Received: 16 November 2022 | Accepted: 18 January 2023

DOI: 10.1111/bjh.18679

BJHaem
BRITISH JOURNAL OF HAEMATOLOGY

ORIGINAL PAPER

Plasma protein and venous thromboembolism: prospective cohort and mendelian randomisation analyses

Shuai Yuan[1] | **Olga E. Titova**[2] | **Ke Zhang**[3,4] | **Wanglong Gou**[3,4] | **Tessa Schillemans**[1] | **Pradeep Natarajan**[5,6,7] | **Jie Chen**[8] | **Xue Li**[8] | **Agneta Åkesson**[1] | **Maria Bruzelius**[9,10] | **Derek Klarin**[11,12] | **Scott M. Damrauer**[13,14] | **Susanna C. Larsson**[1,2]

[1]Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden

[2]Unit of Medical Epidemiology, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

[3]Key Laboratory of Growth Regulation and Translational Research of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, China

[4]Westlake Intelligent Biomarker Discovery Lab, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China

[5]Program in Medical and Population Genetics and the Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[6]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

[7]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA

[8]Department of Big Data in Health Science School of Public Health, Center of Clinical Big Data and Analytics of The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[9]Department of Medicine, Solna, Karolinska Institute, Stockholm, Sweden

[10]Coagulation Unit, Department of Hematology, Karolinska University Hospital, Stockholm, Sweden

[11]VA Palo Alto Healthcare System, Palo Alto, California, USA

[12]Department of Surgery, Stanford University School of Medicine, Palo Alto, California, USA

[13]Corporal Michael Crescenz Veterans Affairs Medical Center, Philadelphia, Pennsylvania, USA

[14]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence
Shuai Yuan and Susanna C. Larsson, Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden.
Email: shuai.yuan@ki.se; susanna.larsson@ki.se

Funding information
Forskningsrådet om Hälsa, Arbetsliv och Välfärd, Grant/Award Number: 2018-00123; Hjärt-Lungfonden, Grant/Award Number: 20190247; Karolinska Institutet, Grant/Award Number: 2020-01842; Vetenskapsrådet, Grant/Award Number: 2019-00977; Swedish Research Council; Department of Veterans Affairs

Summary

We conducted cohort and Mendelian randomisation (MR) analyses to examine the associations of circulating proteins with risk of venous thromboembolism (VTE) to provide evidence basis for disease prevention and drug development. Cohort analysis was performed in 11 803 participants without baseline VTE. Cox regression was used to estimate the associations between 257 proteins and VTE risk. A machine-learning model was constructed to compare the importance of identified proteins and traditional risk factors. Genetic association data on VTE were obtained from a genome-wide meta-analysis (26 066 cases and 624 053 controls) and FinnGen (14 454 cases and 294 700 controls). The cohort analysis, including 353 incident VTE cases diagnosed during a 6.6-year follow-up, identified 21 proteins associated with VTE risk after false discovery rate correction. The machine-learning model indicated that body mass index and von Willebrand factor (vWF) made the same as well as most of the contributions to the overall model prediction. MR analysis found that genetically predicted levels of vWF, SERPINE1 (plasminogen activator inhibitor 1, known as PAI-1), EPHB4 (ephrin type-B receptor 4), TYRO3 (tyrosine-protein kinase receptor TYRO3), TNFRSF11A (tumour necrosis factor receptor superfamily member 11A), and BOC (brother of CDO) were causally associated with VTE risk.

KEYWORDS
cohort, mendelian randomisation, protein, venous thromboembolism

INTRODUCTION

Venous thromboembolism (VTE) is a common thrombotic vascular disease that afflicts 1–2 per 1000 people each year and is responsible for high healthcare costs and increased disability-adjusted life-years globally.[1,2] Compared to atherosclerotic diseases, VTE appeared to be less preventable from the perspective of modifiable risk factors.[3,4] Except for adiposity[5] and cigarette-smoking,[6,7] traditional cardiovascular risk factors, such as high glycaemic status,[8] blood pressure[4] and blood lipid levels,[9] have not been consistently associated with the risk of developing VTE. Thus, trying to identify additional factors closely associated with the onset of VTE——especially biomarkers——may be of importance in facilitating the disease prophylaxis, optimising the clinical management, as well as benefiting the development of drugs treating VTE.[10]

Proteins acting as the principal regulators of molecular pathways have been targeted for the development of the vast majority of drugs.[11] The circulating proteome measured by varying assays includes cell- and tissue-generated circulating proteins that are secreted into circulation or released during cell damage or turnover.[12] Several studies explored the associations of circulating proteins with VTE,[13–16] but most of these analyses were based on the case–control design with a sample size where residual confounding, reverse causality and misclassification can hinder causal inference.

Mendelian randomisation is an epidemiological approach that can strengthen causal inference by utilising genetic variants as instrumental variables for the exposure (e.g., circulating proteins).[17] For MR analysis on proteins, the *cis*-variant in the protein-encoding gene (known as the protein quantitative trait loci, pQTL) has been widely used as a genetic instrument to estimate the causal effects of the circulating protein on health outcomes,[18] which desirably satisfies three key assumptions of MR (i.e., relevance, independence and exclusion assumptions[19]). Here, we conducted cohort and MR analyses to explore the association of blood proteins with the risk of developing VTE by using protein and genetic data from Swedish, U.S., U.K. and Finnish populations.

METHODS

Study design

Figure 1 shows study design. We firstly explored the association of 257 blood proteins with incident VTE risk in 11 803 Swedish adults. We then used machine-learning to compare the importance of identified proteins with known risk factors for VTE. MR analysis was used to strengthen causal inference for the association between identified proteins and VTE risk by using data from large-scale genetic studies.

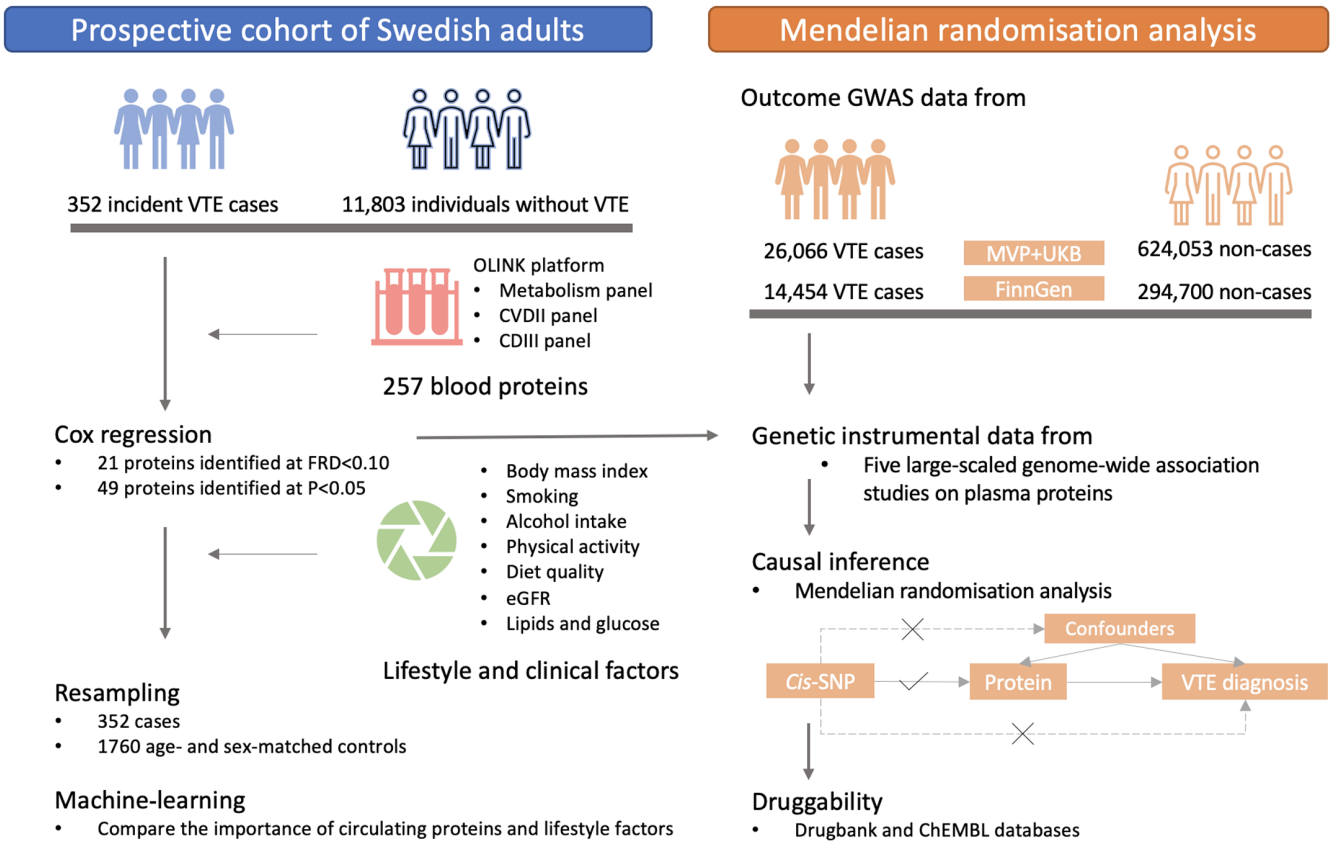


FIGURE 1 Study design. FDR, false discovery rate; GWAS, genome-wide association study; MVP, the Million Veteran Program; SNP, single-nucleotide polymorphism; UKB, UK Biobank; VTE, venous thromboembolism.

Finally, two databases were searched to evaluate the druggability of the identified proteins.

Cohort analysis

Study population

This study was based on the clinical sub-cohorts of the Swedish Infrastructure for Medical Population-based Life-course and Environmental Research (SIMPLER) that includes the Swedish Mammography Cohort (SMC) and Cohort of Swedish Men (COSM). The clinical examinations in sub-cohorts of SMC and COSM were conducted in 2003–2009 and 2010–2019, respectively. Samples of fasting blood were collected at health examinations and participants were asked to fill in questionnaires on diet, health and lifestyle. After removing 511 individuals with baseline VTE from 12 314 participants with available protein data, we included 11 803 participants from the two sub-cohorts in the analysis. Detailed information on two sub-cohorts and questionnaires can be found on the SIMPLER website (https://www.simpler4health.se/).

Proteomic profiling

Venous blood samples were collected after a 12-h overnight fast and immediately centrifuged and stored at −80°C until analysis. In total, 276 protein biomarkers were analysed using three high-throughput multiplex immunoassays: the Olink Proseek Multiplex CVD II, CVD III and Metabolism (Olink Bioscience). Each assay measured 92 selected cardiovascular disease- or metabolism-related proteins. The platform provides normalised protein expression values on a log2 scale standardised per analysis plate. Values below the limit of detection (LOD) were provided by the manufacturer and used as a protein selection criterium. The analyses were performed at SciLifeLab, Uppsala University, Sweden.[20] In this analysis, we excluded 19 proteins with more than 50% samples below the LOD (Table S1). A small portion of specimens was set to missing given the analysed sample did not pass the manufacturer's quality control (3.6%, 0.8% and 0.7% for CVD II, CVD III and metabolism panels, respectively). The 257 proteins included can be found in Table S2.

Case ascertainment and follow-up

VTE cases were ascertained by a medical diagnosis of VTE and its subtypes, either as the primary or contributing causes with diagnostic information (codes of International Classification of Diseases-9 and -10) from the Swedish National Patient Register, which covers nearly all hospital-based inpatient and outpatient care.[21] Dates of death were obtained from the Swedish Death Registry. Individuals were followed up from the baseline until the date of diagnosis of VTE, date of death, or end of follow-up (i.e., 31 December 2019), whichever came first.

Demographical, lifestyle and clinical factors

Information on age, sex, education attainment, smoking, alcohol consumption, physical activity, and diet quality was obtained from self-administrated questionnaires. Diet quality was assessed by a modified Dietary Approaches to Stop Hypertension (mDASH) score.[22] Body mass index (BMI, weight/height squared), estimated glomerular filtration rate (eGFR), levels of blood lipids and glucose, and blood pressure were measured in the health exam. We obtained data on baseline diagnosis of cardiovascular disease including coronary artery disease, heart failure, stroke, and atrial fibrillation from Swedish National Patient Register. Detailed information on above factors is shown in Table S3.

Cox regression analysis

Missing protein and demographic values, lifestyle, and clinical factors were dealt with by using multiple imputation by chained equations (20 imputation cycles). Cox proportional hazards regression model with age as the underlying time scale was used to estimate the hazard ratios (HRs) and corresponding 95% confidence intervals (CIs) of the association between circulating proteins and incident VTE risk. The assumption of proportionality was met as indicated by Schoenfeld residuals. Two models were used: (a) Model 1 adjusted for sex and plate, and (b) Model 2 adjusted for sex, plate, BMI,[5] educational attainment,[23] baseline cardiovascular disease,[24] smoking status,[22] alcohol consumption,[25] physical activity,[22] mDASH score,[22] eGFR,[26] low- and high-density lipoprotein cholesterol,[27] triglycerides,[27] blood pressure,[28] and blood glucose levels.[29] We performed a sensitivity analysis excluding individuals with baseline cardiovascular disease (n = 2548) to test the robustness of the associations. The statistical tests were two-sided, and the analyses were performed in Stata/SE (version 15.0; StataCorp). The false discovery rate (FDR) was used to account for multiple testing. FDR < 0.1 was considered statistically significant.

Machine-learning

The aim of machine learning is to compare the importance between identified protein biomarkers and traditional lifestyle factors using a model with relaxed assumptions. To keep a balance between the numbers of cases and controls in a machine-learning model, we resampled each incident VTE case with five age- and sex-matched controls using the MatchIt package.[30] The comparative importance of identified proteins and the traditional risk factor for VTE was examined by using a gradient-boosting framework, the Light Gradient Boosting Machine (LightGBM).[31] The factors'

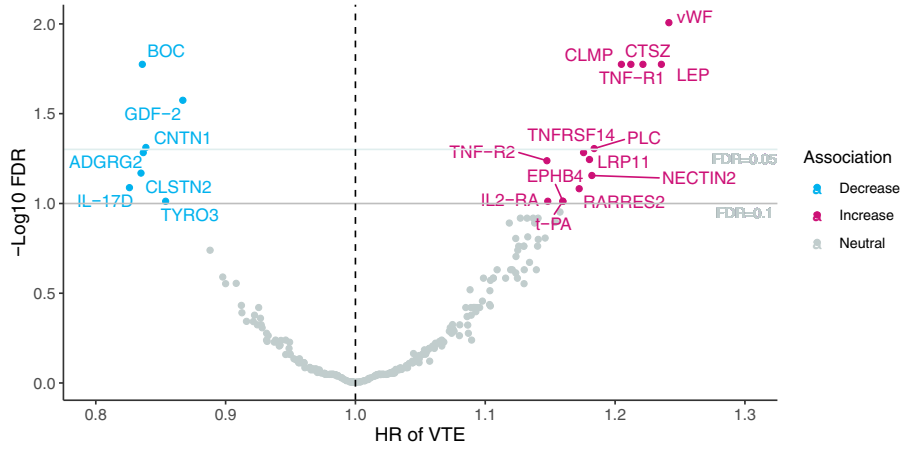


FIGURE 2 Volcano plot showing results from protein-wide Cox regression on the association between 257 circulating proteins and risk of incident VTE in 11 803 Swedish adults. FDR, false discovery rate; HR, hazard ratio; VTE, venous thromboembolism.

predictive performance was evaluated by using tenfold cross validation. We used SHapley Additive exPlanations (SHAP) to estimate the average contribution of each predictive factor to the overall model predictions.[32] Traditional risk factor for VTE included BMI,[5] smoking,[22] alcohol consumption,[25] physical activity,[22] mDASH diet,[22] eGFR,[26] and blood lipids,[27] blood pressure[28] and glucose levels.[29] The analysis was conducted using R software 4.1.1.

MR analysis

To examine the causality of the association between proteins and VTE in a comprehensive way and reduce the possibility of a Type 2 error caused by a heavy multiple-testing burden in cohort analysis, we included proteins associated with VTE at the nominal significance level in the cohort analysis to MR analysis.

VTE outcome data sources

Summary-level data on the associations of DNA sequence variants with the VTE risk were obtained from a genome-wide meta-analysis study of the Million Veteran Program and UK Biobank, which include 26 066 cases and 624 053 controls of European ancestry.[27] Genetic data on VTE from the FinnGen R7 study data release that comprised 14 454 cases and 294 700 controls, was used for independent replication.[33] Detailed information, such as case definition and covariates, is presented in Table S4.

Genetic instrument selection

Cis-SNPs (single nucleotide polymorphisms) associated with proteins at the genome-wide significance level ($p < 5 \times 10^{-8}$) were employed as instrumental variables with summary-level statistic data (β coefficients and standard errors) obtained from six genome-wide association studies (GWASs).[34–39] Protein data were analysed by performing a SomaLogic assay in four studies[34,36–38] and Olink in two studies.[35,39] Missing SNP was replaced by a proxy SNP with strong linkage disequilibrium ($r^2 \geq 0.8$). Proteins without genetic instruments were removed from the analysis. Detailed information on used GWASs is presented in Table S5. Instrument variables are listed in Table S6.

MR statistical analysis

The F statistic was used to measure the strength of the genetic instrument. The Wald ratio method was used to calculate causal estimates of the associations between proteins and VTE risk (i.e., the β coefficient for the effect of the SNP on VTE divided by the β coefficient for the effect of the SNP on the protein).[40] The standard error of the ratio estimate was estimated using the delta method.[40] Estimates for each protein from the GWAS meta-analysis and FinnGen were combined using the fixed-effect meta-analysis method. The analysis was conducted using R software 4.1.1.

Evaluation of druggability

Drugbank (https://go.drugbank.com/) and ChEMBL (https://www.ebi.ac.uk/chembl/) databases were searched to evaluate the druggability of identified proteins in MR analysis. We documented the drug name and the process of drug development for identified proteins.

RESULTS

Cohort analysis

During a median follow-up of 6.6-years, 352 incident VTE cases were diagnosed. The baseline characteristics of participants by incident VTE are shown in Table S7.

After multiple testing (FDR < 0.1, Model 1), increased levels of 14 proteins were associated with increased VTE risk and increased levels of seven proteins were associated with decreased VTE risk (Figure 2 and Table 1). The associations remained in the multivariable adjusted analysis (Model 2, Table 1). The associations were also stable in a sensitivity analysis excluding participants with baseline cardiovascular disease, albeit with larger CIs (Table S8). In detail, higher levels of vWF (von Willebrand factor), LEP (leptin), CTSZ (cathepsin Z), TNF-R1 (tumour necrosis factor receptor 1), CLMP (CXADR-like membrane protein), PLC (perlecan), TNFRSF14 (tumour necrosis factor receptor superfamily member 14), LRP11 (low-density lipoprotein receptor-related protein 11), TNF-R2 (tumour necrosis factor receptor 2), NECTIN2 (nectin-2), RARRES2 (retinoic acid receptor responder protein 2), EPHB4 (ephrin type-B receptor 4), IL2-RA (interleukin-2 receptor subunit alpha), and t-PA (tissue-type plasminogen activator) were associated with an increased risk of VTE (Figure 2 and Table 1). Higher levels of BOC (brother of CDO), GDF-2 (growth-differentiation factor 2), CNTN1 (contactin-1), ADGRG2 (adhesion G-protein coupled receptor G2), CLSTN2 (calsyntenin-2), IL-17D (interleukin-17D), and TYRO3 (tyrosine-protein kinase receptor TYRO3) were associated with a decreased risk of VTE (Figure 2 and Table 1). The associations between the other 239 studied proteins and VTE risk are displayed in Table S9.

Comparative importance of proteins and traditional factors

Among 21 proteins and 10 traditional risk factors, body mass index ranked as the greatest risk with the highest SHAP value, followed by vWF (Figure 3). BOC, CLSTN2, IL2-RA, GDF-2, and CNTN1 showed greater importance than eGFR (Figure 3).

MR analysis

Among the 49 identified proteins associated with VTE at a nominal significance level, five proteins (ADGRG2, APLP1, CD93, GDF-2 and IL-6) were removed from MR analysis due to the lack of genetic instruments. The F statistic for all SNPs was over 10, indicating a good strength of genetic instruments used (Table S6). In the analysis of VTE data from the

TABLE 1 Hazard ratios and 95% confidence intervals of incident VTE for 21 identified associations with proteins at FRD < 0.1.

	Model 1				Model 2		
Protein	**HR**	**95% CI**	***p***	**FDR**	**HR**	**95% CI**	***p***
vWF	1.24	1.12–1.38	3.78 E-05	0.010	1.23	1.11–1.37	9.39 E-05
BOC	0.84	0.76–0.92	1.30 E-04	0.017	0.84	0.76–0.92	2.41 E-04
LEP	1.24	1.1–1.39	3.88 E-04	0.017	1.22	1.08–1.39	0.002
CTSZ	1.22	1.09–1.36	3.69 E-04	0.017	1.22	1.09–1.38	0.001
TNF-R1	1.21	1.09–1.35	3.67 E-04	0.017	1.23	1.09–1.39	0.001
CLMP	1.21	1.09–1.33	2.25 E-04	0.017	1.20	1.07–1.33	0.001
GDF-2	0.87	0.8–0.94	0.001	0.027	0.86	0.79–0.93	0.000
CNTN1	0.84	0.75–0.93	0.002	0.049	0.84	0.75–0.94	0.002
PLC	1.18	1.07–1.32	0.002	0.049	1.17	1.04–1.32	0.010
TNFRSF14	1.18	1.06–1.3	0.002	0.052	1.19	1.05–1.33	0.004
ADGRG2	0.84	0.75–0.94	0.002	0.052	0.83	0.74–0.94	0.003
LRP11	1.18	1.06–1.31	0.003	0.057	1.17	1.04–1.31	0.008
TNF-R2	1.15	1.05–1.26	0.003	0.058	1.15	1.04–1.26	0.006
CLSTN2	0.84	0.74–0.94	0.004	0.068	0.83	0.73–0.93	0.002
NECTIN2	1.18	1.05–1.32	0.004	0.070	1.18	1.04–1.34	0.009
IL-17D	0.83	0.72–0.94	0.005	0.082	0.82	0.72–0.94	0.005
RARRES2	1.17	1.05–1.31	0.005	0.083	1.17	1.03–1.32	0.014
EPHB4	1.16	1.04–1.29	0.007	0.097	1.16	1.03–1.3	0.011
IL2-RA	1.15	1.04–1.27	0.007	0.097	1.15	1.03–1.28	0.010
t-PA	1.16	1.04–1.29	0.008	0.097	1.15	1.02–1.29	0.027
TYRO3	0.85	0.76–0.96	0.008	0.097	0.85	0.76–0.96	0.007

Note: CI, confidence interval; FDR, false discovery rate; HR, hazard ratio. Model 1 adjusted for age, sex, and plate; Model 2 adjusted for age, sex, body mass index, plate, education attainment, baseline cardiovascular disease, smoking status, alcohol consumption, physical activity, mDASH diet score, eGFR, low- and high-density lipoprotein cholesterol, triglycerides, blood pressure, and blood glucose levels. Full name of proteins can be found in Table S2.

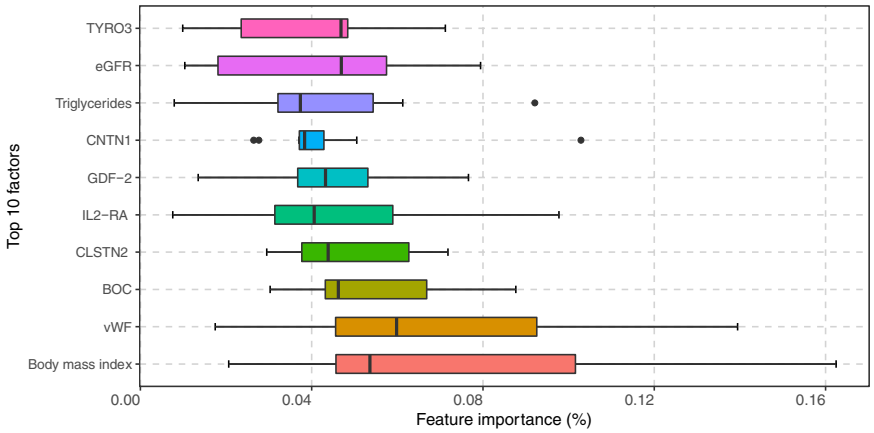
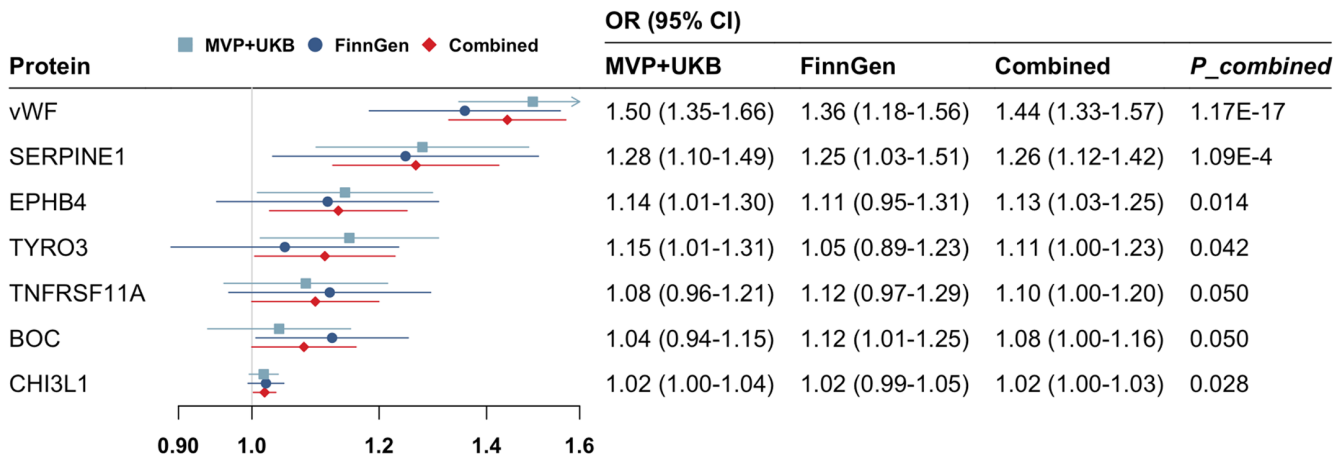


FIGURE 3 Top 10 VTE predictors among 21 identified proteins, five lifestyle factors, and five clinical features based on Light Gradient Boosting Machine analysis.



Protein	OR (95% CI) MVP+UKB	FinnGen	Combined	*P_combined*
vWF	1.50 (1.35-1.66)	1.36 (1.18-1.56)	1.44 (1.33-1.57)	1.17E-17
SERPINE1	1.28 (1.10-1.49)	1.25 (1.03-1.51)	1.26 (1.12-1.42)	1.09E-4
EPHB4	1.14 (1.01-1.30)	1.11 (0.95-1.31)	1.13 (1.03-1.25)	0.014
TYRO3	1.15 (1.01-1.31)	1.05 (0.89-1.23)	1.11 (1.00-1.23)	0.042
TNFRSF11A	1.08 (0.96-1.21)	1.12 (0.97-1.29)	1.10 (1.00-1.20)	0.050
BOC	1.04 (0.94-1.15)	1.12 (1.01-1.25)	1.08 (1.00-1.16)	0.050
CHI3L1	1.02 (1.00-1.04)	1.02 (0.99-1.05)	1.02 (1.00-1.03)	0.028

FIGURE 4 Potential causal associations of circulating proteins with risk of venous thromboembolism (VTE) in Mendelian randomisation analysis. CI, confidence interval; MVP, the Million Veteran Program; OR, odds ratio; UKB, UK Biobank.

meta-analysis of MVP and UKB, higher genetically predicted levels of vWF, SERPINE1 (plasminogen activator inhibitor 1, named as PAI in cohort analysis), TYRO3, and EPHB4 were associated with an increased risk of VTE (Figure 4 and Table S10). The association for vWF and SERPINE1 were replicated in the analysis of FinnGen (Figure 4 and Table S10). Genetically predicted low levels of PILRB (paired immunoglobulin-like Type 2 receptor beta) and higher levels of BOC were associated with an increased risk of VTE in FinnGen (Figure 4 and Table S10). In the meta-analysis using both MVP-UKB and FinnGen, the odds ratio of VTE per standard deviation increase in genetically predicted levels of protein was 1.44 (95% CI 1.33–1.57) for vWF, 1.26 (95% 1.12–1.42) for SERPINE1, 1.13 (95% 1.03–1.25) for EPHB4, 1.11 (95% CI 1.00–1.23) for TYRO3, 1.10 (95% CI 1.00–1.20) for TNFRSF11A (tumour necrosis factor receptor superfamily member 11A), 1.08 (95% 1.00–1.16) for BOC, and 1.02 (95% CI 1.00–1.03) for CHI3L1 (chitinase-3-like protein 1) (Figure 4). However, the association for CHI3L1 did not persist using another genetic instrument (Table S10). The MR associations for other studied proteins can be found in Table S10.

Assessment of druggability

Among the seven proteins potentially causally associated with VTE in the meta-analysis, five proteins have been targeted for drug development (Table S11). Several drugs targeted at vWF have been approved to treat bleeding and thrombosis-related disorders. Some thrombolytic drugs targeting SERPINE1 had been approved for use but were withdrawn due to lack of efficacy or side-effects, like Drotrecogin alfa and Urokinase. Fostamatinib targeting TYRO3 was designed to treat thrombocytopenia and was approved in April 2018 by the US Food and Drug Administration. Drugs targeting EPHB4 and TNFRSF11A aim to treat specific cancers. No drug information was found for BOC and CHI3L1.

DISCUSSION

Using cohort, machine-learning and MR designs, the study examined the association of 257 circulating proteins with VTE risk, using data from several European populations. In

the cohort analysis, we identified 21 proteins associated with incident VTE. Machine-learning analysis found that BMI and vWF shared an identical highest ranking concerning the contribution to the prediction model. MR analysis indicated causal associations of vWF, SERPINE1, EPHB4, TYRO3, TNFRSF11A, BOC and CHI3L1 with VTE risk. However, for TYRO3 and BOC, the direction of the association with VTE differed between cohort and MR analyses. The association for CHI3L1 was unrobust using different genetic instruments. Several drugs targeting these causal proteins, except for BOC and CHI3L1, have been approved or are under investigation, although not always for thrombotic outcomes. This discovery work provided directions for future mechanistic and clinical studies to deepen the understanding of the molecular aetiology of VTE as well as to facilitate drug development for VTE.

vWF has been associated with VTE risk in many cohort studies and genetic analyses.[15,41–43] In line with these studies, our study confirmed the causal role of elevated vWF levels in the development of VTE, which serves as a proof of principle and positive control for the methodological approach. Additionally, the novel results of our machine-learning analysis convey equal importance of vWF levels and body mass index for VTE onset. High BMI has been shown to be a causal risk factor for VTE in our previous study.[5] Thus, this finding may imply that monitoring vWF levels is as important as maintaining a healthy BMI for lowering VTE risk. This finding may help develop clinical treatment for and management of VTE. Furthermore, vWF has also been proposed as an appealing therapeutic target for VTE treatment. There are drugs of nonspecific (heparin and aurintricarboxylic acid) and specific (caplacizumab) vWF antagonists that have been approved to treat thrombotic disorders.[44] In addition, a new drug targeting vWF (Egaptivon pegol) is currently under investigation.

vWF and factor VIII are strongly correlated from biological and genetic perspectives.[42,45] An animal study in vivo found that vWF and factor VIII were independently required for thrombosis formation in injured veins.[46] However, whether vWF is causally associated with the risk of VTE independently of factor VIII in humans remained unclear. A previous MR study based on data from a large consortium failed to demonstrate a causal association of VWF levels with VTE that was independent of factor VIII levels because no genetic loci were independently associated with VWF levels, unlike with factor VIII levels.[42] Given the lack of data on factor VIII, our cohort study could not explore the independent associations of vWF and factor VIII with the risk of VTE. Thus, whether vWF is causally associated with VTE independent of factor VIII needs to be clarified in future studies.

Increased levels of SERPINE1, known as PAI-1, explained the positive association between elevated plasma clot lysis time (hypofibrinolysis) and VTE risk in a case–control study which included 770 patients and 743 control.[47] This association was also observed in a recent case–control study of 383 VTE cases and 782 age- and sex-matched controls in Norway.[48] Even though the genetic association of *PAI-1* gene with VTE risk was inconsistent,[49,50] our prospective cohort and MR analyses indicated a potential causal adverse effect of high levels of PAI-1 on the development of VTE. Notably, even though some drugs which influence PAI-1 levels have been approved to treat thrombotic disorders, several drugs targeting PAI-1 have been withdrawn owing to limited treatment efficacy.[51]

Even though we identified possible causal roles of EPHB4, TYRO3, TNFRSF11A, BOC, and CHI3L1 in VTE, these are novel findings that need confirmation. For EPHB4, the protein may regulate blood vascular morphogenesis and could thereby impact thrombosis development.[52] TYRO3 is a part of the Tyro3/Axl/Mer receptor and plays an important role in the growth arrest specific 6 (gas6) signalling pathway for stable platelet aggregation.[53] The TNFRSF11A-realted pathway influenced by wnt signalling has been associated with bone mineralisation and also linked to inflammatory disease,[54] which may indicate the role of tumour necrosis factor-related inflammation in formation of venous thrombosis.[55] CHI3L1, also known as YKL-40, was associated with VTE in cohort analysis but not in MR analysis in a previous study.[56] Our study found a positive association between CHI3L1 and VTE in both analyses with larger statistical power due to larger sample sizes. However, this association was inconsistent due to using different genetic instruments and thus requires further confirmation. Notably, there are discrepancies in the associations for TYRO3 and BOC between cohort and MR analyses, which might be caused by different proteomic profiling assays in two analyses according to a recent genome-wide association study where each of the eight protein targets measured by SomaLogic and Olink platforms shared the same genetic signal but with opposite effect directions for the protein target or its isoforms.[57] In addition, our druggability assessment identified that the primary outcome of EPHB4- and TNFRSF11A-related drugs were specific cancers, which might indicate that these cancer therapies may be vulnerable to thrombotic side-effects.

This study has several strengths, including the prospective cohort and MR analyses with large sample sizes, low missing rate of protein data, genetic instruments from recent genome-wide studies,[34–39] utility of several analytical approaches as triangulation, as well as a comprehensive research of causal proteins in drug databases. However, there are also several limitations that need to be considered when interpretating our results. First, our analyses were confined to European populations and whether our findings can be generalised to other populations needs to be investigated. Second, even though 257 cardiovascular and metabolic proteins were included in the analysis, the study by no means examined all circulating proteins in relation to VTE, which indicated that other important VTE-related proteins might be overlooked.[58] Third, compared to mass spectrometry methods, the Olink protein immunoassay was not able to profile different protein isoforms or posttranslational modifications. However, this method with moderate-to-high specificity and high throughput efficiently measured

the absolute quantifications of proteins in a large sample.[59] Fourth, we were not able to conduct an independent replication cohort analysis and although we used MR analysis to verify associations for identified proteins, not all proteins were measured in the same assays (Olink vs SomaLogic), This may have caused the directional discordance for certain associations between cohort and MR analyses.[57] Thus, future studies using a combination of different proteomic profiling platforms are needed to confirm the direction of these associations.[60]

In conclusion, several protein biomarkers were associated with the development of incident VTE in our cohort study and several of these (vWF, SERPINE1, EPHB4 and TNFRSF11A) were similarly associated in the MR analysis, which indicates potential causality. Our results also indicate that circulating vWF is as important as BMI for VTE prophylaxis. Trials on drugs targeting identified proteins may inform the efficacy and side-effects on thrombotic disorders. Notably, this study is exploratory, with findings that need to be verified in other studies.

AUTHOR CONTRIBUTIONS

Shuai Yuan and Susanna C. Larsson conceived and designed the study. Shuai Yuan, Ke Zhang, Olga E. Titova, Pradeep Natarajan, Derek Klarin, Scott M. Damrauer, and Susanna C. Larsson undertook the statistical analyses. Shuai Yuan wrote the first draft of the manuscript. All authors provided important comments to the manuscript and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We want to acknowledge the participants and investigators of SIMPLER for providing facilities and experimental support. SIMPLER receives funding through the Swedish Research Council under grant number 2017-00644. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project simpl2020002. We also want to acknowledge the participants and investigators of the VA Million Veteran Program, the UK Biobank study, and the FinnGen study. S.M.D. is supported by IK2-CX001780. This publication does not represent the views of the Department of Veterans Affairs or the United States Government. S.Y. and S.C.L. conceived and designed the study. S.Y., K.Z, O.E.T, P.N., D.K., S.M.D. and S.C.L undertook the statistical analyses. S.Y. wrote the first draft of the manuscript. S.Y., O.E.T, K.Z, W.G., T.S., P.N., J.C., X.L., A.Å., M.B., D.K., S.M.D. and S.C.L interpreted data, provided important comments to the manuscript, and approved the final version of the manuscript.

FUNDING INFORMATION

Funding for this study came from the Karolinska Institutet's Research Foundation Grants (Grant number 2020–01842), the Swedish Research Council (Vetenskapsrådet; Grant Number 2019–00977), the Swedish Research Council for Health, Working Life and Welfare (Forte; 2018–00123), and the Swedish Heart-Lung Foundation (Hjärt-Lungfonden; Grant number 20190247).

CONFLICT OF INTEREST

P.N. reports investigator-initiated grants from Amgen, Apple, AstraZeneca, Boston Scientific, and Novartis and personal fees from Apple, AstraZeneca, Blackstone Life Sciences, Foresite Labs, Novartis, and Roche / Genentech. P.N. is a co-founder of TenSixteen Bio and a scientific advisory board member of Esperion Therapeutics, geneX-well, and TenSixteen Bio. P.N. reports spousal employment at Vertex. All these conflicts are unrelated to the present work.

DATA AVAILABILITY STATEMENT

De-identified SIMPLER data are available for researchers upon application (http://www.simpler4health.se/). Summary data from the MVP + UKB GWAS of VTE can be obtained via dbGAP, accession code no. phs001672.v2.p1. Summary-level data from the FinnGen study of VTE can be obtained via https://finngen.gitbook.io/documentation/.

ETHICS APPROVAL STATEMENT

The Swedish Ethical Review Authority has approved the studies based on data from the SIMPLER cohorts (Dnr 2019–03986). The cited GWAS studies and the FinnGen study had been approved by corresponding ethical permit committees. All participants had signed the consent forms.

PATIENT CONSENT STATEMENT

n/a

PERMISSION TO REPRODUCE MATERIAL FROM OTHER SOURCES

n/a

ORCID

Shuai Yuan https://orcid.org/0000-0001-5055-5627

REFERENCES

1. Ortel TL, Neumann I, Ageno W, Beyth R, Clark NP, Cuker A, et al. American Society of Hematology 2020 guidelines for management of venous thromboembolism: treatment of deep vein thrombosis and pulmonary embolism. Blood Adv. 2020;4(19):4693–738.
2. Heit JA. Epidemiology of venous thromboembolism. Nat Rev Cardiol. 2015;12(8):464–74.
3. Crous-Bou M, Harrington LB, Kabrhel C. Environmental and genetic risk factors associated with venous thromboembolism. Semin Thromb Hemost. 2016;42(8):808–20.
4. Gregson J, Kaptoge S, Bolton T, Pennells L, Willeit P, Burgess S, et al. Cardiovascular risk factors associated with venous thromboembolism. JAMA Cardiol. 2019;4(2):163–73.
5. Yuan S, Bruzelius M, Xiong Y, Håkansson N, Åkesson A, Larsson SC. Overall and abdominal obesity in relation to venous thromboembolism. J Thromb Haemost. 2021;19(2):460–9.
6. Cheng YJ, Liu ZH, Yao FJ, Zeng WT, Zheng DD, Dong YG, et al. Current and former smoking and risk for venous thromboembolism: a systematic review and meta-analysis. PLoS Med. 2013;10(9):e1001515.

7. Larsson SC, Mason AM, Bäck M, Klarin D, Damrauer SM, Million Veteran Program, et al. Genetic predisposition to smoking in relation to 14 cardiovascular diseases. Eur Heart J. 2020;41:3304–10.
8. Yuan S, Mason AM, Burgess S, Larsson SC. Differentiating associations of glycemic traits with atherosclerotic and thrombotic outcomes: mendelian randomization investigation. Diabetes. 2022;71:2222–32.
9. Allara E, Morani G, Carter P, et al. Genetic determinants of lipids and cardiovascular disease outcomes: a wide-angled mendelian randomization investigation. 2019.
10. Letunica N, Van Den Helm S, McCafferty C, et al. Proteomics in thrombosis and hemostasis. Thromb Haemost. 2022;122(7):1076–84.
11. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med. 2017;9(383):eaag1166.
12. Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a multiomics approach. Nat Rev Cardiol. 2021;18(5):313–30.
13. Stachowicz A, Siudut J, Suski M, Olszanecki R, Korbut R, Undas A, et al. Optimization of quantitative proteomic analysis of clots generated from plasma of patients with venous thromboembolism. Clin Proteomics. 2017;14:38.
14. Jensen SB, Hindberg K, Solomon T, Smith EN, Lapek JD Jr, Gonzalez DJ, et al. Discovery of novel plasma biomarkers for future incident venous thromboembolism by untargeted synchronous precursor selection mass spectrometry proteomics. J Thromb Haemost. 2018;16(9):1763–74.
15. Bruzelius M, Iglesias MJ, Hong MG, Sanchez-Rivera L, Gyorgy B, Souto JC, et al. PDGFB, a new candidate plasma biomarker for venous thromboembolism: results from the VEREMA affinity proteomics study. Blood. 2016;128(23):e59–66.
16. Jensen SB, Latysheva N, Hindberg K, Ueland T. Plasma lipopolysaccharide-binding protein is a biomarker for future venous thromboembolism: results from discovery and validation studies. J Intern Med. 2022;292(3):523–35.
17. Burgess S, Thompson SG. Mendelian randomization: methods for using genetic variants in causal estimation. London, UK: Chapman and Hall/CRC; 2015.
18. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet. 2020;52(10):1122–31.
19. Davies NM, Holmes MV, Davey SG. Reading mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
20. Warensjö Lemming E, Byberg L, Stattin K, Ahmad S, Lind L, Elmståhl S, et al. Dietary pattern specific protein biomarkers for cardiovascular disease: a cross-sectional study in 2 independent cohorts. J Am Heart Assoc. 2019;8(11):e011860.
21. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011;11:450.
22. Yuan S, Bruzelius M, Håkansson N, Åkesson A, Larsson SC. Lifestyle factors and venous thromboembolism in two cohort studies. Thromb Res. 2021;202:119–24.
23. Jørgensen H, Horváth-Puhó E, Laugesen K, Braekkan S, Hansen JB, Sørensen HT. Socioeconomic status and risk of incident venous thromboembolism. J Thromb Haemost. 2021;19(12):3051–61.
24. Enga KF, Rye-Holmboe I, Hald EM, Løchen ML, Mathiesen EB, Njølstad I, et al. Atrial fibrillation and future risk of venous thromboembolism:the Tromsø study. J Thromb Haemost. 2015;13(1):10–6.
25. Hansen-Krone IJ, Brækkan SK, Enga KF, Wilsgaard T, Hansen JB. Alcohol consumption, types of alcoholic beverages and risk of venous thromboembolism - the Tromsø study. Thromb Haemost. 2011;106(2):272–8.
26. Yuan S, Bruzelius M, Larsson SC. Causal effect of renal function on venous thromboembolism: a two-sample mendelian randomization investigation. J Thromb Thrombolysis. 2022;53(1):43–50.
27. Klarin D, Busenkell E, Judy R, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. Nat Genet. 2019;51(11):1574–9.
28. Mahmoodi BK, Cushman M, Anne Næss I, Allison MA, Bos WJ, Brækkan SK, et al. Association of Traditional Cardiovascular Risk Factors with Venous Thromboembolism: an individual participant data meta-analysis of prospective studies. Circulation. 2017;135(1):7–16.
29. Ageno W, Becattini C, Brighton T, Selby R, Kamphuisen PW. Cardiovascular risk factors and venous thromboembolism: a meta-analysis. Circulation. 2008;117(1):93–102.
30. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Anal. 2007;15(3):199–236.
31. Ke G, Meng Q, Finley T, et al. Lightgbm: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst. 2017;30:3146–54.
32. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765–74.
33. Kurki MI, Karjalainen J, Palta P, et al. FinnGen: Unique genetic insights from combining isolated population and national health register data. medRxiv. 2022.
34. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, et al. Co-regulatory networks of human serum proteins link genetics to disease. Science. 2018;361(6404):769–73.
35. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman ÅK, Schork A, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. Nat Metab. 2020;2(10):1135–48.
36. Yao C, Chen G, Song C, Keefe J, Mendelson M, Huan T, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nat Commun. 2018;9(1):3268.
37. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun. 2017;8:14357.
38. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73–9.
39. Sun BB, Chiou J, Traylor M, et al. Genetic regulation of the human plasma proteome in 54,306 UK biobank participants. *bioRxiv*. 2022.
40. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for mendelian randomization. Stat Methods Med Res. 2017;26(5):2333–55.
41. Edvardsen MS, Hindberg K, Hansen ES, Morelli VM, Ueland T, Aukrust P, et al. Plasma levels of von Willebrand factor and future risk of incident venous thromboembolism. Blood Adv. 2021;5(1):224–32.
42. Sabater-Lleal M, Huffman JE, de Vries PS, Marten J, Mastrangelo MA, Song C, et al. Genome-wide association transethnic meta-analyses identifies novel associations regulating coagulation factor VIII and von Willebrand factor plasma levels. Circulation. 2019;139(5):620–35.
43. Edvardsen MS, Hansen ES, Hindberg K, Morelli VM, Ueland T, Aukrust P, et al. Combined effects of plasma von Willebrand factor and platelet measures on the risk of incident venous thromboembolism. Blood. 2021;138(22):2269–77.
44. Calabrò P, Gragnano F, Golia E, Grove EL. von Willebrand factor and venous thromboembolism: pathogenic link and therapeutic implications. Semin Thromb Hemost. 2018;44(3):249–60.
45. Lacroix-Desmazes S, Repessé Y, Kaveri SV, Dasgupta S. The role of VWF in the immunogenicity of FVIII. Thromb Res. 2008;122(Suppl 2):S3–6.
46. Chauhan AK, Kisucka J, Lamb CB, Bergmeier W, Wagner DD. von Willebrand factor and factor VIII are independently required to form stable occlusive thrombi in injured veins. Blood. 2007;109(6):2424–9.
47. Meltzer ME, Lisman T, de Groot PG, Meijers JCM, le Cessie S, Doggen CJM, et al. Venous thrombosis risk associated with plasma hypofibrinolysis is explained by elevated plasma levels of TAFI and PAI-1. Blood. 2010;116(1):113–21.
48. Frischmuth T, Hindberg K, Aukrust P, Ueland T, Brækkan SK, Hansen JB, et al. Elevated plasma levels of plasminogen activator

inhibitor-1 are associated with risk of future incident venous thromboembolism. J Thromb Haemost. 2022;20(7):1618–26.
49. Wang J, Wang C, Chen N, Shu C, Guo X, He Y, et al. Association between the plasminogen activator inhibitor-1 4G/5G polymorphism and risk of venous thromboembolism: a meta-analysis. Thromb Res. 2014;134(6):1241–8.
50. Ridker PM, Hennekens CH, Lindpaintner K, Stampfer MJ, Miletich JP. Arterial and venous thrombosis is not associated with the 4G/5G polymorphism in the promoter of the plasminogen activator inhibitor gene in a large cohort of US men. Circulation. 1997;95(1):59–62.
51. Angus DC. Drotrecogin alfa (activated)…A sad final fizzle to a roller-coaster party. Crit Care. 2012;16(1):107 107.
52. Hernandez-Lopez R, Chavez-Gonzalez A, Torres-Barrera P, Moreno-Lorenzana D, Lopez-DiazGuerrero N, Santiago-German D, et al. Reduced proliferation of endothelial colony-forming cells in unprovoked venous thromboembolic disease as a consequence of endothelial dysfunction. PLoS One. 2017;12(9):e0183827.
53. Branchford BR, Stalker TJ, Law L, Acevedo G, Sather S, Brzezinski C, et al. The small-molecule MERTK inhibitor UNC2025 decreases platelet activation and prevents thrombosis. J Thromb Haemost. 2018;16(2):352–63.
54. van Sommeren S, Janse M, Karjalainen J, Fehrmann R, Franke L, Fu J, et al. Extraintestinal manifestations and complications in inflammatory bowel disease: from shared genetics to shared biological pathways. Inflamm Bowel Dis. 2014;20(6):987–94.
55. Yuan S, Carter P, Bruzelius M, Vithayathil M, Kar S, Mason AM, et al. Effects of tumour necrosis factor on cardiovascular disease and cancer: a two-sample mendelian randomization study. EBioMedicine. 2020;59:102956.
56. Kjaergaard AD, Johansen JS, Bojesen SE, Nordestgaard BG. Observationally and genetically high YKL-40 and risk of venous thromboembolism in the general population: cohort and mendelian randomization studies. Arterioscler Thromb Vasc Biol. 2016;36(5):1030–6.
57. Pietzner M, Wheeler E, Carrasco-Zanini J, Kerrison ND, Oerton E, Koprulu M, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. Nat Commun. 2021;12(1):6822.
58. Thibord F, Klarin D, Brody JA, et al. Cross-ancestry investigation of venous thromboembolism genomic predictors. *medRxiv*. 2022.
59. Palstrøm NB, Matthiesen R, Rasmussen LM, Beck HC. Recent developments in clinical plasma proteomics-applied to cardiovascular research. Biomedicine. 2022;10(1):162.
60. Petrera A, von Toerne C, Behler J, Huth C, Thorand B, Hilgendorff A, et al. Multiplatform approach for plasma proteomics: complementarity of Olink proximity extension assay technology to mass spectrometry-based protein profiling. J Proteome Res. 2021;20(1):751–62.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Yuan S, Titova OE, Zhang K, Gou W, Schillemans T, Natarajan P, et al. Plasma protein and venous thromboembolism: prospective cohort and mendelian randomisation analyses. Br J Haematol. 2023;201(4):783–792. https://doi.org/10.1111/bjh.18679