Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Shared genetic influences between blood analyte levels and risk of severe COVID-19

Graphical abstract

Highlights
- Genetic analyses identify pleiotropy between blood analytes and severe COVID-19
- The identified pleiotropies suggest biological or causal relationships
- Blood triglycerides have widespread shared genetic influences with severe COVID-19
- Higher levels of blood triglycerides might increase risk of severe COVID-19

Authors
Hamzeh M. Tanha, Anita Sathyanarayanan, Divya Mehta, Dale R. Nyholt

Correspondence
hamzeh.mesriantanha@hdr.qut.edu.au (H.M.T.), d.nyholt@qut.edu.au (D.R.N.)

In brief
Genetic factors contribute to severe COVID-19 and blood analyte levels. Tanha et al. identify blood analytes associated with the risk of severe COVID-19. The results suggest a strong genetic relationship between higher levels of blood triglycerides and risk of severe COVID-19.
Shared genetic influences between blood analyte levels and risk of severe COVID-19

Hamzeh M. Tanha,1,2,* Anita Sathyanarayanan,1,2 Divya Mehta,1 and Dale R. Nyholt1,3,*
1School of Biomedical Sciences, Faculty of Health, and Centre for Genomics and Personalised Health, Queensland University of Technology (QUT), 60 Musk Avenue, Kelvin Grove, Brisbane, QLD 4059, Australia
2These authors contributed equally
3Lead contact
*Correspondence: hamzeh.mesriantanha@hdr.qut.edu.au (H.M.T.), d.nyholt@qut.edu.au (D.R.N.)
https://doi.org/10.1016/j.celrep.2022.111708

SUMMARY

Genome-wide association studies (GWASs) show that genetic factors contribute to the risk of severe coronavirus disease 2019 (COVID-19) and blood analyte levels. Here, we utilize GWAS summary statistics to study the shared genetic influences (pleiotropy) between severe COVID-19 and 344 blood analytes at the genome, gene, and single-nucleotide polymorphism (SNP) levels. Our pleiotropy analyses genetically link blood levels of 71 analytes to severe COVID-19 in at least one of the three levels of investigation—suggesting shared biological mechanisms or causal relationships. Six analytes (alanine aminotransferase, alkaline phosphatase, apolipoprotein B, C-reactive protein, triglycerides, and urate) display evidence of pleiotropy with severe COVID-19 at all three levels. Causality analyses indicate that higher triglycerides levels causally increase the risk of severe COVID-19, thereby providing important support for the use of lipid-lowering drugs such as statins and fibrates to prevent severe COVID-19.

INTRODUCTION

Coronavirus disease 2019 (COVID-19) due to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus has affected over 254 million people and caused >5.1 million deaths worldwide as of November 2021.1 A wide variability is observed in the clinical manifestations of the disease, with the majority of cases having no or mild symptoms2 and a small proportion having a severe disease with outcomes such as respiratory compromise, acute respiratory distress syndrome, and multiorgan failure and death.3 This variability in disease severity and outcome is attributed to host factors such as sex, age, comorbidities,4 and genetics.5 Epidemiological studies have shown that pre-existing metabolic syndromes including diabetes,6 obesity,7 hypertension, and cardiovascular disease8 are risk factors for developing severe COVID-19.

Metabolic health is assessed through levels of metabolites and clinical chemistry analytes (collectively referred to as “analytes” henceforth), which are influenced by several factors including genetics,9 age,10 drug use,11 diseases,12 exercise,13 and environmental factors.14 Poor metabolic health can dysregulate the innate and adaptive immune responses.15,16 The metabolic state is emerging as the center stage for gaining insights into the COVID-19 host immune response. A recent study examining 598 lipids and 404 polar metabolites levels identified a panel of ten metabolites that can distinguish between healthy controls and COVID-19 cases.17 Another study investigating the blood proteins and metabolites identified 93 proteins and 204 metabolites that correlated with COVID-19 severity.18 Although such studies may identify relevant analytes and provide insights on the COVID-19 disease state, they are susceptible to confounding effects of the disease, medication, environmental effects, and reverse causation on analyte levels.

Genome-wide association studies (GWASs) have identified single-nucleotide polymorphisms (SNPs) associated with the risk of severe COVID-19 compared with controls.19 Similarly, GWASs of metabolomic profiles have identified SNPs associated with blood metabolite levels.9 Identifying the shared genetic influences (pleiotropy) between the risk of severe COVID-19 and levels of blood analytes could circumvent the effects of confounding factors. Different mechanisms of pleiotropy include “horizontal pleiotropy,” where shared genetic factors contribute to both severe COVID-19 risk and levels of analytes, and “vertical pleiotropy,” where shared genetic factors are due to a causal relationship.20 Therefore, pleiotropy between severe COVID-19 and analytes can help interpret severe COVID-19 GWAS findings via identifying underlying biology and prioritize analytes for screening strategies and drug selection. Here, utilizing GWAS summary statistics data, we examined pleiotropy between severe COVID-19 risk and 344 blood analyte levels, including 316 metabolites and 28 traditional clinical chemistry analytes at the genome, gene, and locus (SNP) levels. Lastly, for the analytes exhibiting pleiotropy with the disease, we investigated if the
RESULTS

We sourced curated GWAS summary statistics data for 316 metabolites from our previous study $^{21}$ ($n_{range} = 6,263–24,925$). Briefly, we collated the GWAS summary statistics of 972 metabolites from eight studies and selected 316 metabolites with statistically significant SNP-based heritability ($h^2_{SNP}$) ($h^2_{SNP}$ Z score > 1.64) for further analyses (Tables S1 and S2). In addition, we downloaded GWAS summary statistics of 30 traditional clinical chemistry analytes from UK Biobank GWAS ($\text{http://www.nealelab.is/uk-biobank/}$) and selected 28 traditional clinical chemistry analytes ($n_{range} = 30,565–344,292$) with statistically significant $h^2_{SNP}$ ($h^2_{SNP}$ Z score > 1.64). We conducted quality control (QC) and curation steps for the clinical chemistry analytes similar to our previous study $^{21}$ In total, we examined the pleiotropy between blood levels of 344 analytes and the risk of severe COVID-19. A summary of the characteristics of the GWAS summary statistics for the 344 analytes is provided in Tables S1, S2, and S3.

We downloaded the GWAS summary statistics of severe COVID-19 provided by The Host Genetics Initiative $^{6}$ ($n_{case} = 5,101$; $n_{control} = 1,383,241$). The summary result for the European population, excluding 23andMe data, was selected for analysis as this dataset showed the maximum standardized SNP-based heritability (Z score of $h^2_{SNP}$) (Table S4).

**Genome-level pleiotropy**

We establish the presence of a “genome-level pleiotropy” between two traits based on a significant correlation between all linkage disequilibrium (LD)-independent (i.e., uncorrelated) SNPs across a pair of traits. The correlation is estimated using Pearson correlation ($r$) between the standardized effects (Z score of SNP effect size) of LD-independent SNPs of two traits, a method similar to the SNP effect concordant analysis (SECA) approach. $^{22}$ A significant correlation estimate indicates genetic overlap at the genome level, and its sign informs on the average concordance or discordance of the SNP effects on the two traits across the genome (at the genome level). A positive correlation...
indicates that, overall, the SNP effects across the traits are enriched for concordance, and a negative correlation indicates that the SNP effects across the traits are enriched for discordance. Here, we employ this approach to examine whether genetic influences underlying blood analyte levels are significantly concordant or discordant with the genetic influences underlying severe COVID-19 risk. In total, 109,025 LD-independent SNPs were used for this analysis. The Z score of an SNP effect summarizes the standardized effect of the SNP on the trait. A significant correlation of this measure for independent SNPs between severe COVID-19 and an analyte indicates concordant (positive correlation) or discordant (negative correlation) genome-wide pleiotropy. We emphasize that this approach is not limited to the most significantly associated SNPs and assumes that polygenic traits are influenced by many genetic factors spread across the genome and that although the majority of LD-independent SNPs will not be a (or in LD with a) true risk locus, there will be an enrichment for correlated effects at less significant SNPs because the GWAS traits were insufficiently powered to detect all risk loci at the genome-wide significant level. We also emphasize that significant pleiotropy at the gene level and the locus level (following sections) do not require global SNP effect correlation.

We found 21 analytes with significant genome-wide-pleiotropy with severe COVID-19 (false discovery rate [FDR] < 0.05) (Figure 2A; Table S5). Based on superclass categories, we found eight proteins, six lipids and lipid-like molecules, three organic acids and derivatives, three unknown (i.e., uncharacterized) metabolites, and one organoheterocyclic compound. We found concordant genome-wide pleiotropy between blood levels of 13 analytes and risk of severe COVID-19 and discordant genome-wide pleiotropy between blood levels of eight analytes and risk of severe COVID-19 (Figure 2A). The top five results include a significant genome-level pleiotropy with an increased risk of severe COVID-19 for higher blood levels of C-reactive protein (CRP), urate, and cystatin C and lower blood levels of high-density lipoprotein (HDL) cholesterol and apolipoprotein A.

Gene-level pleiotropy
We defined the “gene-level pleiotropy” as a significant enrichment of the genes associated with severe COVID-19 in the gene set associated with an analyte. To identify the genes associated with severe COVID-19 and the analytes, we conducted MAGMA gene-based analysis.23 Then, we utilized the MAGMA gene results and binomial tests to test for enrichment of genes associated with severe COVID-19 and each analyte.

We identified 38 analytes with significant gene-level pleiotropy with severe COVID-19 (FDR < 0.05; Figure 2B; Table S6). Based on superclass categories, we found 20 lipid and lipid-like molecules, nine proteins, three organoheterocyclic, two organic acids and derivatives, two unknown metabolites, and one homogeneous non-metal compound. Interestingly, genes associated with three analytes showed >10% overlap with genes associated with severe COVID-19, including two unknown metabolites, X-08402 (40.43%) and X-11315 (13.33%), and pyroglutamine (22.86%).

Locus-level pleiotropy (SNP-level pleiotropy)
We defined the “locus-level pleiotropy” based on GWAS-PW analysis results. The method scans 1,703 LD-independent genomic regions25 and estimates the posterior probability for a region to contain an SNP that influences both an analyte and severe COVID-19. We declared the presence of locus-level pleiotropy if there was at least one LD-independent genomic region containing an SNP that affected both an analyte and severe COVID-19 with a posterior probability >0.9 and defined the SNP with the most significant association in the severe COVID-19 GWAS summary statistics in that region as the lead pleiotropic SNP.

We identified 43 analytes with significant locus-level pleiotropy with severe COVID-19 (Figure 2C; Table S7). Based on the superclass categories, there were 27 lipid and lipid-like molecules, eight proteins, three organoheterocyclic compounds, two organic oxygen compounds, one organic acid and derivatives, one homogeneous metal compound, and one unknown metabolite. Overall, six analytes (alkaline phosphatase, aspartate aminotransferase, insulin-like growth factor 1 (IGF-1), SHBG, triglycerides, and urea) shared two regions with severe COVID-19, while the remaining 37 analytes shared only one region.

We found five unique regions, each harboring a pleiotropic SNP that influences blood levels of at least one of the 43 analytes and the risk of severe COVID-19 (Figure 3). In terms of the number of analytes associated with a region, the most commonly shared region was 9q34.13–q34.2 on chromosome 9, shared between 33 analytes and severe COVID-19 (Figure 3A). The lead pleiotropic SNP in this region, rs495828 near ABO, shows genome-wide significant association (p < 5 x 10^{-8}) with blood levels of 23 analytes, nominal association (5 x 10^{-8} < p < 1 x 10^{-3}) with blood levels of ten analytes, and genome-wide suggestive association (p = 6 x 10^{-5}) with severe COVID-19 risk (Figure 3F). The next commonly shared region was 10q23.33, shared between ten analytes and severe COVID-19 (Figure 3B). The lead pleiotropic SNP in this region, rs10786156 in PLCE1, shows genome-wide significant association (p < 5 x 10^{-8}) with blood levels of seven analytes and severe COVID-19 risk. The pleiotropic regions 14q21.1 (Figure 3C) and 19p13.2 (Figure 3D) were shared between two analytes and severe COVID-19 each, while the pleiotropic region on 17q21.33 (Figure 3E) was shared between severe COVID-19 and one analyte. The effects of the lead pleiotropic SNPs in the identified regions on the blood analyte levels and risk of severe COVID-19 obtained from their respective GWAS are shown in Figure 3F.

Causality
Our pleiotropy analyses identified 71 unique analytes genetically associated with severe COVID-19. These findings not only suggest shared biological mechanisms (i.e., horizontal pleiotropy) but also the likelihood of causal influence (i.e., vertical pleiotropy). Hence, we examined the causal influence of all the analytes that exhibited detectable pleiotropy. Prior to conducting the causality analysis, we ensured that the blood levels of the identified analytes have at least nominally significant LD score regression (LDSC)-based27 genetic correlation with severe COVID-19 risk. Following this, we tested for causality using the latent causal variable (LCV) model.28

We estimated the genetic correlation between the blood levels of the identified 71 analytes that showed a significant
pleiotropy with risk of severe COVID-19 (Figure 2). Of these, 12 analytes with a nominally significant (p < 0.05) LDSC genetic correlation ($r_g$) were tested for causality by the LCV model. We identified higher levels of blood triglycerides to have a significant causal influence on an increased risk of severe COVID-19 (Table 1). The high genetic causality proportion (GCP) of 0.82 (SE = 0.14) indicates that 82% of the genetic components of triglycerides are causal for severe COVID-19 disease.

**DISCUSSION**

Here, we found significant pleiotropy between 71 blood analyte levels and severe COVID-19 risk at the genome, gene, or/and "Figure 2. Significant pleiotropy between blood levels of 71 analytes and risk of severe COVID-19 (A) Genome-level pleiotropy defined by a significant Pearson correlation (r) between LD-independent SNPs from blood analytes and severe COVID-19 GWAS summary statistics (FDR < 0.05). Circles represent r, and the error bars indicate the 95% confidence interval. (B) Gene-level pleiotropy defined as a significant enrichment of genes associated across a blood analyte and severe COVID-19 (FDR < 0.05). The bars indicate the proportion of the analyte-associated genes overlapping with severe COVID-19-associated genes. (C) Locus-level pleiotropy defined using GWAS-PW analysis. The bars indicate the proportion of severe COVID-19 loci also influencing the analyte (posterior probability > 0.9)."
Figure 3. Locus-level pleiotropism between blood analyte levels and risk of severe COVID-19

(A–E) Locuszoom plots of the significant locus-level pleiotropies identified by GWAS-PW. The association values of the SNPs are with respect to severe COVID-19. The labeled SNP (purple circle) is the lead SNP in severe COVID-19 GWASs and was selected to represent pleiotropy between severe COVID-19 and the associated analytes.

(F) The effect sizes of the five detected representative pleiotropic lead SNPs for the significantly associated analytes. The lead pleiotropic SNPs rs495828, rs10786156, rs2899909, rs2304256, and rs3785928 were shared between severe COVID-19 and 33, 10, 2, 2, and 1 analyte, respectively.
and the tested liver function analytes. Shared genetic factors contribute to both severe COVID-19 and liver analytes, we found no evidence for a causal relationship (p > 0.05). Collectively, this suggests that only horizontal pleiotropy is present, where shared genetic factors contribute to both severe COVID-19 and the tested liver function analytes.

Table 1. Causality test results for blood analytes with a nominally significant genetic correlation (p < 0.05) with severe COVID-19

| Blood analyte                     | Superclass                           | Class                        | r<sub>g</sub> | SE<sub>r</sub> | p<sub>r</sub> | GCP  | SE<sub>GCP</sub> | p<sub>GCP</sub> | FDR<sub>GCP</sub> |
|----------------------------------|--------------------------------------|------------------------------|---------------|--------------|------------|------|-----------------|----------------|-----------------|
| Triglycerides                    | lipids and lipid-like molecules      | steroids and steroid derivatives | 0.14          | 0.04         | 2.54 x 10^-3 | 0.82 | 0.14            | 4.90 x 10^-7 | 5.89 x 10^-7    |
| C-reactive protein               | protein                              | protein                      | 0.20          | 0.05         | 2.36 x 10^-5 | 0.18 | 0.32            | 5.29 x 10^-1 | 9.63 x 10^-1    |
| Urate                            | organoheterocyclic compounds         | imidazopyrimidines           | 0.15          | 0.04         | 8.89 x 10^-4 | -0.30| 0.41            | 9.63 x 10^-1 | 9.63 x 10^-1    |
| Total cholesterol in small VLDL  | lipids and lipid-like molecules      | lipoprotein lipids           | 0.37          | 0.17         | 2.81 x 10^-2 | 0.34 | 0.47            | 2.18 x 10^-1 | 9.63 x 10^-1    |
| Concentration of very small VLDL particles | lipids and lipid-like molecules | lipid features              | 0.30          | 0.15         | 4.86 x 10^-2 | 0.04 | 0.55            | 9.23 x 10^-1 | 9.63 x 10^-1    |
| Alanine aminotransferase         | protein                              | enzyme                       | 0.13          | 0.05         | 8.24 x 10^-3 | -0.36| 0.40            | 3.78 x 10^-1 | 9.63 x 10^-1    |
| IGF-1                            | protein                              | hormone                      | -0.14         | 0.05         | 2.96 x 10^-3 | 0.01 | 0.57            | 9.01 x 10^-1 | 9.63 x 10^-1    |
| HDL cholesterol                  | lipids and lipid-like molecules      | steroids and steroid derivatives | -0.15         | 0.05         | 2.59 x 10^-3 | 0.03 | 0.25            | 8.91 x 10^-1 | 9.63 x 10^-1    |
| Apolipoprotein A                 | protein                              | lipid-related protein        | -0.11         | 0.05         | 2.36 x 10^-2 | -0.22| 0.47            | 6.79 x 10^-1 | 9.63 x 10^-1    |
| Gamma glutamyltransferase        | protein                              | enzyme                       | 0.09          | 0.04         | 4.66 x 10^-2 | -0.07| 0.56            | 8.00 x 10^-1 | 9.63 x 10^-1    |
| Glycated hemoglobin              | protein                              | protein                      | 0.12          | 0.05         | 2.06 x 10^-2 | -0.02| 0.58            | 7.87 x 10^-1 | 9.63 x 10^-1    |
| Cystatin C                       | organic acids and derivatives         | carboxylic acids and derivatives | 0.16          | 0.05         | 4.97 x 10^-4 | -0.01| 0.57            | 8.98 x 10^-1 | 9.63 x 10^-1    |

Causality testing was conducted only for those analytes with significant LDSC-based genetic correlation (p<0.05). Of these, only triglycerides displayed a significant causal influence on severe COVID-19. Given the current health and economic burden of the COVID-19 pandemic, our results will aid in the prioritization of blood analytes for screening, prognosis, and therapy.

The presence of significant pleiotropy between the risk of severe COVID-19 and blood levels of six analytes (alanine aminotransferase, alkaline phosphatase, apolipoprotein B [ApoB], CRP, triglycerides, and urate) at all three levels of the investigation indicate “widespread genetic pleiotropy.” Higher levels of these six analytes have genome-wide pleiotropy with an increased risk of severe COVID-19. Alanine aminotransferase (AST) and alkaline phosphatase are enzymes commonly found in the liver, and increased levels of these enzymes often indicate liver dysfunction. In line with our results, a previous study examining the relationship between severe COVID-19 and blood levels of liver analytes showed that the level of AST was positively associated with severe disease and that the level at the time of admission could predict the severe course of the disease (specificity = 0.84, 95% confidence interval [CI] = 0.77–0.88). The liver is a major player in the immune system, and inflammation and injured liver can lead to increased susceptibility to infections. However, despite the widespread genetic pleiotropy between severe COVID-19 and liver analytes, we found no evidence for a causal relationship (p > 0.05). Collectively, this suggests that only horizontal pleiotropy is present, where shared genetic factors contribute to both severe COVID-19 and the tested liver function analytes.

ApoB is a major structural protein of the very-low-density lipoprotein (VLDL), intermediate-density lipoprotein (IDL), and low-density lipoprotein (LDL) complexes that carry cholesterol and triglycerides in blood. These complexes are atherogenic; hence, the level of ApoB is a significant indicator of the risk of cardiovascular disease. Also, we found several components of VLDL, IDL, and LDL exhibiting significant pleiotropy with severe COVID-19 at gene and locus levels. The locus-level analysis also identified a shared region on chromosome 19 that harbors the LDLR gene encoding the LDL receptor. The previous observational implication of ApoB in severe COVID-19 is unclear; however, our results strongly suggest that an elevated level of ApoB is associated with severe COVID-19 risk.

CRP is an acute-phase protein involved in the response to inflammatory stimuli such as stress, trauma, and infections; thus, its level is an indicator of the inflammatory state. CRP plays a vital role in the first line of innate host defense and is elevated in the serum of patients with COVID-19 who progressed to severe infection. It is also associated with a respiratory decline in patients with COVID-19. We found widespread genetic pleiotropy and a significant LDSC genetic correlation (r<sub>g</sub>) between increased levels of CRP and severe COVID-19 risk, indicating a substantial shared genetic architecture. This confirms the prognostic role of CRP in severe COVID-19. However, the analyte was not causally associated—suggesting horizontal, and not vertical, pleiotropy.

Urate is a kidney function biomarker, and higher levels in blood suggest decreased kidney function. The kidney is one of the target organs of severe COVID-19, and pre-existing kidney disease is a major player in the immune system, and inflammation and injury of the kidney can lead to increased susceptibility to infections.
risk for COVID-19 infection and death. Interestingly, p = 0.017. Our genetic causality analyses found that higher TRIGlyceride levels increased the risk of severe COVID-19 disease. Despite a positive significant genetic correlation, urate does not causally affect severe COVID-19 (p > 0.05).

Triglycerides, a cardiovascular disease biomarker, also showed widespread genetic pleiotropy with severe COVID-19. It has been observed that among hospitalized patients, patients with endpoints of death or ICU admission had significantly higher levels of triglyceride compared with patients who were discharged or had mild disease. Similarly, another study assessing patients with COVID-19 in normal (≤1.7 mmol/L) or high triglyceride (>1.7 mmol/L) groups identified a significantly higher rate of survival in the normal group compared with the high triglyceride group (absolute survival difference = 2.5%; p = 0.017). Our genetic causality analyses found that higher levels of triglycerides causally increase the risk of severe COVID-19 disease (p = 5 × 10−5). This finding strongly supports the use of lipid-lowering drugs such as statins and fibrates against severe COVID-19. A recent study examining the in-hospital use of statins observed decreased mortality in the statin-treatment group compared with the non-statintreatment group. Retrospective studies, however, produced conflicting results on the protective effect of prior statins use. The conflicting results could partially be explained by the presence of confounders such as dietary factors, which can affect triglyceride levels. Therefore, we expect that prolonged reduction of triglycerides to help prevent severe COVID-19.

One of the analytes we tested was testosterone, a primary male sex hormone, that showed gene- and locus-level pleiotropy with severe COVID-19 and suggests that low levels of testosterone are genetically associated with increased risk of severe disease. Furthermore, we found that lower levels of SHBG, a protein that is positively correlated with testosterone levels, are genetically associated with severe COVID-19 risk. Consistent with our findings, a study has reported testosterone deficiency as a risk factor for severe COVID-19.

Our locus-level pleiotropy analysis identified rs495828 in the ABO blood group gene to be shared between severe COVID-19 and 33 analytes and has been associated with COVID-19 severity. Non-O-blood-type individuals have an increased risk of COVID-19 infection and death. Interestingly, rs495828 in the ABO blood group is also significantly associated with activity of the angiotensin-converting enzyme (ACE) and thereby can have implications in response to ACE inhibitors, a class of anti-hypertension medications that are linked to decreased risk of COVID-19 disease.

Our genetics-based research complements and extends the findings from observational studies on the association between severe COVID-19 and blood analytes. First, our genetically associated analytes are robust to confounding effects of disease, medication, environmental effects, and reverse causation. Second, our genetic causality results indicating that high triglycerides levels increase the risk of severe COVID-19 provide important confirmation and support for the use of triglycerides-lowering drugs to prevent severe COVID-19. Last, the identified pleiotropies between severe COVID-19 and blood analytes provide important insight into the biological factors involved and deliver opportunities to develop improved screening strategies and drug selection (including design of randomized controlled trials) for preventing and treating severe COVID-19.

Limitations of the study
Our results should be interpreted in the context of some potential limitations. The severe COVID-19 GWAS cohort contains a relatively small number of severe COVID-19 cases (n = 5,101) compared with controls (n = 1,383,241), which limits the identification of significant loci associated with severe COVID-19. The GWAS sample sizes of the blood analyte datasets vary substantially—e.g., a maximum of 24,925 for metabolites and a maximum of 344,292 for traditional clinical chemistry analytes. This disparity could have contributed to the absence of metabolites with significant pleiotropy at all three levels with severe COVID-19 risk. Next, our analyses used only HapMap3 SNPs to reduce potential unreliable SNPs. While restriction to the robustly assayed and well-characterized common HapMap3 SNPs produces reliable genome- and gene-level results, we may have lost shared SNPs in locus-level pleiotropy investigations. Although suggestive loci associated with severe COVID-19 have been identified on chromosome X, GWAS results were not available for chromosome X for many analytes. The GWAS summary statistics used in this study are from the European population except for two studies from the Finnish population.

The inclusion of the Finnish population is acceptable as the focus of this study is to identify polygenic pleiotropy utilizing autosomal HapMap3 common variants that are less likely to significantly differ in frequency across European populations from different countries/regions. In addition, the GWASs adjust for sex, age, and principal components to limit potential bias due to sex and population stratification. However, extending these results to other (non-European) ancestries should be done with caution. While the confounding effects such as food consumption or fasting have a minimized impact on the results due to population-level data, it should be noted that blood levels of analytes (e.g., triglycerides and lipoprotein traits) are affected by fasting or non-fasting state. However, we emphasize that such effects primarily influence the power of the original analytic GWAS to identify genetic factors influencing blood analyte levels—i.e., controlling/adjusting for dietary factors reduces variation due to non-genetic factors (noise) and enriches for variation due to genetic factors, thus increasing the power to detect genetic factors associated with the natural variation in analyte levels in a population sample. And, differences in food consumption or fasting, for example, would not produce false positive association of genetic factors with analyte levels. Lastly, while we employed approaches that are robust to small(ER) GWAS sample sizes, these approaches have their own limitations. For example, the correlation coefficients produced by the correlation approach examine genome-level pleiotropy are small in
magnitude; however, these should not be interpreted as the magnitude of analyte-severe COVID-19 relationships but rather as a significant correlation that establishes a relationship, and the sign of the correlation indicates concordance or discordance of genome-wide SNP effects.

**STAR METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **METHOD DETAILS**
  - GWAS summary statistics of blood analytes
  - GWAS summary statistics for severe COVID-19
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Genome-level pleiotropy
  - Gene-level pleiotropy
  - Locus-level pleiotropy
  - Causality test

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.111708.

**ACKNOWLEDGMENTS**

We thank the participants and researchers involved in COVID-19 and metabolite GWASs. H.M.T. and A.S. are grateful for support from the Queensland University of Technology for their respective QUT Postgraduate Research Scholarships.

**AUTHOR CONTRIBUTIONS**

H.M.T. and D.R.N. conceptualized the study. H.M.T. collected and curated COVID-19 and metabolite GWASs. H.M.T. and A.S. carried out the formal analysis. A.S. wrote the initial draft. H.M.T., D.M., and D.R.N. reviewed and edited the draft. D.R.N. supervised the work.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

Received: November 22, 2021
Revised: April 21, 2022
Accepted: November 1, 2022
Published: November 7, 2022

**REFERENCES**

1. Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20, 533–534.
2. Sakurai, A., Sasaki, T., Kato, S., Hayashi, M., Tsuzuki, S.-i., Ishihara, T., Iwata, M., Morise, Z., and Doi, Y. (2020). Natural history of asymptomatic SARS-CoV-2 infection. N. Engl. J. Med. 383, 885–886.
3. Berlin, D.A., Gulick, R.M., and Martinez, F.J. (2020). Severe covid-19. N. Engl. J. Med. 383, 2451–2460.
4. Zhang, X., Tan, Y., Ling, Y., Lu, G., Liu, F., Yi, Z., Jia, X., Wu, M., Shi, B., and Xu, S. (2020). Viral and host factors related to the clinical outcome of COVID-19. Nature 583, 437–440.
5. COVID-19 Host Genetics Initiative (2021). Mapping the human genetic architecture of COVID-19. Nature 600, 472–477.
6. Zhu, L., She, Z.-G., Cheng, X., Qin, J.-J., Zhang, X.-J., Cai, J., Lei, F., Wang, H., Xie, J., and Wang, W. (2020). Association of blood glucose control and outcomes in patients with COVID-19 and pre-existing type 2 diabetes. Cell Metab. 31, 1068–1077.e3.
7. Lighter, J., Phillips, M., Hochman, S., Sterling, S., Johnson, D., Francois, F., and Stachel, A. (2020). Obesity in patients younger than 60 years is a risk factor for Covid-19 hospital admission. Clin. Infect. Dis. 71, 896–897.
8. Yang, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., Ji, R., Wang, H., Wang, Y., and 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.
9. Rhee, E.P., Ho, J.E., Chen, M.-H., Shen, D., Cheng, S., Larson, M.G., Ghorbani, A., Shi, X., Helenius, I.T., and O’Donnell, C.J. (2013). A genome-wide association study of the human metabolome in a community-based cohort. Cell Metab. 18, 130–143.
10. Gallois, A., Merfford, J., Ko, A., Vaysse, A., Julienne, H., Ala-Korpela, M., Laakso, M., Zaitlen, N., Pajukanta, P., and Aschard, H. (2019). A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. Nat. Commun. 10, 1–13.
11. Wurtz, P., Wang, Q., Soininen, P., Kangas, A.J., Fatemifar, G., Tynkkynen, T., Tiainen, M., Perola, M., Tillin, T., and Hughes, A.D. (2016). Metabolomic profiling of statin use and genetic inhibition of HMG-CoA reductase. J. Am. Coll. Cardiol. 67, 1200–1210.
12. Assafalq, M., Bertini, I., Colangiul, D., Luchinat, C., Schäfer, H., Schütz, B., and Spraul, M. (2008). Evidence of different metabolic phenotypes in humans. Proc. Natl. Acad. Sci. USA 105, 1420–1424.
13. Schranner, D., Kastenmüller, G., Schöfler, M., Römisc-Margl, W., and Wackerhage, H. (2020). Metabolite concentration changes in humans after a bout of exercise: a systematic review of exercise metabolomics studies. Sports Med. Open 6, 1–17.
14. Bermingham, K.M., Brennan, L., Segurado, R., Barron, R.E., Gibney, E.R., and Spraul, M. (2008). Evidence of different metabolic phenotypes in humans. Proc. Natl. Acad. Sci. USA 105, 1420–1424.
15. Cheng, S.-C., Joosten, L.A., and Netea, M.G. (2014). The interplay between central metabolism and innate immune responses. Cytokine Growth Factor Rev. 25, 707–713.
16. Mathis, D., and Shoelson, S.E. (2011). Immune metabolism: an emerging frontier. Nat. Rev. Immunol. 11, 81. https://doi.org/10.1038/nri2922.
17. Song, J.-W., Lam, S.M., Fan, X., Cao, W.-J., Wang, S.-Y., Tian, H., Chua, G.H., Zhang, C., Meng, F.-P., and Xu, Z. (2020). Omics-driven systems interrogation of metabolic dysregulation in COVID-19 pathogenesis. Cell Metabol. 32, 188–202.e5.
18. Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R., and Qian, L. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. Cell 178, 59–72.e15.
19. Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Aribilos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., et al. (2020). Genome-wide association study of severe covid-19 with respiratory failure. N. Engl. J. Med. 383, 1522–1534. https://doi.org/10.1056/NEJMoa202283.
20. van Rheenen, W., Peyrot, W.J., Schork, A.J., Lee, S.H., and Wray, N.R. (2019). Genetic correlations of polygenic disease traits: from theory to practice. Nat. Rev. Genet. 20, 567–581. https://doi.org/10.1038/s41576-019-0137-2.
21. Tanha, H.M., Sathyarayarayan, A., and Nyholt, D.R. (2021). Genetic overlap and causality between blood metabolites and migraine. Am. J. Hum. Genet. 108, 2086–2098. https://doi.org/10.1016/j.ajhg.2021.05.011.
22. Nyholt, D.R. (2014). SECA: SNP effect concordance analysis using genome-wide association summary results. Bioinformatics 30, 2086–2088. https://doi.org/10.1093/bioinformatics/btu171.

23. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. 11, e1004219.

24. Pickrell, J.K., Berisa, T., Liu, J.Z., Séguéral, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. 48, 709–717. https://doi.org/10.1038/ng.3570.

25. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics 32, 283–285. https://doi.org/10.1093/bioinformatics/btv546.

26. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gledt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337. https://doi.org/10.1093/bioinformatics/btp419.

27. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295. https://doi.org/10.1038/ng.3211.

28. O’Connor, L.J., and Price, A.L. (2018). Distinguishing genetic correlation from causation across 62 diseases and complex traits. Nat. Genet. 50, 1728–1734. https://doi.org/10.1038/s41588-018-0255-0.

29. Vánca, S., Hegyi, P.J., Zádor, N., Szakó, L., Vörrendi, N., Ocskay, K., Földi, M., Dembrosvky, F., Domötör, Z.R., and Jánosi, K. (2020). Pre-existing liver diseases and on-admission liver-related laboratory tests in COVID-19: a prognostic accuracy meta-analysis with systematic review. Front. Med. 7, 572115.

30. Robinson, M.W., Harmon, C., and O’Farrell, C. (2016). Liver immunity and its role in inflammation and homeostasis. Cell. Mol. Immunol. 13, 267–276.

31. Albillos, A., Lario, M., and Álvarez-Mon, M. (2014). Cirrhosis-associated immune dysfunction: distinctive features and clinical relevance. J. Hepatol. 67, 1385–1396.

32. Pencina, M.J., D’Agostino, R.B., and Vasan, R.S. (2016). Linkage disequilibrium blocks in human populations. Bioinformatics 32, 260–266. https://doi.org/10.1093/bioinformatics/btv546.

33. Vitanova, K., Bises, D., Pai, M.G.J., Acharjee, A., Bankar, R., Palanivel, V., Sarkar, A., Verma, A., Mukherjee, A., and Choudhury, M. (2021). Proteomics and machine learning approaches reveal a set of prognostic inflammatory biomarker trends predict respiratory decline in COVID-19 patients. Cell Rep. Med. 1, 100144.

34. Luo, X., Zhou, W., Yan, X., Guo, T., Wang, B., Xia, H., Ye, L., Xiong, J., Jiang, Z., and Liu, Y. (2020). Prognostic value of C-reactive protein in patients with coronavirus 2019. Clin. Infect. Dis. 77, 2174–2179.

35. Ali, N. (2020). Elevated level of C-reactive protein may be an early marker to predict risk for severity of COVID-19. J. Med. Virol. 92, 2409–2411.

36. Chen, W., Zheng, K.I., Liu, S., Yan, Z., Xu, C., and Qiao, Z. (2020). Plasma CRP level is positively associated with the severity of COVID-19. Ann. Clin. Microbiol. Antimicrob. 19, 1–7.

37. Domrongsatchaiporn, S., Sritara, P., Kittayakara, C., Sitchantrakul, W., Kritthapol, V., Lolekha, P., Cheepudomwit, S., and Yipintsoi, T. (2005). Risk factors for development of decreased kidney function in a southeast Asian population: a 12-year cohort study. J. Am. Soc. Nephrol. 16, 791–799.

38. Fythe, J.E., Assimom, M.M., Tugman, M.J., Chang, E.H., Gupta, S., Shah, J., Sosa, M.A., Renaghan, A.D., Melamed, M.L., and Wilson, F.P. (2021). Characteristics and outcomes of individuals with pre-existing kidney disease who were admitted to intensive care units in the United States. Am. J. Kidney Dis. 77, 190–203.e1.

39. Chen, B., Lu, C., Gu, H.-Q., Li, Y., Zhang, G., Liu, J., Luo, X., Zhang, L., Hu, Y., and Lan, X. (2021). Serum uric acid concentrations and risk of adverse outcomes in patients with COVID-19. Front. Endocrinol. 12, 633767.

40. Perks, B., Wiggins, R.C., Gbadegesin, R., Viangos, C.N., Seelew, D., Numbere, C., Gurg, P., Verma, R., Chaib, H., and Hoskins, B.E. (2020). Potential cloning uncover mutations in PLCε1 responsible for a nephropenic syndrom variant that may be reversible. Nat. Genet. 48, 1397–1405.

41. Divers, J., Ma, L., Brown, W.M., Palmer, N.D., Choi, Y., Israni, A.K., Pa- stan, S.O., Julian, B.A., Gaston, R.S., and Hicks, P.J. (2020). Genome-wide association study for time to failure of kidney transplants from African American deceased donors. Clin. Transplant. 34, e13827.

42. Bellia, A., Andreadi, A., Giudice, L., De Taddeo, S., Maiorino, A., D’Ippolito, I., Giorgino, F.M., Ruotolo, V., Romano, M., and Magrini, A. (2021). Athero- genic dyslipidemia on admission is associated with poorer outcome in people with and without diabetes hospitalized for COVID-19. Diabetes Care 44, 2149–2157.

43. Masana, L., Correig, E., Ibarretxe, D., Anoro, E., Arroyo, J.A., Jericò, G., Guerrero, C., Naf, S., Pardo, A., and Pereira, V. (2021). Low HDL and high triglycerides predict COVID-19 severity. Sci. Rep. 11, 1–9.

44. Zhong, P., Wang, Z., and Du, Z. (2021). Serum triglyceride levels and related factors as prognostic indicators in COVID-19 patients: a retrospective study. Immun. Inflamm. Dis. 9, 1055–1060.

45. Zhang, X.-J., Qin, J.-J., Cheng, X., Shen, L., Zhao, Y.-C., Yuan, Y., Lei, F., Chen, M.-M., Yang, H., and Bai, L. (2020). In-hospital use of statins is associated with a reduced risk of mortality among individuals with COVID-19. Cell Metabol. 32, 176–187.e4.

46. Tan, W.-Y., Young, B.E., Lye, D.C., Chew, D.E., and Dalan, R. (2020). Statin use is associated with lower disease severity in COVID-19 infection. Sci. Rep. 10, 1–7.

47. Miticchiaone, G., Schiavone, M., Curnis, A., Arca, M., Antinori, S., Gasperetti, A., Mascoli, G., Severino, P., Sabato, F., and Caracciolo, M.M. (2021). Impact of prior statin use on clinical outcomes in COVID-19 patients: data from tertiary referral hospitals during COVID-19 pandemic in Italy. J. Clin. Lipidol. 15, 68–78.

48. Butt, J.H., Gerds, T.A., Schou, M., Kragholm, K., Phelps, M., Havers-Borger sen, E., Yafasova, A., Gislason, G.H., Torp-Pedersen, C., and Keber, L. (2020). Association between statin use and outcomes in patients with coronavirus disease 2019 (COVID-19): a nationwide cohort study. BMJ Open 10, e044421.

49. Carlistrom, K., Eriksson, A., Stege, R., and Rannevik, G. (1990). Relationship between serum testosterone and sex hormone-binding globulin in adult men with intact or absent gonadal function. Int. J. Androl. 13, 67–73.

50. Lancer, L., Burkert, F.R., Thommes, L., Egger, A., Hoermann, G., Kaser, S., Pinggera, G.M., Aniker, M., Griesmacher, A., and Weiss, G. (2021).
Testosterone deficiency is a risk factor for severe COVID-19. Front. Endocrinol. 12, 731.

57. Zietz, M., Zucker, J., and Tatonetti, N.P. (2020). Associations between blood type and COVID-19 infection, intubation, and death. Nat. Commun. 11, 1–6.

58. Chung, C.M., Wang, R.Y., Chen, J.W., Fann, C.S., Leu, H.B., Ho, H.Y., Ting, C.T., Lin, T.H., Sheu, S.H., Tsai, W.C., et al. (2010). A genome-wide association study identifies new loci for ACE activity: potential implications for response to ACE inhibitor. Pharmacogenomics J. 10, 537–544. https://doi.org/10.1038/tjp.2009.70.

59. Li, M., Wang, Y., Ndiwane, N., Orner, M.B., Palacios, N., Mittler, B., Berlowitz, D., Kazis, L.E., and Xia, W. (2021). The association of COVID-19 occurrence and severity with the use of angiotensin converting enzyme inhibitors or angiotensin-II receptor blockers in patients with hypertension. PLoS One 16, e0248652. https://doi.org/10.1371/journal.pone.0248652.

60. Hippisley-Cox, J., Tan, P.S., and Coupland, C. (2020). Risk of severe COVID-19 disease with ACE inhibitors and angiotensin receptor blockers: cohort study including 8.3 million people. Heart. https://doi.org/10.1136/heartjnl-2020-318312.

61. Locke, A.E., Steinberg, K.M., Chiang, C.W.K., Service, S.K., Havulinna, A.S., Stell, L., Pirinen, M., Abel, H.J., Chiang, C.C., Fulton, R.S., et al. (2019). Exome sequencing of Finnish isolates enhances rare-variant association power. Nature 572, 323–328. https://doi.org/10.1038/s41586-019-1457-z.

62. Sidhu, D., and Naugler, C. (2012). Fasting time and lipid levels in a community-based population: a cross-sectional study. Arch. Intern. Med. 172, 1707–1710. https://doi.org/10.1001/archinternmed.2012.3708.

63. Kettunen, J., Demirkan, A., Wurtz, P., Draisma, H.H., Haller, T., Rawal, R., Vaarhorst, A., Kangas, A.J., Lyytikäinen, L.P., Pirinen, M., et al. (2016). Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat. Commun. 7, 11122. https://doi.org/10.1038/ncomms11122.

64. Draisma, H.H.M., Pool, R., Kobli, M., Jansen, R., Petersen, A.K., Vaarhorst, A.A.M., Yet, I., Haller, T., Demirkan, A., Esko, T., et al. (2015). Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nat. Commun. 6, 7208. https://doi.org/10.1038/ncomms8208.

65. Shin, S.Y., Fauman, E.B., Petersen, A.K., Krumsieck, J., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.P., et al. (2014). An atlas of genetic influences on human blood metabolites. Nat. Genet. 46, 543–550. https://doi.org/10.1038/ng.2982.

66. Chang, C.C., Chow, C.C., Teller, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4. https://doi.org/10.1186/s13742-015-0047-8.

67. Julienne, H., Shi, H., Pasaniqu, B., and Aschard, H. (2019). RAISS: robust and accurate imputation from summary statistics. Bioinformatics 35, 4837–4839. https://doi.org/10.1093/bioinformatics/btz466.

68. Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., et al. (2018). Hmdb 4.0: the human metabolome database for 2018. Nucleic Acids Res. 46, D608–D617. https://doi.org/10.1093/nar/gkx1089.

69. GWAS round 2 results released on 1st August 2018. http://www.nealelab.is/uk-biobank.

70. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241. https://doi.org/10.1038/ng.3406.

71. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65. https://doi.org/10.1038/nature11632.

72. Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dernitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., et al. (2010). Integrating common and rare genetic variation in diverse human populations. Nature 467, 52–58. https://doi.org/10.1038/nature09298.

73. Hill, W.D., Davies, G., Harris, S.E., Hagnenaars, S.P., Liewald, D.C., Penke, L., Gale, C.R., and Deary, I. (2016). Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions. Transl. Psychiatry 6, e980.

74. Pickrell, J.K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am. J. Hum. Genet. 94, 559–573. https://doi.org/10.1016/j.ajhg.2014.03.004.

75. Nyholt, D.R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am. J. Hum. Genet. 74, 765–769.
STAR METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Deposited data      |        |            |
| 316 curated blood metabolite GWASs | Tanha et al. | N/A |
| Blood metabolite GWAS resource 1 | Gallois et al. | http://statgen.pasteur.fr/Download.html |
| Blood metabolite GWAS resource 2 | Locke et al. | ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/LockeAE_31367044_GCST008673 |
| Blood metabolite GWAS resource 3 | Kettunen et al. | http://www.computationalmedicine.fi/data.php?dataset=NMR_GWAS |
| Blood metabolite GWAS resource 4 | Draisma et al. | http://www.tweelingenregister.org/engagebiocratesgwama |
| Blood metabolite GWAS resource 5 | Shin et al. | http://metabolomics.helmholtz-muenchen.de/gwas/ |
| 28 clinical chemistries | N/A | http://www.nealelab.is/uk-biobank |
| Severe COVID-19 | COVID-19 Host Genetics Initiative, 2021 | https://www.covid19hg.org/ |

Software and algorithms

| SOFTWARE or ALGORITHM | SOURCE | IDENTIFIER |
|----------------------|--------|------------|
| GWAS-PW (v0.21)     | Pickrell et al. | https://github.com/joepickrell/gwas-pw |
| LCV (v1.0)          | O'Connor and Price | https://github.com/lukejoconnor/LCV |
| LDSC (v1.0.1)       | Bulik-Sullivan et al. | https://github.com/bulik/ldsc |
| LocusZoom online tool (Legacy single plot service) | Pruim et al. | http://csg.sph.umich.edu/locuszoom |
| MAGMA (v1.07b)      | de Leeuw et al. | https://ctg.cncr.nl/software/magma |
| PLINK (v1.9)        | Chang et al. | https://www.cog-genomics.org/plink2 |
| R (v4.0.3)          | The Comprehensive R Archive Network | https://www.r-project.org/ |
| RAISS (v1.0)        | Julienne et al. | https://statistical-genetics.pages.pasteur.fr/raiss/ |
| Other               |        |            |
| Human Metabolome Database (HMDB) | Wishart et al. | https://hmdb.ca/ |

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and codes should be directed to and will be fulfilled by the lead contact, Dale R Nyholt (d.nyholt@qut.edu.au).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.
- This paper does not report original code. The Linux command-line scripts and R functions used to format GWAS summary statistics data and to perform statistical analyses are available from the lead contact upon request.
- Additional Supplemental Items are available from Mendeley Data at https://doi.org/10.17632/srh5kxh44d.1. Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.
METHOD DETAILS

GWAS summary statistics of blood analytes
Curated GWAS summary statistics for 316 metabolites were obtained from our previously published study (Tables S1, S2, and S3). In this previous study, the 316 metabolites were identified after curation of GWAS summary statistics obtained from five studies (Table S2). The curation steps involved the selection of GWAS summary statistics with sufficient statistical power using polygenic $h^2_{SNP}$ Z score calculated using LDSC (https://data.broadinstitute.org/alkesgroup/LDSCORE/) ($h^2_{SNP}$ Z score > 1.64), imputation of HapMap3 SNPs using RAISS, and annotation of the superclass and class information based on the Human Metabolome Database (HMDB). For lipid-associated analytes (e.g., triglycerides in VLDL) that are not reported in the HMDB were categorized into the “lipids and lipid-like molecules” superclass. For more details on the GWAS summary statistics characteristics including sample size and estimates of $h^2_{SNP}$ based on superclass and class of the metabolites, refer to Table S3.

GWAS summary statistics for 30 traditional blood clinical chemistry analytes were downloaded from http://www.nealelab.is/uk-biobank/, which are the round 2 results of UK Biobank GWAS released on 1 August 2018 by Ben Neale’s lab (Table S2). Similar to the metabolite GWAS summary statistics, we selected GWAS summary statistics based on polygenic $h^2_{SNP}$ Z score calculated using LDSC ($h^2_{SNP}$ Z score > 1.64), imputation of HapMap3 SNPs using RAISS, estimation of SNP effects, and annotation of the superclass and class information. Selection of GWASs based on $h^2_{SNP}$ Z score resulted in 28 clinical chemistry analytes for further analyses (Table S1). Further information on the clinical chemistry analytes and their estimated $h^2_{SNP}$ measures are provided in Tables S1 and S3.

It should be noted that the imputed GWASs of severe COVID-19 and 344 blood analytes (316 metabolites and 28 clinical chemistry analytes) were used for estimating pleiotropy at the genome-, gene-, and locus-levels between COVID-19 and analytes. However, LDSC (for estimating $h^2_{SNP}$ and genetic correlation) and the latent causal variable (LCV) model analyses were performed using the original GWAS summary statistics data. This is because LDSC and LCV both use LD scores which are correlated with imputation quality, thus, using RAISS-imputed GWAS summary results could affect LDSC and LCV results.

GWAS summary statistics for severe COVID-19
GWAS summary statistics of severe COVID-19 outcome were downloaded from the COVID-19 Host Genetics Initiative website (link: https://www.covid19hg.org; Release 5, download date: 18 Jan 2021). Detailed information on the study groups, meta-analysis, and results for other outcomes are available on the website. Four GWAS summary statistics files are available based on the exclusion of 23andMe and UK Biobank datasets for severe COVID-19. To identify the best GWAS summary statistics, we estimated the polygenic $h^2_{SNP}$ using LDSC. The linkage disequilibrium (LD) information for HapMap3 SNPs obtained from the 1000 Genomes Project Phase 3 (1000G) European population was used. Quality control of the GWAS summary statistics data included filtering SNPs with imputation quality < 0.9 and minor allele frequency (MAF) < 0.01. The estimated heritability measures for the four GWAS summary statistics are provided in Table S4. In this study, we chose the GWAS summary statistics with the largest Z score of $h^2_{SNP}$ as this score is proportional to sample size, SNP-based heritability and proportion of causal variants present in the data. Accordingly, the COVID19_HGI_A2_ALL_eur_leave_23andme_202101107.b37.txt.gz GWAS summary statistics were chosen for pleiotropy analysis in this study. This dataset includes 5,101 severe COVID-19 cases and 1,383,241 controls and estimated the SNP effects for 939,094 genotyped SNPs.

Following this, summary statistics imputation of HapMap3 SNPs was performed using the Robust and Accurate Imputation from Summary Statistics (RAISS, https://statistical-genetics.pages.pasteur.fr/raiss/). The method leverages the LD between neighboring existing SNPs to impute the standardised effects (Z score of SNP effect) of missing SNPs. To increase imputation quality, we limited our imputation to only HapMap3 common SNPs (MAF ≥ 0.01) and filtered imputed SNPs with $R^2 < 0.6$. To include ambiguous SNPs (G/C and A/T) in our analyses, they were removed prior to imputation and then imputed using neighboring SNPs, making these SNPs uniform across all studied GWAS summary statistics. We independently estimated the standard errors for imputed SNPs using the allele frequencies from the 1000 Genome Phase 3 reference panel for the European population (N = 503) and the reported sample size of the GWAS. The effect sizes (Z) of SNPs were estimated as a product of the imputed Z scores and calculated standard errors.

QUANTIFICATION AND STATISTICAL ANALYSIS

Genome-level pleiotropy
The estimation of genome-level pleiotropy was conducted similar to the SNP effect concordant analysis (SECA) approach. Our approach involved the identification of LD-independent SNPs, an additional iteration of QC based on independent SNPs, correlation estimation and lastly, correction for multiple testing. The LD-independent SNPs in the RAISS-imputed severe COVID-19 GWAS summary statistics were identified and extracted using the p value informed LD-clumping approach in PLINK 1.9 with the flags –clump-r2 0.1 –clump-kb 10,000 and the LD reference panel as LD information of common SNPs of 1000 Genome Phase 3 reference panel for European population (N = 503). This resulted in 109,025 independent SNPs with the smallest COVID-19 GWAS p values at $R^2 < 0.1$ within their 10 Mb flanking regions. Summary statistics data for the same 109,025 independent SNPs were extracted from the blood analytes GWAS summary statistics. The next QC steps involved discarding GWAS summary statistics with missing SNPs of more
than 20% of the 109,025 independent SNPs resulting in 303 well-powered analyte GWASs. The Pearson correlations (r) between the independent SNPs Z scores of severe COVID-19 and 303 analytes were estimated. Rather than correct the results for 303 multiple tests, we calculated the number of study-wide independent tests to be 188 using Matrix Spectral Decomposition analysis of the 303 × 303 correlation matrix (calculated from pairwise correlation analysis between the independent SNP Z scores of the 303 analytes). Thus, the correlation p values were adjusted for 188 independent tests using the “p.adjust” function in R to identify study-wide significant correlations with an FDR <0.05.

Gene-level pleiotropy
The estimation of gene-level pleiotropy included annotation of SNPs to genes, identifying analytes and severe COVID-19 associated genes, estimation of enrichment of severe COVID-19 associated genes in analyte associated gene sets, and correction for multiple hypothesis testing. The SNPs were annotated to Ensembl genes of the GRCh37 reference build using a location-based annotation approach where an SNP was assigned to a gene if it was located within ±500 kb from the flanking gene boundaries. Overall, the SNPs were assigned to 18,562 autosomal protein-coding genes. Following annotation, gene-level association testing was conducted for all traits (COVID-19 and 344 analytes) using MAGMA v1.07b software (https://ctg.cncr.nl/software/magma).23 MAGMA estimates the association p value and the Z score for each gene by aggregating the association p values of all the gene-associated SNPs while taking into account the number of SNPs, LD between the SNPs, and gene size. The LD information was obtained from the 1000G European reference panel. The gene association significances were corrected for multiple testing using p.adjust and genes with FDR <0.05 were selected as statistically significant trait-associated genes.

To identify the statistically significant analytes that share gene-level pleiotropy with severe COVID-19, we tested if the number of analyte-associated genes overlapping with the severe COVID-19 associated genes was greater than expected. The enrichment was tested using a one-sided binomial test where the expected (null) was defined as the proportion of significant genes out of all genes tested in MAGMA analysis for severe COVID-19 (i.e., 139/18,562) and the observed as the proportion of genes overlapping the analyte-associated gene set. The binomial p values were adjusted using the p.adjust and analytes with FDR <0.05 were considered as sharing significant gene-level pleiotropy with severe COVID-19.

Locus-level pleiotropy
The estimation of locus-level pleiotropy was conducted using the pairwise GWAS approach (GWAS-PW, https://github.com/joepickrell/gwas-pw) approach.24 The GWAS-PW method divides the genome into 1,703 non-overlapping LD-independent blocks and for each block, estimates the posterior probability for four different scenarios: the genomic region block (i) contains an SNP that influences only severe COVID-19; (ii) contains an SNP that influences only the analyte; (iii) contains an SNP that influences both severe COVID-19 and the analyte; and (iv) contains an SNP that influences severe COVID-19 and a separate SNP that influences the analyte. The probability for each model is estimated against a null model which is that the block has no SNP associated with either severe COVID-19 or the analyte. Scenario (iii) indicates locus-level pleiotropy. Hence, we selected the analytes with a posterior probability >0.9 for scenario (iii) as the analytes with significant locus-level pleiotropy with severe COVID-19. After identifying pleiotropic regions, in each pleiotropic region, we selected the SNP with the most significant association in the severe COVID-19 GWAS as the lead pleiotropic locus. Next, we examined association significance of the pleiotropic locus in the analyte GWAS and ensured that it was at least nominally significant (p < 0.05). Regions with non-significant pleiotropic loci were not selected as significant regions exhibiting locus-level pleiotropy.

Causality test
The causal relationship (vertical pleiotropy) between severe COVID-19 and blood analytes was investigated using the LCV model approach.26 Genetic causality was investigated only for those analytes that showed significant pleiotropy with severe COVID-19 in at least one of the genome-, gene-, or locus-levels. However, prior to conducting LCV analysis, we tested the LDSC genetic correlation (r_g)27 between severe COVID-19 and the selected analytes and as suggested by the authors of the LCV method, we selected the analytes with at least nominally significant genetic correlation (p < 0.05) for LCV analysis.

The LCV model involves the development of a latent variable that mediates genetic correlation between two traits of interest and has a causal effect on each trait. In this model, trait 1 is said to be genetically causal for trait 2 if trait 1 is genetically correlated with the latent variable more than trait 2. Based on the genetic correlation, trait 1 is fully genetically causal or partially genetically causal for trait 2. If fully genetically causal, all the genetic components of trait 1 are perfectly correlated with the latent variable. If partially genetically correlated, part of the genetic components of trait 1 is correlated with the latent variable. The causality is quantified through calculation of a genetic causality proportion (GCP) where GCP = 0 implies no partial causality and GCP = 1 implies full genetic causality. For the study-wide analysis, we correct the LCV significances for multiple testing using p.adjust and select those with FDR <0.05.