A Phenome-Wide Association Study of genes associated with COVID-19 severity reveals shared genetics with complex diseases in the Million Veteran Program

Anurag Verma, PhD[1,9¶]; Noah Tsao, BS[1,11¶]; Lauren Thomann, MPH[2]; Yuk-Lam Ho[7]; Rotonya Carr, MD[1,10]; Dana Crawford, PhD[19,22,23]; Jimmy T. Efird, PhD[16]; Jennifer Huffman, PhD[2]; Adriana Hung, MD[17,20]; Kerry Ivey, PhD[9,21,22]; Sudha Iyengar, PhD[4,22]; Michael Levin, MD[10]; Shiuh-Wen Luoh, MD, PhD[3,24]; Julie Lynch, PhD[5]; Pradeep Natarajan, MD, MMSc[2,15,31,32]; Saiju Pyarajan, PhD[2,12]; Alexander Bick MD, PhD[2,28]; Lauren Costa[7]; Giulio Genovese, PhD[32,33,34]; Richard Hauger, MD[35]; Ravi Madduri[36,37]; Gita Pathak PhD[6,42], Renato Polimanti PhD[6,42], Benjamin Voight, PhD[1,9,26,27]; Marijana Vujkovic, PhD[1,10]; Maryam Zekavat, BSc[2,38,39]; Hongyu Zhao, PhD[6,39,40]; Marylyn D Ritchie, PhD[3]; *VA Million Veteran Program COVID-19 Science Initiative*, Kyong-Mi Chang, MD[1,10]; Kelly Cho, PhD[7]; Juan P. Casas, MD, PhD[7,13]; Philip S. Tsao, PhD[8,41]; J. Michael Gaziano, MD, MPH[7,13]; Christopher O'Donnell, MD, MPH [2,12,13]; Scott Damrauer, MD[1,11]*, Katherine Liao, MD, MPH[2,12,13]*

[1] Corporal Michael Crescenz VA Medical Center, Philadelphia, Philadelphia, USA
[2] VA Boston Healthcare System, Boston, Massachusetts, USA
[3] VA Portland Health Care System, Portland OR, USA
[4] Louis Stokes Cleveland VA Medical Center, Cleveland, OH , USA
[5] VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, Utah, USA.
[6] VA Connecticut Healthcare System, West Haven, CT, USA
[7] Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, Massachusetts, USA
[8] VA Palo Alto Health Care System, Palo Alto, California, USA
[9] Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
[10] Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[11] Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[12] Harvard Medical School, Boston, Massachusetts, USA.
[13] Brigham and Women's Hospital, Boston, Massachusetts, USA.
[14] Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA.
[15] Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
[16] Cardiovascular Medicine Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
[17] Division of Hematology and Medical Oncology, Knight Cancer Institute, Oregon Health and Science University, Portland, Oregon, USA
[18] College of Nursing and Health Sciences, University of Massachusetts, Lowell, Massachusetts, USA.
[19] Cleveland Institute for Computational Biology, Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA.
[20] Cooperative Studies Program Epidemiology Center, Health Services Research and Development, DVAHCS (Duke University Affiliate), Durham, North Carolina, USA.
[21] Tennessee Valley Healthcare System (Nashville VA) & Vanderbilt University, Nashville, Tennessee, USA.
[22] Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA.
[23] Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio, USA
[24] Division of Hematology and Medical Oncology, Knight Cancer Institute, Oregon Health and Science University, Portland, Oregon, USA
[25] Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.
[26] Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[27] Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[28] Vanderbilt University, Nashville, Tennessee, USA

[29] South Australian Health and Medical Research Institute, Infection and Immunity Theme, South Australia, Australia
[30] Harvard T.H. Chan School of Public Health, Department of Nutrition, Massachusetts, USA
[31] Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA.
[32] Program in Medical and Population Genetics and the Cardiovascular Disease Initiative, Broad Institute of Harvard & MIT, Cambridge, Massachusetts, USA
[33] Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[34] Department of Genetics, Harvard Medical School, Boston, MA, USA
[35] Department of Psychiatry, University of California, San Diego, La Jolla, CA; Center of Excellence for Stress and Mental Health, VA San Diego Healthcare System, San Diego, CA, USA
[36] University of Chicago Consortium for Advanced Science and Engineering, The University of Chicago, Chicago, Illinois, USA
[37] Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois.
[38] Broad Institute of MIT & Harvard, Cambridge, MA, USA
[39] Yale School of Medicine New Haven, CT, USA
[40] Computational Biology and Bioinformatics Program, Yale University, New Haven, CT, USA
[41] Department of Medicine (Cardiovascular Medicine), Stanford University School of Medicine, Stanford, CA, USA
[42] Department of Psychiatry, Yale School of Medicine, Connecticut, USA

Joint Authorship
¶These authors contributed equally to this work
*These authors jointly supervised this work
Correspondence
Anurag Verma: anurag.verma@pennmedicine.upenn.edu, Katherine Liao: kliao@bwh.harvard.edu

Abstract

Background There are certain pre-existing conditions which leads in increased risk of coronavirus disease 2019 (COVID-19) severity and mortality. The objective of this study is to determine shared genetic architecture between COVID-19 severity and other medical conditions using electronic health record (EHR) data from diverse patient populations.

Methods and Findings: We conducted Phenome-wide association study (PheWAS) of genetic variants associated with severe COVID-19 in two biobanks with EHR and genomic data: 1) Veteran Affairs (VA) Million Veteran Program (MVP), 2) United Kingdom Biobank (UKBB). Genetic variants associated with critical illness (n=48) or hospitalization (n=39) due to COVID-19 from COVID-19 Host Genetics Initiative genome-wide association studies. Phenotypes defined by International Classification of Diseases (ICD) codes mapped to clinically relevant groups using published PheWAS methods; pre-COVID-19 data used to avoid potential confounding. Among 455,683 US Veterans from MVP, variants associated with severe COVID-19 tested for association across 1,559 phenotypes; 353,365 UK Biobank participants, and 1,064 phenotypes tested. Genetic variants at *ABO* locus (rs550057, rs505922) associated with the largest number of phenotypes ($n_{rs55057}$= 53 and $n_{rs505922}$=61); strongest association with venous embolism, odds ratio $(OR)_{rs550057}$ 1.27 (p=5.28 x 10^{-116}), and thrombosis $OR_{rs505922}$ 1.31, p=3.5 x10^{-183}. Among 67 respiratory conditions tested, only idiopathic pulmonary fibrosis, OR $_{rs2277732}$ 1.17, p=1.3410^{-05}, and asthma $OR_{rs143334143}$ 0.94, p=2.31 x10^{-04}, shared variants with severe COVID-19. The *RAVER1* locus (rs74956615) associated with reduced risk for autoimmune conditions, e.g. psoriasis OR 0.71, p= 1.53 x10^{-22}, rheumatoid arthritis, OR 0.78, p=1.04 x 10^{-09}; findings replicated in UKBB. A known functional missense variant (rs34536443, *TYK2*) in the region had the highest linkage disequilibrium with rs74956615, suggesting signal was likely

from *TYK2*. In MVP, PheWAS results stratified by genetic ancestry did not demonstrate significant difference in associations across ancestry.

Conclusions Shared genetic architecture between COVID-19 severity and conditions related to underlying risk factors for severe and poor COVID-19 outcomes; associations similar across genetic ancestries. Divergent association between inflammatory conditions and severe COVID may be explained by known pathway impairing signaling of inflammatory cytokines, reducing risk for autoimmunity; same pathway reduces type 1 interferon signaling, critical for viral host defense. Caution needed when targeting pathways that may balance immune tolerance and immunodeficiency to treat COVID-19.

Introduction

A respiratory disease coronavirus disease-2019 (COVID-19) caused by a novel coronavirus was identified in December 2019,[1] now known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). As of March 2021, the COVID-19 pandemic has resulted in the loss of over 2.6 million lives worldwide.[2] Identifying host genetic variants associated with severe disease from COVID-19 can identify key pathways important in the pathogenesis of SARS-CoV2. International efforts such as the COVID-19 Host Genetics Initiative (HGI)[3] have meta-analyzed genome-wide association study (GWAS) summary statistics at regular intervals to identify novel genetic associations with COVID-19 severity. Thus far, ten independent variants associated with COVID-19 severity at genome-wide significance have been identified, most notably at the *ABO* locus.[4] These GWASs have also identified variations in genes involving inflammatory cytokines and interferon signaling pathways such as *IFNAR2*, *TYK2*, and *DPP9*.[4]

The unprecedented availability of genome-wide data for COVID-19 provides an opportunity to study clinical conditions that share genetic risk factors for COVID-19 severity. Examining known conditions, each with a body of knowledge regarding important pathways and targets, may in turn improve our understanding of pathways relevant for COVID-19 severity and inform the development of novel treatments against this pathogen. The Phenome-Wide Association Study (PheWAS) is an approach for simultaneously testing genetic variants' association with a wide spectrum of conditions and phenotypes.[5] Mega-biobanks such as the Veteran's Affairs (VA) Million Veterans Program (MVP) and the UK Biobank (UKBB), both with genotypic data on over 500,000 participants linked with Electronic Health Record (EHR) data containing rich phenotypic data, enables large-scale PheWAS. Moreover, MVP has the

highest racial and ethnic diversity of the major biobanks affording an opportunity to compare whether associations are similar across genetic ancestries.[6]

The objective of this study was to use existing clinical EHR data to identify conditions that share genetic variants with COVID-19 severity using the disease-agnostic PheWAS approach. Since COVID-19 is a new condition, identifying existing conditions which share genetic susceptibility may allow us to leverage existing knowledge from these known conditions to provide context regarding important pathways for COVID-19 severity.

Methods

Data sources

The VA MVP is a national cohort launched in 2011 designed to study the contributions of genetics, lifestyle, and military exposures to health and disease among US Veterans.[6] Blood biospecimens were collected for DNA isolation and genotyping, and the biorepository was linked with the VA EHR, which includes diagnosis codes (International Classification of Diseases ninth revision [ICD-9] and tenth revision [ICD-10]) for all Veterans followed in the healthcare system up to September 2019. The single nucleotide polymorphism (SNP) data in the MVP cohort was generated using a custom Thermo Fisher Axiom genotyping platform called MVP 1.0. The quality control steps and genotyping imputation using 1000 Genomes cosmopolitan reference panel on the MVP cohort has been described previously.[7]

UKBB is a population-based cohort of approximately 500,000 subjects from the UK between 49 and 69 years of age and served as the replication cohort. Participants were enrolled between 2006 and 2010 and consented for the prospective collection of health records and lifestyle,

imaging, and genetic data. The phenotype data included ICD-10 diagnosis codes up to March 2019.

Genetic variant selection

An overview of the analytic workflow is outlined in Fig 1. Variants were derived from the COVID-19 HGI GWAS meta-analysis release v4[3]. In this study, we analyzed the following HGI GWAS summary statistics: 1) hospitalized and critically ill COVID-19 vs. population controls denoted as "A2" in HGI, and referred to as "critical COVID" in this study, and 2) hospitalized because of COVID-19 vs. population controls, denoted as "B2" in HGI, referred to as "hospitalized COVID" in this study[3]. For each GWAS, variants with a Benjamini-Hochberg false discover rate (FDR) corrected p-value < 0.1 were selected as candidate lead SNPs (853 associated with critical COVID, and 902 associated with hospitalized COVID). Variants with r^2 <0.1 were clustered within a 250 kb region according to 1000 Genomes phase 3 transethnic reference panel[8], resulting in 48 independent variants associated with critical COVID and 39 variants associated with hospitalized COVID summary statistics. The lead variants from each set of GWAS summary statistics are available in eTable 1.

Outcomes

For both MVP and UKBB, clinical data prior to the onset of the COVID-19 pandemic were used to reduce potential confounding bias from SARS-CoV-2 infection on existing conditions. Briefly, phenotypes were defined by phecodes, used in prior PheWAS studies[5,9]. Each phecode represents ICD codes grouped into clinically relevant phenotypes for clinical studies. For example, the phecode "deep venous thrombosis" includes "venous embolism of deep vessels of the distal lower extremities," and "deep venous thrombosis of the proximal lower extremity,"

both of which have distinct ICD codes. Using this approach, all ICD codes for all Veterans in MVP were extracted and each assigned a phenotype defined by a phecode. ICD-9 and ICD-10 codes were mapped to 1876 phecodes, as previously described.[5,9]

For each phecode, participants with ≥2 phecode-mapped ICD-9 or ICD-10 codes were defined as cases, whereas those with no instance of a phecode-mapped ICD-9 or ICD-10 code were defined as controls. Based on our previous simulation studies of ICD codes from EHRs, we excluded phecodes with < 200 cases from the analysis[10] and we applied this threshold in each ancestry group. Thus, 1,520 (EUR), 1206 (AFR), 858 (HIS), 176 (ASN) phecodes in the MVP cohort were analyzed. The same process was applied to UKBB, resulting in 1,0,64 phecodes from 1,615 phecodes.

Phenome-wide association studies

The primary PheWAS analysis used SNPs identified from the HGI GWAS of critical and hospitalized COVID, and tested association of these SNPs with phenotypes extracted from the EHR data prior to the COVID-19 pandemic. Logistic regression using PLINK2 to examine SNPs' association with phecodes and firth regression was applied when logistic regression model failed to converge. Regression models were adjusted for sex, age (at enrollment), age squared, and the first 20 principal components. Genetic ancestry was determined using the HARE method for four major groups: African (AFR), Asian (ASN), Hispanic (HIS), and European (EUR) ancestry[11]. Ancestry-specific PheWAS was first performed in these four groups, and then summary data were meta-analyzed using an inverse-variance weighted fixed-effects model implemented in the PheWAS R package[9]. We assessed heterogeneity using I^2 and excluded any results with excess heterogeneity ($I^2 > 40\%$).

To address multiple testing, FDR $p < 0.1$ was considered significant. In the PheWAS analysis, $p < 2.44 \times 10^{-4}$ (critical COVID lead variants) and $p < 3.18 \times 10^{-4}$ (hospitalized COVID lead variants) were considered significant. In the main manuscript we highlight PheWAS significant associations using FDR < 0.1 and an effect size associated with increased or reduced risk for a condition by 10%, with complete PheWAS results in eTable 2 and eTable 3. Replication of MVP PheWAS results were performed in UKBB among individuals of European descent due to small sample size of other ancestries in UKBB.

Results

We studied 453,699 MVP participants, with mean age 68 years (SD), 89% male, with 32% participants from non-European ancestry (Table 1). The PheWAS was performed on 48 genetic variants associated with critical COVID-19, and 39 genetic variants associated with hospitalized COVID, across 1,559 phenotypes.

From the trans-ethnic meta-analysis, we identified 109 phenotypes significantly associated with critical COVID GWAS-identified variants, and 118 associations with hospitalized COVID GWAS-identified lead variants (FDR, $p<0.1$). Five variants were common to both severe and hospitalized COVID and had significant PheWAS associations, namely, variations nearest to the genes *ABO* (rs550057 and rs505922), *DPP9* (rs2277732), *HLA-DPB1* (rs143334143), and *CCHCR1* (rs9501257) (eTable 2 and eTable 3).

<u>Association of ABO loci with known risk factors and outcomes related to COVID-19 severity</u>

In the transethnic meta-analysis, the phenotypes with the strongest association with variants near *ABO* locus (rs550057 and rs505922) was “venous embolism and thrombosis” ($OR_{critical_PheWAS}$ =

1.27 [1.24 - 1.29], $P_{critical_PheWAS} = 5.28 \times 10^{-116}$; $OR_{hospitalized_PheWAS} = 1.31$ [1.29 - 1.34], $P_{hospitalized_PheWAS} = 3.53 \times 10^{-183}$, Fig 2). The ABO loci had the largest number of significant PheWAS association findings, accounting for 48% (53/109) of significant phenotype associations in the critical COVID PheWAS, and 51% (61/118) in the hospitalized COVID PheWAS. As expected, conditions associated with the *ABO* locus, such as venous embolism and thrombosis, hypercoagulable state, type 2 diabetes, coronary atherosclerosis, have been reported as risk factors for or are complications associated with COVID-19 severity and mortality (Fig 2, eTable 2 and eTable 3).

<u>Associations between variants associated with COVID-19 severity and respiratory conditions and infections</u>

Among respiratory conditions, only two diseases had significant associations (FDR < 0.1) shared with genetic variants associated with severe COVID-19. The most significant association was observed between rs2277732 (*DPP9*) and idiopathic fibrosing alveolitis (OR = 1.17 [1.08 – 1.26]; $P = 5.34 \times 10^{-5}$), also known as idiopathic pulmonary fibrosis (IPF). The association between *DPP9* variants and IPF has been reported in previous studies.[12] However, the association of genetic variants with other respiratory conditions may represent novel findings: the association of intronic variant rs143334143 in *CCHCR1* with asthma (OR = 0.94 [0.92 - 0.97], $P = 2.31 \times 10^{-4}$). We did not detect associations between any of the variants and either other respiratory conditions or infectious diseases. (eTable 3).

<u>Associations between variants associated with COVID-19 severity and reduced risk for certain phenotypes</u>

A subset of variants associated with severe COVID were also associated with a reduced risk for certain conditions, many of which were autoimmune conditions (Table 2). The rs74956615-A allele of *RAVER1*, a lead variant from the hospitalized COVID GWAS was associated with a reduced risk for psoriasis (OR = 0.71 [0.66- 0.76], $P = 1.5 \times 10^{-22}$), rheumatoid arthritis (RA) (OR = 0.78 [0.72 - 0.84], $P = 4.6 \times 10^{-09}$), and psoriatic arthropathy (OR = 0.60 [0.50 - 0.71], $P = 3.4 \times 10^{-09}$). Similarly, rs11085727-T at the *TYK2* locus identified as a lead variant in the HGI critical COVID GWAS was associated with a reduced risk of psoriasis (OR = 0.88 [0.85 - 0.90], $P = 1.35 \times 10^{-17}$) and psoriatic arthropathy (OR = 0.81 [0.75 -0.86], $P = 9.54 \times 10^{-10}$).

The association of the *RAVER1* variant with psoriasis and RA replicated in UKBB. Among the 353,365 EUR UKBB participants, the *RAVER1* variant was associated with a decreased risk for psoriasis and RA ($OR_{psoriasis} = 0.69$, $P = 2.88 \times 10^{-05}$; $OR_{RA} = 0.67$, $P = 9.15 \times 10^{-07}$). RAVER1 was mentioned in prior RA GWAS' with the signal attributed to *TYK2*.[13] The *TYK2* signal has been mapped to a protein-coding variant (rs34536443) with predicted partial loss of function (LOF) leading to a reduced risk of psoriasis, psoriatic arthropathy, and RA as well as other autoimmune inflammatory conditions[14,15]. The RAVER1 variant (rs74956615) in this study is in high linkage disequilibrium (LD) (r2 = 0.7) with this *TYK2* protein-coding variant (S1 Fig), with the same associations of reduced risk of psoriasis, psoriatic arthropathy, and RA. This suggests that the *RAVER1* signal from our analysis is coupled with the protein-coding variant in *TYK2* (rs34536443). The *TYK2* variant (rs11085727) associated with critical COVID was not in LD (r2 =0.07) with the previously identified protein-coding variant in *TYK2* (rs34536443). To further support these findings, we conducted PheWAS of rs34536443 in MVP and observed similar association of reduced risk with aforementioned autoimmune inflammatory conditions (S2 Fig).

The PheWAS analyses performed across genetic ancestries observed similar findings as the overall meta-analysis with fewer associations within each ancestry group due to smaller population sizes (S3 Fig). Of note, SNP rs1396186 associated with hospitalized COVID didn't had associations with obesity in meta-analysis (P=0.06; I^2 = 84%). However, among Veterans of African ancestry, this association was significant (OR = 1.13 [1.07 – 1.19]; $P = 8.79 \times 10^{-06}$) but was not observed in the larger population of European descent (S3 Fig, panel B, S3 Table). Another SNP, that passed rs1754680 was significantly associated with obesity among EUR ancestry participants but not in other ancestry populations (S3 Table).

Discussion

In this large-scale PheWAS, we identified shared genetic architecture between variants associated with severe COVID-19 and other complex conditions using data from MVP, one of the largest and most diverse biobanks in the world, with replication in UKBB. Broadly, these risk alleles identified conditions associated with risk factors for severe COVID-19 manifestations such as T2D, ischemic heart disease across all ancestries, with a potentially stronger association with obesity among Veterans of African ancestry. Phenotypes associated with poor outcomes related to COVID-19 infection, including deep venous thrombosis and other thrombotic complications, also shared variants associated with severe COVID-19. In contrast, among respiratory conditions, only idiopathic pulmonary fibrosis and asthma shared genetic risk factors, with the notable absence of an association with COPD and other respiratory infections. Finally, we observed that variants associated with severe COVID-19 were associated with a reduced risk of autoimmune inflammatory conditions, such as psoriasis, psoriatic arthritis, and RA.

A classic GWAS tests the association between millions of genetic variants with the presence or absence of one phenotype, e.g., GWAS of deep venous thrombosis. In the COVID-19 HGI GWAS, the "phenotype" was patients hospitalized for or critically ill from COVID-19. Clinically, this population includes a mixture of patients with a complex list of medical conditions at high risk for severe COVID complications and those who had actual complications from COVID-19. Thus, we would anticipate that many of the significant phenotypes would be associated with risk factors such as obesity and deep venous thrombosis. Notably, the clinical data used in this study pre-dates the emergence of COVID-19 to reduce potential confounding bias that can occur in a population infected with SARS-CoV-2, e.g. interaction between COVID-19 and type 2 diabetes. Additionally, our findings suggest that the PheWAS approach can be a useful tool to identify clinical factors related to emerging infectious diseases regarding severity or complications when genomic data are available.

The PheWAS results of SNPs in the *ABO* locus served as a positive control for this study. Genetic variations in *ABO* are an established risk factor for COVID-19 severity. Patients with blood group A have a higher risk of requiring mechanical ventilation and extended ICU stay compared with patients with blood group O.[16] These same variations at *ABO* had known associations with a spectrum of blood coagulation disorders identified in studies pre-dating COVID-19.[17–19] The PheWAS of ABO variants identified associations with increased risk of deep vein thrombosis, pulmonary embolism, and other circulatory disorders, in line with prior studies, and recent studies among patients hospitalized with COVID-19.[20–24]

Autoimmune conditions shared genetic variants with severe COVID-19, the direction of effect was for a reduced odd of these conditions. Based on existing data, we believe known protein-

coding variants in *TYK2* (rs34536443, 1140A) drove the majority of the signal from *RAVER1* with reduced risk of psoriasis, psoriatic arthritis, and RA; this signal is consistent with prior reported associations of a protective effect of *TYK2* with these same conditions.[13,14,15] Moreover, recent studies exploring potentially important pathways for COVID-19 progression have highlighted TYK2 and its downstream signaling pathway through type 1 interferons as a potential target for treatment.[25]

The existing literature can help explain the dual association between reduced risk of autoimmune conditions such as psoriasis and RA and increased risk of severe COVID via *TYK2*. *TYK2*, a member of the Janus Kinase (JAK) family of genes, plays a key role in cytokine signal transduction and the inflammatory response, particularly via IL-12, IL-23, and is also important for IL-6 and IL-10 signaling (Fig 3).[26] Importantly, TYK2 has a central role in type 1 interferon signaling, part of the innate immune response blocking the spread of a virus from infected to uninfected cells. Of interest is the *TYK2* variant represented by the SNP rs34536443 (P1104A) resulting in partial LOF and reduced risk for several autoimmune disorders such as RA, type 1 diabetes, psoriasis, and psoriatic arthritis without evidence of significant immunodeficiency.[14,27–30] However, humans with complete *TYK2* LOF have clinically significant immunodeficiency with increased susceptibility to mycobacterial and viral infections.[26,31]

The balance between immune tolerance and ability to mount a robust early inflammatory response, and viral host defense may in part explain the mixed results of clinical trials using immunosuppressive therapy in COVID-19 to date[25]. Corticosteroid therapy may be beneficial in later stages of COVID-19 infection but detrimental if used early in the disease presumably when an early host defense response is important.[32] Blocking the IL-6 pathway with tocilizumab did

not reduce mortality among hospitalized COVID-19 patients, but may be beneficial among later-stage pulmonary damage[33,34]. The evidence to date suggests that the ability to mount an early inflammatory response is an important aspect in the control of COVID-19 infection and that dampening inflammation early in the disease course may be detrimental.

Limitations

We note several limitations. First, the PheWAS was designed as a broad screen to test for potentially clinically relevant associations between genes and phenotypes, with limited power to detect associations among uncommon conditions, and when further stratified by genetic ancestry. MVP is not a general US population-based cohort; however, results were replicated in UKBB, an independent general population cohort. Findings from this study suggest that variants associated with severe COVID-19 are also associated with reduced risk of having an autoimmune inflammatory condition. However, the results cannot provide information on the impact of actual SARS-CoV-2 infection in these individuals after diagnosis of an autoimmune disease. Finally, the data strongly suggest but cannot confirm that the signal observed from *RAVER1* arises from the previously identified and studied *TYK2* variant.

Conclusions

This PheWAS of genetic variants reported to associate with severe COVID-19 demonstrated shared genetic architecture between COVID-19 severity and known underlying risk factors for both severe COVID-19 and poor COVID-19 outcomes, rather than susceptibility to other viral infections. Overall, the associations observed were generally consistent across genetic ancestries, with a potential stronger association with obesity among Veterans of African ancestry compared to European ancestry. The divergent association between severe COVID-19 and autoimmune

inflammatory conditions sheds light on the concept of a balance between immune tolerance and immunodeficiency. This balance will be important when considering therapeutic targets for COVID-19 therapies where pathways may control both inflammation and the viral host response.

Funding

This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by award MVP035. This publication does not represent the views of the Department of Veteran Affairs or the United States Government. R.M.C. is supported by NIH grants R01 AA026302 and P30 DK0503060. K.P.L. is supported by NIH P30 AR072577, and the Harold and Duval Bowen Fund.

Conflict of Interest

RMC has received research support from Intercept Pharmaceuticals, Inc and Merck & Co. MDR is on the scientific advisory board for Goldfinch Bio and Cipherome. CJO is an employee of Novartis Institute for Biomedical Research. PN reports grant support from Amgen, Apple, AstraZeneca, Boston Scientific, and Novartis, personal fees from Apple, AstraZeneca, Blackstone Life Sciences, Genentech, and Novartis, and spousal employment at Vertex, all unrelated to the present work.

Acknowledgements

We are grateful to our Veterans for their contributions to MVP. Full acknowledgements for the VA Million Veteran Program COVID-19 Science Initiative can be found in the supplementary methods. We would like to thank the Host Genetic Initiative for making their data publicly available (https://www.covid19hg.org/acknowledgements/). We would also like to acknowledge Dr. Madhuvika Murgan for her insightful suggestion on the Fig 3. We used genotype and phenotype data from the UK Biobank under project 32133.

References

1. CDC. About COVID-19 - CDC. : https://www.cdc.gov/coronavirus/2019-ncov/cdcresponse/about-COVID-19.html

2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020;20(5):533-534. doi:10.1016/S1473-3099(20)30120-1

3. The COVID-19 Host Genetics Initiative, ganna andrea. *Mapping the Human Genetic Architecture of COVID-19 by Worldwide Meta-Analysis*. Genetic and Genomic Medicine; 2021. doi:10.1101/2021.03.10.21252820

4. The GenOMICC Investigators, The ISARIC4C Investigators, The COVID-19 Human Genetics Initiative, et al. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2021;591(7848):92-98. doi:10.1038/s41586-020-03065-y

5. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010;26(9):1205-1210. doi:10.1093/bioinformatics/btq126

6. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214-223. doi:10.1016/j.jclinepi.2015.09.016

7. Hunter-Zinck H, Shi Y, Li M, et al. Genotyping Array Design and Data Quality Control in the Million Veteran Program. *Am J Hum Genet*. 2020;106(4):535-548. doi:10.1016/j.ajhg.2020.03.004

8. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

9. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics*. 2014;30(16):2375-2376. doi:10.1093/bioinformatics/btu197

10. Verma A, Bradford Y, Dudek S, et al. A simulation study investigating power estimates in phenome-wide association studies. *BMC Bioinformatics*. 2018;19(1):120. doi:10.1186/s12859-018-2135-0

11. Fang H, Hui Q, Lynch J, et al. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. *Am J Hum Genet*. 2019;105(4):763-772. doi:10.1016/j.ajhg.2019.08.012

12. Allen RJ, Guillen-Guio B, Oldham JM, et al. Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2020;201(5):564-574. doi:10.1164/rccm.201905-1017OC

13. Eyre S, Bowes J, Diogo D, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet*. 2012;44(12):1336-1340. doi:10.1038/ng.2462

14. Diogo D, Bastarache L, Liao KP, et al. TYK2 Protein-Coding Variants Protect against Rheumatoid Arthritis and Autoimmunity, with No Evidence of Major Pleiotropic Effects on Non-Autoimmune Complex Traits. Chiorini JA, ed. *PLOS ONE*. 2015;10(4):e0122271. doi:10.1371/journal.pone.0122271

15. Dendrou CA, Cortes A, Shipman L, et al. Resolving TYK2 locus genotype-to-phenotype differences in autoimmunity. *Sci Transl Med*. 2016;8(363):363ra149-363ra149. doi:10.1126/scitranslmed.aag1974

16. Hoiland RL, Fergusson NA, Mitra AR, et al. The association of ABO blood group with indices of disease severity and multiorgan dysfunction in COVID-19. *Blood Adv*. 2020;4(20):4981-4989. doi:10.1182/bloodadvances.2020002623

17. Zietz M, Zucker J, Tatonetti NP. Associations between blood type and COVID-19 infection, intubation, and death. *Nat Commun*. 2020;11(1):5761. doi:10.1038/s41467-020-19623-x

18. Paranjpe I, Fuster V, Lala A, et al. Association of Treatment Dose Anticoagulation With In-Hospital Survival Among Hospitalized Patients With COVID-19. *J Am Coll Cardiol*. 2020;76(1):122-124. doi:10.1016/j.jacc.2020.05.001

19. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*. Published online February 24, 2020. doi:10.1001/jama.2020.2648

20. Matsunaga H, Ito K, Akiyama M, et al. Transethnic Meta-Analysis of Genome-Wide Association Studies Identifies Three New Loci and Characterizes Population-Specific Differences for Coronary Artery Disease. *Circ Genomic Precis Med*. 2020;13(3):e002670. doi:10.1161/CIRCGEN.119.002670

21. Plagnol V, Howson JMM, Smyth DJ, et al. Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS Genet*. 2011;7(8):e1002216. doi:10.1371/journal.pgen.1002216

22. Reilly MP, Li M, He J, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet Lond Engl*. 2011;377(9763):383-392. doi:10.1016/S0140-6736(10)61996-4

23. Trégouët D-A, 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(21):5298-5303. doi:10.1182/blood-2008-11-190389

24. Larson NB, Bell EJ, Decker PA, et al. ABO blood group associations with markers of endothelial dysfunction in the Multi-Ethnic Study of Atherosclerosis. *Atherosclerosis*. 2016;251:422-429. doi:10.1016/j.atherosclerosis.2016.05.049

25. Gaziano L, Giambartolomei C, Pereira AC, et al. *Actionable Druggable Genome-Wide Mendelian Randomization Identifies Repurposing Opportunities for COVID-19*. Epidemiology; 2020. doi:10.1101/2020.11.19.20234120

26. Nemoto M, Hattori H, Maeda N, et al. Compound heterozygous TYK2 mutations underlie primary immunodeficiency with T-cell lymphopenia. *Sci Rep*. 2018;8(1):6956. doi:10.1038/s41598-018-25260-8

27. Hellquist A, Järvinen TM, Koskenmies S, et al. Evidence for Genetic Association and Interaction Between the *TYK2* and *IRF5* Genes in Systemic Lupus Erythematosus. *J Rheumatol*. 2009;36(8):1631-1638. doi:10.3899/jrheum.081160

28. Sigurdsson S, Nordmark G, Göring HHH, et al. Polymorphisms in the Tyrosine Kinase 2 and Interferon Regulatory Factor 5 Genes Are Associated with Systemic Lupus Erythematosus. *Am J Hum Genet*. 2005;76(3):528-537. doi:10.1086/428480

29. , The Wellcome Trust Case–Control Consortium (WTCCC) and Alastair Compston, Ban M, Goris A, et al. Replication analysis identifies TYK2 as a multiple sclerosis susceptibility factor. *Eur J Hum Genet*. 2009;17(10):1309-1313. doi:10.1038/ejhg.2009.41

30. Cunninghame Graham DS, Morris DL, Bhangale TR, et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with Systemic Lupus Erythematosus. McCarthy MI, ed. *PLoS Genet*. 2011;7(10):e1002341. doi:10.1371/journal.pgen.1002341

31. Watford WT, O'Shea JJ. Human Tyk2 Kinase Deficiency: Another Primary Immunodeficiency Syndrome. *Immunity*. 2006;25(5):695-697. doi:10.1016/j.immuni.2006.10.007

32. The RECOVERY Collaborative Group. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med*. 2021;384(8):693-704. doi:10.1056/NEJMoa2021436

33. Stone JH, Frigault MJ, Serling-Boyd NJ, et al. Efficacy of Tocilizumab in Patients Hospitalized with Covid-19. *N Engl J Med*. 2020;383(24):2333-2344. doi:10.1056/NEJMoa2028836

34. Salama C, Han J, Yau L, et al. Tocilizumab in Patients Hospitalized with Covid-19 Pneumonia. *N Engl J Med*. 2021;384(1):20-30. doi:10.1056/NEJMoa2030340

Tables and Figures

Table 1. Patient characteristics of Million Veteran Program participants

Characteristics	**Million Veteran Program**
	Number (%)
Total Patients	455,683
Male	403,923 (89)
Genetic Ancestry	
European	312,282 (68)
African	84,523 (18.5)
Hispanic	33,348 (7.3)
Asian	4,692 (1.02)
Other	20,568 (4.5)
Risk Factors	
Obesity (phecode = 278)	195,989 (43)
Hypertension (phecode = 401.1)	318,131 (69.8)
Type 2 Diabetes (phecode = 250.2)	161,544 (35.4)
Coronary Artery Disease (phecode = 411.4)	110,564 (24.2)
Chronic Kidney Disease (phecode = 585.2)	64,002 (14.04)

Table 2. Phenotypes sharing association with variants also associated with severe COVID-19 infection, with reduced odds of disease listed in order of p-value*.

Phenotype	OR (95% CI)	p-value	Gene (SNP)	SNP (MAF)	COVID-severity
Psoriasis	0.72 [0.67 - 0.77]	1.5×10^{-22}	*RAVER1*	rs74956615-**A** (0.02)	Both
	0.88 [0.86 – 0.91]	1.35×10^{-17}	*TYK2*	rs11085727-**T** (0.3)	Critical
Rosacea	0.84 [0.81 - 0.89]	2.5×10^{-12}	*HLA-DPB1*	rs9501257-**G** (0.2)	Critical
Seborrheic dermatitis	0.86 [0.83 - 0.9]	1.8×10^{-10}	*RAVER1*	rs74956615-**A** (0.02)	Hospitalized
Psoriatic arthropathy	0.81 [0.76 - 0.87]	3.5×10^{-09}	*TYK2*	rs11085727-**T** (0.3)	Critical
	0.60 [0.51 – 0.71]	9.5×10^{-10}	*RAVER1*	rs74956615-**A** (0.02)	Hospitalized
Rheumatoid arthritis	0.79 [0.73 - 0.85]	1.0×10^{-09}	*RAVER1*	rs74956615-**A** (0.02)	Hospitalized
Vitiligo	0.69 [0.6 - 0.81]	2.4×10^{-06}	*CCHCR1*	rs143334143-**A** (0.09)	Critical
Sarcoidosis	0.73 [0.64 - 0.83]	4.3×10^{-06}	*CCHCR1*	rs143334143-**A** (0.09)	Hospitalized
	0.63 [0.51 – 0.78]	3.1×10^{-05}	*RAVER1*	rs74956615-**A** (0.02)	Hospitalized
Other disorders of liver	0.64 [0.53 - 0.78]	8.2×10^{-06}	*AC113414.1*	rs79833209-**T** (0.01)	Hospitalized
Hemorrhage from GI ulcer	0.85 [0.79 - 0.92]	4.0×10^{-05}	*ABO*	rs550057-**T** (0.2)	Critical
Other specified disorders of plasma protein metabolism	0.80 [0.72 - 0.89]	5.61×10^{-05}	*PTCHD4*	rs4715051-**G** (0.5)	Critical
Specific nonpsychotic mental disorders attributable to brain damage	0.82 [0.75 - 0.9]	6.0×10^{-05}	*PDE8A*	rs200678047-**A** (0.1)	Critical

*OR<0.9 and P<10-5 shown in table, full results in supplementary; if multiple related conditions, e.g. psoriasis, psoriasis vulgaris, psoriasis and related disorders, description with lowest p-value selected shown in table.

Fig 1. Overview of variant selection and PheWAS analysis design.

Fig 2. PheWAS results of candidate SNPs from GWAS of critically ill and hospitalized COVID-19. Significant associations between 48 SNPs from critical ill COVID GWAS (A) and 39 SNPs from hospitalized COVID (C) and EHR derived phenotypes in the Million Veteran Program. The phenotypes are represented on the x-axis and ordered by broader disease categories. The red line

denotes the significance threshold using false discovery rate of 10% using the Benjamini-Hochberg procedure. The description of phenotypes is highlighted for the associations with FDR < 0.1 and odds ratio < 0.90 or odds ratio > 1.10. B) and D) A heatmap plot of SNPs with at least one significant association (FDR < 0.1). The direction of effect disease risk is represented by odds ratio. A red color indicates increased risk and blue color indicated reduced risk. The results with odds ratio < 0.90 or odds ratio > 1.10 are shown.

Fig 3. Conceptual model of the relationship between *TYK2* (rs34536133) with reduced risk for psoriasis, psoriatic arthritis, rheumatoid arthritis, and simultaneous increased risk of COVID-19 severity.

Supplementary Figures and Tables

S1 Fig. LDproxy plot showing *RAVER1* SNP rs74956615 has highest LD with missense variant rs34536443 (*TYK2*).

S2 Fig. The PheWAS results of rs34536133 in MVP.

S3 Fig. The PheWAS results of 48 SNPs from critical ill COVID GWAS by each ancestry a) European ancestry, b) African ancestry, c) Hispanic ancestry, and d) Asian ancestry.

S4 Fig. The PheWAS results of 39 SNPs from hospitalized COVID GWAS by each ancestry a) European ancestry, b) African ancestry, c) Hispanic ancestry, and d) Asian ancestry.

S1 Table. List of lead variants from critical ill and hospitalized COVID GWAS included in the study.

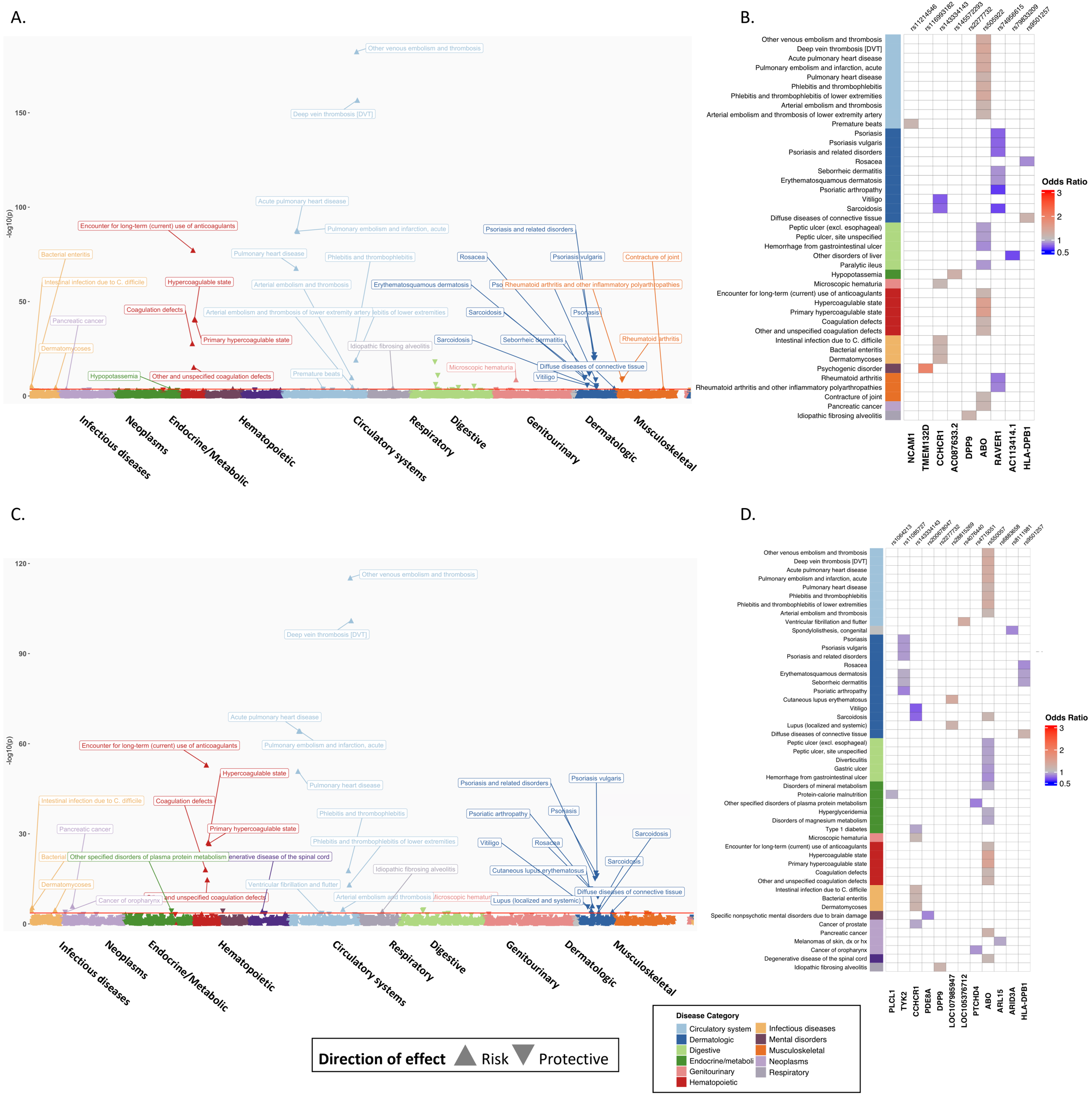
S2 Table. PheWAS results summary statistics by each ancestry and meta-analysis across four ancestry for 48 lead SNPs identified from critical ill COVID GWAS.

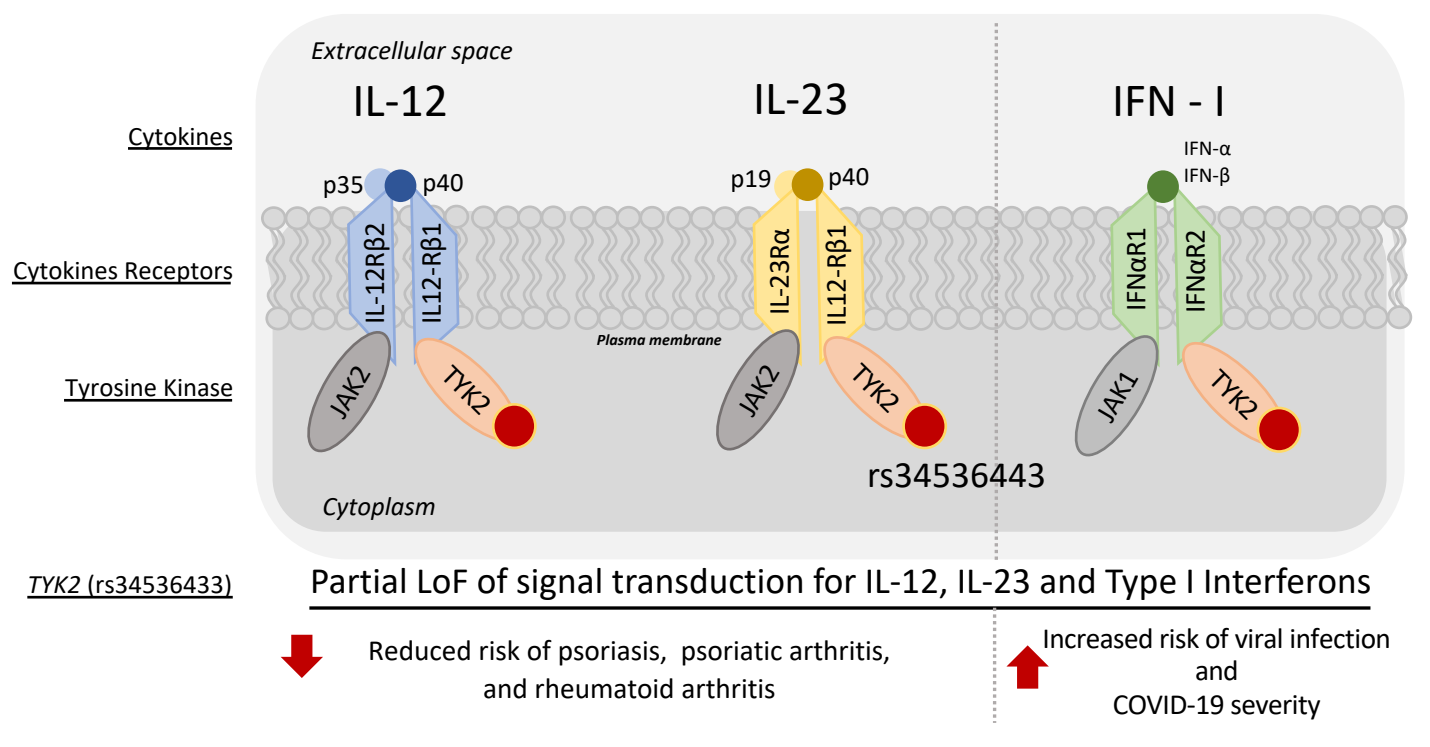
S3 Table. PheWAS results summary statistics by each ancestry and meta-analysis across four ancestry for 39 lead SNPs identified from hospitalized COVID GWAS.

COVID-19 GWAS Summary Statistics	Select COVID-19 Lead Variants	MVP Dataset	PheWAS
❖ Critically ill vs Population (A1*) ❖ Hospitalized vs Population (B2*)	❖ SNPs clumped and pruned using 1000G reference panel ❖ FDR = 0.1 ❖ $R^2 < 0.1$ ❖ Distance threshold = 250kb	**GENOTYPE DATA** ❖ 1000G imputed data ❖ Ancestry defined by HARE algorithm ❖ **n=455,683**	❖ PheWAS by ancestry ❖ Logistic or firth regression for association between SNPs with phecodes ❖ Covariates: 1st 20 PCs, age, age^2, sex
	LEAD LOCI ❖ **n=48** from Critically ill vs Population GWAS (A1) ❖ **n=39** from Hospitalized vs Population GWAS (B2)	**PHENOTYPE DATA** ❖ ICD9/10 codes mapped to phecodes ❖ Case-control definition ❖ Case: ≥2 phecodes ❖ Control: 0 phecodes ❖ n<200 phecodes excluded by ancestry ❖ Total phecodes#, **n=1520** (EUR), **1206** (AFR), **858** (HIS), **176** (ASN)	❖ Transethnic meta-analysis ❖ Replication in UK Biobank

* From Host Genetics Initiative;

EUR = European Ancestry, AFR = African Ancestry, HIS= Hispanic Ancestry, ASN = Asian Ancestry


A.
-log10(p)
Other venous embolism and thrombosis
Deep vein thrombosis [DVT]
Acute pulmonary heart disease
Pulmonary embolism and infarction, acute
Pulmonary heart disease
Phlebitis and thrombophlebitis
Arterial embolism and thrombosis
Arterial embolism and thrombosis of lower extremity artery
Phlebitis of lower extremities
Idiopathic fibrosing alveolitis
Premature beats
Microscopic hematuria
Encounter for long-term (current) use of anticoagulants
Hypercoagulable state
Primary hypercoagulable state
Coagulation defects
Other and unspecified coagulation defects
Bacterial enteritis
Intestinal infection due to C. difficile
Pancreatic cancer
Dermatomycoses
Hypopotassemia
Psoriasis and related disorders
Rosacea
Psoriasis vulgaris
Contracture of joint
Erythematosquamous dermatosis
Rheumatoid arthritis and other inflammatory polyarthropathies
Sarcoidosis
Psoriasis
Seborrheic dermatitis
Rheumatoid arthritis
Diffuse diseases of connective tissue
Vitiligo
Infectious diseases
Neoplasms
Endocrine/Metabolic
Hematopoietic
Circulatory systems
Respiratory
Digestive
Genitourinary
Dermatologic
Musculoskeletal
B.
rs11214546
rs116993182
rs143334143
rs145572293
rs2277732
rs505922
rs74956615
rs79833209
rs9501257
Odds Ratio
NCAM1
TMEM132D
CCHCR1
AC087633.2
DPP9
ABO
RAVER1
AC113414.1
HLA-DPB1
C.
-log10(p)
Other specified disorders of plasma protein metabolism
Cancer of oropharynx
Degenerative disease of the spinal cord
Ventricular fibrillation and flutter
Lupus (localized and systemic)
Cutaneous lupus erythematosus
Psoriatic arthropathy
D.
rs1064213
rs11085727
rs143334143
rs200678047
rs2277732
rs28815269
rs4076440
rs4715051
rs550057
rs6883658
rs8111981
rs9501257
PLCL1
TYK2
CCHCR1
PDE8A
DPP9
LOC107985947
LOC105376712
PTCHD4
ABO
ARL15
ARID3A
HLA-DPB1
Direction of effect
Risk
Protective
Disease Category
Circulatory system
Dermatologic
Digestive
Endocrine/metaboli
Genitourinary
Hematopoietic
Infectious diseases
Mental disorders
Musculoskeletal
Neoplasms
Respiratory


Extracellular space
IL-12
IL-23
IFN - I
Cytokines
p35
p40
p19
p40
IFN-α
IFN-β
Cytokines Receptors
IL-12Rβ2
IL12-Rβ1
IL-23Rα
IL12-Rβ1
IFNαR1
IFNαR2
Plasma membrane
Tyrosine Kinase
JAK2
TYK2
JAK2
TYK2
JAK1
TYK2
rs34536443
Cytoplasm
TYK2 (rs34536433)
Partial LoF of signal transduction for IL-12, IL-23 and Type I Interferons
Reduced risk of psoriasis, psoriatic arthritis, and rheumatoid arthritis
Increased risk of viral infection and COVID-19 severity