Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Medical Hypotheses 147 (2021) 110485

Contents lists available at ScienceDirect

Medical Hypotheses

journal homepage: www.elsevier.com/locate/mehy

Host genomics of COVID-19: Evidence point towards Alpha 1 antitrypsin deficiency as a putative risk factor for higher mortality rate

Atanu Kumar Dutta *, Kalyan Goswami

Department of Biochemistry, All India Institute of Medical Sciences, Kalyani, India

ARTICLE INFO

Keywords:
COVID-19
TMPRSS2
Alpha 1 antitrypsin

ABSTRACT

Corona Virus Disease 2019 (COVID-19) has emerged as a pandemic leading to unprecedented disruption of global health and economy. Transmembrane protease serine 2 (TMPRSS2) has been found to be critical in priming the viral spike protein and the host ACE2 receptor before the virus enters into the host cell. Recent studies have experimentally demonstrated that Alpha 1 antitrypsin (encoded by SERPINA1 gene) is an inhibitor of TMPRSS2 and provided support to the already approved therapy as a candidate for COVID-19. Interestingly Alpha 1 antitrypsin deficiency is common among Europeans. Here we have provided in silico evidence that Alpha 1 antitrypsin can interact with TMPRSS2 and both of them are co-expressed in the human liver and lung. We then analyzed the gnomAD dataset to show that Europeans and Latinos have a substantially higher carrier frequency of Alpha 1 Antitrypsin Deficiency (~12%) compared to other large ethnicities. Therefore, we hypothesize that Alpha 1 antitrypsin deficiency might be a risk factor for severe infection with SARS-CoV-2. We propose Alpha 1 antitrypsin status as a potential prognostic predictor of COVID-19 outcome.

Introduction

In December 2019 China reported a new Coronavirus emerging from Wuhan city. Since then the virus has traveled to all corners of the world with 29,593,883 confirmed infections and 935,446 deaths globally as of 16.09.2020 (https://coronavirus.jhu.edu/map.html). As per available data, five out of the top ten countries with the highest COVID-19 related fatality/million population are from Europe and rest of the five are from Latin America (https://www.worldometers.info/coronavirus/) (Supplementary Table 1).

Understandings the cause behind the difference would need an investigation of the trinity – agent, host, and environment. Concerning the virus, SARS-CoV-2 uses the ACE2 receptor for entry and the serine protease TMPRSS2 for S protein priming [1]. In addition, the host genetics is also likely to play a major part since Asians have fared much better in terms of morbidity and mortality of COVID-19 compared to other ethnicities like Europeans. Moreover, there is a significant difference in terms of morbidity and mortality of COVID-19 in different ethnicities which can hardly be explained by environmental factors alone. TMPRSS2 also being under the regulation of androgen [2] also might explain higher infection as well as the mortality rate of COVID-19 among the males across the ethnicities and geographic regions. Notably, recent large GWAS studies have revealed several emerging chemokine genes like *CXCR6*, *CCR1* and *SLC6A20* gene coding the amino acid transporter SIT1 which is known to interact with ACE2 receptor [3].

The hypothesis

Alpha 1 antitrypsin deficiency might be a risk factor for severe infection with COVID-19

Concerning other genetic predisposing factors, of particular interest is a study that experimentally demonstrated that Alpha 1 antitrypsin is an inhibitor of Transmembrane protease serine 2 (TMPRSS2) enzyme in a dose-dependent manner and this may explain the preventive effect of COVID-19 infection naïve plasma for viral entry in model systems [1,4]. Indeed, the Structure-based computational predictions of protein-protein interactions (Struct2Net – http://cb.csail.mit.edu/cb/struct2net/webserver) also showed the probability of interaction to be 0.386 using the 1k9o and 1ezx PDB crystal structures. Which corresponds to 95% specificity for this interaction which is very close to the suggested specificity of 96% (corresponding to suggested threshold of

* Corresponding author at: Department of Biochemistry, All India Institute of Medical Sciences, Kalyani, NH-34 Connector, Basantapur, Saguna, Kalyani, West Bengal 741245, India.
E-mail address: atanu.dutta05@gmail.com (A.K. Dutta).

https://doi.org/10.1016/j.mehy.2021.110485
Received 16 September 2020; Received in revised form 5 December 2020; Accepted 3 January 2021
Available online 9 January 2021

0.4 – http://cb.csail.mit.edu/cb/struct2net/webserver/sens_spec.csv). Given the Alpha 1 Antitrypsin undergoes significant modifications including glycosylation before it interacts with any of the serine proteases and TMPRSS2 is also glycosylated the actual specificity of their interaction may be higher than predicted as glycosylation in both the partners is known to increase the probability of interaction [5]. Also, the GTEx database showed a co-expression of *TMPRSS2* and *SERPINA1* (Alpha 1 antitrypsin) in the human liver and lung [6]. Further studies are required to investigate if *TMPRSS2* upregulates the expression of *SERPINA1*. Thus, these preliminary experimental and in Silico pieces of evidence opens up a new possibility of repurposing the existing therapy for Alpha 1 antitrypsin deficiency for COVID-19.

Evaluation of the hypothesis

Though it is well known that Alpha 1 antitrypsin deficiency is highly prevalent in Europeans to date most of the studies have focused their effort on the most common deficiency variant rs17580. Thus, we undertook an unbiased approach of a comprehensive evaluation of the gnomAD database which contains exome and genome data of 71,702 individuals as on 5th May 2020. Briefly, all variant data of the SERPINA1 gene were downloaded and annotated using the Varsome (https://varsome.com/) and Ensembl Variant Effect Predictor (https://asia.ensembl.org/info/docs/tools/vep/index.html) web interfaces. Of the total 466 variants in the SERPINA1 gene, 218 were either loss of function or non-synonymous variants. After we used variant filtering as per ACMG 2015 criteria [7] 32 variants remained which were either pathogenic or likely pathogenic (Table 1). Out of these 13 variants were observed only once highlighting the long tail in the variant frequency distribution of most monogenic diseases. We manually corroborated our findings with the Clinvar database (https://www.ncbi.nlm.nih.gov/clinvar/) to ensure there is enough clinical or functional evidence available for variant pathogenicity. We then added up the minor allele frequencies of these 32 variants across all ethnicities to obtain the q value. Using the Hardy Weinberg equilibrium for an autosomal recessive inheritance, we calculated the 2pq value to estimate the population carrier frequencies and percentages for all available ethnicities. Derivation of population carrier frequency data is listed in Table 2. We also identified 60 variants of unknown significance which are listed in Supplementary Table 2. More clinical reports or functional analysis would be required to resolve these variants. The correlation of geographical prevalence of Alpha 1 Antitrypsin deficiency with severe COVID-1 SARS-CoV-2 infection prevalence in Italy [8] serves to highlight the utility of our approach in supporting the hypothesis. Considering the possibility of genetic admixture in the gnomAD data we have also reviewed the literature for country specific prevalence estimates. One such study from Germany [9] using health insurance data showed a prevalence of Alpha 1 Antitrypsin Deficiency of 29.36 per 1,00,000 in those ≥30 years of age. Based on our estimates based on European data ($q = 0.066241$) the projected prevalence in Germany would be (q^2) 43.87 per 1,00,000. Considering Latin American nations previous studies suggested a population carrier frequency of the commonest SERPINA1 mutation rs17580 (PiS) of 33.5 per 1000 population in Peru [10] which is pretty close to the gnomAD frequency of 32.3 per 1000 Latino/Admixed American population (https://gnomad.broadinstitute.org/variant/rs17580?dataset=gnomad_r2_1). gnomAD database had 1042 exomes from Latino/Admixed American population which constituted 17.8% of exomes in gnomAD database which also contained data from the TOPMED project. Therefore, we can argue that estimates based on genome databases can be broadly good alternatives when population-based data is scarce.

Table 1
Pathogenic and likely pathogenic variants.

Sl. No.	rs ID	Protein Consequence	Transcript Consequence
1	rs17580	p.Glu288Val	c.863A>T
2		p.Glu366Lys	c.1096G>A
3	rs28931570	p.Arg63Cys	c.187C>T
4	rs1425620023	p.Thr346SerfsTer7	c.1036_1037insGT
5	rs1175196821	p.Glu347GlyfsTer12	c.1040_1041del
6	rs112661131		c.646+2T>C
7	rs775982338	p.Phe76del	c.227_229del
8	rs750779440		c.917+1G>A
9		p.Asn402IlefsTer12	c.1205del
10		p.Glu400AspfsTer44	c.1199_1200insC
11	rs746155701	p.Thr320Ala	c.958A>G
12	rs11558261	p.Gly139Ser	c.415G>A
13	rs1240316149	p.Ala379ProfsTer19	c.1135del
14		p.Asn289LysfsTer2	c.866dup
15	rs756773408	p.Gly216Cys	c.646G>T
16		p.Glu175LysfsTer35	c.522del
17			c.-23-1G>C
18			c.-23-2A>G
19	rs764325655	p.Glu387ArgfsTer14	c.1158dup
20	rs921982028	p.Thr204SerfsTer11	c.611_612del
21	rs751235320		c.646+1G>T
22	rs267606950	p.Tyr184Ter	c.552del
23	rs776846218	p.Glu387ArgfsTer11	c.1158del
24	rs751343534	p.Ala371TrpfsTer28	c.1111_1115del
25	rs754737768	p.Ala371AsnfsTer31	c.1110_1111insAACA
26			c.1065+2T>G
27	rs199422211	p.Lys241Ter	c.721A>T
28		p.Gly216AlafsTer22	c.647del
29		p.Tyr184Ter	c.552C>G
30		p.Pro393Leu	c.1178C>T
31	rs28931569	p.Leu65Pro	c.194T>C
32	rs1445192595	p.Trp218Ter	c.654G>A

Consequences of the hypothesis and discussion

We thereby estimated that other than the small Amish population who have the highest carrier rate for Alpha 1 antitrypsin deficiency (22%) both Europeans and Latinos have substantially high carrier rates (12%) (Table 2). Incidentally, the later ethnicities are the predominant population group of nine out of ten countries having the highest case fatality rate. Smoking habits may further aggravate the impact of Alpha 1 antitrypsin deficiency to induce further lung damage (3). Serine protease inhibitor family or SERPIN is an important class of proteins that plays a crucial role in the homeostasis of blood coagulation [9]; intriguingly, autopsy reports of the deceased showed that the DIC is an important pathogenic factor [10].

Thus, we believe that there is substantial evidence to warrant further studies involving measurement of serum SERPIN levels including alpha 1 antitrypsin status along with the smoking history in COVID-19 patients and also study of therapeutic efficacy of Alpha 1 antitrypsin in animal models of this infection. Indeed there is already a clinical trial already underway in Ireland investigating [11,12] the efficacy of Alpha 1 antitrypsin pharmacological treatment on inflammatory marker levels (https://www.clinicaltrialsarena.com/news/ireland-alpha-1-antitrypsin-trial/). As of now, based on the available evidence we advocate, the potential role of Alpha 1 antitrypsin and another SERPIN status as a predictor of the disease outcome. Considering the huge burden on medical infrastructure and the socio-economic state due to the SARS-

Table 2
Calculation of carrier frequencies for ethnicities.

Ethnicity	p	q	2pq	2pq (%)
African	0.986686	0.013314	0.026273	2.627285
Amish	0.873333	0.126667	0.221244	22.12444
Latino	0.936201	0.063799	0.119457	11.94568
Ashkenazi Jewish	0.979844	0.020156	0.0395	3.950031
East Asian	0.998724	0.001276	0.002549	0.254939
Finnish	0.972334	0.027666	0.053802	5.380208
European	0.933759	0.066241	0.123707	12.37069
Other	0.940576	0.059424	0.111786	11.17862
South Asian	0.997706	0.002294	0.004577	0.457663

CoV-2 pandemic associated condition, such predictors can help to segregate the specific cohort of susceptible people only for focused medical attention.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mehy.2021.110485.

References

[1] Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 2020;181:271. https://doi.org/10.1016/j.cell.2020.02.052.

[2] Clinckemalie L, Spans L, Dubois V, Laurent M, Helsen C, Joniau S, et al. Androgen regulation of the TMPRSS2 gene and the effect of a SNP in an androgen response element. Mol Endocrinol 2013;27:2028–40. https://doi.org/10.1210/me.2013-1098.

[3] Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide association study of severe Covid-19 with respiratory failure. N Engl J Med 2020. https://doi.org/10.1056/NEJMoa2020283. NEJMoa2020283.

[4] Oguntuyo KY, Stevens CS, Siddiquey MN, Schilke RM, Woolard MD, Zhang H, et al. In plain sight: the role of alpha-1-antitrypsin in COVID-19 pathogenesis and therapeutics. Microbiology 2020. https://doi.org/10.1101/2020.08.14.248880.

[5] Pidathala S. Overview of the structural and functional impact of protein glycosylation in mediating protein – protein interactions. Trends Carbohydr Res 2015;7:30–40.

[6] GTEx Multi Gene Query. GTEx Portal 2020. https://gtexportal.org/home/multiGeneQueryPage/PRSS2,SERPINA1.

[7] On behalf of the ACMG Laboratory Quality Assurance Committee, Richards S, Aziz N, Bale S, Bick D, Das S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–23. https://doi.org/10.1038/gim.2015.30.

[8] Vianello A, Braccioni F. Geographical overlap between Alpha-1 antitrypsin deficiency and COVID-19 infection in Italy: casual or causal? Arch Bronconeumol 2020;56:609–10. https://doi.org/10.1016/j.arbres.2020.05.015.

[9] Greulich T, Nell C, Hohmann D, Grebe M, Janciauskiene S, Koczulla AR, et al. The prevalence of diagnosed α_1-antitrypsin deficiency and its comorbidities: results from a large population-based database. Eur Respir J 2017;49:1600154. https://doi.org/10.1183/13993003.00154-2016.

[10] De Serres FJ, Blanco I, Fernández-Bustillo E. Estimates of PI*S and PI*Z Alpha-1 antitrypsin deficiency alleles prevalence in the Caribbean and North, Central and South America. Monaldi Arch Chest Dis 2016;71. https://doi.org/10.4081/monaldi.2009.354.

[11] Loyola MB, Reis TTA, Oliveira GXLM, Fonseca Palmeira J, Argañaraz GA, Argañaraz ER. Alpha-1-antitrypsin: a possible host protective factor against Covid-19. Rev Med Virol 2020. https://doi.org/10.1002/rmv.2157.

[12] Wichmann D, Sperhake J-P, Lütgehetmann M, Steurer S, Edler C, Heinemann A, et al. Autopsy findings and venous thromboembolism in patients with COVID-19: a prospective cohort study. Ann Intern Med 2020;173:268–77. https://doi.org/10.7326/M20-2003.