The Journal of Infectious Diseases

BRIEF REPORT

Associations Between Genetically Predicted Protein Levels and COVID-19 Severity

Jingjing Zhu,[1] Chong Wu,[2] and Lang Wu[1]

[1]Cancer Epidemiology Division, Population Sciences in the Pacific Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, Hawaii, USA, and [2]Department of Statistics, Florida State University, Tallahassee, Florida, USA

It is critical to identify potential causal targets for SARS-CoV-2, which may guide drug repurposing options. We assessed the associations between genetically predicted protein levels and COVID-19 severity. Leveraging data from the COVID-19 Host Genetics Initiative comparing 6492 hospitalized COVID-19 patients and 1 012 809 controls, we identified 18 proteins with genetically predicted levels to be associated with COVID-19 severity at a false discovery rate of <0.05, including 12 that showed an association even after Bonferroni correction. Of the 18 proteins, 6 showed positive associations and 12 showed inverse associations. In conclusion, we identified 18 candidate proteins for COVID-19 severity.

Keywords. proteins; COVID-19 severity; genetic instruments.

Coronavirus disease 2019 (COVID-19) has become a global pandemic and brings a huge public health burden. Previous work has identified that specific proteins such as ACE2 and DC-SIGN are essential for the entry of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) into human cells [1, 2]. While remdesivir that blocks such targets has been approved for emergence use to treat COVID-19, currently, there is no cure for COVID-19, highlighting a critical need to identify additional causal targets for guiding more drug repurposing, a strategy for identifying new medical uses of existing drugs. A causal target is expected to be causally associated with COVID-19 severity. However, identifying causal targets is very challenging due to the inherent limitations of conventional study designs and insufficient biologic understanding of many human proteins.

To reduce these limitations, we leveraged genetic variants associated with blood protein levels as instruments to assess the associations between genetically predicted protein levels and COVID-19 severity. Because of the random assortment of alleles transferred from parents to offspring during gamete formation, this approach should be less susceptible to selection bias, reverse causation, and confounding effects [3]. Over the past few years, genome-wide association studies (GWAS) have identified hundreds of protein quantitative trait loci (pQTL) [4, 5]. Many of these genetic variants can serve as strong instrumental variables for evaluating the associations of genetically predicted protein levels with COVID-19 severity in a sufficiently powered study. In this study, we leveraged the enriched data from 6492 hospitalized COVID-19 patients and 1 012 809 population controls included in the COVID-19 Host Genetics Initiative (HGI) for discovery [6].

METHODS

Instrumental Variables for Blood Protein Levels

We extracted genetic instruments for blood proteins based on a comprehensive GWAS of 2731 and 831 healthy European-ancestry participants from the INTERVAL study [7]. Detailed information for the instrument determination has been described in our previous work [3, 8, 9]. In brief, in this study, the genetic associations between 1927 variants and 1478 proteins showed a meta-analysis of $P < 1.5 \times 10^{-11}$ after combing results from the two subcohorts, as well as the consistent direction of effect and nominal significance ($P < .05$) in the 2 subcohorts. We used these pQTLs (Sun et al Supplementary Table 4 [7]) to construct the instrumental variables. We only retained independent variants from each other ($r^2 < 0.1$ based on 1000 Genomes Project phase 3 version 5 data of European populations) for each protein.

Genetic Association Datasets for COVID-19 Severity

For evaluation of the association with COVID-19 severity, we used summary statistics data of the most recent version of GWAS analyses from the COVID-19 HGI (release 4 alpha, September 2020) [6]. Detailed information on participating studies, quality control, and analyses has been provided on the COVID-19 HGI website (https://www.covid19hg.org/results/). In brief, data from 6492 hospitalized COVID-19 patients and 1 012 809 population controls from studies of the Amsterdam University Medical Center COVID study group, Biobanque Quebec COVID19, COVID19-Host(a)ge, GEN-COVID, reCOVID, deCODE, FinnGen, Genetic Modifiers for COVID-19 Related Illness, Genetic Determinants of COVID-19 Complications in the Brazilian Population, Genetics of COVID-Related Manifestation, Penn Medicine Biobank, Qatar Genome Program, Determining the Molecular Pathways and Genetic Predisposition of the Acute Inflammatory Process

Received 11 September 2020; editorial decision 13 October 2020; accepted 14 October 2020; published online October 21, 2020.

Correspondence: Lang Wu, PhD, Cancer Epidemiology Division, Population Sciences in the Pacific Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, HI, 96813 (lwu@cc.hawaii.edu).

The Journal of Infectious Diseases® 2020;XX:1–4

© The Author(s) 2020. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail: journals.permissions@oup.com.

DOI: 10.1093/infdis/jiaa660

Caused by SARS-CoV-2, Ancestry, The genetic predisposition to Severe COVID-19, genomiCC, Genes and Health, and UK Biobank were used. Hospitalized COVID-19 cases represented patients with (1) laboratory-confirmed SARS-CoV-2 infection (RNA and/or serology based); and (2) hospitalization due to corona-related symptoms. Controls represent those that were not cases. The majority of the included subjects were European, with a small proportion of other ethnic groups (756 cases and 1637 controls admixed American; 69 cases and 6500 controls East Asian; 62 cases and 27 353 controls South Asian; and 66 cases and 8536 controls African; and 60 cases and 13 360 controls Arab). Only variants with imputation quality > 0.6 were retained. Meta-analysis of individual studies was performed with inverse variance weighting.

Ethics Committee Approval

Participating studies of the COVID-19 HGI have been approved by ethics committees of the involved institutes [6].

Association Analysis Between Genetically Predicted Protein Levels and COVID-19 Severity

We applied the widely used inverse variance weighted method [10] to estimate the associations of examined proteins with COVID-19 severity. In brief, the β coefficient of the association between genetically predicted protein levels and COVID-19 severity was estimated using $\sum_i \beta_{i,GX} * \beta_{i,GY} * \sigma_{i,GY}^{-2} / (\sum_i \beta_{i,GX}^2 * \sigma_{i,GY}^{-2})$, and the corresponding standard error was estimated using $1/(\sum_i \beta_{i,GX}^2 * \sigma_{i,GY}^{-2})^{0.5}$. $\beta_{i,GX}$ represented the β coefficient of the association between i th single-nucleotide polymorphism (SNP) and each protein of interest; and $\beta_{i,GY}$ and $\sigma_{i,GY}$ represented the β coefficient and standard error, respectively, for the association between i th SNP and COVID-19 severity. The association odds ratio (OR), confidence interval, and P value were further estimated based on the calculated β coefficient and standard error. A Benjamini-Hochberg false discovery rate of <0.05 was used to adjust for multiple comparisons.

RESULTS

In the analysis of the COVID-19 HGI dataset, of the 1357 proteins assessed, we identified 18 proteins with genetically predicted levels to be associated with COVID-19 severity at a false discovery rate of < 0.05 (Table 1 and Supplementary Table 1). A positive association between predicted protein level and COVID-19 severity was detected for DC-SIGN, BGAT, B3GN2, C1GLC, SCF, and FA20B (ORs ranging from 1.09 to 1.66). Conversely, an association between a lower predicted protein level and increased COVID-19 severity was identified for ST4S6, IGF-I sR, Endoglin, sICAM-2, LIF sR,

Table 1. Significant Protein–COVID-19 Severity Associations

Protein		Protein-Encoding Gene	Region	Instrument Variants	Type of pQTL	COVID-19 Host Genetics Initiative			
						OR[a]	95% CI[a]	*P* Value	FDR *P* Value[b]
sE-Selectin	E-selectin	*SELE*	1q24.2	rs2519093	trans	0.88	.83–.93	7.29×10^{-6}	.001
FA20B	Glycosaminoglycan xylosylkinase	*FAM20B*	1q25.2	rs587729126	trans	1.66	1.26–2.20	3.22×10^{-4}	.03
B3GN2	*N*-acetyllactosaminide β-1,3-*N*-acetylglucosaminyltransferase 2	*B3GN2*	2p15	rs2519093	trans	1.66	1.33–2.06	7.29×10^{-6}	.001
LIF sR	Leukemia inhibitory factor receptor	*LIFR*	5p13.1	rs635634	trans	0.60	.47–.75	8.71×10^{-6}	.001
Met	Hepatocyte growth factor receptor	*MET*	7q31.2	rs635634	trans	0.66	.55–.80	8.71×10^{-6}	.001
Endoglin	Endoglin	*ENG*	9q34.11	rs635634	trans	0.52	.39–.70	8.71×10^{-6}	.001
BGAT	Histo-blood group ABO system transferase	*ABO*	9q34.2	rs505922	cis	1.09	1.06–1.14	1.34×10^{-6}	9.10×10^{-4}
ST4S6	Carbohydrate sulfotransferase 15	*CHST15*	10q26.13	rs550057	trans	0.62	.51–.76	2.40×10^{-6}	.001
COX8A	Cytochrome c oxidase subunit 8A, mitochondrial	*COX8A*	11q13.1	rs2232613	trans	0.82	.75–.90	3.79×10^{-5}	.004
SCF	Kit ligand	*KITLG*	12q21.32	rs6065904	trans	1.54	1.20–1.97	6.33×10^{-4}	.05
OAS1	2′-5′-oligoadenylate synthase 1	*OAS1*	12q24.13	rs4767027, rs62143197	cis, trans	0.75	.64–.87	2.45×10^{-4}	.02
IGF-I sR	Insulin-like growth factor 1 receptor	*IGF1R*	15q26.3	rs635634	trans	0.49	.36–.67	8.71×10^{-6}	.001
AT2A3	Sarcoplasmic/endoplasmic reticulum calcium ATPase 3	*ATP2A3*	17p13.2	rs6065904	trans	0.65	.51–.83	6.33×10^{-4}	.05
sICAM-2	Intercellular adhesion molecule 2	*ICAM2*	17q23.3	rs587729126	trans	0.56	.41–.77	3.22×10^{-4}	.03
IR	Insulin receptor	*INSR*	19p13.2	rs507666	trans	0.80	.72–.88	8.04×10^{-6}	.001
DC-SIGN	CD209 antigen	*CD209*	19p13.2	rs505922	trans	1.15	1.09–1.22	1.34×10^{-6}	9.10×10^{-4}
IL-3 Ra	Interleukin-3 receptor subunit α	*IL3RA*	Xp22.33	rs2519093	trans	0.83	.77–.90	7.29×10^{-6}	.001
C1GLC	C1GALT1-specific chaperone 1	*C1GALT1C1*	Xq24	rs7787942, rs2519093	trans, trans	1.24	1.13–1.36	9.88×10^{-6}	.001

Abbreviations: CI, confidence interval; FDR, false discovery rate; OR, odds ratio; pQTL, protein quantitative trait loci.

[a]OR and CI per 1 standard deviation increase in genetically predicted protein levels and *P* value are derived from association analyses of 6492 hospitalized patients and 1 012 809 population controls (2-sided).

[b]FDR *P* value, FDR adjusted *P* value. Associations with FDR $P \leq .05$ considered statistically significant.

AT2A3, Met, OAS1, IR, COX8A, IL-3 Ra, and sE-Selectin (ORs ranging from 0.49 to 0.88). For 12 of the proteins (DC-SIGN, BGAT, IGF-I sR, Endoglin, LIF sR, Met, IR, IL-3 Ra, sE-Selectin, B3GN2, C1GLC, and ST4S6), their associations were significant even at the Bonferroni-corrected threshold ($0.05/1357 = 3.68 \times 10^{-5}$).

DISCUSSION

This is the first study to evaluate the associations of genetically predicted protein levels with COVID-19 severity using GWAS-identified pQTLs as instruments. We identified 18 proteins that demonstrated a statistically significant association. Our study provides novel information to improve the understanding of potentially causal molecular targets for SARS-CoV-2, and the identified promising proteins could potentially guide drug repurposing efforts, which holds the promise of significantly reducing the public health burden of COVID-19.

Of the identified proteins, DC-SIGN has been newly reported to act as a receptor for SARS-CoV-2 and is differentially expressed in lung and kidney epithelial and endothelial cells [1]. BGAT is the basis of the ABO blood group system. In a recent GWAS for COVID-19 severity, a significant association signal has been reported for the ABO blood group locus [11]. Interestingly, it was also identified that blood group A was associated with higher risk of acquiring COVID-19 while blood group O was associated with lower risk [11]. More in-depth work to better characterize the exact roles of other identified proteins is needed.

Previous research suggests that ACE2, TMPRSS2, and L-SIGN are also essential for the entry of SARS-CoV-2 into human cells [1, 2]. Of these, TMPRSS2 and L-SIGN, are not measured in the INTERVAL study. For ACE2, although it is measured in the INTERVAL, there was no corresponding pQTL identified in the study [7]; thus it was not investigated in the current study.

There are several potential limitations in our study. Firstly, in GWAS of the COVID-19 HGI, the population composition is mixed. However, a majority of the included subjects were European, and it is expected that this will not significantly influence our findings. Secondly, aligned with the above point, the current study primarily focuses on analyses of Europeans. Whether the identified proteins also demonstrate associations in other ethnic groups require further investigation. However, the foundation of disease biology should be similar across populations of different ethnic backgrounds, thus it is anticipated that findings of this study should be generalizable to other populations of non-Europeans. Thirdly, due to the inclusion of multiple studies from different countries in COVID-19 HGI, it is possible that the included cases, although all were hospitalized COVID-19 patients, are not entirely homogeneous. It is possible that the criteria for hospitalization of COVID-19 patients are different across different regions, thus measurement errors may exist in this study. Fourthly, the possibility of a pleiotropy effect cannot be excluded. For example, rs505922 is the instrument for BGAT and DC-SIGN. Previous studies have also identified associations between this variant and several other outcomes, such as type 2 diabetes, pancreatic cancer, and venous thromboembolism [12–14]. It is known that type 2 diabetes is associated with increased risk of severe COVID-19 outcomes [15]. Further studies are needed to validate our identified protein–COVID-19 associations. Fifthly, our analysis could be constrained by the pQTLs identified in previous GWAS of protein levels. As discussed above, we were not able to evaluate some important COVID-19–associated proteins. We anticipate that additional protein targets could be identified when further pQTLs are reported. More comprehensive genetic prediction models for protein levels could provide improved power to characterize additional COVID-19–associated proteins. Furthermore, the pQTL instruments used in this study are based on blood tissue. Blood tissue could reflect the systematic pattern of the body, as well as capture immune-related pathways which play a vital role in the host response to viral infection. On the other hand, it is known that the relevant tissue for the entry of SARS-CoV-2 into humans is lung, and future work that uses genetic instruments generated in lung tissue would be useful to identify additional promising targets. Lastly, in our study, the SARS-CoV-2 infection status was largely unknown for the control participants. On the other hand, if infected subjects were included in the control group, it is expected that this would only bias our results toward the null.

Compared with conventional observational studies, the design using genetic instruments could potentially avoid many biases and confounding issues existing in traditional studies. It is anticipated that COVID-19 GWAS datasets involving much larger sample sizes will be available in the near future. Well-conducted proteome-wide association studies using genetic instruments are warranted to identify additional proteins that are potentially related to COVID-19 severity. Such findings will be critical to guide drug repurposing efforts to reduce the COVID-19 burden.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. We acknowledge the COVID-19 Host Genetics Initiative for making their COVID-19 GWAS summary statistics available.

Disclaimer. The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Financial support. This work was supported by the University of Hawaii Cancer Center and Florida State University.

Potential conflicts of interest. All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Amraie R, Napoleon MA, Yin W, et al. CD209L/L-SIGN and CD209/DC-SIGN act as receptors for SARS-CoV-2 and are differentially expressed in lung and kidney epithelial and endothelial cells. bioRxiv, doi: 10.1101/2020.06.22.165803, **23** June **2020**, preprint: not peer reviewed.
2. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell **2020**; 181:271–80.e8.
3. Wu L, Shu X, Bao J, et al; PRACTICAL, CRUK, BPC3, CAPS, PEGASUS Consortia. Analysis of over 140 000 European descendants identifies genetically predicted blood protein biomarkers associated with prostate cancer risk. Cancer Res **2019**; 79:4592–8.
4. Enroth S, Johansson A, Enroth SB, Gyllensten U. Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. Nat Commun **2014**; 5:4684.
5. Suhre K, Arnold M, Bhagwat AM, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun **2017**; 8:14357.
6. COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur J Hum Genet **2020**; 28:715–8.
7. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. Nature **2018**; 558:73–9.
8. Shu X, Bao J, Wu L, et al. Evaluation of associations between genetically predicted circulating protein biomarkers and breast cancer risk. Int J Cancer **2020**; 146:2130–8.
9. Zhu J, Shu X, Guo X, et al. Associations between genetically predicted blood protein biomarkers and pancreatic cancer risk. Cancer Epidemiol Biomarkers Prev **2020**; 29:1501–8.
10. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol **2013**; 37:658–65.
11. Severe COVID-19 GWAS Group, Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide association study of severe COVID-19 with respiratory failure. N Engl J Med **2020**; 383:1522–34.
12. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet **2009**; 41:986–90.
13. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet **2018**; 50:1505–13.
14. Trégouët DA, Heath S, Saut N, et al. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. Blood **2009**; 113:5298–303.
15. Apicella M, Campopiano MC, Mantuano M, Mazoni L, Coppelli A, Del Prato S. COVID-19 in people with diabetes: understanding the reasons for worse outcomes. Lancet Diabetes Endocrinol **2020**; 8:782–92.