Proteomic insights into modifiable risk of venous thromboembolism and cardiovascular comorbidities

Running head: Protein and thrombosis

Shuai Yuan¹, Fengzhe Xu², Han Zhang³, Jie Chen³, Xixian Ruan⁴, Yuying Li⁵, Stephen Burgess⁶,⁷, Agneta Åkesson¹, Xue Li³, Dipender Gill⁸, Susanna C. Larsson¹,⁹

Affiliations
¹ Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.
² Key Laboratory of Growth Regulation and Translational Research of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, China.
³ 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.
⁴ Department of Gastroenterology, The Third Xiangya Hospital, Central South University, Changsha, China
⁵ Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
⁶ MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.
⁷ Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
⁸ Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom.
⁹ Unit of Medical Epidemiology, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden.

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

Essentials
- The protein pathways linking modifiable factors (MF) to VTE are scarcely explored.
- We used Mendelian randomization to investigate mediation of blood proteins in MF-VTE links.
- Several proteins mediated the associations of obesity, smoking, and insomnia with VTE.
- Many VTE-associated proteins had druggable potentials and pleiotropy on cardiovascular diseases.
Abstract

Introduction: Venous thromboembolism (VTE) has been associated with several modifiable factors (MFs) and cardiovascular comorbidities. However, the mechanisms are largely unknown. We aimed to decipher proteomic pathways underlying the associations of VTE with MFs and cardiovascular comorbidities.

Methods: A two-stage network Mendelian randomization (MR) analysis was conducted to explore the associations between 15 MFs and 1,151 blood proteins with VTE using data from a genome-wide meta-analysis including 81,190 VTE. We used protein data from 35,559 individuals as the discovery analysis and from two independent studies including 10,708 and 54,219 participants respectively as the replication analyses. Based on identified proteins, we assessed the druggability and examined the cardiovascular pleiotropy.

Results: The network MR analyses identified 10 MF-VTE, 86 MF-protein, and 34 protein-VTE associations. These associations were overall consistent in the replication analyses. Thirty-eight pathways with directionally consistent direct and indirect effects in the MF-protein-VTE pathway were identified. Low-density lipoprotein receptor-related protein 12 (LRP12, 34.3%-58.1%) and coagulation factor XI (20.6%-39.6%) mediated most of the associations between three obesity indicators and VTE. Likewise, coagulation factor XI mediated most of the smoking-VTE association (40%; 95% confidence interval 20%-60%) and insomnia-VTE association (27%; 95% confidence interval 5%-49%). Many VTE-associated proteins were highly druggable for thrombotic conditions. Five proteins (interleukin-6 receptor subunit alpha, LRP12, prothrombin, angiopoietin-1, and low-density lipoprotein receptor-related protein 4) were associated with VTE and its cardiovascular comorbidities.

Conclusions: This study suggested that coagulation factor XI, a druggable target, is an important mediator of the associations of obesity, smoking, and insomnia with VTE risk.

Keywords: modifiable; proteomics; venous thrombosis
Introduction

Venous thromboembolism (VTE) affects one to two individuals per 1000 in Europe and the USA, and has been ranked top among vascular diseases with regard to prevalence, morbidity, and mortality [1]. Although multifactorial in etiology, the risk of VTE has been fortunately found to be modifiable [1]. Mainly for unprovoked VTE, reducing obesity [2] and adopting a healthy lifestyle [3] appear to be associated with a lower risk of VTE. Even though the classic Virchow Triad theory may help clarify underlying mechanisms behind these associations, data are scarce to illuminate the molecular pathways linking these modifiable factors to VTE. Given that blood proteins have been revealed in thrombosis and coagulation [4, 5] and may be altered by obesity and lifestyle factors [6], we hypothesize that the modifiable risk of VTE may be partly attributed to changes in proteomic features. A clear and comprehensive appraisal of these pathways not only benefits understanding of pathogenesis of VTE but also provides evidence support for better disease prevention and management in a precision manner.

In clinical practice, antithrombotic therapy is used for treating VTE. Despite being effective, the risk of major bleeding during antithrombotic treatment has been found to be considerable [7]. In addition, no current tools can precisely predict bleeding risk [8]. Thus, identifying new targets for attenuating thrombosis with the potential for less bleeding is of great importance for VTE drug development. Blood proteins, usually the principal regulators of molecular pathways, are always treated as therapeutic targets [9]. By conducting a comprehensive investigation on protein-VTE associations, this study aims to pinpoint potential drug targets by integrating clinical, genetic and
omics data from large databases of human data. Cardiovascular risk has been found to be elevated among VTE patients [10]. Thus, we also aim to clarify proteomic etiology of this.

Mendelian randomization (MR) analysis is an epidemiological approach that can reinforce association inference using observational genetic data [11]. Leveraging genetic variants as an instrumental variable for the exposure, the MR design can minimize confounding since genetic variants are randomly assorted at conception and assumed to be unassociated with confounder [11]. This process mimics the randomization process in randomized controlled trials. The MR design can also diminish reverse causation as germline genotypes cannot be modified by the onset or progression of disease [11]. There are three important assumptions of MR. First, the genetic variants used as the instrumental variables should be strongly associated with the exposure (e.g., blood proteins and modifiable factors in this study). Second, the genetic instruments should not be associated with any confounders. Third, the genetic variants used as the instrumental variables should influence the outcome only through the exposure not via other alternative pathways. The violation of the third assumption is known as pleiotropy, among which the unbalanced horizontal type is a major challenge for MR studies [12]. Regarding investigations on the proteome-disease associations, MR design has been usually observed to satisfy key assumptions, particularly when utilizing cis-variants in protein-encoding genes as genetic instruments for the proteins [9, 13]. For phenotypic traits with multiple genetic variants as the instrumental variables, several methods have been developed to generate indication of horizontal pleiotropy and to minimize this by removing pleiotropic genetic instruments [14]. Overall, considering the merits of MR design as well as genetic data available for modifiable factors and
blood proteins, we here conducted this study to disentangle the proteomic pathways linking modifiable factors to VTE as well as to explore shared protein etiologies between VTE and its cardiovascular comorbidities.

**Methods**

**Study design**

*Figure 1* shows the schematic study design of traditional and network MR approaches. In this study, we first conducted proteome-wide MR and colocalization analyses to identify VTE-associated blood proteins. In the proteome-wide MR on VTE, we first selected suitable genetic variants as instrumental variables to mimic the levels of blood proteins with data from the corresponding protein GWASs and then used the MR approach to estimate the associations between a wide range of blood protein and the risk of VTE. Likewise, we performed MR analyses to explore the VTE-associated modifiable factors and further the associations between VTE-associated modifiable factors and VTE-associated blood proteins. Based on this two-stage network MR design, we estimated mediation effects of proteins in the associations between modifiable factors and VTE. We performed reverse MR analysis to rule out reverse causation and conducted replication analyses using data from independent protein genome-wide association studies. Based on identified proteins, we performed two downstream analyses: 1) druggability assessment, and 2) pleiotropic associations with cardiovascular comorbidities. The study used publicly available data generated in published studies (*Table S1*). These studies had been approved by corresponding ethical committee and the participants signed informed consent forms.
**Blood protein data**

In the discovery also primary analysis, summary-level data for 4,907 blood proteins were obtained from genome-wide association studies (GWASs) in 35,559 Icelanders (mean age of 55 years and 50% of women) [15]. Plasma proteins were measured using the SomaScan version 4 assay (SomaLogic). Information on genotyping, imputation, and quality control is described in detail in the original paper [15]. The levels of proteins were under the rank-inverse normal transformation by age, sex, and sample age and standardized. Associations of genetic variants with protein levels were estimated using the linear mixed model. For MR analysis, we selected index $cis$-SNPs associated with protein levels at the genome-wide significance level ($P <5\times 10^{-8}$) as the IVs for proteins. Regarding identified proteins associated with VTE in the discovery analysis, we used two independent protein GWAS sources, the Fenland study and the UK Biobank Pharma Proteomics Project (UKB-PPP), for replication. GWAS summary-level data and IVs were extracted from the Fenland study including 10,708 participants whose blood proteins were profiled using the SomaScan version 4 assay [16]. We additionally selected IVs from a GWAS in the UKB-PPP (N=54,219) [17] to confirm the association where available. In UKB-PPP, proteomic profiling was performed by Olink platform. In both replication data, the IVs were selected under the same criteria as previously described.

**VTE data**

Summary-level data on VTE were extracted from a genome-wide meta-analysis of 81,190 cases and 1,419,671 controls of European ancestry from six cohorts. [18] VTE cases were defined by
hospital or register records (International Classification of Diseases (ICD)-9 or ICD-10). Detailed information on genotyping, imputation, and quality control at the participant-level and gene-level is described somewhere else [18]. Associations of DNA variants with odds of VTE were estimated by logistic regression with at least age (or year of birth), sex and principal components as covariates. We selected single nucleotide polymorphisms (SNPs) strongly associated with VTE at the genome-wide significance level \( P<5\times10^{-8} \) and then clumped these selected SNPs by setting linkage disequilibrium \( r^2 \) threshold at 0.01, which lead to the remaining SNPs as the instrumental variables (IVs) to proxy the genetic liability to VTE in the reverse MR analysis.

**Modifiable factor data**

Our study primarily focused on understanding the links between previously identified and potentially modifiable risk factors, notably obesity and lifestyle habits, with VTE susceptibility, which thus included adiposity indicators (body mass index, waist-to-hip ratio, and visceral adiposity) [2] and lifestyle factors (smoking initiation [19], lifetime smoking index [19], alcohol intake [20], coffee, and caffeine consumption [21], moderate-to-vigorous physical activity [22], and leisure screen time [23]). Recognizing the American Heart Association's emphasis on sleep as a pivotal factor for cardiometabolic health [24], we also incorporated sleep-related traits [25], including sleep duration, short and long sleep duration, daytime napping, and insomnia. Data sources for included modifiable traits are clarified in Table S1. To select genetics IVs, we first extracted SNPs associated with each trait at \( P<5\times10^{-8} \) from the corresponding GWAS. SNPs with high linkage disequilibrium \( (r^2>0.01) \) were pruned and the SNP in linkage disequilibrium with the
lowest \( P \) value was retained. The remaining SNPs were selected as the IVs to proxy the effects of the above modifiable factors.

**Druggability assessment of VTE-associated proteins**

To assess the druggability of the identified VTE-associated proteins, we searched the DrugBank, Dependency Map, the Connectivity Map, the ChEMBL databases. We documented information on drug name and the process of drug development, and classified these protein targets into four classes: 1) approved (at least one drug targeting the protein has been approved); 2) in clinical trials (at least one drug targeting the protein is currently studied in clinical trials); 3) preclinical (at least one drug targeting the protein is in preclinical pipelines); 4) druggable (proteins cannot be identified in drug databases but listed as druggable targets); and 5) currently not listed as druggable.

**Pleiotropic effects of VTE-associated proteins on cardiovascular comorbidities**

To explore VTE-associated cardiovascular comorbidities, we performed a polygenic risk score-Phenome-wide association study (PRS-PheWAS) in the UK Biobank study. The IVs \((P < 5 \times 10^{-8} \text{ and } r^2 < 0.01)\) for VTE were selected from the genome-wide meta-analysis of 81,190 cases and 1,419,671 controls [18]. We included a wide range of clinical outcomes in this analysis and aimed to explore which systems are most influenced by genetic susceptibility to VTE, which may provide support for selecting major comorbidities in the following analysis. Detailed descriptions for PRS construction and PRS-PheWAS analysis, are described in **Supplementary methods**. We also used the MR-Base (a database and analytical platform for MR, [https://www.mrbase.org/](https://www.mrbase.org/)) to confirm
the associations of VTE with common cardiovascular diseases, including coronary artery disease, myocardial infarction, ischemic stroke, atrial fibrillation, heart failure, and peripheral artery disease. Data sources for these outcomes in the MR-Base can be found in Table S1. We performed MR analysis to examine the associations of VTE-associated proteins with these cardiovascular comorbidities and summarized cardiovascular effects of proteins shared by VTE and studied cardiovascular comorbidities using the Open Targets Genetics database [26].

**Statistical analysis**

The $F$ statistic was calculated for each protein to measure the strength of used IV. Proteins without SNPs in the VTE dataset or with IV of $F < 10$ were removed from the analysis. For the MR analysis of protein-VTE associations, the odds ratio (OR) and corresponding confidence interval (CI) of the association were estimated by the Wald ratio test and the delta method, respectively. For the MR analysis of modifiable risk factors and reverse MR analysis, we used the inverse variance weighted method under the multiplicative random effects as the primary analysis, which was supplemented by three sensitivity analyses, including the weighted median, MR-Egger, and MR-PRESSO to examine the robustness of the results and detect potential horizontal pleiotropy. For mediation estimation, the proportion mediated by a protein was calculated as the estimated effect of the modifiable factor on protein levels multiplied by the estimated effect of protein levels on VTE. The propagation of error method (also known as the Delta method) was employed to estimate the standard error associated with the mediation effect [27]. The method is based on the principle that errors in measurements or estimates can propagate and influence the precision of derived quantities. In the context of this MR analysis, the propagation of error method provides
uncertainty estimates surrounding the mediation effect of protein in the association between modifiable risk factors and VTE risk. For proteins, we performed colocalization analysis based on a Bayesian model to test whether the protein and VTE share the same causal variant in the encoding gene region (Supplementary methods) and used a webtool (https://genemania.org/) to map the network among identified proteins [28]. In the primary analysis, we used Bonferroni method or FDR (false discovery rate; Benjamini-Hochberg method) correction to adjust for multiple testing. In the replication analysis, the association with the $P$ value < 0.05 was deemed significant. The analyses were performed using TwoSampleMR, MendelianRandomization, and coloc packages in R software (4.4.1).

**Results**

**Proteome-wide MR analysis identified 34 plasma proteins associated with VTE**

After removing proteins without SNPs in the outcome data or with weak instruments ($F$ statistic <10), the proteome-wide MR analysis included a total of 1151 plasma proteins. The results of all 1151 proteins are shown in Table S2. After Bonferroni correction, genetically predicted levels of 34 proteins were identified to be associated with VTE ($P < 0.05/1151$; Figure 2A). Per standard deviation (SD) increase in genetically predicted protein levels, the OR of VTE ranged from 0.55 (95% CI 0.45-0.68) for PROS1 (Protein S) to 3.27 (95% CI 2.66-4.01) for LPR12 (low-density lipoprotein receptor-related protein 12) (Figure 2B). Among 34 protein-VTE associations, 23 had strong colocalization support with $PH4 >0.8$ and one association had medium colocalization support with $0.8 > PH4 >0.5$ (Figure 2C). Colocalization analysis additionally identified 9 proteins with strong colocalization support with VTE albeit without MR support (Table S3). The network
of VTE-associated proteins is shown in Figure S1. Among these 34 proteins, 34 and 10 proteins had IVs from the Fenland study and UKB-PPP, respectively. We could not obtain IVs for the rest 24 proteins in UKB-PPP due to a smaller number of proteins measured by Olink in this study. For proteins with available IVs, we replicated 31 (91.2%) associations using IVs for protein from the Fenland study and 9 (90%) associations using IVs for proteins from the UKB-PPP (Figure 2D, Table S4 and S5). In the reverse MR analysis, we found no evidence of associations of genetic liability to VTE with the levels of identified blood proteins after multiple testing correction (Table S6).

**Over 10 proteins had a high potential of druggability**

We collected data on clinical trials in four drug databases for drugs targeting VTE-associated proteins identified in the proteome-wide MR analysis to reveal the druggability. We found 12 proteins being targets for drugs approved and nine of them were used to treat thrombotic conditions (Table S7). Most of these proteins were regarded druggable even though in different stages of druggability exploration.

**Ten modifiable factors were associated with VTE**

Genetically proxied 10 out of 15 modifiable factors were associated with VTE after Bonferroni correction (Figure 3). Genetic predisposition to obesity, cigarette smoking, sedentary lifestyle, short sleep duration, daytime napping, and insomnia was associated with an increased risk of VTE (Figure 3). The associations were consistent in sensitivity analyses (Table S8). Heterogeneity was observed in most associations; however, limited indication of horizontal pleiotropy was observed
by MR-Egger intercept test ($P > 0.05$; Table S8). We observed no evidence of associations of genetic liability to VTE with identified modifiable risk factors in the reverse MR analysis (Table S9).

**Associations between VTE-associated modifiable factors and VTE-associated proteins**

In the MR analyses on the associations between VTE-associated modifiable factors and VTE-associated proteins, we set the significance level at the nominal level to reveal as many potential mediation signals as possible. In total, 86 pairs of associations were identified (Table S10). Among VTE-associated modifiable factors, genetically predicted obesity indicators were associated with 28 VTE-associated proteins. Among VTE-associated proteins, there were genetically predicted 7 modifiable factors associated with ADAMTS-like protein 2, 6 with anthrax toxin receptor 2 (ANTR2), and 5 with coagulation factor XI. Most of these associations were observed to be at least directionally consistent in the replication analysis using protein data from the Fenland study (Table S11).

**Mediation of proteins in the associations between modifiable factors and VTE**

We estimated mediation of 38 modifiable factor-protein-VTE combinations where the direction of the total effect (beta of the modifiable factor-VTE association) was in line with the direction of the effect through the mediator (beta of the modifiable factor-protein association $\times$ beta of the protein-VTE association) (Table S12). Thirty of 38 combinations were related to obesity indicators and five proteins (annexin II [Annexin A2], coagulation factor XI, Kininostatin-1, LRP12, and Prekallikrein [plasma kallikrein]) showed consistent mediation effects on the associations of three obesity indicators with VTE risk (Figure 4A). We observed networks (e.g., co-expression, physical
interactions, and pathway) between these proteins mediating the obesity-VTE association (Figure S1). Concerning the magnitude of mediation, LPR12, coagulation factor XI, and prothrombin ranked top for the association between obesity and VTE (Figure 4B). Genetically predicted levels of five proteins mediated the association between cigarette smoking and VTE with the highest mediation for coagulation factor XI (40%; 95% CI 20-60%; Figure 4B). Likewise, genetically predicted levels of three proteins mediated the association between sleep-related traits and VTE with the highest mediation for coagulation factor XI (27%; 95% CI 5-49%; Figure 4B). Overall, genetically predicted levels of annexin II and coagulation factor XI mediated the most associations between studied modifiable factors and VTE (Figure 4C). There were limited networks between annexin II and coagulation factor XI (Figure S1).

**Pleiotropic effects of VTE-associated proteins on cardiovascular comorbidities**

We identified cardiovascular comorbidities associated with VTE by PRS-PheWAS and MR analyses (Figure S2 and Table S13). In the further analysis on the association between VTE-associated proteins and cardiovascular comorbidities, genetically predicted levels of IL-6 sRa (interleukin-6 receptor subunit alpha) were inversely associated with VTE and 4 cardiovascular diseases (Figure 5). Genetically predicted levels of LRP12, prothrombin, angiopoietin-1, and LRP4 (low-density lipoprotein receptor-related protein 4) were positively associated with VTE and 3 cardiovascular diseases (Figure 5). Genetically predicted the above five proteins were also associated with other cardiovascular diseases with the direction being mostly consistent with the associations for VTE (Figure S3).
Discussion

In this study, we first performed a protein-wide MR analysis on VTE, which identified genetically predicted levels of over 30 blood proteins with potential roles in the development of VTE. We then conducted a two-stage network MR analysis to explore the protein pathways linking modifiable factors to VTE. We found that genetically predicted levels of several proteins, in particular annexin II and coagulation factor XI, mediated the MR associations of obesity, smoking, and insomnia with VTE. Many VTE-associated proteins were found to be highly druggable with effects on coagulation-related conditions. We also revealed that LRP12, prothrombin, angiopoietin-1, LRP4, and IL-6sRA were commonly associated with VTE and its cardiovascular comorbidities, which implies the shared etiologies between them. Our findings were overall consistent across the analyses using data from independent protein GWASs and limited reverse causality was observed. Taken together, our findings may facilitate the understanding of the protein pathogenesis of VTE and better guide VTE prophylaxis among feathered populations, such as obese individuals and smokers. The shared proteins between VTE and its cardiovascular comorbidities may identify therapeutic targets across the spectrum of VTE associated cardiovascular multi-morbidity.

Concerning the associations between blood proteins and VTE risk, our study is in line with previous studies and revealed additional signals. A proteome-wide MR study based on 81,669 VTE cases of multiple ancestries identified 23 proteins associated with VTE [4], many of which were confirmed in our study, like that for coagulation factor XI, prekallikrein, prothrombin, and protein S with well-defined function in thrombosis. Consistent with findings from another
protein-wide study [29], we also identified associations of VTE with genetically predicted levels of kininogen 1 and protein C. For other identified proteins in relation to thrombosis in our MR analysis, we noticed supporting evidences from previous studies on phospholipase C gamma 2 (PLCG2) [30], angiopoietin-1 (ANGPT1) [31, 32], glycoprotein VI (GPVI) [33], tyrosine kinase Syk (SYK) [34], N-terminal pro-BNP [35], metalloproteinase inhibitor 3 (TIMP-3) [36], extracellular matrix protein 1 (ECM1) [37], ADAMTS13 [38], protease nexin-1 (SERPINE2) [39], annexin II [40], IL-6 sRa [41], plasma protease C1 inhibitor [42], fibrinogen g-chain dimer [43], Trem-like transcript 1 protein [44], ubiquitin-associated and SH3 domain-containing protein B (UBASH3B, also known as T-cell ubiquitin ligand 2) [45], and regulator of G-protein signaling 18 (RGS18) [46]. Despite limited direct evidence linking VTE with LPR12, this lipid metabolism-related protein has been associated with platelet internalization [47] and vascular endothelial function [48], which thus may be indirectly associated with thrombus formation. Likewise, currently, there is scant evidence concerning the relationship between ADAMTS-like protein 2 (ADAMTSL2) and VTE. However, it is worth noting that ADAMTSL2 belongs to the same protein superfamily as ADAMTS13 [49] with a clear role in thrombosis. We found few studies on the associations of VTE with LRP4, dermatopontin (DPT), endoplasmic reticulum aminopeptidase 2 (ERAP2), apolipoprotein L3 (APOL3), microtubule affinity-regulating kinase 3 (MARK3), lactase-phlorizin hydrolase (LPH), protein phosphatase 1 regulatory subunit 14A (PPP1R14A), alpha-(1,6)-fucosyltransferase (FUT8), or anthrax toxin receptor 2 (ANTXR2). These proteins are related to lipid metabolism, cancer, or immune response. More studies are needed to confirm these associations.
We prioritized protein drug targets by including clinical data from drug databases. Several protein targets, like annexin II, prothrombin, protein C, protein S, ADAMTS13, coagulation factor XI, and fibrinogen g-chain dimer, were reported to have corresponding approved drugs to treat coagulation-related conditions. We also supported glycoprotein VI (GPVI) as a highly druggable target with less bleeding risk, which is under investigation clinical trials [50]. Besides, this study hinted the repurposing values of some proteins approved for other diseases, like prekallikrein for hereditary angioedema [51] and IL-6 sRa for inflammatory diseases [52], in VTE treatment.

Being different from atherosclerotic cardiovascular outcomes,[53] VTE appears to be associated with fewer lifestyle-related factors, mainly obesity [2], smoking [54], and physical inactivity [22], and possibly poor sleep habits [55]. This study found consistent results for obesity [2], smoking [54], and insomnia [55], and reinforced the causality of the associations of physical inactivity and short sleep with VTE for the first time. More interestingly, our two-stage network MR analysis identified genetically predicted levels of some important proteins mediating these lifestyle-VTE associations. For example, genetically predicted levels of annexin II and coagulation factor XI mediated the associations between different modifiable factors (i.e., obesity, smoking, and insomnia) with VTE. The two proteins shared no clear networks, which explained the associations with lifestyle factors in varying magnitudes. In addition, genetically predicted levels of LPR12, coagulation factor XI, and prothrombin consistently mediated most of association between different obesity indicators and VTE. Some networks between these proteins were observed, which explains a high proportion of the obesity-VTE association explained by these proteins and may shed a light in future studies on interactions of these proteins. Identification of these
pathways might not only deepen insight into the pathology underlying the development of VTE, particularly unprovoked VTE from the protein angle, but also guide VTE treatment given that some of these protein mediators are druggable.

The risk of atherosclerotic cardiovascular diseases appears to increase among VTE patients.[10] Even without direct evidence illuminating detailed pathways linking VTE and excessive risk of cardiovascular comorbidities, this study revealed some proteins, like genetically predicted levels of LRP12, prothrombin, angiopoietin-1, LRP4, and IL-6 sRa, shared by etiologies of VTE and other cardiovascular diseases, which may provide clues for cardiovascular risk management among VTE patients. For example, antiplatelet and antiinflammation therapy at a low dose may be useful for both cardiovascular disease and VTE prevention. However, these benefits of this strategy should be carefully assessed against the bleeding risk.

The strengths of this study included MR design that minimized confounding and reverse causation, use of the largest GWAS data on VTE to ensure optimal statistical power, consideration of a large number of blood proteins, and investigation using complementary methods. Limitations of the study deserve to be discussed when interpreting the results. First, even though involving that many proteins, this analysis might miss proteins without suitable genetic instruments or overlook weak associations due to inadequate power. Second, this study was confined to individuals of European ancestry, which might limit the generalizability of our results to other populations. Third, even though the associations for modifiable factors as the exposures proxied by multiple genetic variants were consistent and with limited indication of unbalanced horizontal pleiotropy in the
sensitivity analyses, we could not completely rule out the possibility of horizontal pleiotropy. Forth, the analyses were based on summary-level data, which did not allow the examination of sex- or age-specific effects or stratification by provoking factors.

In summary, this study suggested that genetically predicted levels of several proteins mediated the positive associations of obesity, smoking, short sleep, and insomnia with VTE risk and some of these protein mediators had highly druggable potentials. These findings may benefit the development of VTE prophylaxis and treatment in the high-risk populations.

**Authorship contributions**

S.Y. and S.C.L. conceived and designed the study. S.Y. and F.X. undertook the statistical analyses. S.Y. wrote the first draft of the manuscript. All authors provided important comments to the manuscript and approved the final version of the manuscript.

**Acknowledgments**

We want to acknowledge participants and researchers of all included genome-wide association studies for sharing summary-level data.

**Funding sources**

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). DG is supported by the British Heart Foundation Centre of Research Excellence (RE/18/4/34215) at Imperial College.

**Disclosure**

All authors declare no conflict of interest.

**Data availability**

The study was based on publicly summary-level data that can be downloaded from the cited genome-wide association studies.

**Code availability**

Publicly available software was used to perform the analyses. Code is available from the corresponding author upon reasonable request.

**References**


Figure legends

**Figure 1.** The schematic study design of traditional and network MR approaches. Abbreviations: MR, Mendelian randomization; GWAS, genome-wide association study; PRS-PheWAS, polygenic risk score-phenome-wide association study.

**Figure 2.** Mendelian randomization (MR) and colocalization analyses on the associations between blood proteins and the risk of venous thromboembolism (VTE). Abbreviations: OR, odds ratio; SD, standard deviation. A. the volcano plot of results of the proteome-wide MR analysis of VTE using the discovery deCODE protein data. B. forest plot of identified MR associations between blood proteins and VTE risk using the discovery deCODE protein data. C. results of colocalization analysis based on deCODE protein data. D. comparison of associations in the discovery analysis based on deCODE protein data and replication analyses based on Fenland and UKB-PPP protein data. The protein-VTE association successfully replicated using genetic instruments from Fenland or UKB-PPP was marked by a star sign (*). The full name of proteins listed can be found in Table S2.

**Figure 3.** Associations of modifiable factors with the risk of venous thromboembolism. Abbreviations: CI, confidence interval; FDR, false discovery rate; MVPA, moderate-to-vigorous physical activity; OR, odds ratio.

**Figure 4.** Mediation effects of proteins in the associations between modifiable factors and venous thromboembolism (VTE) risk. A. protein pathways linking obesity, smoking and sleep-related traits to VTE. B. the proportion of association between the modifiable factor and VTE mediated by a protein. C. the count of protein mediators among all identified pathways. The full name of proteins listed can be found in Table S2.

**Figure 5.** Cardiovascular pleiotropy of 34 VTE-associated proteins. Abbreviations: MR, Mendelian randomization; VTE, venous thromboembolism. Grey squares indicate missing values. False discovery rate (FDR) correction was used in this analysis for each outcome. *** indicates FDR <0.001; ** indicates 0.001<FDR <0.01; and * indicates 0.01<FDR <0.05. The full name of proteins listed can be found in Table S2.