	bicolor damselfish	XP_008290762.1	K	D	P	E	I	No
227	Gekko japonicus	XP_015273067.1	S	D	P	H	I	No
228	great blue-spot	XP_020781598.1	K	D	R	E	I	No
229	blue tit	XP_023774184.1	T	D	D	L	T	No
230	Siamese fighting fish	XP_028999570.1	S	D	P	E	I	No
231	willow flycatcher	XP_027757151.1	E	D	N	L	I	No
232	live sharksucker	XP_029354066.1	K	D	P	E	I	No
233	Buceros rhinoceros	XP_010136813.1	Q	H	D	L	T	No
234	Kea	XP_010012481.1	.	.	.	.	.	NA
235	Burmese python	XP_007431942.2	T	D	E	T	I	No
236	Tibetan ground sloth	XP_005516712.1	T	D	D	L	T	No
237	jewelled blenny	XP_029949252.1	S	D	P	E	I	No
238	Cape golden molly	XP_006835673.1	S	N	S	T	I	Yes
239	great roundleaf	XP_019522936.1	Q	N	A	T	I	Yes
240	Macqueen's buck	XP_010120523.1	Q	D	A	L	T	No
241	cuckoo roller	XP_009954393.1	Q	D	A	L	T	No
242	little egret	XP_009638257.1	E	D	D	A	T	No
243	burrowing owl	XP_026705725.1	Q	D	A	L	T	No
244	ruff	XP_014815705.1	Q	D	D	L	T	No
245	Apteryx mantelli	XP_013805736.1	K	D	D	L	I	No
246	zig-zag eel	XP_026175949.1	R	D	P	E	I	No
247	Indian glassy fish	XP_028257887.1	T	D	P	E	I	No
248	large yellow crocodile	XP_010730146.1	K	N	P	I	I	No
249	Aquila chrysaetos	XP_011587755.1	Q	D	D	L	T	No
250	tufted duck	XP_032058386.1	Q	D	P	L	L	No
251	Aquila chrysaetos	XP_029855025.1	Q	D	D	L	T	No
252	Myotis davidii	XP_006775273.1	Q	N	P	T	I	Yes
253	prairie deer mouse	XP_006973269.1	Q	N	L	I	I	No
254	yellow-throated	XP_010084373.1	Q	D	V	L	T	No

	A	B	C	D	E	F	G	H
3	Common name	Refseq ID	Residue #90	Residue #90	Residue #91	Residue #92	Residue #93	Motif NxT/S present?
255	tongue sole	XP_016887914.1	K	D	P	E	I	No
256	Chinese tree shrew	XP_006164754.1	Q	D	T	T	E	No
257	chuck-will's-wid	XP_010169238.1	Q	D	A	L	I	No
258	pike-perch	XP_031162227.1	K	D	P	E	I	No
259	dingo	XP_025292925.1	Q	D	S	T	V	No
260	Miniopterus na	XP_016058453.1	Q	N	S	S	T	Yes
261	Bengalese finch	XP_021388026.1	A	D	D	P	T	No
262	denticle herring	XP_028837781.1	T	D	P	T	N	No
263	Cichlid	XP_005724169.1	N	D	L	E	I	No
264	Okarito brown	XP_025942946.1	K	D	D	L	I	No
265	Balaenoptera a	XP_028020351.1	Q	N	L	T	L	Yes
266	striped catfish	XP_026803610.1	S	D	Q	E	I	No
267	blue-crowned n	XP_017667729.1	E	D	N	L	I	No
268	golden-collared	XP_017939494.2	D	D	N	L	I	No
269	Saker falcon	XP_005443093.2	Q	D	A	L	T	No
270	Coquerel's sifa	XP_012494185.1	Q	N	V	T	V	Yes
271	Anser cygnoide	XP_013039300.1	Q	D	P	L	I	No
272	White-ruffed m	XP_027494818.1	E	D	N	L	I	No
273	Wild Bactrian c	XP_006194263.1	Q	N	V	T	L	Yes
274	wolf-eel	XP_031702716.1	N	D	T	K	I	No
275	blunt-snouted c	XP_028297875.1	T	D	L	G	I	No
276	Struthio camelu	XP_009667495.1	N	D	D	L	I	No
277	Grammomys s	XP_028617961.1	K	T	P	I	I	No
278	pinecone soldie	XP_029904152.1	T	D	P	E	I	No
279	Ugandan red C	XP_023054821.1	Q	N	L	T	V	Yes
280	Wire-tailed mar	XP_027593974.1	D	D	N	L	I	No
281	Damara mole-r	XP_010643477.1	Q	N	L	T	V	Yes
282	Corvus cornix c	XP_010392735.2	T	D	D	L	T	No
283	Upper Galilee r	XP_008839098.1	Q	D	L	V	I	No
284	New Caledonia	XP_031956594.1	T	D	D	L	T	No
285	Orycteropus af	XP_007951028.1	S	N	S	T	I	Yes
286	yellow catfish	XP_027024524.1	S	D	P	E	T	No

	A	B	C	D	E	F	G	H
3	Common name	Refseq ID	Residue #90	Residue #90	Residue #91	Residue #92	Residue #93	Motif NxT/S present?
287	Sinocyclocheilus	XP_016345325.1	S	D	P	L	I	No
288	Paramormyrops	XP_023679669.1	T	D	P	T	I	No
289	Yangtze finless	XP_024599894.1	R	N	P	T	L	No
290	Cebus capucinus	XP_017367865.1	Q	N	L	T	V	Yes
291	Goodes thornbill	XP_030407881.1	T	D	P	I	V	No
292	Seriola lalandi	XP_023257445.1	K	N	P	E	I	No
293	Philippine tarsier	XP_008062810.1	Q	N	S	T	I	Yes
294	Kakapo	XP_030332639.1	E	D	A	L	T	No
295	budgerigar	XP_005151516.1	Q	D	A	L	T	No
296	southern white	XP_004435206.1	Q	N	V	T	V	Yes
297	Florida manatee	XP_004386381.1	S	S	S	V	I	No
298	Enhydra lutris	XP_022374078.1	Q	D	P	I	N	No

	A	B	C	D
1	Supplementary Table 4 - Structure evaluation findings			
2				
3	Mutation	Categorization	Allele Count	Rationale
4	S19P	Direct Enhancing	64	In PDB 6LZG, S19 forms backbone interactions with RBD A475 and backbone interactions with E25 of the same helix. Although these interactions may be stabilizing, prolines are well suited for N-terminal capping of helices and the energetic cost of binding could outstrip the moderate gain achieved through one polar contact. By maintaining more of the helix, S19P could remove this slight steric hindrance where the RBD ridge protrudes.
5	I21V	Indirect Enhancing	2	Likely weakens hydrophobic interaction with ACE2 P84 and reduces steric constraints for the bulky ACE2 Y83 to establish hydrophobic interactions with RBD F486 and polar contact with RBD N487. May also facilitate S19 interaction with the RBD.
6	T27A	Direct Enhancing	2	T27A removes side-backbone and backbone-backbone interactions. This could endow a1 with greater flexibility to accommodate the RBD ridge binding loop.
7	K26R	Direct Enhancing	797	K26R destabilizes N90 glycan while stabilizing intrahelical interaction with ACE2 D30. The N90 glycan may sterically hinder viral recognition.
8	T92I	Direct Enhancing	2	Prevents N90 glycosylation by mutating NxS/T glycosylation motif. Additionally, T92 is a helix capping residue. T92I would likely destabilize helix and presentation of the N90-linked glycan.
9	E23K	Indirect Enhancing	1	E23 is important for helical fold and likely positions K26 away for interaction with the N90 glycan.
10	M383T	Indirect Enhancing	1	M383T appears to stabilize the overall fold with sidechain polar contacts with I379 and Q380.
11	K31R	Direct disrupting	1	Loss of polar contact with CoV2 RBD Q493. This residue is distinct from CoV and represents a strengthened interaction.
12	N33I	Indirect disrupting	1	N33 forms backbone-backbone and sidechain-backbone contacts with ACE2 L29. Coordinates polar contacts with an H ₂ O with ACE2 Q96 which helps to appropriately position N90 containing helix. Loss of these interactions could destabilize the central concave portion of ACE2 a1 helix resulting in loss of RBD affinity.
13	H34R	Direct disrupting	1	H34 establishes a polar contact with RBD Y453 and intramolecularly contacts D30. H34R would abolish both of those interactions critical for affinity.
14	E35K	Direct disrupting	3	E35K likely results in loss of polar contact with RBD Q493.

	A	B	C	D
15	E37K	Indirect disrupting	8	E37 participates in numerous contacts, including an intramolecular salt-bridge with ACE2 R393 and a sidechain-mainchain polar contact with K353 on the ACE2 b-hairpin that contributes to the binding interface. E37K may unfix the position of b-hairpin relative to a1 such that K353 no longer, participates in a polar contact with RBD G396.
16	D38V	Direct disrupting	1	Loss of two strong polar contacts with CoV2 RBD Y449.
17	Y50F	Direct disrupting	2	Loss of polar contact with RBD N487but may retain some hydrophobic packing with F486.
18	N51S	Indirect disrupting	1	Located on the C-terminal end of a1, N51 potentially helps anchor the helix through backbone-backbone polar contact with ACE2 L359 and backbone-backbone sidechain-backbone interactions with S47. Loss of these interactions could result destabilize a1 concavity which may alter affinity.
19	K68E	Indirect disrupting	2	Located on the a2 helix, K68 may be important for helix-helix packing. K68E could disrupt the packing critical for the binding interface.
20	F72V	Indirect disrupting	1	Located on the a2 helix, F72 participates in hydrophobic interactions and appears important in positioning of a1 helix.
21	Y83H	Direct disrupting	2	Loss of polar contact with RBD N487
22	G326E	Direct disrupting	1	Induces steric clash with RBD G502
23	G352V	Direct disrupting	1	Steric clash with RBD Y505
24	D355N	Indirect disrupting	2	D355 is critical for the stability of the b-hairpin loop that forms part of the binding interface.
25	Q388L	Indirect disrupting	4	Q388 stabilizes a loop proximal to the ACE2-RBD interface through sidechain-backbone polar contact with ACE2 L560. Q388L would diminish this interaction.