

Genome-wide cross-trait analysis and Mendelian randomization reveal a shared genetic etiology and causality between COVID-19 and venous thromboembolism

Xin Huang^{1,6}, Minhao Yao^{2,6}, Peixin Tian², Jason Y. Y. Wong³, Zilin Li⁴, Zhonghua Liu⁵✉ & Jie V. Zhao¹✉

Venous thromboembolism occurs in up to one-third of patients with COVID-19. Venous thromboembolism and COVID-19 may share a common genetic architecture, which has not been clarified. To fill this gap, we leverage summary-level genetic data from the latest COVID-19 host genetics consortium and UK Biobank and examine the shared genetic etiology and causal relationship between COVID-19 and venous thromboembolism. The cross-trait and co-localization analyses identify 2, 3, and 4 shared loci between venous thromboembolism and severe COVID-19, COVID-19 hospitalization, SARS-CoV-2 infection respectively, which are mapped to *ABO*, *ADAMTS13*, *FUT2* genes involved in coagulation functions. Enrichment analysis supports shared biological processes between COVID-19 and venous thromboembolism related to coagulation and immunity. Bi-directional Mendelian randomization suggests that venous thromboembolism was associated with higher risk of three COVID-19 traits, and SARS-CoV-2 infection was associated with a higher risk of venous thromboembolism. Our study provides timely evidence for the genetic etiology between COVID-19 and venous thromboembolism (VTE). Our findings contribute to the understanding of COVID-19 and VTE etiology and provide insights into the prevention and comorbidity management of COVID-19.

¹School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China. ²Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China. ³Epidemiology and Community Health Branch, National Heart Lung and Blood Institute, Bethesda, MD, USA. ⁴Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, IN, USA. ⁵Department of Biostatistics, Columbia University, New York, NY, USA. ⁶These authors contributed equally: Xin Huang, Minhao Yao. ✉email: z12509@cumc.columbia.edu; janezhao@hku.hk

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, has led to a worldwide pandemic since March 2020, and caused repeated waves of outbreaks across the globe. Venous thromboembolism (VTE) is one common and serious comorbidity. Notably, thrombotic events occur in up to one-third of patients with COVID-19¹. People with more pronounced thrombotic symptoms are more likely to develop severe COVID-19², and thrombotic complications are a well-established predictor of mortality in people with COVID-19¹. Consistently, several genome-wide association studies (GWASs) of COVID-19-related traits (severe COVID-19, COVID-19 hospitalization, and SARS-CoV-2 infection) have identified genetic loci in *ABO*^{3–6}, an established gene related to thrombosis. This evidence suggests that VTE and COVID-19 may share common genetic architecture. Identifying shared genetic factors contributing to both COVID-19 and VTE can provide novel insights into disease pathogenesis and pinpoint targets for therapeutic development or drug repurposing. However, to our knowledge, this problem has not been comprehensively investigated.

Meanwhile, the causal relationship between VTE and COVID-19 has not been clarified. On the one hand, VTE predicts COVID-19 severity and mortality^{1,2}. On the other hand, the cytokine storm and excessive inflammation caused by SARS-CoV-2 infection are hypothesized to lead to systemic coagulation dysfunction;^{7–10} therefore, SARS-CoV-2 infection may increase the risk of VTE. Many institutional evidence-based guidelines supported the use of prophylactic treatments such as anticoagulants for thromboprophylaxis in COVID-19 patients¹¹. However, the recommendation is mainly based on observational studies, which might be subject to unmeasured confounding bias. To resolve this issue, Mendelian randomization (MR) study uses genetic variants as instruments for causal inference and can provide unbiased causal effect estimates even in the presence of unmeasured confounding¹². One previous MR study in a relatively smaller GWAS for COVID-19 suggested that genetically predicted VTE is associated with a higher risk of COVID-19 hospitalization and SARS-CoV-2 infection¹³, however, the association with severe COVID-19 is uncertain possibly due to the limited sample size. The reverse association, i.e., the association of COVID-19 with VTE has not been examined. The recently updated GWAS for COVID-19 with doubled sample size provided a well-powered study dataset to investigate the relationship between VTE and COVID-19.

Taken together, to fill this knowledge gap, we performed a genome-wide cross-trait analysis and colocalization analysis based on summary statistics from GWASs of VTE and COVID-19 and identified shared loci between VTE and three COVID-19-related traits. We also conducted an enrichment analysis and found the shared genes were enriched for expression in the lung tissue and involved in coagulation and immunity. We also applied bi-directional Mendelian randomization to study the causal relationship between VTE and three COVID-19-related traits.

Results

Genetic correlation of VTE with COVID-19-related traits. The heritability (h^2) estimated by Linkage disequilibrium score regression (LDSC) analysis suggested severe COVID-19, COVID-19 hospitalization, SARS-CoV-2 infection, and VTE are heritable ($P < 0.05$ shown in Supplementary Table 1). We found positive overall genetic correlation of VTE with COVID-19 hospitalization ($r_g = 0.2320$, P value = 0.0092) (shown in Fig. 1 and Supplementary Table 2). The genetic correlation between VTE and severe COVID-19 and SARS-CoV-2 was positive but not significant at p value of 0.05 level. Further partitioned LDSC analysis

found that severe COVID-19, COVID-19 hospitalization, and SARS-CoV-2 infection are genetically correlated with VTE in five (Fetal DNase I hypersensitivity sites (DHS), H3K4me1, H3K9ac H3K27ac and transcription factor-binding site (TFBS)), 11 (except for super-enhancers regions) and 10 (except for conserved and super-enhancers regions) of the 12 functional categories, respectively (Fig. 2 and Supplementary Data 7).

Multi-trait analysis of GWAS (MTAG). Based on the COVID-19 host genetics initiative (HGI) updated GWAS meta-analysis (release 7), there were 45 genome-wide significant ($P < 5 \times 10^{-8}$) and uncorrelated ($r^2 < 0.01$) loci for severe COVID-19, 46 for COVID-19 hospitalization and 26 for SARS-CoV-2 infection (Supplementary Data 1), and 12 for VTE in the GWAS from UK Biobank (Supplementary Data 2). Compared with these previously identified SNPs, using MTAG incorporating information from GWAS of COVID-19 and VTE we did not identify additional genome-wide significant loci for COVID-19 traits or VTE.

As shown in Table 1, we identified 2 genetic loci associated with both VTE and severe COVID-19, 3 with both VTE and COVID-19 hospitalization, and 4 with both VTE and SARS-CoV-2 infection ($P_{\text{meta}} < 5 \times 10^{-8}$; single trait $P < 5 \times 10^{-3}$). In line with previous studies^{3–6,14,15}, we identified *ABO* gene and *FUT2* gene, which contributed to both VTE and COVID-19. The strongest association signals were localized on or near the *ABO* gene (index SNP: rs11244061 for severe COVID-19 and COVID-19 hospitalization; rs550057 for SARS-CoV-2 infection) at locus 9q34.2. We also identified *ADAMTS13*, which has not been reported yet. Figure 3 displays the Manhattan plots of these results (also shown in Supplementary Data 8–10).

Fine-mapping and colocalization analysis identify shared causal variants. For each loci associated with VTE and COVID-19, Supplementary Tables 3–5 listed all SNPs within 500 kb of these variants in the 99% credible sets. Co-localization analysis shows that all loci identified in cross-trait meta-analysis had the conditional posterior probability of colocalization > 0.8 , (Table 1), indicating that they are shared loci between the two traits.

GTEX tissue-specific expression analysis (TSEA) and over-representation enrichment analysis of shared genes. The potential associated genes identified for severe COVID-19 and COVID-19 hospitalization with VTE were significantly enriched for expression in the lung tissue; however, no highly enriched tissues were found for SARS-CoV-2 infection and VTE (Fig. 4 and Supplementary data 11). Gene ontology (GO) analysis highlighted several significant shared biological processes between three COVID-19 traits and VTE, such as “calcium-mediated signaling”, “second-messenger-mediated signaling”, “chemokine-mediated signaling pathway”, “response to chemokine”, “response to type I interferon” (Supplementary Tables 6–8).

Bi-directional MR analysis. We found that genetically predicted VTE was positively associated with higher risk of all three COVID-19-related traits (OR = 1.11 for severe COVID-19, 95% CI: 1.06–1.17, P value = 1.51×10^{-5} ; OR = 1.10 for COVID-19 hospitalization, 95% CI: 1.06–1.14, P value = 1.33×10^{-7} ; OR = 1.06 for SARS-CoV-2 infection, 95% CI: 1.04–1.09, P value = 1.13×10^{-5}) (Fig. 5 and Supplementary Data 12). MR-Egger intercept test did not indicate directional pleiotropy (Supplementary Table 9), which supported the validity of the findings. MR-PRESSO detected outliers in the associations of VTE with three COVID-19 traits (shown in Supplemental Data 3), and the positive associations of VTE with severe COVID-19 and COVID-19 hospitalization remained after removing these

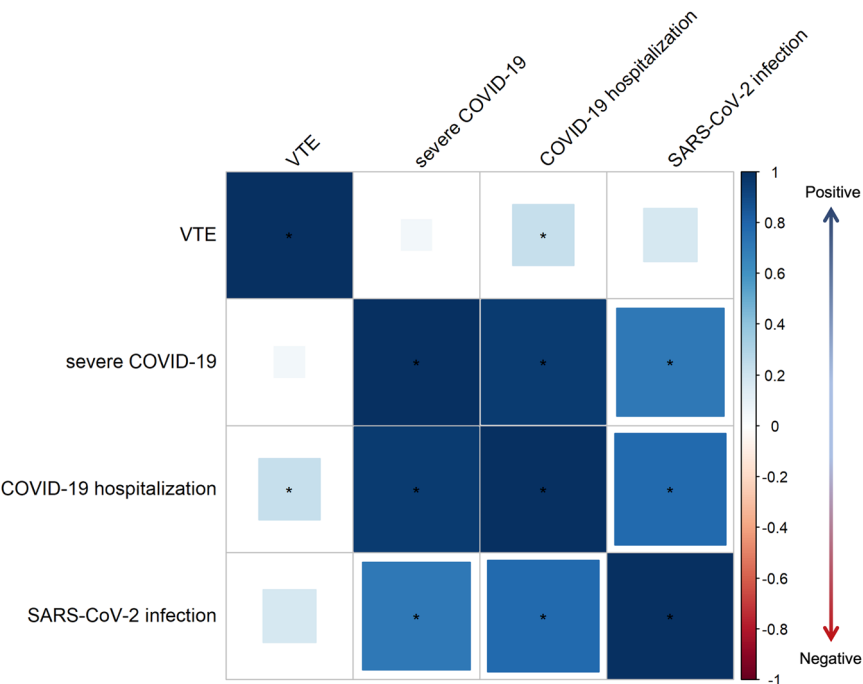


Fig. 1 Genetic correlation of VTE with COVID-19-related traits. Results were derived from linkage disequilibrium score regression (LDSC) analysis. Vertical axis represents the genetic correlation estimate R_g, with red indicating a negative correlation and blue indicating a positive correlation. The size of the colored squares is proportional to the P value, where larger squares represent a smaller P value. Asterisk represents significance (P < 0.05). The presented data are available in Supplementary Table 2.

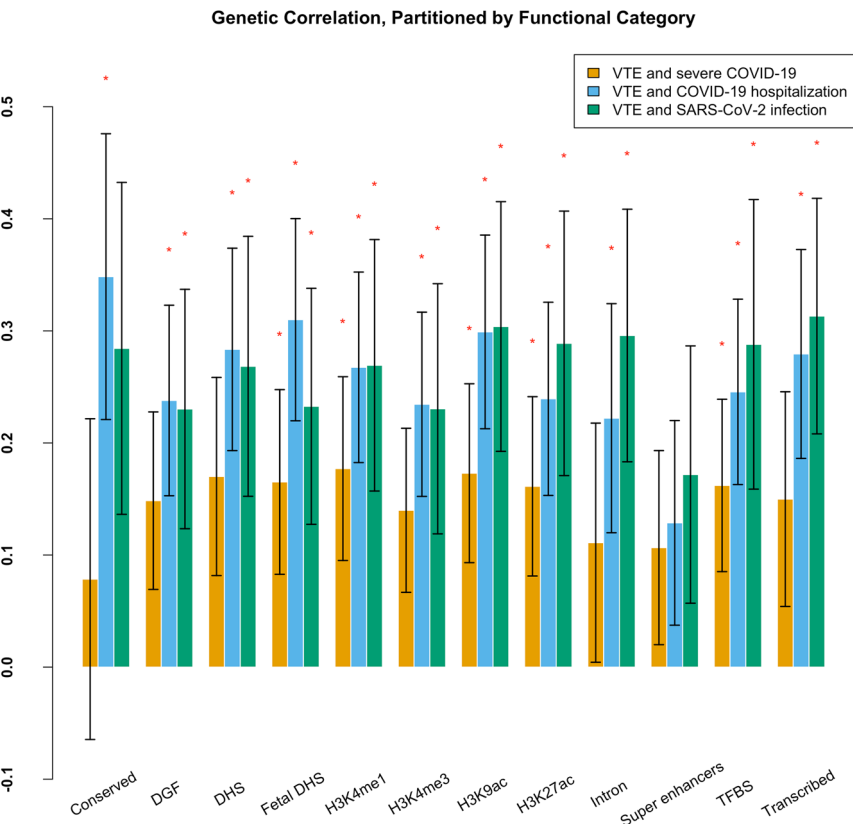


Fig. 2 Partitioned genetic correlation between VTE and COVID-19-related traits. Results were derived from partitioned linkage disequilibrium score regression (LDSC) analysis. The x axis represents the 12 functional categories, and the y axis represents the estimated partitioned genetic correlation. The significant functional categories (P < 0.05) are starred. Error bars represent the standard error of genetic correlation estimates. DGF DNase I digital genomic footprinting, DHS DNase I hypersensitivity site, TFBS transcription factor binding sites. The presented data are available in Supplementary Data 7.

Table 1 Genome-wide significant loci associated with the COVID-19 traits and VTE in cross-trait meta-analysis.												
SNP	CHR	A1	A2	Representative gene	BETA_1	SE_1	P_1	BETA_2	SE_2	P_2	P_META	Conditional posterior probability of colocalization (P(H ₄)/ (P(H ₃) + P(H ₄))
Severe COVID-19 and VTE												
rs11244061	9	T	C	near ABO	0.9528	0.3582	1.36E + 08	0.1145	0.1076	1.33E + 08	3.55E-26	8.55E-01
rs149181677	9	T	C	ADAMTS13	0.9785	0.3064	1.36E + 08	0.0526	0.1192	1.33E + 08	2.07E-10	8.55E-01
COVID-19 hospitalization and VTE												
rs11244061	9	T	C	near ABO	0.9528	0.3582	1.36E + 08	0.1070	0.0994	1.33E + 08	3.64E-28	9.00E-01
rs149181677	9	T	C	ADAMTS13	0.9785	0.3064	1.36E + 08	0.0530	0.0978	1.33E + 08	1.09E-10	9.00E-01
rs8176686	9	C	C	ABO	0.6813	-0.1300	1.36E + 08	0.4499	-0.0709	1.33E + 08	1.89E-14	9.00E-01
SARS-CoV-2 infection and VTE												
rs492602	19	G	A	FUT2	0.6754	0.0927	4.92E + 07	0.4707	-0.0357	4.87E + 07	1.42E-10	8.23E-01
rs550057	9	T	C	ABO	0.7830	0.3176	1.36E + 08	0.7402	-0.0963	1.33E + 08	1.56E-51	9.84E-01
rs71503180	9	A	G	near ABO	0.9757	-0.1391	1.36E + 08	0.0693	-0.0506	1.33E + 08	3.05E-10	9.84E-01
rs9411367	9	T	C	near ABO	0.6740	0.1155	1.36E + 08	0.1948	0.0380	1.33E + 08	3.20E-12	9.84E-01
CHR chromosome, A1 other allele, A2 effect allele. BETA_1, SE_1, and P_1 represent the estimated effects, standard errors, and p values in the GWAS of VTE. BETA_2, SE_2, and P_2 represent the estimated effects, standard errors and, p values in the GWASs of COVID-19-related traits. P_META represents the p value of the cross-trait meta-analysis.												

outliers (Supplementary Figure 1). In the reverse direction, we found a positive association of genetically predicted SARS-CoV-2 infection with higher risk of VTE (OR = 1.75, 95% CI: 1.18–2.58, *P* value = 4.92×10^{-3}), and null associations of genetically predicted severe COVID-19 or COVID-19 hospitalization with VTE (Fig. 6 and Supplementary Data 12). These results are robust to different MR methods (Supplementary Figure 2). And we obtained similar bi-directional associations between VTE and COVID-19 if we used COVID-19 HGI release 7 excluding data from UK Biobank (Supplementary Table 10).

Discussion

In this study, we comprehensively investigated the shared genetic etiology between three COVID-19 related traits and VTE based on the latest data from COVID-19 HGI and UK Biobank. We found a significant positive genetic correlation of VTE with COVID-19 hospitalization and identified specific shared loci for VTE with each COVID-19 trait in *ABO*, *ADAMTS13* and *FUT2* genes, which was involved in coagulation. Enrichment analysis suggested that genes potentially associated with VTE, and COVID-19 were enriched for expression in the lung tissue, and supported pathways related to coagulation and immunity. Moreover, we examined their causal relationship using bi-directional MR, which suggested that VTE may increase the risk of severe COVID-19, COVID-19 hospitalization, and SARS-CoV-2 infection, and interestingly, SARS-CoV-2 infection may increase the risk of VTE.

LDSC analysis in our study found genetic correlation between VTE and COVID-19, which provided genetic evidence of a significant positive genetic correlation between VTE and COVID-19 hospitalization. Although LDSC analysis did not find significant overall genetic correlation between severe COVID-19 and VTE, however partitioned LDSC analysis found severe COVID-19 and VTE was positively correlated in some functional categories, including TFBS, Fetal DHS, H3K4me1, H3K9ac and H3K27ac, which are associated with the control of transcription and the status of *cis*-regulatory elements such as promoters and enhancers within gene regulatory regions^{16–19}.

Notably, we identified several shared genetic loci between VTE and COVID-19 using cross-trait meta-analysis and colocalization analysis. These loci were located on or near the *ABO*, *FUT2*, and *ADAMTS13* genes, which are well-established VTE related genes^{20,21}. Our findings are consistent with previous genetic studies showing *ABO* gene is associated with severe COVID-19 and SARS-CoV-2 infection^{3–5,14}, possibly by regulating thrombosis^{3,22}, as *ABO* gene is responsible for post-translational glycosylation of coagulation factors. In line with the role of *ABO* gene, we found that *FUT2* gene, a fucosyltransferase gene involved in ABO blood group antigen synthesis, was associated with both VTE and SARS-CoV-2 infection. We also found a shared gene, *ADAMTS13*, encoding the protein ADAMTS13, which has not been previously reported in COVID-19 GWAS. Defects in this gene are associated with blood clotting and thrombotic thrombocytopenic purpura²³, and a previous MR study showed genetically predicted ADAMTS13 level is associated with severe COVID-19²⁴.

In addition, findings from enrichment analysis suggest immune function might be involved in the shared etiology. In the enrichment analysis, we found several shared pathways related to immune function, such as calcium-mediated signaling and chemokine/interferon related response. Previous studies suggested that calcium signaling may play a role in hemostasis and thrombosis^{25,26}. Calcium signaling is also of paramount importance in immune cells^{27,28}, and chemokine and interferon play a major role in activating host immune and inflammatory

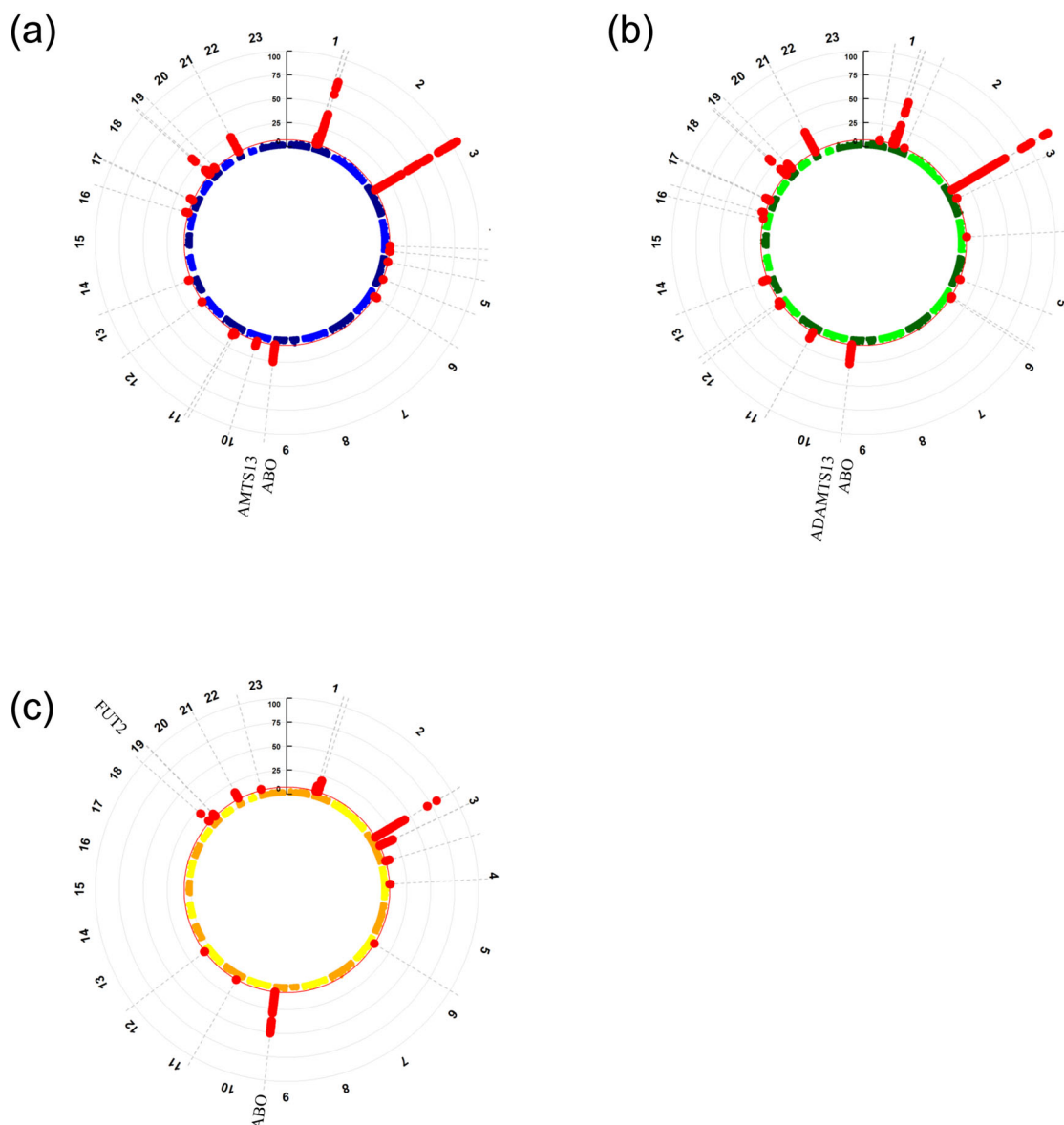


Fig. 3 Circular Manhattan plot of cross-trait analysis. a VTE and severe COVID-19; **b** VTE and COVID-19 hospitalization; **c** VTE and SARS-CoV-2 infection. Each point represents a SNP, and significant loci with meta-analysis P value $< 5 \times 10^{-8}$ are colored in red. Shared loci with meta-analysis P value $< 5 \times 10^{-8}$ and single-trait P value $< 5 \times 10^{-3}$ are labeled with the mapped genes. SNPs are arranged according to the chromosome position. The presented data are available in Supplementary Data 8–10.

responses^{29–31}. Our TSEA reported that shared genes for severe COVID-19 and COVID-19 hospitalization with VTE were mainly enriched for gene expressions in the lung tissue. Consistently, pulmonary vascular endothelial injury and immunothrombosis (most of which occur within the lung microvessels) are key drivers of severe events after SARS-CoV-2 infection, such as acute respiratory failure and ARDS^{32,33}. So, these evidence suggests that the common pathways between COVID-19 and VTE may relate to immunity, endothelial cell function, and coagulation.

Our study also provided timely evidence regarding the causal relationship between VTE and COVID-19 traits. Interestingly, our MR analysis showed a positive association between genetic susceptibility for VTE and the risk of severe COVID-19, COVID-19 hospitalization and SARS-CoV-2 infection, as well as a positive association of SARS-CoV-2 infection with VTE. Despite the statistical significance, our finding still needs to be interpreted with caution as we cannot exclude that this positive association

may have occurred by chance. However, the positive association was also shown in a previous MR study using COVID-19 HGI release 5 data which showed that genetically predicted VTE was associated with higher risk of COVID-19 hospitalization and SARS-CoV-2 infection¹³. Using COVID-19 HGI release 7 with doubled sample size of release 5, we added by showing a positive association of genetically predicted VTE with higher risk of severe COVID-19. In addition, there are concerns that partial overlap of the samples may bias the estimates, however, we obtained consistent associations when we used COVID-19 HGI release 7 excluding data from UK Biobank. We also considered that the observed associations may be mediated by coagulation factors, as VTE is associated with coagulation factors³⁴, and coagulation is a risk factor of COVID-19^{35,36}. Although a previous MR study on 12 coagulation factors and COVID-19 severity found that VWF was link to COVID-19 severity using COVID-19 HGI release 5²⁴, we found null associations of these coagulation factors with three COVID-19 traits using the latest and much larger GWAS of

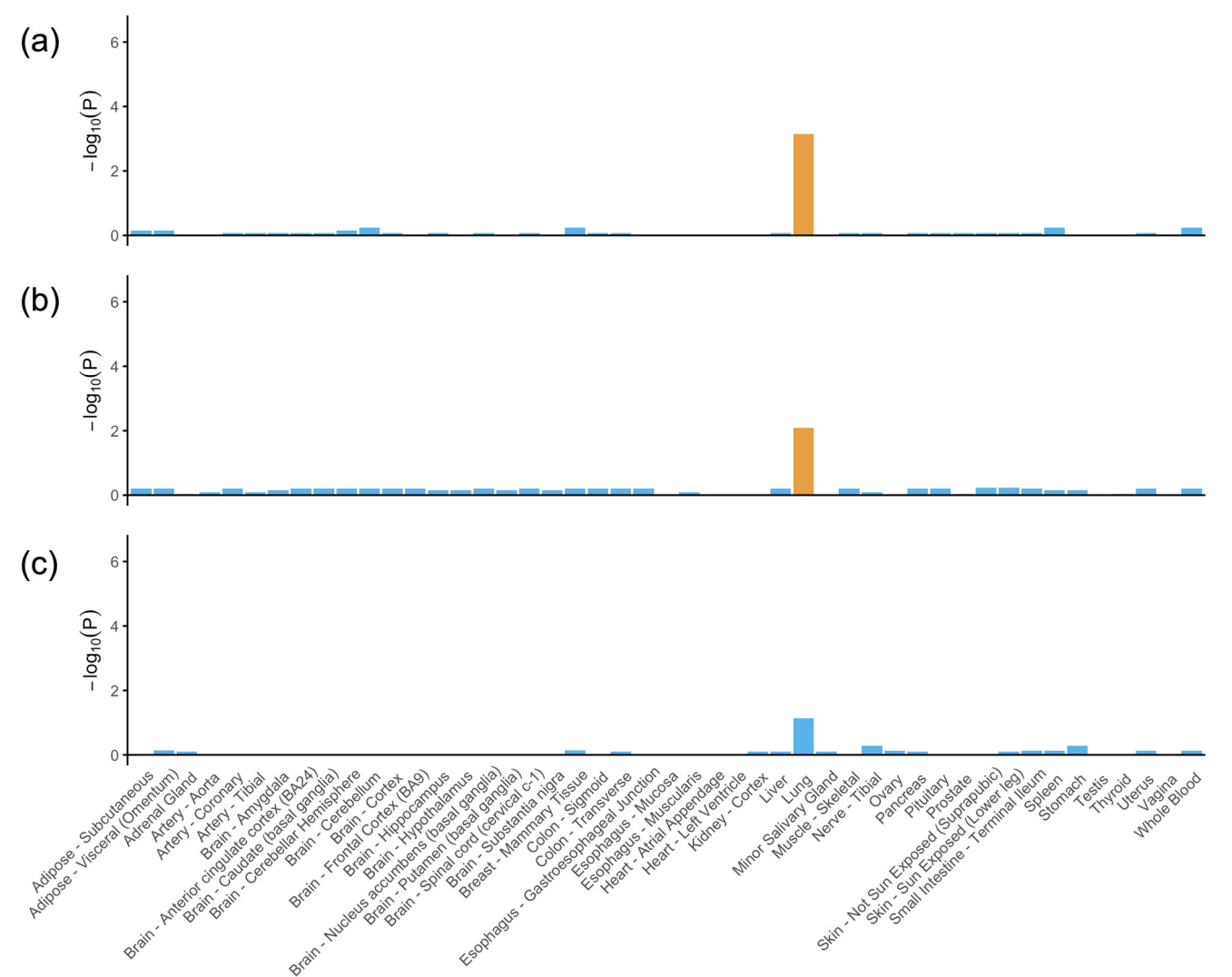


Fig. 4 GTEx tissue enrichment analysis for expression of shared significant genes (meta-analysis P value $< 5 \times 10^{-6}$). **a** VTE and severe COVID-19; **b** VTE and COVID-19 hospitalization; **c** VTE and SARS-CoV-2 infection. P values of Fisher’s exact test after Benjamin–Hochberg correction are presented in $-\log_{10}$ scale. Orange represents significant tissue enrichment (Lung, P value = 7.16×10^{-4} for severe COVID-19 and VTE, P value = 8.20×10^{-3} for COVID-19 hospitalization and VTE). The presented data are available in Supplementary Data 11.

Direction	No. of SNPs	OR(95%CI)	P-value
VTE to Severe Covid	98	1.11(1.06,1.17)	1.51e-05
VTE to Covid Hospitalization	99	1.10(1.06,1.14)	1.33e-07
VTE to SARS-Cov2 Infection	102	1.06(1.04,1.09)	1.13e-05

Fig. 5 Mendelian randomization analysis on the association of genetically predicted VTE with the risk of COVID-19-related traits using inverse variance weighting (IVW). The estimates are presented as odds ratios (OR) with 95% confidence intervals (CI). The presented data are available in Supplementary Data 12.

COVID-19, COVID-19 HGI release 7 (Supplementary Figure 3). Further studies are needed to explore whether there are other potential mediators. Our study delivers an important information that VTE possibly increase the risk of COVID-19, which matters in the consideration of COVID-19 management strategies, especially for people

with VTE. For example, taking vaccination to lower the risk for severe COVID-19 in people with VTE. Meanwhile, our MR analysis suggested that genetically predicted SARS-CoV-2 infection was possibly associated with a higher risk of VTE. Partly consistent with our results, in a cohort of 153,760 individuals with COVID-19, as well as two sets of control cohorts with 5,637,647

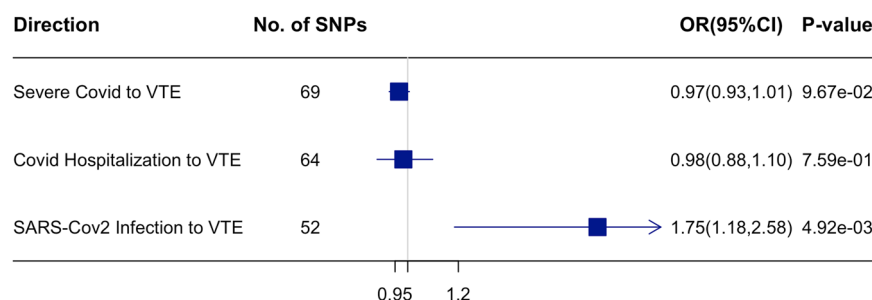


Fig. 6 Mendelian randomization analysis on the association of genetically predicted COVID-19-related traits with the risk of VTE using IVW. The estimates are presented as odds ratios (OR) with 95% confidence intervals (CI). An arrow represents the estimate that is out of the boundary in this direction. The presented data are available in Supplementary Data 12.

(contemporary controls) and 5,859,411 (historical controls) individuals, people with SARS-CoV-2 infection had higher risk of cardiovascular events, including VTE³⁷. Nevertheless, a replication study of MR analysis in a larger GWAS of VTE will be worthwhile.

To the best of our knowledge, our study is the first to use large-scale genetic data to explore the shared genetic architecture between COVID-19 related traits and VTE, providing timely evidence and more insights into the genetic etiology between them. The GWAS summary-level data for COVID-19 was obtained from the COVID19 HGI release 7 summary statistics, the largest and latest GWAS of COVID-19 available, which doubles the sample size from the previous version, and improves study power.

We also acknowledge several limitations of our study. First, the GWAS data for COVID-19 and VTE used in this study were derived from the European population, so the associations may not be generalizable to other ancestries. Second, although GWAS summary statistics conducted study-specific quality control, misclassification of COVID-19 might exist. Third, the summary statistics limit us to assess sex and age-specific genetic effects. Fourth, the GWAS of COVID-19 related traits might be conducted at different time periods, so there might be differences in the SARS-CoV-2 infection strain. However, our study does not aim to assess the association with specific strain of SARS-CoV-2 infection. Fifth, genetic instruments for COVID-19 only capture a small proportion of variance. Although MR analysis provided unconfounded estimates, it is less precise than conventional observational studies. We observed a positive association of genetically predicted SARS-CoV-2 infection with VTE using inverse variance weighting and MR-RAPS, but the association included the null in weighted median, MR-Egger, MR-PRESSO and weighted mode, which may be due to the wider confidence intervals. Replication in a larger GWAS will be needed for validation. Finally, although our study provides evidence of genetic correlation and genetic overlap between COVID-19 and VTE, the underlying biological mechanisms are still unclear. For example, we identified rs149181677, the variant located on the *ADAMTS13* gene, was shared between COVID-19 and VTE. Although this SNP was associated with the expression of *ADAMTS13* genes³⁸, we cannot fully confirm the link between this genetic variant with the protein ADAMTS13, as there is no study assessing the association between this genetic variant with ADAMTS13. Future studies are warranted to test the link.

In conclusion, our findings provided important evidence of genetic correlations between severe COVID-19, COVID-19 hospitalization, SARS-CoV-2 infection and VTE, and highlighted their common genetic architecture, with shared genes closely related to coagulation and immunity. We also found a causal association of VTE with COVID-19. Our work contributes to the understanding of COVID-19 and VTE etiology and provides

more insights into the prevention and comorbidity management of COVID-19.

Methods

Study population. The GWAS summary statistics for COVID-19 of European ancestry were provided by the COVID-19 host genetics initiative (COVID-19 HGI) round 7 (<https://www.covid19hg.org/results/>, release date: 08 April 2022). COVID-19 HGI is the largest GWAS of COVID-19 to date, combining data from over 3 million individual samples across 82 large cohort studies in 35 countries. We included three COVID-19 related traits: (1) Severe COVID-19, defined as COVID-19-confirmed individuals with very severe respiratory symptoms or those who died from the disease (up to 13,769 cases and 1,072,442 controls); (2) COVID-19 hospitalization defined as individuals who were hospitalized for related infection symptoms, with laboratory-confirmed SARS-CoV-2 infection (up to 32,519 cases and 2,062,805 controls); (3) SARS-CoV-2 infection defined as all individuals who reported positive (laboratory diagnosis, physician diagnosis or self-report) for SARS-CoV-2 infection (up to 122,616 cases and 2,475,240 controls). For VTE, we used summary statistics on the GWAS of VTE (3900 cases and 369,592 controls of European ancestry) from the UK Biobank provided by Lee Lab (<https://www.leelabsg.org/resources>).

Ethics approval. Participants for all studies included in COVID-19 HGI were recruited following protocols approved by the local institutional review boards and informed consent was obtained where required⁴. The UK Biobank has already received the ethical approval from North West Multi-center Research Ethics Committee (MREC), which covers the UK. It also got the approval from the Patient Information Advisory Group (PIAG) in England and Wales and from the Community Health Index Advisory Group (CHIAG) in Scotland. This study is an analysis using publicly available summary data that does not require additional ethical approval.

Linkage disequilibrium score regression (LDSC) analysis. LDSC analysis was conducted to assess the heritability for a single trait and genome-wide genetic correlations between two traits, where genome-wide associations were used in the calculation. The analysis was conducted using the LDSC software based on the GWAS summary statistics. LD scores of 1000 G European ancestry was used as reference³⁹.

An estimate of the heritability or genetic correlation was obtained by regressing the χ^2 statistics or the products of z scores on LD scores, respectively. LDSC can also correct for the inflation of test statistics caused by polygenicity⁴⁰. Under the polygenic model, the expected χ^2 statistic and the LD score l_j of SNP j follows the linear relationship:³⁹

$$E[\chi_j^2 | l_j] = \frac{N h^2 l_j}{M} + N a + 1 \quad (1)$$

where N is the sample size, M is the number of SNPs, h^2 is the heritability, and a measures the confounding bias. Therefore, an estimate of the heritability can be obtained by regressing the χ^2 statistics calculated from GWAS on the LD scores from a reference panel.

If we have two studies for two polygenic traits, there is a similar relationship between the product of the z -scores from two studies and the LD scores:⁴¹

$$E[z_{1j} z_{2j} | l_j] = \frac{\sqrt{N_1 N_2} \rho_g l_j}{M} + \frac{\rho N_s}{\sqrt{N_1 N_2}} \quad (2)$$

where N_i is the sample size for study i , ρ_g is the genetic covariance, N_s is the number of overlapping samples, and ρ is the phenotypic correlation. If we regress the product of the z -scores from two GWASs on the LD scores from a reference panel, we can get an estimate of the genetic covariance between two traits. The

genetic correlation then can be obtained by $r_g = \rho_g / \sqrt{h_1^2 h_2^2}$, where h_i^2 is the heritability of trait i .

Partitioned LDSC analysis. We performed partitioned LDSC analysis to estimate the genetic correlation between two traits within each of the following 12 functional categories:¹⁷ conserved region, DNaseI digital genomic footprinting region (DGF), DNase I hypersensitivity sites (DHS), fetal DHS, H3K4me1, H3K4me3, H3K9ac and H3K27ac, intron region, super enhancers, transcription factor-binding site (TFBS) and transcribed region. The re-calculated LD scores of the SNPs classified in each specific annotation category allowed us to find out which functional categories accounted for the majority of the overall genetic correlation.

Multi-trait analysis of GWAS. We applied MTAG, an approach for meta-analysis of summary statistics from GWAS of different traits robust to sample overlap, to identify novel loci with strong signals for COVID-19, and to detect genome-wide significant loci between VTE and three COVID-19 traits^{42–44}.

MTAG can improve the effect estimates for each COVID-19 trait by incorporating the weighted sum of GWAS estimates for VTE. Let β_j be the true effects of SNP j on multiple traits with $E[\beta_j] = \mathbf{0}$ and $\text{Var}(\beta_j) = \mathbf{\Omega}$. MTAG assumes that $\mathbf{\Omega}$ is the same for all SNPs. Based on the moment condition, the MTAG estimates can be expressed as a weighted sum of the GWAS estimates:

$$\hat{\beta}_{\text{MTAG},j,t} = \frac{\frac{\omega_{jt}}{\omega_{jt}} \left(\mathbf{\Omega} - \frac{\omega_{jt} \omega_{jt}'}{\omega_{jt}} + \mathbf{\Sigma}_j \right)^{-1}}{\frac{\omega_{jt}}{\omega_{jt}} \left(\mathbf{\Omega} - \frac{\omega_{jt} \omega_{jt}'}{\omega_{jt}} + \mathbf{\Sigma}_j \right)^{-1} \frac{\omega_{jt}}{\omega_{jt}}} \hat{\beta}_j \quad (3)$$

where ω_t is the t -th column of $\mathbf{\Omega}$, ω_{tt} is the t -th diagonal entry of $\mathbf{\Omega}$, $\mathbf{\Sigma}_j$ is the covariance of the GWAS estimates $\hat{\beta}_j$. Therefore, we can check whether there are novel loci that are extracted from the MTAG estimates and not identified using GWAS.

We also used MTAG to conduct genome-wide cross-trait meta-analysis, which utilizes sample size-weighted, fixed-effect model together with genetic covariance modeling from all sources to combine evidence of the genome-wide association between individual variants for VTE and COVID-19. The equation for the MTAG estimates can be simplified to:

$$\hat{\beta}_{\text{MTAG},j,t} = \frac{\mathbf{1}' \mathbf{\Sigma}_j^{-1}}{\mathbf{1}' \mathbf{\Sigma}_j^{-1} \mathbf{1}} \hat{\beta} \quad (4)$$

Under this assumption, MTAG can summarize the effect estimates from two GWASs and produce cross-trait effect estimates, which allows us to detect the shared significant loci between two traits.

Using cross-trait meta-analysis, we identified SNPs reaching genome-wide significance ($P_{\text{meta}} < 5 \times 10^{-8}$) in both traits and suggestive significance (single trait $P < 5 \times 10^{-3}$) in single traits⁴⁵. We further applied the PLINK clumping function (r^2 threshold = 0.01, distance = 500 kb) to ensure the independence of selected SNPs, i.e., SNPs in linkage disequilibrium that have a correlation >0.01 with the most significant SNP within a distance of 500 kb will be pruned⁴⁶.

Fine-mapping and co-localization analysis. For each locus associated with two traits, we used the Bayesian fine-mapping algorithm that assumes a multinomial likelihood for the joint distribution of the phenotypes and genotypes to identify a 99%-credible set of causal variants within 500 kb of the index SNP^{47,48}. This algorithm uses a flat prior with steepest descent approximation and only requires summary statistics. Then, we conducted the co-localization analysis to check whether the loci identified in the cross-trait meta-analysis are causal variants shared between COVID-19 and VTE. The SNP causality between two traits in a region can be assigned to one of the following five hypotheses⁴⁹:

- H_0 : No association with either trait.
- H_1 : Association with trait 1 only.
- H_2 : Association with trait 2 only.
- H_3 : Association with both traits, two independent SNPs.
- H_4 : Association with both traits, one shared SNP.

With the R package ‘coloc’, we can calculate the posterior probability for each hypothesis based on the summary statistics. In colocalization, we calculated the conditional posterior probability of colocalization ($P(H_4)/(P(H_3) + P(H_4))$)⁴⁹. If a SNP has the probability larger than 0.8, we labeled it as a co-localized genetic variant.

TSEA and over-representation enrichment analysis. We also performed TSEA with the HUGO Gene Nomenclature Committee (HGNC) name on genes that correspond with the associated loci with meta-analysis P value $< 5 \times 10^{-6}$ in cross-trait meta-analysis to test whether these genes are over-expressed in a specific tissue. We used the R package ‘deTS’, which uses GTEx RNA-seq data and the ENCODE panel as a reference panel, and calculates the corresponding z score for

each tissue⁵⁰. We also used the WEB-based GENE SeT Analysis Toolkit to further assess the overrepresented enrichment of the same genes in Gene ontology (GO) biological process⁵¹.

Mendelian randomization (MR) analysis. We conducted a bidirectional MR analysis to assess the causal association between each COVID-19 related trait and VTE. Genetic variants for VTE were obtained from a large GWAS of VTE restricted to Europeans (23,151 cases, 553,439 controls)⁵². We extracted genome-wide significant ($P < 5 \times 10^{-8}$) and uncorrelated ($r^2 < 0.2$) SNPs, and then excluded six SNPs in *ABO*, *FUT2*, *VWF* genes (Supplemental Data 3) as these gene regions have pleiotropic associations with multiple traits, including COVID-19. 107 SNPs have been selected as genetic instruments for VTE (Supplementary Data 3)⁵². The genetic associations of these SNPs with COVID-19-related traits were obtained from COVID-19 HGI release 7 GWAS meta-analysis, as shown in study population. Using the same process of identifying instruments ($P < 5 \times 10^{-8}$, $r^2 < 0.2$), and after excluding 2, 2, and 8 SNPs in *ABO*, *FUT2*, or *VWF* genes, we obtained 87, 83, and 61 SNPs as instruments for severe COVID-19, COVID-19 hospitalization and SARS-CoV-2 infection respectively based on COVID-19 HGI release 7 GWAS meta-analysis (Supplementary Data 4–6). The genetic associations of these SNPs with VTE were obtained from the GWAS of VTE (3900 cases and 369,592 controls of European ancestry) in the UK Biobank. We used inverse variance weighted (IVW) analysis in the main analysis⁵³, and used several other MR methods that are robust to pleiotropy, including the weighted median⁵⁴, MR-Egger⁵⁵, MR-PRESSO⁵⁶, MR-RAPS⁵⁷, and weighted mode⁵⁸ in sensitivity analysis. Considering sample overlap, we also repeated the bi-directional analysis in the sensitivity analysis using the GWAS summary statistics for COVID-19 excluding data from UK Biobank provided by COVID-19 HGI release 7 (<https://www.covid19hg.org/results/>).

Statistics and reproducibility. We used the following software and packages: R and R packages: R (version 4.0.5), TwoSampleMR (version 0.5.6), forestplot (version 2.0.1), mr.raps (version 0.2), dplyr (version 1.0.5), CMplot (version 3.6.2), corplot (version 0.92), ggplot2 (version 3.3.5), deTS (version 1.0), WebGestaltR (version 0.4.4), coloc (version 3.2-1), data.table (version 1.14.2), gprofiler2 (version 0.2.1). Python and Python-based software: Python (version 2.7.5 and 3.6.3), LDSC (version 1.0.1), MTAG (version 1.0.8). We used publicly available GWAS data with up to 13,769 cases and 1,072,442 controls for severe COVID-19; up to 32,519 cases and 2,062,805 controls for COVID-19 hospitalization; up to 122,616 cases and 2,475,240 controls for SARS-CoV-2 infection. To enable us to test the reproducibility of causal associations between VTE and three COVID-19 traits obtained from MR analysis, we used COVID-19 data excluding the UK Biobank for further analysis and obtained similar results (Supplementary Table 10).

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Genetic associations with VTE were from the UK biobank GWAS results provided by Lee Lab (<https://www.leelabsg.org/resources>). Genetic associations with severe COVID-19, COVID-19 hospitalization and SARS-CoV-2 infection were obtained from COVID-19 host genetics consortium GWAS meta-analyses round 7, downloaded from https://storage.googleapis.com/covid19-hg-public/freeze_7/results/20220403/pop_spec/sumstats/COVID19_HGI_A2_ALL_eur_leave23andme_20220403.tsv.gz; https://storage.googleapis.com/covid19-hg-public/freeze_7/results/20220403/pop_spec/sumstats/COVID19_HGI_B2_ALL_eur_leave23andme_20220403.tsv.gz; https://storage.googleapis.com/covid19-hg-public/freeze_7/results/20220403/pop_spec/sumstats/COVID19_HGI_C2_ALL_eur_leave23andme_20220403.tsv.gz. In the MR sensitivity analysis, Genetic associations with severe COVID-19, COVID-19 hospitalization and SARS-CoV-2 infection excluding data from UK Biobank were obtained from https://storage.googleapis.com/covid19-hg-public/freeze_7/results/20220403/leave_one_out/sumstats/COVID19_HGI_A2_ALL_leave_23andme_and_UKBB_20220403.tsv.gz; https://storage.googleapis.com/covid19-hg-public/freeze_7/results/20220403/leave_one_out/sumstats/COVID19_HGI_B2_ALL_leave_23andme_and_UKBB_20220403.tsv.gz; https://storage.googleapis.com/covid19-hg-public/freeze_7/results/20220403/leave_one_out/sumstats/COVID19_HGI_C2_ALL_leave_23andme_and_UKBB_20220403.tsv.gz. Source data underlying Figs. 2–6 are provided in Supplementary Data 7–12

Code availability

We used publicly available software for analysis in this study. Here, we listed the URLs (some of which are online methods) for the software: TwoSampleMR, forestplot, mr.raps, deTS, WebGestaltR, coloc, gprofiler2, LDSC, MTAG. Detailed code book, and analytic code will be made available upon request pending application and approval by the corresponding author.

Received: 1 July 2022; Accepted: 4 April 2023;

Published online: 21 April 2023

References

- Malas, M. B. et al. Thromboembolism risk of COVID-19 is high and associated with a higher risk of mortality: a systematic review and meta-analysis. *eClinicalMedicine* **29**, 100639 (2020).
- Zhou, F. et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054–1062 (2020).
- Severe Covid, G. G. et al. Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
- Niemi, M. E. K. et al. Mapping the human genetic architecture of COVID-19. *Nature* **600**, 472–477 (2021).
- Shelton, J. F. et al. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* **53**, 801–808 (2021).
- Kousathanas, A. et al. Whole genome sequencing reveals host factors underlying critical Covid-19. *Nature* **607**, 97–103 (2022).
- Middeldorp, S. et al. Incidence of venous thromboembolism in hospitalized patients with COVID-19. *J. Thromb. Haemost.* **18**, 1995–2002 (2020).
- Tang, N., Li, D., Wang, X. & Sun, Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J. Thromb. Haemost.* **18**, 844–847 (2020).
- Driggin, E. et al. Cardiovascular considerations for patients, health care workers, and health systems during the COVID-19 pandemic. *J. Am. Coll. Cardiol.* **75**, 2352–2371 (2020).
- Mehta, P. et al. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet* **395**, 1033–1034 (2020).
- Cuker, A. et al. American Society of Hematology living guidelines on the use of anticoagulation for thromboprophylaxis in patients with COVID-19: July 2021 update on post-discharge thromboprophylaxis. *Blood Adv.* **6**, 664–671 (2021).
- Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
- Xiang, Y., Chau, C.K.-L., Qiu, J., Rao, S. & So, H.-C. Exploring causal relationships between COVID-19 and cardiometabolic disorders: a bi-directional Mendelian randomization study. *medRxiv*, <https://www.medrxiv.org/content/10.1101/2021.03.20.21254008v1> (2021).
- Roberts, G.H.L. et al. Ancestry DNA COVID-19 host genetic study identifies three novel loci. *medRxiv*, <https://www.medrxiv.org/content/10.1101/2020.10.06.20205864v1> (2020).
- Thibord, F. et al. Cross-ancestry investigation of venous thromboembolism genomic predictors. *Circulation* **146**, 1225–1242 (2022).
- Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- Trynka, G. et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45**, 124–130 (2013).
- Vinson, C., Chatterjee, R. & Fitzgerald, P. Transcription factor binding sites and other features in human and *Drosophila* proximal promoters. *Subcell. Biochem.* **52**, 205–222 (2011).
- Heit, J. A. et al. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *J. Thromb. Haemost.* **10**, 1521–1531 (2012).
- Lotta, L. A. et al. Next-generation sequencing study finds an excess of rare, coding single-nucleotide variants of ADAMTS13 in patients with deep vein thrombosis. *J. Thromb. Haemost.* **11**, 1228–1239 (2013).
- Wang, L. et al. An atlas connecting shared genetic architecture of human diseases and molecular phenotypes provides insight into COVID-19 susceptibility. *Genome Med.* **13**, 83–83 (2021).
- Tsai, H. M. ADAMTS13 and microvascular thrombosis. *Expert Rev. Cardiovasc. Ther.* **4**, 813–825 (2006).
- Zhou, Y. et al. Coagulation factors and the incidence of COVID-19 severity: mendelian randomization analyses and supporting evidence. *Signal Transduct. Target. Ther.* **6**, 222–222 (2021).
- Nesbitt, W. S. et al. Intercellular calcium communication regulates platelet aggregation and thrombus growth. *J. Cell Biol.* **160**, 1151–1161 (2003).
- Varga-Szabo, D., Braun, A. & Nieswandt, B. Calcium signaling in platelets. *J. Thromb. Haemost.* **7**, 1057–1066 (2009).
- Vig, M. & Kinet, J.-P. Calcium signaling in immune cells. *Nat. Immunol.* **10**, 21–27 (2009).
- Trebak, M. & Kinet, J.-P. Calcium signalling in T cells. *Nat. Rev. Immunol.* **19**, 154–169 (2019).
- Moser, B. & Willmann, K. Chemokines: role in inflammation and immune surveillance. *Ann. Rheum. Dis.* **63**, ii84 (2004).
- McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nat. Rev. Immunol.* **15**, 87–103 (2015).
- Charo, I. F. & Ransohoff, R. M. The many roles of chemokines and chemokine receptors in inflammation. *N. Engl. J. Med.* **354**, 610–621 (2006).
- McFadyen, J. D., Stevens, H. & Peter, K. The emerging threat of (Micro) thrombosis in COVID-19 and its therapeutic implications. *Circ. Res.* **127**, 571–587 (2020).
- Bonaventura, A. et al. Endothelial dysfunction and immunothrombosis as key pathogenic mechanisms in COVID-19. *Nat. Rev. Immunol.* **21**, 319–329 (2021).
- Sabater-Lleal, M. et al. Genome-wide association transethnic meta-analyses identifies novel associations regulating coagulation factor VIII and von Willebrand factor plasma levels. *Circulation* **139**, 620–635 (2019).
- Arachchillage, D. R. J. & Laffan, M. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J. Thromb. Haemost.* **18**, 1233–1234 (2020).
- Liao, D. et al. Haematological characteristics and risk factors in the classification and prognosis evaluation of COVID-19: a retrospective cohort study. *Lancet Haematol.* **7**, e671–e678 (2020).
- Xie, Y., Xu, E., Bowe, B. & Al-Aly, Z. Long-term cardiovascular outcomes of COVID-19. *Nat. Med.* **28**, 583–590 (2022).
- Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
- Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
- Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
- Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
- Chung, J. et al. Genome-wide association study of cerebral small vessel disease reveals established and novel loci. *Brain* **142**, 3176–3189 (2019).
- Guo, Y. et al. A genome-wide cross-phenotype meta-analysis of the association of blood pressure with migraine. *Nat. Commun.* **11**, 3368 (2020).
- Bone, W. P. et al. Multi-trait association studies discover pleiotropic loci between Alzheimer's disease and cardiometabolic traits. *Alzheimers Res. Ther.* **13**, 34 (2021).
- Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- Huang, H. et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173–178 (2017).
- Farh, K. K. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
- Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genet.* **10**, e1004383 (2014).
- Pei, G., Dai, Y., Zhao, Z. & Jia, P. deTS: tissue-specific enrichment analysis to decode tissue specificity. *Bioinformatics* **35**, 3842–3845 (2019).
- Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based Gene SeT Analysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
- Klarin, D. et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat. Genet.* **51**, 1574–1579 (2019).
- Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
- Emdin, C. A., Khera, A. V. & Kathiresan, S. Mendelian randomization. *JAMA* **318**, 1925–1926 (2017).
- Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
- Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
- Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D. S. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann. Stat.* **48**, 1742–1769 (2020).
- Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).

Acknowledgements

This research was conducted using the summary statistics from COVID-19 host genetics consortium and UK Biobank. The authors would like to thank all participants in the study and investigators for sharing valuable data. The authors reported no funding was received for this study.

Author contributions

The manuscript was drafted by X.H., with great help of J.V.Z. and Z.L. (edited by J.Y.Y.W. and Z.L.L.). Data analysis was conducted by M.Y. and P.T. X.H. and M.Y. interpreted the results, with the help of J.Y.Y.W., Z.L.L., Z.L., and J.V.Z. J.V.Z. developed the study conception, Z.L. designed the analysis with discussion with J.V.Z., J.V.Z. and Z.L. critically revised the manuscript for important intellectual content. All authors read 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/s42003-023-04805-2>.

Correspondence and requests for materials should be addressed to Zhonghua Liu or Jie V. Zhao.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Kaoru Ito and George Inglis. Peer reviewer reports are available.

Reprints and permission information is available at <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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023