

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Brain Behavior and Immunity

journal homepage: [www.elsevier.com/locate/ybrbi](http://www.elsevier.com/locate/ybrbi)

## Cognitive disorders associated with hospitalization of COVID-19: Results from an observational cohort study

Jingqi Zhou<sup>a,b</sup>, Chang Liu<sup>a,c</sup>, Yitang Sun<sup>a</sup>, Weishan Huang<sup>d,e</sup>, Kaixiong Ye<sup>a,f,\*</sup><sup>a</sup> Department of Genetics, Franklin College of Arts and Sciences, University of Georgia, Athens, GA, USA<sup>b</sup> School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, PR China<sup>c</sup> College of Life Sciences, Wuhan University, Wuhan, PR China<sup>d</sup> Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Baton Rouge, LA, USA<sup>e</sup> Department of Microbiology and Immunology, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA<sup>f</sup> Institute of Bioinformatics, University of Georgia, Athens, GA, USA

## ARTICLE INFO

## Keywords:

COVID-19  
Dementia  
Alzheimer's disease  
Risk factors  
Pre-existing conditions  
Cognitive disorders

## ABSTRACT

**Introduction:** Our understanding of risk factors for COVID-19, including pre-existing medical conditions and genetic variations, is limited. To what extent the pre-existing clinical condition and genetic background have implications for COVID-19 still needs to be explored.

**Methods:** Our study included 389,620 participants of European descent from the UK Biobank, of whom 3,884 received the COVID-19 test and 1,091 were tested positive for COVID-19. We examined the association of COVID-19 status with an extensive list of 974 medical conditions and 30 blood biomarkers. Additionally, we tested the association of genetic variants in two key genes related to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, angiotensin-converting enzyme 2 (*ACE2*) and transmembrane protease serine 2 (*TMPRSS2*), with COVID-19 or any other phenotypes.

**Results:** The most significant risk factors for COVID-19 include Alzheimer's disease (OR = 2.29, 95% CI: 1.25–4.16), dementia (OR = 2.16, 95% CI: 1.36–3.42), and the overall category of delirium, dementia, amnestic and other cognitive disorders (OR = 1.90, 95% CI: 1.24–2.90). Evidence suggesting associations of genetic variants in SARS-CoV-2 infection-related genes with COVID-19 (rs7282236, OR = 1.33, 95% CI: 1.14–1.54,  $p = 2.31 \times 10^{-4}$ ) and other phenotypes, such as an immune deficiency ( $p = 5.65 \times 10^{-5}$ ) and prostate cancer ( $p = 1.1 \times 10^{-5}$ ), was obtained.

**Conclusions:** Our unbiased and extensive search identified pre-existing Alzheimer's disease and dementia as top risk factors for hospital admission due to COVID-19, highlighting the importance of providing special protective care for patients with cognitive disorders during this pandemic. We also obtained evidence suggesting a direct association of genetic variants with COVID-19.

## 1. Introduction

The pandemic of coronavirus disease 2019 (COVID-19), which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has led to at least 6 million confirmed cases of infection and 370,000 deaths by the end of May 2020 (World Health Organization, 2020). A growing number of studies have started to accumulate information on the common characteristics of patients with COVID-19 and risk factors

for viral infection and disease progression (Grasselli et al., 2020; Richardson et al., 2020; Docherty et al., 2020). Older age, male sex, African ethnicity, a lower socioeconomic status, and some pre-existing medical conditions (e.g., chronic kidney diseases and obesity) have been repeatedly associated with a positive COVID-19 test and/or adverse outcomes (Docherty et al., 2020; Niedzwiedz et al., 2020; Raisi-Estabragh et al., 2005; de Lusignan et al., 2020; Kumar et al., 2020; Petrilli et al., 2020). The vast majority of these existing studies relied on

**Abbreviations:** Apo(a), apolipoprotein A; COVID-19, coronavirus disease 2019; GWAS, genome-wide association study; HDL, high-density lipoprotein cholesterol; ICD, International Classification of Diseases; LDL, low-density lipoprotein cholesterol; OR, odds ratio; PheWAS, phenome-wide association study; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SNP, single nucleotide polymorphism; SD, standard deviation; TC, total cholesterol; UKB, UK Biobank.

\* Corresponding author at: Department of Genetics, Franklin College of Arts and Sciences, University of Georgia, Athens, GA 30602, USA.

E-mail address: [Kaixiong.Ye@uga.edu](mailto:Kaixiong.Ye@uga.edu) (K. Ye).

<https://doi.org/10.1016/j.bbi.2020.10.019>

Received 23 July 2020; Received in revised form 30 September 2020; Accepted 19 October 2020

Available online 24 October 2020

0889-1591/© 2020 The Author(s).

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

recently assembled cohorts of patients with COVID-19, and due to the time constraints, the sample size and the number of risk factors evaluated are both limited (de Lusignan et al., 2020; Kumar et al., 2020; Petrilli et al., 2020; Yang et al., 2020; Guo et al., 2020; Bianchetti et al., 2020). Moreover, candidate factors were usually chosen based on the ease of collection, clinical experience, and prior publications, with some medical conditions (e.g., diabetes, respiratory and cardiovascular conditions) receiving much more attention than others (e.g., cognitive disorders) (de Lusignan et al., 2020; Kumar et al., 2020; Petrilli et al., 2020; Yang et al., 2020; Guo et al., 2020; Bianchetti et al., 2020). More risk factors likely remain to be tested and identified. The addition of the COVID-19 status to previously existing population cohorts with comprehensive electronic health records, such as the UK Biobank (UKB) (Bycroft et al., 2018), provide a valuable opportunity to perform an unbiased and exhaustive search across all available phenotypes (i.e., the phenome) to identify novel risk factors.

Another likely source of individual differences in responses to the SARS-CoV-2 is genetic variations. These variations may directly affect virus entry and replication, and/or indirectly predispose individuals to medical conditions that exacerbate COVID-19 progression. For instance, the *ApoE* e4 genotype, a known genetic risk factor for both dementia and Alzheimer's disease, was recently shown to be associated with an increased risk of a severe COVID-19 infection (Kuo et al., 2020). While efforts are still ongoing to assemble COVID-19 cohorts and to sequence patient genomes in order to map host genetic determinants of susceptibility and severity (Murray et al., 2020), we are able to leverage deep genomic and phenotyping data in existing biobanks to evaluate the clinical effects of genetic variants in human genes known to be indispensable for SARS-CoV-2 infection. Angiotensin converting enzyme 2, encoded by *ACE2*, is the cell surface receptor for the viral spike (S) protein (Zhou et al., 2020), while transmembrane protease serine 2, encoded by *TMPRSS2*, is essential for priming S protein-mediated membrane fusion (Hoffmann et al., 2020). Genetic variants located around *TMPRSS2* have been previously associated with prostate cancer (Al Olama et al., 2014), heart failure and coronary heart disease (He et al., 2016). However, no associations have been reported for *ACE2* in the GWAS Catalog (Buniello et al., 2019), likely because *ACE2* is located on the X chromosome, a part of the genome that is commonly neglected in genome-wide analysis (Chang et al., 2014). Similarly, in the existing UKB GWAS database, GeneAtlas, no information was available for the sex chromosome and thus *ACE2* (Canela-Xandri et al., 2018). Very recent efforts to investigate the clinical effects of *ACE2* and *TMPRSS2* did not identify associations that reach genome-wide statistical significance (Curtis, 2005; Cirulli et al., 2004; Lopera et al., 2004). Extensive searches including more phenotypes in cohorts with much larger sample sizes will likely reveal novel findings.

UKB is a large population-based prospective study established to investigate genetic and environmental determinants of human diseases. More than 500,000 middle-aged participants were recruited between 2006 and 2010, for whom deep genomic and phenotyping data were collected, including genome-wide genotypes, physical measurements, sociodemographic factors, lifestyle indicators, biomarkers in blood and urine, and linkage to medical records (Bycroft et al., 2018). Recently, Public Health England provided COVID-19 test results for UKB participants. Since testing was initially prioritized to patients in the hospital or with a severe respiratory illness, test positivity may indicate severe disease in the UK (Iacobucci, 2020). In this data-driven phenome-wide association study (PheWAS), we leverage the extensive UKB resource to identify (i) pre-existing medical conditions that are overrepresented in patients with COVID-19, and (ii) the clinical effects of genetic variants in *ACE2* and *TMPRSS2*, including their direct associations with COVID-19.

## 2. Methods

### 2.1. UK Biobank cohort

UK Biobank (UKB) is a large population-based prospective study established for investigations of genetic and nongenetic determinants of diseases in middle- and old-aged adults. More than 500,000 individuals aged 40–69 years were recruited between 2006 and 2010, all of whom underwent baseline measurements, donated biological materials, and provided access to their medical records. The project was approved by the North West Multi-Centre Research Ethics Committee and appropriate informed consent was obtained from participants (Bycroft et al., 2018). Data used in the project were accessed through an approved application to UKB (Application ID: 48818). We analyzed data from participants of self-reported European ancestry and excluded individuals from this analysis if they enrolled outside of England, died prior to September 2019, or the self-reported sex was not consistent with genetic information. In total, 389,620 participants were included in this analysis. COVID-19 laboratory test results reported for UKB participants in England from March 16 to May 18, 2020, were included. During this period, COVID-19 testing was largely restricted to hospitalized patients with serious illness who required active medical intervention; therefore, these data are regarded as a proxy for hospitalization for COVID-19 in England only (Iacobucci, 2020).

### 2.2. Phenotypes

We analyzed three sets of phenotypes (i.e., inpatient hospital records, cancer registry, and death registry) available in the UKB database. We used the International Classification of Diseases (ICD) versions 9 and 10 to identify cases in the hospital episode statistics, with both incident and prevalent cases included. Self-reported diagnoses were not considered. Diagnoses of ICD9/ICD10 for phenotypic analyses were mapped to the PheCODE grouping system. Compared with ICD codes, phecodes have been shown to closely align with disease categories commonly used in clinical practice and genomic studies (Wu et al., 2019). For each disease category represented by a phecode, we recoded participants with the phecode as cases, whereas participants without the target phecode or its parent or child phecodes were classified as controls. Analysis was limited to phecodes that had enough cases in order to generate more than 80% statistical power. The numbers of cases and controls in analyses of phenotypes are shown in Table S1. Sex-stratified analyses were performed in males and females, separately. In addition to these phenotypes, 30 biomarkers, measured in blood samples collected at recruitment, were included in association analyses (Bycroft et al., 2018).

### 2.3. Tag SNPs for *ACE2* and *TMPRSS2*

Tag SNPs, capturing haplotype structures and common genetic variants in the regulatory and coding regions of *ACE2* and *TMPRSS2* were selected based on the whole-genome sequencing data of 91 British individuals from the 1,000 Genomes Project (Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, 2015). For each gene, genetic variants fulfilling all the following criteria were included in the analysis: 1) 5 kb upstream or downstream of the coding region or associated with the expression of the gene in any tissue in the GTEx project (Consortium, 2017); 2) biallelic SNPs; 3) minor allele frequencies  $\geq 5\%$ . Tag SNPs were further selected using the Tagger function in Haploview 4.2 with  $r^2 > 0.5$  (Barrett et al., 2005). Seventeen tag SNPs were selected for *ACE2* and 31 were selected for *TMPRSS2*. Once these tag SNPs were identified, we tested their associations with hospitalization for COVID-19 and phenotypes across the phenome in the UK Biobank.

## 2.4. Statistical analysis

Baseline characteristics are presented as the number (percentage) of participants for categorical variables and the mean (standard deviation) for continuous variables. We performed a logistic regression analysis to estimate the association between each phenotype and hospitalization with a positive COVID-19 test, while correcting for age, sex, body mass index (BMI), assessment center, and 10 genetic principal components. Two types of control samples were used: all other UKB participants who had not yet been tested or tested negative and only participants who tested negative. The effect of each categorical phenotype was measured as an odds ratio (OR). In the blood biomarker analysis, biomarker levels were normalized, and the ORs correspond to each one standard deviation (SD) increase in biomarker levels. We performed a logistic regression analysis of each pair of SNPs and phenotypes, with cases and controls defined as above, while adjusting for age, sex, assessment center, type of genotyping array, and the top 10 genetic principal components, to evaluate the associations between tag SNPs in *ACE2* and *TMPRSS2* with all possible phenotypes.

The associations between tag SNPs and COVID-19 test positivity were analyzed using all other UKB participants as the control group. Sex-stratified analyses were separately conducted in males and females, separately, with the same covariates except sex. We applied the Bonferroni correction for the number of phenotypes evaluated in the comparison between COVID-19 patients and other UKB participants. Statistical analyses were performed using R (version 3.6.2), and the PheWAS package was used to facilitate the phenome-wide association analysis (Carroll et al., 2014).

## 3. Results

### 3.1. Baseline characteristics of the study population

After quality control and the exclusion of confounding factors, such as residence outside of England, death prior to September 2019, and

**Table 1**  
Baseline characteristics of UK Biobank participants by COVID-19 test positivity.

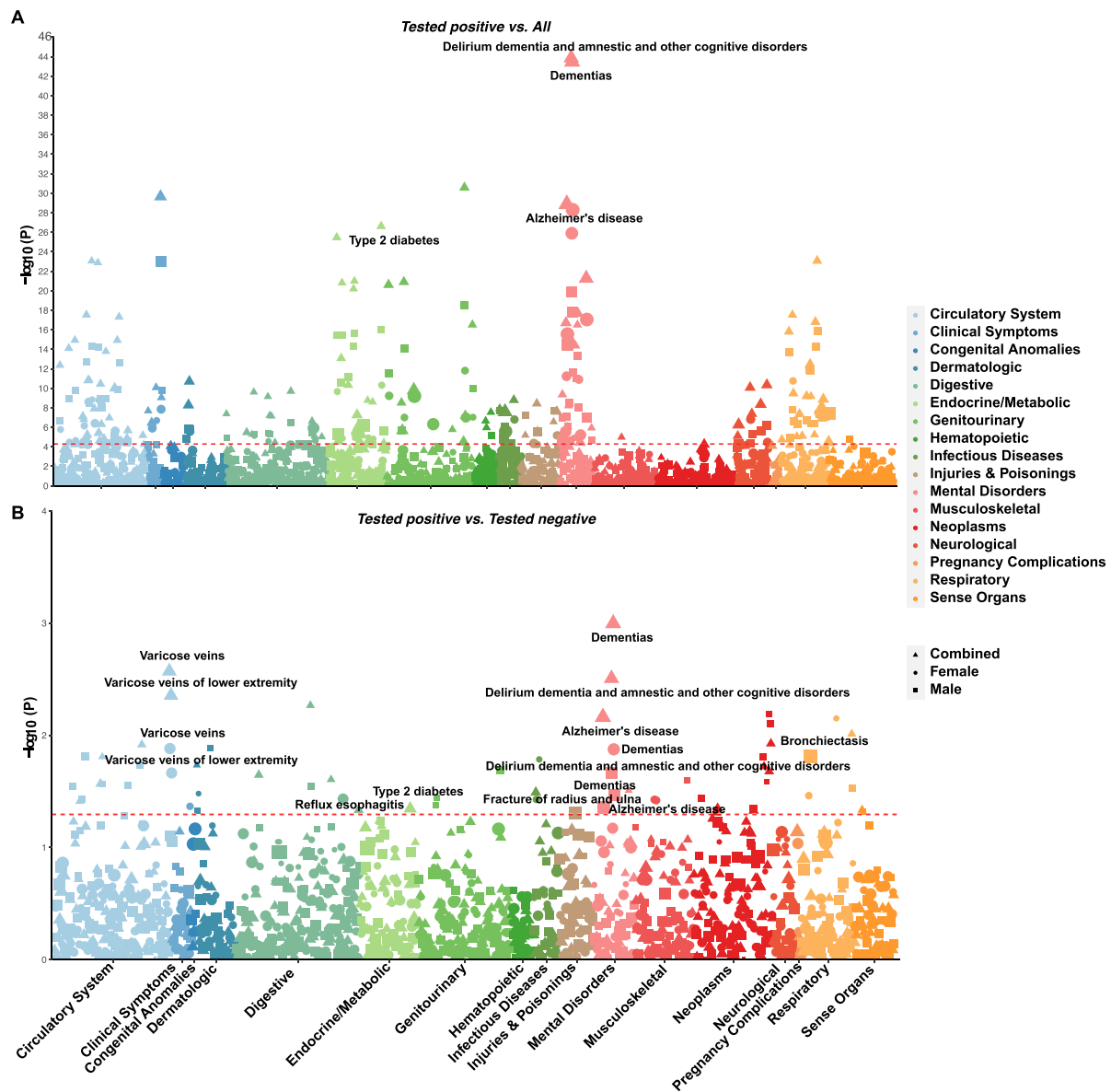
Characteristic	All	Test	COVID-19 Test
	Participants	Negative	Positive
	(N = 389,620)	(N = 2,793)	(N = 1,091)
Sex			
Male	176,151	1,343	582
Female	213,469	1,450	509
Age at Recruitment	56.62(8.03)	57.71(8.59)	57.17(9.23)
BMI	27.25(4.95)	27.91(5.88)	28.52(5.85)
Smoking status			
not answer	1,368	13	8
Never	211,495	1,282	492
Previous	138,423	1,110	471
Current	38,315	388	120
Alcohol use			
not answer	329	6	2
Never	12,195	101	30
Previous	12,885	144	67
Current	364,211	2,542	992
Pre-existing conditions			
Type 2 diabetes	19,732	450	144
Hypertension	86,337	306	394
Obesity	15,361	214	83
Dementias	992	45	34
Alzheimer's disease	509	25	20
Biomarkers			
HDL cholesterol	1.45 (0.38)	1.43 (0.37)	1.35 (0.38)
Apolipoprotein A	1.55 (0.26)	1.52 (0.27)	1.48 (0.24)
Triglycerides	1.74 (1.01)	1.76 (1.02)	1.86 (1.03)
Rheumatoid factor	24.35 (19.6)	22.1 (19.6)	27.9 (18.6)

inconsistent sex information, a total of 389,620 self-reported European participants were included in our study (Tables 1). Of these participants, 3,884 (0.997%) patients were tested for COVID-19, and 1,091 of them (19.75%) were tested positive at least once while in the hospital. Compared to all other UKB participants (i.e., untested or tested negative), patients who tested positive for COVID-19 were older ( $p = 0.024$ ), tend to be male ( $p = 7.33 \times 10^{-8}$ ), had a higher BMI ( $p = 7.27 \times 10^{-18}$ ) and were previous or current smokers ( $p = 2.84 \times 10^{-5}$ ). Compared to participants who were tested negative, patients with COVID-19 still had a higher BMI ( $p = 3.9 \times 10^{-3}$ ) and tend to be male ( $p = 3.2 \times 10^{-3}$ ) (Table S2 in supporting information). Age, sex, and BMI were included as covariates in our analyses.

### 3.2. Phenome-wide association study for COVID-19

We performed an exhaustive association analysis across 974 phenotypes and 30 blood biomarkers to identify pre-existing medical conditions that are overrepresented in patients with COVID-19, with all other UKB participants and COVID-19-negative individuals served as controls. Compared to all other UKB participants, a wide range of pre-existing conditions were overrepresented in patients with COVID-19, even after the Bonferroni correction (Fig. 1A and Table S3 in the supporting information). Some of the most significant associations included the overall category of delirium, dementia, amnesic and other cognitive disorders ( $p = 1.36 \times 10^{-44}$ ), dementia ( $p = 3.48 \times 10^{-44}$ ), renal failure ( $p = 2.63 \times 10^{-31}$ ), Alzheimer's disease ( $p = 1.33 \times 10^{-29}$ ), type 2 diabetes ( $p = 2.45 \times 10^{-27}$ ), pneumonia ( $p = 8.43 \times 10^{-24}$ ), hypertension ( $p = 1.31 \times 10^{-23}$ ), and hyperlipidemia ( $p = 9.94 \times 10^{-22}$ ). Since these overrepresented conditions may only reflect sampling bias in individuals who received COVID-19 tests, we made further correction against this bias by comparing patients with COVID-19 to participants who were tested negative (Fig. 1B and Fig. 2 and Table S4 in supporting information). Some phenotypes were consistently overrepresented in patients with COVID-19: the overall category of delirium, dementia, amnesic and other cognitive disorders (OR = 1.90, 95% CI: 1.24–2.90 in combined samples; OR = 2.06, 95% CI: 1.11–3.81 in males), dementia (OR = 2.16, 95% CI: 1.36–3.42 in combined samples; OR = 2.05, 95% CI: 1.05–3.98 in males; OR = 2.24, 95% CI: 1.18–4.24 in females), Alzheimer's disease (OR = 2.29, 95% CI: 1.25–4.16 in combined samples; OR = 2.40, 95% CI: 1.02–5.62 in males), and type 2 diabetes (OR = 1.25, 95% CI: 1.00–1.55 in combined samples). The comparison to participants who were tested negative also revealed the following novel comorbidities that were overrepresented in patients with COVID-19: bronchiectasis (OR = 2.95, 95% CI: 1.23–7.05 in males), varicose veins (OR = 1.71, 95% CI: 1.20–2.42 in combined samples; OR = 1.81, 95% CI: 1.13–2.89 in females), varicose veins in the lower extremities (OR = 1.69, 95% CI: 1.18–2.42 in combined samples; OR = 1.74, 95% CI: 1.08–2.80 in females), reflux esophagitis (OR = 1.65, 95% CI: 1.03–2.63 in females), fracture of the clavicle or scapula (OR = 8.40, 95% CI: 1.61–43.43 in females), and fracture of the radius and ulna (OR = 2.5, 95% CI: 1.00–6.22 in males) (Fig. 2).

In the blood biomarker analysis, at the nominal significance level, four biomarkers were different between patients with COVID-19 and participants who were tested negative (Fig. 3 and Table S5 in the supporting information). Each SD increase in high-density lipoprotein cholesterol (HDL) and apolipoprotein A (Apo(a)) levels was associated with reduced risks of COVID-19 (OR = 0.82, 95% CI: 0.75–0.90; OR = 0.85, 95% CI: 0.78–0.92, respectively). On the other hand, rheumatoid factor and triglyceride levels were associated with increased risks (OR = 1.33, 95% CI: 1.03–1.72; OR = 1.08, 95% CI: 1.00–1.16, respectively). Overall, our extensive phenome-wide search highlighted multiple pre-existing medical conditions, particularly Alzheimer's disease and dementia, as risk factors for COVID-19.



**Fig. 1.** Manhattan plots showing the associations between COVID-19 and pre-existing comorbidities. (A) Comparison between COVID-19 patients and all other UKB participants. The significance threshold with Bonferroni correction ( $p < 5.13 \times 10^{-5}$ ) is represented by the red dashed line. Labeled phenotypes are over-represented in COVID-19 patients and remain nominally significant after the correction for bias in the testing sample. (B) Comparison between COVID-19 patients and those tested negative. Nominal significance ( $p < 0.05$ ) is indicated by the red dashed line. Phenotypes over-represented in COVID-19 patients are labeled. Results based on males, females, and combined samples are shown with different shapes. The size of the shape is proportional to the odds ratio. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 3.3. Phenome-wide association study of *ACE2* and *TMPRSS2*

Genetic variants in human genes that mediate SARS-CoV-2 infection (e.g., *ACE2* and *TMPRSS2*) may directly affect viral susceptibility or indirectly influence pre-existing medical conditions. We evaluated the direct associations between common genetic variants in these two genes and COVID-19 test positivity to assess the former possibility. Seventeen and 31 tag SNPs were selected to capture haplotype structures and common genetic variants in the regulatory and coding regions of *ACE2* and *TMPRSS2*, respectively. We did not identify associations reaching the genome-wide significance cutoff ( $p < 5 \times 10^{-8}$ , Table S6 in the supporting information). However, five tag SNPs for *TMPRSS2* were associated with COVID-19 test positivity at the nominally significant level ( $p < 0.05$ ) in both analyses using all other UKB participants and individuals who were tested negative as the controls (Table 2). When

comparing patients with COVID-19 to participants who were tested negative, the association of SNP rs7282236 (A/G) passed the Bonferroni correction cutoff (Fig. 4). This SNP had an alternative allele frequency of 75.1% in all other UKB participants, 74.2% in those tested negative, and 77.7% in patients with COVID-19, corresponding to an increased risk of COVID-19 (OR = 1.2, 95% CI: 1.06–1.36,  $p = 3.30 \times 10^{-3}$ ; OR = 1.33, 95% CI: 1.14–1.54,  $p = 2.31 \times 10^{-4}$ , respectively). Collectively, these association signals suggest a possible role of *TMPRSS2* genetic variants in modulating the risk of COVID-19.

We systematically evaluated associations between tag SNPs and 848 phenotypes to broadly evaluate the clinical effects of these two genes. We performed separate analyses of males, females, and combined samples. Although no associations reached the genome-wide significance cutoff, suggestive associations were identified with the Bonferroni correction for the number of phenotypes tested ( $p < 5.9 \times 10^{-5}$ , Fig. 5



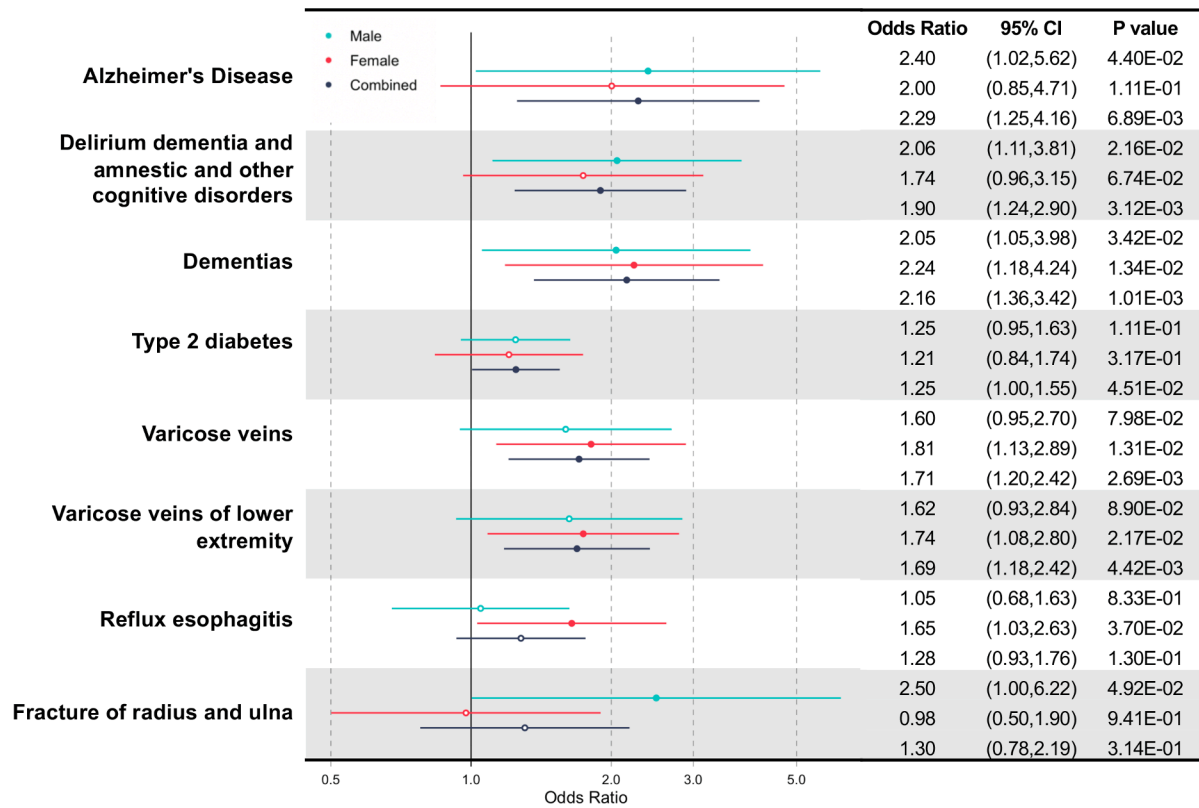


Fig. 2. Forest plot for pre-existing conditions over-represented in COVID-19 patients when compared to tested negatives. Point estimates of ORs are represented by a filled or hollow circle, while horizontal lines indicate the 95% confidence intervals. Filled circles indicate statistical significance. Results for males, females, and combined samples are shown in different colors. The corresponding statistics were shown on the right.

and Tables S7 and S8 in the supporting information). For *ACE2*, only one suggestive association was identified in all analyses, namely, immune deficiency ( $p = 5.65 \times 10^{-5}$ ) in the combined analysis. For *TMPRSS2*, the only phenotype reaching the cutoff value in both combined and female-specific analyses was atypical inflammatory spondylopathies ( $p = 4.2 \times 10^{-5}$  and  $4.3 \times 10^{-6}$ , respectively). In males, four suggestive associations were identified: noninfectious gastroenteritis ( $p = 1.1 \times 10^{-5}$ ), prostate cancer ( $p = 1.1 \times 10^{-5}$ ), symptoms involving the head and neck ( $p = 2.8 \times 10^{-5}$ ), and neoplasm of uncertain behavior ( $p = 2.91 \times 10^{-5}$ ).

#### 4. Discussion

This study leverages the existing extensive genomic and phenotyping data and the recent COVID-19 test results in the UKB to identify risk factors for COVID-19 and to evaluate the clinical effects of genetic variants in key human genes on regulating SARS-CoV-2 infection. Our findings highlighted multiple pre-existing medical conditions as risk factors for COVID-19: dementia, Alzheimer's disease, general cognitive disorders, and type 2 diabetes. In addition, genetic variants in genes related to SARS-CoV-2 infection were found to have suggestive associations with hospitalized COVID-19 and other phenotypes, such as immune deficiency and prostate cancer.

The most significant and consistent risk factors we identified are cognitive disorders, consistent with a few prognostic studies investigating smaller clinical samples. A study of 627 patients with COVID-19 in Northern Italy showed that dementia and its progressive stages were associated with mortality and that these patients commonly exhibited neurological symptoms of delirium and a worsening functional status (Bianchetti et al., 2020). In another study of 214 patients in Wuhan, China, neurological symptoms including acute cerebrovascular diseases, impaired consciousness, and skeletal muscle injury, were observed in 36.5% of patients with COVID-19 and were more common (45.5%) in

patients with a severe illness (Mao et al., 2020). Interestingly, dementia was commonly observed in inpatients with COVID-19 and associated with COVID-19 in models adjusted for demographic characteristics, and smoking and drinking status (Tables S9 supporting information). As the COVID-19 pandemic progresses, reports of neurological manifestations are increasing (Mao et al., 2020; Ding et al., 2020). These manifestations may be direct effects of tissue damages caused by viral infection and replication in the nervous system or indirect effects due to neural immunopathology caused by exuberant unresponsive immune responses triggered by the viruses, or a combination of both direct and indirect effects of the infection, manifesting the neurological complications of the systemic effects of COVID-19 (Ellul et al., 2020). Key issues of SARS-CoV-2 infection and its associated neuropathology include the routes of viral entry, tissue tropism, immune responses, as well as immunopathology in the nervous system (Wu et al., 2020). SARS-CoV-2 may enter the brain through the olfactory bulb. Studies of intranasal injection in mice have shown that the human coronavirus invades the central nervous system through infected white blood cells that cross the blood-brain barrier (Desforges et al., 2019). Additionally, SARS-CoV-2 is found in cerebral vascular endothelial cells that binds to angiotensin-converting enzyme 2 receptor (Yan et al., 2020). Therefore, internal damage to the central nervous system may be directly caused by the virus or the systemic infection in the body. The detailed characteristics of inflammatory infiltrates must be determined to correctly interpret the mechanisms underlying the over-representation of cognitive disorders in patients with COVID-19.

To date, data on pre-existing dementia and COVID-19 hospitalization are limited, although dementia affects more than 40 million people worldwide (Ritchie et al., 2016). As age is one of the greatest risk factors for dementia and cognitive disorders, the vast majority of patients with Alzheimer's disease is aged 70 years or older. When the sample was stratified into four groups by age, we only observed associations

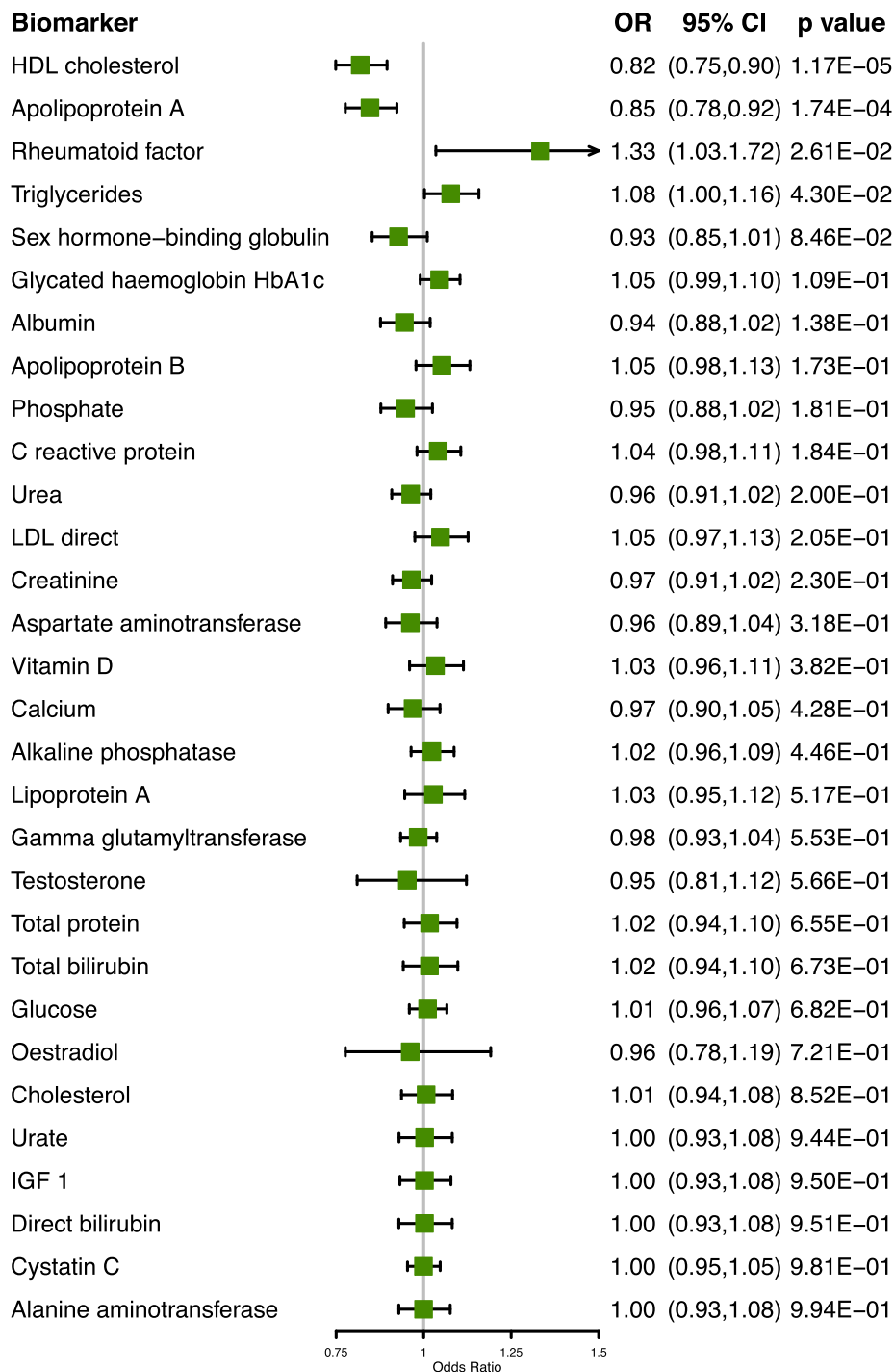


Fig. 3. A Forest plot showing the associations between blood biomarkers and COVID-19. Point estimates of ORs are represented by the filled squares while horizontal lines indicate the 95% confidence intervals. Biomarkers are ordered by their statistical significance.

between an increased risk of COVID-19 and dementia or cognitive disorders in groups older than 70 years, while the relatively younger age group (<70) did not contain a sufficient number of individuals with dementia or cognitive disorders (Table S10 in the supporting information). We also observed a qualitatively similar result when our association model adjusted by the age at 2020 (Table S11 in the supporting information). Based on our findings, cognitive disorders are likely risk comorbidities in older groups and their associated susceptibility to severe COVID-19 is not merely a result of an older age. Another possible explanation for the finding that more individuals with mental disorders suffer from COVID-19 is that they are at a higher risk of viral infection

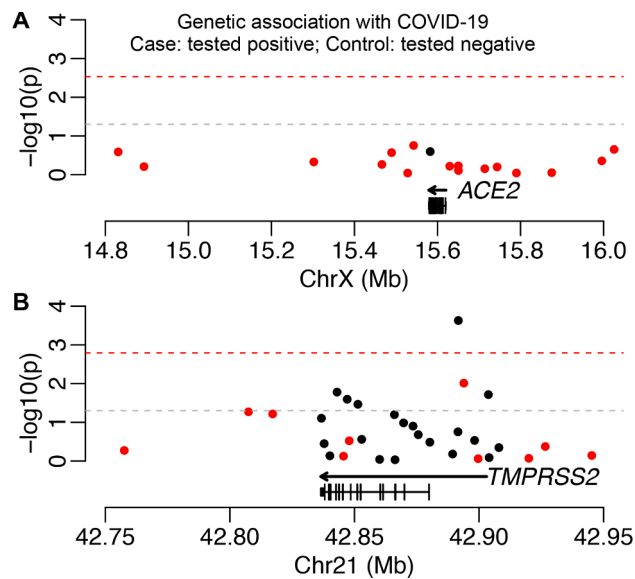
because of their limited self-care ability and their frequent interactions with care providers. Overall, these results should help stimulate COVID-19 research on the special needs of patients with these cognitive conditions. Given the different risks faced by the elderly living with different styles, a more comprehensive strategy with precise approaches of primary prevention may be desirable during this and similar pandemics.

The comparison between patients with COVID-19 and participants who were tested negative also revealed associations of four blood biomarkers and multiple novel comorbidities with COVID-19. Among the four blood biomarkers, three are indicators of cardiovascular health,

**Table 2**List of tag SNPs at *TMPRSS2* associated with COVID-19 at the nominal significance level.

SNP	Position on Chr21	REF/ALT	EAF (Non/Neg/Pos)	Pos vs. Non		Pos vs. Neg	
				OR (95% CI)	p value	OR (95% CI)	p value
rs7282236	41519797	A/G	0.751/0.742/0.777	1.203 (1.06,1.36)	0.0033	1.327 (1.14,1.54)	0.00023
rs114837856	41475211	A/T	0.494/0.492/0.477	0.856 (0.76,0.97)	0.013	0.842 (0.72,0.98)	0.025
rs56695953	41497808	A/G	0.824/0.824/0.807	0.864 (0.76,0.98)	0.027	0.874 (0.74,1.03)	0.104
rs8134657	41521981	A/G	0.905/0.910/0.891	0.839 (0.71,0.98)	0.033	0.765 (0.63,0.94)	0.009
rs915823	41479527	C/A	0.798/0.796/0.808	1.140 (1.00,1.29)	0.050	1.186 (1.01,1.39)	0.033
rs35050484	41471079	A/G	0.963/0.967/0.952	0.784 (0.61,1.00)	0.050	0.681 (0.50,0.93)	0.016
rs56379149	41531825	T/G	0.529/0.538/0.508	0.905(0.82,1.00)	0.055	0.863 (0.76,0.98)	0.018

SNP: single nucleotide polymorphism; REF: reference allele; ALT: alternative allele; EAF: effect allele frequency (alternative allele is the effect allele); Non: non-positive, all other UK Biobank participant that are untested or tested negative; Neg: COVID-19 tested negative; Pos: COVID-19 tested positive; OR: odds ratio; CI: confidence interval.

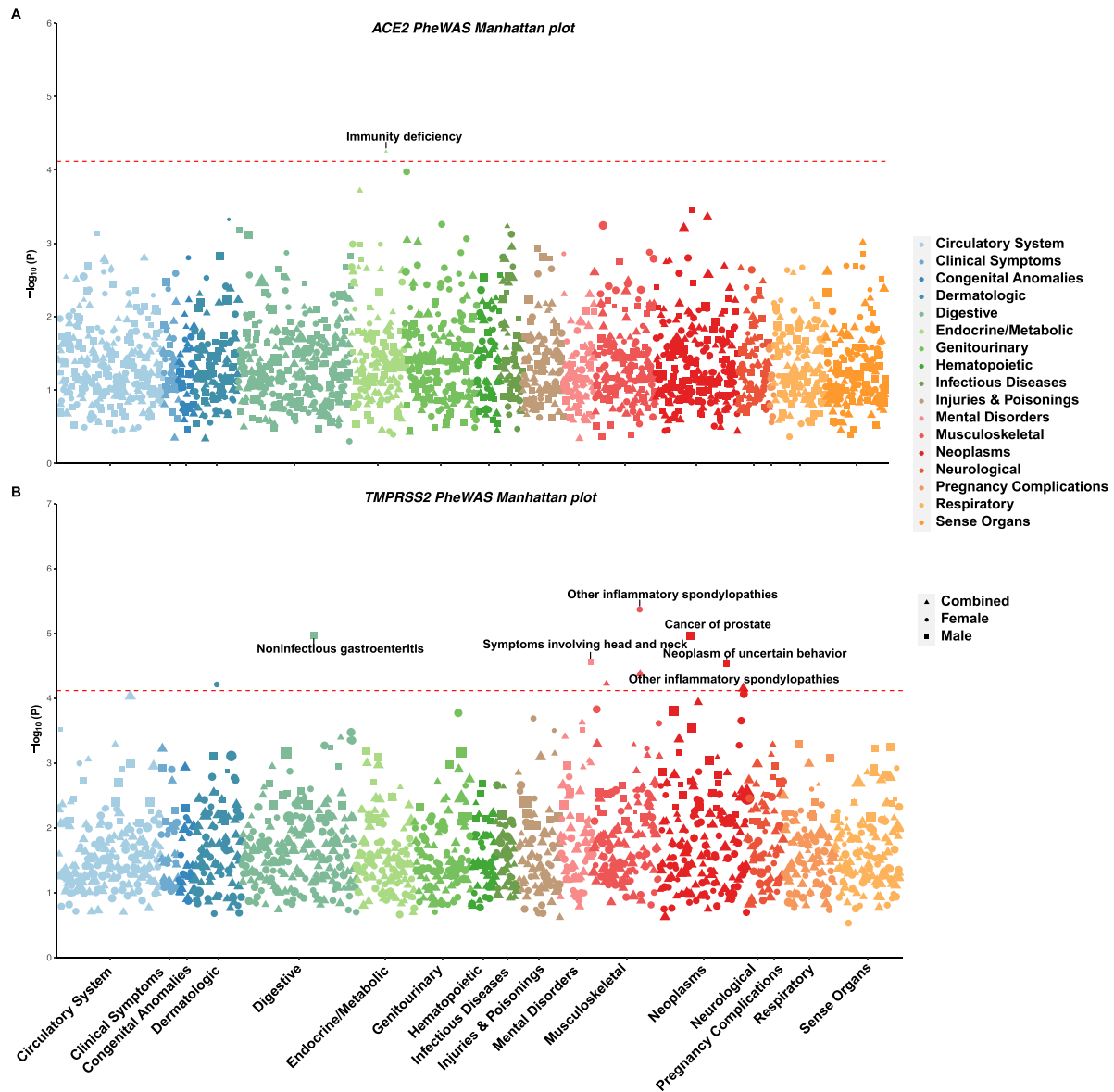


**Fig. 4.** Associations between (A) *ACE2* and (B) *TMPRSS2* genetic variants and COVID-19. Each dot is a tag SNP. SNPs associated with gene expression are indicated in red. The nominal significance cutoff ( $p < 0.05$ ) is represented by the gray dashed line, while the significance threshold with Bonferroni correction for the number of SNPs tested is indicated by the red dashed line. The model of the longest transcript is shown at the bottom while the gene region, covering all transcripts, is shown as an arrow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

including HDL, Apo(a), and triglycerides. These findings consistently identified an association between deteriorating cardiovascular health (i.e., decreased HDL and Apo(a) levels, but increased triglyceride levels) and a higher risk of COVID-19 test positivity. We explored the association between these significant risk biomarkers and a wide range of neurological symptoms. The three indicators related to cardiovascular health were associated with multiple neurological phenotypes, including peripheral nerve disorders, headache syndromes, migraine, inflammatory and toxic neuropathy, sleep apnea, and sleep disorders (Bonferroni-corrected  $p$  value  $< 0.001$ ), while no neurological symptoms were associated with the level of rheumatoid factor (Table S12 in the supporting information). Thus, mental health is closely associated with indicators of cardiovascular health. While these blood biomarkers were measured and collected a decade before this pandemic, our results might indicate potential applications for more personalized COVID-19 prevention efforts, particularly among middle-aged healthy individuals. Although we can speculate about potential connections of our results with the current knowledge of COVID-19, longitudinal and well-characterized data from patients are needed for further exploration.

Importantly, these risk factors were identified not only by comparing patients with COVID-19 to the background cohort but also to individuals tested negative, thus correcting for sampling bias. During the study period, COVID-19 testing was prioritized for high-risk groups, particularly when the testing capacity was limited, and some symptoms or pre-existing conditions (e.g., pneumonia) were overrepresented in the individuals who underwent testing due to the selection process (Atkins et al., 2020; Ho et al., 2004). The polymerase chain reaction test used in the UK had a false negative rate between 2% and 29% on initial testing (Watson et al., 2020). The correction of this bias is critical to identify true COVID-19 risk factors (Griffith et al., 2020). A comparison of patients with positive and negative tests may therefore exclude other reasons for hospital admission for symptoms resembling COVID-19, as well as false negatives. Our finding of type 2 diabetes as a risk factor for a positive COVID-19 test is also consistent with previous studies (de Lusignan et al., 2020; Kumar et al., 2020). Novel risk factors that have not been reported in previous studies include bronchiectasis, varicose veins, reflux esophagitis, fracture of the clavicle or scapula, and fracture of the radius and ulna. These risk factors may exacerbate COVID-19 progression, or patients with these pre-existing conditions may be more frequently exposed to infection. Future studies are needed to elucidate the underlying mechanisms of these associations. Many pre-existing conditions with significant associations in the comparison of COVID-19 patients to the rest of the UKB sample were not found to be significant in the comparison to those tested negative. Interpretation on these pre-existing conditions should be taken with caution, as the control group of those tested negative has a much smaller sample size and may not have enough statistical power.

Our phenome-wide association study of *ACE2* and *TMPRSS2* revealed evidence suggesting associations with COVID-19 test positivity and other medical conditions. None of the associations reached the genome-wide significance cutoff, which is consistent with very recent studies (Cirulli et al., 2004; Lopera et al., 2004). However, our study identified associations that passed the stringent Bonferroni correction cutoff. In terms of a direct association with COVID-19, one tag SNP in *TMPRSS2*, rs7282236, is associated with COVID-19 test positivity, regardless of which control group was used. A very recent study used exome sequencing data from 49,953 UKB subjects and 74 patients with COVID-19 to evaluate the contributions of rare coding variants in *ACE2* and *TMPRSS2* to COVID-19, but did not find associations (Curtis, 2005). Our focus on common coding and regulatory variants in the almost full UKB cohort of participants with COVID-19 patients may have facilitated our novel discovery. In terms of broader clinical effects, we observed one association (i.e., immune deficiency) with *ACE2* that met the Bonferroni correction cutoff. More associations were identified for *TMPRSS2*, including atypical inflammatory spondylopathies, noninfectious gastroenteritis, prostate cancer, symptoms involving the head and neck, and neoplasm of uncertain behavior. Notably, the association of *TMPRSS2* with prostate cancer has been previously identified (Al Olama et al., 2014), supporting the validity of our findings. It is possible that these genetic variants are associated with other COVID-19-relevant



**Fig. 5.** Manhattan plots showing the associations between (A) *ACE2* and (B) *TMPRSS2* genetic variants and disease status. The significance threshold with Bonferroni correction ( $p < 5.82 \times 10^{-5}$ ) is represented by the red dashed line. For each phenotype, only the p value from the most significant tag SNP is shown. Results for males, females, and combined samples are shown in different shapes. The size of the shape is proportional to the odds ratio. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

phenotypes, which were not available in UK Biobank, such as specific immune cell types or cytokine levels. Among the analyzed hypotheses, the most interesting signal is the relation between genetic variants in *TMPRSS2* and COVID-19 test positivity. It is of great interest to evaluate if these genetic variants are associated with different degrees of severity or different disease manifestations in COVID-19 patients.

Our study has strengths and limitations. UKB is a large prospective cohort with extensive genomic and phenotyping information, enabling a hypothesis-free phenome-wide scan for COVID-19 risk factors. The availability of the background cohort and a subgroup of individuals who were tested for COVID-19 allow us to compare patients with COVID-19 to the general population while simultaneously correcting for sampling bias. This relatively small sample size renders the analysis susceptible to collider bias. We applied two methods to adjust for collider bias and examined the sensitivity of our study: i) where the selected sample is nested within the complete UKB dataset that comprises samples

representative of the target population, or ii) where the dataset consists of only the tested samples. However, it is difficult to estimate the extent of sample selection, and even if that parameter were known, we would be unable to prove that it has been fully explained by any method. Collider bias might also arise because of the original selection in the UK Biobank, which include more healthy and well-educated participants. In addition to pre-existing medical conditions and biomarkers, our study evaluated the possible role of genetic factors in COVID-19 by studying candidate genes. Our phenome-wide analysis of the two key genes related to SARS-CoV-2 infection also provided clinical insights into their biological functions. Another limitation of our study was the inability to provide additional information about the specific symptoms or outcomes of the patients with COVID-19. The physical measurements, biomarker levels and medical conditions were either measured at recruitment or retrieved from medical records, and therefore they may not accurately reflect the current health status. To reduce the potential



confounding effect of ethnicity, which is well-known to affect COVID-19-related health disparity, our analysis was restricted to participants of European descent, a group with the biggest sample size. Future studies with large sample sizes are urgently needed for other ethnicities. Last, our study is associative in nature and was unable to address the causal roles of risk factors. Our findings identified associations of an extensive range of pre-existing conditions and genetic variants as being associated with COVID-19. However, the real causal effects of risk factors on COVID-19 susceptibility are likely to vary by genetic background, lifestyle, and social connectedness and are presumably more complicated than indicated by our population-level screen.

## 5. Conclusion

Overall, our unbiased phenome-wide study in UK Biobank confirmed known and identified novel risk factors for COVID-19, including dementia, Alzheimer's disease, type 2 diabetes, blood biomarkers of cardiovascular health, and genetic variants in *TMPRSS2*. These systematic discoveries provide insights into the management, prevention, and treatment of COVID-19 during future phases of the outbreak, while highlighting an urgent need of special protective care for patients with cognitive disorders.

## 6. Ethics approval and consent to participate

UK Biobank was approved by the North West Multi-Centre Research Ethics Committee and appropriate informed consent was obtained from participants. Data used in the project was accessed through an approved application to UK Biobank (Application ID: 48818).

## Author contributions

JZ, WH, and KY conceptualized the study idea and design. JZ, CL, and YS curated the data and performed formal analysis. JZ and KY prepared the figures. JZ and KY wrote the original draft of the manuscript. All authors reviewed and edited the first draft. KY acquired the funding and supervised the study.

## 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.

## Acknowledgments

We would like to thank all UK Biobank participants and administrators for data access. We also want to express our gratitude to all other Ye lab members for stimulating discussions and especially to Michael Francis for managing the UK Biobank data in the lab. KY is supported by the University of Georgia Research Foundation. Funding sources had no involvement in the conception, design, analysis, or presentation of this work.

## Appendix A. Supplementary data

The summary statistics supporting the conclusions of this article are included within the article and its additional files. Individual-level genetic and phenotypic data are available from the UK Biobank (<https://www.ukbiobank.ac.uk/register-apply/>) through application. Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbi.2020.10.019>.

## References

- Al Olama, A.A., Kote-Jarai, Z., Berndt, S.I., Conti, D.V., Schumacher, F., Han, Y., Benlloch, S., Hazelett, D.J., Wang, Z., Saunders, E., et al., 2014. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat. Genet.* 46 (10), 1103–1109.
- Atkins, J.L., Masoli J.A., Delgado, J., Pilling, L.C., Kuo C.-L.C., Kuchel G., Melzer, D., 2020. Preexisting comorbidities predicting severe COVID-19 in older adults in the uk biobank community cohort. *medRxiv*, 2020: 2020.2005.20092700.
- Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21 (2), 263–265.
- Bianchetti, A., Rozzini, R., Guerini, F., Boffelli, S., Ranieri, P., Minelli, G., Bianchetti, L., Trabucchi, M., 2020. Clinical Presentation of COVID19 in Dementia Patients. *J. Nutr. Health Aging* 1–3.
- Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al., 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 47 (D1), D1005–D1012.
- Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al., 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562 (7726), 203–209.
- Canela-Xandri, O., Rawlik, K., Tenesa, A., 2018. An atlas of genetic associations in UK Biobank. *Nat. Genet.* 50 (11), 1593–1599.
- Carroll, R.J., Bastarache, L., Denny, J.C., 2014. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* 30 (16), 2375–2376.
- Chang, D., Gao, F., Slavney, A., Ma, L., Waldman, Y.Y., Sams, A.J., Billing-Ross, P., Madar, A., Spritz, R., Keinan, A., 2014. Accounting for eXcentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PLoS One* 9 (12), e113684.
- Cirulli, E.T., Riffle, S., Bolze, A., Washington, N.L., 2004. Revealing variants in SARS-CoV-2 interaction domain of ACE2 and loss of function intolerance through analysis of >200,000 exomes. *bioRxiv* 2020 (2020), 2007.
- Consortium G.T., Laboratory D.A., 2017. Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida et al: Genetic effects on gene expression across human tissues. *Nature*, 550(7675):204-213.
- Curtis, D., 2005. Coding variants in ACE2 and TMPRSS2 are not major drivers of COVID-19 severity in UK Biobank subjects. *medRxiv* 2020 (2020), 2001.
- de Lusignan, S., Dorward, J., Correa, A., Jones, N., Akinyemi, O., Amirthalingam, G., Andrews, N., Byford, R., Dabrera, G., Elliot, A., et al., 2020. Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: a cross-sectional study. *Lancet Infect Dis*.
- Desforges, M., Le Coupanec, A., Dubeau, P., Bourgoin, A., Lajoie, L., Dube, M., Talbot, P.J., 2019. Human Coronaviruses and Other Respiratory Viruses: Underestimated Opportunistic Pathogens of the Central Nervous System? *Viruses* 12 (1).
- World Health Organization, 2020. Coronavirus disease (COVID-19) pandemic [<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/>]. (Accessed 31 May 2020).
- Ding H, Yin S, Cheng Y, Cai Y, Huang W, Deng W, 2020. Neurologic manifestations of nonhospitalized patients with COVID-19 in Wuhan, China. *MedComm* Doi: 10.1002/mco1002.1013.
- Docherty, A.B., Harrison, E.M., Green, C.A., Hardwick, H.E., Pius, R., Norman, L., Holden, K.A., Read, J.M., Dondelinger, F., Carson, G., et al., 2020. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 369, m1985.
- Ellul, M.A., Benjamin, L., Singh, B., Lant, S., Michael, B.D., Easton, A., Kneen, R., Defres, S., Sejvar, J., Solomon, T., 2020. Neurological associations of COVID-19. *Lancet Neurol.* 19 (9), 767–783.
- Genomes Project C., Auton A., Brooks L.D., Durbin R.M., Garrison E.P., Kang H.M., Korbel J.O., Marchini J.L., McCarthy S., McVean G.A. et al. 2015. A global reference for human genetic variation. *Nature*, 526 (7571), 68-74.
- Grasselli, G., Zangrillo, A., Zanella, A., Antonelli, M., Cabrini, L., Castelli, A., Cereda, D., Coluccello, A., Foti, G., Fumagalli, R., et al., 2020. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA*.
- Griffith G., Morris T.T., Tudball M., Herbert A., Mancano G., Pike L., Sharp G.C., Palmer T.M., Davey Smith G., Tilling K. et al. 2020. Collider bias undermines our understanding of COVID-19 disease risk and severity. *medRxiv* 2020.2005.2004.20090506.
- Guo, T., Fan, Y., Chen, M., Wu, X., Zhang, L., He, T., Wang, H., Wan, J., Wang, X., Lu, Z., 2020. Cardiovascular Implications of Fatal Outcomes of Patients With Coronavirus Disease 2019 (COVID-19). *JAMA Cardiol*.
- He, L., Kernogitski, Y., Kulminkaya, I., Loika, Y., Arbee, K.G., Loiko, E., Bagley, O., Duan, M., Yashkin, A., Ukraintseva, S.V., et al., 2016. Pleiotropic meta-analyses of longitudinal studies discover novel genetic variants associated with age-related diseases. *Front. Genet.* 7, 179.
- Ho, F.K., Celis-Morales, C.A., Gray, S.R., Katikireddi, S.V., Niedzwiedz, C.L., Hastie, C., Lyall, D.M., Ferguson, L.D., Berry, C., Mackay, D.F., et al., 2004. Modifiable and non-modifiable risk factors for COVID-19: results from UK Biobank. *medRxiv* 2020 (2020), 2028.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al., 2020. SARS-CoV-2 cell

- entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181 (2), 271–280 e278.
- Iacobucci, G., 2020. Covid-19: What is the UK's testing strategy? *BMJ* 368, m1222.
- Kumar, A., Arora, A., Sharma, P., Anikhindi, S.A., Bansal, N., Singla, V., Khare, S., Srivastava, A., 2020. Is diabetes mellitus associated with mortality and severity of COVID-19? A meta-analysis. *Diabetes Metab. Syndr.* 14 (4), 535–545.
- Kuo, C.-L., Pilling, L.C., Atkins, J.L., Masoli, J.A.H., Delgado, J., Kuchel, G.A., Melzer, D., 2020. APOE e4 genotype predicts severe COVID-19 in the UK Biobank community cohort. *J. Gerontol. Series A*.
- Lopera, E., van der Graaf, A., Lanting, P., van der Geest, M., Fu, J., Swertz, M., Franke, L., Wijmenga, C., Deelen, P., Zernakova, A., et al., 2004. Lack of association between genetic variants at ACE2 and TMPRSS2 genes involved in SARS-CoV-2 infection and human quantitative phenotypes. *medRxiv* 2020 (2020), 2022.
- Mao L, Jin H, Wang M, Hu Y, Chen S, He Q, Chang J, Hong C, Zhou Y, Wang D et al., 2020. Neurologic Manifestations of Hospitalized Patients With Coronavirus Disease 2019 in Wuhan, China. *JAMA Neurol.*
- Mao, L., Jin, H., Wang, M., Hu, Y., Chen, S., He, Q., Chang, J., Hong, C., Zhou, Y., Wang, D., et al., 2020. Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan, China. *JAMA Neurol.* 77 (6), 683–690.
- Murray, M.F., Kenny, E.E., Ritchie, M.D., Rader, D.J., Bale, A.E., Giovanni, M.A., Abul-Husn, N.S., 2020. COVID-19 outcomes and the human genome. *Genet. Med.*
- Niedzwiedz, C.L., O'Donnell, C.A., Jani, B.D., Demou, E., Ho, F.K., Celis-Morales, C., Nicholl, B.I., Mair, F.S., Welsh, P., Sattar, N., et al., 2020. Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank. *BMC Medicine* 18 (1), 160.
- Petrilli, C.M., Jones, S.A., Yang, J., Rajagopalan, H., O'Donnell, L., Chernyak, Y., Tobin, K.A., Cerfolio, R.J., Francois, F., Horwitz, L.I., 2020. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. *BMJ* 369, m1966.
- Raisi-Estabragh, Z., McCracken, C., Ardissino, M., Bethell, M.S., Cooper, J., Cooper, C., Harvey, N.C., Petersen, S.E., 2005. Non-white ethnicity, male sex, and higher body mass index, but not medications acting on the renin-angiotensin system are associated with coronavirus disease 2019 (COVID-19) hospitalisation: review of the first 669 Cases From The UK Biobank. *medRxiv* 2020 (2020), 2010.
- Richardson, S., Hirsch, J.S., Narasimhan, M., Crawford, J.M., McGinn, T., Davidson, K. W., the Northwell, C-R.C., Barnaby, D.P., Becker, L.B., Chelico, J.D., et al., 2020. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*.
- Ritchie, C.W., Molinuevo, J.L., Truyen, L., Satlin, A., Van der Geysen, S., Lovestone, S., 2016. European Prevention of Alzheimer's Dementia C: Development of interventions for the secondary prevention of Alzheimer's dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. *Lancet Psych.* 3 (2), 179–186.
- Watson, J., Whiting, P.F., Brush, J.E., 2020. Interpreting a covid-19 test result. *BMJ* 369, m1808.
- Wu, P., Gifford, A., Meng, X., Li, X., Campbell, H., Varley, T., Zhao, J., Carroll, R., Bastarache, L., Denny, J.C., et al., 2019. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med. Inform.* 7 (4), e14325.
- Wu, Y., Xu, X., Chen, Z., Duan, J., Hashimoto, K., Yang, L., Liu, C., Yang, C., 2020. Nervous system involvement after infection with COVID-19 and other coronaviruses. *Brain Behav. Immun.* 87, 18–22.
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., Zhou, Q., 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367 (6485), 1444–1448.
- Yang, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., Ji, R., Wang, H., Wang, Y., Zhou, Y., 2020. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int. J. Infect. Dis.* 94, 91–95.
- Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579 (7798), 270–273.