1 **Main Manuscript for**

2 **A 3D Structural Interactome to Explore the Impact of Evolutionary Divergence,**
3 **Population Variation, and Small-molecule Drugs on SARS-CoV-2-Human Protein-**
4 **Protein Interactions**

5 Shayne D. Wierbowski[1,2], Siqi Liang[1,2], You Chen[2,3], Nicole M. Andre[4], Steven M. Lipkin[5], Gary R.
6 Whittaker[4,6], Haiyuan Yu[1,2,†]

7 1. Department of Computational Biology, Cornell University, Ithaca, NY 14853, USA

8 2. Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY 14853, USA

9 3. Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA

10 4. Department of Microbiology and Immunology, Cornell University, Ithaca NY 14853, USA

11 5. Department of Medicine, Weill-Cornell Medicine, New York, NY 10021

12 6. Master of Public Health Program, Cornell University, Ithaca NY 14853, USA

13 † Correspondence should be addressed to H.Y. (haiyuan.yu@cornell.edu)

14 **Abstract**

15   The recent COVID-19 pandemic has sparked a global public health crisis. Vital to the development of

16   informed treatments for this disease is a comprehensive understanding of the molecular interactions

17   involved in disease pathology. One lens through which we can better understand this pathology is

18   through the network of protein-protein interactions between its viral agent, SARS-CoV-2, and its human

19   host. For instance, increased infectivity of SARS-CoV-2 compared to SARS-CoV can be explained by

20   rapid evolution along the interface between the Spike protein and its human receptor (ACE2) leading to

21   increased binding affinity. Sequence divergences that modulate other protein-protein interactions may

22   further explain differences in transmission and virulence in this novel coronavirus. To facilitate these

23   comparisons, we combined homology-based structural modeling with the ECLAIR pipeline for interface

24   prediction at residue resolution, and molecular docking with PyRosetta. This enabled us to compile a

25   novel 3D structural interactome meta-analysis for the published interactome network between SARS-

26   CoV-2 and human. This resource includes docked structures for all interactions with protein structures,

27   enrichment analysis of variation along interfaces, predicted $\Delta\Delta G$ between SARS-CoV and SARS-CoV-

28   2 variants for each interaction, predicted impact of natural human population variation on binding affinity,

29   and a further prioritized set of drug repurposing candidates predicted to overlap with protein interfaces[†].

30   All predictions are available online[†] for easy access and are continually updated when new interactions

31   are published.

32

33   † Some sections of this pre-print have been redacted to comply with current bioRxiv policy restricting

34   the dissemination of purely in silico results predicting potential therapies for SARS-CoV-2 that have not

35   undergone thorough peer-review. The results section titled "Prioritization of Candidate Inhibitors of

36   SARS-CoV-2-Human Interactions Through Binding Site Comparison," Figure 4, Supplemental Table 9,

37   and all links to our web resource have been removed. Blank headers left in place to preserve structure

38   and item numbering. Our full manuscript will be published in an appropriate journal following peer-

39   review.

Main Text

Introduction

The ongoing global COVID-19 pandemic caused by the infection of SARS-CoV-2 has to date infected more than 30 million people and caused more than 940,000 deaths worldwide[1]. Coronaviruses are a family of enveloped viruses that cause respiratory and/or enteric tract infections in a wide range of avian and mammalian hosts[2]. To date, seven well characterized coronaviruses infect humans[3-5] with severity ranging from a mild respiratory illness to severe pneumonia and acute respiratory distress syndrome (ARDS). Among these, SARS-CoV-2 is unique in its characteristics being both highly transmissible and capable of causing severe disease in a subset of individuals; whereas other human coronaviruses are either highly transmissible yet generally not highly pathogenic (e.g. HCoV-229E, HCoV-OC43) or highly pathogenic but poorly transmissible (SARS-CoV and MERS-CoV). SARS-CoV-2 is also unique in its range of infection and pathogenesis[6, 7]. While the vast majority of infected individuals (~25-35%) experience only mild or minimal symptoms, ~1-2% of infected patients die primarily from severe respiratory failure and acute respiratory distress syndrome[8, 9]. The differences in morbidity, hospitalization, and mortality among different ethnic groups—particularly Blacks and Hispanics[10-15]—could not be fully explained by cardiometabolic, socioeconomic, or behavioral factors, suggesting that human genetic variation may significantly impact SARS-CoV-2 pathogenicity. Therefore, insights into the evolution of SARS-CoV-2, its markedly elevated transmission rate relative to SARS-CoV, and dynamic range of symptoms are currently of key areas of interest. These traits are likely driven by differences in the mechanism of pathology and interactions between the virus and its host cells, but their specific causes are yet to be characterized.

One avenue to better understand the mechanisms of either viral- or bacterial- infection and pathology is through studying the network of protein-protein interactions that occur between a pathogen and its host. Viral-human interactome maps have previously been compiled for SARS-CoV[16], HIV[17], Ebola virus[18], and Dengue and Zika viruses[19] among others. Recently, an affinity-purification mass-spectrometry approach has been applied to 29 SARS-CoV-2 proteins uncovering 332 viral-human interactions[20]. These inter-species interactions play central roles in disease progression including, acting as facilitators of pathogen entry into host cells[21-26], inducing an inhibitory effect on host response proteins and pathways[27-29], and hijacking cell signaling or metabolism to accelerate cellular—and

68   consequentially viral—replication[30-32]. Understanding the structure and dynamics of these interactions

69   can provide critical insights into their effect on the cell. For instance, determination of the structure of

70   an interaction between poxvirus chemokine inhibitor vCCI and human MIP-1β revealed that the viral-

71   human binding interface occluded domains vital to chemokine homodimerization, receptor binding, and

72   interactions with GAG, thus explaining how this interaction elicits an inhibitory effect on chemokine

73   signaling[29]. Additionally, the dynamics of an interaction between a herpesvirus cyclin and human cdk2

74   revealed that although the binding interface was distinct from that between human cyclin A and cdk2,

75   the net conformational impact on cdk2 effectively mimicked that of the native interaction leading to

76   dysregulated cell cycle progression[31].

77   Because protein-protein interactions are responsible for mediating the majority of protein

78   function[33-35], targeted disruption of these interactions by small molecule inhibitors that compete for the

79   same binding site can offer a precise toolkit to modulate cellular function[33, 35-38]. For example, BCL-2

80   inhibitors that displace bound anti-apoptotic BCL-X interactors can treat chronic lymphocytic leukemia

81   pathogenesis[39, 40]. Targeted disruption of protein interactions can be particularly effective in viral

82   networks due to their small proteomes with highly optimized function, and potent inhibitors of key

83   interactions have been developed against viral proteins. Targeted disruption of viral complexes—

84   particularly those responsible for viral replication—has been successful in vaccinia virus[41] and human

85   papilloma virus therapies[42, 43]. In particular, disruption of viral-host protein-protein interactions explicitly

86   involved in early viral infection is an important therapeutic strategy. Discovery that a population variant in

87   the membrane protein CCR5 conferred resistance to HIV-1 by disrupting its interaction with the viral

88   envelope glycoprotein led to the development of Maraviroc as an FDA approved treatment for HIV-1

89   that functions by blocking the interface for this interaction[23, 44].

90   Here we apply a comprehensive interactome modeling framework to construct a 3D structural

91   interactome between SARS-CoV-2 and human protein based on the current interactome map published

92   by Gordon *et al*.[20] Our framework consists of homology modeling to maximize structural coverage of the

93   SARS-CoV-2 proteome, applying our previous ECLAIR classifier[45] to identify interface residues for the

94   whole SARS-CoV-2-human interactome, followed by atomic resolution interface modeling through

95   guided docking in PyRosetta[46]. We additionally carried out in-silico scanning mutagenesis to predict the

96   impact of mutations on interaction binding affinity and performed a comparison of protein-protein and

97   protein-drug binding sites. We compile all results from our structural interactome into a user-friendly

98   web server allowing for quick exploration of individual interactions or bulk download and analysis of the

99   whole dataset. Further, we explore the utility of our interactome modeling approach in identifying key

100 interactions undergoing evolution along viral protein interfaces, highlighting population variants on

101 human interfaces that could modulate the strength of viral-host interactions to confer protection from or

102 susceptibility to COVID-19, and prioritizing drug candidates predicted to bind competitively at viral-

103 human interaction interfaces.

104 **Results**

105 <u>Enrichment of divergence between SARS-CoV and SARS-CoV-2 at spike-ACE2 binding interface</u>

106 To highlight the utility of computational and structural approaches to model the SARS-CoV-2-human

107 interactome, we first examined the interaction between the SARS-CoV-2 spike protein (S) and human

108 angiotensin-converting enzyme 2 (ACE2) (**Fig 1.a**). This interaction is key for viral entry into human

109 cells[3] and is the only viral-human interaction with solved crystal structures available in both SARS-CoV[47]

110 and SARS-CoV-2[48-50]. Comparison between SARS-CoV and SARS-CoV-2 revealed that sequence

111 divergence of the S protein was highly enriched at the S-ACE2 interaction interface (**Fig 1.a**;

112 Log2OddsRatio=2.82, p=1.97e-5), indicating functional evolution around this interaction. To explore the

113 functional impact of these mutations on this interaction, we leveraged the Rosetta energy function[51] to

114 estimate the change in binding affinity ($\Delta\Delta G$) between the SARS-CoV and SARS-CoV-2 versions of the

115 S-ACE2 interaction (**Fig 1.b** and **1.c**). The predicted negative $\Delta\Delta G$ value of -14.66 Rosetta Energy Units

116 (REU) indicates an increased binding affinity using the SARS-CoV-2 S protein driven by better optimized

117 solvation and hydrogen bonding potential fulfillment along the ACE2 interface. Our result is consistent

118 with the hypothesis that increased stability of the S-ACE2 interaction is one of the key reasons for

119 elevated transmission of SARS-CoV-2[52]. Moreover, recent experimental energy kinetics assays have

120 shown that SARS-CoV-2 S protein binds ACE2 with 10-20-fold higher affinity than that of SARS-CoV S

121 protein[53] supporting the conclusions from our computational modeling.

122   Medical experts have also noted a wide range in severity of and susceptibility to SARS-CoV-2

123   among individuals[6, 7, 54]. Several hypotheses for genetic predisposition models have been proposed

124   including that expression quantitative trait loci (eQTLs) may up- or down-regulate host response genes

125   and that functional coding variants may alter viral-human interactions[55, 56]. For instance, a recent RNA-

126   sequencing analysis suggested that higher expression of ACE2 in Asian males effectively provides more

127   routes of entry for the virus and could explain increased susceptibility among this population[57].

128   Alternatively, missense population variants in ACE2 could modulate susceptibility to infection by

129   strengthening or weakening the S-ACE2 interaction. A total of six missense population variants reported

130   in gnomAD[58] localize to the S-ACE2 interface. Using a mutation scanning pipeline in PyRosetta[59, 60] we

131   predicted the impact on binding affinity for each variant (**Fig 1.d**). The three population variants

132   predicted to have the largest impact on S-ACE2 binding affinity— ACE2_E37K ($\Delta\Delta G=1.50$), ACE2_M82I

133   ($\Delta\Delta G=2.95$), and ACE2_G326E ($\Delta\Delta G=5.74$)—were all consistent with previous experimental screens

134   which identified them as putative protective variants exhibiting decreased binding of ACE2 to S[61, 62].

135   Cumulatively, our results highlight the usefulness of a 3D structural interactome modeling approach in

136   identifying interactions and mutations important for viral infection, pathogenesis, and transmission.

137   <u>Constructing the 3D Structural SARS-CoV-2-Human Interactome</u>

138   After successfully applying our modeling approaches to recapitulate the effect of mutations along the S-

139   ACE2 interface, we expanded our efforts to compile a complete 3D structural interactome between

140   SARS-CoV-2 and human proteins. An early interactome screen by Gordon *et al.*[20] uncovered 332 viral-

141   human interactions that provide the foundation for our 3D interactome. To model these interactions, we

142   first constructed homology models for 18 of 29 SARS-CoV-2 proteins with suitable templates

143   (**Supplemental Figure 1**). Then, we predicted the interface residues for each interaction using our

144   **E**nsemble **C**lassifier **L**earning **A**lgorithm to predict **I**nterface **R**esidues (ECLAIR) framework[45]. In total,

145   our pipeline identified 692 interface residues across 18 SARS-CoV-2 proteins with an average 24.7

146   residues per interface and 6,763 across 190 human proteins with an average 20.3 residues per

147   interface. A summary of the classifier utilization, prediction confidences, and interface size is provided

148   in **Supplemental Figure 2**.

149   In order to add a structural component to our interactome map, and thereby enable modeling of

150   the binding affinity for these interactions, we additionally performed docking in PyRosetta using our

151   ECLAIR interface likelihood predictions to refine the search space (**Supplemental Figure 3**). Structural

152   models were available for both the human and SARS-CoV-2 proteins in 250 out of 332 interactions

153   (75%) making them amenable to guided docking experiments. After performing up to 50 independent

154   docking experiments for each interaction and retaining the top-scored conformation we report 959

155   interface residues across 18 SARS-CoV-2 proteins with an average 18.2 residues per interface and

156   4,483 across 250 human proteins with an average 17.9 residues per interface (**Supplemental Figure

157   2.g**). For all analyses, interface annotations provided from docking experiments were prioritized over

158   our ECLAIR predictions. The full interface annotations from our ECLAIR and docking predictions are

159   available in **Supplemental Table 1** and **Supplemental Table 2** respectively.

160   <u>Perturbation of Human Proteins by Disease Mutations and Binding of SARS-CoV-2 Interactions Occur</u>

161   <u>at Distinct Sites</u>

162   After constructing the 3D interactome between SARS-CoV-2 and human, we first looked for evidence

163   of interface-specific variation by mapping both gnomAD[58] reported human population variants

164   (**Supplemental Table 3**) and sequence divergences between SARS-CoV and SARS-CoV-2

165   (**Supplemental Table 4**) onto the predicted interfaces. In general, conserved residues have been shown

166   to cluster at protein-protein interfaces[63], and a recent analysis of SARS-CoV-2 structure and evolution

167   likewise concluded that highly conserved surface residues were likely to drive protein-protein

168   interactions[64]. Consistent with these prior findings at an interactome-wide level, we observed significant

169   depletion for both viral and human variation along the predicted interfaces comparable to that observed

170   on solved human-human interfaces (**Fig 2.a**).

171   Nonetheless, considering each interaction individually, our analysis uncovered a 13 interaction

172   interfaces enriched for human population variants (**Fig 2.b**), and 7 enriched for recent viral sequence

173   divergences (**Fig 2.c**). A breakdown of variant enrichment on each interface is provide in **Supplemental

174   Table 5**. The individual viral interfaces showing an unexpected degree of variation may—like the

175   previously discussed S-ACE2 interface—be indicative of recent functional evolution around the viral-

176   human interaction. Considering the slower rate of evolution in humans, enrichment of population

177   variants along the human interfaces is unlikely to be a selective response to the virus. Rather, these

178   interfaces with high population variation along the interfaces may represent edges in the interactome

179   whose strength may fluctuate among individuals or between populations. Alternatively, enrichment and

180   depletion of variation along the human-viral interfaces could help distinguish viral proteins that bind

181   along existing—and therefore conserved—human-human interfaces from those that bind using novel

182   interfaces—that would be less likely to be under selective pressure.

183   To further explore the functional impact of naturally occurring variants on the human interactors

184   of SARS-CoV-2, we considered variants with phenotypic associations as reported in HGMD[65], ClinVar[66]

185   or the NHGRI-EBI GWAS Catalog[67]. Interactors of SARS-CoV-2 were significantly more likely than the

186   rest of the human proteome to harbor phenotypic variants in each of these databases (**Fig 2.d**). Notably,

187   among the individual disease categories enriched in this gene set, several were consistent with reported

188   comorbidities including heart disease, respiratory tract disease, and metabolic disease[68, 69] (**Fig 2.e**;

189   **Supplemental Table 6**). Disruption of native protein-protein interactions is one mechanism of disease

190   pathology, and disease mutations are known to be enriched along protein interfaces[70, 71]. Human

191   population variants on the predicted human-viral interface were more likely to be annotated as

192   deleterious by SIFT[72] and PolyPhen[73] but showed identical allele frequency distributions compared to

193   those off the interfaces (**Supplemental Figure 4**). However, mapping annotated disease mutations onto

194   the protein interfaces only revealed significant enrichment along known human-human interfaces; no

195   such enrichment was found on human-viral interfaces (**Fig 2.f**). This is likely because unlike with human-

196   human interactions, mutations disrupting human-viral interactions would not disrupt natural cell function,

197   and therefore would be unlikely to be pathogenic. Our finding that disease mutations and viral proteins

198   affect human proteins at distinct sites is consistent with a two-hit hypothesis of comorbidities whereby

199   proteins whose function is already affected by genetic background may be further compromised by viral

200   infection.

201   <u>Analysis of Binding Affinity Changes Between SARS-CoV and SARS-CoV-2 Helps Identify Key Interactions</u>

202   We next sought to explore the impact of sequence divergence in SARS-CoV-2 relative to SARS-CoV on

203   viral-human interactions. Mutations between the two viruses were identified by pairwise alignment and the

204   impacts of these mutations on the binding energy (ΔΔG) for 250 interactions amenable to docking were

205   predicted using a PyRosetta pipeline[46, 59, 60]. Although the binding energy for most interactions was

206   unchanged—either because no mutations occurred near the interface or because the mutations that did

207   had marginal effect—we observed an increased likelihood of the divergence from SARS-CoV to SARS-

208   CoV-2 resulting in decreased binding energy (i.e. more stable interaction) (**Fig 3.a**; **Supplemental Table**

209   **7**). The significant outliers in these ΔΔG predictions can help pinpoint key differences between the viral-

210   human interactomes of SARS-CoV and SARS-CoV-2. We further note a wide range of affinity impacts

211   among various human interactors of a single viral protein (**Fig 3.d**) and hypothesize that these differences

212   may help identify the most important interactions.

213   To further explore the significance of these changes in interaction affinity, we considered those

214   interactions with the largest decrease in binding energy; corresponding to the largest predicted increase in

215   affinity. Specifically, we highlight the interaction between coronavirus orf9c and human mitochondrial NADH

216   Dehydrogenase (Ubiquinone) 1 Alpha Subcomplex, Assembly Factor 1 (NDUFAF1) (**Fig 3.e**; ΔΔG=-21.7

217   REU). Two mutations contribute to the increased stability of this interaction in SARS-CoV-2. During the

218   transition from the unbound to bound states of orf9c, His-34 rotates inwards to accommodate NDUFAF1

219   resulting steric tensions with orf9c residue 61. The mutation from the bulkier Val-61 (SARS-CoV) to Ala-61

220   (SARS-CoV-2) helps alleviate this tension resulting in overall more favorable energy state transition.

221   Second, the mutation from the polar Glu-60 (SARS-CoV) to the nonpolar Met-60 (SARS-CoV-2) contributes

222   to overall better solvation energy along the interface involving residue 60 of orf9c and the Ser-148, Glu-

223   149, Val-150 stretch of NDUFAF1.

224   Although the precise role of coronavirus orf9c (sometimes annotated as orf14) has not yet been

225   characterized, experiments in SARS-CoV have shown that it localizes to vesicular components[74] and

226   interactome data from SARS-CoV-2 reveal that it targets host mitochondrial proteins and could impact

227   proteins responsible for modulating IkB kinase and NF-kB signaling pathways[20]. NDUFAF1 is a chaperone

228   protein involved in the assembly of the mitochondrial complex I[75, 76]. RNAi screens have associated

229   knockdown of NDUFAF1 with increased vaccinia virus infection[77]. Moreover, properly assembled

230   mitochondrial complex I—including NDUFAF1—has been shown to be indispensable for reactive oxygen

231   species based signaling to trigger expression of interleukins 2 and 4 and activate signaling of TLR7[78], a

232   SARS-CoV-2 ARDS predisposition gene[79]. Recent reports have suggested that the virus may hijack host

233   metabolism through complex I to increase replication[80]. Additionally, NDUFAF1 harbors disease mutations

234   for complex I deficiency and hypertrophic cardiomyopathy that may be linked to SARS-CoV-2 comorbidities.

235   Our result predicting increased stability of the interaction between orf9c and NDUFAF1 in SARS-CoV-2

236   relative to SARS-CoV suggests a stronger impact on complex I assembly and potential disruption of

237   downstream pathways.

238   We further identified several interactions with significantly altered binding energy in SARS-CoV-2

239   with potential links to cellular response to viral infection. In orf9c we additionally predicted increased affinity

240   with a Golgi signaling protein, TMED5 ($\Delta\Delta G$=-21 REU). Phenotypic screens have linked TMED5 to altered

241   viral infectivity[77, 81, 82] and decreased interleukin 8 secretion[83]. Although no known mechanisms directly link

242   TMED5 to immune response, one of its interactors, TMED2[84], is known to potentiate interferon response

243   to viral infection[85]. Interaction with TMED5 may additionally be linked to viral transport within and secretion

244   from its host[86]. Finally, the mitochondrial protein TOMM70 was predicted to bind orf9b with greater affinity

245   ($\Delta\Delta G$=-11 REU) in SARS-CoV-2. TOMM70 has been linked to mitochondrial antiviral signaling through

246   downstream activation of interferon regulatory factors[87].

247   <u>Predicting the Impact of Human Populations Variants on SARS-CoV-2-Human Protein Interactions</u>

248   A dynamic range of patient responses and symptoms have been reported for SARS-CoV-2 infection. In

249   previously studied viruses including HIV, underlying genetic variation can explain up to 15% of variation in

250   patient response and overall viral load[88]. Moreover, previous high-throughput experiments suggest that, up

251   to 10.5% of missense population variants can disrupt native protein-protein interactions[89]. Therefore, we

252   hypothesize that some fraction of patient response to SARS-CoV-2 can be explained by missense

253   variations and their impact on viral-human interactions. To explore this hypothesis, we performed in-silico

254   scanning mutagenesis along all docked interfaces in PyRosetta. We identified as binding energy hotspot

255   mutations all mutations with a predicted $\Delta\Delta G$ at least one standard deviation away from the mean for

256   identical amino acid substitutions across the rest of the interface. In total, out of 2,241 population variants

257   on eligible interfaces, 161 (7.2%) were identified as hotspots predicted to disrupt interaction stability, and

258   116 (5.2%) were identified as hotspots predicted to contribute to interaction stability (**Fig 3.b**). Most of the

259   hotspot mutations were predicted to be driven by solvation or repulsive forces, with disruptive hotspots

260   primarily being driven by repulsive forces and stabilizing hotspots primarily being driven by solvation forces

261   (**Fig 3.c**). Results summarizing the predicted impact of all 2,241 population variants on the docked

262   interfaces are provided in **Supplemental Table 8**.

263   Prioritization of Candidate Inhibitors of SARS-CoV-2-Human Interactions Through Binding Site Comparison

264   A Web Server to Present the SARS-CoV-2-Human 3D Structural Interactome

265   In order to present the results from our experiments as a comprehensive easy-access resource to the

266   general public, we constructed the SARS-CoV-2 human structural interactome web. All structures used

267   for modeling, interface predictions, raw docking outputs, mutation binding impacts, and analyses

268   described herein are directly available for download through our downloads page. Our homepage allows

269   users quick navigation through the reported interactome to view results summarizing specific

270   interactions of interest. Aside from providing a 3D view of the interaction and predicted interface

271   residues, our results page provides four main functionalities (**Fig 5**).

272   The interface comparison panel (**Fig 5 top left**) provides a linear representation of protein

273   sequence with interface annotations. Linear representations of all other known or predicted interfaces

274   from the same protein are shown for comparison along with navigation links to explore other interaction

275   partners either within the site (for viral-human interactions) or through InteractomeINSIDER[45] (for

276   human-human interactions). In particular, from the human perspective, this comparison may reveal

277   biologically meaningful insights about the interface overlap and possible competition between viral and

278   human interactors.

279   The mutations panel (**Fig 5 top right**) provides information on variation in each interaction

280   partner. For the viral side, mutations are reported based on divergence from the SARS-CoV version of

281   the protein. For the human side, all known population variants as reported in gnomAD are listed[58].

282   Additionally, we provide a log odds ratio describing the enrichment or depletion of variation along the

283   interface. These results may help highlight interactions undergoing functional evolution along the viral

284   interface.

285   For each interaction amenable to docking, the ΔΔG Information panel (**Fig 5 lower left**)

286   compiles the results from our in-silico scanning mutagenesis to report binding affinity predictions for all

287   possible mutations across the docked interface. Individual mutations and colored by their z-score

288   normalized ΔΔG prediction. The results can be toggled to only show the impacts of known variants. On

289   the viral side, a cumulative ΔΔG value is provided describing the predicted change in binding affinity

290   between the SARS-CoV and SARS-CoV-2 versions of the protein.

291   The current version of the SARS-CoV-2 human structural interactome web server describes 332

292   viral-human interactions reported by Gordon *et al*.[20]. We will continue support for the web server with

293   periodic updates as additional interactome screens between SARS-CoV-2 and human are published.

294   As we update, a navigation option to select between the current or previous stable releases of the web

295   server will be provided.

296   **Discussion**

297   Overall, we present a comprehensive resource to explore the SARS-CoV-2-human protein-protein

298   interactome map in a structural context. Analysis through this framework allows us to consider the recent

299   evolution of SARS-CoV-2 in the context of its interactome map and to prioritize for further functional

300   characterization key interactions. Likewise, our consideration of underlying variation in the human

301   proteins that interact with SARS-CoV-2 may be valuable in explaining differences in response to

302   infection. We particularly note that our observation that perturbation from underlying disease mutations

303   and viral protein binding occur at distinct sites on the protein is of clinical interest. Further investigation

304   into the combined role of these two sources of perturbation to better understand the mechanisms linked

305   to comorbidities is warranted.

306   However, our work is not without limitation. Firstly, we note that although structural coverage

307   from our homology modelling of SARS-CoV-2 proteins was robust (**Supplemental Figure 1**), the same

308   could not be universally said of the human proteins. Although guided molecular docking was always

309   done to orient the most likely interface residues on each structure towards each other, protein-protein

310   docking using incomplete protein models introduces some bias and low coverage may exclude some

311   true interface residues. For this reason, the initial ECLAIR interface annotations—which are less subject

312   to structural coverage limitations—may provide orthogonal value. We additionally note that direct

313   quantitative interpretation of predicted ΔΔG values using the Rosetta scoring function is often difficult

314   since different term weights can be used in different setups. Although the same scoring function was

315   used for all predictions described here, the relative magnitude of each term may change based on the

316   size and composition of different proteins between interactions. For these reasons, we only employ a

317   relative qualitative comparison of similar predictions when interpreting our scanning mutagenesis

318   results. Moreover, the structure optimization after each mutation is applied focuses on side-chain

319   repacking, therefore, our results focus only on mutations at or near the interface where the impact of

320   side-chain repacking could be measured. We expect there may be some mutations with significant

321   impact on binding affinity that act through refolding or other allosteric effects that are missed by our

322   method.

323   Perhaps most importantly, we emphasize the importance of further experimental

324   characterization to confirm the predictions made here. Nonetheless we believe our 3D Structural SARS-

325   CoV-2-human Interactome web server will prove to be a key resource in informing hypothesis driven

326   exploration of the mechanisms of SARS-CoV-2 pathology and host response. The scope, and potential

327   impacts of our webserver will continue to grow as we incorporate the results of ongoing and future

328   interactome screens between SARS-CoV-2 and human. Additionally, we note that our 3D structural

329   interactome framework can be rapidly deployed to analyze future viruses.

330   **Online Methods**

331   Generation of Homology Models for SARS-CoV-2 Proteome

332   Homology-based modeling of all 29 SARS-CoV-2 proteins was performed in Modeller[90] using a multiple

333   template modeling procedure. In brief, a list of candidate template structures for each protein to be

334   modelled was obtained by running BLAST[91] against a reference containing all sequences in the Protein

335   Data Bank (PDB)[92]. Templates were filtered to only retain those with at least 30% identify to the protein

336   to be modelled, and the remaining templates were ranked using a weighted combination of percent

337   identity and coverage as described previously[93]. To compile the final set of overlapping templates for

338   modeling, first the top ranked template was selected as a seed. Overlapping templates were iteratively

339   added to the set so long as 1) the new template increased the overall coverage by at least 10%, and 2)

340   the new template retained a total percent identity no more than 25% worse than the initial seed template.

341   Pairwise alignments between the protein to be modelled and the template set were generated using a

342   Modeller alignment object with default settings. Alignments were manually trimmed to remove any

343   regions with large gaps (at least 5 gaps in the alignment in a 10 residue window). Finally, alignment was

344   carried out using the Modeller automodel function.

345   This approach generated suitable homology models for 18 out of 29 proteins. A visual

346   representation of each structure and templates used is provided in **Supplemental Figure 1**.

347   Phase One Interface Prediction Using ECLAIR

348   Interface predictions for all 332 interactions reported by Gordon *et al.*[20] were made in two phases. In

349   phase one, we leveraged our previously validated **E**nsemble **C**lassifier **L**earning **A**lgorithm to

350   predict **I**nterface **R**esidues (ECLAIR)[45] to perform an initial prediction of likely interface residues across

351   all interactions. ECLAIR compiles five sets of features; biophysical, conservation, coevolution,

352   structural, and docking. In brief, biophysical features are compiled using a windowed average of several

353   ExPASy ProtScales[94], conservation features are derived from the Jensen-Shannon divergence[95, 96]

354   across all available homologs for each protein, coevolution features between interacting proteins are

355   derived from direct coupling analysis (DCA)[97] and statistical coupling analysis (SCA)[98] among paired

356   homologs, structural features are obtained by calculating the solvent accessible surface area of

357   available PDB[92] or ModBase[99] models using NACCESS[100], and docking features are derived from an

358   average of docking predictions made using Zdock[101].

359   To accommodate predictions between SARS-CoV-2 and human, slight alterations were made.

360   First calculation of coevolution features was impossible because the calculation of SCA and DCA

361   requires analysis of multiple sequence alignments from paired homologs of both interacting proteins

362   across at least 50 species. Because we here consider inter-species interaction, no one species could

363   contain homologs of both the viral and human proteins. Second, the calculation of conservation features

364   for the viral proteins were modified to account for conservation between both related viral species and

365   various strains that have been sequenced in a single species. We typically only include one homolog

366   per species in these calculations, but expanded our criteria here because availability of the protein

367   conservation feature is a requirement for all of our higher-confidence classifiers. Finally, the calculation

368   of structural features for the viral protein were overruled to use the manually provided homology models

369   instead of pulling structures from the PDB or ModBase. A visual summary of the ECLAIR interface

370   predictions is presented in **Supplemental Figure 2** and the initial prediction results are provided in

371   **Supplemental Table 1**.

372   <u>Phase Two Interface Prediction Using Guided PyRosetta Docking</u>

373   Interface predictions for all 332 interactions reported by Gordon *et al.*[20] were made in two phases. In

374   phase two, we generated atomic resolution models of 250 interactions by leveraging the Rosetta scoring

375   function[51] and prior probabilities obtained from ECLAIR predictions to perform guided docking. The

376   remaining 82 interactions were missing reliable 3D models for at least one of their members and

377   therefore were not amenable to docking. A schematic summary of our docking methodology is presented

378   in **Supplemental Figure 3**.

379   Using the predicted interface probabilities reported by ECLAIR, we set up the initial docking

380   conformation to explore a restricted search space for each docking simulation. In cases where multiple

381   structures were available for the human protein, all structures were weighted based on the ECLAIR

382   scores for the covered residues in each structure to maximize both coverage age inclusion of likely

383   interface residues. For each protein in the interaction, we performed a linear regression classification in

384   scikit-learn[102] to optimally separate the likely interface residues from likely non-interface residues. The

385   plane defined by this linear regression served as a reference to orient the structures along the y-axis

386   with their most probable interface sides facing each other. The two chains were centered at (0, 0) along

387   the x-z plane and separated a distance of 5 Å along the y-axis. For each docking attempt, a series of

388   random perturbations from these initial conformations were made to search the nearby space. First, the

389   human protein was rotated up to 360° along the y-axis to allow full exploration of different rotations of

390   the two interfaces relative to each other. Second, to apply some flexibility to the plane predicting the

391   interface vs. non-interface sides of each protein, up to 30° of rotation along the x- and z- axis were

392   allowed for both the viral and human proteins. Finally, a random translation up to 5 Å in magnitude was

393   applied to the human protein along the x-z plane so that the docking could explore contact points other

394   than the center of masses along these axes.

395   After initializing these guided starting conformations, docking was simulated in PyRosetta[46]

396   using a modified version of the protein-protein docking methodology provided by Gray 2006[103]. The

397   initial demo (https://graylab.jhu.edu/pyrosetta/downloads/scripts/demo/D100_Docking.py) takes two

398   chains from a co-crystal structure, applies a random perturbation, and re-docks them. Because

399   randomized initial orientation was already handled as described previously, these steps were removed

400   from our docking runs. In brief, the protein models were converted to centroid representation, slid into

401   contact using the "interchain_cen" scoring function, and converted back to full atom representation,

402   before having their side-chains optimized using the predefined "docking" and "docking_min" scoring

403   functions. Up to 50 iterations of this guided docking were performed for each interaction, and the docked

404   conformation resulting in the lowest Rosetta energy score was retained. The final docked interface

405   annotations are provided in **Supplemental Table 2**.

406   <u>Definition of Interface Residues</u>

407   To annotate interface residues from atomic resolution docked models, we used a previously described

408   and established definition for interface residues[45]. In brief, the solvent accessible surface area (SASA)

409   for both bound and unbound docked structures was calculated using NACCESS.[100] We define as

410   interface residue, any residue that is both 1) at the surface of a protein (defined as ≥ 15% relative

411   accessibility) and 2) in contact with the interacting chain (defined as any residue whose absolute

412   accessibility decreased by ≥ 1.0 Å^2).

413   <u>Compilation of Viral SARS-CoV to SARS-CoV-2 Mutations and Human Population Variants</u>

414   For analysis of genetic variation that may impact the viral-human interactome, two sets of mutations

415   were compiled; 1) viral mutations between SARS-CoV and SARS-CoV-2, and 2) human population

416   variants on proteins shown to interact with viral proteins.

417   For viral mutations, variations between SARS-CoV and SARS-CoV-2 versions of each proteins

418   were collected. The representative sequences used for the 16 proteins in the SARS-CoV proteome were

419   taken from the Proteomes section of UniProt (ID UP000000354)[104, 105]. Sequences for 29 SARS-CoV-2

420   proteins were obtained from the annotations by Gordon *et al.*[20] based on genbank accession

421   MN985325[106, 107]. Notably, UniProt accessions for the SARS-CoV proteome report two sequences for

422   the uncleaved ORF1a and ORF1a-b which correspond to NSP1 through NSP16 in SARS-CoV-2.

423   Variations between SARS-CoV and SARS-CoV-2 were reported after pairwise Needleman Wench

424   alignment[108, 109] (using Blosum62 scoring matrix, gap open penalty of 10 and gap extension penalty of

425   0.5) between the corresponding protein sequences in each species. A total of 1,003 missense variants

426   were detected among 23 SARS-CoV-2 proteins. No variations were reported for orf3b, orf8, or orf10

427   because no suitable alignment could be made with a SARS-CoV sequence. Additionally, orf7b, nsp3,

428   and nsp16 were excluded from this set because they were not involved in any viral-human interactions.

429   The full list of SARS-CoV-2 mutations is reported in **Supplemental Table 4**.

430   Human population variants in all 332 human proteins shown to interact with SARS-CoV-2

431   proteins were obtained from gnomAD[58]. Programmatic queries to fetch all variants for a given gene were

432   performed using gnomAD's graphQL API. For details on performing gnomAD queries in this manner see

433   the   gnomad-api   github   page   (https://github.com/broadinstitute/gnomad-

434   browser/tree/master/projects/gnomad-api). Population variants were filtered to only retain missense

435   variants. In order to map these gnomAD DNA-level SNPs to equivalent protein-level UniProt

436   annotations, we used the Ensembl Variant Effect Predictor (VEP)[110]. After all mapping was completed,

437   all variants were parsed to make sure that the reported reference amino acid and position matched up

438   with the UniProt sequence. Roughly 95.6% of cases matched and the remaining variants that could not

439   reliably be mapped to UniProt coordinates were dropped from our dataset. In total 127,528 human

440   population variants were curated. The full list of human population variants from GnomAD is reported in

441   **Supplemental Table 3**.

442   <u>Log Odds Enrichment Calculations</u>

443   In order to determine if viral mutations or human populations variants were enriched at predicted

444   interaction interfaces, odds ratios were calculated as described previously[111]. All odds ratios were $\log_2$

445   transformed to maintain symmetry between enriched and depleted values. For this particular use case,

446   the equation for the odds ratio was…

447
$$OR = \log_2 \frac{VariantInterfaceResidues \,/\, VariantNonInterfaceResidues}{NonVariantInterfaceResidues \,/\, NonVariantNonInterfaceResidues}$$

448   where "Variant" vs. "NonVariant" distinguishes positions in a protein sequence that respectively did and

449   did not overlap with the set of variants, and "InterfaceResidues" vs. "NonInterfaceResidues"

450   distinguishes positions in a protein sequence that respectively did and did not overlap with the predicted

451   interface for a given interaction. To avoid arbitrary odds ratio inflation or depletion from missing data, in

452   all cases where the interface residues were predicted by molecular docking, the odds ratio was altered

453   to only account for positions that were included in the structural models used for docking. Enrichment

454   calculations for disease mutations and overlap between drug and protein binding sites were calculated

455   in the same manner with appropriately adjusted case and exposure categories.

456   <u>Curation of Disease Associated Variants</u>

457   To explore whether human proteins interacting with SARS-CoV-2 proteins were enriched for disease or

458   trait associated variants, three datasets were curated; the Human Gene Mutation Database (HGMD)[65],

459   ClinVar[66], and the NHGRI-EBI GWAS Catalog[67]. Disease annotations from HGMD and ClinVar were

460   obtained directly from their respective downloads pages and mapped to UniProt. For overall enrichment

461   of individual disease terms among all human proteins interacting with SARS-CoV-2, disease terms were

462   linked in an ontology based on the NCBI MedGen term relationships

463   (https://ftp.ncbi.nlm.nih.gov/pub/medgen/MGREL.RRF.gz). When calculating enrichment, counts for

464   each term were propagated through all parent nodes up to a singular root node. Significant terms were

465   reported as the most general term with no more significant ancestor term (**Supplemental Table 6**, sheet

466   1). Raw enrichment values for all terms are also predicted (**Supplemental Table 6**, sheet 2).

467   For curation of disease and trait associations from NHGRI-EBI GWAS Catalog

468   (http://www.ebi.ac.uk/gwas/)[67], lead SNPs (p-value<5e-8) for all diseases/traits were retrieved on June

469   16, 2020. Proxy SNPs in high linkage disequilibrium (LD) (Parameters: $R^2 > 0.8$; pop: "ALL") for

470   individual lead SNPs were obtained through programmatic queries to the LDproxy API[112], which used

471   phase 3 haplotype data from the 1000 Genomes Project as reference for calculating pairwise metrics

472   of LD. Both lead SNPs and proxy SNPs were filtered to only retain missense variants.

473   Gene-level enrich calculations for disease / trait associated mutations in known interactors for

474   SARS-CoV-2 and variant-level enrichment at the predicted interaction interfaces (either human-human

475   interactions or human-viral interactions) were performed as described above.

476   <u>Estimation of ΔΔG from variation at the interfaces</u>

477   In order to predict the impact of variation at viral-human interaction interfaces on binding affinity, two

478   sets of ΔΔG predictions were made using PyRosetta[46]. First, to explore the overall binding energy

479   contributions of each interface residue, and to predict the impact of all possible mutations along the

480   interface, a scanning mutation ΔΔG approach was implemented based on protocols described

481   previously[59, 60]. The implementation is essentially identical to a demo provided by the PyRosetta

482   documentation (https://graylab.jhu.edu/pyrosetta/downloads/scripts/demo/D090_Ala_scan.py) with one

483   major exception: interface residues in the original demo are defined using an 8.0 Å inter-chain distance

484   cutoff, which we overrule with our definition described above. In brief, we iterate over all interface

485   residues for a given docked structure. For each position, the binding for the interaction from the wildtype

486   structure is first estimated. The energy for the complex state is estimated following a

487   PackRotamersMover optimization constrained to only adjust the side-chains of residues within 8.0 Å of

488   the residue to be mutated. To estimate energy from the unbound state, the chains are first separated

489   500.0 Å to eliminate any interchain energy contributions, the structures are then optimized and scored

490   identically to the complex state. The difference between these two values provides the binding energy

491   for the wildtype structure.

492    $$\Delta G_{WT} = E_{complex} - E_{unbound}$$

493   To estimate the binding energy for all 19 amino acid mutations possible at the given position, each

494   mutation is made iteratively, and the ΔG_{Mut} is estimated identically to the wildtype except using the

495   mutated structures. Finally, the overall impact on binding energy for each mutation is calculated as the

496   difference between these two binding energies.

497    $$\Delta\Delta G = \Delta G_{Mut} - \Delta G_{WT}$$

498   The scoring function used for these calculations is as described previously[59] using the following weights;

499   fa_atr=0.44, fa_rep=0.07, fa_sol=1.0, hbond_bb_sc=0.5, hbond_sc=1.0. To account for the

500   stochasticity of the PackRotamersMover optimization between trials, all $\Delta\Delta G$ values are reported from

501   an average of 10 independent trials which yielded overall low standard deviation between trials. To

502   check whether an individual mutation's average $\Delta\Delta G$ was significantly non-zero, a two-sided z-test

503   between the 10 independent trials was performed. To account for average impact other same amino

504   acid mutations at other positions along the interface, each average $\Delta\Delta G$ was z-normalized relative to

505   the rest of the interface and outliers were called at ≥ 1 standard deviation away from the mean. Mutations

506   that passed both criteria were identified as significant interface binding affinity hotspots.

507   Second, estimates of the overall impact of the cumulative set of mutations between SARS-CoV

508   and SARS-CoV-2 were made using the same general framework. For each interaction, the binding

509   energy for the SARS-CoV-2 version of the interaction was estimated identically to the wildtype binding

510   energy. The binding energy for the SARS-CoV version of the interaction was estimated after applying

511   all mutations between the two viruses (for simplicity, only amino acid substitutions were applied, a

512   minority of mutations that comprised insertions or deletions could not be modelled). The only difference

513   compared to the single mutation $\Delta\Delta G$ calculation was that multiple mutations were applied at once, and

514   consequentially, the interface packing was done allowing side-chain rotamer optimization for all residues

515   within 8.0 Å of any of the mutated residues. The $\Delta\Delta G$ values were calculated such that an interaction

516   predicted to be more stable (lower binding energy) in the SARS-CoV-2 version of the interaction

517   compared to the SARS-CoV version of the interaction would have a negative $\Delta\Delta G$.

518 $$\Delta\Delta G = \Delta G_{SARSCoV2} - \Delta G_{SARSCoV1}$$

519 To account to stochasticity between trials for these predictions (which notably had a much larger impact

520 likely due to the decreased constraints on rotamer optimization in these cases), this set of ΔΔG values

521 was reported as an average of 50 trials. Significant outliers for overall binding affinity change from

522 SARS-CoV to SARS-CoV-2 were called based on similar criteria to the individual mutations, except the

523 z-score normalization was performed relative to all other interactions.

524 <u>Protein-ligand Docking Using Smina</u>

525 The previous viral-human interactome screen by Gordon *et al.*[20] reported 76 candidate drugs targeting

526 the 332 human proteins. In order to further prioritize this list and search for drugs that share a binding

527 site with the viral interactor, we performed protein-ligand docking. Among this list, 71 interaction-drug

528 pairs involving 49 unique drugs that were amenable to docking were identified. To prep for docking, 3D

529 structures for all ligands were first generated using Open Babel[113] and the command:

530
```
obabel -:"[SMILES_STRING]" --gen3d -opdb -O [OUT_FILE] -d
```

531 Protein-ligand docking was executed using smina[114] with the following parameters. The autobox_ligand

532 option was turned on and centered around the receptor PDB file with an autobox_add border size of 10

533 Å. The exhaustiveness was set to 40 to increase the number of independent stochastic sampling

534 trajectories and increase the likelihood of identifying a global minimum. The num_modes was set to

535 1000 to so that a multitude of lower ranked models could be compared for internal purposes. To reduce

536 real wall time each docking process was run using 5 CPU cores, although it should be noted runtime is

537 affected in a strictly linear manner and CPU cores has no impact on net CPU time. Finally, each protein-

538 ligand docking command was repeated 45 times with a unique seed on each interaction. The final smina

539 command used was as follows:

540
```
smina -r [RECEPTOR] -l [LIGAND] --autobox_ligand [RECEPTOR] --autobox_add 10
```

541
```
  -o [OUT_FILE] --exhaustiveness 40 --num_modes 1000 --cpu 5 --seed [SEED]
```

542 Multiple trials for the same docking problem were performed in order to saturate the ligand binding

543 search space as thoroughly as possible. Practically speaking this batched docking approach was only

544 taken so that time spent docking would cycle through each protein-drug pair equally, thus producing

545   results from each pair at an even rate. The net effect is merely as if one docking process had been run

546   for each protein-drug pair with exhaustiveness parameter set to ~2,000. It should be noted that this level

547   of sampling was overkill relative to common practices and usually a single run with exhaustiveness

548   ranging from 30-50 would be suffient[114]. Docked conformation from all iterations on each protein-drug

549   pair were compiled into a final set of up to 10 of the best scoring poses. To retain poses that cover

550   different low-energy binding sites, these poses were selected such that the center of mass of each

551   docked pose added must be at least 1 Å away from the center of mass of any of the higher ranked

552   poses. All results described in this manuscript are reported based on only the top ranked pose for each

553   protein-drug pair. Protein residues involved in the drug binding site were annotated using the same

554   criteria used to define interface residues. Of note, to include ligand molecules as part of the NACCESS

555   solvent accessible surface area calculations, the Record Type for all ligand atoms must be manually

556   changed from HETATM to ATOM.

References

1.   *COVID-19 Dashboard by the Center for Systems Science and Engineering at Johns Hopkins University*. 2020 [cited 2020 August 26, 2020]; Available from: https://coronavirus.jhu.edu/map.html.

2.   Fehr, A.R. and S. Perlman, *Coronaviruses: an overview of their replication and pathogenesis*. Methods Mol Biol, 2015. **1282**: p. 1-23.

3.   Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin*. Nature, 2020. **579**(7798): p. 270-273.

4.   McIntosh, K. and S. Perlman, *Coronaviruses, Including Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS)*. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, 2015: p. 1928-1936.e2.

5.   Zhou, H., et al., *A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein*. Curr Biol, 2020. **30**(11): p. 2196-2203 e3.

6.   Gupta, A., et al., *Extrapulmonary manifestations of COVID-19*. Nat Med, 2020. **26**(7): p. 1017-1032.

7.   Wang, D., et al., *Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China*. JAMA, 2020.

8.   Yang, X., et al., *Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study*. Lancet Respir Med, 2020. **8**(5): p. 475-481.

9.   Zhou, F., et al., *Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study*. Lancet, 2020. **395**(10229): p. 1054-1062.

10.   Palaiodimos, L., et al., *Severe obesity, increasing age and male sex are independently associated with worse in-hospital outcomes, and higher in-hospital mortality, in a cohort of patients with COVID-19 in the Bronx, New York.* Metabolism, 2020. **108**: p. 154262.

11.   Ferdinand, K.C. and S.A. Nasser, *African-American COVID-19 Mortality: A Sentinel Event.* J Am Coll Cardiol, 2020. **75**(21): p. 2746-2748.

12.   Killerby, M.E., et al., *Characteristics Associated with Hospitalization Among Patients with COVID-19 - Metropolitan Atlanta, Georgia, March-April 2020.* MMWR Morb Mortal Wkly Rep, 2020. **69**(25): p. 790-794.

13.   Raisi-Estabragh, Z., et al., *Greater risk of severe COVID-19 in Black, Asian and Minority Ethnic populations is not explained by cardiometabolic, socioeconomic or behavioural factors, or by 25(OH)-vitamin D status: study of 1326 cases from the UK Biobank.* J Public Health (Oxf), 2020.

14.   Moore, J.T., et al., *Disparities in Incidence of COVID-19 Among Underrepresented Racial/Ethnic Groups in Counties Identified as Hotspots During June 5-18, 2020 - 22 States, February-June 2020.* MMWR Morb Mortal Wkly Rep, 2020. **69**(33): p. 1122-1126.

15.   Mahajan, U.V. and M. Larkins-Pettigrew, *Racial demographics and COVID-19 confirmed cases and deaths: a correlational analysis of 2886 US counties.* J Public Health (Oxf), 2020. **42**(3): p. 445-447.

16.   Pfefferle, S., et al., *The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors.* PLoS Pathog, 2011. **7**(10): p. e1002331.

17.   Jager, S., et al., *Global landscape of HIV-human protein complexes.* Nature, 2011. **481**(7381): p. 365-70.

18.   Batra, J., et al., *Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication.* Cell, 2018. **175**(7): p. 1917-1930 e13.

19.   Shah, P.S., et al., *Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis.* Cell, 2018. **175**(7): p. 1931-1945 e18.

20.   Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.* Nature, 2020. **583**(7816): p. 459-468.

21.   Niemann, H.H., et al., *Structure of the human receptor tyrosine kinase met in complex with the Listeria invasion protein InlB.* Cell, 2007. **130**(2): p. 235-46.

22.   Hoffmann, M., et al., *SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.* Cell, 2020. **181**(2): p. 271-280 e8.

23.   Xu, G.G., J. Guo, and Y. Wu, *Chemokine receptor CCR5 antagonist maraviroc: medicinal chemistry and clinical applications.* Curr Top Med Chem, 2014. **14**(13): p. 1504-14.

24.   Hayouka, Z., et al., *Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium.* Proc Natl Acad Sci U S A, 2007. **104**(20): p. 8316-21.

25.   Peat, T.S., et al., *Small molecule inhibitors of the LEDGF site of human immunodeficiency virus integrase identified by fragment screening and structure based design.* PLoS One, 2012. **7**(7): p. e40147.

26.   Maginnis, M.S., *Virus-Receptor Interactions: The Key to Cellular Invasion.* J Mol Biol, 2018. **430**(17): p. 2590-2611.

27. Daczkowski, C.M., et al., *Structural Insights into the Interaction of Coronavirus Papain-Like Proteases and Interferon-Stimulated Gene Product 15 from Different Species.* J Mol Biol, 2017. **429**(11): p. 1661-1683.

28. Yao, J., et al., *Mechanism of inhibition of retromer transport by the bacterial effector RidL.* Proc Natl Acad Sci U S A, 2018. **115**(7): p. E1446-E1454.

29. Zhang, L., et al., *Solution structure of the complex between poxvirus-encoded CC chemokine inhibitor vCCI and human MIP-1beta.* Proc Natl Acad Sci U S A, 2006. **103**(38): p. 13985-90.

30. Jonker, H.R., et al., *Structural properties of the promiscuous VP16 activation domain.* Biochemistry, 2005. **44**(3): p. 827-39.

31. Card, G.L., et al., *Crystal structure of a gamma-herpesvirus cyclin-cdk complex.* EMBO J, 2000. **19**(12): p. 2877-88.

32. Smith, M., R. Honce, and S. Schultz-Cherry, *Metabolic Syndrome and Viral Pathogenesis: Lessons from Influenza and Coronaviruses.* J Virol, 2020.

33. Vidal, M., *A unifying view of 21st century systems biology.* FEBS Lett, 2009. **583**(24): p. 3891-4.

34. Robinson, C.V., A. Sali, and W. Baumeister, *The molecular sociology of the cell.* Nature, 2007. **450**(7172): p. 973-82.

35. Barabasi, A.L., N. Gulbahce, and J. Loscalzo, *Network medicine: a network-based approach to human disease.* Nat Rev Genet, 2011. **12**(1): p. 56-68.

36. Scott, D.E., et al., *Small molecules, big targets: drug discovery faces the protein-protein interaction challenge.* Nat Rev Drug Discov, 2016. **15**(8): p. 533-50.

37. Arkin, M.R., Y. Tang, and J.A. Wells, *Small-molecule inhibitors of protein-protein interactions: progressing toward the reality.* Chem Biol, 2014. **21**(9): p. 1102-14.

38. Rooklin, D., et al., *AlphaSpace: Fragment-Centric Topographical Mapping To Target Protein-Protein Interaction Interfaces.* J Chem Inf Model, 2015. **55**(8): p. 1585-99.

39. Lampson, B.L. and M.S. Davids, *The Development and Current Use of BCL-2 Inhibitors for the Treatment of Chronic Lymphocytic Leukemia.* Curr Hematol Malig Rep, 2017. **12**(1): p. 11-19.

40. *VENCLEXTA combination regimens for CLL work through 2 distinct cytotoxic mechanisms of action*. 2019; Available from: https://www.venclextahcp.com/cll/venclexta-efficacy/mechanism-of-action.html.

41. Schormann, N., et al., *Identification of protein-protein interaction inhibitors targeting vaccinia virus processivity factor for development of antiviral agents.* Antimicrob Agents Chemother, 2011. **55**(11): p. 5054-62.

42. White, P.W., et al., *Inhibition of human papillomavirus DNA replication by small molecule antagonists of the E1-E2 protein interaction.* J Biol Chem, 2003. **278**(29): p. 26765-72.

43. Goudreau, N., et al., *Optimization and determination of the absolute configuration of a series of potent inhibitors of human papillomavirus type-11 E1-E2 protein-protein interaction: a combined medicinal chemistry, NMR and computational chemistry approach.* Bioorg Med Chem, 2007. **15**(7): p. 2690-700.

44. Brito, A.F. and J.W. Pinney, *Protein-Protein Interactions in Virus-Host Systems.* Front Microbiol, 2017. **8**: p. 1557.

45. Meyer, M.J., et al., *Interactome INSIDER: a structural interactome browser for genomic studies.* Nat Methods, 2018. **15**(2): p. 107-114.

46. Chaudhury, S., S. Lyskov, and J.J. Gray, *PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta.* Bioinformatics, 2010. **26**(5): p. 689-91.

47. Kirchdoerfer, R.N., et al., *Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis.* Sci Rep, 2018. **8**(1): p. 15701.

48. Wang, Q., et al., *Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2.* Cell, 2020. **181**(4): p. 894-904 e9.

49. Wrobel, A.G., et al., *SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects.* Nat Struct Mol Biol, 2020. **27**(8): p. 763-767.

50. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.* Cell, 2020. **181**(2): p. 281-292 e6.

51. Alford, R.F., et al., *The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design.* J Chem Theory Comput, 2017. **13**(6): p. 3031-3048.

52. Shang, J., et al., *Structural basis of receptor recognition by SARS-CoV-2.* Nature, 2020. **581**(7807): p. 221-224.

53. Wrapp, D., et al., *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.* Science, 2020. **367**(6483): p. 1260-1263.

54. Jordan, R.E. and P. Adab, *Who is most likely to be infected with SARS-CoV-2?* The Lancet Infectious Diseases, 2020. **20**(9): p. 995-996.

55. Cao, Y., et al., *Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations.* Cell Discov, 2020. **6**: p. 11.

56. Darbeheshti, F. and N. Rezaei, *Genetic predisposition models to COVID-19 infection.* Med Hypotheses, 2020. **142**: p. 109818.

57. Zhao, Y., et al., 2020.

58. Karczewski, K.J., et al., *The mutational constraint spectrum quantified from variation in 141,456 humans.* Nature, 2020. **581**(7809): p. 434-443.

59. Kortemme, T. and D. Baker, *A simple physical model for binding energy hot spots in protein-protein complexes.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14116-21.

60. Shulman-Peleg, A., et al., *Spatial chemical conservation of hot spot interactions in protein-protein complexes.* BMC Biol, 2007. **5**: p. 43.

61. Stawiski, B. and T. Kania, *Tests of Concrete Strength across the Thickness of Industrial Floor Using the Ultrasonic Method with Exponential Spot Heads.* Materials (Basel), 2020. **13**(9).

62. Procko, E., *The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2.* bioRxiv, 2020.

63. Guharoy, M. and P. Chakrabarti, *Conserved residue clusters at protein-protein interfaces and their use in binding site identification.* BMC Bioinformatics, 2010. **11**: p. 286.

64. Gupta, R., et al., *SARS-CoV2 (COVID-19) Structural/Evolution Dynamicome: Insights into functional evolution and human genomics.* bioRxiv, 2020.

65. Stenson, P.D., et al., *Human Gene Mutation Database (HGMD): 2003 update.* Hum Mutat, 2003. **21**(6): p. 577-81.

66. Landrum, M.J., et al., *ClinVar: improving access to variant interpretations and supporting evidence.* Nucleic Acids Res, 2018. **46**(D1): p. D1062-D1067.

67.   Buniello, A., et al., *The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D1005-D1012.

68.   Killerby, M.E., et al., *Characteristics Associated with Hospitalization Among Patients with COVID-19 - Metropolitan Atlanta, Georgia, March-April 2020.* MMWR Morb Mortal Wkly Rep, 2020. **69**(25): p. 790-794.

69.   Yang, J., et al., *Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis.* Int J Infect Dis, 2020. **94**: p. 91-95.

70.   Sahni, N., et al., *Widespread macromolecular interaction perturbations in human genetic disorders.* Cell, 2015. **161**(3): p. 647-660.

71.   Wang, X., et al., *Three-dimensional reconstruction of protein networks provides insight into human genetic disease.* Nat Biotechnol, 2012. **30**(2): p. 159-64.

72.   Sim, N.L., et al., *SIFT web server: predicting effects of amino acid substitutions on proteins.* Nucleic Acids Res, 2012. **40**(Web Server issue): p. W452-7.

73.   Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.

74.   von Brunn, A., et al., *Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome.* PLoS One, 2007. **2**(5): p. e459.

75.   Vogel, R.O., et al., *Human mitochondrial complex I assembly is mediated by NDUFAF1.* FEBS J, 2005. **272**(20): p. 5317-26.

76.   Janssen, R., et al., *CIA30 complex I assembly factor: a candidate for human complex I deficiency?* Hum Genet, 2002. **110**(3): p. 264-70.

77.   Sivan, G., et al., *Human genome-wide RNAi screen reveals a role for nuclear pore proteins in poxvirus morphogenesis.* Proc Natl Acad Sci U S A, 2013. **110**(9): p. 3519-24.

78.   Kaminski, M.M., et al., *Mitochondrial reactive oxygen species control T cell activation by regulating IL-2 and IL-4 expression: mechanism of ciprofloxacin-mediated immunosuppression.* J Immunol, 2010. **184**(9): p. 4827-41.

79.   van der Made, C.I., et al., *Presence of Genetic Variants Among Young Men With Severe COVID-19.* JAMA, 2020.

80.   Tutuncuoglu, B., et al., *The Landscape of Human Cancer Proteins Targeted by SARS-CoV-2.* Cancer Discov, 2020. **10**(7): p. 916-921.

81.   Lipovsky, A., et al., *Genome-wide siRNA screen identifies the retromer as a cellular entry factor for human papillomavirus.* Proc Natl Acad Sci U S A, 2013. **110**(18): p. 7452-7.

82.   Filone, C.M., et al., *The master regulator of the cellular stress response (HSF1) is critical for orthopoxvirus infection.* PLoS Pathog, 2014. **10**(2): p. e1003904.

83.   Warner, N., et al., *A genome-wide small interfering RNA (siRNA) screen reveals nuclear factor-kappaB (NF-kappaB)-independent regulators of NOD2-induced interleukin-8 (IL-8) secretion.* J Biol Chem, 2014. **289**(41): p. 28213-24.

84.   Huttlin, E.L., et al., *Architecture of the human interactome defines protein communities and disease networks.* Nature, 2017. **545**(7655): p. 505-509.

85.   Sun, M.S., et al., *TMED2 Potentiates Cellular IFN Responses to DNA Viruses by Reinforcing MITA Dimerization and Facilitating Its Trafficking.* Cell Rep, 2018. **25**(11): p. 3086-3098 e3.

86.   Sicari, D., et al., *Role of the early secretory pathway in SARS-CoV-2 infection.* J Cell Biol, 2020. **219**(9).

87.   Liu, X.Y., et al., *Tom70 mediates activation of interferon regulatory factor 3 on mitochondria.* Cell Res, 2010. **20**(9): p. 994-1011.

88.   Fellay, J., et al., *A whole-genome association study of major determinants for host control of HIV-1.* Science, 2007. **317**(5840): p. 944-7.

89.   Fragoza, R., et al., *Extensive disruption of protein interactions by genetic variants across the allele frequency spectrum in human populations.* Nat Commun, 2019. **10**(1): p. 4141.

90.   Eswar, N., et al., *Comparative protein structure modeling using Modeller.* Curr Protoc Bioinformatics, 2006. **Chapter 5**: p. Unit-5 6.

91.   Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.

92.   Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.

93.   Mosca, R., A. Ceol, and P. Aloy, *Interactome3D: adding structural details to protein networks.* Nat Methods, 2013. **10**(1): p. 47-53.

94.   Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server.* (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005), 2005: p. 571-607.

95.   Lin, J., *Divergence measures based on the Shannon entropy.* IEEE Transactions on Information Theory, 1991. **37**(1): p. 145-151.

96.   Capra, J.A. and M. Singh, *Predicting functionally important residues from sequence conservation.* Bioinformatics, 2007. **23**(15): p. 1875-82.

97.   Morcos, F., et al., *Direct coupling analysis for protein contact prediction.* Methods Mol Biol, 2014. **1137**: p. 55-70.

98.   Lockless, S.W. and R. Ranganathan, *Evolutionarily conserved pathways of energetic connectivity in protein families.* Science, 1999. **286**(5438): p. 295-9.

99.   Pieper, U., et al., *ModBase, a database of annotated comparative protein structure models and associated resources.* Nucleic Acids Res, 2014. **42**(Database issue): p. D336-46.

100.   Lee, B. and F.M. Richards, *The interpretation of protein structures: estimation of static accessibility.* J Mol Biol, 1971. **55**(3): p. 379-400.

101.   Pierce, B.G., Y. Hourai, and Z. Weng, *Accelerating protein docking in ZDOCK using an advanced 3D convolution library.* PLoS One, 2011. **6**(9): p. e24657.

102.   Pedregosa, F., et al., *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 2011. **12**(1532-4435): p. 2825–2830.

103.   Gray, J.J., *High-resolution protein-protein docking.* Curr Opin Struct Biol, 2006. **16**(2): p. 183-93.

104.   UniProt, C., *UniProt: a worldwide hub of protein knowledge.* Nucleic Acids Res, 2019. **47**(D1): p. D506-D515.

105.   He, R., et al., *Analysis of multimerization of the SARS coronavirus nucleocapsid protein.* Biochem Biophys Res Commun, 2004. **316**(2): p. 476-83.

106.   Wu, F., et al., *A new coronavirus associated with human respiratory disease in China.* Nature, 2020. **579**(7798): p. 265-269.

107.   Chan, J.F., et al., *Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan.* Emerg Microbes Infect, 2020. **9**(1): p. 221-236.

108. Needleman, S.B. and C.D. Wunsch, *A general method applicable to the search for similarities in the amino acid sequence of two proteins.* J Mol Biol, 1970. **48**(3): p. 443-53.

109. Henikoff, S. and J.G. Henikoff, *Amino acid substitution matrices from protein blocks.* Proc Natl Acad Sci U S A, 1992. **89**(22): p. 10915-9.

110. McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.

111. Szumilas, M., *Explaining odds ratios.* J Can Acad Child Adolesc Psychiatry, 2010. **19**(3): p. 227-9.

112. Machiela, M.J. and S.J. Chanock, *LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants.* Bioinformatics, 2015. **31**(21): p. 3555-7.

113. O'Boyle, N.M., et al., *Open Babel: An open chemical toolbox.* J Cheminform, 2011. **3**: p. 33.

114. Koes, D.R., M.P. Baumgartner, and C.J. Camacho, *Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise.* J Chem Inf Model, 2013. **53**(8): p. 1893-904.

557   **Supplemental Table Headings**

558   **Supplemental Table 1. List of ECLAIR predicted interface residues.**

559   **Supplemental Table 2. List of guided docking annotated interface residues.**

560   **Supplemental Table 3. List of human population variants reported by gnomAD.**

561   **Supplemental Table 4. List of sequence divergences between SARS-CoV and SARS-CoV-2.**

562   **Supplemental Table 5. Enrichment for sequence variation on SARS-CoV-2-human interfaces.**

563   **Supplemental Table 6. Enrichment for individual disease terms in human interactors of SARS-**

564   **CoV-2.**

565   **Supplemental Table 7. Predicted ΔΔG between SARS-CoV and SARS-CoV-2 versions of all**

566   **docked interactions.**

567   **Supplemental Table 8. Predicted ΔΔG impact of all human population variants at the interface.**

568   **Supplemental Table 9. List of all predicted drug-target binding sites .**

Figure 1
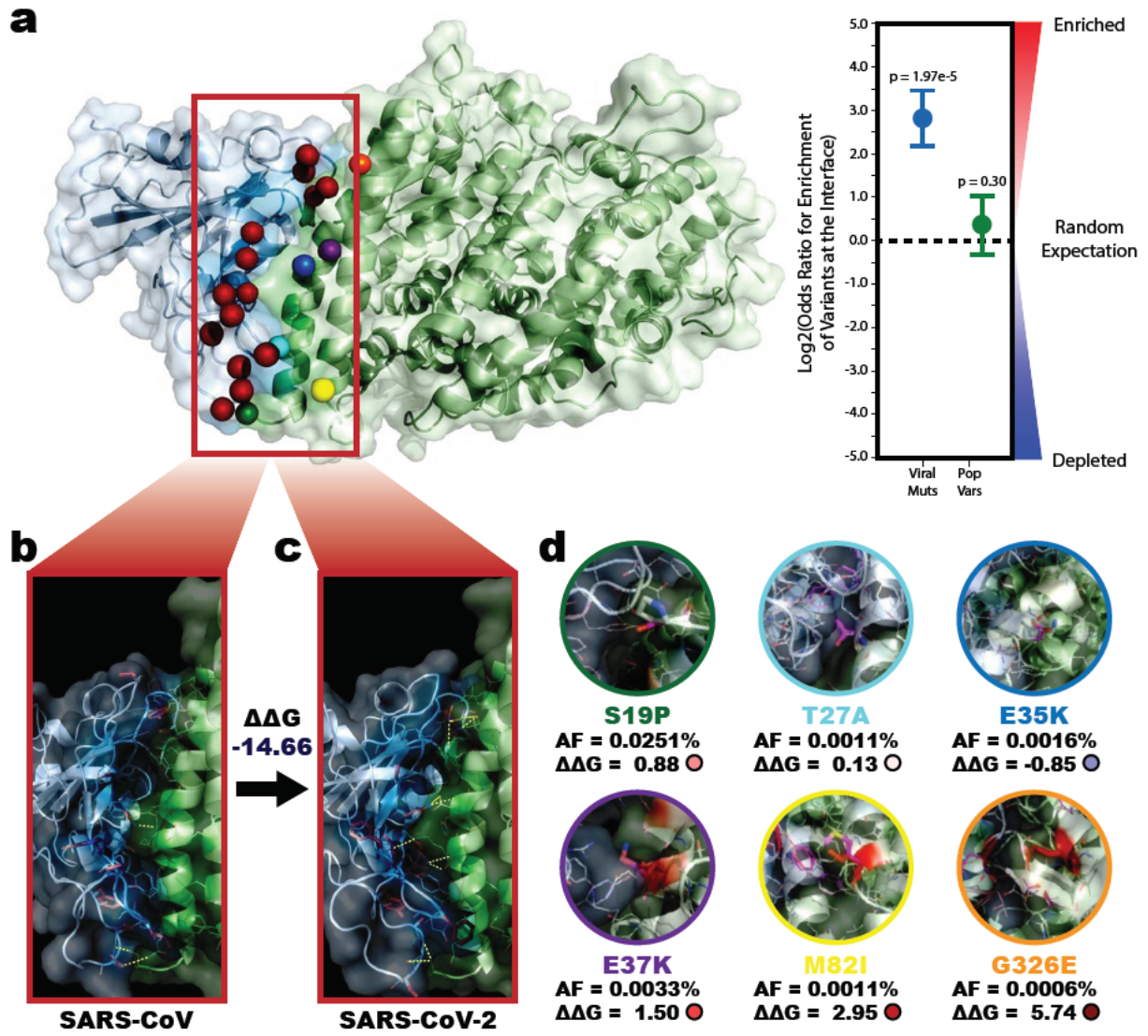
569 **Figure 1. Enrichment and predicted impact of divergences between SARS-CoV and SARS-CoV-2**

570 **along the S-ACE2 interface.**

571 **a,** Co-crystal structure of the interaction between SARS-CoV-2 Spike protein (S) with human ACE2

572 (PDB 6LZG). All 15 sequence divergences between SARS-CoV and SARS-CoV-2 Spike protein

573 interfaces are highlighted as red spheres while all 6 population variants on the ACE2 protein interface

574 are highlighted as green (ACE2_S19P), cyan (ACE2_T27A), blue (ACE2_E35K), purple (ACE2_E37K),

575 yellow (ACE2_M82I), and orange (ACE2_G326E) spheres. Enrichment of these variants on the interface

576 are reported for SARS-CoV-2 (Log2OR=2.82, p=1.97e-5) and human (Log2OR=0.38, p=0.30) shown to

577 the right. Error bars indicate ± SE. **b, c,** Zoomed in interface views for the SARS-CoV S-ACE2 structure

578 (PDB 6CS2) and SARS-CoV-2 S-ACE2 structure (PDB 6LZG). Sequence divergences between the two

579 Spike proteins are highlighted as red sticks. Inter-protein polar contacts that contribute to stabilizing the

580 interaction are shown as yellow dashed lines. The binding energy (ΔG) of each interaction was

581 estimated using PyRosetta and the change in this binding energy ($\Delta\Delta G$) is reported. The negative value

582 ($\Delta\Delta G$=-14.66 Rosetta Energy Units (REU)) indicates the interaction is more stable (lower energy) in the

583 SARS-CoV-2 version of the interaction. **d,** Comparison of the impact of each of the ACE2 population

584 variants. Mutated structures containing the population variant are shown over the original structure

585 (magenta). The mutated residue is shown as sticks. Residues whose predicted impact on binding energy

586 was affected are colored from blue (decreased $\Delta\Delta G$) to white (no change) to red (increased $\Delta\Delta G$). The

587 gnomAD reported allele frequency and predicted $\Delta\Delta G$ for each mutation are reported. Outlines of each

588 structures are consistent with the spheres in **a**.

Figure 2
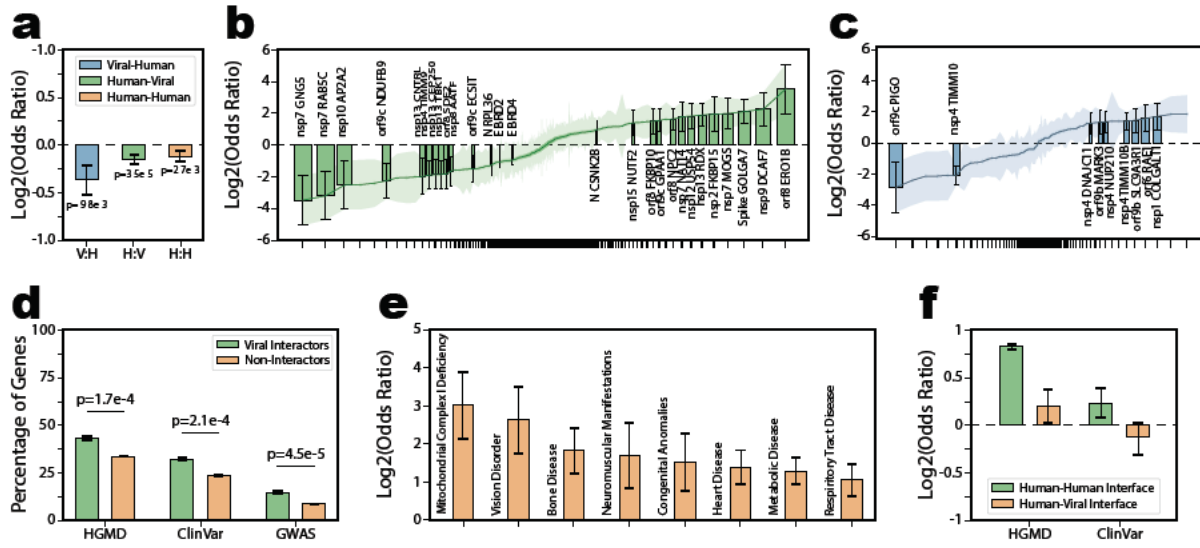
589 **Figure 2. Enrichment of sequence divergences and disease mutations across all SARS-CoV-2-**

590 **Human interaction interfaces.**

591 **a,** Overall enrichment for viral sequence divergence along viral-human (V:H, Log2OR=-0.37, p=9.76e-

592 3) interfaces and human population variants along human-viral (H:V, Log2OR=-0.16, p=3.47e-5) or

593 human-human (H:H, Log2OR=-0.13, p=2.67e-3) interfaces at an interactome level. Error bars indicate

594 ± SE. **b, c,** Individual enrichment for human population variants and viral sequence divergences

595 respectively on all SARS-CoV-2-human interaction interfaces. Interactions are sorted from most

596 depleted to most enriched. The x-axis is skewed to compress non-significant interactions. Only

597 interfaces with statistically significant enrichment or depletion are shown as bars and labeled. The

598 remainder are shown as Log2OR (line) and error bars (colored area). Error bars indicate ± SE. **d,**

599 Comparison of the percentage of human genes that interact with (green) or do not interact with (orange)

600 SARS-CoV-2 that contain disease annotations in HGDM (Log2OR=0.57, p=1.70e-4), ClinVar

601 (Log2OR=0.64, p=1.05e-4), and GWAS (Log2OR=0.89, p=4.54e-5) respectively. Genes targeted by

602 SARS-CoV-2 proteins were significantly more likely to harbor disease mutations than non-interactors.

603 Error bars indicate ± SE. **e,** A sample of individual disease terms enriched in human genes targeted by

604 SARS-CoV-2. Full results are reported in **Supplemental Table 6**. Error bars indicate ± SE. **f,**

605 Comparison of the enrichment of HGDM, ClinVar, and GWAS annotated mutations on human-vial

606 interfaces or human-human interfaces for the same gene set. Although disease mutations were enriched

607 on human-human interfaces (HGMD, Log2OR=0.82, p<1e-20; ClinVar, Log2OR=-0.13, p=0.24), no

608 enrichment was observed on human-viral interfaces (HGMD, Log2OR=0.21, p=0.13; ClinVar,

609 Log2OR=0.51, p<e-20). The GWAS category was removed from this analysis because most lead GWAS

610 SNPs occurred in non-coding regions. Error bars indicate ± SE.

Figure 3
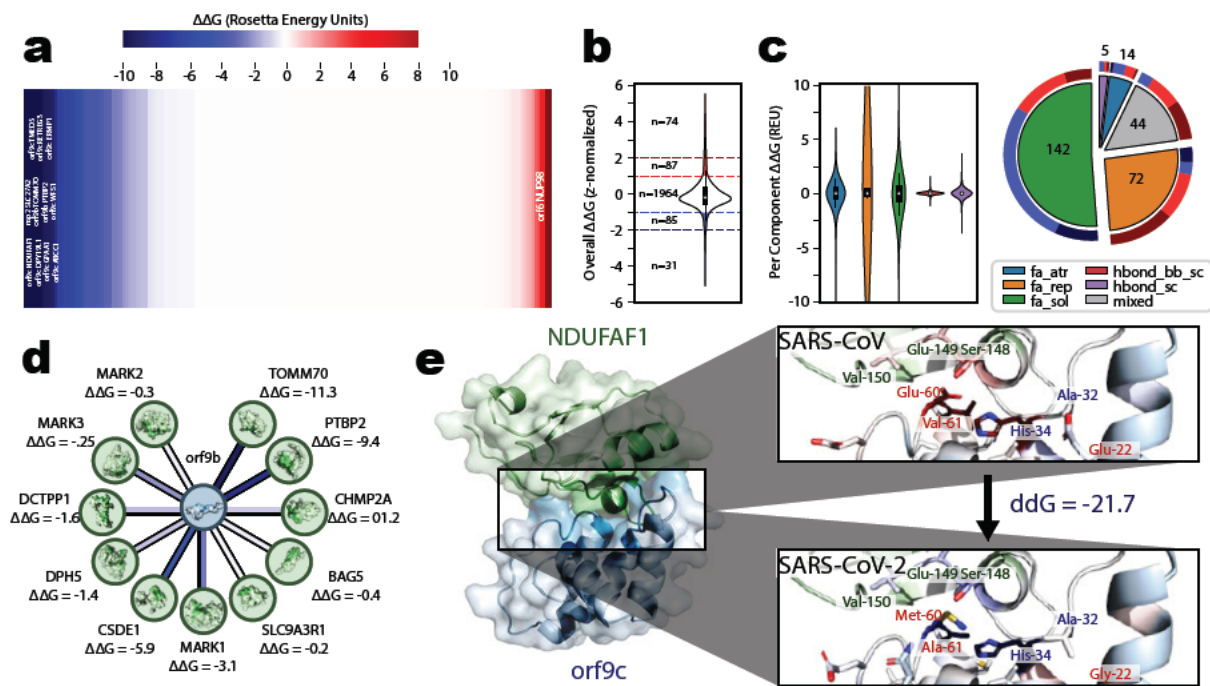
611   **Figure 3. Predicted impact of sequence divergences on the binding affinity of SARS-CoV-2-**

612   **Human interactions.**

613   **a,** Predicted changes in binding affinity from sequence divergences between SARS-CoV and SARS-

614   CoV-2 on SARS-CoV-2-human interactions were predicted in PyRosetta. An overall representation of

615   these ΔΔG predictions is reported (mean=-1.40 REU, std=6.16 REU) with interactions sorted from those

616   with the largest decrease in binding energy (most stabilized relative to SARS-CoV) to those with the

617   largest increase in binding energy (most destabilized relative to SARS-CoV). Outlier interactions with

618   the change in binding energy at least one standard deviation from the mean are labeled. **b,** Distribution

619   of the predicted binding affinity change from all human population variants on a SARS-CoV-2-human

620   interface. Values were z-score normalized across each interface. Interface binding energy hotspots

621   were binned as strongly disruptive (z-score ≥ 2, n=74), disruptive (1 ≤ z-score < 2, n=87), stabilizing (-

622   2 < z-score ≤ -1, n=85), or strongly stabilizing (z-score ≤ -2, n=31). All other population variants (-1 < z-

623   score < 1, n=1,964) showed minimal impact of binding affinity. **c,** Breakdown of the contribution of each

624   term in the PyRosetta energy function used for in-silico scanning mutagenesis for all population variants.

625   Distributions for each term are shown on the left. A breakdown of which term contributed most heavily

626   to the classification of all 277 interface hotspot population variants is shown on the right. **d,** Individual

627   SARS-CoV-2-human interactions involving the same viral protein can have distinct interfaces with

628   distinct predicted changes in binding affinity between SARS-CoV and SARS-CoV-2 versions of the

629   protein. An example involving orf9b is highlighted where some interactions (e.g. TOMM70 and PTBP2)

630   are predicted to be more stabilized in SARS-CoV-2 whereas others (e.g. BAG5, SLC9A3R1, and

631   MARK2) are predicted to me unaffected. **e,** Docked structure for the interaction between SARS-CoV-2

632   orf9c and human NDUFAF1, alongside comparisons of the predicted interface using SARS-CoV (top)

633   or SARS-CoV-2 (bottom) orf9c. Interface residues are colored by their predicted energy contribution

634   from blue (stabilizing) to white (no impact) to red (destabilizing). Residues that differ between SARS-

635   CoV and SARS-CoV-2 are labeled in red, while other residues with a major contribution to the binding

636   affinity are labeled in green (NDUFAF1) or blue (orf9c). The overall predicted change in binding energy

637   (ΔΔG=-21.7 REU) suggests the interaction is more stable (lower energy) in the SARS-CoV-2 version of

638   the interaction.

639 **Figure 4. Drug Docking and Prioritization of SARS-CoV-2-Human Interaction Inhibitors**
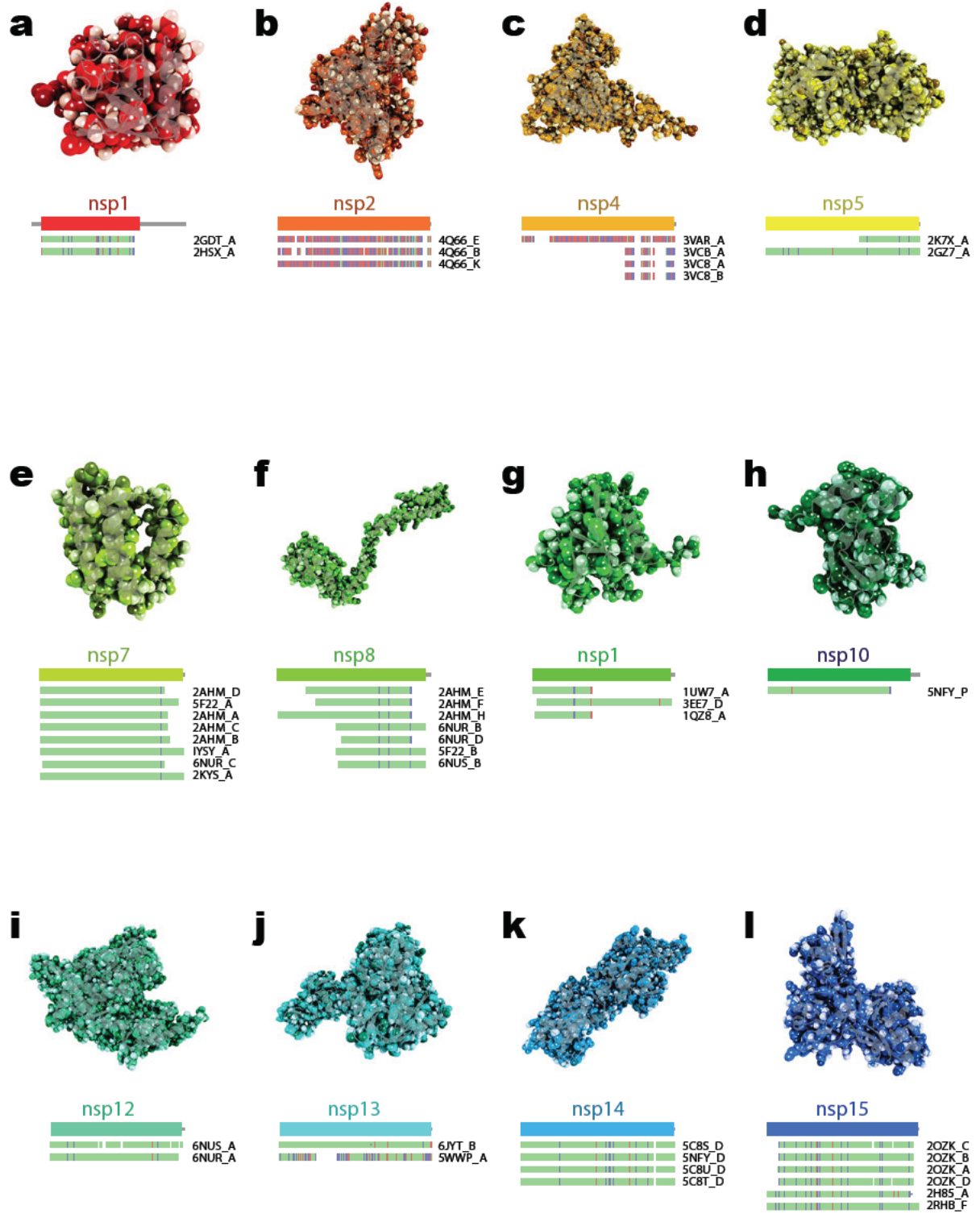
Figure 5

640 **Figure 5. 3D-SARS2 Structural Interactome Browser Overview.**

641 Overview of the main results page for exploring a given interaction in our 3D-SARS2 structural

642 interactome browser. The main display contains information for both the SARS-CoV-2 and human

643 proteins including structural displays for either the docked or single crystal structures as well as a table

644 summarizing the interface residues for both proteins. Interface residues are colored dark blue and dark

645 green for the viral and human proteins respectively. By default the page will display the docked structure

646 if available. The display can be toggled between docked structures and single structures using the

647 button in the bottom middle. When single structures display is selected residues will instead be colored

648 based on the initial ECLAIR interface definition. Four categories of expandable panels containing

649 additional analyses are provided. **upper left,** The interface view shows a linear representation of the

650 protein sequence with interface residues annotated in dark blue or dark green. Interfaces for other

651 interactors of the protein are shown underneath for easy comparison. **upper right,** The mutations panel

652 summarizes either human population variants or viral sequence divergences on the protein. Mutations

653 on the interface are labeled. **lower left,** The ΔΔG information panel summarizes the results from in-

654 silico mutagenesis scanning along the interface. Results for each mutation are z-score normalized

655 relative to the rest of the interface and colored on a blue (negative ΔΔG, stabilizing) to yellow (minimal

656 impact) to red (positive ΔΔG, destabilizing). The heatmap can be filtered to only show values
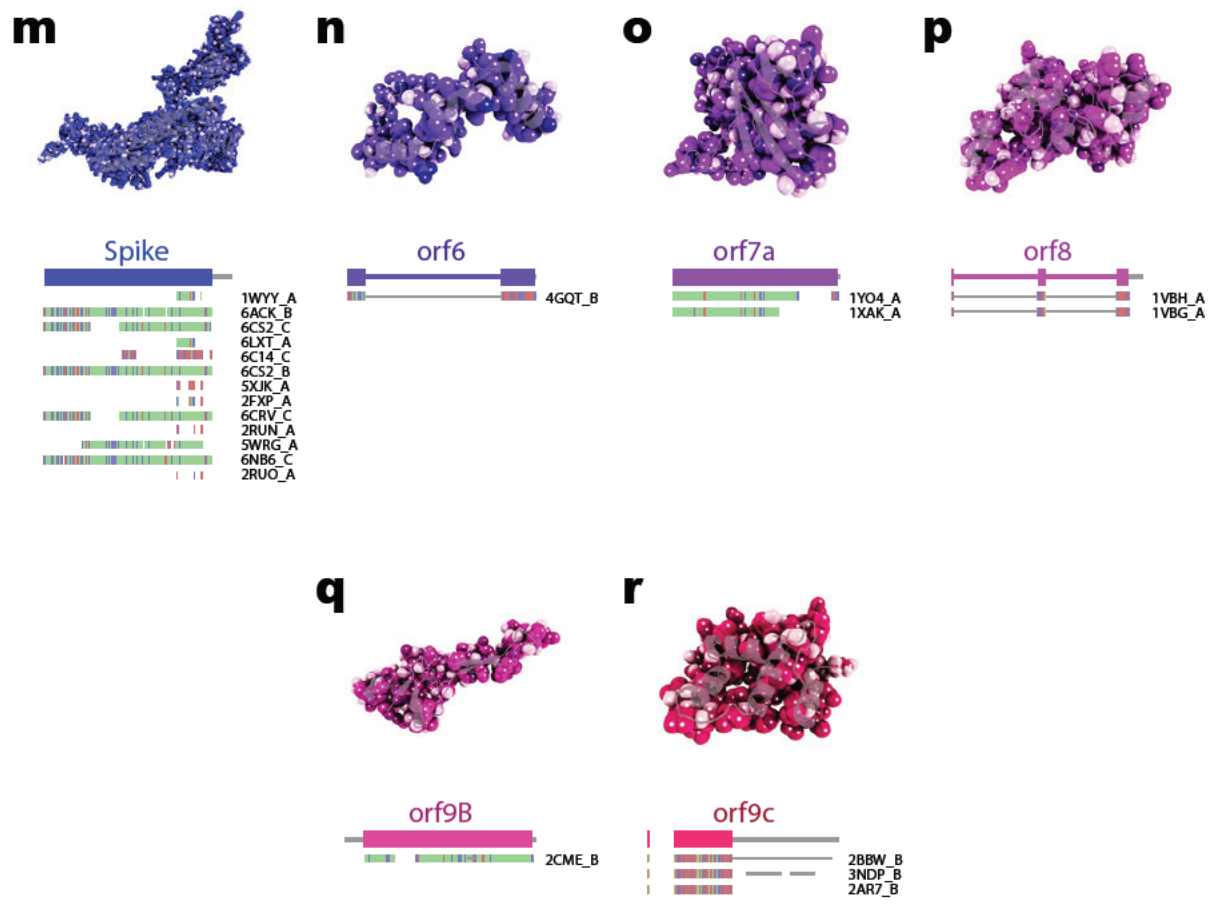
657 corresponding to known mutations on the interface.

Supplemental Figure 1

a nsp1

2GDT_A
2HSX_A

b nsp2

4Q66_E
4Q66_B
4Q66_K

c nsp4

3VAR_A
3VCB_A
3VCB_A
3VC8_B

d nsp5

2K7X_A
2GZ7_A

e nsp7

2AHM_D
5F22_A
2AHM_A
2AHM_C
2AHM_B
IYSY_A
6NUR_C
2KYS_A

f nsp8

2AHM_E
2AHM_F
2AHM_H
6NUR_B
6NUR_D
5F22_B
6NUS_B

g nsp1

1UW7_A
3EE7_D
1QZ8_A

h nsp10

5NFY_P

i nsp12

6NUS_A
6NUR_A

j nsp13

6JYT_B
5WWP_A

k nsp14

5C8S_D
5NFY_D
5C8U_D
5C8T_D

l nsp15

2OZK_C
2OZK_B
2OZK_A
2OZK_D
2H85_A
2RHB_F

Supplemental Figure 1

m Spike

1WYY_A
6ACK_B
6CS2_C
6LXT_A
6C14_C
6CS2_B
5XJK_A
2FXP_A
6CRV_C
2RUN_A
5WRG_A
6NB6_C
2RUO_A

n orf6

4GQT_B

o orf7a

1YO4_A
1XAK_A

p orf8

1VBH_A
1VBG_A

q orf9B

2CME_B

r orf9c

2BBW_B
3NDP_B
2AR7_B

658   **Supplemental Figure 1. Homology modeling for SARS-CoV-2 proteins.**

659   **a-r,** Homology models for 18 SARS-CoV-2 proteins amenable to homology modeling. The first bar under

660   each model indicates the overall coverage of the SARS-CoV-2 protein. The underlying bars indicate

661   template utilization across the structure. The template bars are colored based on the alignment between

662   the template and the corresponding SARS-CoV-2 where green indicates identical sequence, blue

663   indicates a mismatch with positive BLOSUM substitution score, and red indicates a mismatch with a

664   negative BLOSUM substitution score.

Supplemental Figure 2
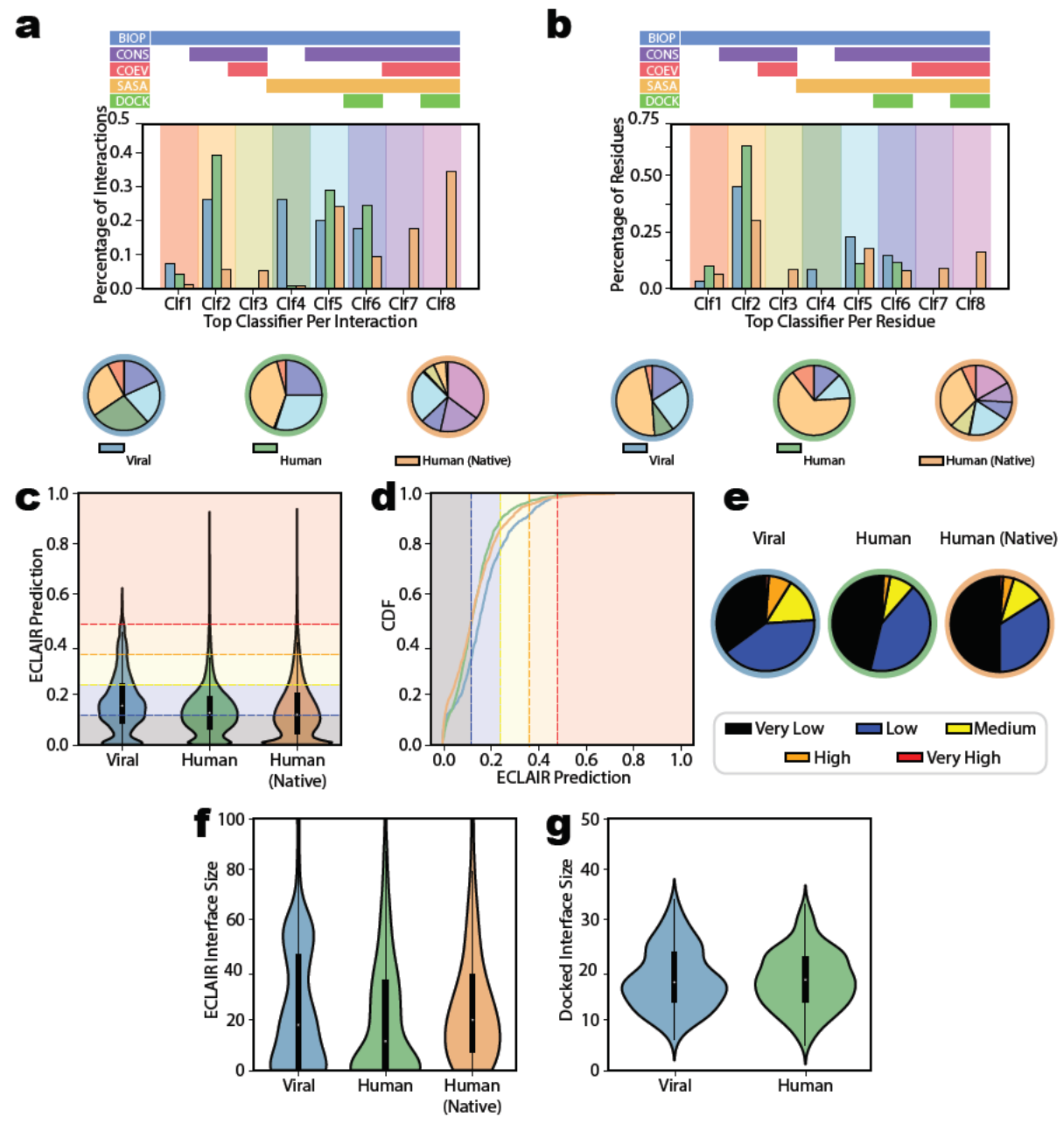
665 **Supplemental Figure 2. Summary of ECLAIR and guided docking interface predictions.**

666 **a,** Breakdown of best ECLAIR classifier used in interface prediction per interaction. The top section

667 indicates which features are used by each classifier. The total breakdown of classifier usage for

668 prediction of viral interfaces (blue) and human interfaces on human-viral (green) or native human-human

669 (orange) interactions are shown either in bar plot (middle) or as pie charts (bottom). **b,** Identical

670 breakdown, but reported at a per-residue rather than per-interaction level. **c, d,** Summary of ECLAIR

671 prediction scores across viral, human, and native human-human interfaces presented as either a raw

672 distribution or a cumulative density respectively. Plots are colored based on corresponding ECLAIR

673 confidence bins in each range of the plot. **e,** Breakdown of the overall ECLAIR confidence bin

674 representation among the three classes. Overall, ECLAIR classification predictions were generally

675 similar between the three classes of interface, although a higher fraction of predictions made on viral

676 interfaces were made with high or medium confidence. **f,** Distribution of the size of all interfaces defined

677 by ECLAIR compared among the three classes. **g,** Distribution of viral or human interfaces defined by

678 guided docking.

Supplemental Figure 3
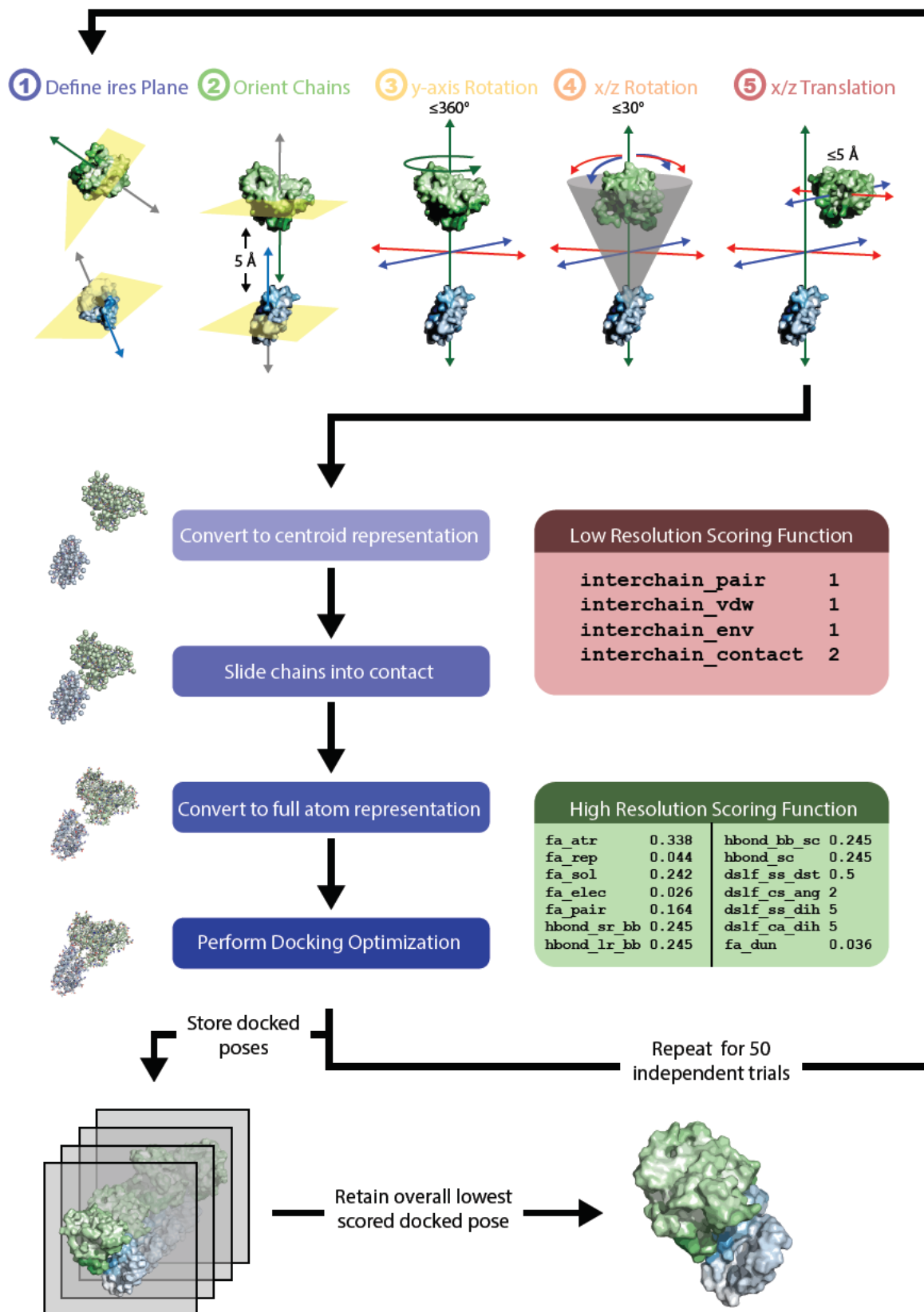
679 **Supplemental Figure 3. Visual representation of the guided docking protocol used.**

680 For each SARS-CoV-2-human interaction with 3D structure available for both proteins, 50 independent

681 guiding docking trials were used to select a final docked configuration. The structure for the viral protein

682 is colored from white to blue with darker blue corresponding to higher ECLAIR prediction. The structure

683 for the human protein is colored similarly using a green to white gradient. Initial semi-random docked

684 configurations were generated using five steps. First a plane separating ECLAIR predicted likely

685 interface from likely non-interface residues was drawn to divide each protein. Second, the two protein

686 chains were separated 5 Å apart on the y-axis using the previously defined plane to orient the likely

687 interface sides of each protein towards each other. Third, the human protein was randomly rotated up

688 to 360° along the y-axis to sample different orientations of the two interfaces relative to each other.

689 Fourth, the human protein was randomly rotated up to 30° along the x- and z-axes with the point of

690 rotation centered on the viral protein. Fifth a random translation up to 5 Å was applied to the human

691 protein along the x- and z-axes. The first two steps constitute the ECLAIR-based guiding of the docking

692 space. The last three steps serve to randomly perturb the initial docking configurations to sample the

693 space near the ECLAIR predicted interface. After this initial perturbation docking is performed using a

694 combination of low resolution and high-resolution scoring. During the low-resolution scoring, the proteins

695 are initially converted to a centroid representation and slid into contact. During the high-resolution

696 scoring, the proteins are converted back to a full-atom representation and the contact and side-chain

697 packing is optimized. The best (lowest) scored docked pose is retained and used for docked interface

698 definition.

Supplemental Figure 4
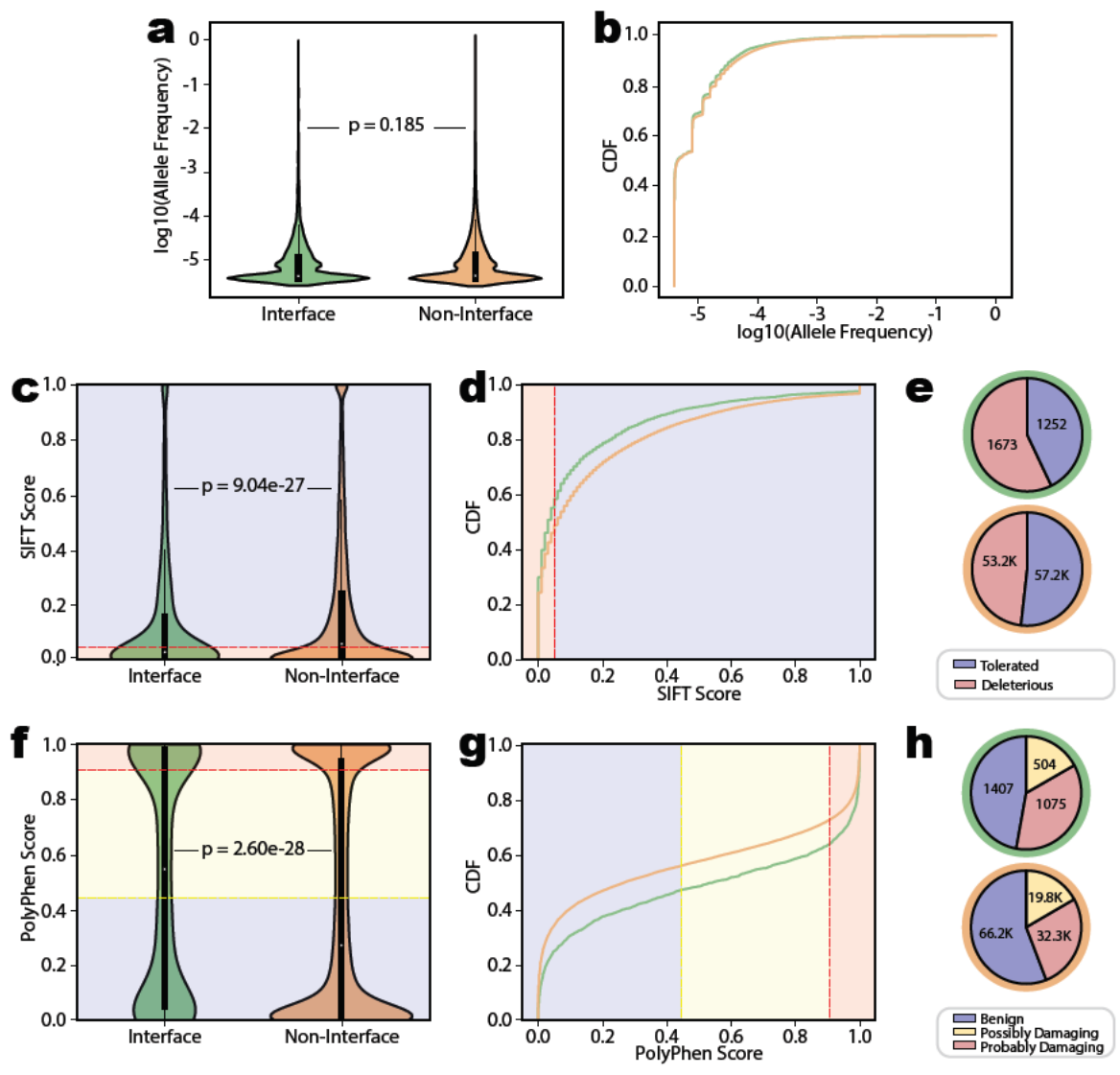
699 **Supplemental Figure 4. Summary of human population variant frequency and deleteriousness.**

700 **a, b,** Summary of allele frequency for human population variants either on or off the predicted human-

701 viral interface presented as either a raw distribution or a cumulative density respectively. Variants in

702 either category had roughly identical allele frequency distributions. **c, d,** Summary of the SIFT

703 deleteriousness score for human population variants either on or off the predicted human-viral interface

704 presented as either a raw distribution or a cumulative density respectively. Plots are colored based on

705 the split between SIFT tolerated and deleterious categories. **e,** Pie chart breakdown of these categories.

706 Pie char outlines distinguish interface (green) from non-interface (orange). Population variants on the

707 interface were significantly more likely to be classified deleterious. **f, g,** Summary of the PolyPhen

708 deleteriousness score for human population variants either on or off the predicted human-viral interface

709 presented as either a raw distribution or a cumulative density respectively. Plots are colored based on

710 the split between PolyPhen benign, possibly damaging, and probably damaging categories. **e,** Pie chart

711 breakdown of these categories. Pie char outlines distinguish interface (green) from non-interface

712 (orange). Population variants on the interface were significantly more likely to be classified probably

713 damaging.