No evidence for allelic association between Covid-19 and ACE2 genetic variants by direct exome sequencing in 99 SARS-CoV-2 positive patients

Authors: Antonio Novelli,[1] Michela Biancolella,[2,3] Paola Borgiani,[4] Dario Cocciadiferro,[1] Vito Luigi Colona,[3] Maria Rosaria D'Apice[3], Paola Rogliani[5], Salvatore Zaffina[6], Francesca Leonardis[7], Andrea Campana[8], Massimiliano Raponi[9], Massimo Andreoni[10,11], Sandro Grelli[12], and Giuseppe Novelli,[3,4,12,14]*

[1]Laboratory of Medical Genetics, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy

[2]Department of Biology, Tor Vergata University of Rome, 00133 Rome, Italy

[3]Medical Genetics Laboratory, Tor Vergata Hospital, Rome, Italy

[4]Department of Biomedicine and Prevention, Tor Vergata University of Rome, 00133 Rome, Italy

[5]Unit of Respiratory Medicine, Department of Experimental Medicine, University of Rome "Tor Vergata", Rome, Italy

[6]Occupational Medicine, Bambino Gesù Children's Hospital , IRCCS , Rome , Italy

[7]Intensive Care Unit, Tor Vergata University Hospital, Rome, Italy

[8]Department of Pediatrics, IRCCS "Bambino Gesù" Children's Hospital, Rome, Italy

[10]Department of Systems Medicine, University of Rome Tor Vergata, Rome, Italy

[11]Infectious Diseases Clinic, Policlinico Tor Vergata, Rome, Italy

[12]Dept. of Experimental Medicine and Biochemical Sciences, University of Rome Tor Vergata, Italy

[13]IRCCS Neuromed, Pozzilli (IS), Italy.

[14]Department of Pharmacology, School of Medicine, University of Nevada, Reno, NV 89557, USA

*Correspondence: novelli@med.uniroma2.it (G.N.)

Abstract

Background

Coronaviruses (CoV) are a large family of viruses that are common in people and many animal species. Animal coronaviruses rarely infect humans with the exceptions of the Middle East Respiratory Syndrome (MERS-CoV), the Severe acute respiratory syndrome coronavirus (SARS-CoV), and now SARS-CoV-2, which is the cause of the ongoing pandemic of coronavirus disease 2019 (COVID-19). Many studies suggested that genetic variants in *ACE2* gene may influence the host susceptibility/resistance to SARS-CoV-2 virus according to the functional role of ACE2 in human pathophysiology. However, all these studies have been conducted *in silico* based on epidemiological and population data. We therefore investigated the occurrence of *ACE2* variants in a cohort of 99 Italian unrelated individuals clinically diagnosed with coronavirus disease 19 (COVID-19) to experimental demonstrate allelic association with disease severity.

Methods

By whole-exome sequencing we analysed 99 DNA samples of severely and extremely severely COVID-19 patients hospitalized at the University Hospital of Rome "Tor Vergata" and Bambino Gesù Hospital in Rome.

Results

We identified three different germline variants, one intronic (c.439+4G>A) and two missense (c.2158A>G, p.Asn720Asp; c.1888G>C, p.Asp630His), in 26 patients with a similar frequency between male and female and a not statistically different frequency, except for c.1888G>C, (p.Asp630His) with the ethnically matched populations (EUR).

Conclusions

Our results suggest that there is not any *ACE2* exonic allelic association with disease severity. It is possible that rare susceptibility alleles are located in the non-coding region of the gene able to control *ACE2* gene activity. It is therefore of interest, to explore the existence of *ACE2* susceptibility alleles to SARS-Co-V2 in these regulatory regions. In addition, we found no significant evidence that *ACE2* alleles is associated with disease severity/sex bias in the Italian population.

Introduction

Coronaviruses (CoV) are a large family of viruses that are common in people and many animal species, including camels, cattle, cats, and bats. Animal coronaviruses rarely infect humans and then spread between people with the exceptions of the Middle East Respiratory Syndrome (MERS-CoV), the Severe acute respiratory syndrome coronavirus (SARS-CoV), and now SARS-CoV-2, which is the cause of the ongoing pandemic of coronavirus disease 2019 (COVID-19) (1; 2). SARS-CoV-2 utilizes an extensively glycosylated spike (S) protein that protrudes from the viral surface to bind to angiotensin-converting enzyme 2 (ACE2) to mediate host-cell entry (3). ACE2-binding affinity of the receptor-binding domain (RBD) in S1 subunit of S protein of SARS-CoV-2 is 10- to 20-fold higher than that of SARS-CoV, which may contribute to the higher infectivity and transmissibility of SARS-CoV-2 (4). Many studies suggested that genetic variants in *ACE2* gene may influence the host susceptibility/resistance to SARS-CoV-2 virus according to the functional role of ACE2 in human pathophysiology (2). However, all these studies have been conducted *in silico* based on epidemiological and population data. We therefore investigated the occurrence of *ACE2* variants in a cohort of 99 Italian SARS-CoV-2 -positive patients by direct exome sequencing.

Results and Discussion

We analysed by exome sequencing 99 DNA samples of COVID-19 patients hospitalized at the University Hospital of Rome "Tor Vergata" in Rome and Bambino Gesù Hospital, Rome. All patients were clinically diagnosed as COVID-19 based on clinical suspicion and confirmed by viral RNA detection at oropharyngeal and nasopharyngeal swabs. All the selected patients were severely or extremely severely affected. Nine patients in the extremely severe group are passed away. The majority of the enrolled patients were males (61males, 38 females). Median age was 63 years (range: 2–92 years), 64 patients were under 65 years old. Ten patients were children (median age was 11.5 years) showing a severe form of the disease but none of them have Kawasaki-like syndrome (6). The analytical procedure received approval by the local ethics committee at University Hospital of Rome Tor Vergata (protocol no. 50/20).

We identified in *ACE2* gene, three different germline variants, one intronic (c.439+4G>A) and two missense (c.2158A>G, p.Asn720Asp; c.1888G>C, p.Asp630His), in 26 patients (14 females and 12 males). The frequency of the three variants identified are similar between male and female patients suggesting that there is not gender effects underlying the frequency distribution of *ACE2* variants. **(Table 1)**.

GnomAD database analysis revealed that the identified *ACE2* variants existed with a cumulative frequency of 0.2289 in ethnically matched populations (EUR). The frequency of the variants

3

detected in our examined cohort (0.2119) was not statistically different (Table 1). A significant difference was detected only for the c.1888G>C (p.Asp630His) observed in a heterozygous female (p=0.0067) (Table 1). The allelic frequency of this variant in GnomAD for the EUR reference population is 0.000036 confirming that this is a very rare allele. To predict the potential impact of this variant on the protein we used different tools (PolyPhen2, Mutation Taster, SIFT, MetaLR_pred, and MetaSVM_pred.). The in-silico analyses gave conflicting computational verdict because 3 benign predictions vs. 2 pathogenic predictions. The sequence alignment of the ACE2 protein with its orthologous proteins shows that the wild type residue is not highly conserved in species implying an irrelevant functional or structural role of this residue in the ACE protein. Concerning the others two variants, the recurrent c.439+4A>G (rs2285666) intronic variant has been previously reported by Strafella et al. (5) and by Asselta et al.(4) in two different Italian cohort representative of the country's population. The variant is located in the intron 3 in a splice site region of the gene. However, using Human Splicing Finder (HSF) the analysis showed no significant splicing alterations.

The missense variant c.2158A>G (p.Asn720Asp) was found in two patients, 1 female in heterozygous state and 1 male, with a frequency in line with the frequency reported for the European non-Finnish population in the GnomAD database. It is in the C-terminal domain that is not involved in the SARS-CoV-2 S protein interaction. The *in-silico* analyses to predict the potential impact of this variant on the protein gave benign computational verdict because 4 benign predictions vs 1 pathogenic prediction.

Numerous *in silico* data suggested that the ACE2 variants in structural part of the protein could have an impact on the pathogen binding dynamics or increase the quantitative expression of ACE2. All these studies were carried out on an epidemiological basis of population allele frequencies deposited in the various databases available. We systematically analyzed the *ACE2* coding-region variants in a representative cohort of Italian patients severely affected by COVID-19 in order to identify rare and causative predisposing alleles. Although we have identified a variant (p.Asp630His), very rare in European population, in a single patient affected by COVID-19, we do not believe that there is an enrichment of *ACE2* coding mutant alleles in the population of Italian patients affected by COVID-19. Our results confirm and extend that *ACE2* is a gene with low allelic frequencies of missense variants as expected on the basis of GnomAD population data. In fact, we provide evidence that the rate of amino acid changes in the binding region with SARS-Co-V2 and at the protein cleavage sites is very low. This suggests that these regions have been under evolutionary pressure, probably for the essential catalytic role of ACE2 as transmembrane carboxypeptidase. It is possible that rare susceptibility alleles are located in the non-coding region

of the gene able to control *ACE2* gene activity. Mutant alleles in noncoding DNA can cause alteration in expression levels or disturbing the timing of the expression (1). These variations concern enhancers, promoters, insulators and silencers or regions that provide instructions for producing functional RNA molecules, such as transfer RNA, miRNAs or long non-coding RNA. It is therefore of interest, to explore the existence of *ACE2* susceptibility alleles to SARS-Co-V2 in these regulatory regions. Interestingly very recently Bunyavanich et al. (10) have showed age-dependent expression of ACE2 gene in nasal epithelium, highlighting how the different level of expression of ACE2 may be the reason for explaining why COVID-19 is less frequent in children (10). Several studies have shown that *ACE2* gene undergoes the action of at least four miRNAs: miR-200c, let-7b, miR-1246, and miR-125b (11-14). Genes coding for these miRNAs could reveal variations able to modulate the expression and therefore produce significant quantitative differences in the ACE2 protein.

References

1. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg. 2020 Apr;76:71-6.

2. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol. 2016 Aug;14(8):523-34.

3. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol. 2020 Apr;5(4):562-9.

4. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020 Mar 13;367(6483):1260-3.

5. Hussain M, Jabeen N, Raza F, Shabbir S, Baig AA, Amanullah A, et al. Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. J Med Virol. 2020 Apr 6.

6. Verdoni L, Mazza A, Gervasoni A, Martelli L, Ruggeri M, Ciuffreda M, et al. An outbreak of severe Kawasaki-like disease at the Italian epicentre of the SARS-CoV-2 epidemic: an observational cohort study. Lancet. 2020 May 13.

7. Strafella C. CV, Termine A.,Barati S.,Gambardella S.,Borgiani P. ,Caltagirone C.,Novelli G., Emiliano Giardina E.,Cascella R. Analisys of ace2 genetic variability among populations highlights a possible link with covid19-related neurological complications. 15 May 2020, PREPRINT (Version 1) available at Research Square [+https://doiorg/1021203/rs3rs-28871/v1+]. 2020.

8. Asselta R., Paraboschi EM., Mantovani A., Duga S. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. medRxiv 2020033020047878; doi: https://doiorg/101101/2020033020047878.

9. Gloss BS, Dinger ME. Realizing the significance of noncoding functionality in clinical genomics. Exp Mol Med. 2018 Aug 7;50(8):97.

10. Bunyavanich S, Do A, Vicencio A. Nasal Gene Expression of Angiotensin-Converting Enzyme 2 in Children and Adults. JAMA. 2020 May 20.

11. Liu Q, Du J, Yu X, Xu J, Huang F, Li X, et al. miRNA-200c-3p is crucial in acute respiratory distress syndrome. Cell Discov. 2017;3:17021.

12. Zhang R, Su H, Ma X, Xu X, Liang L, Ma G, et al. MiRNA let-7b promotes the development of hypoxic pulmonary hypertension by targeting ACE2. Am J Physiol Lung Cell Mol Physiol. 2019 Mar 1;316(3):L547-L57.

13. Fang Y, Gao F, Hao J, Liu Z. microRNA-1246 mediates lipopolysaccharide-induced pulmonary endothelial cell apoptosis and acute lung injury by targeting angiotensin-converting enzyme 2. Am J Transl Res. 2017;9(3):1287-96.

14. Huang YF, Zhang Y, Liu CX, Huang J, Ding GH. microRNA-125b contributes to high glucose-induced reactive oxygen species generation and apoptosis in HK-2 renal tubular epithelial cells by targeting angiotensin-converting enzyme 2. Eur Rev Med Pharmacol Sci. 2016 Oct;20(19):4055-62.

Table 1 - Comparison of allelic counts (variant vs WT alleles) between our Italian population of SARS-CoV2 Positive Patients and Europeans (GnomAD database)

Nr.	dbSNP	Position (Hg19)	Coding	Protein	Genotype (N)	Variant type	Allelic count	Allelic count EUR (GnomAD)	Allelic Frequeny	Allelic Frequency EUR (GnomAD)	Fisher Test
1	**rs140312271**	chrX:15588426	c.1888G>C	p.Asp630His	Het (1)	Missense variant	1 vs 136	3 vs 81439	1/137 = **0,0073**	3/81442 = **3,68x10**$^{-5}$	P = 0,0067
2	**rs2285666**	chrX:15610348	c.439+4G>A		Hom (4) Het (7) Hem (12)	Splice region variant	27 vs 110	17240 vs 68924	27/137 = **0,197**	17240/86164 = **0,2001**	P = 0,92 NS
3	**rs41303171**	chrX:15582298	c.2158A>G	p.Asn720Asp	Het (1) Hem (1)	Missense variant	2 vs 135	2195 vs 84871	2/137 = **0,0146**	2195/87066 = **0,0251**	P = 0,59 NS
Cumulative frequencies (variants 1, 2, 3) The allelic count of Europeans wild-type alleles (GnomAD database) in cumulative frequencies was calculated as the average of the number of wild-type alleles of the three genetic variants, since the 3 variants has been genotyped in a number of individuals slightly different							30 vs 107	19438 vs 78411	30/137 = **0,2189**	19438/84890 = **0,2289**	P = 0,55

Acknowledgements

Declaration of Interests

The authors declare no competing interests.

Web Resources

GnomAD, https://gnomad.broadinstitute.org/
PolyPhen2, http://genetics.bwh.harvard.edu/pph2/
Mutation Taster, http://www.mutationtaster.org/
SIFT, https://sift.bii.a-star.edu.sg/