Investigation of the genetic variation in ACE2 on the structural recognition by the novel coronavirus (SARS-CoV-2)

Xingyi Guo[1*], Zhishan Chen[1], Yumin Xia[2], Weiqiang Lin[3*], Hongzhi Li[4*]

Affiliations:

[1]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center; Department of Biomedical Informatics and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville 37203, TN, USA.

[2]Department of Dermatology, The Second Affiliated Hospital of Xi'an Jiaotong University, 157 Xiwu Road, Xi'an 710004, China.

[3]The Fourth Affiliated Hospital, Institute of Translational Medicine, Zhejiang University School of Medicine, Jinhua 322000, China.

[4]Department of molecular medicine, City of Hope National Medical Center, Duarte 91008, California, USA.

* Corresponding author contact information:
Dr. Xingyi Guo

Vanderbilt University School of Medicine

xingyi.guo@vumc.org

Dr. Weiqiang Lin

Zhejiang University School of Medicine

wlin@zju.edu.cn

Dr. HongzhiLi

City of Hope National Medical Center

holi@coh.org

Abstract

Background: The outbreak of coronavirus disease (COVID-19) was caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), through its surface spike glycoprotein (S-protein) recognition on the receptor Angiotensin-converting enzyme 2 (ACE2) in humans. However, it remains unclear how genetic variations in ACE2 may affect its function and structure, and consequently alter the recognition by SARS-CoV-2.

Methods: We have systemically characterized missense variants in the gene ACE2 using data from the Genome Aggregation Database (gnomAD; N = 141,456). To investigate the putative deleterious role of missense variants, six existing functional prediction tools were applied to evaluate their impact. We further analyzed structural flexibility of ACE2 and the protein-protein interface with the S-protein of SARS-CoV-2 using our developed Legion Interfaces Analysis (LiAn) program.

Results: Here, we characterized a total of 12 ACE2 putative deleterious missense variants. Of those 12 variants, we further showed that p.His378Arg could directly weaken the binding of catalytic metal atom to decrease ACE2 activity and p.Ser19Pro could distort the most important helix to the S-protein. Another seven missense variants may affect secondary structures (i.e. p.Gly211Arg; p.Asp206Gly; p.Arg219Cys; p.Arg219His, p.Lys341Arg, p.Ile468Val, and p.Ser547Cys), whereas p.Ile468Val with AF = 0.01 is only present in Asian.

Conclusions: We provide strong evidence of putative deleterious missense variants in ACE2 that are present in specific populations, which could disrupt the function and structure of ACE2. These findings provide novel insight into the genetic variation in ACE2 which may affect the SARS-CoV-2 recognition and infection, and COVID-19 susceptibility and treatment.

Keywords: COVID-19, ACE2, SARS-CoV-2, S-protein, missense

Background

The outbreak of the coronavirus disease 2019 (COVID-19), caused by a novel (new) coronavirus (SARS-CoV-2), has been characterized as a global pandemic[1-5]. COVID-19 is rapidly spreading across the world and affecting all populations. It has been documented that the S-protein of SARS-CoV-2 plays a key role in the recognition to the peptidase domain (PD) of the Angiotensin converting enzyme **(**ACE2) in humans [6, 7]. The three-dimensional protein structures of SARS-CoV-2 have recently been determined, which provide important insight into the treatment of the disease, such as vaccine development, antibody design and drug discovery[7]. The first X-ray crystallization structure of 3CLpro was resolved by Liu *et al* at 2.16Å resolution (Protein data bank (PDB) id 6lu7). The virus S-protein structure was first observed by Wrapp*et al* at 3.46 Å resolution by electron microscopy (PDB id 6vsb) [8]. The first full-length S-protein in complex with human ACE2 Cryo-EM structure was observed by Xihu University [7], and at almost the same time, the X-ray structure of the S-protein RBD domain in complex with ACE2 was solved by Tsinghua University at 2.45 Å resolution [9]. Two X-ray structures of S2 subunit have been determined by Zhu *et al* (PDB id 6lxt and 6lvn). In addition to SARS-CoV-2, several three-dimensional structures of ACE2, especially in complex with SARS S-protein, have been solved. It shows that ACE2 structure is flexible to toggle between open and close states when it binds an inhibitor or virus S-protein. Genetic variation, especially deleterious missense variants in these flexible regions, may affect its function and structure, and consequently alter the recognition by SARS-CoV-2. Thus, it's important to systematically characterize and evaluate potentially deleterious variants in ACE2, which may affect SARS-CoV-2 recognition and infection, and COVID-19 susceptibility and treatment.

Methods:

Characterization of genetic variants in ACE2 from the genome Aggregation Database

The genome Aggregation Database (gnomAD v2.1.1) has provided summary data (i.e. allele counts) for germline variants from 125,748 WES and 15,708 whole-genome sequences from unrelated individuals, sequenced as part of various disease-specific and population genetic studies, through the website browser http://gnomad.broadinstitute.org/. We characterized all genetic variants located in ACE2

genes after removing those with low quality control or had a MAF = 0. We next focused on the top 20 ACE2 missense variants with a MAF > 8×10^{-5}.

Variant annotation, bioinformatics and statistical analyses

The ANNOVAR tool was applied to annotate missense and disruptive variants. Disruptive variants were defined by nonsense, splice-site and frameshift. To further evaluate the functional impact of missense variants, we annotated each variant with the possible impact of an amino acid substitution on the structure/function from five protein prediction algorithms, including Polyphen-2 HumDiv, Poplyphen HumVar, Sorting Intolerant From Tolerant (SIFT), logistic regression test scores and MutationTaster. All of these analyses were implemented using the WGS Annotator (WGSA) via Amazon Web Service (AWS).

Protein structure analysis for ACE2 and the interaction between COVID-19 spike glycoprotein and ACE2

The protein structures are downloaded from the RCSB Protein Data Bank or from the authors' website. The protein structure figures and animations are produced by PyMol and its Morph function (The PyMoL Molecular Graphics System, Version 2.0 Schrödinger, LLC). The 2-dimensional interaction diagrams are produced by Schrödinger Maestro software. The 3-dimensional interaction plots are generated by our in-house developed LiAn (Legion Interfaces Analysis) program, which can calculate and display protein-ligand or protein-protein interactions (such as hydrogen bond, salt-bridge, water-bridge, π-interactions, hydrophobic interactions, halogen bond, *etc.*) for single protein structure or massive structures from molecular dynamics simulations. The LiAn program also integrates protein-protein interface analysis, protein structural clustering, protein interaction energy calculations, and fixed water predictions to analyze large amount of protein structures automatically.

Results

To illustrate how genetic variation may affect the structure, we analyzed the structural interactions between ACE2 and SARS-Cov-2. As displayed in Figure 2A,B and Video S1, we demonstrated that ACE2 has two states, *i.e.,* open and closed, for its native and ligand-binding states through a large hinge-bending motion [10]. In open state, ACE2 opens wide from its active site to wait for a ligand to enter. When the ligand enters ACE2

active site, it triggers ACE2 to close the active slot. Most SARS binding structures (e.g. PDB ids 2ajf, 3d0g, 3kbh, 3scl) show that S-protein binds the open/native state of ACE2. However, as depicted in Figure 2C,D and Video S2, the two monomers in PDB 3scl display that SARS spike proteins can bind either in an open state or in a closed state of ACE2, which implies that the conformational change of ACE2 can be triggered either by an inhibitor from an inner active site or S-protein from an outer PPI (protein-protein interface) site. The huge conformational change of the two states can be up to 14Å distance shift between K341 and T129. The two N-terminal helices (S19-N53, I54-M82) that contact SARS-Cov-2 S-proteins are among the most flexible regions. The hinge movement of the helices pivots on the loop region of Y83-N90. We also observed that the protein-protein interface of the SARS-CoV-2 spike glycoprotein to ACE2 has more hydrophilic residues that hydrophobic ones. The residues in ACE2 within distance of 3Å to S-protein are Q24, H34, D38, Y41, Q42, Y83 and K353 to G446, Y449, Y453, N487, T500 and G502. The PPI interface binds with six hydrogen bonds (Q24-N487, Q42-G446, Q42-Q498, K353-G502), a network of π-cation interactions (Y41-Q498, Y41-N501, Q42-Y449, Y83-N487, Q493-H34), one π-stacking interaction (Y83-F486), and only one hydrophobic interaction pair (M82-F486). In summary, the structural flexibility of ACE2 implies that its structure could be distorted by potentially deleterious missense variants with the altered amino acids in ACE2, which may consequently affect its binding efficiency to the S-protein in the virus.

We next analyzed germline coding variants in ACE2 from the gnomAD and performed functional predications using six existing bioinformatics tools (see Methods). We characterized a total of 12 ACE2 putative deleterious missense variants, whereas the top variants with functional disruptions predicted by all tools included p.Leu731Phe (rs147311723, allele frequency [AF] = 0.01 in African), p.Arg219Cys (rs372272603, AF = 7×10^{-4} in European), p.Ser547Cys (rs373025684, AF = 4×10^{-4} in European) and p.His378Arg (rs142984500, AF = 2×10^{-4} in European) (Table 1). Of those, we observed that these variants showed low frequency or rare in all populations (Figure 3A; Table 1). AF of the characterized variants varied in populations, whereas a majority of them showed population specificities (Figure 3A). In particular, we observed that two variants with low frequency were present in African (rs147311723) and in East Asian (s191860450), respectively (Figure 3B). The top AF of the other missense variants were present including, in African (rs73635825, rs138390800, and rs149039346), East Asian

(rs191860450), South Asian (rs148771870 and rs751603885) and Europeans (rs148771870) (Figure 3B).

We further analyzed the structural flexibility of ACE2 and its interaction with the RBD of S-protein of SARS-CoV-2 using the two or three-dimensional interaction diagrams (see Methods) for nine missense variants on eight residues, as displayed in Figure 1A. We showed that p.His378Arg could directly weaken the binding of catalytic metal atoms to decrease ACE2 activity, and p.Ser19Pro (rs73635825, AF = 3×10^{-3} in African) could distort the most important helix to interact with the S-protein.

H378R: As shown in Figure 1B, H378 is a key residue to fix the catalytic metal atom together with H378, E375 and E402. Its mutation to longer arginine will break the chelation network to Zn atoms, which could result in weakening its peptidase activity. Meanwhile, H378 also stabilizes the structure of the catalytic center via hydrogen bond and π-interaction with E402 and H401. Thus, the H378R mutant could reduce ACE2 peptidase function and destabilize the ACE2 structure.

S19P: S19 is the first N-terminal residue that can be shown in an X-ray structure, as displayed in Figure 1C. It locates at the beginning of helix S19-I54, which is one of the most important regions to contact virus S-protein. Its backbone forms hydrogen bonds with E23 and Q24 to stabilize the helical structure. It may also interact with the S477 in SARS-CoV-2 S-protein through weak hydrophilic interaction. Proline has poor helix-forming propensities, as it either breaks or kinks the helix [11]. Therefore, S19P mutation could destabilize the helix structure and weaken its binding to S-protein.

We also showed another seven missense variants that may affect secondary structures (i.e. p.Gly211Arg/rs148771870; p.Asp206Gly/rs142443432; p.Arg219Cys/rs759590772; p.Arg219His/rs759590772, p.Lys341Arg/rs138390800, p.Ile468Val/rs191860450, and p.Ser547Cys/rs373025684), whereas p.Ile468Val/rs191860450 with AF = 0.01 is only present in Asian.

G211R: G211 is at the turn point of a loop, as depicted in Figure 4A. Its neighboring V212 has strong hydrophobic interaction with L91 to stabilize the ACE2 structures across secondary structures. Its mutation to long and positive arginine is not favorable for the loop turning. Moreover, its arginine mutation also introduces hydrophilic group to this region, which may weaken the important hydrophobic interaction pair of V212-L91. Therefore, G211R mutation may destabilize the ACE2 structure.

D206G: D206 is on a helix of Y199-Y207 to stabilize multiple secondary structures via a hydrogen bond to N397/E398, as depicted in Figure 4B. Its mutation to glycine may affect the ACE2 inhibitor binding site allosterically, as it may disturb the location of the helix of E398-A413, which is essential for the binding of the catalytic zinc atom.

R219C/H: R219 is a key center residue to stabilize three helices, i.e., D157-Y196, D198-E208, and R219-Y252. As displayed in Figure 4C, R219 has a strong salt-bridge, hydrogen bond and charge interaction with D201 and E208, together with cation-π interaction, with Y196 across secondary structures to stabilize protein. Its mutation to cysteine or histidine will interrupt the strong interactions and destabilize the protein structure.

K341R: K341 is on a loop to stabilize another loop from the backbone hydrogen bond to I54, as shown in Figure 4D. It stabilizes the loop structure via a strong hydrogen bond to N338. Its mutation to longer arginine may weaken this hydrogen bond and slightly destabilize the loop structure.

I468V: As demonstrated in Figure 4E, I468 locates at the turn point of a loop to stabilize two helical structures (helices D431-K465 and W473-V485) via hydrophobic interactions to W459, M462, P469 and W473, together with a hydrogen bond of its backbone to W473. The π-stacking interaction from the pair of W459 and W473 is an important interaction to bundle the two helices. I468 chaperones the pair interaction by fixing the positions of two tryptophans. Consequently, its mutation to valine, which is shorter in side-chain and weaker in hydrophobic interaction, may slightly weaken the contact of two helices and destabilize the protein structure.

S547C: As displayed in Figure 4F, S547 stabilizes local helix S547-G561 through hydrogen bonds to A550 and G551. Its mutation to cysteine may weaken the hydrogen bond to A550 from hydroxyl side-chain to the thiol group, which in turn destabilizes the helical structure slightly.

Discussion

As shown in Figure 1, ACE2 has flexibility in its structure when it binds an inhibitor or virus S-protein. Therefore, the conformational change could be triggered by altered amino acids as well. Although some of missense variants we analyzed are not directly located on the PPI surface, the altered amino acids could affect, and most probably weaken, the binding of virus S-protein allosterically. Since the binding of the inhibitor

inside the active site triggers ACE2 to enter a closed state from an open state through a huge conformational change, the altered amino acids of active site residue could cause a structural change of ACE2 more easily. Thus, the H378R amino acid change may not only reduce ACE2 peptidase activity, but also change the structure of a PPI area to affect S-protein binding. When S19 mutates to the helix “killer” proline, it may destabilize the most important helix to contact with S-protein, which may result in decreasing the binding affinity to SARS-CoV-2. For SARS, the 24QAK to 24KAE mutant of ACE2 slightly inhibits interaction with SARS-CoV-2 spike glycoprotein [12]. G211R, D206G, R219C/H, K341R, and I468V may affect the interactions across secondary structures. Therefore, their mutations may destabilize the local structure significantly. S547C may only affect the stability of one secondary structure, which may have a minor effect on the S-protein binding. It should also be noted that our predicted deletious variants in protein structure lack of experimental validation. Further exploration would be required to further confirm their potential effects on ACE2 function.

Conclusions

In this study, we characterized a total of 12 putative deleterious missense variants in the gene ACE2. Of those, we further provided strong evidence of nine missense variants that may disrupt the flexible regions of ACE2 protein structure or its protein-protein interaction with the RBD of S-protein of SARS-CoV-2. Results from this study highlight an important role of deleterious missense variants in the gene ACE2 that are present in the specific populations, which may affect SARS-CoV-2 recognition and infection. These variants could be important for the development of appropriate strategies of COVID-19 prevention, control and treatment to distinguish individuals between carrying and non-carrying those deleterious variants. Our findings may also provide a clue to partially explain why there were substantial discrepancies about the morbidity and mortality in regional disparity and distinct populations[1-5].

Abbreviations

COVID-19: Coronavirus disease

SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2

S-protein: Spike glycoprotein

ACE2: Angiotensin-converting enzyme 2

gnomAD: Genome Aggregation Database

LiAn: Legion Interfaces Analysis
PD: Peptidase domain
PDB: Protein data bank
PPI: Protein-protein interface
AF: Allele frequency

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and materials
Characterized genetic data and source codes that are used in this work are available from Github (https://github.com/XingyiGuo/ACE2).

Declaration of interests
We declare no competing interests.

Funding
Not applicable.

Author Contributions
XG conceived and designed the study. XG, WL and HL performed the data, bioinformatic and protein structure analyses. XG, WL and HL wrote the manuscript with contributions from ZC and YX. All authors have reviewed and approved the content of the article.

Acknowledgements
We thank the gnomAD and RCSB Protein Data Bank for providing valuable data resources for this study. We also thank Marshal Younger for assistance with editing and manuscript preparation.

References

1. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; **395**(10223): 497-506.
2. Liang W, Guan W, Chen R, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. *The Lancet Oncology* 2020; **21**(3): 335-7.
3. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *Jama* 2020.
4. Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *The Lancet Respiratory medicine* 2020.
5. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England journal of medicine* 2020; **382**(8): 727-33.
6. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 2005; **309**(5742): 1864-8.
7. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science* 2020.
8. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020.
9. Lan J, Ge J, Yu J, et al. Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the ACE2 receptor. *bioRxiv: https://doiorg/101101/20200219956235* 2020.
10. Towler P, Staker B, Prasad SG, et al. ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *The Journal of biological chemistry* 2004; **279**(17): 17996-8007.
11. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical journal* 1998; **75**(1): 422-7.
12. Li W, Zhang C, Sui J, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *The EMBO journal* 2005; **24**(8): 1634-43.

Table 1. Characterization of putative deleterious missense variants in ACE2.

rsID	Position (hg19)	Ref	Alt	AA change[1]	AF							D[6]	S[7]
					Combined	African	Latino	Asian[2]	European[3]	European[4]	Asian[5]		
rs147311723	15582265	G	A	**Leu731Phe**	1.4×10^{-3}	0.014	4.8×10^{-4}	0	0	3.3×10^{-5}	0	**6**	N
rs148771870	15607532	C	T	Gly211Arg	1.3×10^{-3}	2.7×10^{-4}	2.1×10^{-4}	0	1.6×10^{-3}	1.9×10^{-3}	1.9×10^{-3}	2	Y
rs191860450	15593829	T	C	Ile468Val	8.4×10^{-4}	0	0	0.011	0	2.2×10^{-5}	5.6×10^{-5}	4	Y
rs149039346	15584416	A	G	Ser692Pro	5.6×10^{-4}	5.8×10^{-3}	3.6×10^{-5}	0	0	2.2×10^{-5}	5.3×10^{-5}	4	N
rs138390800	15599392	T	C	Lys341Arg	4.0×10^{-4}	3.9×10^{-3}	1.8×10^{-4}	0	0	0	0	2	Y
rs372272603	15607508	G	A	**Arg219Cys**	3.5×10^{-4}	1.6×10^{-4}	0	0	0	7.0×10^{-4}	1.6×10^{-4}	**6**	Y
rs73635825	15618980	A	G	Ser19Pro	3.1×10^{-4}	3.3×10^{-3}	0	0	0	0	0	3	Y
rs142443432	15607546	T	C	Asp206Gly	3.0×10^{-4}	1.1×10^{-4}	3.6×10^{-5}	0	0	6.3×10^{-4}	0	2	Y
rs751603885	15584401	T	C	Arg697Gly	2.5×10^{-4}	0	0	0	0	0	2.4×10^{-3}	4	N
rs373025684	15590348	G	C	**Ser547Cys**	2.1×10^{-4}	0	1.4×10^{-4}	0	5.4×10^{-5}	3.9×10^{-4}	0	**6**	Y
rs142984500	15596376	T	C	**His378Arg**	8.8×10^{-5}	0	0	0	0	1.9×10^{-4}	0	**6**	Y
rs759590772	15607507	C	T	Arg219His	9.8×10^{-5}	0	0	0	0	1.2×10^{-5}	8.9×10^{-4}	5	Y

"1": AA refers to amino acid
"2": East Asian
"3": Finnish
"4": European not including Finnish
"5": South Asian
"6": Number of tools predicted the variant to be deleterious; the top four variants with predicted functional disruptions highlighted in bold
"7": "Y" refers to The variant likely affecting the protein structure of ACE2, otherwise "N"

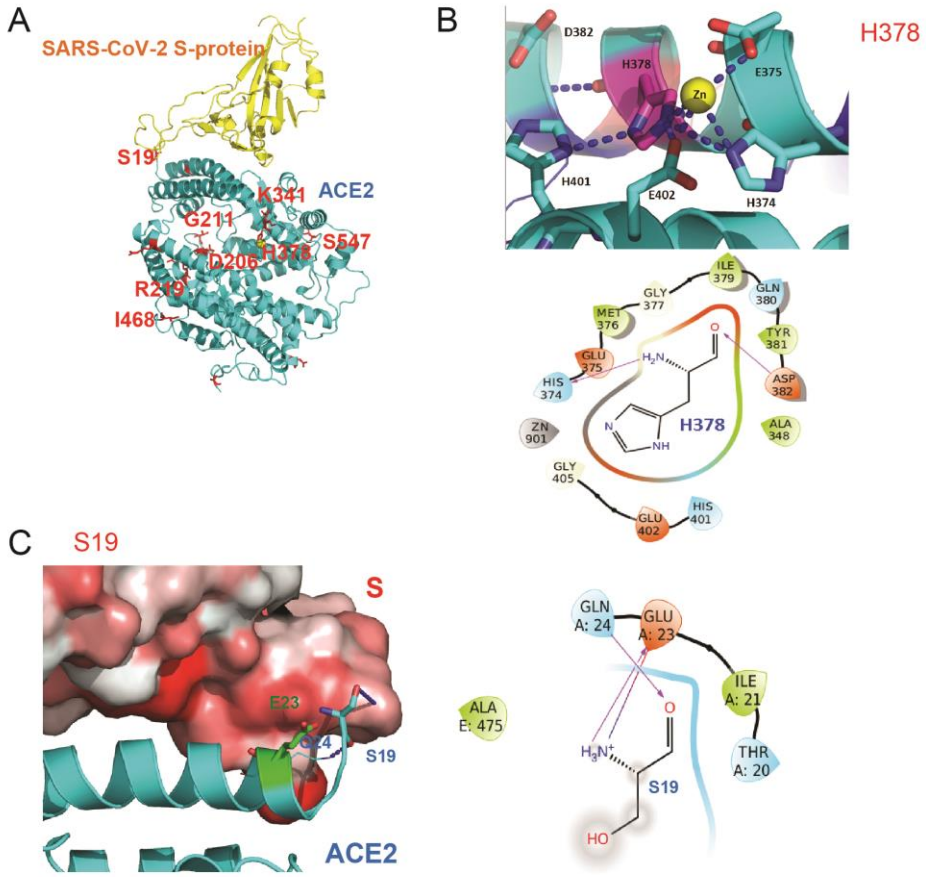

Figure 1. **Interaction diagrams for H378R (rs142984500) and S19P (rs73635825)**. (A) Structural positions of the ACE2 altered amino acids from the nine ACE2 missense variants (colored red). The SARS-Cov-2 S-protein is colored yellow. (B) Interactions of

H378R and (C) Interactions of S19P. Hydrogen bonds are depicted as blue dots. Π-interactions are depicted as yellow arrows. Hydrophobic interactions are shown as grey arrows. Water molecules from water-bridge are displayed as red spheres.

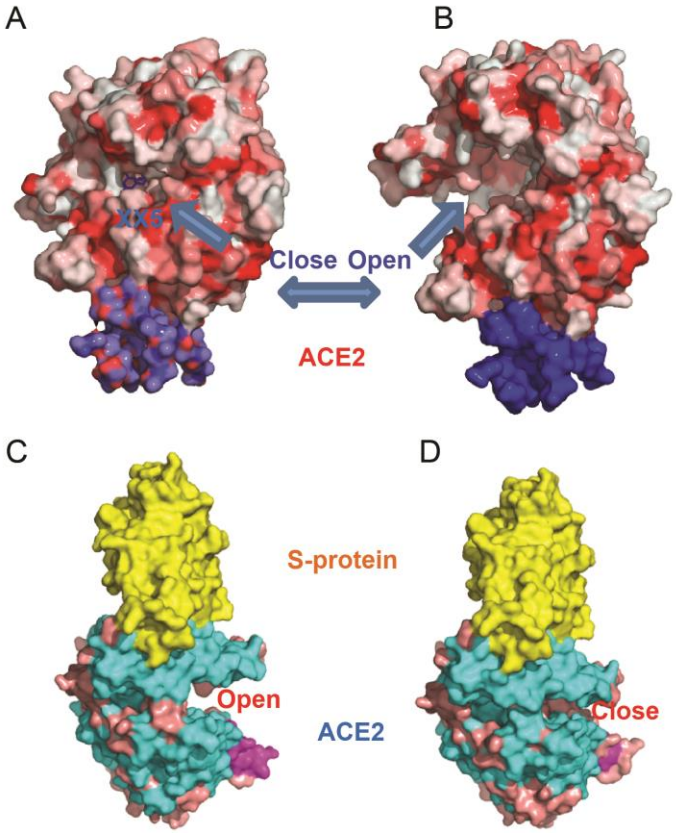

Figure 2. **Open/Closed state of ACE2 and S-protein binding**. (A) Closed state when ACE2 binds MLN-4760 (XX5) inhibitor (PDB1r4l). (B) Open/native state of ACE2 (PDB 1r42). (C) Most PDB structures show that SARS S-protein (colored yellow) binds the

open state of ACE2, as from one monomer of PDB 3scl. (D) S-protein shows binding to closed state of ACE2 from another monomer of PDB 3scl.

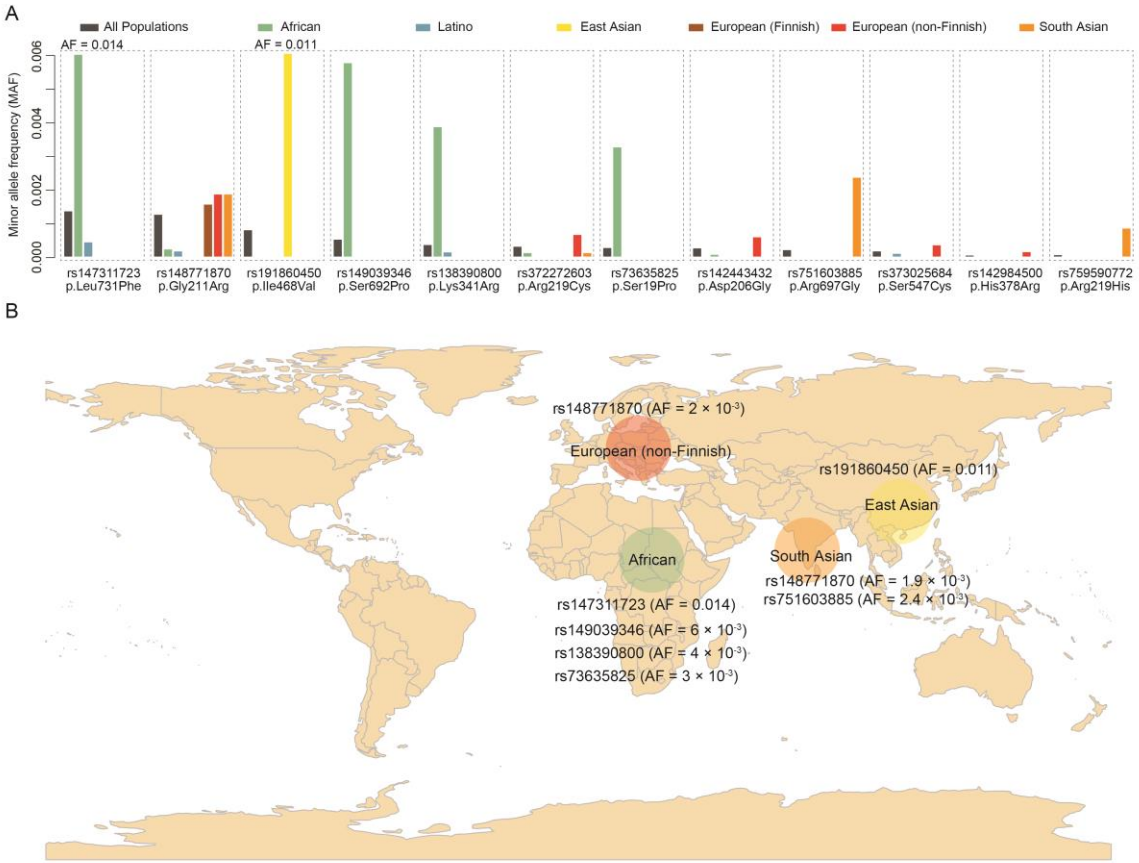

Figure 3. **Distribution of the characterized 12 ACE2 missense variants in different populations**. (A) AF for each missense variant in different populations. (B) The top AF of the missense variants present in African (rs73635825, rs138390800, rs149039346

and rs147311723), East Asian (rs191860450), South Asian (rs148771870 and rs751603885) and European (rs148771870) populations.

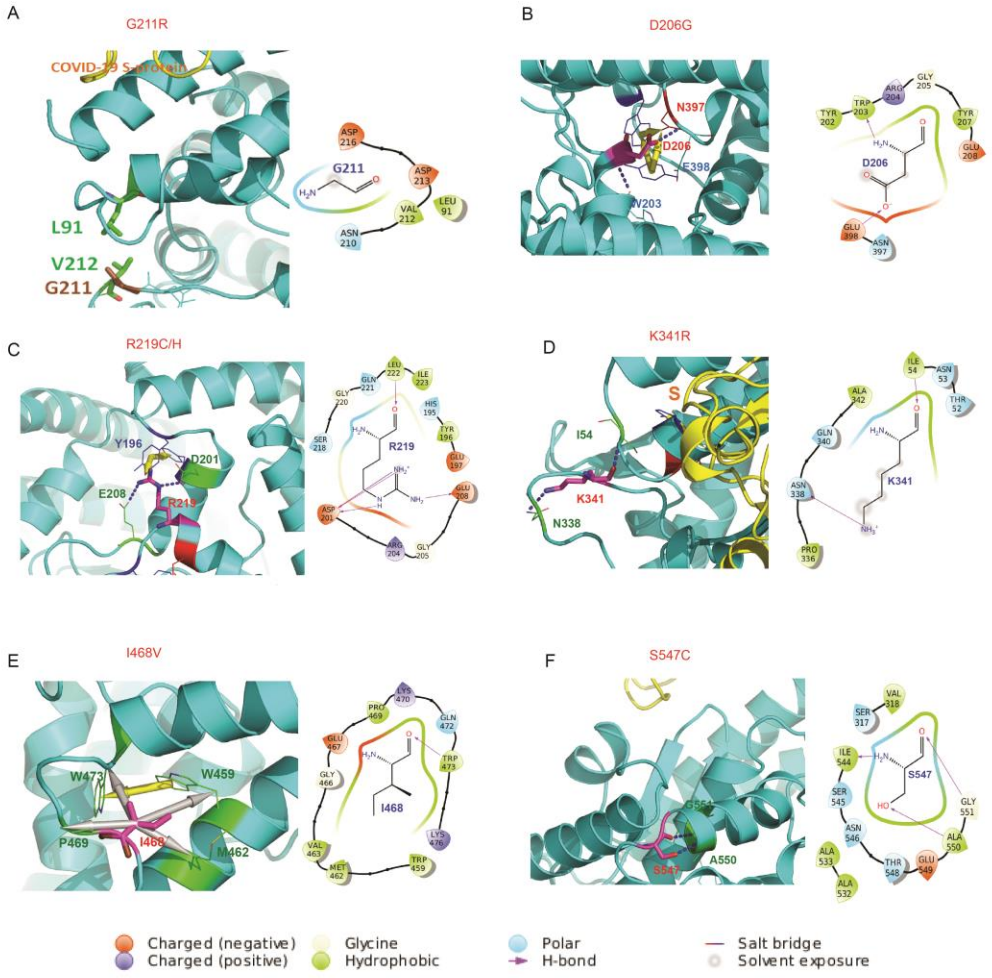

Figure 4. **Interaction diagrams for the six residuals from seven *ACE2* missense variants.** (A) Interactions of G211R, (B)Interactions of D206G, (C) Interactions of R219 C/H (D) Interactions of K341R, (E) Interactions of I468V and (F) Interactions of S547C. Hydrogen bonds are depicted as blue dots. Π-interactions are depicted as yellow arrows. Hydrophobic interactions are shown as grey arrows. Water molecules from water-bridge are displayed as red spheres.