








RESEARCH LETTER

Phenome-Wide Association Study of Severe COVID-19 Genetic Risk Variants

Jessica A. Regan , MD; Jawan W. Abdulrahim , MD; Nathan A. Bihlmeyer , PhD; Carol Haynes , BA; Lydia Coulter Kwee , PhD; Manesh R. Patel , MD; Svati H. Shah , MD, MS, MHS

CCOVID 19 exhibits clinical heterogeneity, ranging from mild flulike illness to severe respiratory failure. The biologic underpinnings for this heterogeneity are unclear, although genetic factors for risk of severe COVID 19 have been identified. To date, the COVID 19 Host Genetics Initiative has identified 10 genome-wide significant loci associated with severe COVID 19 infection.¹ As greater comorbidities have been observed in patients with severe COVID 19 infection,² we aimed to identify comorbidities associated with these genetic loci using a phenome-wide association study (PheWAS), to better understand potential conjoint genetic risk of severe COVID 19 and comorbidities mediated by these variants. One such PheWAS study has been conducted for the 3p21.31 locus³; however, additional phenotypic associations and broader implications of risk for COVID 19 genetic loci have not been described.

This study included 247 488 unrelated White participants from UK Biobank, a prospective population-based cohort in the United Kingdom with genetic and phenotypic data collected on individuals aged 40 to 69 years.⁴ Directly genotyped or imputed data for the genetic loci of interest were obtained from either the UK Biobank or the UK BiLEVE Axiom array. The 10 severe COVID 19 associated loci studied here are rs11385942 (3p21.31, risk allele GA),⁵ rs1886814 (*FOXP4*, C), rs72711165 (*TMEM65*, C), rs657152 (*ABO*, A), rs10735079 (*OAS1*, A), rs1819040 (*KANSL1*, A), rs77534576 (*TAC4*, T), rs74956615 (*TYK2*, A), rs2109069 (*DPP9*, A), and rs2236757 (*IFNAR2*, G). Phenotype data were derived from *International Classification of*

Diseases, Ninth Revision (ICD-9) and *Tenth Revision (ICD-10)* codes from primary care, hospitalizations, and death-related data. Data were accessed through approved UK Biobank applications (IDs 48785, 65043). The validation cohort consisted of 2247 White individuals from CATHGEN (Catheterization Genetics), a study of sequential individuals who underwent cardiac catheterization at Duke University Medical Center (Durham, NC) between 2001 and 2010. Genotypes were obtained using the Illumina Human Omni1-Quad Infinium Bead Chip and imputed with Minimac4 using 1000G phase 3 reference panels. Phenotype data were derived from electronic health record data from 2001 to 2020. Both studies were approved by the Duke Institutional Review Board and all participants provided informed consent. All data are available upon request.

In UK Biobank, 1402 phenotypic outcomes (minimum ≥ 20 occurrences) were included, and 866 in CATHGEN were included. The R PheWAS package (v0.99.0.5-5) was used to perform logistic regression for each outcome adjusted for age, genotyping array, sex, and 5 principal components (R v4.0.2). Significant phenotypes were considered at a false discovery rate-adjusted q-value < 0.05 in UK Biobank and nominally ($P < 0.1$) in CATHGEN for validation.

Four of the 10 tested genetic loci showed significant phenotype associations in UK Biobank after false discovery rate adjustment. Vascular dementia was associated with rs72711165 (*TMEM65*) (odds ratio [OR], 5.66; 95% CI, 2.21–11.85; $q = 0.049$), but did not validate in CATHGEN (Figure – Panel A). There were 40 associations with the rs657152 risk allele (*ABO*), including

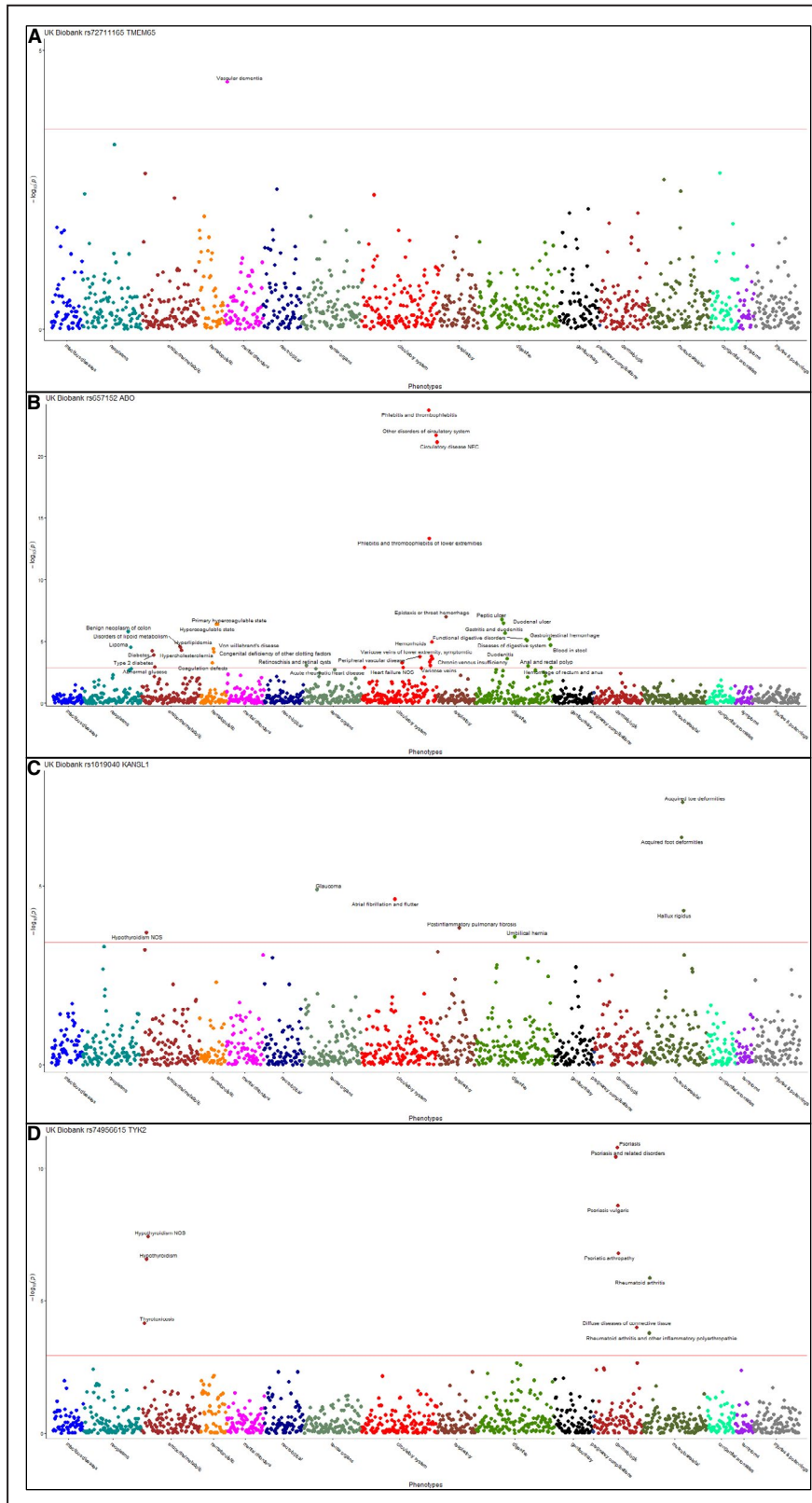
Key Words: cardiovascular disease ■ COVID 19 ■ genetics ■ inflammation

Correspondence to: Svati H. Shah, MD, MS, MHS, 300 N. Duke Street, DUMC Box 104775, Durham, NC 27701. E-mail: svati.shah@duke.edu

For Sources of Funding and Disclosures, see page 3.

© 2022 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

JAHA is available at: www.ahajournals.org/journal/jaha



novel associations with greater odds of heart failure (OR, 1.09; 95% CI, 1.03–1.14; $q=0.046$), diabetes (OR, 1.05; 95% CI, 1.02–1.07; $q=0.004$) and hypercholesterolemia (OR, 1.04; 95% CI, 1.02–1.06; $q=0.004$); and

lower odds of gastrointestinal disorders including duodenal ulcer (OR, 0.88; 95% CI, 0.84–0.92; $q=6.3 \times 10^{-5}$, Figure – Panel B) with nongroup O blood types. Of these, 34 out of 40 were available in CATHGEN, but

Figure 1. Significant phenotypic associations with COVID 19 risk alleles in UK Biobank.

Shown are the results of the significant findings of the phenome-wide association study for severe COVID 19 single nucleotide polymorphisms. The x-axes correspond to the different groups of phenotypes analyzed and the y-axes correspond to the negative logarithm *P* values for these analyses. The red line corresponds to statistical significance level at a false discovery rate <0.05, phenotypes meeting statistical significance are annotated. **A**, Results for rs72711165 8q24.13 *TMEM65*. **B**, Results for rs657152 9q34.2 *ABO*. **C**, Results for rs1819040 17q21.31 *KANSL1*. **D**, Results for rs74956615 19p13.2 *TYK2*. NOS indicates not otherwise specified.

none of these findings validated. Eight phenotypes associated with rs1819040 (*KANSL1*), including atrial fibrillation and flutter (OR, 1.07; 95% CI, 1.04–1.10; $q=0.0084$) and pulmonary fibrosis (OR, 0.80; 95% CI, 0.71–0.89; $q=0.035$) (Figure – Panel C); only glaucoma validated in CATHGEN ($P<0.1$). These results suggest that genetic predisposition for these cardiovascular and endocrine phenotypes may amplify the risk of adverse COVID 19 outcomes but may also have broader long-term health implications.

Ten phenotypes associated with rs74956615 (*TYK2*), all with lower odds associated with the COVID 19 risk allele (Figure – Panel D), including psoriatic arthropathy (OR, 0.31; 95% CI, 0.20–0.47; $q=4.5\times 10^{-5}$), rheumatoid arthritis (OR, 0.83; 95% CI, 0.64–0.83; $q=0.0003$) and thyrotoxicosis (OR, 0.77; 95% CI, 0.68–0.87; $q=0.01$). Seven of these phenotypes nominally validated in CATHGEN: psoriasis, rheumatoid arthritis, and hypothyroidism (all $P<0.1$). COVID 19-related genetic variants suggest the importance of host antiviral defense mechanisms and inflammatory signaling. *TYK2* is implicated in psoriasis via Th17 responses and interferon- α signaling. At the *TYK2* locus, we clarified prior discordant associations for autoimmune disease, showing for the first time decreased odds of psoriasis associated with rs74956615, which may implicate a distinct impact of this allele on *TYK2* gene function from what has been identified in prior genome-wide association study analysis of psoriasis.

Using an unbiased PheWAS approach to clinical diagnoses in a large data set, we identified novel phenotypic associations between risk alleles for severe COVID 19 infection with relevant comorbidities. These associations suggest that individuals carrying these genetic markers, known for their role in blood traits, host antiviral response and inflammation, may have modified risk of cardiovascular disease, as well as autoimmune and inflammatory disorders, which in turn increases risk of severe COVID 19. Alternatively,

these genetic risk loci may have pleiotropic effects on these diseases and COVID 19 related complications. Limitations to this study include the underpowered validation sample and restriction to individuals of European ancestry.

ARTICLE INFORMATION

Received September 16, 2021; accepted January 3, 2022.

Affiliations

Duke Molecular Physiology Institute, Durham, NC (J.A.R., J.W.A., N.A.B., C.H., L.C.K., S.H.S.); Duke University Department of Medicine, Durham, NC (J.A.R., J.W.A., M.R.P., S.H.S.); and Duke Clinical Research Institute, Durham, NC (M.R.P., S.H.S.).

Acknowledgments

We thank the UK Biobank and CATHGEN participants whose data contributed to this work.

Sources of Funding

This research was supported by grants from the National Heart, Lung, and Blood Institute (1R38HL143612 [Regan] and 1R21-AI158786 [Shah]).

Disclosures

None.

REFERENCES

1. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature*. 2021;600:472–477. doi: 10.1038/s41586-021-03767-x
2. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*. 2020;584:430–436. doi: 10.1038/s41586-020-2521-4
3. Zhou J, Sun Y, Huang W, Ye K. Altered blood cell traits underlie a major genetic locus of severe COVID-19. *J Gerontol A Biol Sci Med Sci*. 2021;76:e147–e154. doi: 10.1093/gerona/glab035
4. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203–209. doi: 10.1038/s41586-018-0579-z
5. Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, Fernández J, Prati D, Baselli G, Asselta R, et al. Genomewide Association Study of Severe Covid-19 with respiratory failure. *N Engl J Med*. 2020;383:1522–1534. doi: 10.1056/NEJMo a2020283