



OPEN

## Shared genomic architecture between COVID-19 severity and numerous clinical and physiologic parameters revealed by LD score regression analysis

Jinyoung Byun<sup>1,2,6</sup>, Younghun Han<sup>1,2,6</sup>, Kyle M. Walsh<sup>3,6</sup>, Amy S. Park<sup>1</sup>, Melissa L. Bondy<sup>4</sup> & Christopher I. Amos<sup>1,2,5</sup>

The COVID-19 pandemic has produced broad clinical manifestations, from asymptomatic infection to hospitalization and death. Despite progress from genomic and clinical epidemiology research, risk factors for developing severe COVID-19 are incompletely understood and identification of modifiable risk factors is desperately needed. We conducted linkage disequilibrium score regression (LDSR) analysis to estimate cross-trait genetic correlation between COVID-19 severity and various polygenic phenotypes. To attenuate the genetic contribution of smoking and BMI, we further conducted sensitivity analyses by pruning genomic regions associated with smoking/BMI and repeating LDSR analyses. We identified robust positive associations between the genetic architecture of severe COVID-19 and both BMI and smoking. We observed strong positive genetic correlation ( $r_g$ ) with diabetes ( $r_g = 0.25$ ) and shortness of breath walking on level ground ( $r_g = 0.28$ ) and novel protective associations with vitamin E ( $r_g = -0.53$ ), calcium ( $r_g = -0.33$ ), retinol ( $r_g = -0.59$ ), Apolipoprotein A ( $r_g = -0.13$ ), and HDL ( $r_g = -0.17$ ), but no association with vitamin D ( $r_g = -0.02$ ). Removing genomic regions associated with smoking and BMI generally attenuated the associations, but the associations with nutrient biomarkers persisted. This study provides a comprehensive assessment of the shared genetic architecture of COVID-19 severity and numerous clinical/physiologic parameters. Associations with blood and plasma-derived traits identified biomarkers for Mendelian randomization studies to explore causality and nominates therapeutic targets for clinical evaluation.

Since the outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in December 2019, the global pandemic of coronavirus disease 2019 (COVID-19) has resulted in more than 205 million confirmed cases and 4.3 million deaths worldwide as of August 13, 2021<sup>1</sup>. Genome-wide association studies (GWAS) have revealed the genetic underpinnings of complex human traits, helping to elucidate the heritability of susceptibility to both chronic diseases and infectious diseases—including COVID-19<sup>2</sup>. Such complex diseases typically have multifactorial etiologies, with contributions from germline genetic variation and environmental exposures<sup>3</sup>, and they frequently present with related, comorbid medical conditions<sup>4</sup>.

SARS-CoV-2 is a highly contagious respiratory virus that also impacts additional organ systems. Several comorbid risk factors are known to be associated with developing severe COVID-19 illness<sup>5</sup>, including hypertension, cardiovascular disease, chronic obstructive pulmonary disease (COPD), high body mass index (BMI),

<sup>1</sup>Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX, USA. <sup>2</sup>Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, USA. <sup>3</sup>Duke Cancer Institute, Duke University Medical Center, Durham, NC, USA. <sup>4</sup>Department of Epidemiology and Population Health, School of Medicine, Stanford University, Stanford, CA, USA. <sup>5</sup>Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA. <sup>6</sup>These authors contributed equally: Jinyoung Byun, Younghun Han, and Kyle M. Walsh. ✉email: jinyoung.byun@bcm.edu; chrisa@bcm.edu

Strata	Trait	Sample size	SNPs
A2 very severe, respiratory-confirmed COVID-19 patients versus population-based controls	COVID-19 A2	707,407	1,140,193
B2 hospitalized COVID-19 patients versus population-based controls	COVID-19 B2	1,206,629	1,141,302
<b>Exclusion of smoking-associated genomic regions</b>			
A2 very severe, respiratory-confirmed COVID-19 patients versus population-based controls	COVID-19 A2 $\perp$ Smoke	707,407	1,001,866
B2 hospitalized COVID-19 patients versus population-based controls	COVID-19 B2 $\perp$ Smoke	1,206,629	1,003,099
<b>Exclusion of BMI-associated genomic regions</b>			
A2 very severe, respiratory-confirmed COVID-19 patients versus population-based controls	COVID-19 A2 $\perp$ BMI	707,407	856,012
B2 hospitalized COVID-19 patients versus population-based controls	COVID-19 B2 $\perp$ BMI	1,206,629	856,864

**Table 1.** Study description.

and type 2 diabetes<sup>6</sup>. Diabetes, associated with immunosuppression and vascular and renal complications, has emerged as a critical comorbidity among severely ill COVID-19 patients<sup>7</sup>.

Leveraging GWAS data, we conducted cross-trait genetic correlation analyses to examine the sum effect of pleiotropy across all causal loci to determine if we could reveal shared genetic correlations between multiple polygenic traits. In addition, we determined the directionality of these associations, and whether the genetic architecture of two traits are correlated or anti-correlated<sup>8</sup>. Examining the shared genetic architecture underlying co-occurrence of disease can help elucidate their common genetic etiology<sup>4</sup>, but there has been no comprehensive evaluation of the shared genetic architecture between COVID-19 disease severity and additional diseases or traits. Previous analyses of lung cancer risk reveal shared genetic architecture with emphysema, a common co-morbid condition, as well as with cigarette consumption, the leading cause of lung cancer<sup>4</sup>. In the setting of COVID-19 disease severity, such an evaluation has similar potential to identify co-morbid conditions and to uncover traits that are causally involved in worsening clinical course.

To identify polygenic traits sharing underlying genetic etiology with COVID-19 disease severity, we utilized summary statistics from prior GWAS and conducted cross-trait linkage disequilibrium (LD) score regression (LDSR) analyses<sup>9,10</sup>. Results provide fundamental knowledge on traits and conditions that share genetic underpinnings with COVID-19 disease severity, reveal potential risk factors for developing severe COVID-19 disease subsequent to SARS-CoV-2 infection, and implicate several modifiable factors that merit further study and may ultimately help improve patient outcomes.

## Methods

**GWAS summary statistics for COVID-19 critical illness and hospitalization.** We downloaded the GWAS summary statistics (COVID19-hg GWAS meta-analyses round 5, released on January 18, 2021; <https://www.covid19hg.org/results/r5/>) from the COVID-19 Host Genetics Initiative (COVID-19 HGI)<sup>11–13</sup>, comprising (1) A2 (critical illness)<sup>13</sup>: 4,606 very severe, respiratory-confirmed COVID-19 patients versus 702,801 population-based controls (A2\_ALL\_eur\_leave\_23andme) and (2) B2 (hospitalization)<sup>13</sup>: 9,373 hospitalized COVID-19 patients versus 1,197,256 population-based controls (B2\_ALL\_eur\_leave\_23andme) (Table 1 and Supplementary Tables 1 and 2). While the summary statistics of COVID-19-hg GWAS meta-analyses across multiple populations have been deposited at the COVID-19 HGI<sup>11,13</sup>, we restricted analyses to European-ancestry subjects (to align with the ancestral background of participants in GWAS of traits used in our downstream LDSR analyses) and did not include the 23andMe cohort (due to the data-use constraint, which makes only the top 10,000 SNPs publicly available)<sup>14</sup>. All methods were performed in accordance with the relevant guidelines and regulations.

**GWAS summary statistics for additional traits.** To estimate cross-trait genetic correlation patterns between COVID-19 disease severity and multiple polygenic traits, we harmonized publicly available GWAS summary-level data from the UK Biobank (UKBB), a prospective population-based cohort study consisting of ~500,000 individuals, aged 40–69 years, who were recruited in the United Kingdom between 2006 and 2010<sup>15,16</sup>. All methods were carried out in accordance with relevant guidelines and regulations.

GWAS summary-level data used for the LDSR analyses of UKBB traits are from publicly-posted results generated by the Neale lab (<http://www.nealelab.is/uk-biobank/>). These association analyses are adjusted with the first 20 principal components, which adjust for sources of population level variability in genetic allele frequencies. The GWAS summary-level data of UKBB used in our study are restricted to “British ancestry” using the first 6 principal components to determine “British ancestry” and further filtered by self-reported ethnicity with “white-British”, “Irish”, or “White”. The sample sizes and more details for the tested traits are shown in Supplementary Table 2.

**Estimating SNP-heritability and cross-trait genetic correlation of COVID-19.** LD score regression analysis with 1000 Genomes Project European (EUR) samples as a reference for pattern of genome-wide LD quantifies the co-heritability of diverse traits<sup>4,9,10,17</sup> using GWAS summary statistics for common genetic variants (i.e., SNPs). In brief, LDSR method regresses  $\chi^2$  statistics from GWAS on LD scores, allowing the estimation

of genetic correlation without bias due to population stratification or cryptic relatedness<sup>4,9,10,18,19</sup>. By regressing SNP-level associations for two traits, (i.e., the product of Z scores,  $Z_{\text{COVID19\_A2}} \times Z_{\text{UKBB\_BMI}}$ ) and weighting each SNP by its LD Score (an estimate of the amount of total genetic variation tagged by each variant), one can estimate the magnitude and direction of shared genomic architecture between these traits. To control the multiple testing burden, we restricted analyses to the tested UKBB traits showing heritability  $\geq 1\%$  and for which prior studies have suggested correlations between COVID-19 and risk for severe outcomes, or traits that were correlated with traits that have been associated with severe outcomes. We conservatively set the test-wise level of significance after Bonferroni correction to be  $0.05/(6 \times 64)$ , adjusting for analysis of COVID-19 severity (A2 and B2) with 64 UKBB traits, with and without removal of BMI and Smoking SNPs. We first implemented the command option of LD Score (<https://github.com/bulik/ldsc>; ldsc v1.0.1) with “munge\_sumstats.py” to generate the “sumstats” format from the GWAS summary statistics after  $\sim 1.14$  M HapMap3 SNPs with MAF  $> 1\%$  were selected for the analysis as recommended. Multi-allelic SNPs and the major histocompatibility complex (MHC) region (Chr6:25–34 Mb) were excluded from summary statistics because of the complex and unusual LD pattern and genetic architecture of the MHC region<sup>4</sup>. We then applied “ldsc.py -rg covid19.A2.sumstats.gz, trait1.sumstats.gz-ref-ld-chr eur\_w\_ld\_chr/-w-ld-chr eur\_w\_ld\_chr/-out covid19.A2.triat1”.

**Exclusion of genomic regions related to smoking behavior and BMI.** Although a clearer picture is emerging, the contribution of cigarette smoking to COVID-19 disease severity remains incompletely understood, with most studies suggesting increased disease severity among former smokers versus never-smokers, but some studies observing a protective effect for current smoking<sup>20</sup> and others showing an increased risk for more severe symptoms in smokers<sup>21</sup>. Since smoking behaviors are heritable traits that correlate with many other complex diseases, we performed sensitivity analyses by excluding chromosomal regions ( $\pm 500$  kb) around 473 SNPs previously associated with various smoking behaviors ( $\perp$ Smoke) to attenuate the genetic contribution of smoking-related variants<sup>4</sup>. The removed genomic regions related to cigarettes per day, smoking initiation, smoking cessation, initiation age of regular smoking, and nicotine dependence (Supplementary Tables 3 and 4).

Although obesity increases risk of systemic inflammation, pulmonary clots, stroke, and myocardial infarction, it remains unclear whether reported associations between BMI and COVID-19 disease severity are confounded by socioeconomic status or concurrent health issues. We performed sensitivity analyses by excluding genomic regions ( $\pm 500$  kb) around 941 SNPs previously associated with BMI ( $\perp$ BMI) to attenuate the genetic contribution of BMI-related variants (Supplementary Tables 4 and 5).

## Results

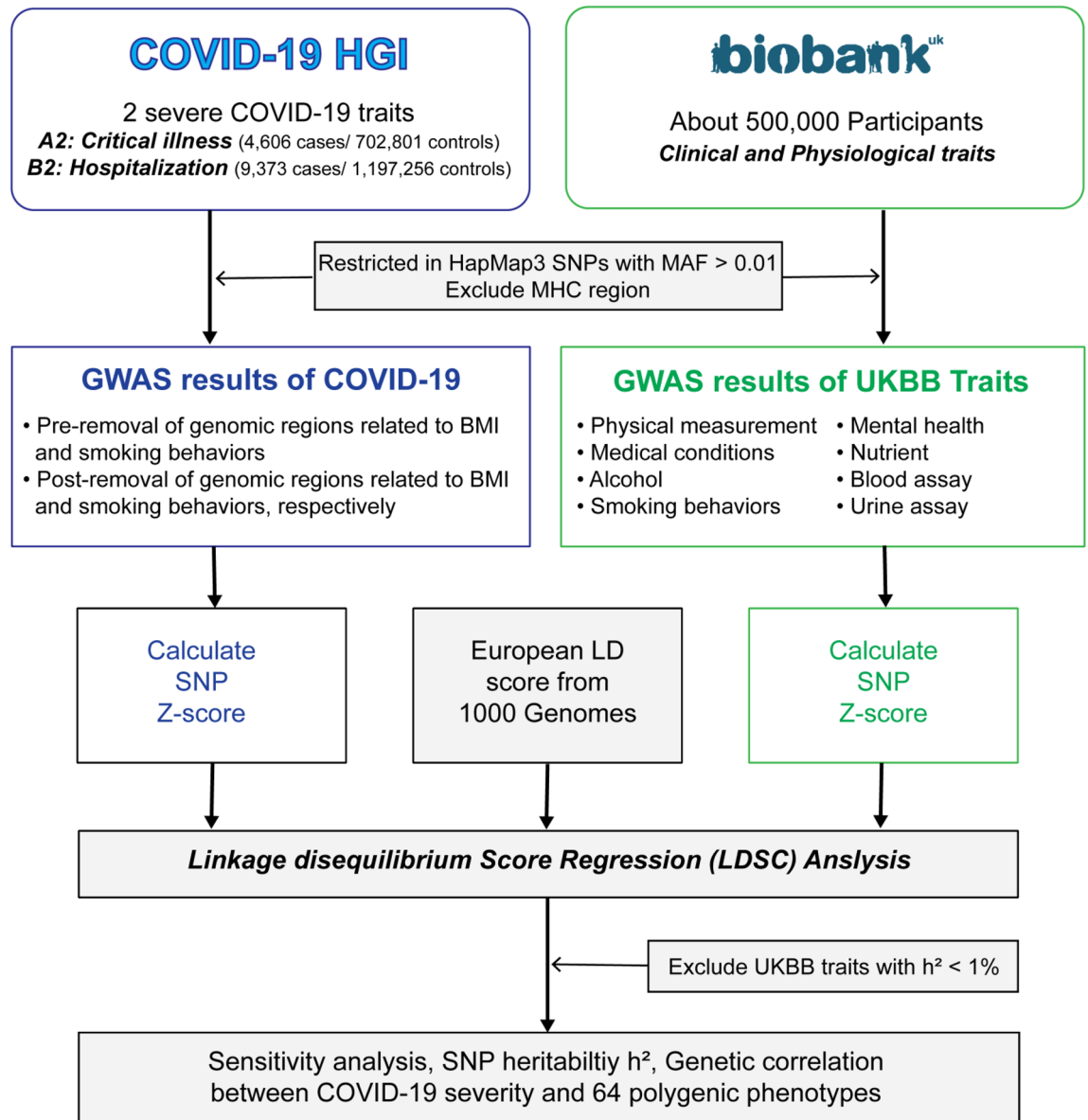
We implemented cross-trait LDSR analysis to examine shared genetic contributions to COVID-19 disease severity and multiple clinical and epidemiologic traits using pairwise genetic correlations (rg) and the observed-scale heritability ( $h^2$ , representing the proportion of phenotypic variance explained by all common SNPs). The flow chart presented in Fig. 1 summarizes the steps from data preparation to LDSR analysis for COVID-19 severity versus 64 polygenic traits we studied. A prior GWAS analysis<sup>13</sup> of very severe, respiratory-confirmed COVID-19 (phenotype A2: critical illness; 4606 cases, 702,801 controls in only European descent) identified loci on chromosomes 3, 12, 17, 19, and 21 that reached genome-wide statistical significance ( $P < 5.0 \times 10^{-8}$  shown in the red horizontal line), with a genomic inflation factor of 1.047, and an estimated  $h^2$  of 0.35%. Sensitivity analysis excluding chromosomal regions known to be associated with smoking reduced the genomic inflation factor to 1.041 and  $h^2$  to 0.34%. Sensitivity analysis excluding chromosomal regions known to be associated with BMI increased the genomic inflation factor to 1.050 and  $h^2$  to 0.35% (Fig. 2, Supplementary Table 6).

A prior GWAS analysis of hospitalized COVID-19 (phenotype B2: hospitalization; 9373 cases, 1,197,256 controls in only European-descent) identified loci on chromosomes 3, 12, 19, and 21 at genome-wide statistical significance, with a genomic inflation factor of 1.041, and an estimated  $h^2$  of 0.19% (Fig. 2). Sensitivity analysis excluding chromosomal regions known to be associated with smoking reduced the genomic inflation factor to 1.038 and  $h^2$  to 0.19%. Sensitivity analysis excluding chromosomal regions known to be associated with BMI reduced the genomic inflation factor to 1.035 and  $h^2$  to 0.17% (Fig. 2, Supplementary Table 6).

Using these GWAS results, we next performed LDSR analyses with two phenotypes for COVID-19 severity (COVID-19 A2, and COVID-19 B2) considering four phenotypes for exclusions of genomic regions related to BMI and Smoking (COVID19\_A2 $\perp$ BMI, COVID19\_A2 $\perp$ Smoke, COVID19\_B2 $\perp$ BMI, and COVID19\_B2 $\perp$ Smoke) and 64 UKBB polygenic traits that had SNP array-based heritability ( $h^2$ )  $\geq 1\%$  (to maximize study power and to provide reliable inferences). Twenty-three diverse traits showed moderate to strong co-heritability with COVID-19 disease severity (Table 2, Supplementary Table 7), including several at Bonferroni-corrected significance level ( $P < 1.30 \times 10^{-4}$ ). Very severe, respiratory-confirmed COVID-19 illness (A2) and COVID-19 hospitalization (B2) showed strong genomic correlation with traits related to adiposity, diabetes, digestive diseases, smoking behaviors, hematologic traits, and selected nutrient levels (Fig. 3, Supplementary Table 7).

Among physical traits, the genetic architecture of COVID-19 disease severity was positively correlated with BMI ( $rg_{\text{COVID19\_A2}} = 0.20$ ,  $P_{\text{COVID19\_A2}} = 1.51 \times 10^{-5}$ ;  $rg_{\text{COVID19\_B2}} = 0.34$ ,  $P_{\text{COVID19\_B2}} = 1.99 \times 10^{-8}$ ), weight ( $rg_{\text{COVID19\_A2}} = 0.17$ ,  $P_{\text{COVID19\_A2}} = 1.24 \times 10^{-4}$ ;  $rg_{\text{COVID19\_B2}} = 0.27$ ,  $P_{\text{COVID19\_B2}} = 7.23 \times 10^{-7}$ ), and whole body fat mass ( $rg_{\text{COVID19\_A2}} = 0.20$ ,  $P_{\text{COVID19\_A2}} = 7.81 \times 10^{-6}$ ;  $rg_{\text{COVID19\_B2}} = 0.33$ ,  $P_{\text{COVID19\_B2}} = 2.24 \times 10^{-8}$ ). After excluding genomic regions previously associated with BMI, both BMI and whole body fat mass continued to show strongly significant positive correlation with COVID-19 disease severity ( $rg_{\text{COVID19\_A2}\perp\text{BMI}} = 0.17$  and  $P_{\text{COVID19\_A2}\perp\text{BMI}} = 2.37 \times 10^{-3}$ ;  $rg_{\text{COVID19\_B2}\perp\text{BMI}} = 0.28$  and  $P_{\text{COVID19\_B2}\perp\text{BMI}} = 1.82 \times 10^{-5}$ ).

Among medical conditions, the genetic architecture of COVID-19 disease severity was positively correlated with shortness of breath walking on level ground ( $rg_{\text{COVID19\_A2}} = 0.28$ ,  $P_{\text{COVID19\_A2}} = 2.87 \times 10^{-3}$ ;  $rg_{\text{COVID19\_B2}} = 0.43$ ,  $P_{\text{COVID19\_B2}} = 4.56 \times 10^{-5}$ ), diabetes ( $rg_{\text{COVID19\_A2}} = 0.54$ ,  $P_{\text{COVID19\_A2}} = 7.10 \times 10^{-4}$ ;

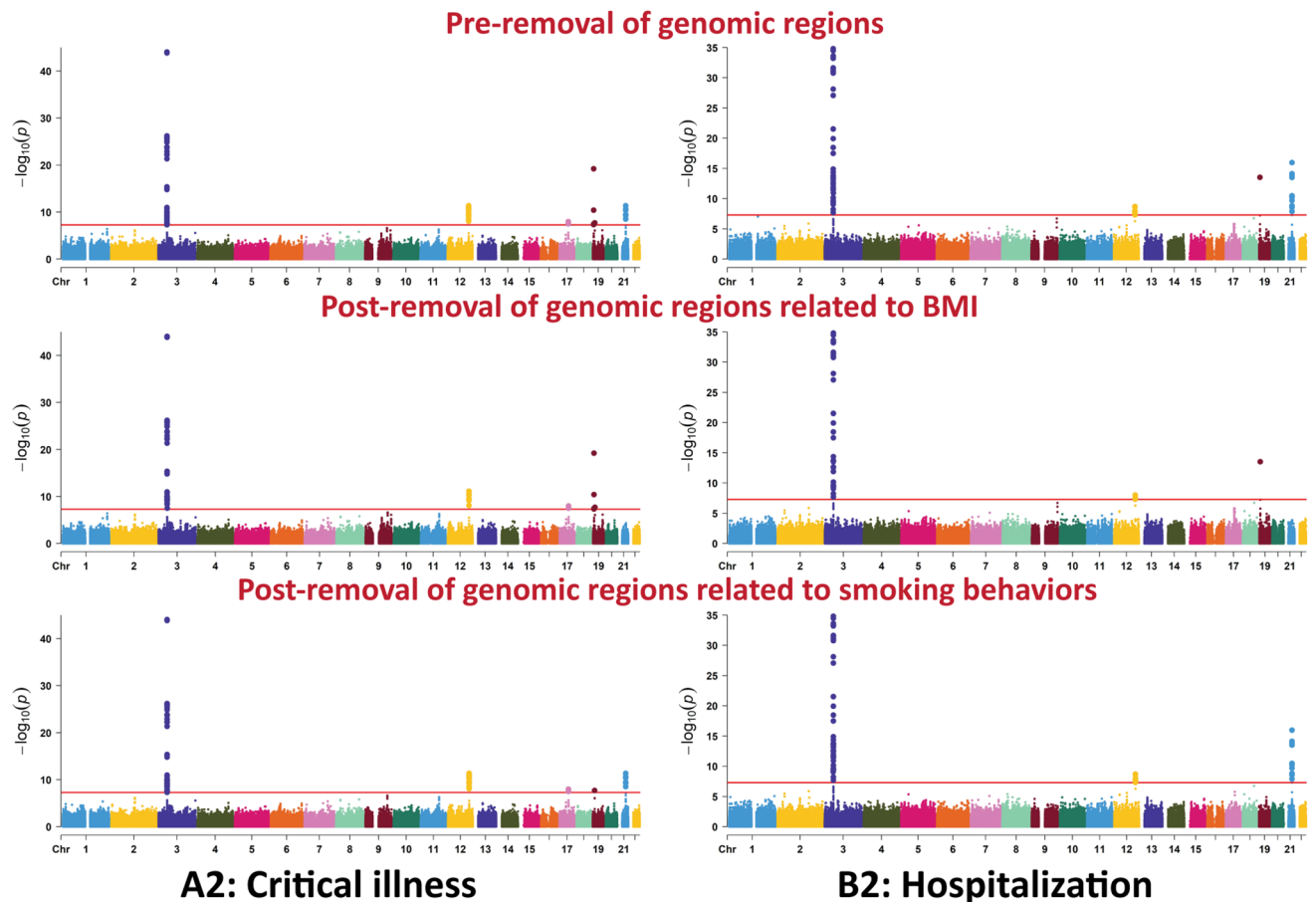


**Figure 1.** Flow chart of the analytical workflow in the study.

$rg_{\text{COVID19\_B2}} = 0.31$ ,  $P_{\text{COVID19\_B2}} = 3.39 \times 10^{-5}$ ), diverticulosis ( $rg_{\text{COVID19\_A2}} = 0.30$ ,  $P_{\text{COVID19\_A2}} = 4.31 \times 10^{-4}$ ;  $rg_{\text{COVID19\_B2}} = 0.38$ ,  $P_{\text{COVID19\_B2}} = 1.66 \times 10^{-5}$ ), diseases of the digestive system ( $rg_{\text{COVID19\_A2}} = 0.20$ ,  $P_{\text{COVID19\_A2}} = 3.48 \times 10^{-3}$ ;  $rg_{\text{COVID19\_B2}} = 0.45$ ,  $P_{\text{COVID19\_B2}} = 1.86 \times 10^{-7}$ ), and diseases of the musculoskeletal system and connective tissue ( $rg_{\text{COVID19\_A2}} = 0.24$ ,  $P_{\text{COVID19\_A2}} = 4.84 \times 10^{-4}$ ;  $rg_{\text{COVID19\_B2}} = 0.34$ ,  $P_{\text{COVID19\_B2}} = 3.54 \times 10^{-6}$ ). Excluding genomic regions associated with smoking behaviors or BMI generally attenuated these correlations, although most remained nominally associated at  $P < 0.05$  and the association between COVID-19 hospitalization and diseases of the digestive system remained significant at Bonferroni-corrected levels after exclusion of smoking-associated loci ( $rg_{\text{COVID19\_B2LSmoke}} = 0.38$ ,  $P_{\text{COVID19\_B2LSmoke}} = 2.30 \times 10^{-5}$ ).

Among smoking behaviors, current tobacco smoking ( $rg_{\text{COVID19\_B2}} = 0.34$ ,  $P_{\text{COVID19\_B2}} = 2.01 \times 10^{-6}$ ) and exposure to tobacco smoke at home ( $rg_{\text{COVID19\_B2}} = 0.47$ ,  $P_{\text{COVID19\_B2}} = 1.73 \times 10^{-5}$ ) presented strongly significant positive genetic correlation with COVID-19 hospitalization, which were only modestly attenuated when removing known smoking-associated loci from analysis. Current tobacco smoking was more modestly associated with severe, respiratory-confirmed COVID-19 illness COVID ( $rg_{\text{COVID19\_A2}} = 0.13$ ,  $P_{\text{COVID19\_B2}} = 0.021$ ) and this association became non-significant after removing smoking-associated loci from analysis (Table 2), supporting a link between known smoking risk loci and risk for severe COVID-19 outcomes.

Examining hematologic traits, both high light scatter reticulocyte percentage and count were significantly positively correlated with COVID-19 hospitalization, as was immature reticulocyte fraction. These traits were also positively correlated with severe COVID-19 illness, but not at Bonferroni-corrected levels of statistical significance. C reactive protein levels were also positively correlated with COVID-19 disease severity. Interestingly, serum (not urinary) albumin was negatively correlated with COVID-19 disease severity at nominal statistical



**Figure 2.** Manhattan plots of the COVID-19 GWAS meta-analysis for pre- and post-removal of genomic regions associated with smoking behaviors and BMI in European descent population. A2: very severe respiratory confirmed COVID-19 cases versus population: 4606 cases and 702,801 controls, B2: hospitalized COVID-19 cases versus population: 9373 cases and 1,197,256 controls.

significance ( $r_{g_{\text{COVID19\_A2}}} = -0.12$ ,  $P_{\text{COVID19\_A2}} = 0.026$ ;  $r_{g_{\text{COVID19\_B2}}} = -0.16$ ,  $P_{\text{COVID19\_B2}} = 0.011$ ), as were HDL, apolipoprotein A levels, and levels of serum IGF-1 (Table 2).

We also examined the pairwise genetic relationship between COVID-19 disease severity and nutrient-related traits in UKB. Although we were not able to observe any significant associations between COVID-19 critical illness and hospitalization and nutrient-related traits at Bonferroni-corrected levels, we identified suggestive negative correlations with magnesium ( $r_{g_{\text{COVID19\_A2}}} = -0.39$ ,  $P_{\text{COVID19\_A2}} = 2.28 \times 10^{-3}$ ;  $r_{g_{\text{COVID19\_B2}}} = -0.36$ ,  $P_{\text{COVID19\_B2}} = 5.17 \times 10^{-3}$ ), retinol ( $r_{g_{\text{COVID19\_A2}}} = -0.59$ ,  $P_{\text{COVID19\_A2}} = 0.041$ ;  $r_{g_{\text{COVID19\_B2}}} = -0.59$ ,  $P_{\text{COVID19\_B2}} = 0.029$ ), and vitamin E ( $r_{g_{\text{COVID19\_A2}}} = -0.53$ ,  $P_{\text{COVID19\_A2}} = 2.16 \times 10^{-3}$ ;  $r_{g_{\text{COVID19\_B2}}} = -0.53$ ,  $P_{\text{COVID19\_B2}} = 3.10 \times 10^{-3}$ ) (Table 2 and Supplementary Table 7). Vitamin D levels were not associated with risk for severe COVID-19 ( $r_{g_{\text{COVID19\_A2}}} = -0.023$ ,  $P_{\text{COVID19\_A2}} = 0.67$ ;  $r_{g_{\text{COVID19\_B2}}} = -0.043$ ,  $P_{\text{COVID19\_B2}} = 0.44$ ).

## Discussion

We investigated the genetic correlations between COVID-19 disease severity (A2:critical illness and B2:hospitalization) with a variety of clinical and physiologic traits using summary-level GWAS data from extremely large patient cohorts, observing shared genomic architecture with a number of illnesses and biomarkers of somatic well-being. We identify a suite of medical conditions and physiological traits that appear to share the genetic architecture with that of COVID-19 severity. Many of these traits overlap those previously identified in the large databases of COVID-19 patient outcomes, including traits related to adiposity, kidney function, and pulmonary insufficiency. We also identified additional traits that have received comparatively little attention, such as blood and serum levels of several vitamins and nutrients. Although our datasets are quite large (COVID-19 severity GWAS  $n = 707,407$  and  $1,206,629$  for critical illness (A2) and hospitalization (B2), respectively; UKBB GWAS  $n = 361,194$ ), larger datasets would likely identify many of these same associations and could potentially bring some of the nominally associated associations to a corrected level of statistical significance.

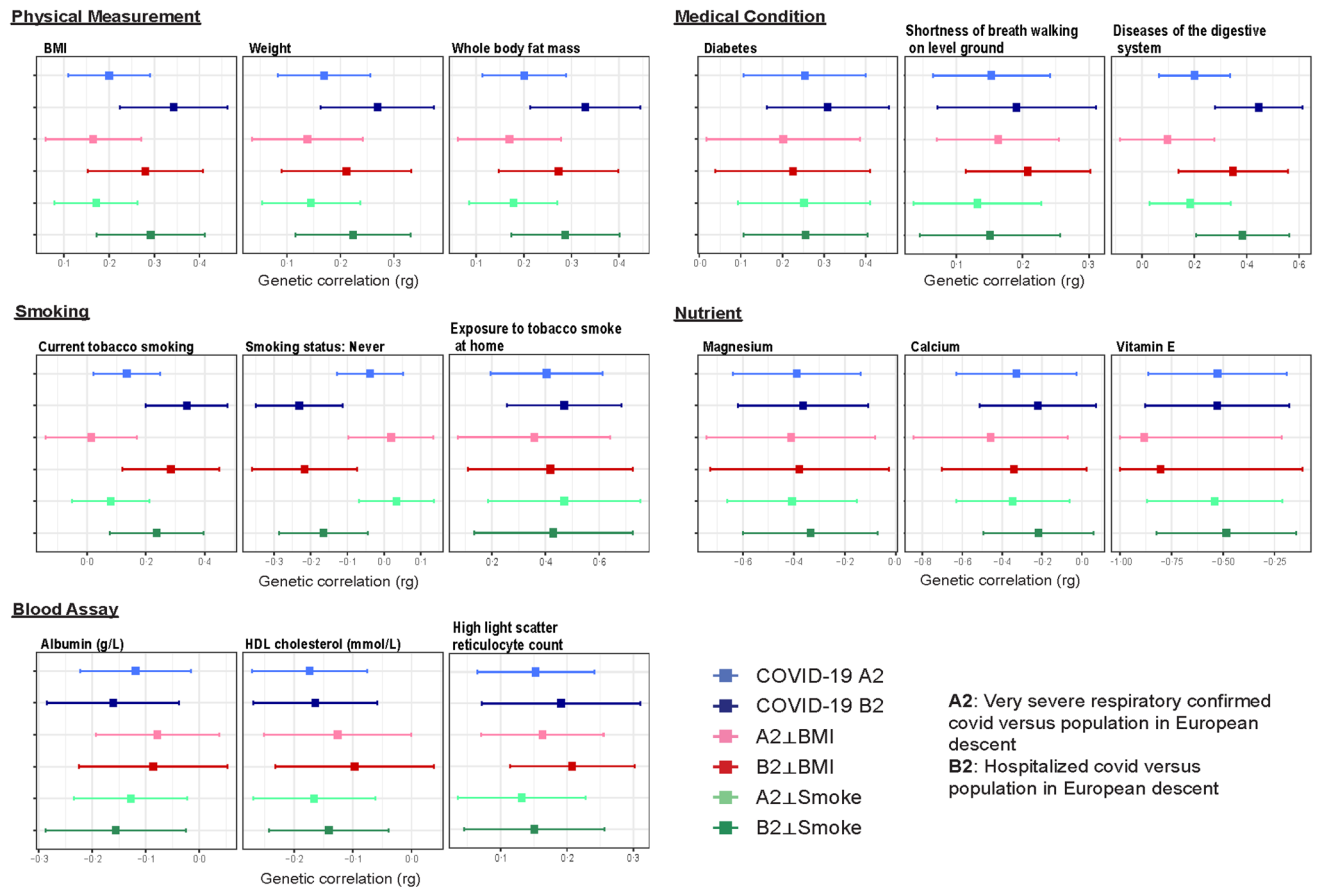
Using an orthogonal genomics-driven approach that complements previous COVID-19 clinical epidemiology research, we confirm a link between the development of severe COVID-19 illness and both elevated BMI and diabetes. We also clarify associations with current smoking status, observing that it was positively correlated with COVID-19 disease severity, and note new associations with diverticulosis and reticulocyte traits. Additionally,

Traits	COVID-19 critical illness (A2)						COVID-19 hospitalization (B2)					
	A2		A2⊥BMI		A2⊥Smoke		B2		B2⊥BMI		B2⊥Smoke	
	rg	<i>P</i>	rg	<i>P</i>	rg	<i>P</i>	rg	<i>P</i>	rg	<i>P</i>	rg	<i>P</i>
<b>Physical measurement</b>												
BMI	0.200	<b>1.51 × 10<sup>-5</sup></b>	0.165	2.37 × 10 <sup>-3</sup>	0.171	3.07 × 10 <sup>-4</sup>	0.343	<b>1.99 × 10<sup>-8</sup></b>	0.280	<b>1.82 × 10<sup>-5</sup></b>	0.292	<b>1.42 × 10<sup>-6</sup></b>
Weight	0.169	<b>1.24 × 10<sup>-4</sup></b>	0.138	9.60 × 10 <sup>-3</sup>	0.145	2.24 × 10 <sup>-3</sup>	0.269	<b>7.23 × 10<sup>-7</sup></b>	0.211	6.86 × 10 <sup>-4</sup>	0.223	<b>4.23 × 10<sup>-5</sup></b>
Whole body fat mass	0.201	<b>7.81 × 10<sup>-6</sup></b>	0.170	2.00 × 10 <sup>-3</sup>	0.178	1.49 × 10 <sup>-4</sup>	0.329	<b>2.24 × 10<sup>-8</sup></b>	0.273	<b>2.32 × 10<sup>-5</sup></b>	0.287	<b>8.51 × 10<sup>-7</sup></b>
<b>Medical condition</b>												
Shortness of breath walking on level ground	0.283	2.87 × 10 <sup>-3</sup>	0.198	0.107	0.272	8.33 × 10 <sup>-3</sup>	0.433	<b>4.56 × 10<sup>-5</sup></b>	0.372	5.94 × 10 <sup>-3</sup>	0.395	4.01 × 10 <sup>-4</sup>
Diabetes diagnosed by doctor	0.254	7.10 × 10 <sup>-4</sup>	0.202	0.032	0.252	1.95 × 10 <sup>-3</sup>	0.309	<b>3.39 × 10<sup>-5</sup></b>	0.225	0.018	0.256	8.00 × 10 <sup>-4</sup>
Diagnoses—main ICD10: K57 Diverticular disease of intestine	0.297	4.31 × 10 <sup>-4</sup>	0.228	0.019	0.315	6.64 × 10 <sup>-4</sup>	0.380	<b>1.66 × 10<sup>-5</sup></b>	0.305	2.84 × 10 <sup>-3</sup>	0.355	2.01 × 10 <sup>-4</sup>
Diseases of the digestive system	0.201	3.48 × 10 <sup>-3</sup>	0.097	0.292	0.183	0.020	0.446	<b>1.86 × 10<sup>-7</sup></b>	0.348	1.14 × 10 <sup>-3</sup>	0.384	<b>2.30 × 10<sup>-5</sup></b>
Diseases of the musculoskeletal system and connective tissue	0.243	4.84 × 10 <sup>-4</sup>	0.121	0.170	0.223	5.96 × 10 <sup>-3</sup>	0.338	<b>3.54 × 10<sup>-6</sup></b>	0.174	0.038	0.272	4.62 × 10 <sup>-4</sup>
<b>Smoking</b>												
Current tobacco smoking	0.135	0.021	0.014	0.858	0.081	0.223	0.339	<b>2.01 × 10<sup>-6</sup></b>	0.285	6.57 × 10 <sup>-4</sup>	0.237	3.33 × 10 <sup>-3</sup>
Exposure to tobacco smoke at home	0.404	1.67 × 10 <sup>-4</sup>	0.358	0.013	0.470	1.16 × 10 <sup>-3</sup>	0.470	<b>1.73 × 10<sup>-5</sup></b>	0.418	7.79 × 10 <sup>-3</sup>	0.429	4.62 × 10 <sup>-3</sup>
<b>Nutrient</b>												
Magnesium	-0.389	2.28 × 10 <sup>-3</sup>	-0.411	0.015	-0.407	1.81 × 10 <sup>-3</sup>	-0.364	5.17 × 10 <sup>-3</sup>	-0.378	0.035	-0.335	0.013
Calcium	-0.329	0.033	-0.457	0.020	-0.347	0.017	-0.222	0.137	-0.340	0.067	-0.219	0.121
Retinol	-0.591	0.041	-0.888	0.027	-0.484	0.081	-0.588	0.029	-1.00	0.018	-0.473	0.076
Vitamin E	-0.527	2.16 × 10 <sup>-3</sup>	-0.883	0.010	-0.540	1.26 × 10 <sup>-3</sup>	-0.528	3.10 × 10 <sup>-3</sup>	-0.804	0.022	-0.484	5.08 × 10 <sup>-3</sup>
<b>Blood Assay</b>												
Albumin (g/L)	-0.119	0.026	-0.071	0.229	-0.128	0.017	-0.161	0.011	-0.086	0.225	-0.156	0.020
Apolipoprotein A (g/L)	-0.134	6.17 × 10 <sup>-3</sup>	-0.093	0.138	-0.126	0.016	-0.118	0.024	-0.061	0.369	-0.101	0.047
C-reactive protein (mg/L)	0.187	1.88 × 10 <sup>-4</sup>	0.139	0.015	0.218	1.50 × 10 <sup>-4</sup>	0.278	3.04 × 10 <sup>-4</sup>	0.237	2.79 × 10 <sup>-4</sup>	0.294	<b>2.98 × 10<sup>-5</sup></b>
HDL cholesterol	-0.174	5.20 × 10 <sup>-4</sup>	-0.126	0.048	-0.166	1.60 × 10 <sup>-3</sup>	-0.164	2.25 × 10 <sup>-3</sup>	-0.097	0.161	-0.141	6.93 × 10 <sup>-3</sup>
High light scatter reticulocyte count	0.153	7.13 × 10 <sup>-4</sup>	0.163	5.33 × 10 <sup>-4</sup>	0.132	6.74 × 10 <sup>-3</sup>	0.208	<b>1.23 × 10<sup>-5</sup></b>	0.191	1.83 × 10 <sup>-3</sup>	0.151	4.73 × 10 <sup>-3</sup>
High light scatter reticulocyte percentage	0.155	4.13 × 10 <sup>-4</sup>	0.168	1.78 × 10 <sup>-4</sup>	0.131	5.68 × 10 <sup>-3</sup>	0.209	<b>7.66 × 10<sup>-6</sup></b>	0.198	8.04 × 10 <sup>-4</sup>	0.150	5.12 × 10 <sup>-3</sup>
IGF-1 (nmol/L)	-0.148	4.99 × 10 <sup>-3</sup>	-0.155	0.017	-0.153	5.64 × 10 <sup>-3</sup>	-0.090	0.077	-0.116	0.064	-0.083	0.118
Immature reticulocyte fraction	0.173	1.58 × 10 <sup>-3</sup>	0.156	0.010	0.155	6.05 × 10 <sup>-3</sup>	0.264	<b>1.05 × 10<sup>-5</sup></b>	0.219	3.92 × 10 <sup>-3</sup>	0.199	3.59 × 10 <sup>-3</sup>
Platelet distribution width	-0.108	0.017	-0.130	0.011	-0.109	0.020	-0.038	0.440	-0.060	0.256	-0.038	0.432

**Table 2.** Cross-trait genetic correlations of COVID-19 on inclusion/exclusion of genomic regions associated with BMI and smoking. *P*-values in bold indicates  $P \leq 1.30 \times 10^{-4}$ . COVID19\_A2, very severe respiratory confirmed covid versus population including whole genomic regions; A2⊥BMI, very severe respiratory confirmed covid versus population with exclusion of genomic regions related to BMI; A2⊥Smoke, very severe respiratory confirmed covid versus population with exclusion of genomic regions related to smoking behaviors; COVID19\_B2, hospitalized covid versus population including whole genomic regions; B2⊥BMI, hospitalized covid versus population with exclusion of genomic regions related to BMI; B2⊥Smoke, hospitalized covid versus population with exclusion of genomic regions related to smoking behaviors.

we observe a suggestive association between increased disease severity and reduced levels of IGF-1—a marker of nutritional status—and additional suggestive protective associations with magnesium, retinol, and vitamin E levels.

COVID-19 is primarily a respiratory illness. We observed that higher forced vital capacity (FVC) was negatively (protectively) associated with COVID-19 disease severity and observed a strongly positive correlation between the genetic architecture of ‘shortness of breath while walking on level ground’ and development of severe COVID-19 illness. Chest pain and discomfort have previously been associated with COVID-19 hospitalization and the U.S. Centers for Disease Control and Prevention (CDC) announced that individuals with chronic lung diseases including emphysema, chronic bronchitis, COPD, and interstitial lung disease are at high risk for becoming critically ill from SARS-CoV-2<sup>1</sup>. Our study demonstrates a positive correlation between the genetic architecture of these risk factors and COVID-19 disease severity through LDSR analyses. In this study, a differential diagnosis of COPD was strongly positively correlated with COVID-19 hospitalization, regardless of the exclusion of genomic regions related to BMI and smoking behaviors. Since chronic inflammation is an important feature in developing both emphysema and bronchitis, these finding suggest a potential shared genetic contribution between COPD and COVID-19 hospitalization separate from the contributions of known BMI and



**Figure 3.** The pairwise genetic correlation of COVID-19 disease severity and selected traits.

smoking-related variants. Variants located in immune-related genes and contributing to increased pulmonary inflammation could be evaluated in future work.

Traits related to smoking behaviors were generally associated with increased COVID-19 disease severity in our analyses, including current smoking, exposure to tobacco smoke either at home or outside home, in utero tobacco smoke exposure, and cumulative pack-years. Conversely, never-smoker status showed negative genomic correlation with COVID-19 disease severity. Although UKBB does not delineate former smokers in ascribing smoking status, our analyses indicate that the genetic determinants of current smoking are associated with increased COVID-19 disease severity and do not support the clinical observations that current smoking may protect against severe COVID-19 illness.

Given the lack of a COVID-19 vaccine during the first year of the pandemic and continued supply scarcity in numerous regions, many studies have sought to identify alternative strategies to minimize risk of developing severe COVID-19 following SARS-CoV-2 infection and also to treat severe COVID-19. In addition to evaluations of existing pharmacologic agents (*e.g.*, ivermectin, hydroxychloroquine, azithromycin, and dexamethasone), vitamin and nutrient supplementation has been widely studied. Global mortality rate differences associated with latitude and clinical observations of low serum 25-hydroxyvitamin D levels among hospitalized COVID-19 patients has perhaps garnered greatest attention<sup>22</sup>, but we did not observe a significant association between genetic determinants of vitamin D levels and COVID-19 severity. However, we observed nominally significant protective effects for less-studied nutrient-related traits, including magnesium, calcium, retinol, and vitamin E. A combined vitamin D/magnesium/vitamin B12 combination was associated with a reduction in the proportion of elderly COVID-19 patients requiring oxygen support and intensive care support in a small prospective cohort<sup>23</sup>, and lower plasma retinol levels have also been observed in hospitalized COVID-19 patients<sup>24</sup>. We did not observe a significant association between serum Vitamin D levels and risk for COVID-19 or severe outcomes. Vitamin E levels have not been widely examined in the context of COVID-19, but deficiency is frequently associated with intestinal malabsorption rather than dietary insufficiency and thus may reinforce the observed genetic correlation between COVID-19 disease severity and diverticular disease in our analyses. In our study, we do observe an association between higher levels. Further, we observe protective associations for both HDL and serum concentration of apolipoprotein A, a major component of the HDL complex involved in clearing fat. HDL is involved in vitamin E absorption and contains approximately 40% of circulating  $\alpha$ -tocopherol, the main dietary source of vitamin E<sup>25</sup>.

Integration and harmonization of extant large-scale GWAS datasets has become a popular approach to reveal novel epidemiologic associations. Still, access to individual-level GWAS datasets remains limited, because of data

use restrictions. The LDSR method does not require individual-level genotype data or LD pruning and can quantify the shared genetic architecture of traits having undergone GWAS analysis. However, LDSR analysis assumes absence of population stratification in the underlying summary statistics used and necessitates incorporation of GWAS data from populations expected to have similar genomic architecture. This assumption restricted our analysis to use of GWAS data from British-ancestry individuals, limiting our ability to make conclusions about the shared genetic architectures among other racial/ethnic groups. Given that COVID-19 disease severity has been associated with racial/ethnic background, as well as socioeconomic status and somatic well-being, it is imperative that efforts be made to enrich future genetic epidemiology studies for participants of non-European descent to expand generalizability of results. Interpretation of our results is also limited by the strong correlation between many of the traits studied with BMI and smoking behaviors. Although we made efforts to limit the impact of BMI and smoking-associated genetic variation by excluding known loci from LDSR analysis, such sensitivity analyses cannot account for polygenic contributions not yet having reached genome-wide statistical significance in prior research.

To estimate cross-trait genetic correlation, we restricted the range of UKBB traits with an arbitrary threshold of  $h^2 \geq 1\%$  to improve reliability. For instance, there were additional subtypes of diabetes derived from the various medical records in UKBB and their estimates of SNP-heritability showed  $h^2_{\text{Type 1 Diabetes}} = 0.3\%$  and  $h^2_{\text{Type 2 Diabetes}} = 0.4\%$ . Therefore, we did not include results from them. The type of diabetes reported in Table 1, described as “Diabetes diagnosed by doctor” (UKBB Field Identifier:2443), is not specified for the type of diabetes, but given the age of participants and the general prevalence of T2D versus T1D, the association between diabetes and COVID-19 severity is ostensibly driven by the shared genetic architecture between COVID-19 severity and the genetic architecture of T2D. Furthermore, LDSR analysis relies on the common genetic variants with  $\text{MAF} > 1\%$  and therefore it can fail to capture the SNP-heritability on the observed scale due to underlying low-frequency or rare variants<sup>4</sup>. If a polygenic trait in UKBB shows a significant genetic correlation with COVID-19 severity, this does not imply a causal association. Both the tested trait and COVID-19 severity risk may be jointly influenced by an unmodeled trait that is independently associated with each<sup>19</sup>. Although our study relies on associations with common genetic variants (generally,  $\text{MAF} > 1\%$ ), the inclusion of additional rare genetic variants might be valuable as it could increase the overall trait heritability being modeled. Inferences from LDSR rely on normality assumptions that may be violated when rare variants are studied, and we therefore restricted analysis to more common variants. Additionally, LDSR does not explicitly model confounding effects which can arise when studying multiple correlated traits. Therefore, the method identifies novel associations that can be further studied using Mendelian Randomization or direct analyses of the nominated trait phenotypes for further confirmation of causal relationships. LDSR is a useful approach for identifying potential novel associations that will warrant further epidemiological analysis to tease apart causal associations from associations that are influenced by confounding.

Our findings support previously identified risk factors for severe COVID-19 illness, including elevated BMI, diabetes, and numerous pulmonary conditions (e.g., COPD, reduced FEV, shortness of breath during mild activity). We also observe protective associations between the genetic underpinnings of COVID-19 severity and that of non-smoking, serum albumin, apolipoprotein A, HDL cholesterol level, and several nutrients. Further studies using Mendelian randomization approaches may help to dissect causal associations between COVID-19 disease severity and these traits, potentially nominating targets for therapeutic intervention.

## Data availability

The datasets supporting the conclusions of this article are publicly available in the COVID-19 HGI website for COVID-19 severity GWAS summary statistics [<https://www.covid19hg.org/results/r5/>] and Neale’s lab repository for UK Biobank GWAS summary statistics [[https://github.com/Nealelab/UK\\_Biobank\\_GWAS](https://github.com/Nealelab/UK_Biobank_GWAS)], and therefore no approvals were required.

Received: 12 October 2021; Accepted: 19 January 2022

Published online: 03 February 2022

## References

1. WHO COVID-19 Dashboard. Geneva: World Health Organization, 2020. Available online: <https://covid19.who.int/> (last cited: June 1, 2021).
2. Severe Covid GG *et al.* Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534. <https://doi.org/10.1056/NEJMoa2020283> (2020).
3. Becker, F. *et al.* Genetic testing and common disorders in a public health framework: How to assess relevance and possibilities. Background document to the ESHG recommendations on genetic testing and common disorders. *Eur. J. Hum. Genet.* **19**(Suppl 1), S6–44. <https://doi.org/10.1038/ejhg.2010.249> (2011).
4. Byun, J. *et al.* The shared genetic architectures between lung cancer and multiple polygenic phenotypes in genome-wide association studies. *Cancer Epidemiol. Biomark. Prev.* **30**, 1156–1164. <https://doi.org/10.1158/1055-9965.EPI-20-1635> (2021).
5. Das, S. *et al.* Role of comorbidities like diabetes on severe acute respiratory syndrome coronavirus-2: A review. *Life Sci* **258**, 118202. <https://doi.org/10.1016/j.lfs.2020.118202> (2020).
6. Karagiannidis, C. *et al.* Case characteristics, resource use, and outcomes of 10 021 patients with COVID-19 admitted to 920 German hospitals: An observational study. *Lancet Respir. Med.* **8**, 853–862. [https://doi.org/10.1016/S2213-2600\(20\)30316-7](https://doi.org/10.1016/S2213-2600(20)30316-7) (2020).
7. Fang, L., Karakiulakis, G. & Roth, M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir. Med.* **8**, e21. [https://doi.org/10.1016/S2213-2600\(20\)30116-8](https://doi.org/10.1016/S2213-2600(20)30116-8) (2020).
8. van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H. & Wray, N. R. Genetic correlations of polygenic disease traits: From theory to practice. *Nat. Rev. Genet.* **20**, 567–581. <https://doi.org/10.1038/s41576-019-0137-z> (2019).
9. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241. <https://doi.org/10.1038/ng.3406> (2015).



10. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295. <https://doi.org/10.1038/ng.3211> (2015).
11. The COVID-19 Host Genetics Initiative. <https://www.covid19hg.org/> (2021).
12. COVID-19 Host Genetics Initiative. The COVID-19 host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* **28**, 715–718. <https://doi.org/10.1038/s41431-020-0636-6> (2020).
13. Covid-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* <https://doi.org/10.1038/s41586-021-03767-x> (2021).
14. Fadista, J. *et al.* Shared genetic etiology between idiopathic pulmonary fibrosis and COVID-19 severity. *EBioMedicine* **65**, 103277. <https://doi.org/10.1016/j.ebiom.2021.103277> (2021).
15. Neale Lab - UK Biobank GWAS. <http://www.nealelab.is/uk-biobank/> (2018).
16. Larsson, S. C. *et al.* Smoking, alcohol consumption, and cancer: A mendelian randomisation study in UK Biobank and international genetic consortia participants. *PLoS Med.* **17**, e1003178. <https://doi.org/10.1371/journal.pmed.1003178> (2020).
17. Duncan, L. E. *et al.* Genetic correlation profile of schizophrenia mirrors epidemiological results and suggests link between polygenic and rare variant (22q11.2) cases of schizophrenia. *Schizophr. Bull.* **44**, 1350–1361. <https://doi.org/10.1093/schbul/sbx174> (2018).
18. Ostrom, Q. T. *et al.* Partitioned glioma heritability shows subtype-specific enrichment in immune cells. *Neuro. Oncol.* <https://doi.org/10.1093/neuonc/noab072> (2021).
19. Pettit, R. W. *et al.* The shared genetic architecture between epidemiological and behavioral traits with lung cancer. *Sci. Rep.* **11**, 17559. <https://doi.org/10.1038/s41598-021-96685-x> (2021).
20. Puebla Neira, D. *et al.* Smoking and risk of COVID-19 hospitalization. *Respir. Med.* **182**, 106414. <https://doi.org/10.1016/j.rmed.2021.106414> (2021).
21. Hopkinson, N. S. *et al.* Current smoking and COVID-19 risk: Results from a population symptom app in over 2.4 million people. *Thorax* **76**, 714–722. <https://doi.org/10.1136/thoraxjnl-2020-216422> (2021).
22. Oscanoa, T. J. *et al.* The relationship between the severity and mortality of SARS-CoV-2 infection and 25-hydroxyvitamin D concentration: A metaanalysis. *Adv. Respir. Med.* **89**, 145–157. <https://doi.org/10.5603/ARM.a2021.0037> (2021).
23. Tan, C. W. *et al.* Cohort study to evaluate the effect of vitamin D, magnesium, and vitamin B12 in combination on progression to severe outcomes in older patients with coronavirus (COVID-19). *Nutrition* **79–80**, 111017. <https://doi.org/10.1016/j.nut.2020.111017> (2020).
24. Tepaspe, P. R. *et al.* Vitamin A plasma levels in COVID-19 patients: A prospective multicenter study and hypothesis. *Nutrients* <https://doi.org/10.3390/nu13072173> (2021).
25. Traber, M. G. Mechanisms for the prevention of vitamin E excess. *J. Lipid Res.* **54**, 2295–2306. <https://doi.org/10.1194/jlr.R032946> (2013).

## Acknowledgements

We thank all individuals who have contributed their samples and clinical data for this study, and we also thank all members in the *COVID-19 Host Genetics Initiative* community for sharing GWAS summary statistics of COVID-19.

## Author contributions

J.B. and Y.H. conceived the study design. J.B. prepared the first draft of this manuscript. Y.H. contributed to the data preparation and performed the analyses. A.S.P. and Y.H. assisted the visualization of the findings. J.B., Y.H., K.M., M.L.B., and C.I.A. contributed to the interpretation of the presented analysis. J.B., Y.H., and K.M.W. wrote the manuscript. J.B. and C.I.A. supervised this study. All authors reviewed and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-05832-5>.

**Correspondence** and requests for materials should be addressed to J.B. or C.I.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022