

## ACE 2 Coding Variants: A Potential X-linked Risk Factor for COVID-19 Disease

William T. Gibson<sup>1,2</sup>, Daniel M Evans<sup>3</sup>, Jianghong An<sup>4</sup> and Steven JM Jones<sup>1,4</sup>

### Affiliations:

<sup>1</sup>Department of Medical Genetics, University of British Columbia, C201-4500 Oak St, Vancouver, British Columbia, V6H 3N1, Canada

<sup>2</sup>Provincial Medical Genetics Program, BC Women's Hospital and Health Centre, 4500 Oak Street, Vancouver, British Columbia, V6H 3N1, Canada

<sup>3</sup>Centre for Applied Neurogenetics, University of British Columbia, Vancouver, Canada

<sup>4</sup>Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, Canada

### ORCID iDs:

William T Gibson: 0000-0002-6552-402X

Daniel M Evans: 0000-0002-8188-8667

Jianghong An: 0000-0002-6299-1327

Steven JM Jones: 0000-0003-3394-2208

### Abstract

Viral genetic variants are widely known to influence disease progression among infected humans. Given the recent and rapid emergence of pandemic SARS-CoV-2 infection, the cause of COVID-19 disease, viral protein variants have attracted research interest. However, little has yet been written about genetic risk factors among human hosts. Human genetic variation has proven to affect disease progression and outcome for important diseases such as HIV infection and malaria infestation. The fact that the human ACE2 protein is encoded on the X chromosome means that males who carry rare ACE2 coding variants will express those variants in all ACE2-expressing cells, whereas females will typically express those variants in a mosaic distribution determined by early X-inactivation events. This sex-based difference in ACE2 expression has unique implications for epidemiological studies designed to assess host genetic factors influencing progression from asymptomatic SARS-coV-2 infection to COVID-19. Here we present theoretical modelling of rare ACE2 coding variants documented to occur naturally in several human superpopulations and subpopulations, and show that rare variants predicted to affect the binding of ACE2 to the SARS-CoV-2 spike protein exist in people. Though the rs4646116 (p.Lys26Arg) allele is found in 1 in 70 Ashkenazi Jewish males, and in 1 in 172

39 non-Finnish European males, this allele is found at higher frequencies in females. Furthermore, the  
40 class of missense ACE2 alleles predicted to affect SARS-CoV-2 binding are found in aggregate among  
41 1.43% and 2.16% of Ashkenazi males and females, respectively, as well as in 0.58% and 1.24% of  
42 European males and females outside of Finland. These alleles are rarer in other population groups, and  
43 almost absent from East Asians genotyped to date.

44         Though we are aware that full genome-wide and exome-wide sequencing studies may  
45 ultimately be required to assess human genetic susceptibility to SARS-CoV-2 fully, we argue on the  
46 basis of strong prior probabilities that genotyping of this class of alleles is justified in cases of atypical  
47 SARS-CoV-2 diseases, such as asymptomatic super-spreaders (if any are identified), and in  
48 neonatal/paediatric-onset COVID-19 disease. Even relatively rare susceptibility factors (1% or fewer  
49 carriers) may become quantitatively important in the context of hundreds of thousands of infections. A  
50 small number of asymptomatic carriers, or a small number of super-spreaders, or a small segment of  
51 the population that is disproportionately likely to require intensive care, can magnify the medical,  
52 social and economic impacts of a pandemic of this size. The speed of the pandemic and the large  
53 number of affected cases worldwide justify efforts to identify all possible risk factors for adverse  
54 outcomes, including efforts to identify genetic susceptibility factors in human hosts.

55

## 56 **Introduction**

57         SARS-CoV-2 emerged as a human pathogen in late 2019.<sup>1</sup> Though SARS-CoV-2<sup>2</sup> infection is  
58 believed to be asymptomatic in some human hosts,<sup>3</sup> COVID-19 (coronavirus disease-2019)<sup>4</sup> is a  
59 serious respiratory illness with fatality rates that vary from 1.4% of diagnosed cases<sup>5</sup> to 50% or more  
60 among patients requiring ICU treatment.<sup>6</sup> Age and preexisting comorbidities have been described as  
61 risk factors for more serious disease,<sup>7,8</sup> and environmental factors such as initial dose of virions and  
62 availability of critical care are also believed important. Early data from the pandemic has suggested that  
63 males might account for a greater proportion of cases<sup>5</sup> and/or suffer from higher COVID-19 morbidity  
64 and mortality,<sup>9</sup> though if this is ultimately shown to be true, the reasons behind it are unclear.

65         Given the recent and rapid emergence of the COVID-19 global pandemic, viral genetic and  
66 protein variants have attracted immediate early attention as putative risk factors for disease progression,  
67 though little has been written about genetic risk factors among human hosts. Human genetic variation  
68 has proven to affect disease progression and outcome for important infectious diseases; variation in  
69 CCR5 has been shown to affect the possibility of progression from HIV infection to AIDS,<sup>10-14</sup> and

70 human host genetic variation has long been recognized as mitigating the outcome of infestation by  
71 malarial parasites (reviewed by Kariuki and Williams<sup>15</sup>).

72 The human ACE2 protein has recently been shown to serve as the binding site and point-of-  
73 entry of SARS-CoV-2 into human cells.<sup>16</sup> ACE2 had also previously been shown to bind SARS-CoV,<sup>17</sup>  
74 a related virus that caused serious respiratory infections in humans in the early 2000s. The viral Spike  
75 glycoproteins (“S” proteins) of SARS-CoV<sup>18</sup> and SARS-CoV-2<sup>19</sup> bind physically to the peptidase  
76 domain of ACE2 using their receptor binding domains (RBD), thereby enabling viral internalization.

77 We note here that the *ACE2* gene is located on the X-chromosome; human males are  
78 hemizygous for the gene, expressing only one haplotype from their maternally-derived allele. Females,  
79 by contrast, have two X chromosomes and generally express both their paternal and maternal alleles in  
80 a mosaic distribution that is determined by early X-inactivation. Since we have prior knowledge that  
81 the ACE2 protein is bound directly by the viral S protein, and since prior protein-protein interaction  
82 modelling has identified specific amino acid residues within ACE2 that are bound directly by SARS-  
83 CoV-2, we propose here that coding variants at these specific sites are likely to be a genetic risk factor  
84 for progression from SARS-CoV-2 infection to COVID-19 disease, and we further propose that this  
85 risk would affect human males and females differently.

## 86 87 **Methods**

88 The gnomAD database of human genetic variation catalogues coding variants from 141,456  
89 adults without childhood-onset neurodevelopmental syndromes.<sup>20</sup> Though little phenotypic detail is  
90 available on participants, this database serves as a convenient reference of “background” human  
91 genetic variation. We accessed the ACE2 entry in the gnomAD database (v.2.1.1) on March 30, 2020  
92 and downloaded separate .csv files of the predicted missense variants and predicted Loss-of-Function  
93 (pLoF) variants in human ACE2. We manually counted the number of coding missense variants (242  
94 variants) and pLoF variants (5 variants). We estimated sex-specific prevalences of rare and common  
95 coding ACE2 variants among the 76,702 male gnomAD participants (67,961 exomes and 8,741  
96 genomes) by totalling hemizygotes for all variants, and among the 64,754 female gnomAD participants  
97 (57,787 exomes and 6,967 genomes) by totalling all variants, then subtracting the total number of  
98 hemizygotes and homozygotes for each variant, such that females homozygous for any variant were  
99 counted only once. We then obtained more granular allele count and frequency information from VCF  
100 files downloadable from gnomAD v2.1. We calculated Minor Allele Frequencies (MAFs) within each

101 global superpopulation (e.g. Africans, South Asians) represented in gnomAD, and for specific  
102 subpopulations (e.g. Southern Europeans) where these were available. Among males, the prevalence of  
103 a rare X-linked variant is equivalent to the MAF for that allele. For females, the prevalence of a rare X-  
104 linked variant is equivalent to twice the MAF for that allele, because females have two X chromosomes  
105 (and hence, twice the number of X-linked alleles). We retrieved CADD scores from a downloaded  
106 version of CADD v1.4; the Phred-scaled CADD score<sup>21</sup> is reported for position and allele combination.  
107 We also used the online tool LIST<sup>22</sup> to estimate deleteriousness of coding amino acid substitutions in  
108 ACE2. Three-dimensional modelling used MolSoft.<sup>23</sup> We also used the method of Schapira et al.<sup>24</sup> to  
109 estimate the change in binding energy that would accrue for the 15 naturally-occurring missense  
110 variants.

## 111 **Results**

### 112 **ACE2 is Generally Intolerant to Loss-of-function Variants**

113 ACE2 is annotated as being intolerant to pLoF single-nucleotide variants (SNVs), with 3 pLoF  
114 SNVs observed, yet 31 such variants expected based on the size of the gene. The two stop-gain  
115 mutations (p.Leu116Ter and p.Leu656Ter) and the frameshift variant (p.Gly422ValfsTer15) are not  
116 seen among hemizygotes. The two splice variants that are seen in the hemizygous state may allow  
117 some correctly-spliced product to be made; allowing for that possibility, ACE2 variants confidently  
118 predicted to abrogate protein expression were not observed in the hemizygous state among the 76,702  
119 male gnomAD participants (67,961 exomes and 8,741 genomes).

### 120 **ACE2 Missense Variants Occur at and Near Residues that Bind SARS-CoV-2**

121 The gnomAD database catalogues 242 missense variants in ACE2, of which only 15 are  
122 predicted to lie at or near the Spike protein binding site of ACE2. The 67,961 male exomes and 8,741  
123 male genomes in the gnomAD cohort sampled 76,702 X chromosomes, whereas the 57,787 female  
124 exomes and 6,967 female genomes sampled 64,754 X chromosomes (total number of X chromosomes  
125 originally sampled: 206,210). However, the actual sample size of chromosomes for each variant is  
126 smaller, because not all X chromosomes genotyped successfully at each site (Supplementary Table 1).  
127 The range varied from 158,104 X chromosomes genotyped successfully for p.Val488Ala to 183,374 X  
128 chromosomes that genotyped successfully for p.Glu35Lys. Variant frequencies in Supplementary Table  
129 1 were calculated using the number of alternate allele counts as the numerator and the number of  
130 chromosomes genotyped successfully at the locus as the denominator. Aggregating the allele  
131 frequencies of ACE2 variants located at the SARS-CoV-2 binding site (i.e. excluding p.Asn720Asp  
132 because it is not at the binding site) yields a frequency of 0.0041, which estimates that roughly 3.9

133 males per 1000 and 8.5 females per 1000 would harbour a rare allele that might affect the binding of  
134 SARS-CoV-2 to cells that express ACE2. Importantly, these rare alleles are not distributed evenly  
135 across human subpopulations, as outlined below.

136 The only common missense Single-Nucleotide Polymorphism (SNP) is rs41303171, predicted  
137 to encode p.Asn720Asp (UniProt Accession Number Q9BYF1). This SNP was observed in 1.7% of  
138 male gnomAD samples and in 3.1%-3.2% of female participants overall. Calculated allele frequencies  
139 among females vary slightly depending on how the counts are done. The downloadable .csv file lists  
140 the rs41303171 variant allele as observed 3,054 times: from that we can subtract 1,026 hemizygotes  
141 and 34 homozygotes (to avoid counting those females twice), yielding 1,994 heterozygotes +  
142 homozygotes among 64,754 female samples, such that 3.08% of female samples are estimated to have  
143 at least one p.Asn720Asp. This method assumes that all samples genotyped successfully at that site. To  
144 take genotyping efficiency into account we downloaded the relevant .vcf files and divided the Alternate  
145 Allele Count (1,809) by the Allele Number at that locus (112,288 alleles genotyped successfully among  
146 females) to yield a MAF of 1.6%, so 3.2% of females have at least one p.Asn720Asp allele. This allele  
147 would be frequent enough to be detected by Genome-Wide Association Studies (GWAS), either  
148 directly or by imputation, if this variant or one nearby were involved in the disease under study; to date  
149 no GWAS have flagged this locus.<sup>25</sup> However, rs41303171 does not lie in a domain of the protein that  
150 is predicted to interact with SARS-CoV-2, so it is not an obvious candidate site for this particular host-  
151 pathogen interaction.

152 The various superpopulations and subpopulations in gnomAD each offer different sample sizes.  
153 For rare X-linked variants it is best to deconvolute allele frequencies by both sex and subpopulation, to  
154 achieve a fuller picture of the prevalence of hemizygous males in individual regions or countries. Any  
155 disease risk signal conferred by a functional X-linked allele will otherwise be diluted by lumping  
156 hemizygous males in with heterozygous female carriers, who are likely at a lower risk due to mosaic  
157 expression of the variant. After rs41303171, rs4646116 is the next-most-common missense ACE2  
158 variant (global MAF 0.4%); it is predicted to encode p.Lys26Arg. Our 3D modelling predicted  
159 p.Lys26Arg to lie immediately beside the ACE2-Spike protein interface (albeit with the lysine side  
160 chain pointing away from the interface). Prior work on ACE2 binding to SARS-CoV<sup>18, 26</sup> and recent  
161 work on ACE2 binding to SARS-CoV-2<sup>27, 28</sup> by others has identified three main S protein binding  
162 domains within ACE2. The largest includes Gln24, Thr27, Phe28, Asp30, Lys31, His34, Glu35,  
163 Asp38, Tyr41, Gln42 and Leu45. The smallest comprises Met82 and Tyr83, and the last includes

164 Asn330, Lys353, Gly354, Asp355 and Arg357. According to Li et al,<sup>29</sup> residues Asp30 to Tyr41 make  
165 up the alpha-1 ridge, residues Met82 to Val93 make up the Loop and alpha-3 ridge, and residues  
166 Lys353 to Arg357 make up the Loop and beta-5-helix.

167 We estimated a high prior probability that human germline coding variants affecting these key  
168 binding residues would also affect the affinity of the endogenous human ACE2 protein for the viral  
169 spike protein. By visual inspection of gnomAD's ACE2 missense variant list, we conservatively  
170 considered p.Lys26Glu, p.Lys26Arg, p.Thr27Ala, p.Glu35Lys, p.Glu37Lys, p.Phe40Leu, p.Ser43Arg,  
171 p.Met82Ile, p.Pro84Thr, p.Gly326Glu, p.Glu329Gly, p.Gly352Val, p.Asp355Asn and p.Val488Ala to  
172 be at or near the ACE2-S protein interface. All of these variants occur naturally in humans, and could  
173 plausibly affect risk for progression to COVID-19 after an initial encounter with SARS-CoV-2. We  
174 modelled the interface between ACE2 and the viral spike protein as shown in Figure 1.

For Peer Review



215 Considering this class of ACE2 variants as a group, their estimated aggregate prevalence is 3.9  
216 per 1000 males and 8.5 per 1000 females. Rare deleterious X-linked alleles will be depleted in live  
217 male adults, and observed disproportionately among heterozygous carrier females. Thus, the  
218 observation that this class of alleles is more than twice as common among females as in males may  
219 reflect some degree of historical selection against these alleles. We went on to calculate subpopulation-  
220 specific frequencies by sex, as .vcf files downloadable from the gnomAD interface provide the  
221 breakdown of exomes and genomes by subpopulation and by sex. rs4646116 (p.Lys26Arg) is found at  
222 polymorphic frequencies (MAF 1.43%, or 1 in 70) in Ashkenazi Jewish males, and at lower frequency  
223 among non-Finnish European males (MAF 0.58%, or 1 in 172). In Southern Europeans, not segregated  
224 by sex, the MAF is 0.34% (1 in 300). In males of Latino ancestry the MAF is 0.28% (1 in 360), and the  
225 frequency is lower among African Males (0.13% or 1 in 770), South Asian males (0.08% or 1 in 1,250)  
226 and Finnish males (0.07% or 1 in 1,430). This allele is at very low frequencies in East Asian  
227 participants, being observed only once in a heterozygous female (MAF of 0.01%). The allele was  
228 absent from Korean (n=2887 chromosomes) and Japanese (n=122 chromosomes). The remaining  
229 alleles p.Lys26Glu, p.Thr27Ala, p.Glu35Lys, p.Glu37Lys, p.Phe40Leu, p.Ser43Arg, p.Met82Ile,  
230 p.Pro84Thr, p.Gly326Glu, p.Glu329Gly, p.Gly352Val, p.Asp355Asn and p.Val488Ala are encountered  
231 very infrequently (between 0 and 4 times in each of the gnomAD subpopulations). However, since all  
232 of these naturally-occurring alleles occur at the binding interface, they may legitimately be grouped  
233 together to estimate allelic burden arising from this class of alleles. The fact that rare X-linked alleles  
234 are observed twice as often in females means that 2.1% of Ashkenazi Jewish females (1 in 48), carries  
235 a missense variant predicted to affect binding of SARS-CoV-2, as do 1.2% of non-Finnish European  
236 females. Similar calculations show that 0.74% of Latino females (1 in 135), 0.48% of South Asian  
237 females (1 in 208), 0.28% of African females (1 in 360) and 0.13% of Finnish females (1 in 770) carry  
238 a rare ACE2 allele that may affect binding to SARS-CoV-2.

239 The CADD scores among this class of alleles range from 0.017 to 24.9. We provide these  
240 scores here for interest. LIST scores for predicting the deleteriousness of all possible missense ACE2  
241 alleles are available at <https://list.msl.ubc.ca/proteins/Q9BYF1>. We offer the caveat that scores of  
242 deleteriousness are difficult to interpret in this context, because they are not based on specific  
243 characteristics of a protein's ability to bind to coevolved (or in this case, newly-evolved) partners.

#### 244 **Human ACE2 Missense Variants May Increase or Decrease its Binding to SARS-CoV-2**

245 Applying the method of Schapira et al.,<sup>24</sup> we estimated the change in binding energy of the 15  
246 missense variants (Supplementary Table 2). Glu37Lys increased the binding the most, followed by



247 Thr27Ala, Lys329Gly, and Lys26Glu. Val488 and Asn720 are relatively distant from the interface, so  
248 they may contribute to the binding of other partners, but not likely the S protein. Several variants were  
249 predicted to weaken binding between ACE2 and SARS-CoV-2 spike protein: Asn720Asp weakened  
250 binding somewhat, with Ser43Arg, Gly326Glu, Met82Ile and Lys26Arg all weakening binding  
251 progressively more. The fact that Lys26Glu and Lys26Arg seem to have opposing effects suggests that  
252 the bulkier arginine side chain may not pack as tightly with viral S protein, whereas replacing the basic  
253 lysine with an acidic glutamate residue may disrupt important polar interactions.<sup>30</sup> Brielle et al. note  
254 that the Phe486 of SARS-CoV-2 is critical for binding ACE2 residues Leu79 and Met82, thereby  
255 stabilizing the S protein-ACE2 interaction.<sup>31</sup> Methionine 82 interacts with Phe486 of the viral protein's  
256 RBD via van der Waals forces, so an isoleucine at this position may weaken this interaction.<sup>30</sup> Yan et  
257 al. note that Arg426 of SARS-CoV forms a salt bridge with Glu329 on ACE2, and that SARS-CoV-2  
258 has Asn439 at the analogous position, which cannot form this salt bridge. We speculate here that  
259 replacing glutamate 329 with a smaller glycine side chain in the p.Glu329Gly variant may allow easier  
260 binding of the SARS-CoV-2 spike protein's Asn439, though this remains to be tested.

## 261 Discussion

262 At the time of this writing, SARS-CoV-2 has infected over 1,250,000 people, with over 335,000  
263 cases in the United States, approximately 130,000 cases in each of Italy and Spain, over 100,000 cases  
264 in Germany and nearly as many again in France. China has reported 3,333 deaths among 82,602  
265 cases;<sup>32</sup> since new cases have dropped precipitously there, we may estimate a fatality rate among  
266 recognized cases at 4%. However, Spain has reported over 12,500 deaths, and Italy has reported nearly  
267 16,000 deaths. Though a large number of asymptomatic and/or unreported cases would dilute the true  
268 case-fatality rate, fatality rates of 9.6% and 12% among the subset of cases that are recognized and  
269 reported are significant. The speed of the pandemic and the large number of affected cases worldwide  
270 justify efforts to identify all possible risk factors for adverse outcomes, including efforts to identify  
271 genetic susceptibility factors in human hosts. Even relatively rare susceptibility factors (1% or fewer  
272 carriers) may become quantitatively important in the context of hundreds of thousands of infections. A  
273 small number of asymptomatic carriers, or a small number of super-spreaders, or a small segment of  
274 the population that is disproportionately likely to require intensive care, can magnify the medical,  
275 social and economic impacts of a pandemic of this size.

276 The genetic component of host-pathogen interaction studies forms a subset of gene-by-  
277 environment interaction studies. These latter studies require 10<sup>5</sup> or more participants and sensitive

278 measures of environmental exposures, and typically identify tagging SNPs linked to loci with small  
279 effect sizes. These methods, their attendant complexities, and the conservative genomewide thresholds  
280 they require are justified by the fact that previous candidate-gene studies had extremely high false-  
281 discovery rates,<sup>33</sup> largely due to the use of convenience samples that generated artefactual signals  
282 owing to population stratification (i.e. the fact that SNP allele frequencies vary between ancestral  
283 populations for historical reasons that are unrelated to heritable risk for disease). However, given the  
284 truly staggering number of SARS-CoV-2 infections and the rapid progression from COVID-19 to death  
285 (often within 14 days), such labour-intensive gene-agnostic methods are unlikely to yield clinically-  
286 actionable results during the current pandemic.

287 For these reasons, we argue that prior knowledge of SARS-CoV biology and emerging  
288 knowledge of SARS-CoV-2 similarities to SARS-CoV justify the assessment of variation within  
289 human ACE2 as an exceptionally strong candidate gene for host response to SARS-CoV-2 infection.  
290 The fact that ACE2 is X-linked means that rare variants that enhanced SARS-CoV-2 binding *in vivo*  
291 would likely increase susceptibility to COVID-19 among males. Other functional effects are possible  
292 also; for example, female variant carriers might have a shorter asymptomatic period, or be more likely  
293 to develop symptoms, irrespective of viral shedding status. Given that a limited number of coding  
294 variants in ACE2 are predicted to affect the binding of SARS-CoV-2 spike protein, these variants could  
295 be genotyped rapidly in early-onset cases (e.g. paediatric cases, or even cases under age 40) as a pilot  
296 study in the United States and/or in European countries where case numbers are high, case fatality  
297 appears to be high and the rare alleles are likely to be found. The delivery of viral nucleic acid testing  
298 to affected populations is a more urgent need, but given that many viral samples are already being  
299 collected, the addition of a rapid assay targeting the SNVs listed here would be an efficient means to  
300 gather data to test this hypothesis. Notably, tests of rare variant burden do lose power rapidly if neutral  
301 variants and variants with opposing functional effects are grouped together, so we have restricted  
302 ourselves to variants with prior evidence of involvement in SARS-CoV-2 Spike protein binding. There  
303 are many more missense ACE2 variants in gnomAD; the aggregate crude prevalence of all rare (i.e.  
304 excluding the common p.Asn720Asp) missense variants in gnomAD males is 2.49% and in gnomAD  
305 females is 5.81%. Furthermore, among patients with extreme outcomes such as death from COVID-19  
306 in childhood (or asymptomatic super-spreaders, if any are found), sequencing of the entire ACE2  
307 coding region might conceivably identify ultra-rare “private” variants not catalogued in gnomAD,  
308 enabling assessment of the role of these variants in the conversion from initial SARS-CoV-2 infection  
309 to COVID-19. A properly-powered genome-wide study with correlation of genetic variants to multiple

310 other clinical risk factors and outcomes is also advisable, but cannot realistically be accomplished in  
311 the near term.

312 Procko has done unbiased mutagenesis of the ACE2 regions that bind to the RBD of SRS-  
313 COV-2,<sup>34</sup> and noted that substitution of Thr27 with hydrophobic residues (e.g. alanine) is expected to  
314 increase the ability of aromatic residues of the S protein to pack into the ACE2 interface. Procko's *in*  
315 *vitro* experiments flagged Asn 90 and Thr92 as sites at which nonsynonymous variants should increase  
316 binding of the viral RBD motif. The p.Thr92Ile variant (rs763395248) occurs naturally in Southern  
317 Europeans (2 out of 7040 X chromosomes in gnomAD, both in males, MAF of  $2.8 \times 10^{-4}$ ). This *in vitro*  
318 work offers proof-of-principle for our hypothesis that naturally-occurring human ACE2 variants are  
319 likely to affect SARS-CoV-2 infection kinetics *in vivo*. At the very least, further *in vitro* functional  
320 assays of viral binding and replication within cell lines transfected to express human variant ACE2  
321 proteins are urgently justified. Furthermore, given recent data that human recombinant soluble ACE2  
322 (hrsACE2) can block early stages of SARS-CoV-2 infections,<sup>35</sup> any naturally-occurring human ACE2  
323 variant that bound more tightly to the viral spike protein might serve as a better "decoy" for the viral  
324 protein, and would be a candidate for novel hrsACE2 molecules with therapeutic potential.

325 Whether or not these studies can be done in time to affect patient outcomes during the current  
326 pandemic, the fact that three pathogenic coronaviruses (SARS-CoV,<sup>36</sup> MERS-CoV<sup>37</sup> and SARS-CoV-  
327 2<sup>38</sup>) have emerged in the past two decades, two of which bind and enter via ACE2, augurs that  
328 additional coronaviruses that bind to ACE2 are likely to emerge in the future. Thus, knowledge of this  
329 specific host-pathogen interaction at the molecular level is important to have at this time.

330

331

332 **References:**

- 333 1. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin.  
334 *Nature* **579**, 270-273 (2020).
- 335 2. Gorbalenya Alexander, E. *et al.* The species Severe acute respiratory syndrome-related coronavirus:  
336 classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* **5**, 536-544 (2020).
- 337 3. Hoelscher, M. *et al.* Transmission of 2019-nCoV Infection from an Asymptomatic Contact in  
338 Germany. *N. Engl. J. Med.* **382**, 970-971 (2020).
- 339 4. WHO. Novel Coronavirus (2019-nCoV) Situation Report – 22. *World Health Organization* (2020).
- 340 5. Guan, W. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.*  
341 (2020).
- 342 6. Bhatraju, P. K. *et al.* Covid-19 in Critically Ill Patients in the Seattle Region - Case Series. *N. Engl.*  
343 *J. Med.* (2020).
- 344 7. Chinese Center for Disease Control and Prevention. Novel Coronavirus Pneumonia Emergency  
345 Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel  
346 coronavirus diseases (COVID-19) in China [Chinese]. *Chinese Journal of Epidemiology* **41**, 145–51  
347 (2020).
- 348 8. CDC COVID-19 Response Team. Severe Outcomes Among Patients with Coronavirus Disease 2019  
349 (COVID-19) — United States, February 12–March 16, 2020. *Morbidity and Mortality Weekly Report*  
350 **69**, 343-346 (2020).
- 351 9. Nasiri, M. *et al.* COVID-19 clinical characteristics, and sex-specific risk of mortality: Systematic  
352 Review and Meta-Analysis. (2020). *bioRxiv*
- 353 10. Doranz, B. J. *et al.* A Dual-Tropic Primary HIV-1 Isolate That Uses Fusin and the  $\beta$ -Chemokine  
354 Receptors CKR-5, CKR-3, and CKR-2b as Fusion Cofactors. *Cell* **85**, 1149-1158 (1996).
- 355 11. Choe, H. *et al.* The  $\beta$ -Chemokine Receptors CCR3 and CCR5 Facilitate Infection by Primary HIV-  
356 1 Isolates. *Cell* **85**, 1135-1148 (1996).
- 357 12. Dragic, T. *et al.* HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5.  
358 *Nature* **381**, 667-673 (1996).
- 359 13. Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-  
360 exposed individuals to HIV-1 infection *Cell* **86**, 367-77 (1996).
- 361 14. Dean, M. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the  
362 CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study,  
363 Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* **273**, 1856-  
364 1862 (1996).

- 365 15. Kariuki, S. N. & Williams, T. N. Human genetics and malaria resistance. *Human Genetics* (2020).
- 366 16. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by  
367 a Clinically Proven Protease Inhibitor. *Cell* (2020).
- 368 17. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus.  
369 *Nature* **426**, 450-4 (2003).
- 370 18. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS Coronavirus Spike Receptor-  
371 Binding Domain Complexed with Receptor. *Science* **309**, 1864-1868 (2005).
- 372 19. Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.  
373 *Cell* (2020).
- 374 20. Karczewski, K. J., Francioli, L. C. & Tiao, G. Variation across 141,456 human exomes and  
375 genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes.  
376 *bioRxiv* (2019).
- 377 21. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic  
378 variants. *Nat. Genet.* **46**, 310-5 (2014).
- 379 22. Malhis, N., Jones, S. & Gsponer, J. Improved measures for evolutionary conservation that exploit  
380 taxonomy distances. *Nature Communications* **10** (2019).
- 381 23. Abaygan, R., Totrov, M. & Kuznetsov, D. ICM—A new method for protein modeling and design:  
382 Applications to docking and structure prediction from the distorted native conformatio *Journal of*  
383 *Computational Chemistry* **15**, 488-506 (1994).
- 384 24. Schapira, M., Totrov, M. & Abagyan, R. Prediction of the binding energy for small molecules,  
385 peptides and proteins. *Journal of Molecular Recognition* **12**, 177–190 (1999).
- 386 25. <https://www.ebi.ac.uk/gwas/>. Accessed 30 March, 2020.
- 387 26. Ke Wang *et al.* SARS-CoV-2 invades host cells via a novel route: CD147-spike protein. (2020).
- 388 27. Milanetti, E. *et al.* In-Silico evidence for two receptors based strategy of SARS-CoV-2. *bioRxiv*  
389 (2020).
- 390 28. Lupala, C. S. *et al.* Computational simulations reveal the binding dynamics between human ACE2  
391 and the receptor binding domain of SARS-CoV-2 spike protein. *bioRxiv* (2020).
- 392 29. Li, W. *et al.* Receptor and viral determinants of SARS coronavirus adaptation to human ACE2.  
393 *EMBO J* **24**, 1634–1643 (2005).
- 394 30. Yan, R. *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2.  
395 *Science* **367**, 1444-1448 (2020).

- 396 31. Brielle, E. S., Schneidman-Duhovny, D. & Linial, M. Title: The SARS-CoV-2 exerts a distinctive  
397 strategy for interacting with the ACE2 human receptor. (2020).
- 398 32. Dong, E. & Du, H. G., L. An interactive web-based dashboard to track COVID-19 in real time.  
399 *Lancet Infectious Diseases* (2020).
- 400 33. John P. A. Ioannidis, Robert Tarone & Joseph K. McLaughlin. The False-positive to False-negative  
401 Ratio in Epidemiologic Studies. *Epidemiology* **22**, 450-456 (2011).
- 402 34. Procko, E. The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS  
403 coronavirus2. (2020). *bioRxiv*
- 404 35. Monteil, V. *et al.* Inhibition of SARS-CoV-2 infections in engineered human tissues using clinical-  
405 grade soluble human ACE2. (2020).
- 406 36. Ksiazek, T. G. *et al.* A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N.*  
407 *Engl. J. Med.* **348**, 1953-1966 (2003).
- 408 37. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of  
409 a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. *N. Engl. J. Med.* **394**, 1814-1820  
410 (2013).
- 411 38. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New*  
412 *England Journal of Medicine* **382**, 727-733 (2020).
- 413