medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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.

Proteogenomic analyses identify coagulation factor XI as a thromboinflammatory mediator of long COVID

Art Schuermans[1,2,3,*], Andreas Verstraete[1,4], Vilma Lammi[5], Tomoko Nakanishi[6-10], Maddalena Ardissino[11,12], Jef Van den Eynde[1], Benjamin B. Sun[13], Marios K. Georgakis[2,14,15], Johan Van Weyenbergh[16], Adam J. Lewandowski[17], Betty Raman[18], Hanna M. Ollila[2,3,5,19], Stephen Burgess[20], Pradeep Natarajan[2,3,21] Michael C. Honigberg[2,3,21], Kathleen Freson[1], Thomas Vanassche[1,4], Peter Verhamme[1,4].

[1] KU Leuven Department of Cardiovascular Sciences, Center for Molecular and Vascular Biology, Leuven, Belgium

[2] Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[3] Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

[4] UZ Leuven Department of Cardiovascular Diseases, Leuven, Belgium

[5] Institute for Molecular Medicine Finland, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland

[6] Department of Human Genetics, McGill University, Montréal, QC, Canada

[7] Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, QC, Canada

[8] Kyoto-McGill International Collaborative Program in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[9] Department of Genome Informatics, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan

[10] Japan Society for the Promotion of Science, Tokyo, Japan

[11] BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[12] National Heart and Lung Institute, Imperial College London, London, UK

[13] Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

[14] Institute for Stroke and Dementia Research, Ludwig Maximilian University of Munich, Munich, Germany

[15] Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

[16] Laboratory of Clinical and Epidemiological Virology, Rega Institute for Medical Research, KU Leuven, Leuven, Belgium

[17] Nuffield Department of Population Health, University of Oxford, Oxford, UK

[18] Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

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/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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] Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, USA

[20] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

[21] Department of Medicine, Harvard Medical School, Boston, MA, USA

[*]Correspondence to:

Art Schuermans

KU Leuven Department of Cardiovascular Sciences

Center of Molecular and Vascular Biology

Herestraat 49 (box 911)

3000 Leuven, Belgium

Tel: +32 16 343491

Fax: +32 16 343493

Email: aschuerm@broadinstitute.org

Manuscript word count (excluding methods): 3,126

Abstract word count: 148

Figure/table count: 4

Reference count: 54

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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

SARS-CoV-2 infection can result in long COVID, characterized by post-acute symptoms from multiple organ systems. Current hypotheses on mechanisms underlying long COVID include persistent inflammation and dysregulated coagulation; however, precise mechanisms and causal mediators remain unclear. Here, we tested the associations of genetic instruments for 49 complement and coagulation factors from the *UK Biobank* (*N*=34,557) with long COVID in the *Long COVID Host Genetics Initiative* (*N*=997,600). Primary analyses revealed that genetically predicted higher factor XI increased long COVID risk (odds ratio, 1.17 [95% confidence interval, 1.08-1.27] per standard deviation; $P=1.7\times10^{-4}$). This association was robust to sensitivity analyses using pleiotropy-robust methods and different genetic instruments and was replicated using proteogenomic data from an Icelandic cohort. Genetically predicted factor XI was also associated with venous thromboembolism, but not with acute COVID-19 or long COVID-resembling conditions. Collectively, these findings provide genetic evidence implicating factor XI in the biology of long COVID.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

Main

Infection with SARS-CoV-2 leads to different acute clinical manifestations, ranging from mild respiratory disease to multi-organ dysfunction. It is becoming increasingly clear that infection with SARS-CoV-2 also has chronic effects, with epidemiological data suggesting that 5-10% of acutely infected individuals experience persistent complaints three months after infection[1]. These post-acute symptoms, also referred to as “long COVID”, are not limited to the respiratory tract but involve various organ systems[2]. Common manifestations include physical (e.g., fatigue, post-exertional malaise, and shortness of breath), cognitive (e.g., memory loss and cognitive impairment), and psychological (e.g., anxiety and depression) symptoms[3]. Up to 85% of long COVID patients report ongoing complaints even after one year[4,5]. Despite its high prevalence and impact on quality of life, clinical strategies for the prevention and treatment of long COVID remain limited.

The lack of disease-specific therapies for long COVID can be partially ascribed to an incomplete understanding of its molecular drivers. Currently hypothesized mechanisms underlying long COVID include immune dysregulation, virus-induced inflammation, and endothelial dysfunction[3]. For instance, in a prospective cohort study of 1,837 adults hospitalized with COVID-19, selected blood markers of coagulation (i.e., fibrinogen and D-dimer relative to C-reactive protein) predicted cognitive defects at 6-12 months after infection[6]. Similarly, a recent proteomic study of 113 individuals infected with SARS-CoV-2 revealed that the development of long COVID was associated with changes to blood proteins indicative of persistent complement activation and altered coagulation[7]. In the same study, circulating levels of factor XI (FXI) were elevated in those with compared to those without long COVID, pointing towards thromboinflammation as a possible mechanism contributing to long COVID development[7,8]. Despite these findings, it remains unclear whether the complement and coagulation pathways are causally involved in long COVID.

Recent work has identified genetic variants associated with plasma protein levels[9,10], facilitating the identification of causal mechanisms for human diseases using Mendelian randomization (MR)[11,12]. MR is an epidemiological method that uses the random assortment of alleles at conception, which leads to an effective randomization of individuals to higher or lower genetically predicted values for phenotypes such as changes in plasma protein levels, enabling causal inference[13,14]. This effective randomization limits the risk of bias due to confounding and reverse causation and can therefore provide support for a causal relationship between an exposure and outcome of interest[15,16]. Previous MR analyses have identified causal mediators of acute COVID-19[17,18], with clinical trial data confirming some of these (e.g., the interleukin-6 receptor) as effective drug targets in human patients[19,20]. Given the limitations and high costs associated with traditional methods for identifying drug targets,

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

genetic approaches such as MR may also help prioritize new causal biomarkers and therapeutic targets for long COVID[21].

Here, we used MR to test the hypothesis that the complement and coagulation systems contribute causally to long COVID. We constructed genetic instruments for plasma proteins involved in the complement and/or coagulation pathways and estimated their causal effects on the development of long COVID. Sensitivity analyses probed the robustness of the observed associations and replication analyses evaluated whether these associations persisted using separate datasets. Finally, we evaluated the effects of the identified proteins on different conditions to gain insights into the biological mechanisms behind the observed associations.

Results

Overall analysis approach

The study design is shown in **Fig. 1**. First, we identified all complement- and coagulation-related proteins using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database[22]. For each protein, we selected near-uncorrelated genetic variants (linkage disequilibrium R^2<0.1) associated with circulating levels of that protein ($P<1\times10^{-6}$) near the corresponding protein-encoding gene (within 200 kilobases). These genetic variants are also referred to as *cis*-protein quantitative trait loci (*cis*-pQTLs) and facilitate adherence to the core assumptions of MR [12,23,24] (**Fig. 1**; **Supplemental Methods**). For each protein, all *cis*-pQTLs were combined into one statistical entity, i.e., a genetic instrument (**Supplemental Methods**).

Once all genetic instruments were constructed, we tested the association of each protein's genetic instrument with long COVID. The present study used a two-sample MR approach, which uses genetic data from one genome-wide association study (GWAS) to construct a genetic instrument for the exposure of interest (e.g., the UK Biobank) and genetic data from a second, separate GWAS to test the association of this genetic instrument with the outcome of interest (e.g., the Long COVID Host Genetics Initiative)[15,25].

Characteristics of the included cohorts

Genetic instruments for circulating complement and coagulation factors were constructed in the UK Biobank. The characteristics of this cohort and its participants have been described previously[9]. In brief, genetic instruments were built using proteogenomic data from 34,557 participants with European ancestry of the 46,595 participants in total that were randomly selected for proteomic measurements. Of those, approximately 48.8% were male and the mean age was 56.7 years[9].

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

The associations of each protein's genetic instruments with long COVID was tested in the Long COVID Host Genetics Initiative, characteristics of which are provided in **Supplementary Table 1** and have been described previously[26]. While the Long COVID Host Genetics Initiative included 25 cohorts (cases: *n*=6,450; controls: *n*=1,093,995), primary analyses excluded participants who met the criteria for long COVID but did not have a prior test-verified SARS-CoV-2 infection and therefore only used data from 12 cohorts (cases: *n*=3,018; controls: *n*=994,582). Of the 12 included cohorts, the majority (83.3%) used questionnaires rather than electronic health data to ascertain long COVID (**Supplementary Table 1**). Approximately 99.1% of participants included in these cohorts had European ancestry, 48.0% were male, and the weighted mean age across cohorts was 62.3 years[26]. Around 30.5% of long COVID cases had a history of hospitalization due to COVID-19.

Construction of genetic instruments for complement- and coagulation-related proteins

A total of 69 candidate complement- and coagulation-related proteins were identified using the KEGG pathway database[22] (**Supplementary Table 2**). Of those, 49 (71.0%) were available and had at least one valid *cis*-pQTL (i.e., a genetic variant within 200 kilobases of the protein-encoding gene with $P<1\times10^{-6}$ and linkage disequilibrium $R^2<0.1$) based on proteogenomic data from up to 34,557 UK Biobank participants (**Supplementary Tables 2-3**). Variants used for each protein-specific genetic instrument are listed in **Supplementary Table 4**. All estimated *F*-statistics were >20, indicating low risk of weak instrument bias.

Genetic associations of complement- and coagulation-related proteins with long COVID

Primary analyses identified one protein that was genetically associated with long COVID at Bonferroni-adjusted $P<1.0\times10^{-3}$ considering 49 tested proteins (**Fig. 2**; **Supplementary Table 5**). Higher genetically predicted levels of plasma FXI were associated with higher odds of having long COVID (OR, 1.17 [95% CI, 1.08-1.27] per SD; $P=1.7\times10^{-4}$).

Although not reaching Bonferroni significance, another 14 proteins had nominally significant associations with long COVID. Higher genetically predicted levels of the factor II receptor (F2R, also known as PAR1), complement C1s subcomponent (C1S) and C1q subcomponent subunit A (C1QA), α-2-antiplasmin (SERPINF2), C4b-binding protein (β-chain; C4BPB), carboxypeptidase 2 (CPB2), and the urokinase plasminogen activator surface receptor (PLAUR) were nominally associated with higher odds of having long COVID. In contrast, higher genetically predicted levels of factor XII (F12), thrombomodulin (THBD), heparin cofactor 2 (SERPIND1), tissue factor pathway inhibitor (TFPI), complement factor I (CFI), complement factor C7 (C7), and C1 inhibitor (SERPING1) were nominally associated with lower odds of having long COVID.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

Genetic associations of FXI with long COVID in sensitivity and replication analyses

To evaluate the role of horizontal pleiotropy (i.e., effects of a genetic instrument on the outcome outside of its effects on the exposure) in the genetic association of circulating FXI with long COVID, we tested the association of genetically predicted plasma FXI with long COVID using MR-Egger (**Supplemental Fig. 1**). Analysis using MR-Egger suggested a pleiotropy-robust association of genetically predicted plasma FXI with long COVID (OR, 1.18 [95% CI, 1.04-1.34] per SD; P=8.1×10^{-3}), with a nonsignificant MR-Egger intercept term ($P_{intercept}$=0.84), suggesting that *cis*-pQTLs associated with FXI affect long COVID risk through plasma levels of FXI rather than through other mediators. We also performed sensitivity analyses using different linkage disequilibrium R^2 (R^2<0.001/R^2<0.01/R^2<0.1/R^2<0.2) and P-value (P<1×10^{-4}/P<1×10^{-6}/P<1×10^{-8}) thresholds to construct genetic instruments. The association of genetically predicted FXI with long COVID was directionally consistent across all tested instrument construction parameters (**Fig. 3**; **Supplementary Table 6**) and persisted when removing a single potentially reverse causal genetic variant using Steiger filtering (OR, 1.15 [95% CI, 1.07-1.24] per SD; P=2.5×10^{-4}).

Replication analyses tested whether the observed associations persisted against stricter criteria to define long COVID controls (restricting to proven controls; i.e., all participants with proven SARS-CoV-2 infection who did not meet long COVID criteria[26]) as well as a different proteomic profiling method (the aptamer-based SomaScan platform instead of Olink) in an external cohort including 35,559 Icelanders genotyped by deCODE[10]. Using stricter long COVID control criteria, FXI was associated with 1.15-fold odds of having long COVID (95% CI, 1.05-1.26; P=1.9×10^{-3}) per SD increase in genetically predicted levels. Using proteogenomic data from deCODE, genetically predicted FXI was associated with 1.17-fold odds of having long COVID (95% CI, 1.08-1.27; P=1.5×10^{-4}) per SD. Additional models using these replication datasets, as well as the aforementioned set of R^2 and P-value thresholds to construct genetic instruments, also yielded directionally consistent results (**Fig. 3**; **Supplementary Tables 7-8**).

We also performed colocalization analyses to test for distinct or shared causal genetic variants between plasma FXI and long COVID (see also **Supplementary Methods**). Considering the fact that colocalization methods are developed in the context of GWASs rather than *cis*-MR studies[27], and given previous analyses providing inconclusive colocalization evidence for the majority of strong and robust *cis*-MR associations[12], the main purpose of these analyses was to exclude the possibility that there were distinct causal variants. Reassuringly, there was no evidence for distinct causal variants between FXI and long COVID (posterior probability for distinct causal variants [H_3]: 7.5%; posterior probability for shared causal variants [H_4]: 1.7%).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

Association of predicted FXI levels with venous thromboembolism, bleeding, and long COVID-resembling phenotypes

Next, we tested the associations of genetically predicted FXI levels with bleeding and venous thromboembolism. In line with clinical evidence showing that FXI levels are associated with venous thromboembolism[28,29], higher genetically predicted FXI levels were associated with a higher risk of venous thromboembolism (OR, 1.71 [95% CI, 1.62-1.80] per SD; $P=5.6\times10^{-91}$), further supporting the use of *cis*-MR to identify causal associations of circulating FXI with clinical conditions. There was no genetic association of plasma FXI levels with risk of bleeding (OR, 0.99 [95% CI, 0.97-1.01] per SD; $P=0.42$).

In addition, to gain insights into the mechanisms driving the association of FXI with long COVID, we tested the associations of genetically predicted FXI levels with long COVID-resembling phenotypes. Genetically predicted levels of circulating FXI showed no association with acute COVID-19 (OR, 0.96 [95% CI, 0.90-1.03] per SD; $P=0.25$). This indicates that the genetic association between FXI levels and long COVID is not driven by an increased susceptibility to acute symptoms resulting from SARS-CoV-2 infection. Additionally, genetically predicted levels of FXI were not associated with a higher risk of long COVID-resembling conditions or symptoms such as memory loss or mood disorders (**Fig. 4**; **Supplementary Table 9**). Postviral fatigue was the only condition with an effect estimate within the 95% CI of long COVID (OR, 1.13 [95% CI, 0.88-1.44] per SD; $P=0.35$), although this association was likely underpowered due to a small number of cases in FinnGen's GWAS of postviral fatigue (cases: $n=195$; controls: $n=382{,}198$). These findings support the notion that the genetic association of plasma FXI with long COVID represents a specific association, wherein FXI plays a role in the pathophysiology of long COVID (and potentially other postviral conditions) rather than the occurrence of its individual, non-specific clinical symptoms.

Discussion

Using genetic data from multiple cohorts, we tested the associations of genetically predicted levels of different complement- and coagulation-related proteins with long COVID. Primary analyses identified an association of higher genetically predicted plasma levels of FXI with long COVID. The genetic association of FXI with long COVID was robust to multiple sensitivity analyses using pleiotropy-robust methods and different genetic instruments, and was replicated using data from an external Icelandic cohort. Additional analyses testing the genetic associations of plasma FXI with a range of different phenotypes did not reveal any associations with long COVID-resembling conditions (e.g., memory loss or mood disorders), suggesting a specific role for FXI in the pathophysiology of long COVID. Collectively, these

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

findings provide new mechanistic insights and identify a potential therapeutic target for long COVID.

Our findings suggest FXI is involved in the development of long COVID. FXI is a coagulation factor encoded by the *F11* gene that circulates as an inactive zymogen in the bloodstream until it is cleaved by the activated form of factor XII or thrombin[30]. The cleaved form of FXI activates factor IX, thereby propagating the intrinsic coagulation pathway, also known as the contact activation pathway, and amplifying the coagulation cascade. Epidemiological, preclinical, and clinical studies have established that higher FXI levels increase the risk of thrombosis, prompting the clinical evaluation of FXI inhibitors in phase 3 trials for the prevention of thromboembolic events in various clinical contexts[31,32]. In addition to its established role in promoting thrombosis, FXI has also been implicated in inflammatory and immune-related processes[30,33]. Recent evidence indicates that FXI can induce vascular inflammation without causing overt thrombosis[34]. Experimental data also suggest that pharmacological inhibition of FXI can mitigate inflammation and attenuate coagulopathy in mice with polymicrobial sepsis[35]. Similarly, inhibition of FXI in mice with experimental autoimmune encephalomyelitis leads to improved disease severity as well as reduced vascular disruption and immune cell migration[36]. These findings collectively suggest that FXI is an important thromboinflammatory mediator with effects that extend beyond coagulation, including vascular dysfunction and neuroinflammation.

Furthermore, consistent with the notion that thromboinflammation may contribute to long COVID, prior studies have suggested involvement of thrombosis-related pathways in the pathophysiology of acute COVID-19[37,38]. During the early stages of the pandemic, one in four COVID-19 patients admitted to the intensive care unit experienced venous thromboembolism despite standard dose thromboprophylaxis[39]. Mechanistic research suggests that this heightened risk of venous thromboembolism partially reflects a systemic COVID-19-associated coagulopathy involving immune response dysregulation, endothelial cell dysfunction, and hypercoagulability[38]. For instance, transcriptomic profiling of bronchoalveolar lavage fluid from patients with vs. without acute COVID-19 demonstrated differential regulation of genes encoding coagulation-related proteins (e.g., FXI)[40–42], suggesting involvement of these proteins in COVID-19-associated coagulopathy. In addition, selected blood markers of coagulation and inflammation (i.e., fibrinogen and D-dimer relative to C-reactive protein) have been shown to predict cognitive defects in patients hospitalized for COVID-19 at 6-12 months after infection[6]. Consistent with these findings, a recent study of 113 COVID-19 patients with proteomic profiling and longitudinal follow-up showed that participants who developed long COVID exhibited persistent complement activation (e.g., higher levels of C5bC6 and lower levels of C7-containing complexes) and altered coagulation (e.g., higher levels of von Willebrand factor and FXI) at 6 months after acute infection in

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

comparison with those who did not develop long COVID[7]. Some proteins such as von Willebrand factor and complement C5 did not receive genetic support from the present study, potentially because these proteins may represent markers of disease rather than being causal mediators or because they act downstream of other proteins or external factors with stronger causal effects. In contrast, circulating FXI, which was significantly higher in patients with vs. without long COVID[7], had a strong and robust genetic association with long COVID in the present study. Collectively, these findings suggest that thromboinflammation is involved in SARS-CoV-2-associated disease beyond acute infection and support the notion that FXI contributes to the development of long COVID. Future work should test pharmacological inhibition of FXI in animal models of long COVID and potentially human trials to establish clinical efficacy and evaluate when during the disease course patients would benefit the most from FXI inhibition.

Although this study benefitted from multiple large-scale and well-profiled datasets, as well as a robust framework for causal inference using human genetics, certain limitations must be considered when interpreting the study findings. First, we only examined 49 proteins involved in the complement and coagulation cascades. This targeted approach was chosen based on observational data showing differences in the levels of circulating complement- and coagulation-related proteins in individuals with vs. without long COVID[7] and considering the importance of triangulating evidence from different research approaches for causal inference using MR[15,43]. Nevertheless, there may be other proteins and mediators with causal roles in long COVID. Second, we only tested genetic instruments constructed using data from European-ancestry cohorts, limiting generalizability to other ancestries. Although more than 99% of participants from the Long COVID Host Genetics Initiative's GWAS had European ancestry, participants with other ancestries were also included, potentially leading to bias due to population stratification. Third, the cross-sectional nature of data collection for our analyses precluded longitudinal analyses evaluating the associations of genetically predicted FXI with long COVID in a longitudinal manner. It remains unclear if FXI mediates long COVID by aggravating the acute disease manifestations of COVID-19 or if it represents a separate mechanism that manifests during the months after acute infection. Fourth, any associations of genetically predicted FXI with viral persistence or burden after acute infection could not be evaluated. Fifth, analyses tested the association of genetically predicted FXI with postviral fatigue were underpowered and require further dedicated research. Sixth, because circulating protein levels were quantified as relative normalized protein expression units, derived effect estimates may not be representative of the absolute magnitude of effect. Assuming that plasma FXI activity is normally distributed in the general population with a mean of ~100% and a SD of ~20%[44], and that pharmacological inhibition of FXI leads to a reduction of FXI activity to ~0%[28], extrapolation of our findings (i.e., an OR of 1.17 for long

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

COVID per SD increase in genetically predicted FXI levels) suggests that pharmacological reduction of FXI activity from 100% to 0% could be associated with an OR of 0.46 for long COVID. Nevertheless, this extrapolation remains speculative and true effect sizes remain to be established in intervention trials. Finally, while MR can be used to infer causality in certain exposure-outcome relationships, any causal inference using MR relies on justification of the different MR assumptions. Our study used a *cis*-MR framework facilitating adherence to these assumptions[11,12], included different sensitivity analyses to probe the robustness of the study findings, and replicated these using data from an external cohort. Nevertheless, genetic associations must be validated in randomized controlled trials for clinical translation.

In summary, this study used genetic data from multiple cohorts to investigate the relationship between the complement and coagulation systems and long COVID. Primary analyses identified a strong association of circulating FXI with long COVID risk, which was corroborated using different sensitivity analyses and replicated using data from an external cohort. These findings suggest that FXI is causally involved in long COVID, which has potential implications for the development of therapeutic interventions. Further research is needed to elucidate other molecular mechanisms and explore the effectiveness and safety of targeting circulating FXI for the prevention and/or treatment of long COVID.

Methods

Cis-MR and instrumental variable assumptions

MR analyses can be used to infer causality in certain exposure-outcome relationships, given that these analyses adheres to three key instrumental variable assumptions[15] (**Fig. 1**). First, the tested genetic instrument must comprise genetic variants that are associated with the exposure of interest (i.e., relevance assumption). Second, no unmeasured confounders can affect the association between the genetic variants comprising the tested genetic instrument and the outcome (i.e., independence assumption). Third, there can be no horizontal pleiotropy in the association of the tested genetic instrument with the outcome, meaning that the tested genetic instrument can only affect the outcome through its effects on the exposure of interest (i.e., exclusion restriction assumption).

The present study used a *cis*-MR approach, which tests genetic instruments constructed using variants associated with circulating protein levels that map near the corresponding protein-encoding gene (i.e., *cis*-pQTLs). The use of *cis*-pQTLs has beneficial effects on the adherence to the instrumental variable assumptions[23,24]. Variants located near the protein-encoding gene often have large effects on protein expression in comparison to other traits and are less prone to violating the “no horizontal pleiotropy” assumption than variants located elsewhere in the genome[23,24]. In addition, MR analyses using *cis*-pQTLs reduce the risk of reverse causation because it is more likely that *cis*-pQTLs influence protein

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

levels, which then influence a disease, rather than the disease influencing protein levels through *cis*-pQTLs.

GWAS of circulating complement- and coagulation-related proteins

Complement- or coagulation-related proteins were defined as those included in the "*Complement and coagulation cascades*" gene set from the KEGG database[22] (**Supplementary Table 2**). Genetic association data were obtained for plasma levels of 49 complement- or coagulation-related proteins measured in the UK Biobank[9,45] (**Supplementary Table 3**). GWASs for circulating protein levels included up to 34,557 UK Biobank participants of European ancestry[9]. Blood samples donated by these participants were analyzed for proteomic profiling using the Olink Explore 3072 platform (Olink Proteomics, Inc; Waltham, MA), which uses proximity extension assay technology to quantify protein levels. Additional information on the UK Biobank and circulating protein GWASs can be found in the **Supplementary Methods**.

GWAS of long COVID

Genetic association data for long COVID were obtained from the Long COVID Host Genetics Initiative[26]. Data from a total of 3,018 long COVID cases and 994,582 controls from 11 individual studies were included for primary analyses (**Supplementary Table 1**). Consistent with the genetic ancestry of the exposure cohort, more than 99% of participants had European ancestry. Because the UK Biobank cohort used for primary analyses only included 34,557 participants, the maximal possible sample overlap between the exposure and outcome cohorts was ~3.5%.

Long COVID cases and controls were defined as reported previously[26]. In brief, study participants were classified as long COVID cases if they had a prior test-verified SARS-CoV-2 infection and met the long COVID criteria for a minimum of three months after their SARS-CoV-2 infection or the onset of COVID-19. As reported previously, and aligning with the World Health Organization guidelines[2], individuals with long COVID criteria were defined as those who (1) reported presence of COVID-19 symptoms not attributable to other diagnoses; (2) reported persistent significant limitations in daily activities; or (3) had any diagnosis codes of long COVID in their electronic health records. All cohort-specific criteria to ascertain long COVID are shown in **Supplementary Table 1**.

Primary analyses used population controls to increase statistical power and limit selective inclusion of certain groups within the population, which can introduce collider bias in MR studies. Population controls were defined as those who were not identified as having long COVID using the above-mentioned criteria[26]. Replication analyses used a stricter control definition (i.e., all individuals with a history of SARS-CoV-2 infection but without long

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

COVID criteria [cases: *n*=2,975; controls: *n*=37,935]). Additional details on the long COVID case and control phenotypes, including diagnosis codes and study-specific criteria, can be found in the **Supplementary Methods** and have been described previously[26].

Cis-MR analyses

A full description of genetic instrument construction and MR methods can be found in the **Supplementary Methods**. Genetic instruments for circulating complement- and coagulation-related proteins were constructed using *cis*-pQTLs within 200 kilobases of each protein's encoding gene. Consistent with prior *cis*-MR analyses[11,12,46], all genetic instruments included sub-genome-wide significant ($P<1\times10^{-6}$) and near-uncorrelated (linkage disequilibrium $R^2<0.1$) *cis*-pQTLs. MR analyses were performed using the inverse-variance weighted method accounting for between-variant correlation structure or the Wald ratio, depending on the number of *cis*-pQTLs per genetic instrument[47,48].

Associations were presented as odds ratios (ORs) per standard deviation (SD) increase in genetically predicted protein levels, with statistical significance set at Bonferroni-corrected $P<1.0\times10^{-3}$. The *TwoSampleMR*[49] and *MendelianRandomization*[50] packages in R (version 4.1.0) were used for analyses.

Sensitivity analyses

All proteins with a Bonferroni-significant association in primary analyses were carried forward for sensitivity analyses using different MR approaches and genetic instrument selection parameters. First, to evaluate the possibility of horizontal pleiotropy (i.e., effects of a genetic instrument on the target outcome beyond its effects on the exposure of interest) affecting the observed associations, we performed sensitivity analyses using the pleiotropy-robust MR-Egger method accounting for between-variant correlation structure[47,48]. MR-Egger is an MR method used to test and correct for horizontal pleiotropy within a genetic instrument at the cost of lower precision and statistical power[51]. Second, because *cis*-MR analyses often rely on *cis*-pQTLs with residual correlation[52], additional sensitivity analyses were conducted using genetic instruments constructed with various linkage disequilibrium R^2 thresholds ($R^2<0.001$/$R^2<0.01$/$R^2<0.1$/$R^2<0.2$). Third, as the instruments used in primary analyses were built using sub-genome-wide significant *cis*-pQTLs, genetic associations were further corroborated using various *P*-value thresholds ($P<1\times10^{-4}$/$P<1\times10^{-6}$/$P<1\times10^{-8}$). Fourth, we also performed Steiger filtering, which removes *cis*-pQTLs that explain more variation in the outcome than in the exposure of interest[53], to evaluate the possibility of reverse causation affecting the observed results. Fifth, we performed colocalization analyses to exclude the possibility that there were distinct causal variants between the *cis* loci of the prioritized proteins and long COVID (**Supplementary Methods**).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

Consistent with prior work[12], protein-disease associations were considered robust if (1) the genetic association was Bonferroni-significant in primary analyses, (2) MR-Egger suggested no horizontal pleiotropy ($P \geq 0.05$ for the intercept test or $P < 0.05$ for the intercept test with $P < 0.05$ for the causal test), and (3) MR estimates had a consistent direction of effect across all primary and sensitivity analyses.

Replication analyses

Robust protein-disease associations were carried forward for replication analyses. First, we evaluated whether the observed associations persisted when we used stricter criteria for controls in the Long COVID Host Genetics Initiative GWAS[26]. Unlike our primary analyses, which used population controls (i.e., all study participants without long COVID criteria; n=994,582), these replication analyses only included proven controls (i.e., all study participants with proven SARS-CoV-2 infection but without long COVID criteria; n=37,935). Second, we evaluated whether the observed associations remained consistent using an external proteomics dataset. These analyses used genetic association data from an Icelandic cohort including 35,559 participants who had their circulating proteome profiled using the aptamer-based SomaScan assay (version 4; SomaLogic, Inc.; Boulder, CO) at deCODE (deCODE genetics, Inc.; Reykjavík, Iceland)[10]. Additional details on the deCODE cohort can be found in the **Supplementary Methods**. Two-sided $P < 0.05$ indicated statistical significance for replication analyses.

Cis-MR analyses of bleeding, venous thromboembolism, and long COVID-resembling phenotypes

We tested the genetic associations of all complement and coagulation factors surviving sensitivity analyses with bleeding, venous thromboembolism, acute COVID-19, and a range of “long COVID-resembling” phenotypes. These long COVID-resembling phenotypes included various conditions sharing similar symptoms—yet likely distinct mechanisms—to those observed in long COVID[3]. The phenotypes were chosen based on prior evidence suggesting overlapping symptoms with long COVID[3] and were tested if they had GWAS data available in FinnGen (release 10)[54]. Information on the tested long COVID-resembling phenotypes can be found in **Supplementary Table 10**.

Genetic association data for these phenotypes were obtained from FinnGen (release 10)[54], which includes 412,181 participants with European ancestry. Additional information on the tested phenotypes and FinnGen can be found in the **Supplementary Methods** and **Supplementary Table 10**.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

References

1. Vos, T. *et al.* Estimated Global Proportions of Individuals With Persistent Fatigue, Cognitive, and Respiratory Symptom Clusters Following Symptomatic COVID-19 in 2020 and 2021. *JAMA* **328**, 1604–1615 (2022).
2. Soriano, J. B., Murthy, S., Marshall, J. C., Relan, P. & Diaz, J. V. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis* **22**, e102–e107 (2022).
3. Davis, H. E., McCorkell, L., Vogel, J. M. & Topol, E. J. Long COVID: major findings, mechanisms and recommendations. *Nat Rev Microbiol* **21**, 133–146 (2023).
4. Seeble, J. *et al.* Persistent Symptoms in Adult Patients 1 Year After Coronavirus Disease 2019 (COVID-19): A Prospective Cohort Study. *Clin Infect Dis* **74**, 1191–1198 (2022).
5. Tran, V. T., Porcher, R., Pane, I. & Ravaud, P. Course of post COVID-19 disease symptoms over time in the ComPaRe long COVID prospective e-cohort. *Nat Commun* **13**, 1–6 (2022).
6. Taquet, M. *et al.* Acute blood biomarker profiles predict cognitive deficits 6 and 12 months after COVID-19 hospitalization. *Nat Med* **29**, 2498–2508 (2023).
7. Cervia-Hasler, C. *et al.* Persistent complement dysregulation with signs of thromboinflammation in active Long Covid. *Science (1979)* **383**, 1–18 (2024).
8. Ruf, W. Immune damage in Long Covid. *Science (1979)* **383**, 262–263 (2024).
9. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
10. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet* **53**, 1712–1721 (2021).
11. Henry, A. *et al.* Therapeutic Targets for Heart Failure Identified Using Proteomics and Mendelian Randomization. *Circulation* **145**, 1205–1217 (2022).
12. Schuermans, A. *et al.* Genetic Associations of Circulating Cardiovascular Proteins With Gestational Hypertension and Preeclampsia. *JAMA Cardiol* (2024) doi:10.1001/JAMACARDIO.2023.4994.
13. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. *Genet Epidemiol* **37**, 658–665 (2013).
14. Ardissino, M. *et al.* Birth weight influences cardiac structure, function, and disease risk: evidence of a causal association. *Eur Heart J* (2023) doi:10.1093/EURHEARTJ/EHAD631.
15. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**, 601 (2018).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

16. Sanderson, E. *et al.* Mendelian randomization. *Nature Reviews Methods Primers* **2**, 1–21 (2022).
17. Gaziano, L. *et al.* Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat Med* **27**, 668–676 (2021).
18. Bovijn, J., Lindgren, C. M. & Holmes, M. V. Genetic variants mimicking therapeutic inhibition of IL-6 receptor signaling and risk of COVID-19. *Lancet Rheumatol* **2**, e658–e659 (2020).
19. Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19. *N Engl J Med* **384**, 1491–1502 (2021).
20. Group, T. W. R. E. A. for C.-19 T. (REACT) W. *et al.* Association Between Administration of IL-6 Antagonists and Mortality Among Patients Hospitalized for COVID-19: A Meta-analysis. *JAMA* **326**, 499–518 (2021).
21. Trajanoska, K. *et al.* From target discovery to clinical drug development with human genetics. *Nature* **620**, 737–745 (2023).
22. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
23. Schmidt, A. F. *et al.* Genetic drug target validation using Mendelian randomisation. *Nat Commun* **11**, 1–12 (2020).
24. Swerdlow, D. I. *et al.* Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol* **45**, 1600–1616 (2016).
25. Pierce, B. L. & Burgess, S. Efficient Design for Mendelian Randomization Studies: Subsample and 2-Sample Instrumental Variable Estimators. *Am J Epidemiol* **178**, 1177–1184 (2013).
26. Lammi, V. *et al.* Genome-wide Association Study of Long COVID. *medRxiv* 2023.06.29.23292056 https://www.medrxiv.org/content/10.1101/2023.06.29.23292056v1 (2023) doi:10.1101/2023.06.29.23292056.
27. Zuber, V. *et al.* Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. *Am J Hum Genet* **109**, 767–782 (2022).
28. Verhamme, P. *et al.* Abelacimab for Prevention of Venous Thromboembolism. *N Engl J Med* **385**, 609–617 (2021).
29. Weitz, J. I. *et al.* Milvexian for the Prevention of Venous Thromboembolism. *N Engl J Med* **385**, 2161–2172 (2021).
30. Moellmer, S. A., Puy, C. & McCarty, O. J. T. Biology of factor XI. *Blood* **143**, 1445–1454 (2024).
31. De Caterina, R., Prisco, D. & Eikelboom, J. W. Factor XI inhibitors: cardiovascular perspectives. *Eur Heart J* **44**, 280–292 (2023).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

32. Verstraete, A., Engelen, M. M., Van Edom, C., Vanassche, T. & Verhamme, P. Reshaping Anticoagulation: Factor XI Inhibition in Thrombosis Management. *Hamostaseologie* **44**, 49–58 (2023).
33. Kohs, T. C. L. *et al.* Activation of coagulation FXI promotes endothelial inflammation and amplifies platelet activation in a nonhuman primate model of hyperlipidemia. *Res Pract Thromb Haemost* **8**, 102276 (2024).
34. Kossmann, S. *et al.* Platelet-localized FXI promotes a vascular coagulationinflammatory circuit in arterial hypertension. *Sci Transl Med* **9**, (2017).
35. Tucker, E. I. *et al.* Inhibition of factor XI activation attenuates inflammation and coagulopathy while improving the survival of mouse polymicrobial sepsis. *Blood* **119**, 4762–4768 (2012).
36. Kohs, T. C. L. *et al.* Pharmacological targeting of coagulation factor XI attenuates experimental autoimmune encephalomyelitis in mice. *Metab Brain Dis* **38**, 2383–2391 (2023).
37. Gorog, D. A. *et al.* Current and novel biomarkers of thrombotic risk in COVID-19: a Consensus Statement from the International COVID-19 Thrombosis Biomarkers Colloquium. *Nat Rev Cardiol* **19**, 475–495 (2022).
38. Conway, E. M. *et al.* Understanding COVID-19-associated coagulopathy. *Nat Rev Immunol* **22**, 639–649 (2022).
39. Klok, F. A. *et al.* Incidence of thrombotic complications in critically ill ICU patients with COVID-19. *Thromb Res* **191**, 145–147 (2020).
40. FitzGerald, E. S., Chen, Y., Fitzgerald, K. A. & Jamieson, A. M. Lung epithelial cell transcriptional regulation as a factor in COVID-19–associated coagulopathies. *Am J Respir Cell Mol Biol* **64**, 687–697 (2021).
41. Mast, A. E. *et al.* SARS-CoV-2 suppresses anticoagulant and fibrinolytic gene expression in the lung. *Elife* **10**, (2021).
42. Martens, C. P. *et al.* Dysregulation of the kallikrein-kinin system in bronchoalveolar lavage fluid of patients with severe COVID-19. *EBioMedicine* **83**, 104195 (2022).
43. Lawlor, D. A., Tilling, K. & Smith, G. D. Triangulation in aetiological epidemiology. *Int J Epidemiol* **45**, 1866–1886 (2016).
44. Meijers, J. C. M., Tekelenburg, W. L. H., Bouma, B. N., Bertina, R. M. & Rosendaal, F. R. High Levels of Coagulation Factor XI as a Risk Factor for Venous Thrombosis. *N Engl J Med* **342**, 696–701 (2000).
45. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
46. Schuermans, A. *et al.* Integrative proteomic analyses across common cardiac diseases yield new mechanistic insights and enhanced prediction. *medRxiv*

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

2023.12.19.23300218 https://www.medrxiv.org/content/10.1101/2023.12.19.23300218v1 (2023) doi:10.1101/2023.12.19.23300218.

47. Burgess, S., Zuber, V., Valdes-Marquez, E., Sun, B. B. & Hopewell, J. C. Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. *Genet Epidemiol* **41**, 714 (2017).

48. Burgess, S., Dudbridge, F. & Thompson, S. G. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med* **35**, 1880–1906 (2016).

49. Hemani, G. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).

50. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* **46**, 1734–1739 (2017).

51. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol* **32**, 377 (2017).

52. Gkatzionis, A., Burgess, S. & Newcombe, P. J. Statistical methods for cis-Mendelian randomization with two-sample summary-level data. *Genet Epidemiol* **47**, 3–25 (2023).

53. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* **13**, e1007081 (2017).

54. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

Acknowledgements

All authors want to acknowledge the participants and investigators of the FinnGen study, the UK Biobank, and the deCODE genetics cohort. T.N. is supported by a research fellowship of the Japan Society for the Promotion of Science for Young Scientists (22J30004) and by a Grant-in-Aid for Scientific Research (B) (23H02917). M.K.G. is funded by the German Research Foundation in the form of an Emmy Noether grant (GZ: GE 3461/2-1, ID 512461526). B.R. is supported by a Wellcome Career Development Award fellowship (302210/Z/23/Z). H.M.O. is supported by the Academy of Finland (#1350181) and the NIH (R01AI170850). K.F. is supported by the National Institute of Health (R01HI161365), KU Leuven (C14/23/121), research foundation Flanders (FWO; G072921N). T.V. is supported by a grant from FWO (1843423N).

Ethics declarations

M.K.G. has consulted for Tourmaline Bio and serves in the Editorial Board of Neurology, both not relevant to this work. M.C.H. reports research funding from Genentech, consulting fees from Comanche Biopharma, and advisory board service for Miga Health, all unrelated to the present work. K.F. received research grants from Novo Nordisk and Swedish Orphan Biovitrum AB (Sobi). T.V. has participated in advisory boards and/or acted as a speaker on behalf of Bayer, BMS/Pfizer, Boehringer Ingelheim, Daiichi Sankyo, Leo Pharma, Sanofi Aventis. P.V. has received research funding from Bayer, BMS, Pfizer, and Leo Pharma, and honoraria from Bayer, Pfizer, BMS, Daiichi-Sankyo, Sanofi-Aventis, Leo Pharma, Anthos Therapeutics, and Astra-Zeneca.

Data availability

This study used publicly available summary statistics from genome-wide association studies performed in the UK Biobank (https://biobank.ndph.ox.ac.uk/showcase/; http://ukb-ppp.gwas.eu/) the Long COVID Host Genetics Initiative (https://my.locuszoom.org/gwas/793752/; https://www.medrxiv.org/content/10.1101/2023.06.29.23292056v1), an Icelandic cohort genotyped by deCODE (https://www.decode.com/summarydata/), and FinnGen (release 10; https://r10.finngen.fi/).

Code availability

Code used for the main analyses of this study will be published upon publication of the manuscript.

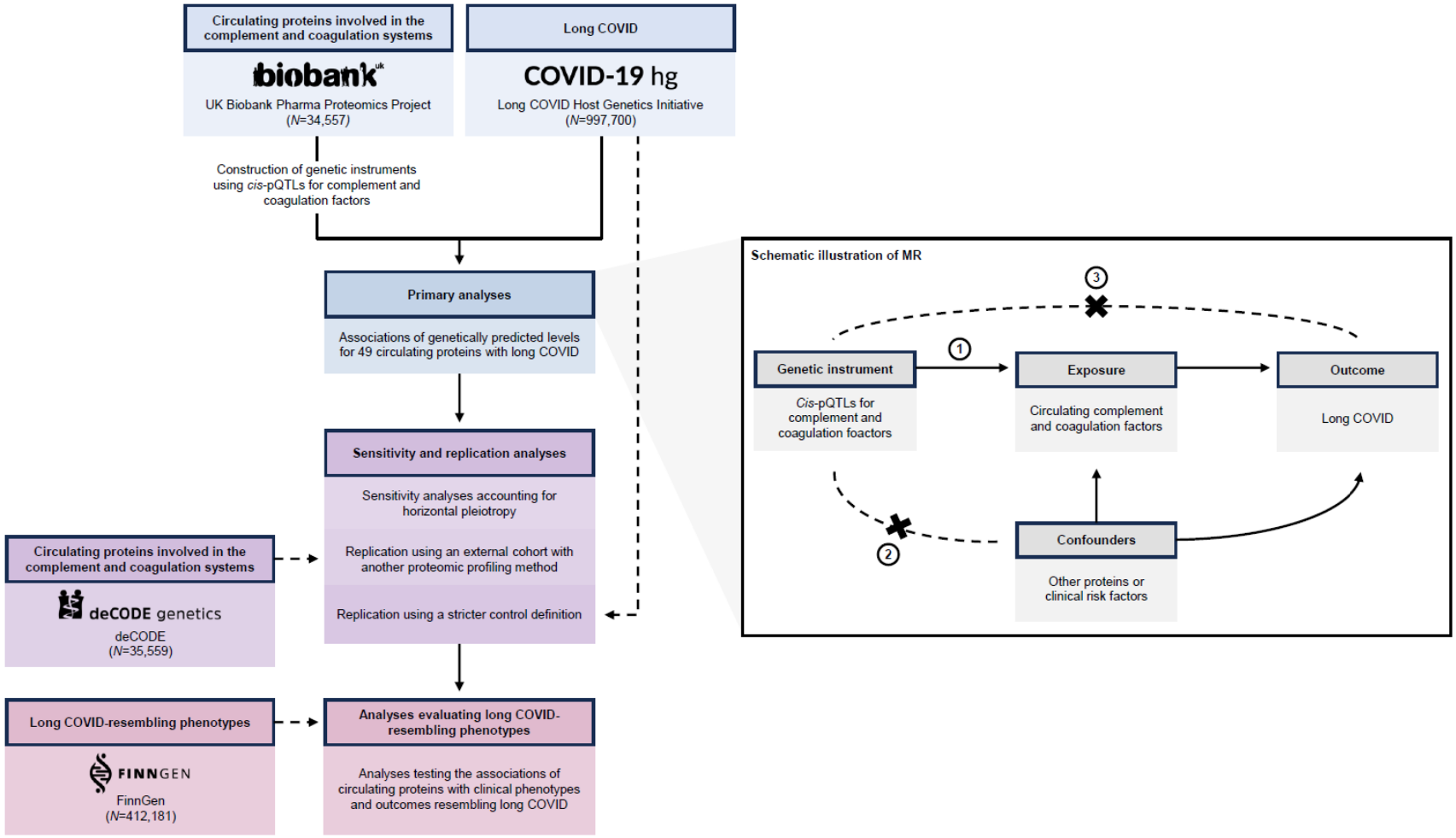
Inclusion and ethics

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

590 Inclusion and ethics standards have been reviewed where applicable.

Figures

Fig. 1. Study design.

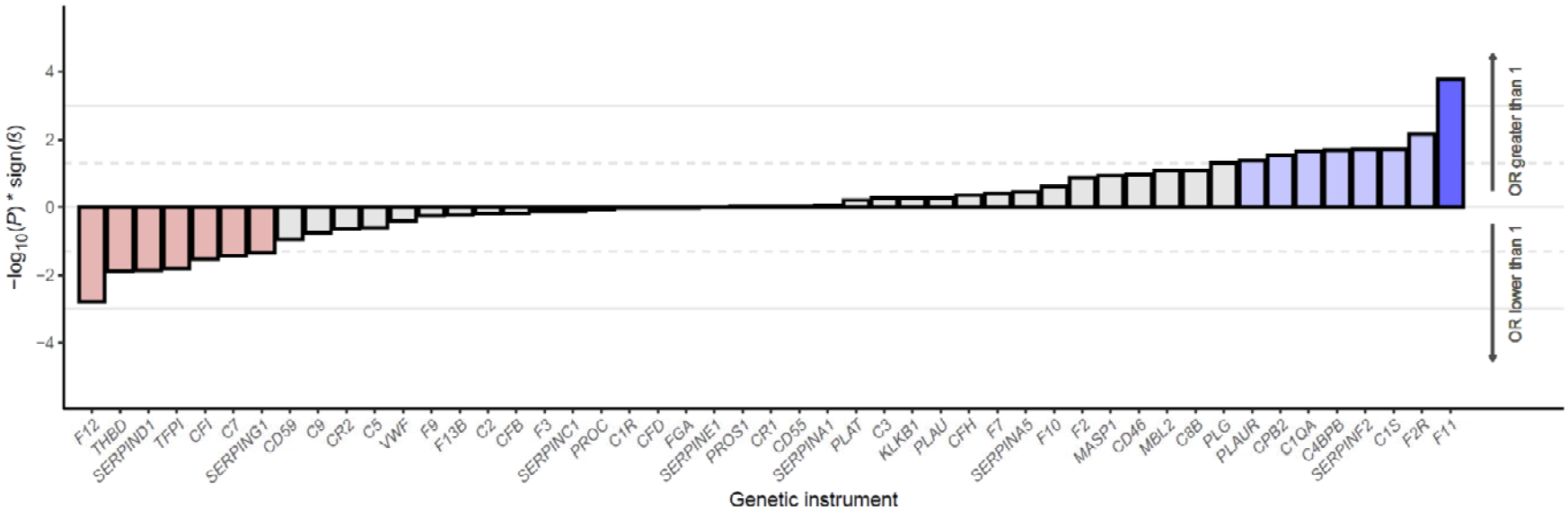


medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The flow chart shows the study design and the data sources. The diagram illustrates the design and three core assumptions of Mendelian randomization (MR) analyses: (1) the tested genetic instrument must be associated with the exposure of interest; (2) there must be no confounders affecting the association between the tested genetic instrument and outcome of interest; (3) the genetic instrument can only affect the outcome through its effect on the exposure of interest (i.e., no horizontal pleiotropy). The use of genetic variants that are associated with circulating protein levels and are located near the corresponding protein-encoding gene, also known as *cis*-protein quantitative trait loci (*cis*-pQTLs), facilitates adherence to these assumptions, as discussed in the **Supplementary Methods**.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

Fig. 2. Primary analyses testing the associations of genetically predicted circulating complement- and coagulation-related protein levels with long COVID.

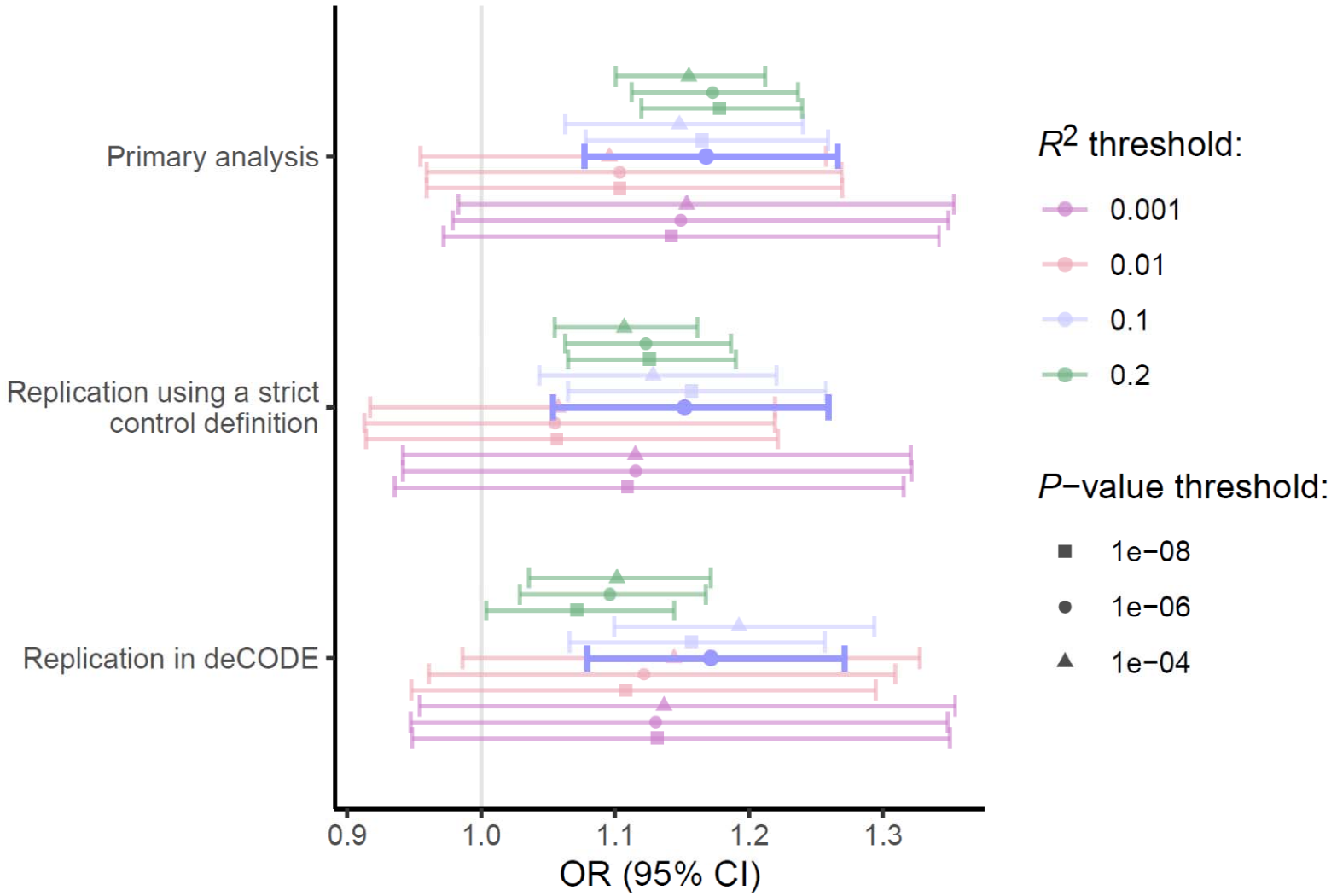


Genetic instruments were constructed for all complement- and coagulation-related proteins measured in up to 34,557 participants from the UK Biobank[9]. The *x*-axis highlights all 49 proteins, for which genetic instruments were constructed using genetic variants, also known as *cis*-protein quantitative trait loci (*cis*-pQTLs), that were strongly associated with circulating protein levels ($P<1\times10^{-6}$) and not in linkage disequilibrium with each other (linkage disequilibrium $R^2<0.1$). Data for long COVID were obtained from the Long COVID Host Genetics Initiative[26]. Associations of genetic instruments for circulating proteins with long COVID were calculated using the inverse-variance weighted Mendelian randomization (MR) method accounting for between-variant correlation. Columns with a negative value on the *y*-axis represent negative associations (i.e., associations for which the regression coefficient [β] was negative) whereas those with a positive value represent positive associations. The height of each bar represents the $\log_{10}$-transformed *P*-value of the corresponding association. Dark blue columns indicate associations reaching

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

Bonferroni significance ($P<1.0\times10^{-3}$); light red and blue columns indicate associations reaching nominal significance ($P<0.05$); grey columns indicate associations not reaching nominal significance ($P\geq0.05$).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.



612 **Fig. 3. Association of genetically predicted circulating factor XI levels with long COVID in sensitivity and replication analyses.**

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

Sensitivity analyses tested the association of genetically predicted factor XI levels with long COVID using the inverse-variance weighted method with genetic instruments constructed using a range of *P*-value (different shapes as indicated) and linkage disequilibrium R^2 (different colors as indicated) thresholds to construct genetic instruments. The *P*-value threshold determines the minimal association strength for genetic variants included in each genetic instrument, whereas the linkage disequilibrium R^2 thresholds determines the maximal allowed correlation between genetic variants in each genetic instrument. In primary analyses, associations were calculated using data from the UK Biobank (with proximity extension assay proteomics) and a broad control definition for long COVID. Replication analyses used an alternative, stricter, control definition[26] ("Replication using a strict control definition") or data from an external cohort with an alternative, aptamer-based, proteomic profiling method[10] ("Replication in deCODE"). Long COVID phenotype definitions are as reported by Lammi *et al.*[26] CI indicates confidence interval; OR, odds ratio.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.

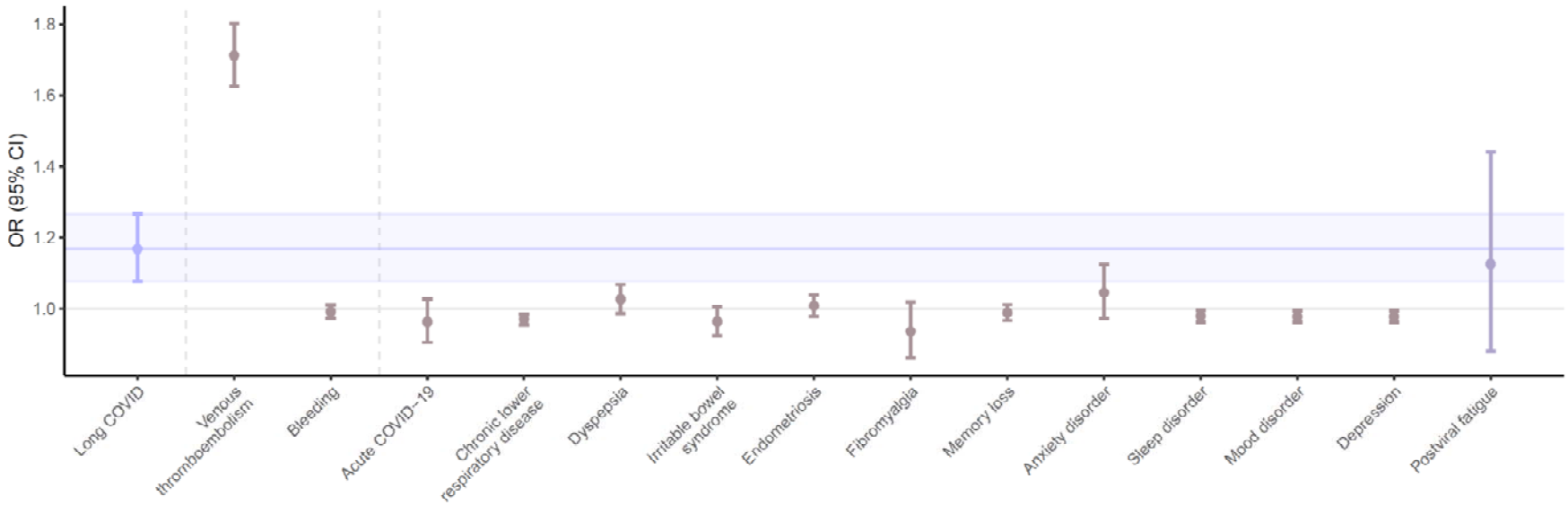
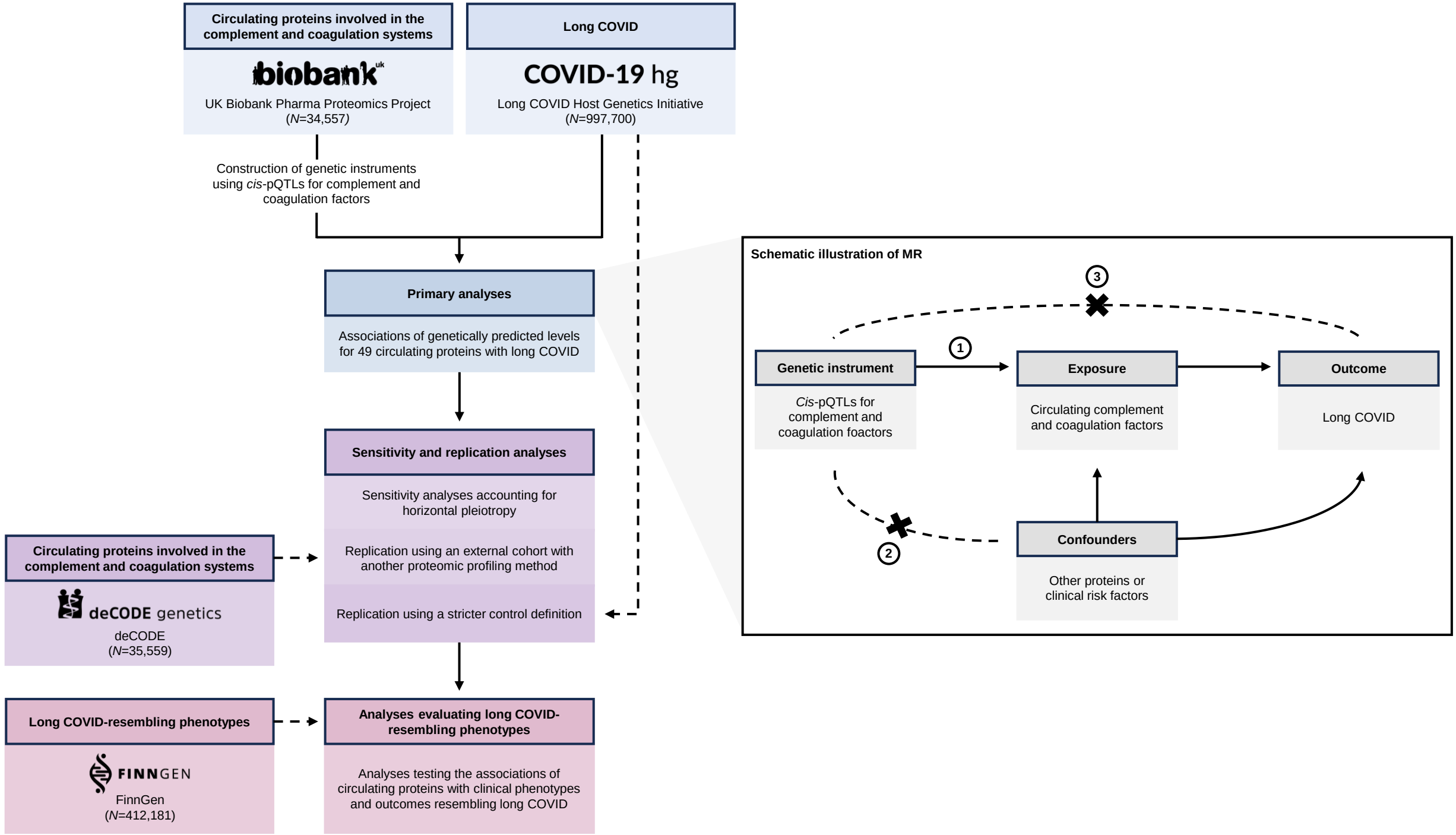
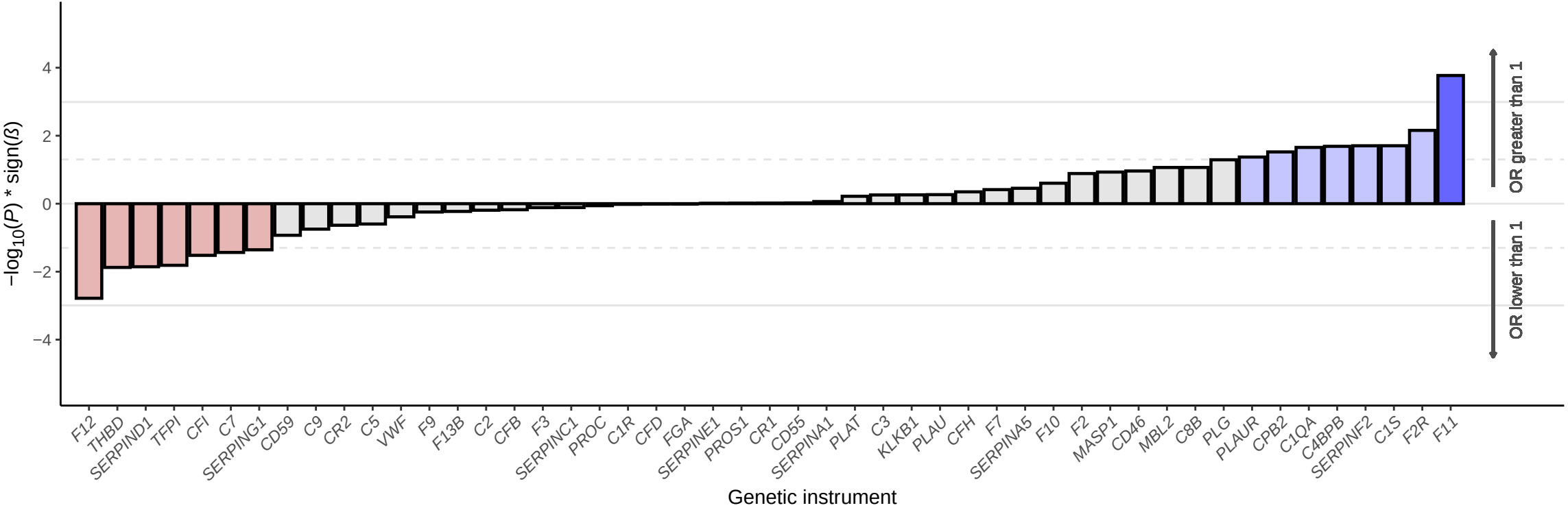
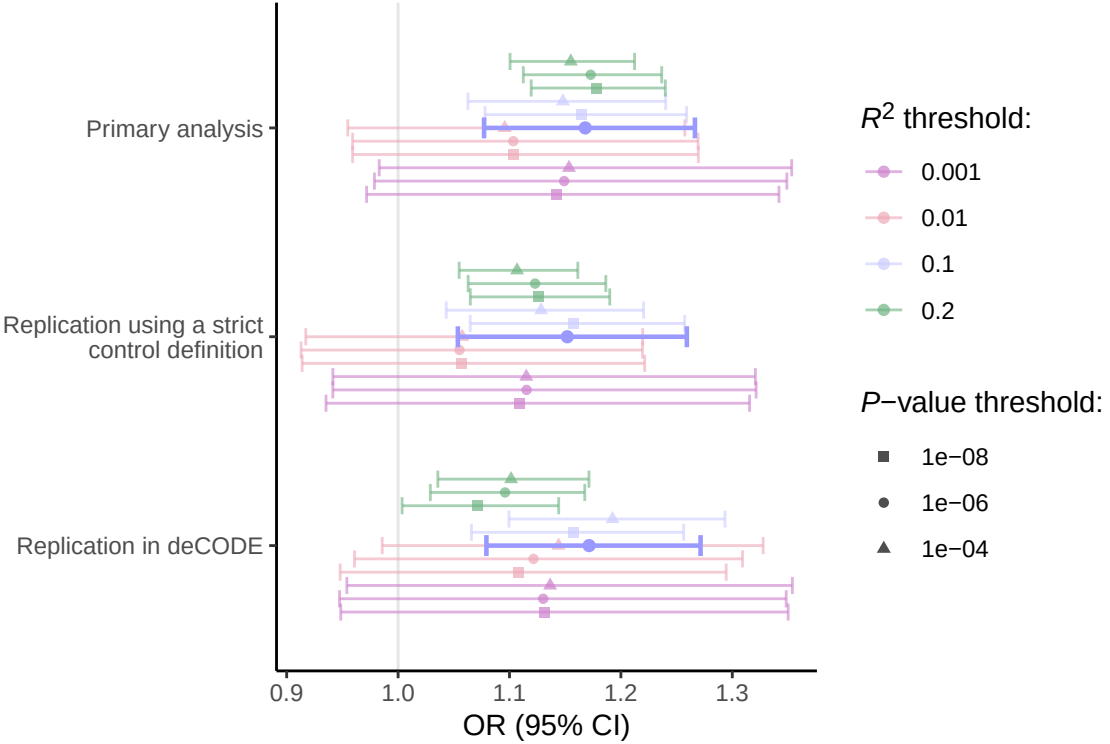


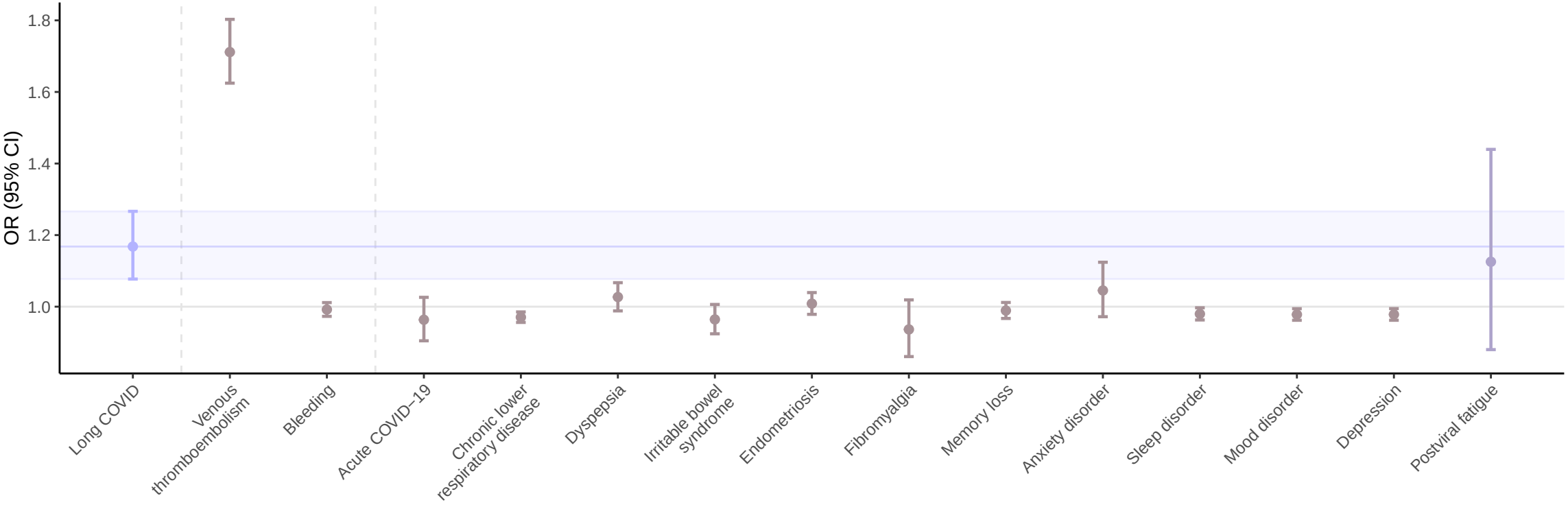
Fig. 4. Associations of genetically predicted circulating factor XI levels with long COVID and other phenotypes. Associations of genetically predicted circulating levels of factor XI (FXI) with long COVID were calculated using the inverse-variance weighted method accounting for between-variant correlation structure. The genetic instrument for FXI was constructed using proteogenomic data from the UK Biobank[9] with a *P*-value threshold of 1×10^{-6} and an R^2 threshold of 0.1. The association of genetically predicted circulating levels of FXI with long COVID was tested using genetic data from the Long COVID Host Genetics Initiative[26]. The associations of genetically predicted circulating levels of FXI with other phenotypes were tested using genetic data from FinnGen[54]. Higher genetically predicted circulating levels of factor XI were significantly associated with a higher risk of long COVID (depicted in blue; the blue band corresponds to the 95% confidence interval of this association). Higher genetically predicted factor XI levels were associated with a higher risk of venous thromboembolism, but not with bleeding (depicted in grey between the dotted lines). Higher genetically predicted factor XI levels were not associated with a higher risk of acute COVID-19 or conditions with clinical manifestations that overlap with long COVID (yet represent separate clinical entities). CI indicates confidence interval; OR, odds ratio.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.17.24307553; this version posted May 18, 2024. 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.


Circulating proteins involved in the complement and coagulation systems
biobank uk
UK Biobank Pharma Proteomics Project
(N=34,557)
Long COVID
COVID-19 hg
Long COVID Host Genetics Initiative
(N=997,700)
Construction of genetic instruments using cis-pQTLs for complement and coagulation factors
Primary analyses
Associations of genetically predicted levels for 49 circulating proteins with long COVID
Sensitivity and replication analyses
Sensitivity analyses accounting for horizontal pleiotropy
Replication using an external cohort with another proteomic profiling method
Replication using a stricter control definition
Circulating proteins involved in the complement and coagulation systems
deCODE genetics
deCODE
(N=35,559)
Long COVID-resembling phenotypes
FINNGEN
FinnGen
(N=412,181)
Analyses evaluating long COVID-resembling phenotypes
Analyses testing the associations of circulating proteins with clinical phenotypes and outcomes resembling long COVID
Schematic illustration of MR
3
1
Genetic instrument
Cis-pQTLs for complement and coagulation foactors
Exposure
Circulating complement and coagulation factors
Outcome
Long COVID
2
Confounders
Other proteins or clinical risk factors


$-\log_{10}(P)$ * sign(β)
4
2
0
−2
−4
OR greater than 1
OR lower than 1
F12
THBD
SERPIND1
TFPI
CFI
C7
SERPING1
CD59
C9
CR2
C5
VWF
F9
F13B
C2
CFB
F3
SERPINC1
PROC
C1R
CFD
FGA
SERPINE1
PROS1
CR1
CD55
SERPINA1
PLAT
C3
KLKB1
PLAU
CFH
F7
SERPINA5
F10
F2
MASP1
CD46
MBL2
C8B
PLG
PLAUR
CPB2
C1QA
C4BPB
SERPINF2
C1S
F2R
F11
Genetic instrument


Primary analysis
Replication using a strict control definition
Replication in deCODE
0.9
1.0
1.1
1.2
1.3
OR (95% CI)
R^2 threshold:
0.001
0.01
0.1
0.2
P−value threshold:
1e−08
1e−06
1e−04


OR (95% CI)
1.8
1.6
1.4
1.2
1.0
Long COVID
Venous thromboembolism
Bleeding
Acute COVID−19
Chronic lower respiratory disease
Dyspepsia
Irritable bowel syndrome
Endometriosis
Fibromyalgia
Memory loss
Anxiety disorder
Sleep disorder
Mood disorder
Depression
Postviral fatigue