

## Human ACE2 receptor polymorphisms and altered susceptibility to SARS-CoV-2

Kushal Suryamohan<sup>1,15</sup>, Devan Diwanji<sup>2,3,15</sup>, Eric W. Stawiski<sup>1,15</sup>, Ravi Gupta<sup>4</sup>, Shane Miersch<sup>5</sup>, Jiang Liu<sup>6</sup>, Chao Chen<sup>5</sup>, Ying-Ping Jiang<sup>6</sup>, Frederic A. Fellouse<sup>7</sup>, J. Fah Sathirapongsasuti<sup>1</sup>, Patrick K. Albers<sup>8</sup>, Tanneeru Deepak<sup>4</sup>, Reza Saberianfar<sup>7</sup>, Aakrosh Ratan<sup>9</sup>, Gavin Washburn<sup>1</sup>, Monika Mis<sup>1</sup>, Devi Santhosh<sup>6</sup>, Sneha Somasekar<sup>10</sup>, G. H. Hiranjith<sup>1</sup>, Derek Vargas<sup>1</sup>, Sangeetha Mohan<sup>11</sup>, Sameer Phalke<sup>11</sup>, Boney Kuriakose<sup>12</sup>, Aju Antony<sup>11</sup>, Mart Ustav Jr<sup>5</sup>, Stephan C. Schuster<sup>13</sup>, Sachdev Sidhu<sup>5</sup>, Jagath R. Junutula<sup>6</sup>, Natalia Jura<sup>2,3</sup>✉ & Somasekar Seshagiri<sup>6,14</sup>✉

COVID-19 is a respiratory illness caused by a novel coronavirus called SARS-CoV-2. The viral spike (S) protein engages the human angiotensin-converting enzyme 2 (ACE2) receptor to invade host cells with ~10–15-fold higher affinity compared to SARS-CoV S-protein, making it highly infectious. Here, we assessed if ACE2 polymorphisms can alter host susceptibility to SARS-CoV-2 by affecting this interaction. We analyzed over 290,000 samples representing >400 population groups from public genomic datasets and identified multiple ACE2 protein-altering variants. Using reported structural data, we identified natural ACE2 variants that could potentially affect virus–host interaction and thereby alter host susceptibility. These include variants S19P, I21V, E23K, K26R, T27A, N64K, T92I, Q102P and H378R that were predicted to increase susceptibility, while variants K31R, N33I, H34R, E35K, E37K, D38V, Y50F, N51S, M62V, K68E, F72V, Y83H, G326E, G352V, D355N, Q388L and D509Y were predicted to be protective variants that show decreased binding to S-protein. Using biochemical assays, we confirmed that K31R and E37K had decreased affinity, and K26R and T92I variants showed increased affinity for S-protein when compared to wildtype ACE2. Consistent with this, soluble ACE2 K26R and T92I were more effective in blocking entry of S-protein pseudotyped virus suggesting that ACE2 variants can modulate susceptibility to SARS-CoV-2.

<sup>1</sup> Research and Development Department, MedGenome Inc, Foster City, CA, USA. <sup>2</sup> Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, USA. <sup>3</sup> Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA, USA. <sup>4</sup> MedGenome Labs Ltd., Bangalore, Karnataka, India. <sup>5</sup> Department of Molecular Genetics, and the Terrence Donnelly Center for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada. <sup>6</sup> ModMab Therapeutics, Foster City, CA, USA. <sup>7</sup> ModMab Therapeutics, Accelerator for Donnelly Collaboration, University of Toronto, Toronto, ON, Canada. <sup>8</sup> Wellcome Sanger Institute, Cambridge, UK. <sup>9</sup> Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. <sup>10</sup> Midwestern University, Glendale, AZ, USA. <sup>11</sup> Department of Molecular Biology, SciGenom Labs Pvt Ltd, Kochi, Kerala, India. <sup>12</sup> AgriGenome Labs Private Ltd, Kochi, Kerala, India. <sup>13</sup> Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, Singapore. <sup>14</sup> SciGenom Research Foundation, Bangalore, Karnataka, India. <sup>15</sup> These authors contributed equally: Kushal Suryamohan, Devan Diwanji, Eric W. Stawiski. ✉email: [natalia.jura@ucsf.edu](mailto:natalia.jura@ucsf.edu); [sekar@sgrf.org](mailto:sekar@sgrf.org)

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<sup>1</sup>. CoVs are enveloped non-segmented positive-sense single stranded RNA viruses and are classified into  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -CoVs<sup>1</sup>. While  $\alpha$ - and  $\beta$ -CoVs infect mammals, the  $\gamma$ - and  $\delta$ -CoVs generally infect birds<sup>1</sup>. Previously,  $\alpha$ -CoVs HCoV-229E and HCoV-NL63, and  $\beta$ -CoVs HCoV-HKU1 and HCoV-OC43 have been found to infect humans leading to mild symptoms<sup>1,2</sup>. More recently, three  $\beta$ -CoVs: severe acute respiratory syndrome coronavirus (SARS-CoV) in 2003<sup>1,3</sup>, Middle-East respiratory syndrome coronavirus in 2012 (MERS-CoV)<sup>1,4</sup>, and SARS-CoV-2 in 2019<sup>5-7</sup> have crossed the species barrier to infect humans resulting in respiratory illnesses including pneumonia that can be fatal.

SARS-CoV-2 is a novel coronavirus (2019-nCoV) first reported in December 2019 and is the cause of an ongoing global pandemic<sup>5-7</sup>. It has infected over 130 million people in 181 countries leading to over 2.8 million deaths as of April, 2021<sup>8</sup>. Sequence analysis of the SARS-CoV-2 genome revealed that it is closer to the bat CoV RaTG13 (96.2% identical) than to SARS-CoV (79.5% identical) that was responsible for the 2003 epidemic, suggesting that this novel virus originated in bats independently before jumping to humans either directly or through a yet to be determined intermediary host<sup>9</sup>.

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<sup>10-15</sup>. 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<sup>1</sup>. 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<sup>1,12,16-18</sup>.

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<sup>19</sup>.

Several structural studies involving the SARS-CoV-2 S-protein RBD and ACE2 PD have identified key residues involved in their interaction<sup>19-22</sup>. 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<sup>20,22</sup>, potentially contributing to the 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 wanted to ascertain if there were natural ACE2 variations that would decrease or increase its affinity to the S-protein RBD and may thus 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<sup>23</sup>.

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 decoy-receptor for treatment of COVID-19.

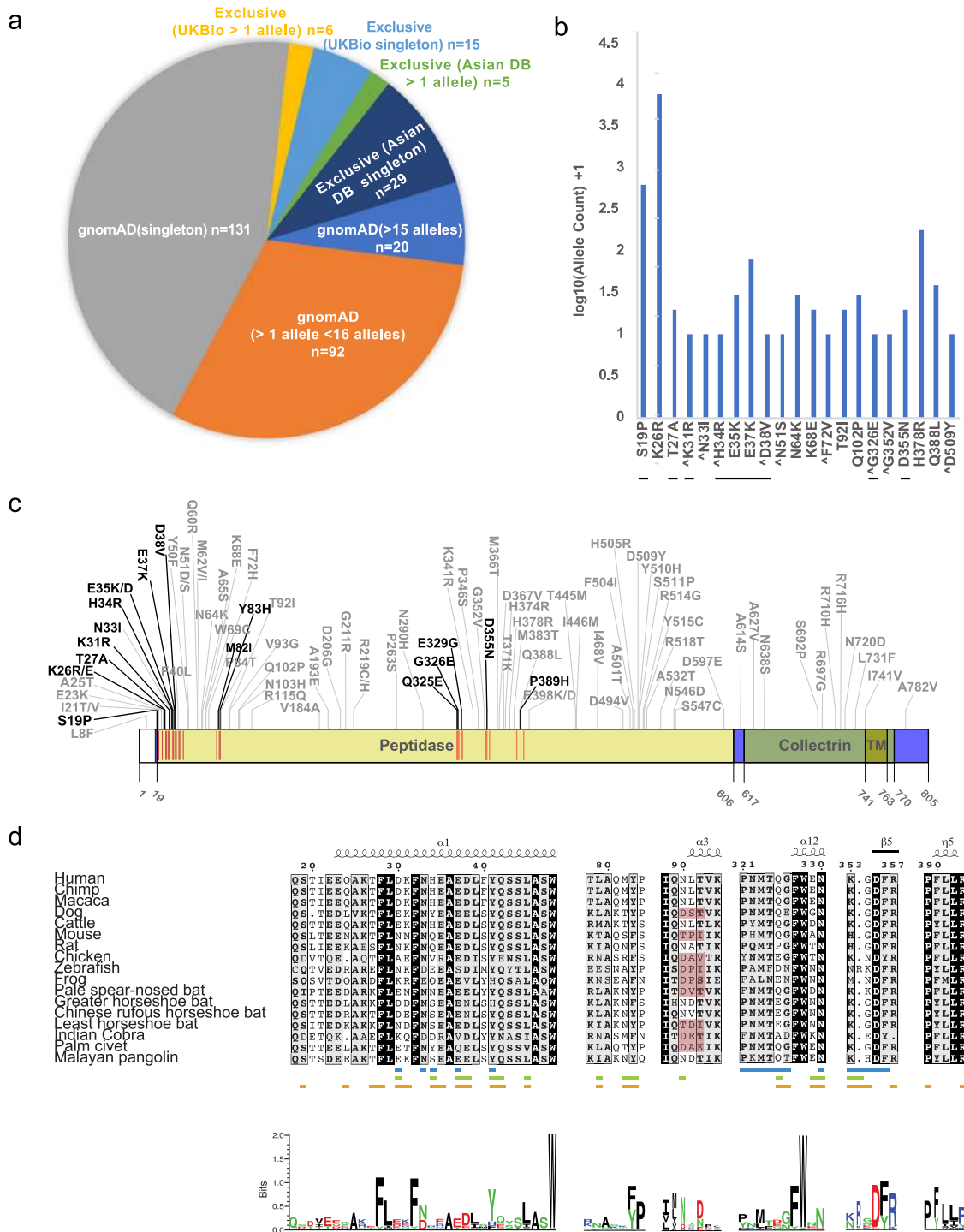
## Results

**Human ACE2 population polymorphism.** The SARS-CoV-2 S-protein interacts with the ACE2 PD to enter human host cells. Analysis of the RBD domain of SARS-CoV-2, SARS-CoV and bat CoV RaTG13 S-proteins identified changes that have increased the affinity of CoV-2 S1 RBD to human ACE2, which likely contributes to its increased infectivity<sup>20,22</sup>. It is very likely that there exists ACE2 variants 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 gnomAD<sup>24</sup>, RotterdamStudy<sup>25</sup>, ALSPAC<sup>26</sup> and Asian-specific databases, namely GenomeAsia100k<sup>27</sup>, TOMMO-3.5kjpnv2<sup>28</sup>, IndiGen (<https://indigen.igib.in/>), and HGDP<sup>29</sup> (Supplementary Data 1). We found a total of 298 unique protein altering variants across 256 codons distributed throughout the 805 amino acid long human ACE2 (Fig. 1a, b and Supplementary Fig. 1, Supplementary Data 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  $p$ -value = 0.0656,  $n = 65$ , Supplementary Fig. 2, see “Methods” section).

Further, genealogical estimation of variant age (GEVA) suggests that ACE2 coding variants are more recent (Supplementary Fig. 3). Although ACE2 has been reported to be highly intolerant of loss-of-function variants ( $pLI = 0.9977$ , gnomAD; Supplementary Fig. 2, see “Methods” section)<sup>24</sup>, we observed 5 predicted LOF singleton alleles (Supplementary Data 1).

Structural studies involving SARS-CoV and SARS-CoV-2 S-protein in complex with human ACE2 have identified three regions in a ~120 amino acid claw-like exposed outer surface of human ACE2 (ACE2-claw) that contributes to its binding to the S-protein<sup>19-22</sup>. The key residues at the ACE2 S-protein-RBD 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 (Fig. 1b). Mutagenesis of four residues, namely M82, Y83, P84 and K353, 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<sup>30</sup>. 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, N331, H34R, E35K, E37K, D38V, N51S, N64K, K68E, F72V, T92I, Q102P, G326E, G352V, D355N, H378R, Q388L, and D509Y (Figs. 1c, d, 2 and Supplementary Data 1). These variants could potentially 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.

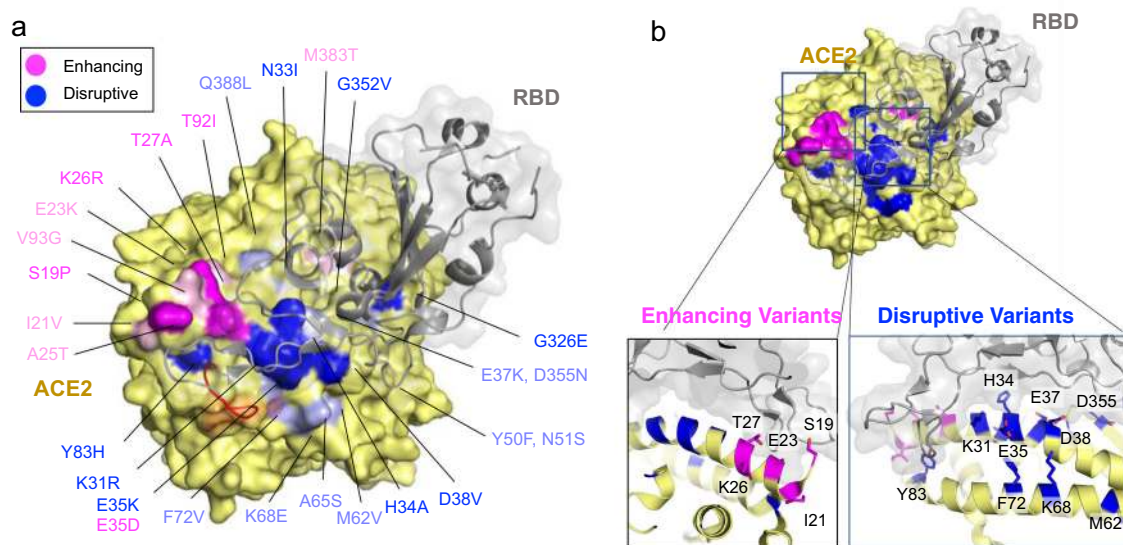
**Structural evaluation of ACE2 polymorphism.** To investigate the effect of the ACE2 polymorphisms on receptor recognition by the SARS-CoV-2 RBD, we modeled the identified ACE2 variants using published cryo-EM and crystal structures of ACE2/SARS-CoV-2 RBD complexes<sup>19-22</sup>. Based on the evaluation of the structures and a functional analysis of a synthetic human ACE2 mutant library for RBD-binding affinity<sup>23</sup>, we broadly classified



**Fig. 1** ACE2 polymorphisms. **a** Pie chart representing protein altering variations in ACE2 by allele count and source. **b** 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. **c** ACE2 protein domain showing positions with polymorphisms that can alter SARS-CoV-2 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 peptidase domain (PD) known to interact with viral S-protein are shown as red vertical lines within the peptidase domain in the ACE2 diagram. **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.

ACE2 polymorphic variants into two categories with respect to their predicted effect on ACE2-RBD-binding as enhancing or disrupting (Fig. 2 and Supplementary Data 2, 3). These two groups of polymorphic variants mapped onto the ACE2 structure

remarkably segregate into two distinct clusters at the ACE2/CoV-2 RBD interface (Fig. 2a). The predicted enhancing variants cluster to the ACE2 surface most proximal to the receptor-binding ridge of CoV-2 RBD (Fig. 2b) whereas the majority of the



**Fig. 2** Human ACE2 polymorphisms mapped to the structure of human ACE2 in complex with the SARS-CoV-2 RBD. Residues in ACE2 showing polymorphic variation in human populations were mapped on to the structure of the ACE2/SARS-CoV-2 RBD (PDB: 6VWV1) and colored according to their effect on the predicted affinity to SARS-CoV-2 RBD. Polymorphisms that were predicted to enhance the binding between ACE2 and the S-protein are colored in magenta. Polymorphisms that are predicted to disrupt the binding between ACE2 and the S-protein are colored in dark blue. The variable loop in the ridge binding motif consisting of residues V483 and E484 is shown in red. This region in the structure (PDB: 6LZG) is zoomed-in to show variants predicted to enhance or disrupt the ACE2 - SARS-CoV-2 interaction.

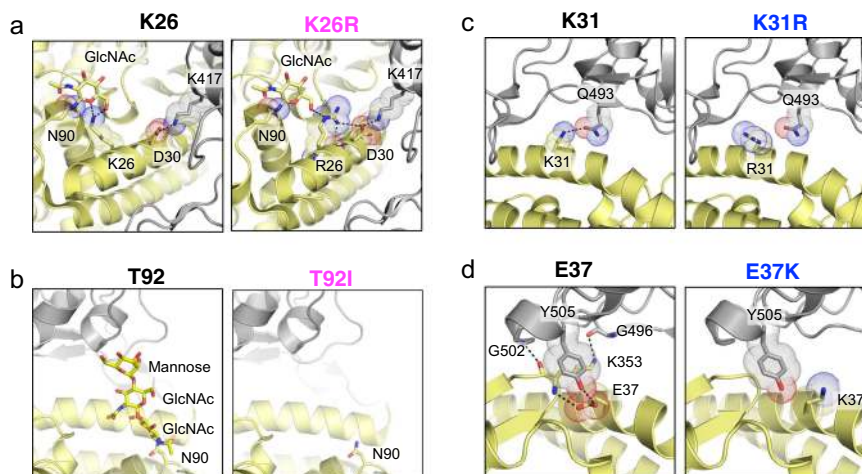
predicted disrupting variants reside centrally on the two major ACE2  $\alpha$ -helices that substantially contribute to the buried surface area at the interface (Fig. 2b).

The spatial segregation of the functionally different ACE2 variants can be structurally explained. The loop conformation in the receptor-binding ridge of SARS-CoV-2 differs from that of SARS-CoV owing to the presence of bulky residues (V483 and E484) in the loop<sup>22</sup>. This feature allows the SARS-CoV-2 loop to extend further towards ACE2 establishing more extensive contacts with the receptor (Fig. 2a). 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  $\alpha$ -helices ( $\alpha 1$  and  $\alpha 2$ ) and are not unique to CoV-2. They encompass what seem to be 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. Based on its structure, K26 establishes polar contacts with the first mannose moiety of the partially resolved ACE2 N90-linked glycan and likely stabilizes the position of the glycan relative to the native protein (Fig. 3a). 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<sup>31</sup>. We predict that K26R would abrogate stabilizing polar contacts with N90, impairing coordination of the glycan (Fig. 3a) and lead to an increase in the affinity of the virus to the ACE2 receptor. Further, 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 (Fig. 3a). In this scenario, the net effect of K26R polymorphism is the stabilization of core  $\alpha$ -helices that increases ACE2 binding affinity to CoV-2 RBD at the cost of glycan rigidity. Another variant likely to enhance affinity through a similar mechanism is T92I. T92I directly targets the glycosylation consensus sequence (<sub>90</sub>NXT<sub>92</sub>). This mutation would eliminate post-translational modification by

the N90-linked glycan providing an opportunity for a more optimal ACE2/RBD interaction (Fig. 3b). The predicted effect of the T27A variant is increased hydrophobicity at the interface which could contribute to an increase in binding affinity. Similar destabilizing patterns are predicted for the variants S19P and E23K (Supplementary Data 3).

The vast majority of predicted disruptive ACE2 polymorphic variants map to the core  $\alpha$ -helical bundle of ACE2 and to residues known to form contacts with the RBD. There are two key hotspot residues in the  $\alpha$ -helical bundle of the ACE2 interface that are important for CoV-2 RBD-binding: K31 and K353. 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<sup>22,32,33</sup>. A recent elegant study<sup>22</sup> 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. Interestingly, K31R is one of the human ACE2 polymorphisms that we identified. 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 (Fig. 3c and Supplementary Fig. 4a) and destabilize the charge-neutralizing interaction with the virus. Similarly, E37, also located in the core  $\alpha$ -helical bundle coordinates polar contacts with RBD Y505, RBD G502, and ACE2 K353 (Fig. 3d). The E37K variant we identified is predicted to disrupt these critical interactions by removing the polar intramolecular interaction with ACE2 K353 and stabilizing contacts with the RBD (Fig. 3d). Thus, individuals carrying K31R or E37K ACE2 variants are predicted to be less susceptible to SARS-CoV-2 infection. While we did not identify any polymorphic variants at residue K353, we detected an ACE2 mutation at D38 (Supplementary Data 1), which forms an electrostatic interaction with K353 (Supplementary Data 3). This mutation, D38V, would compromise the neutralizing effect of the K353-D38 interaction at the interface and is predicted to reduce binding affinity between the virus and the host receptor.



**Fig. 3** Structural basis of interaction between human ACE2 polymorphic variants and SARS-CoV-2 S-protein. **a** K26R, **b** T92I, **c** K31R and **d** E37K. Interaction mapping done using PDB structure 6VW1.

Another recurrent polymorphism in ACE2 maps to residue E35 where the glutamate is replaced by a lysine (E35K; Supplementary Data 2). 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 (Supplementary Fig. 4b). Interestingly, the ACE2 E35 interaction is not conserved between SARS-CoV and SARS-CoV-2 S-proteins<sup>22</sup> and hence we predict that E35K could offer selective protection from the SARS-CoV-2 infection in individuals carrying this variant in a manner akin to H34R (Supplementary Fig. 4c) which similarly also results in a loss of interface polar contacts. Another interesting polymorphism at position 83 results in a Y83H alteration. Notably, in mouse, residue 83 is phenylalanine (Fig. 1d)<sup>34</sup>. Residue F/Y83 underlies a hydrophobic pocket into which F486 from SARS-CoV-2 RBD is inserted (Supplementary Fig. 4d). 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<sup>22</sup>. 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 Fig. 4d), potentially offering selective protection from SARS-CoV-2 infection.

#### Altered affinity of ACE2 variants for SARS-CoV-2 S-protein.

To validate our structural predictions, we measured the effect of select ACE2 polymorphisms on its binding affinity to CoV-2 S-protein. We expressed and purified the S1 subunit of the S-protein, CoV-2 S-RBD, and a trimer stabilized form of S-protein (S-trimer; Supplementary Fig. 5). We also recombinantly produced His-tagged monomeric and Fc-tagged dimeric forms of the extracellular domain of wildtype ACE2 (WT) and variant forms of ACE2 (S19P, K26R, K31R, E37K and T92I; Supplementary Fig. 5). These variants were selected based on their population frequency and the predicted effect on their interaction with S-protein.

We tested the affinity of these ACE2 variants against a panel of S-protein constructs using enzyme-linked immunosorbent assay (ELISA) and/or bio-layer interferometry (BLI; Fig. 4 and Table 1). For ELISAs, we used dimeric ACE2-Fc to assess its binding to the S-protein variants. We found the ACE2-Fc WT dimer bound to the isolated S-RBD (EC50 1.01 nM) and S-trimer (EC50 0.95 nM) more strongly compared to the S1 subunit (EC50 10.4 nM) (Fig. 4a–c and Table 1). This is consistent with previous studies that showed a decreased ACE2 affinity for SARS-CoV S1 subunit

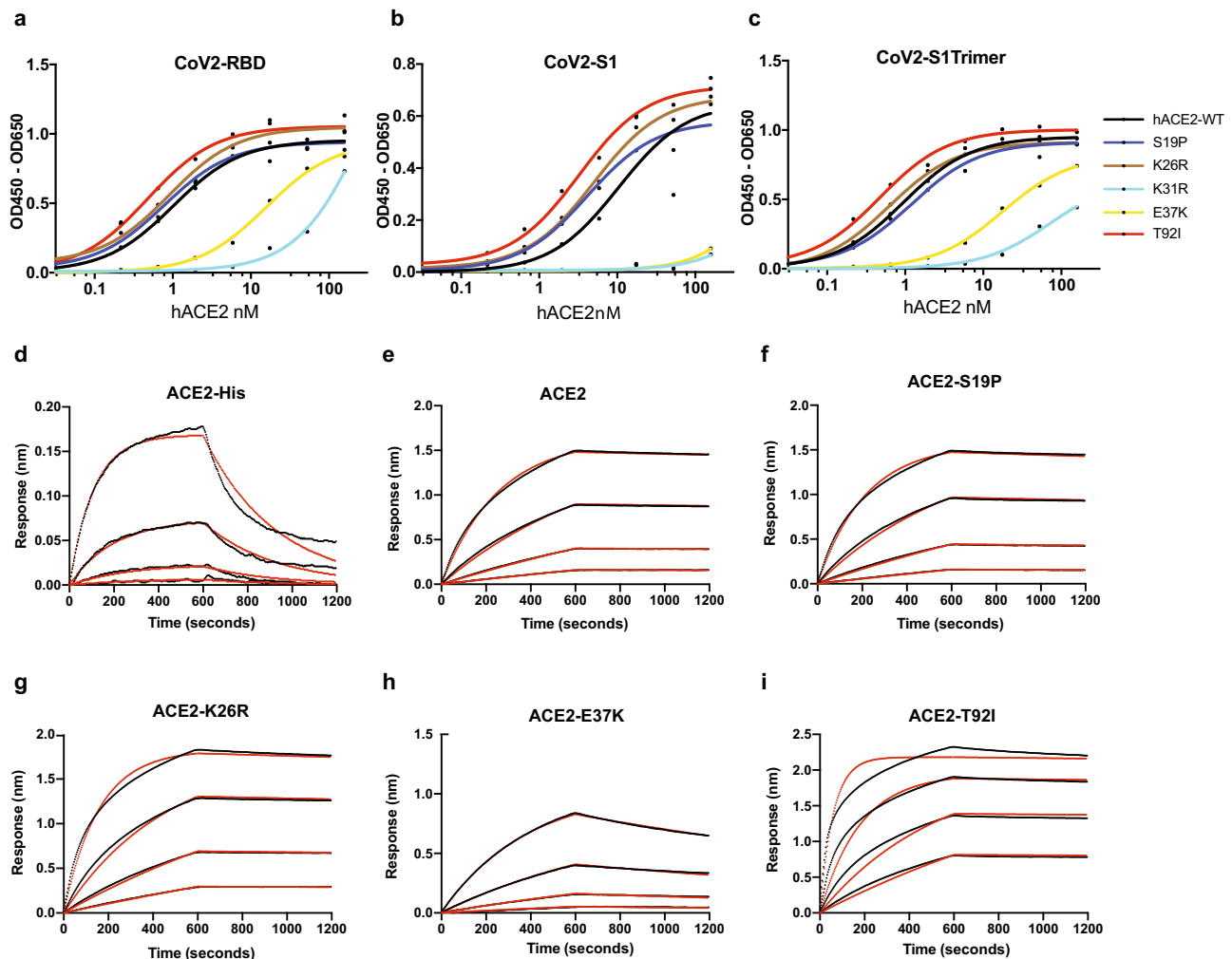
compared to S-RBD, indicating a conformational difference between these variants<sup>12,16,35</sup>. In the trimeric state, in contrast to the monomeric full-length S1-protein, the RBD within the S1 subunit in one or more of the constituent S-proteins is known to adopt a receptor-accessible “RBD-out” conformation, supporting its high affinity for ACE2 that is comparable to that observed for isolated RBD<sup>20,21,36</sup>.

Using BLI, we also tested the interaction between monomeric ACE2 WT or dimeric ACE2-Fc WT and S-RBD and found that ACE2-Fc ( $K_d = 1.21$  nM) bound more tightly to S-RBD compared to monomeric ACE2 ( $K_d = 117$  nM) (Table 1). Recent structural studies show ACE2 interacts with the S-protein as a dimer and hence the observed higher affinity of the ACE2 dimer is likely critical for a productive infection<sup>21,37</sup>.

Next, using ELISA and/or BLI assays, we measured the affinity of the S-RBD or S-trimer for ACE2-Fc variants (Fig. 4 and Table 1). While the S19P ACE2-Fc mutant had marginally increased affinity compared to WT ACE2-Fc ( $K_d$  1.21 vs 1.05 nM), ACE2 K26R mutant showed higher affinity for S-RBD ( $K_d = 0.36$  nM) and S-Trimer (EC50 0.62 nM) (Table 1). The 4- to 5-fold increased affinity of K26R mutation for S-RBD relative to ACE2 WT was in part due to a >2-fold enhancement in its on-rate ( $K_{on}$  0.76 vs  $0.31 \times 10^5$   $M^{-1} s^{-1}$ ) combined with a ~25% reduction in its off-rate (Table 1). Similarly, the T92I, glycosylation site mutant of ACE2-Fc, showed an increased affinity for S-RBD ( $K_d = 0.13$  nM) and S-trimer (EC50 0.47 nM). The pronounced ~9-fold increase in affinity of ACE2 T92I versus ACE2 WT ( $K_d$  0.13 vs 1.21 nM) is consistent with a ~9-fold increase in its on-rate ( $K_{on}$  2.8 vs  $0.31 \times 10^5$   $M^{-1} s^{-1}$ ). In contrast, E37K mutation led to a loss of ACE2 affinity for S-RBD ( $K_d = 22.7$  nM) and S-trimer (EC50 = 17.6 nM). The ~18-fold decrease in affinity of E37K compared to the WT ACE2 is in part due to a ~10-fold increase in its off-rate ( $K_{off}$  41 vs  $3.74 \times 10^{-5}$   $s^{-1}$ ). As, observed with E37K, K31R ACE2-Fc had a decreased affinity for S-RBD (EC50 = 298 nM) and S-trimer (EC50 = 73 nM) when compared to WT ACE2-Fc.

#### Affinity-enhancing ACE2 variants are more effective in virus neutralization.

We tested the ability of Fc-tagged ACE2 WT, and its variant forms, for their ability to block S-protein pseudotyped virus from entering infection-competent cells (Fig. 5 and Table 1). Among the variants tested, ACE2 K26R, consistent with its increased affinity for S-RBD observed in the biochemical assay, was the most effective ( $K_i$  30 nM; ~33 fold greater than WT) in blocking viral entry (Fig. 5 and Table 1). Similarly, ACE2 T92I



**Fig. 4** Binding affinity of SARS-CoV-2 S-RBD, S1 and S-trimer. **a–c** ELISA assay measuring the affinity of indicated ACE2 WT or variants for SARS-CoV-2 S-RBD (**a**), S1 (**b**) and S-trimer (**c**). **d–i** Sensorgrams for the binding of a 3-fold dilution series of monomeric and dimeric ACE2 variants to streptavidin-captured SARS-CoV-2 RBD are provided (black). Fits to the empirical binding curves are shown overlaid (red).

variant was ~20-fold more potent in blocking the pseudovirus compared to ACE2 WT ( $K_i$  53 vs >1000 nM; Fig. 5 and Table 1). Interestingly, though ACE2 S19P showed only a minor increase in affinity for S-RBD in the biochemical assay, it was ~11 fold more effective ( $K_i$  > 92 nM) in blocking viral entry, suggesting that a subtle increase in affinity might be sufficient to alter susceptibility.

## Discussion

The host-virus evolutionary arms race over time leads to natural selection that alters the host and the viral proteins allowing both to increase their fitness<sup>38</sup>. In this context, multiple studies have analyzed and identified the origin, evolution and successful adaptation of the SARS coronaviruses as human pathogens<sup>9,39</sup>. Viral genome sequencing and analysis has identified bats to be the most likely natural host of origin for both SARS-CoV and the recent SARS-CoV-2<sup>9</sup>. 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<sup>19–22,39,40</sup>. These studies have identified S-protein changes that have rendered the human cells permissive to the SARS-CoV and SARS-CoV-2 infection<sup>19–22,40</sup>.

Thus far, the role of variations in human ACE2 receptor in susceptibility to both SARS CoVs had not been comprehensively

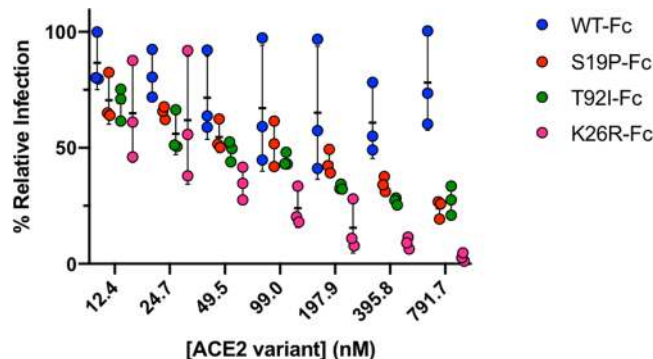
examined. While a recent in silico study analyzed limited ACE2 population variation data set and concluded that these polymorphisms did not confer resistance to the virus<sup>34</sup>, other studies have implicated ACE2 variants in altering binding to S-protein<sup>41–45</sup>. In this study, we comprehensively 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 published protein structures and data from a high-throughput functional mutagenesis screen that used deep sequencing to assess enrichment or depletion of S-protein binding to ACE2 variants (Supplementary Fig. 6), we performed structural modeling to classify ACE2 variants identified in this study based on their effects on susceptibility to SARS-CoV<sup>19–23</sup>.

We identified several ACE2 polymorphic variants that increase ACE2/S-protein interaction including S19P, I21V, E23K, K26R, K26E, T27A, N64K, T92I, Q102P, M383T and H378R (Supplementary Data 1, 2 and 3). Among these, the T92I polymorphism is part of a NxT/S consensus N-glycosylation motif<sup>46</sup>. The ACE2 NxT/S motif, while conserved in 96 out of 296 jawed vertebrates, it is absent or altered in several species, including the civet cat (*Paguma larvata*). The NxT/S motif is altered in several bat species and this includes substitution at N90, presence of a proline at position 91 or any amino acid except serine at T92, any of which will abolish the glycosylation at N90 (Fig. 1d and

**Table 1 SARS-CoV-2 S-protein affinity for ACE2 variants.**

hACE2 variant	ELISA		BLI		PVNA	
	EC50, nM [ACE2-Fc]	S1	K <sub>d</sub> (nM) [S-RBD]	RBD vs ACE2-Fc	K <sub>off</sub> (10 <sup>5</sup> s <sup>-1</sup> )	K <sub>i</sub> (nM)
	S-RBD	S-trimer	RBD vs ACE2-Fc	RBD vs ACE2-His	K <sub>on</sub> (10 <sup>5</sup> M <sup>-1</sup> s <sup>-1</sup> )	ACE2-Fc
WT	1.01 ± 0.04	0.95 ± 0.03	1.21 ± 0.2	117 ± 11	0.26 ± 0.1	>1000
S19P	0.77 ± 0.08	1.20 ± 0.03	1.05 ± 0.1	80 ± 3	0.46 ± 0.1	92 ± 30
K26R	0.89 ± 0.11	0.62 ± 0.06	0.36 ± 0.1	114 ± 8	0.27 ± 0.1	30 ± 13
K31R	2.98 ± 0.64	73 ± 0.07	ND	ND	ND	ND
E37K	15.8 ± 0.03	17.6 ± 0.02	22.7 ± 0.4	NB	NB	NB
T92I	0.48 ± 0.03	0.47 ± 0.04	0.13 ± 0.1	87 ± 6	0.37 ± 0.1	53 ± 14

BLI bio-layer interferometry, PVNA Pseudovirus neutralization assay, NB no binding, ND not determined (for K31R, BLI values could not be determined given the low affinity observed in ELISA).



**Fig. 5 ACE2 variants block pseudotyped SARS-CoV-2 virus infection.** Inhibition of pseudovirus entry into HEK293T cells by ACE2 WT and mutants S19P, T92I, and K26R. A range of seven different concentrations of Fc-tagged ACE2 proteins were mixed with pseudoviruses and the infectivity was measured by the luciferase signal as an indication of the amount of virus entering the cells. Data are presented as the mean ± SD (standard deviation). Sample size *n* = 3.

Supplementary Fig. 7, Supplementary Data 4)<sup>30,31,46,47</sup>. These ACE2 variations are expected to abolish glycosylation at N90<sup>46</sup>. Another mutation that altered the NxT/S motif in human ACE2 to a civet ACE2-like sequence (90-NLTV-93 to DAKI), also expected to abolish the N-glycosylation, was shown to increase the SARS-CoV infectivity and S-protein binding (Fig. 1d)<sup>30</sup>. A previous study showed that the ACE2 N90 renders human cells resistant to civet CoV<sup>30</sup>.

Our structural investigation suggests that the T92I mutation favors improved Cov-2 S-protein binding. Using recombinant T92I mutant ACE2 protein, we showed that it had an increased affinity for S-RBD and also found it to be more effective in blocking virus entry compared to ACE2 WT (Fig. 5 and Table 1). Further, the T92I mutant showed a strong enrichment in a sequencing-based screen for S-protein binders<sup>23</sup>. Thus, the T92I mutation likely renders individuals harboring this mutation more susceptible to the virus. Taken together, these observations suggest that N90 glycosylation site is critical and it could confer protection through glycan shielding. ACE2 N90 glycosylation could also determine the strength and specificity of infection by different CoV viruses.

We also show that another ACE2 residue, K26, plays an important role in controlling the susceptibility to viral infections via a similar mechanism. Our structural analysis has suggested that K26R mutation will weaken coordination of the N90-linked glycan presumably interfering with its ability to shield the host from the viral infection. Our biochemical binding assays confirmed the predicted increased affinity for K26R ACE2 for S-protein. In fact, K26R ACE2 was the most effective among mutants tested for their ability to enhance viral entry, suggesting that this frequent polymorphism, very likely increases susceptibility to SARS-CoV-2 (Fig. 5 and Table 1).

We also found ACE2 variants predicted to reduce ACE2 S-protein interactions and thereby decrease S/ACE2 binding affinity. Variants predicted to reduce ACE2 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 (Supplementary Data 1 and 2). Biochemical binding assays confirmed the decreased affinity of two variants that we tested, K31R and E37K, indicating that these likely are protective polymorphisms.

Overall, we find the ACE2 population variants, that either increase or decrease susceptibility, to be rare, which is consistent with the overall low number of ACE2 receptor population level

polymorphisms (mean  $F_{st}$  0.0167). Also, we did not observe any statistically significant difference 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 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. *ACE2* K26R, predicted to increase susceptibility to SARS-CoV-2, is found in 8 women and 6 men in the UK Biobank exome sequencing data set. Two of the 6 men tested positive for SARS-CoV-2 infection, representing a (non-significant) 2.4-fold increased odds of infection compared to those who do not carry the variants (Fisher's exact  $p$ -value = 0.279). No other variants with predicted binding affinity were found in the UK Biobank participants with both exome sequencing data and COVID-19 test results.

Genetic variation in *ACE2* alone is unlikely to explain the vast variability in infection susceptibility and severity of COVID-19. While a handful of large genome-wide association studies (GWAS) of SARS-CoV-2 infection status have identified additional genetic risk factors<sup>48,49</sup>, the *ACE2* locus shows only weak association in these studies, possibly due to the lack of common variation in the locus. The extremes in COVID-19 clinical symptoms reported range from asymptomatic infected adult individuals to those that show acute respiratory syndrome leading to death<sup>50–52</sup>. This suggests a role for additional factors, including the role of innate and adaptive immunity, besides variation in *ACE2* in modifying disease outcomes.

Currently, there are no approved targeted therapeutics for curing SARS-CoV-2 infection. Therefore, development of therapeutics to treat patients and mitigate the COVID-19 pandemic is urgently needed<sup>52,53</sup>. Several small molecules and neutralizing antibodies for treatment are in development<sup>54,55</sup>. Soluble *ACE2* and *ACE2*-Fc fusion protein have been proposed as decoy SARS-CoV-2 receptor therapeutic<sup>56–58</sup>. Soluble *ACE2*, as a therapy for pulmonary arterial hypertension, has been shown to be safe in early human clinical studies<sup>59,60</sup>. 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. Even though a human recombinant soluble *ACE2* is in clinical trials to treat SARS-CoV-2 infection<sup>61</sup>, a catalytically inactive soluble *ACE2* might be preferred from a safety perspective, as S-protein binding enhances *ACE2*'s carboxypeptidase activity<sup>62</sup>. Additionally, as *ACE2* enzymatic activity modulates multiple biological pathways<sup>63</sup>, a catalytically inactive form should be considered for treating SARS-CoV-2 infection. 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.

Understanding the natural *ACE2* polymorphism spectrum not only provides information on the SARS-CoV-2 susceptibility but can also be used to generate high affinity, rationally designed soluble *ACE2* receptor molecules. Such agents that carry naturally occurring polymorphism(s) will lead to no or low immunogenicity in a drug setting and can be used as a decoy-receptor for treating patients.

## Methods

**Identification of *ACE2* polymorphisms.** We queried multiple genomic databases including gnomAD<sup>24</sup> (<https://gnomad.broadinstitute.org/>), DiverEHR<sup>64</sup>,

RotterdamStudy<sup>25</sup>, ALSPAC<sup>26</sup> and Asian-specific databases which included GenomeAsia100k<sup>27</sup>, HGDP<sup>29</sup>, TOMMO-3.5kjpnv2<sup>28</sup> 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.

**F<sub>st</sub> analysis.** To assess genetic variation in the coding region of *ACE2*, we calculated the fixation index ( $F_{st}$ ) from 2381 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  $F_{st}$  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)<sup>65</sup> and reported the weighted mean  $F_{st}$  as described in Bhatia et al. (2013)<sup>66</sup>; 277 variants were used. Because  $F_{st}$  values vary based on variants used (Bhatia et al. 2013<sup>66</sup>), we calculated the  $F_{st}$  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.

**Genealogical estimation of variant age (GEVA).** We used data from the 1000 Genomes Project<sup>67</sup> to estimate the time of mutation of all variants located within a 1 Mb region around the *ACE2* gene on Chromosome X, from the female-only subset of 1271 individuals (Supplementary Fig. 3a). As previously described<sup>68</sup>, we performed the analysis using an effective population size of  $N_e = 10,000$ , mutation rate  $\mu = 1.2 \times 10^{-8}$ , and with variable recombination rates according to HapMap2<sup>69</sup>. We used the most recent version of GEVA software (<https://github.com/pkalbers/geva/tree/ancallele>), which allowed us to provide external information about predicted ancestral and derived allelic states from Ensembl (release 95) to correct model assumptions for all variants on Chromosome X. Variant age is estimated through pairwise analyses between haplotype sequences which may or may not carry the derived allele at a given variant. We analyzed each variant using a maximum of 5000 concordant pairs (carrier and carrier haplotypes) and 5000 discordant pairs (carrier and non-carrier) to achieve high confidence. We further distinguished variants into non-coding, synonymous, and missense variants using the Ensembl Variant Effect predictor (release 95)<sup>70</sup> and separated variants affecting *ACE2* ( $n = 385$ ) from those outside the *ACE2* gene region ( $n = 9095$ ). The proportion of rare variants ( $\leq 0.1\%$  frequency of the derived allele within the sample) was similar in both groups; 19% and 22%, respectively (Supplementary Fig. 3b). For variants outside the *ACE2* gene, we found that 54% of non-coding variants were estimated to have arisen within the last 1000 generations, compared to 75% of synonymous and 80% of missense variants (Supplementary Fig. 3c). This suggests that past selective pressure may have acted more strongly to prune mutations that occur within the coding region of the genome. We found that this signal was more pronounced for missense mutations affecting *ACE2*, where we found 58% of non-coding and 60% of synonymous variants to be younger than 1000 generations, whereas all missense variants were younger than approximately 800 generations. The average age ( $\pm$ SE) of missense variants affecting *ACE2* was 472 ( $\pm 58$ ) generations, compared to 3016 ( $\pm 2198$ ) generations for variants outside the *ACE2* gene region. However, the low number of coding variants found within the focal 1 MB region for which we were able to estimate the age ( $n = 43$  missense and  $n = 37$  synonymous variants) makes such comparisons difficult.

***ACE2* ortholog sequence analysis.** A total of 295 Human *ACE2* orthologs were obtained from NCBI (Supplementary Data 4 for accession numbers). A snake *ACE2* ortholog protein was obtained from the published Indian cobra genome<sup>71</sup>. Multiple sequence alignment of residues surrounding the *ACE2* NxT/S motif was performed using MCooffee ([www.tcoffee.org](http://www.tcoffee.org)). Phylogenetic trees were constructed using the PhyML webserver ([www.phylogeny.fr](http://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).

**Cloning and protein expression.** Extracellular domain (amino acids 1–615; NP\_001358344) of human *ACE2* (hACE2) WT or variants with a c-terminal 8x-His or human-Fc tag was synthesized (IDT, USA) and cloned into a CMV promoter-driven mammalian expression vector. Human codon optimized CoV-2-S-RBD (amino acids 319–541; YP\_009724390) sequence with a c-terminal 8x His-tag were synthesized and cloned into a CMV promoter-driven mammalian expression vector. The prefusion SARS-CoV-2 S-protein trimer stabilized ectodomain (amino acids 1–1208; YP\_009724390), as previously described<sup>20</sup>, containing K986P, V987P, RRAR to GSAS (residues 682–685) at the furin cleavage site, a C-terminal T4 fibrin trimerization motif, an HRV3C protease cleavage site,



a TwinStrep-tag and a 8x Hi-tag was synthesized and expressed using a CMV promoter. Sequence verified plasmids prepared using NucleoBond® Xtra Midi kit (Takara Bio USA, Inc) were transfected into 293 cells using FectoPro (Polyplus, USA). Proteins were purified from media 3–5 days post transfection using Protein A GraviTrap column or His GraviTrap column (GE Healthcare).

**ELISA affinity studies.** The affinity of S-RBD or S1 for hACE2-Fc WT or variants was measured using a standard ELISA assay. Briefly, purified CoV2-S-RBD (2 µg/mL) or S1 (2 µg/mL) or the prefusion S-protein trimer (2 µg/mL) was coated onto 96-well ELISA plates (Fisher Scientific, #07-000-102) and incubated at 4 °C for 18 h. The coated plates were washed three times with 200 µL of PBST and then blocked with 200 µL of 3% BSA (Sigma-Aldrich # A8327) in PBST (Sigma-Millipore # 524653) and incubated for 1 h at room temperature. After washing the plates three times with 200 µL of PBST an increasing concentration of hACE2-Fc proteins were added and incubated for 1 h at room temperature. The unbound hACE2-Fc was removed by washing the plate three times with 200 µL of PBST. The bound hACE2 was detected using Goat-anti-human-IgG-Fc-HRP (Jackson Immuno Research # 109-035-008; 1:5000 dilution) using 50 µL TMB substrate (Pierce/ThermoFisher Scientific # 34028). After 3 min, the reaction was stopped using 50 µL of 2 N H<sub>2</sub>SO<sub>4</sub>. The optical density of the reaction was measured at 450 nm using a plate reader (Molecular Devices Gemini XPS). The data was analyzed and EC50 was calculate using Prism (GraphPad).

**Biolayer interferometry.** To determine the binding kinetics and estimate the affinity ( $K_d$ ) of ACE2 variants for SARS-CoV-2 S-RBD, biolayer interferometry (BLI) experiments were performed on an Octet HTX instrument (ForteBio, USA) at 1000 rpm and 25 °C<sup>72</sup>. Biotinylated S-RBD protein was first captured on streptavidin biosensors from a 2 µg/mL solution to achieve binding response of 0.4–0.6 nm, followed by a quench step of 180 s with 100 µg/mL biotin. ACE2 variant proteins were diluted with assay buffer (PBS, 1% BSA, 0.05% Tween 20) and 200 nM of an unrelated biotinylated protein of similar size was used as negative control. After equilibrating with assay buffer, the loaded biosensors were dipped for 600 s into wells containing 3-fold serial dilutions of each variant from 200 nM and subsequently were transferred for 600 s back into assay buffer. Binding response data were reference subtracted and were fitted with 1:1 binding model using ForteBio's Data Analysis software 9.0.

**Generation of SARS-CoV-2 pseudovirus.** To generate SARS-CoV-2 pseudovirus, Human embryonic kidney 293 (HEK 293, ATCC® CRL-1573™) cells were seeded at  $0.3 \times 10^6$  cells/well, in DMEM (ThermoFisher Scientific, Cat. No. 11995073) supplemented with 10% FBS and 1% Penicillin-Streptomycin (Gibco, Cat. No. 15070063) and grown overnight at 37 °C, 5% CO<sub>2</sub>. HEK 293 cells were then co-transfected with 1 µg of pNL4-3.luc.R-E- plasmid (luciferase expressing HIV-1 with defective envelop protein) (NIH AIDS Reagent Program, Cat. No. 3418) and 0.06 µg of CMV promoter-driven plasmid encoding SARS-CoV-2 wild-type or mutant spike variants using Lipofectamin™ 2000 transfection reagent (ThermoFisher Scientific, Cat. No. 11668027)<sup>72</sup>. Pseudoviruses harvested from the supernatant at 48 h post-transfection were filter sterilized (0.44 µm, Millipore Sigma, Cat. No. SLHA033SS) and used.

**Pseudovirus entry assay.** HEK 293 cells (ATCC® CRL-1573) stably over-expressing full-length human ACE2 protein were seeded in 96-well white polystyrene microplates (Corning, Cat. No. CLS3610) at  $0.03 \times 10^6$  cells/well in DMEM (10% FBS and 1% Pen-Strep), and grown overnight at 37 °C, 5% CO<sub>2</sub>. To test the inhibition of pseudovirus entry by ACE2 WT or mutants, increasing concentration of Fc-tagged ACE2 proteins were first mixed with pseudoviruses and incubated at room temperature for 10 m<sup>72</sup>. The ACE2 WT or mutant treated pseudovirus mixture was then used to infect cells. The cells were incubated at 37 °C, 5% CO<sub>2</sub> for 6 h, then the medium was replaced with fresh DMEM (10% FBS and 1% Pen-Strep), and then again every 24 h for up to 72 h. To measure the luciferase signal (a proxy for virus uptake), DMEM was removed and cells were replaced in DPBS (ThermoFisher, Cat. No. 14190250) and mixed with an equal volume of ONE-Glo™ EX Luciferase Assay System (Promega, Cat. No. 8130). Relative luciferase units were measured using a BioTek Synergy Neo plate reader (BioTek Instruments Inc.). The data was then analyzed using GraphPad Prism Version 8.4.3 (GraphPad Software, LLC).

**Statistics and reproducibility.** Data were presented as the mean ± SD (standard deviation) of all experiments with  $N$  = number of biological replicates. Data were evaluated with the Wilcoxon test and Fisher's exact test for statistical significance. Means were judged as statistically insignificant when  $p > 0.05$ . Box plots represent the median (center line), the interquartile range (IQR; box limits) and  $1.5 \times$  IQR for the whiskers.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All ACE2 variant data used in this study were obtained from publicly available data sources as described in the methods. The ACE2 variant data is also presented in Supplementary Data 1.

Received: 6 November 2020; Accepted: 23 March 2021;

Published online: 12 April 2021

## References

- Li, F. Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* **3**, 237–261 (2016).
- Graham, R. L. & Baric, R. S. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J. Virol.* **84**, 3134–3146 (2010).
- Holmes, K. V. SARS-associated coronavirus. *N. Engl. J. Med.* **348**, 1948–1951 (2003).
- Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367**, 1814–1820 (2012).
- Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
- Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
- Chan, J. F. et al. 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 (2020).
- JHU. <https://coronavirus.jhu.edu/> (2020).
- Guo, Y. R. et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. *Mil. Med. Res.* **7**, 11 (2020).
- Lin, H. X. et al. 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 (2008).
- 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. *J. Virol.* **94**, <https://doi.org/10.1128/JVI.00127-20> (2020).
- Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* <https://doi.org/10.1016/j.cell.2020.02.052> (2020).
- Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
- Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020).
- Ou, X. et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* **11**, 1620 (2020).
- Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
- Simmons, G. et al. Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc. Natl Acad. Sci. USA* **102**, 11876–11881 (2005).
- Belouzard, S., Chu, V. C. & Whittaker, G. R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl Acad. Sci. USA* **106**, 5871–5876 (2009).
- Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* <https://doi.org/10.1016/j.cell.2020.1002.1058> (2020).
- Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
- Yan, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
- Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* <https://doi.org/10.1038/s41586-020-2179-y> (2020).
- Chan, K. K. et al. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* <https://doi.org/10.1126/science.abc0870> (2020).
- Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
- Ikram, M. A. et al. The Rotterdam Study: 2018 update on objectives, design and main results. *Eur. J. Epidemiol.* **32**, 807–850 (2017).

26. Fraser, A. et al. Cohort profile: the avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
27. GenomeAsia, K. C. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature* **576**, 106–111 (2019).
28. Tadaka, S. et al. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. *Hum. Genome Var.* **6**, 28 (2019).
29. Bergstrom, A. et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, <https://doi.org/10.1126/science.aay5012> (2020).
30. Li, W. et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* **24**, 1634–1643 (2005).
31. Demogines, A., Farzan, M. & Sawyer, S. L. Evidence for ACE2-utilizing coronaviruses (CoVs) related to severe acute respiratory syndrome CoV in bats. *J. Virol.* **86**, 6350–6353 (2012).
32. Li, F. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. *J. Virol.* **82**, 6984–6991 (2008).
33. Wu, K., Peng, G., Wilken, M., Geraghty, R. J. & Li, F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J. Biol. Chem.* **287**, 8904–8911 (2012).
34. 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 (2020).
35. Wong, S. K., Li, W., Moore, M. J., Choe, H. & Farzan, M. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J. Biol. Chem.* **279**, 3197–3201 (2004).
36. Walls, A. C. et al. Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* **176**, 1026–1039.e1015 (2019).
37. Lui, I. et al. Trimeric SARS-CoV-2 Spike interacts with dimeric ACE2 with limited intra-Spike avidity. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.05.21.109157> (2020).
38. Daugherty, M. D. & Malik, H. S. Rules of engagement: molecular insights from host-virus arms races. *Annu. Rev. Genet.* **46**, 677–700 (2012).
39. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0820-9> (2020).
40. Chen, Y., Guo, Y., Pan, Y. & Zhao, Z. J. Structure analysis of the receptor binding of 2019-nCoV. *Biochem. Biophys. Res. Commun.* <https://doi.org/10.1016/j.bbrc.2020.02.071> (2020).
41. Cirulli, E. T., Riffle, S., Bolze, A. & Washington, N. L. Revealing variants in SARS-CoV-2 interaction domain of ACE2 and loss of function intolerance through analysis of >200,000 exomes. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.04.07.030544> (2020).
42. Hussain, M. et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *J. Med. Virol.* <https://doi.org/10.1002/jmv.25832> (2020).
43. Hou, Y. et al. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. *BMC Med.* **18**, 216 (2020).
44. Devaux, C. A., Rolain, J. M. & Raoult, D. ACE2 receptor polymorphism: susceptibility to SARS-CoV-2, hypertension, multi-organ failure, and COVID-19 disease outcome. *J. Microbiol. Immunol. Infect.* **53**, 425–435 (2020).
45. Benetti, E. et al. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. *Eur. J. Hum. Genet.* <https://doi.org/10.1038/s41431-020-0691-z> (2020).
46. Gavel, Y. & von Heijne, G. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng.* **3**, 433–442 (1990).
47. Damas, J. et al. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc. Natl Acad. Sci. USA* **117**, 22311–22322 (2020).
48. Ellinghaus, D. et al. Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa2020283> (2020).
49. Kachuri, L. et al. The landscape of host genetic factors involved in immune response to common viral infections. *Genome Med.* **12**, 93 (2020).
50. Cao, Y., Liu, X., Xiong, L. & Cai, K. Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2: a systematic review and meta-analysis. *J. Med. Virol.* <https://doi.org/10.1002/jmv.25822> (2020).
51. Yuen, K. S., Ye, Z. W., Fung, S. Y., Chan, C. P. & Jin, D. Y. SARS-CoV-2 and COVID-19: The most important research questions. *Cell Biosci.* **10**, 40 (2020).
52. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C. & Di Napoli, R. in *StatPearls* (2020).
53. Jiang, S. Don't rush to deploy COVID-19 vaccines and drugs without sufficient safety guarantees. *Nature* **579**, 321 (2020).
54. Zhou, Y. et al. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* **6**, 14 (2020).
55. Li, G. & De Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nat. Rev. Drug Discov.* **19**, 149–150 (2020).
56. Hofmann, H. et al. 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 (2004).
57. Lei, C. et al. Potent neutralization of 2019 novel coronavirus by recombinant ACE2-Ig. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.02.01.929976> (2020).
58. Kruse, R. L. Therapeutic strategies in an outbreak scenario to treat the novel coronavirus originating in Wuhan, China. *F1000Res* **9**, 72 (2020).
59. Haschke, M. et al. Pharmacokinetics and pharmacodynamics of recombinant human angiotensin-converting enzyme 2 in healthy human subjects. *Clin. Pharmacokinet.* **52**, 783–792 (2013).
60. Guignabert, C., de Man, F. & Lombes, M. ACE2 as therapy for pulmonary arterial hypertension: the good outweighs the bad. *Eur. Respir. J.* **51**, <https://doi.org/10.1183/13993003.00848-2018> (2018).
61. Zoufaly, A. et al. Human recombinant soluble ACE2 in severe COVID-19. *Lancet Respir. Med.* **8**, 1154–1158 (2020).
62. Lu, J. & Sun, P. D. High affinity binding of SARS-CoV-2 spike protein enhances ACE2 carboxypeptidase activity. *J. Biol. Chem.* **295**, 18579–18588 (2020).
63. Arendse, L. B. et al. Novel therapeutic approaches targeting the renin-angiotensin system and associated peptides in hypertension and heart failure. *Pharmacol. Rev.* **71**, 539–570 (2019).
64. Dewey, F. E. et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* **354**, <https://doi.org/10.1126/science.aaf6814> (2016).
65. Wright, S. The genetical structure of populations. *Ann Eugen.* **15**, 323–354 (1951).
66. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: the impact of rare variants. *Genome Research*, **23**, 1514–1521 (2013).
67. Genomes Project, C. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
68. Albers, P. K. & McVean, G. Dating genomic variants and shared ancestry in population-scale sequencing data. *PLoS Biol.* **18**, e3000586 (2020).
69. International HapMap Consortium et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
70. McLaren, W. et al. The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
71. Suryamohan, K. et al. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nat. Genet.* **52**, 106–117 (2020).
72. Miersch, S. et al. Tetraivalent SARS-CoV-2 neutralizing antibodies show enhanced potency and resistance to escape mutations. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.10.31.362848> (2020).

## Acknowledgements

D.D. was supported by an F30 fellowship. Grant number: 1F30CA247147. N.J. was supported by the UCSF Quantitative Biosciences Institute Coronavirus Research Group (QCRG) fellowship. A.R. and S.C.S. were supported by SCELSE.

## Author contributions

Conceptualization by S.C.S. and S.Ses.; methodology by E.W.S., J.R.J., N.J., and S.Ses.; validation by S.Mie., J.L., C.C., Y.P.J., and M.U.; formal analysis by K.S., D.D., E.W.S., R.G., S.Mie., C.C., Y.P.J., F.A.F., J.F.S., P.K.A., T.D., A.R., and N.J.; investigation by S.Mie., J.L., C.C., Y.P.J., R.S., G.W., A.A., S.Moh., S.P., B.K., M.M., D.S., D.V., and M.U.; resources by H.G.H. S.Sid.; data curation by K.S., D.D., E.W.S., R.G., F.A.F., J.F.S., P.K.A., and S.Som.; writing-original draft preparation by K.S., D.D., E.W.S., S.Mie., J.F.S., J.R.J., N.J., and S.Ses.; writing-review and editing by K.S., D.D., E.W.S., R.G., S.Mie., M.M., D.S., J.R.J., N.J., and S.Ses.; visualization by K.S., D.D., E.W.S., S.Mie., Y.P.J., F.A.F., J.F.S., P.K.A., T.D., and M.U.; supervision by S.Sid., J.R.J., N.J. and S.Ses.; project administration by S.Ses.

## Competing interests

Employees of ModMab Therapeutics and MedGenome hold shares/options in their respective companies. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02030-3>.

**Correspondence** and requests for materials should be addressed to N.J. or S.S.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021