medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Shared genetic etiology and causality between COVID-19 and venous thromboembolism: evidence from genome-wide cross trait analysis and bi-directional Mendelian randomization study

Xin Huang [1†], Minhao Yao [2†], Peixin Tian [2], Jason Y.Y. Wong [3], Zhonghua Liu [2*], Jie V. Zhao [1*]

[1]Division of Epidemiology and Biostatistics, School of Public Health, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China

[2] Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China

[3]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA

[†] These authors contributed equally to this work

[*] Corresponding to Dr. Zhonghua Liu (Email: zhhliu@hku.hk, Tel: (852) 39170065; Fax: (852) 2858-9041) and Dr. Jie V. Zhao (Email: janezhao@hku.hk, Tel: (852) 39176739)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Venous thromboembolism (VTE) occurs in up to one third patients with COVID-19. VTE and COVID-19 may share a common genetic architecture, which has not been clarified yet. To fill this gap, we leveraged summary-level genetic data from the latest COVID-19 host genetics consortium and UK Biobank and examined their shared genetic etiology and causality. The cross-trait analysis identified 8, 11, and 7 shared loci between VTE and severe COVID-19, COVID-19 hospitalization, SARS-CoV-2 infection respectively, in 13 genes involved in coagulation and immune function and enriched in the lung. Co-localization analysis identified eight shared loci in *ABO*, *ADAMTS13* and *FUT2* genes. Bi-direction Mendelian randomization suggested that VTE was associated with higher risks of all COVID-19 related traits, and SARS-CoV-2 infection was associated with higher risk of VTE .Our study provided timely evidence and novel insights into the genetic etiology between COVID-19 and VTE.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

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-19[1]. People with more pronounced thrombotic symptoms are more likely to develop severe COVID-19 [2], and thrombotic complications are a well-established predictor of mortality in people with COVID-19 [1]. 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*[3-7], an established gene related to thrombosis. These 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 is hypothesized to lead to the systemic coagulation dysfunction [8-11]; 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 patients [12]. However, the recommendation is mainly based on observational studies, which might be subject to unmeasured confounding bias. To resolve this issue, Mendelian randomization study uses genetic variants as instrument for causal inference and can provide unbiased causal effect estimate even in the presence of unmeasured confounding[13]. One previous Mendelian randomization study in a relatively smaller GWAS for COVID-19 suggested that genetically predicted VTE is associated with higher risk of COVID-

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

19 hospitalization and SARS-COV-2 infection [14], 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 for VTE and three COVID-19 related traits to estimate their genetic correlations and shared genetic components using publicly available summary level data from GWASs of VTE and COVID-19. We also applied bi-directional Mendelian randomization methods to study the causal relationship between VTE and three COVID-19 related traits.

Methods

Study population

The GWAS summary statistics for COVID-19 of European ancestry were provided by the COVID-19 host genetics initiative round 7 (https://www.covid19hg.org/results/, release date: April 08, 2022). 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 (3,900 cases and 369,592 controls of European ancestry) from the UK Biobank. Participating cohorts of COVID-19 and VTE were described in detail in the Supplementary Note.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Linkage disequilibrium score regression (LDSC) analysis

LDSC analysis was conducted to assess the heritability for a single trait and genetic correlations between two traits. The analysis was conducted using the LDSC software based on the GWAS summary statistics. LD scores of 1000G European ancestry was used as reference [15]. An estimate of the heritability or genetic correlation can be 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 polygeneicity [16].

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 (MTAG)

We applied MTAG to identify novel loci with strong signals for COVID-19, and to detect shared genetic variants for VTE and three COVID-19 traits[18-20]. MTAG can improve the effect estimates for each COVID-19 trait by incorporating the weighted sum of GWAS estimates for VTE. We also used MTAG to conduct 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 association between individual variants for VTE and COVID-19.

Fine-mapping and co-localization analysis

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

For the significant shared loci between two traits, we used the Bayesian fine-mapping algorithm to identify a 99%-credible set of causal variants for each of the shared loci within 500kb of the index SNP [21]. Then, we conducted the co-localization analysis to check whether COVID-19 and VTE association signals co-localized at shared loci, by calculating the probability that two traits shared the same causal variant ($P(H_4)$) [22]. If $P(H_4)$ is greater than 0.5 for any SNP, we labelled it as a co-localized genetic variant.

Tissue-specific enrichment analysis (TSEA) and over-representation enrichment analysis

In order to test whether the shared genes are overly expressed in a specific tissue, we performed TSEA with the HUGO Gene Nomenclature Committee (HGNC) name of genes that correspond with the shared loci in cross-trait meta-analysis. 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 tissue[23]. We also used the WEB-based GeEne SeT AnaLysis Toolkit to further assess the overrepresented enrichment of the same shared genes in Gene ontology (GO) biological process [24].

Mendelian randomization (MR) analysis

We conducted a bidirectional MR analysis to assess the causal association between each COVID-19 related trait and VTE. Genetic instruments for VTE were obtained from a published genetic instruments published in a large GWAS of VTE ((23,151 cases, 553,439 controls) [25]. The GWAS provided 297 SNPs (P-value$<1\times10^{-5}$, $r^2<0.2$) as instrument for VTE[25]. 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, based on COVID-19 HGI release 7 GWAS meta-analysis we obtained 221, 282 and 177 SNPs as instruments for severe COVID-19, COVID-19 hospitalization and SARS-CoV-2 infection respectively. The genetic associations of these SNPs with VTE were obtained from the GWAS of VTE (3,900 cases and 369,592 controls of European ancestry) in the UK Biobank. We used inverse variance weighted (IVW) analysis in the main analysis [26], and used several other MR methods that are robust to pleiotropy,

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

including the weighted median [27], MR-Egger [28], MR-PRESSO [29] and MR-RAPS [30] in sensitivity analysis.

The details of methods were provided in the Supplemental Note.

Results

Genetic correlation of VTE with COVID-19 related traits

The heritability (h^2) estimated by LDSC 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 Figure 1 and Supplementary Table 2). The genetic correlation of VTE and severe COVID-19 and SARS-COV-2 were 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 5 (Fetal DHS, H3K4me1, H3K9ac H3K27ac and TFBS) (Figure 2), 11 (except for super enhancers regions) and 10 (except for conserved and super enhancers regions) of the 12 functional categories, respectively.

Multi-Trait Analysis of GWAS (MTAG)

Based on the HGI updated GWAS meta-analysis, there were 45 genome-wide significant ($P<5\times10^{-8}$) and uncorrelated ($r^2<0.01$) loci for severe COVID-19, 46 for COVID-19 hospitalization and 26 for SARS-CoV-2 infection (Supplementary Table 3), and 12 for VTE in the GWAS from UK Biobank (Supplementary Table 4). Using MTAG incorporating information from GWAS of VTE did not identify additional genome-wide significant loci for COVID-19 traits.

As shown in Table 1, we identified 8 shared novel genetic loci associated with both VTE and severe COVID-19, 11 with both VTE and COVID-19 hospitalization and 7 with both VTE and SARS-CoV-2 infection ($P_{meta} < 5\times10^{-8}$; single trait $P <0.05$). Figure 3 displays the Manhattan plots of these results. In line with previous studies [3-

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[6,31,32], we identified *ABO* gene and *FUT2* gene, which contributed to both VTE and COVID-19. Notably, we identified seven novel genes which have not been reported yet, including *LINC00970*, and six protein-coding genes (*TSPAN15*, *ADAMTS13*, *F5*, *DNAJB4*, *SLC39A8* and *OBSCN)*.

For severe COVID-19 with VTE, the strongest association signals were localized to the *LINC00970* gene (index SNP: rs114101204 for severe COVID-19 and VTE, $P_{meta}=1.44\times10^{-37}$) at locus 1q24.2. For COVID-19 hospitalization and SARS-CoV-2 infection with VTE, the strongest association signals were localized on or near the *ABO* gene (index SNP: rs11244061 for COVID-19 hospitalization, $P_{meta}=3.64\times10^{-28}$; rs550057 for SARS-CoV-2 infection, $P_{meta}=1.56\times10^{-51}$) at locus 9q34.2.

Fine-mapping and colocalization analysis identify shared causal variants

For each of the shared loci of VTE with COVID-19, Supplementary Tables 5-7 listed all SNPs within 500 kb of these variants in the 99% credible sets. Co-localization analysis shows that those colocalized genetic loci for VTE and severe COVID-19 were located in *ABO* (index SNPs: rs11244061 and rs138683771), *ADAMTS13* (index SNP: rs149181677), the colocalized loci for VTE and COVID-19 hospitalization were located in *ABO* (index SNPs: rs11244061and rs8176686) and *ADAMTS13* (index SNP: rs149181677), the colocalized loci for VTE and SARS-COV-2 infection were located in *ABO* (index SNPs: rs550057, rs9411367 and rs71503180) and *FUT2* (index SNP: rs492602) (Table 1).

GTEx tissue enrichment analysis and over-representation enrichment analysis of shared genes

The identified shared genes 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 (Figure 4). GO analysis highlighted several significant shared biological processes between three COVID-19 traits and VTE, such as “calcium-mediated signaling”, “second-

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

messenger-mediated signaling", "chemokine-mediated signaling pathway", "response to chemokine", "response to type I interferon" (Supplementary Table 8-10).

Mendelian randomization analysis

We found that genetically predicted VTE was positively associated with higher risk of all three COVID-19 related traits (OR= 1.08 for severe COVID-19, 95%CI: 1.04-1.12, P-value= 3.22×10^{-5}; OR=1.08 for COVID-19 hospitalization, 95%CI: 1.05-1.11, P-value =9.09×10^{-10}; OR=1.05 for SARS-CoV-2 infection, 95%CI: 1.04-1.07, P-value= 1.02×10^{-9}) (Figure 5). In the reverse direction, we found that a positive association of genetically predicted SARS-CoV-2 infection with higher risk of VTE (OR=1.80, 95%CI: 1.41-2.29, P-value= 1.98×10^{-6}), and null associations of genetically predicted severe COVID-19 or COVID-19 hospitalization with VTE (Figure 6). These results are robust to different MR methods (Supplementary Figures 1 and 2).

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. For the first time, we identified 8 shared loci for severe COVID-19, 11 loci for COVID-19 hospitalization and 7 loci for SARS-CoV-2 infection with VTE, with mapped genes involved in coagulation and immune function. Co-localization analysis identified eight shared genetic loci in *ABO*, *ADAMTS13* and *FUT2* genes. Enrichment analysis suggested that those gene expressions were enriched mainly 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.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Our study found genetic correlations between VTE and COVID-19. LDSC analysis showed 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 were 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 [17,33-35].

We identified shared genetic loci between VTE and COVID-19 using cross-trait meta-analysis and colocalization analysis. We identified shared loci in 13 genes (*LINC00970, ABO, ADAMTS13, TSPAN15, NAPSA, IFNAR2, F5, WNT3, DNAJB4, SLC39A8, OBSCN, SLC6A20, FUT2*) and co-localized genetic loci in three genes (*ABO, ADAMTS13* and *FUT2*), involved in coagulation and immune function. Consistent with previous report, we confirmed the role of *ABO, NAPSA, IFNAR2* in severe COVID-19 and COVID-19 hospitalization[3,4,7,36,37], *WNT3* in COVID-19 hospitalization[38], and *SLC6A20* in SARS-CoV-2 infection[4]. Notably, we found several novel genes, including *ADAMTS13, TSPAN15, WNT3, DNAJB4, SLC39A8, OBSCN*, which have not been reported yet. These novel findings provide new insights into genetic etiology of both VTE and COVID-19 related traits, and may help pinpoint therapeutic targets.

Interestingly, we found several genes related to coagulation. In cross-trait meta-analysis, we found a strong shared genetic signal for severe COVID-19 and VTE located at locus 1q24.2 in *LINC00970* close to *NME7* gene. Locus 1q24.2 has been found to be associated with VTE previously [39], but has not been linked to severe COVID-19. This locus (1q24.2) also includes shared genetic variant in *F5* gene. *F5* gene encodes plasma procoagulant factor (F)V, a coagulation factor involved in thrombin activation, and its impaired downregulation is key to thrombosis [40]. Similarly, *ABO*, *FUT2*, and *ADAMST13* are well-established VTE related genes [39,41], and were found in the colocalization analysis. Our findings are consistent with

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

previous genetic studies showing *ABO* gene is associated with severe COVID-19 and SARS-CoV-2 infection[3-5,31], possibly by regulating thrombosis[3,42]. 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 novel shared gene, *ADAMST13*, which encodes protein ADAMST13.. The function of ADAMTS13 might be affected by ABO blood group [43], and a previous MR study showed genetically predicted ADAMST13 is associated with severe COVID-19 [44].

Moreover, several shared genes are involved in immune function. The signals shared by severe COVID-19 and COVID-19 hospitalization with VTE were also mapped to the *IFNAR2* gene at 21q22.11 and *NAPSA* gene at 4q24. *IFNAR2* gene encodes interferon receptor subunit that mediates the early host immune response to viral infection [7]. *NAPSA* gene is associated with damage-associated transient progenitors promoted by inflammation [36,45]. Other shared genes have been known to be associated with innate antiviral defenses (*SLC6A20*) [3,46], immune cell signaling regulation (*WNT3, TSPAN15)*[47,48], inflammatory lung injury *(SLC39A8)*[49] or tumor activation (*TSPAN15, DNAJB4, OBSCN*)[50-52]. These results imply that immune function may be related to both COVID-19 and VTE.

The findings from enrichment analysis also 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 [53,54]. Calcium signaling is also of paramount importance in immune cells [55,56], and chemokine and interferon play a major role in activating host immune and inflammatory responses [57-59]. 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

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

infection, such as acute respiratory failure and ARDS [60,61] 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 it cannot be excluded 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 showing that genetically predicted VTE was associated with higher risk of COVID-19 hospitalization and SARS-CoV-2 infection [14]. 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.

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 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 (contemporary controls) and 5,859,411 (historical controls) individuals, people with SARS-COV-2 infection had higher risk of cardiovascular events, including VTE [62]. Nevertheless, a replication study of Mendelian randomization analysis in a larger GWAS of VTE will be worthwhile.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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 novel 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. Finally, although our study provides evidence of genetic correlation and genetic overlap between COVID-19 and VTE, the underlying biological mechanisms are still unclear, and further studies are still needed for validation.

In conclusion, our findings provided novel 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. Our work contributes to the understanding of COVID-19 and VTE etiology and provides new insights into the prevention and comorbidity management of COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

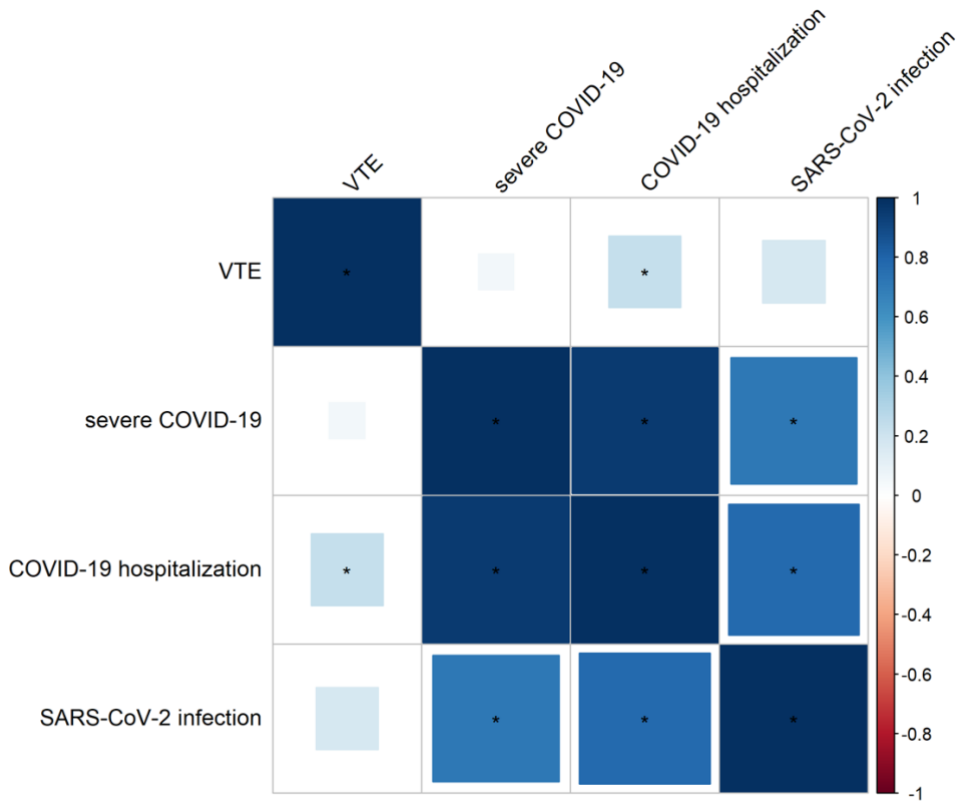

Figure 1 Genetic correlation of VTE with COVID-19 related traits

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

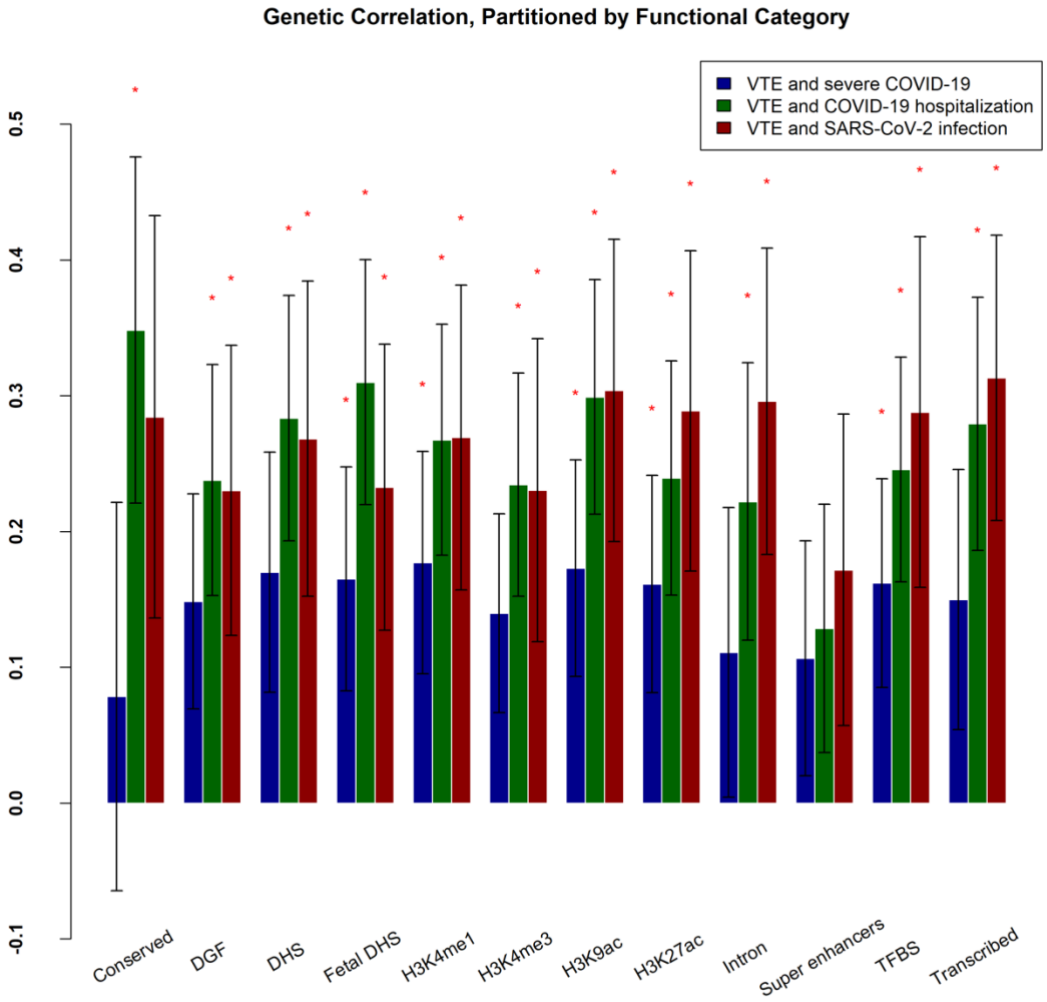

Figure 2 Partitioned genetic correlation between VTE and COVID-19 related traits

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. DGF: DNaseI digital genomic footprinting; DHS: DNase I hypersensitivity site; TFBS: transcription factor binding sites.

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	P_1	P_2	P_META	Coloc H4 Prob
Severe COVID-19 and VTE								
rs114101204	1	A	G	LINC00970	2.16E-47	3.39E-02	1.44E-37	2.25E-04
rs7726201	5	T	C	nan	2.28E-02	1.23E-07	2.17E-08	2.00E-02
rs11244061	9	T	C	ABO	2.66E-26	1.81E-04	3.55E-26	8.55E-01
rs149181677	9	T	C	ADAMTS13	1.39E-08	7.36E-04	2.07E-10	8.55E-01
rs138683771	9	C	A	ABO	6.77E-12	7.22E-03	1.42E-09	8.55E-01
rs75583483	10	G	A	TSPAN15	1.73E-15	4.23E-02	6.04E-14	3.19E-02
rs676314	19	G	A	NAPSA	4.44E-02	1.42E-09	8.13E-10	1.54E-02
rs2834166	21	A	C	IFNAR2	3.25E-02	4.74E-14	4.12E-13	2.33E-03
COVID-19 hospitalization and VTE								
rs114101204	1	A	G	LINC00970	2.16E-47	9.26E-03	1.65E-25	1.29E-03
rs9287090	1	A	G	F5	7.69E-18	2.93E-02	1.61E-11	1.98E-03
rs34517439	1	A	C	DNAJB4	4.50E-02	3.41E-09	4.96E-09	5.94E-03
rs7532924	1	A	C	OBSCN	2.84E-02	2.76E-07	2.97E-08	8.91E-02
rs13135092	4	G	A	SLC39A8	2.04E-02	1.49E-07	1.42E-08	4.94E-02
rs11244061	9	T	C	ABO	2.66E-26	8.70E-10	3.64E-28	9.00E-01
rs8176686	9	C	T	ABO	1.40E-09	1.63E-07	1.89E-14	9.00E-01
rs149181677	9	T	C	ADAMTS13	1.39E-08	7.13E-05	1.09E-10	9.00E-01
rs199501	17	G	A	WNT3	3.52E-02	1.45E-13	6.84E-10	8.26E-03
rs676314	19	G	A	NAPSA	4.44E-02	8.95E-15	4.10E-15	1.55E-02
rs2248202	21	A	C	IFNAR2	3.45E-02	4.45E-15	3.04E-15	2.32E-03

SARS-CoV-2 infection and VTE

rs76896797	3	T	C	SLC6A20	3.42E-02	4.63E-15	2.15E-11	3.60E-03
rs13107325	4	T	C	SLC39A8	2.69E-02	5.14E-08	9.54E-09	4.81E-02
rs550057	9	T	C	ABO	4.25E-38	4.05E-78	1.56E-51	9.84E-01
rs9411367	9	T	C	ABO	4.64E-05	8.09E-11	3.20E-12	9.84E-01
rs71503180	9	A	G	ABO	2.07E-04	6.81E-08	3.05E-10	9.84E-01
rs492602	19	G	A	FUT2	9.18E-06	7.26E-15	1.42E-10	8.18E-01
rs676314	19	G	A	NAPSA	4.44E-02	1.02E-08	3.54E-09	1.54E-02

CHR, chromosome; A1, other allele; A2, effect allele

P_1 represent the p-values in the GWAS of VTE. P_2 represent the p-values in the GWASs of COVID-19 related traits. P_META represents the p-value of the cross-trait meta-analysis.

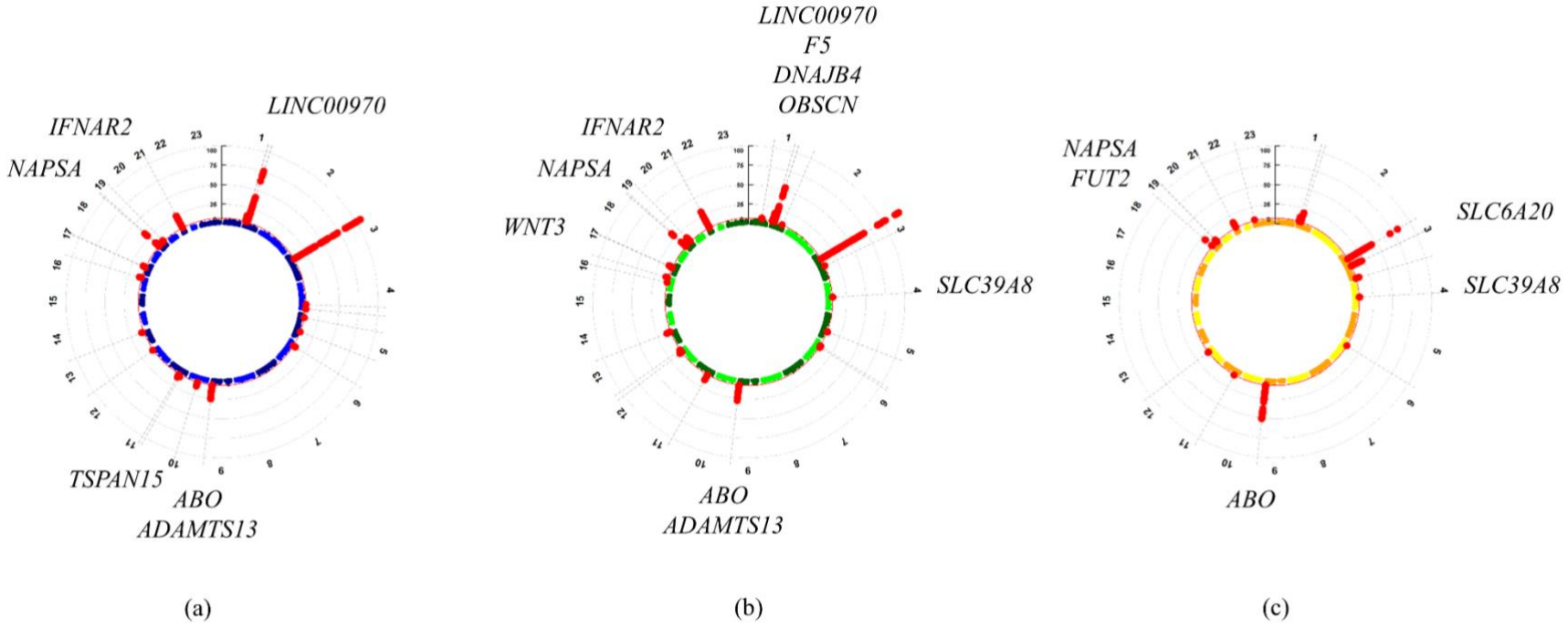

Figure 3 Circular Manhattan plot of cross-trait analysis for (a) VTE and severe COVID-19, (b) VTE and COVID-19 hospitalization, (c) VTE and SARS-CoV-2 infection

Each point represents a SNP, and shared significant loci with meta-analysis P-value < 5×10^{-8} and single-trait P-value < 0.05 are colored in red. SNPs are arranged according to the chromosome position.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

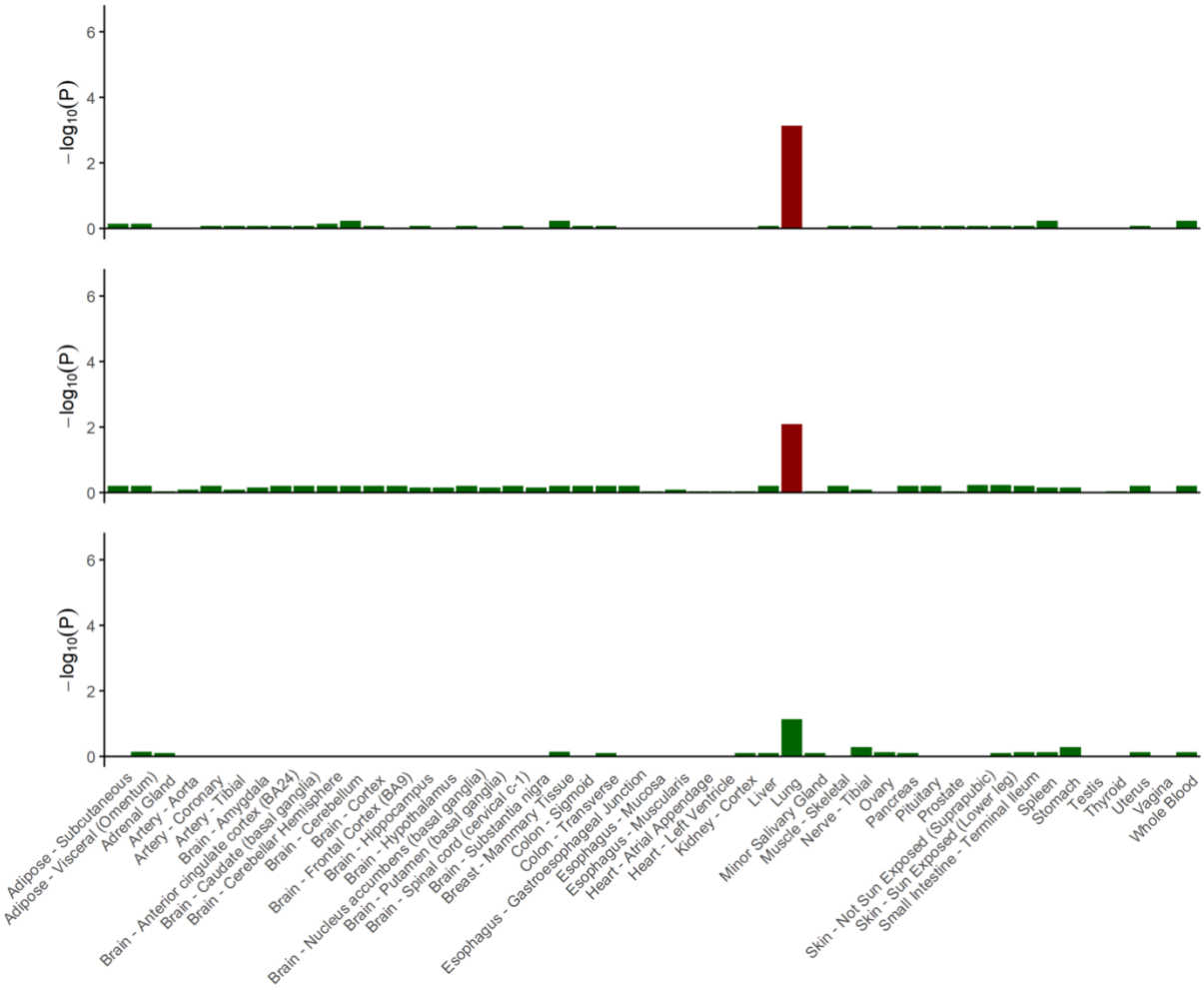

Figure 4 GTEx tissue enrichment analysis for expression of shared significant genes (meta-analysis P-value < 5×10^{-6}) for (a) VTE and severe COVID-19, (b) VTE and COVID-19 hospitalization, (c) VTE and SARS-CoV-2 infection

P-values of Fisher's exact test after Benjamin-Hochberg correction are presented in $-\log_{10}$ scale. Red 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).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

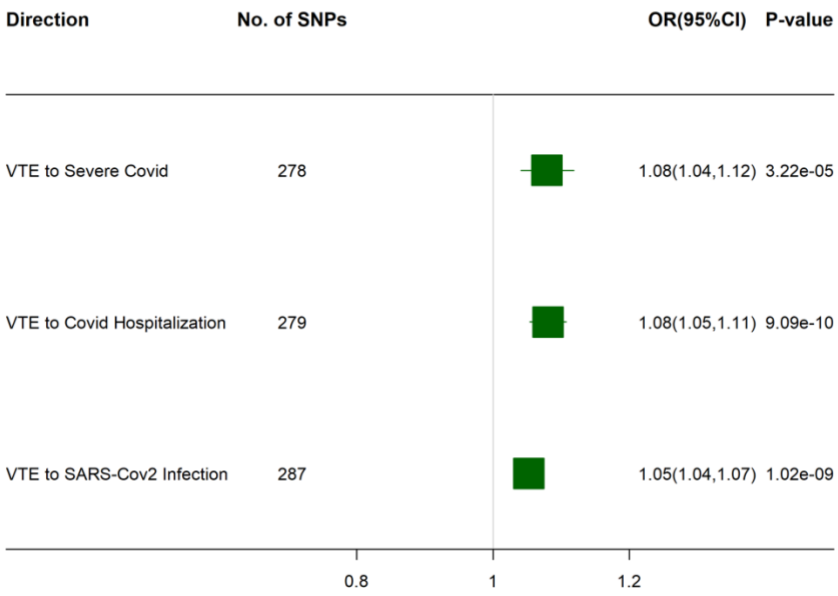

Figure 5 MR analysis of genetically predicted VTE on the risk of COVID-19 related traits using IVW method

Causal effect estimates are presented as odds ratios (OR) with 95% confidence intervals (CI).

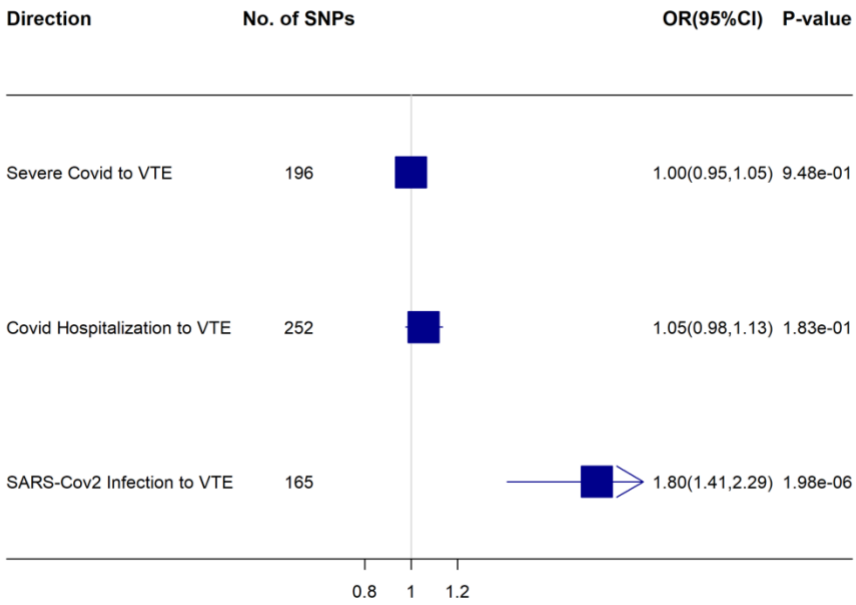

Figure 6 MR analysis of genetically predicted COVID-19 related traits on the risk of VTE using IVW method

Causal effect estimates are presented as odds ratios (OR) with 95% confidence intervals (CI). An arrow represents the estimate is out of boundary in this direction.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Statements and Declarations

Funding

The authors reported no funding received for this study.

Competing Interests

The authors have no competing interests.

Author Contributions

The manuscript was drafted by XH, with great help of JVZ and ZL (edited by JYYW). Data analysis were performed by MY and PT. XH and MY interpreted the results, with the help of JYYW, ZL and JVZ. ZL and JVZ developed the study conception, designed the analysis, and critically revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Ethics approval

The study is an analysis using publicly available summary data that does not require ethical approval.

Acknowledgments

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 the valuable data.

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 was obtained from COVID-19 host genetics consortium GWAS meta-analyses round 7, downloaded from https://www.covid19hg.org/results/r7/.

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Reference:

1. 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**(2020).
2. 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).
3. Severe Covid, G.G. *et al.* Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *The New England journal of medicine* **383**, 1522-1534 (2020).
4. Niemi, M.E.K. *et al.* Mapping the human genetic architecture of COVID-19. *Nature* **600**, 472-477 (2021).
5. 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).
6. Kousathanas, A. *et al.* Whole genome sequencing reveals host factors underlying critical Covid-19. *Nature* (2022).
7. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92-98 (2021).
8. Middeldorp, S. *et al.* Incidence of venous thromboembolism in hospitalized patients with COVID-19. *J Thromb Haemost* **18**, 1995-2002 (2020).
9. 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).
10. 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).
11. Mehta, P. *et al.* COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet* **395**, 1033-1034 (2020).
12. 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 advances*, bloodadvances.2021005945 (2021).
13. 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**, 1133-63 (2008).
14. 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*, 2021.03.20.21254008 (2021).
15. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
16. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics* **19**, 807-812 (2011).
17. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
18. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature genetics* **50**, 229-237 (2018).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

19. Chung, J. *et al.* Genome-wide association study of cerebral small vessel disease reveals established and novel loci. *Brain* **142**, 3176-3189 (2019).
20. Guo, Y. *et al.* A genome-wide cross-phenotype meta-analysis of the association of blood pressure with migraine. *Nature Communications* **11**, 3368 (2020).
21. Farh, K.K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-43 (2015).
22. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
23. Pei, G., Dai, Y., Zhao, Z. & Jia, P. deTS: tissue-specific enrichment analysis to decode tissue specificity. *Bioinformatics* **35**, 3842-3845 (2019).
24. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* **41**, W77-83 (2013).
25. Klarin, D. *et al.* Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nature genetics* **51**, 1574-1579 (2019).
26. Burgess, S., Butterworth, A. & Thompson, S.G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* **37**, 658-65 (2013).
27. Emdin, C.A., Khera, A.V. & Kathiresan, S. Mendelian Randomization. *Jama* **318**, 1925-1926 (2017).
28. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512-25 (2015).
29. 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).
30. 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. *The Annals of Statistics* **48**, 1742-1769, 28 (2020).
31. Roberts, G.H.L. *et al.* AncestryDNA COVID-19 Host Genetic Study Identifies Three Novel Loci. *medRxiv*, 2020.10.06.20205864 (2020).
32. Thibord, F. *et al.* Cross-Ancestry Investigation of Venous Thromboembolism Genomic Predictors. *medRxiv*, 2022.03.04.22271003 (2022).
33. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
34. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature Genetics* **45**, 124-130 (2013).
35. Vinson, C., Chatterjee, R. & Fitzgerald, P. Transcription factor binding sites and other features in human and Drosophila proximal promoters. *Sub-cellular biochemistry* **52**, 205-222 (2011).
36. Degenhardt, F. *et al.* New susceptibility loci for severe COVID-19 by detailed GWAS analysis in European populations. *medRxiv*, 2021.07.21.21260624 (2021).
37. Initiative, C.-H.G. & Ganna, A. Mapping the human genetic architecture of COVID-19: an update. *medRxiv*, 2021.11.08.21265944 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

38. Wu, L., Zhu, J., Liu, D., Sun, Y. & Wu, C. An integrative multiomics analysis identifies putative causal genes for COVID-19 severity. *Genetics in Medicine* **23**, 2076-2086 (2021).
39. Heit, J.A. *et al.* A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *Journal of thrombosis and haemostasis : JTH* **10**, 1521-1531 (2012).
40. Bos, M.H.A. & Camire, R.M. Blood coagulation factors V and VIII: Molecular Mechanisms of Procofactor Activation. *Journal of coagulation disorders* **2**, 19-27 (2010).
41. 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**, 1228-39 (2013).
42. Wang, L. *et al.* An atlas connecting shared genetic architecture of human diseases and molecular phenotypes provides insight into COVID-19 susceptibility. *Genome medicine* **13**, 83-83 (2021).
43. McGrath, R.T. *et al.* Expression of terminal alpha2-6-linked sialic acid on von Willebrand factor specifically enhances proteolysis by ADAMTS13. *Blood* **115**, 2666-73 (2010).
44. Zhou, Y. *et al.* Coagulation factors and the incidence of COVID-19 severity: Mendelian randomization analyses and supporting evidence. *Signal transduction and targeted therapy* **6**, 222-222 (2021).
45. Melms, J.C. *et al.* A molecular single-cell lung atlas of lethal COVID-19. *Nature* **595**, 114-119 (2021).
46. Vuille-dit-Bille, R.N. *et al.* Human intestine luminal ACE2 and amino acid transporter expression increased by ACE-inhibitors. *Amino Acids* **47**, 693-705 (2015).
47. Haseeb, M., Pirzada, R.H., Ain, Q.U. & Choi, S. Wnt Signaling in the Regulation of Immune Cell and Cancer Therapeutics. *Cells* **8**, 1380 (2019).
48. Cabañas, C., Yáñez-Mó, M. & van Spriel, A.B. Editorial: Functional Relevance of Tetraspanins in the Immune System. *Frontiers in Immunology* **10**(2019).
49. Besecker, B. *et al.* The human zinc transporter SLC39A8 (Zip8) is critical in zinc-mediated cytoprotection in lung epithelia. *Am J Physiol Lung Cell Mol Physiol* **294**, L1127-36 (2008).
50. Acun, T. *et al.* HLJ1 (DNAJB4) Gene Is a Novel Biomarker Candidate in Breast Cancer. *Omics : a journal of integrative biology* **21**, 257-265 (2017).
51. Rajendran, B.K. & Deng, C.-X. A comprehensive genomic meta-analysis identifies confirmatory role of OBSCN gene in breast tumorigenesis. *Oncotarget* **8**, 102263-102276 (2017).
52. Sidahmed-Adrar, N. *et al.* Tspan15 Is a New Stemness-Related Marker in Hepatocellular Carcinoma. *Proteomics* **19**, e1900025 (2019).
53. Nesbitt, W.S. *et al.* Intercellular calcium communication regulates platelet aggregation and thrombus growth. *The Journal of cell biology* **160**, 1151-1161 (2003).
54. Varga-Szabo, D., Braun, A. & Nieswandt, B. Calcium signaling in platelets. *J Thromb Haemost* **7**, 1057-66 (2009).
55. Vig, M. & Kinet, J.-P. Calcium signaling in immune cells. *Nature immunology* **10**, 21-27 (2009).

medRxiv preprint doi: https://doi.org/10.1101/2022.05.21.22275413; this version posted May 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

56. Trebak, M. & Kinet, J.-P. Calcium signalling in T cells. *Nature reviews. Immunology* **19**, 154-169 (2019).
57. Moser, B. & Willimann, K. Chemokines: role in inflammation and immune surveillance. *Annals of the Rheumatic Diseases* **63**, ii84 (2004).
58. McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nature Reviews Immunology* **15**, 87-103 (2015).
59. Charo, I.F. & Ransohoff, R.M. The Many Roles of Chemokines and Chemokine Receptors in Inflammation. *New England Journal of Medicine* **354**, 610-621 (2006).
60. 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).
61. Bonaventura, A. *et al.* Endothelial dysfunction and immunothrombosis as key pathogenic mechanisms in COVID-19. *Nature Reviews Immunology* **21**, 319-329 (2021).
62. Xie, Y., Xu, E., Bowe, B. & Al-Aly, Z. Long-term cardiovascular outcomes of COVID-19. *Nat Med* **28**, 583-590 (2022).