ARIC Manuscript Proposal #3687

PC Reviewed: 8/11/20	**Status: _____**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Proteomic-Based Examination of Two Gene Variants Related to Covid-19 Respiratory Failure

b. Abbreviated Title (Length 59 characters): Proteomics of Covid-19-related respiratory failure variants

2. Writing Group: Brian Steffen, Jim Pankow, Weihong Tang, Ryan Demmer, Faye Norby, Pamela Lutsey, Logan Cowan, Joe Coresh. We welcome any further nominations.

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __

First author: Brian Steffen
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 S. 2nd St., Suite 300
Minneapolis, MN 55454
Phone: 612-626-0371
E-mail: steff293@umn.edu

Senior author: Weihong Tang
Address: Division of Epidemiology and Community Health
University of Minnesota
1300 S. 2nd St., Suite 300
Minneapolis, MN 55454

3. Timeline: Begin analysis in July, 2020
Draft of paper by October, 2020

4. Rationale:

The outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corresponding coronavirus disease 2019 (Covid-19) were first reported in December of 2019. Within seven months, Covid-19 has resulted in over half a million deaths (1) and a cumulative loss of global GDP in the trillions ($) (2). Initial descriptive reports have shown greater relative risks of Covid-19 mortality and disease severity among males, the elderly, and in those with comorbidities including diabetes, hypertension, and underlying respiratory or cardiovascular disease (3). However, in absolute terms, large numbers of cases have been observed in young or otherwise healthy individuals resulting in hospitalizations, acute respiratory failure, or death. It is therefore likely that other factors are involved in Covid-19 pathogenicity—particularly with respiratory failure. Genetic factors were recently suggested to have a role.

To identify host genetic risk factors that contribute to severe Covid-19 typified by respiratory failure, a genome-wide association study was conducted in 1610 cases and 2205 controls among Italian and Spanish individuals (4). Two independent gene variants were found to be associated with greater likelihoods of developing severe Covid-19 with respiratory failure—rs11385942 (OR: 1.77; 95% CI: 1.48–2.11; $p=1.15\times10^{-10}$) and rs657152 (OR: 1.32; 95% CI: 1.20–1.47; $p=4.95\times10^{-8}$). Critically, this latter variant was previously shown to be associated with greater risk of SARS-CoV-1 infection (5) and is related to ABO blood type. Indeed, a sub-analysis showed that individuals with blood type A were likely to develop Covid-19-related respiratory failure, while those with blood type O were less susceptible (4). The mechanism that explains this phenomenon remains unclear but may involve the von Willebrand factor protein (4). In contrast to the localized rs657152 polymorphism, the rs11385942 variant at locus 3p21.31 spans a cluster of six genes: *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCR1*. The risk allele was associated with lower expression of *CXCR6* and greater expressions of *SLC6A20* and *LZTFL1*. While the mechanisms through which these genes may directly or indirectly influence respiratory failure development are unclear, *CXCR6* or *SLC6A20* may be involved based on their encoded proteins. The investigators called for further studies to determine the functional consequences of the detected associations—this serves as the basis for the current proposal.

The identification of these gene loci (4) represents a critical step in advancing our understanding of severe Covid-19 and an opportunity to elucidate its underlying mechanisms. Using a protein quantitative trait loci approach, we propose to examine the proteomic signatures of carriers and non-carriers of variants rs11385942 and rs657152 in a sample of 9,358 Atherosclerosis Risk in Communities study participants. Proteomic outcomes associated with these variants may provide evidence of the mechanisms that affect susceptibility to Covid-19 respiratory failure. However, it must be acknowledged that blood samples in ARIC were collected prior to the pandemic, and our approach is predicated on carriers of these variants showing relevant physiological differences in their respective plasma proteomes in the absence of infection. We will be unable to evaluate the possibility of a gene-environment interaction whereby differences in proteomes can only be detected once individuals are infected or

otherwise develop Covid-19. With that stated, genetic variant-proteome associations may identify targets or pathways through which risk for respiratory failure is increased and provide information about disease pathogenesis as well as potential pharmacological targets for treatment.

5. Main Hypothesis/Study Questions:

Main Hypothesis: Variants rs11385942 and rs657152 will be associated with plasma proteomic targets, which may suggest pathways/mechanisms of COVID-19-related respiratory failure pathogenesis.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Design: cross-sectional. Proteomic data were measured at Visit 3 samples using SOMAscan v 4.0. This approach quantifies 4,871 human proteins through an aptamer-based platform. The substantial protein coverage of SOMAscan v 4.0 and the over 11,000 ARIC participants with available data provide the opportunity to characterize proteomic signatures. Visit 3 was selected for the primary analysis due to the larger sample size, fewer comorbidities, and less extensive medication use.

Inclusion/exclusion: those with missing or low quality proteomic or gene variant data, missing covariates, or individuals who identified as other than white or Black will be excluded. These criteria result in a final sample of 9,358 study participants.

Exposure: gene variants rs11385942 and rs657152

Outcome: natural log2-transformed protein levels will serve as the outcome variable, and ARIC analytic recommendations for the SomaScan proteomic data will be followed. Non-human proteins and proteins with unacceptable QC will be removed, e.g. that that have large CVs (> 20%), poor reproducibility between the blind duplicate pairs, or non-specific binding.

Ingenuity Pathway Analysis will be used to assist in interpreting the proteomic findings and identify potential signaling pathways, molecular networks, upstream regulators, and biological functions of proteins found to be associated with the variant exposures.

Secondary outcomes: For plasma proteins found to relate to rs11385942 and rs657152, we propose to conduct secondary analyses to determine whether the identified proteins may be related to other respiratory and infection outcomes including declining pulmonary function as

determined from tests at visits 1, 2, and 5 as well as pulmonary infections, infection-related hospitalizations (N=2,701) and mortality (N=523).

Data analysis: the primary analysis will be cross-sectional in design and multiple linear regression analysis will determine associations between each SNP and protein levels. For each SNP, the number copy of the risk allele as reported in the GWAS by Ellinghaus et al. (2020) will be modeled. We will examine associations in Black and white individuals separately and subsequently conduct a meta-analysis of the two samples. Environmental factors are not expected to affect observations; however, covariates will include age at visit 3, sex, principal components of ancestry, estimated glomerular filtration rate to control for kidney function, and field center. Statistical significance will be determined after Bonferroni correction for the number of proteins and SNPs tested.

Secondary analyses will be prospective in design and estimate associations between each protein identified in the primary analysis with: 1) pulmonary function test outcomes; 2) pulmonary infections; 3) infection-related hospitalizations and mortality. Cox proportional hazards regression will be conducted with adjustments for age, sex, race, center, smoking status, alcohol consumption, years of education, hypertension, immunosuppressive drugs, and diabetes as defined by self-reported physician diagnosis of diabetes; taking an antidiabetic drug; a fasting glucose ≥126 mg/dL. Interactions by age category, race and sex will be evaluated by including cross-product terms in the models. The proportional hazards assumption will be tested.

Limitations: As stated above, Visit 3 samples in ARIC were collected prior to the COVID-19 pandemic, and our approach relies on the detection of associations between the exposures and plasma protein outcomes in the absence of infection. We cannot evaluate the possibility of a gene-environment interaction in which protein expressions only diverge upon SARS-CoV-2 infection or otherwise development of Covid-19.

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes __x__ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? ____ Yes ____ No
(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? __x__ Yes ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"? __x__ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at:

___x___ Yes _____ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

No investigators have so far proposed to examine proteomic signatures associated with these phenotypic PRSs.

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? _x___ Yes ____ No

11.b. If yes, is the proposal

_x__ A. primarily the result of an ancillary study (list number* _AS2017.27___)

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

*ancillary studies are listed by number at https://www2.cscc.unc.edu/aric/approved-ancillary-studies

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*; published online Feb 19.
2. Makridis CA, Hartley JS. The Cost of COVID-19: A Rough Estimate of the 2020 US GDP Impact. . *Special Edition Policy Brief* (2020).
3. Yang J, Zheng Y, Gou X, 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;94:91-95.
4. Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure [published online ahead of print, 2020 Jun 17]. N Engl J Med. 2020;NEJMoa2020283. doi:10.1056/NEJMoa2020283
5. Cheng Y, Cheng G, Chui CH, et al. ABO blood group and susceptibility to severe acute respiratory syndrome [published correction appears in JAMA. 2005 Aug 17;294(7):794. Cheng, Yufeng [corrected to Cheng, Yunfeng]]. JAMA. 2005;293(12):1450-1451.