Thrombosis Journal

RESEARCH

Open Access

Therapeutic targets for venous thromboemblism: proteome-wide mendelian randomization and colocalization analyses

Xiaolong Li[1], Cheng-hao Yang[2], Min Zhou[2*] and Min-yi Yin[3*]

Abstract

Background Venous thromboembolism (VTE) is a significant global health issue, yet effective therapeutic targets for its prevention and treatment remain elusive. This study aimed to identify plasma proteins causally associated with VTE risk using proteome-wide Mendelian randomization (MR) and colocalization analyses.

Methods We utilized genome-wide association study (GWAS) data from the UK Biobank and FinnGen cohorts, encompassing 38,573 VTE cases and 946,373 controls. Plasma protein levels were quantified using Olink technology in the UK Biobank Pharma Proteomics Project (UKB-PPP) and SomaScan in the deCODE Health study. MR analysis was performed to assess causal relationships, followed by colocalization analysis to evaluate shared genetic variants. Functional enrichment analyses and molecular docking were conducted to explore biological mechanisms and predict potential therapeutic compounds.

Results Eight proteins showed significant associations with VTE risk after Bonferroni correction ($p < 3.19 \times 10^{-5}$). Odds ratios ranged from 0.98 (95% CI: 0.98–0.99) for PLEK to 1.03 (95% CI: 1.02–1.04) for LRP12. Strong colocalization evidence (PH4 ≥ 0.8) was found for LRP12, F11, PLCG2, and ABO. Molecular docking identified promising drug candidates including valine, folic acid, ibrutinib, and simvastatin, with valine showing the strongest binding energy (-32.057 kcal/mol).

Conclusions This study highlights novel therapeutic targets for VTE and provides insights into potential drug candidates. These findings offer a foundation for future research and drug development aimed at reducing VTE risk.

Keywords Venous thromboembolism, Proteomics, Mendelian randomization, Colocalization analysis, Molecular docking, Drug discovery

*Correspondence:
Min Zhou
zhouminvasc@163.com
Min-yi Yin
yinminyi9@126.com
[1] The Department of Thoracic and Cardiovascular Surgery, The first College of Clinical Medical Science, China Three Gorges University, Yichang Central People's Hospital, Yichang, China
[2] Department of Vascular Surgery, Zhongshan Hospital, Institute of Vascular Surgery, Fudan University, National Clinical Research Center for Interventional Medicine, Shanghai, China
[3] Department of Vascular Surgery, Shanghai Ninth People's Hospital, Shanghai JiaoTong University School of Medicine, Shanghai, China

Introduction

Venous thromboembolism (VTE), comprising deep vein thrombosis and pulmonary embolism, represents a significant global health burden with an annual incidence of 1–2 per 1000 individuals [1]. Despite advances in anticoagulation therapy, VTE remains a leading cause of cardiovascular mortality and morbidity worldwide, with substantial healthcare costs and reduced quality of life for affected individuals [2, 3]. The pathogenesis of VTE involves complex interactions between genetic and environmental factors, traditionally explained by Virchow's triad: hypercoagulability, endothelial injury, and blood

stasis [4]. While conventional risk factors such as surgery, immobilization, and cancer are well-established, growing evidence suggests that genetic factors play a crucial role in VTE susceptibility [5]. Recent genome-wide association studies (GWAS) have identified numerous genetic loci associated with VTE risk, highlighting the genetic complexity of this condition [6]. The emergence of high-throughput proteomics technologies has revolutionized our understanding of disease mechanisms by enabling comprehensive analysis of protein expression and regulation [7]. Large-scale proteomics studies, coupled with genetic data, provide unprecedented opportunities to identify novel therapeutic targets through systematic evaluation of protein-disease associations [8]. In particular, the integration of proteomics with genetic studies allows for the identification of proteins that may have causal relationships with disease outcomes [9]. Mendelian randomization (MR) analysis has emerged as a powerful tool for investigating causal relationships between biomarkers and diseases by using genetic variants as instrumental variables [10]. This approach helps minimize confounding and reverse causation, common limitations in observational studies. Furthermore, the addition of colocalization analysis helps distinguish true causal associations from genetic linkage, providing more reliable evidence for potential therapeutic targets [11]. Despite these advances, a comprehensive investigation of plasma proteins as therapeutic targets for VTE using these modern methodological approaches has not been previously conducted. This study aims to identify and validate novel protein targets for VTE prevention and treatment through an integrated approach combining proteome-wide MR, colocalization analyses, and molecular docking studies.

Research methods

Data sources for plasma proteins

We utilized index cis-acting single nucleotide polymorphisms (cis-SNPs) reaching genome-wide significance ($p < 5 \times 10-^{8}$) as instrumental variables for plasma protein quantitative trait loci (pQTLs) from two independent proteome-wide association studies. Protein-coding variants were defined as SNPs located within ±1 Mb of their corresponding gene loci, with linkage disequilibrium (LD) patterns determined using the 1000 Genomes European reference panel. The analysis incorporated data from 54,306 UK Biobank participants in the Pharma Proteomics Project (UKB-PPP) [12], where 1,463 plasma proteins were quantified using Olink technology, yielding 1,161 proteins with suitable cis-SNPs for two-sample Mendelian randomization (MR) analysis. Comparatively, the deCODE Health study employed SomaScan platform to measure 4,907 proteomic aptamers in 35,559 Icelandic individuals, identifying instrumentally valid cis-SNPs for 1,423 plasma proteins [13]. Notably, 509 proteins demonstrated overlapping cis-SNP associations between both cohorts, representing shared genetic instruments across platforms.

Data sources for VTE

GWAS data for VTE were assessed from UK biobank database (https://www.ebi.ac.uk/gwas/studies/GCST90038607) [14]. The total sample size for VTE was 484,598, comprising 12,240 patients with VTE and 472,358 controls [14]. We used the latest release data on VTE from the FinnGen study R12 in this analysis, which comprised 26,333 cases 474,015 controls. There were no sample overlaps between two outcome datasets. In MR analysis, we treated the UK biobank as the discovery study and the FinnGen R12 study as the replication.To increase the power, we meta-analyzed the two GWASs and performed colocalization analysis based on the GWAS meta-analysis data (38,573 cases and 946,373 controls). The GWAS meta-analysis was performed by the METAL software. Genetic variants associated with VTE at $p< 5\times 10 -8$ in this GWAS meta-analysis and with low linkage disequilibrium ($R2 <0.001$) were selected as the instrument variable for VTE in the reverse MR analysis. To assess potential technical biases from different proteomics platforms, we performed a systematic comparison of results for proteins measured on both Olink and SomaScan platforms. Among our identified proteins significantly associated with VTE, eight were measured in both data sources and were used to evaluate cross-platform consistency. For each of these proteins, we compared the direction and magnitude of effect estimates (odds ratios) obtained when using genetic instruments derived from each platform. We calculated Pearson correlation coefficients between effect estimates from both platforms to quantify agreement. Additionally, for each protein, we computed the absolute percentage difference in effect estimates to assess measurement discrepancies. Technical differences between platforms were considered as follows: Olink employs antibody-based proximity extension assays with high specificity, while SomaScan uses aptamer-based technology with broader coverage but potentially different binding properties. To mitigate platform-specific biases, we prioritized results from genetic instruments derived from the platform with stronger instrument strength (higher F-statistics) when proteins were available in both datasets. For our primary findings, we conducted sensitivity analyses using instruments from the alternative platform to ensure robustness. Proteins showing directional inconsistency between platforms

were flagged for additional scrutiny and are clearly indicated in our results.

This cross-platform evaluation approach is consistent with recent large-scale proteome-wide MR studies [15], which have demonstrated that despite technical differences between platforms, causal protein associations showing consistency across measurement technologies are more likely to represent true biological relationships rather than technical artifacts.

MR analysis

When protein-related SNPs were not available in the outcome datasets, we substituted them with proxy SNPs showing high linkage disequilibrium ($R2 \geq 0.8$), using the 1000 Genomes European reference panel [16]. SNPs lacking suitable proxies were excluded [16]. For clumping to obtain independent cis-variants, we applied a stringent LD threshold of $R^2 < 0.001$. Additionally, we excluded variants with minor allele frequency (MAF) < 0.01 to ensure robust genetic instruments. To evaluate instrumental variable strength, we computed F statistics. For analyzing the relationships between plasma proteins and outcomes, we employed the Wald ratio to calculate odds ratios (ORs) and implemented the delta method to determine corresponding confidence intervals (CIs). Results from dual outcome datasets for individual proteins were combined using fixed-effect meta-analysis. The MR effect estimates were standardized to reflect one standard deviation increase in genetically predicted protein levels. Multiple testing correction was implemented using the Bonferroni method for VTE analyses. To validate our findings and address potential platform-specific biases, we conducted sensitivity analyses using genetic instruments derived from the UKB-PPP (Olink platform) for the eight proteins significantly associated with VTE in our primary analysis. This approach allowed us to assess whether the identified associations were robust across different proteomics measurement technologies. All MR analyses were executed using TwoSampleMR and MendelianRandomization packages in R version 4.4.1.

Colocalization analysis

We performed colocalization analysis to evaluate whether the identified associations between proteins and VTE were influenced by linkage disequilibrium. This analysis was conducted using a Bayesian framework that assesses the support for five mutually exclusive hypotheses: 1) no association with either trait; 2) association with trait 1 only; 3) association with trait 2 only; 4) both traits are associated, but distinct causal variants underlie the two traits; and 5) both traits are associated, and the same causal variant is shared between them [17]. A posterior probability was calculated for each hypothesis (H0, H1, H2, H3, and H4). In this study, we set the prior probabilities as follows: the probability of the SNP being associated with trait 1 only (p1) was 1×10^{-4}, the probability of the SNP being associated with trait 2 only (p2) was 1×10^{-4}, and the probability of the SNP being associated with both traits (p12) was 1×10^{-5}. Strong evidence of colocalization was defined as a posterior probability for shared causal variants (PH4) ≥ 0.8, while moderate evidence was defined as $0.5 < \text{PH4} < 0.8$. The analysis was performed using the coloc package in R software (version 4.4.1). To address the limitations of traditional colocalization approaches, which cannot detect scenarios where exposure and outcome traits share multiple causal variants, we employed the SuSiE (Sum of Single Effects) method [15]. This approach integrates proteomics GWAS summary statistics and genetic correlation matrix reference panels based on individuals of European ancestry from the 1000 Genomes Project Phase 3 to identify multiple causal variants.

GeneMANIA analysis

The GeneMANIA platform [18] (https://genemania.org/) synthesizes multiple functional genomics datasets—including genetic interactions, pathway annotations, and co-expression profiles—with supplementary gene-function associations to systematically explore the biological mechanisms of target genes [19].

Molecular docking

Molecular docking techniques were employed to investigate the interactions between drug candidates and their target proteins. The three-dimensional structures of drug candidates were retrieved from the PubChem Compound Database, while the target protein structures were obtained from the RCSB Protein Data Bank [20]. To prepare for semi-flexible docking simulations, protein structures underwent preprocessing in PyMOL 2.4, which included the removal of water molecules and non-essential ligands, followed by the addition of hydrogen atoms. AutoDock Tools 1.5.6 was utilized to convert the prepared structures into PDBQT format files, which are required for docking simulations. The molecular docking analysis was subsequently performed using AutoDock Vina 1.2.2, with binding energies below −5 kcal/mol considered to indicate effective ligand-receptor interactions, and values below −7 kcal/mol suggesting strong binding activity between the drug candidates and their protein targets [21].

Drug candidate identification process

Our drug discovery pipeline followed a systematic multi-step approach to identify potential therapeutic compounds for VTE. First, we assessed the druggability of

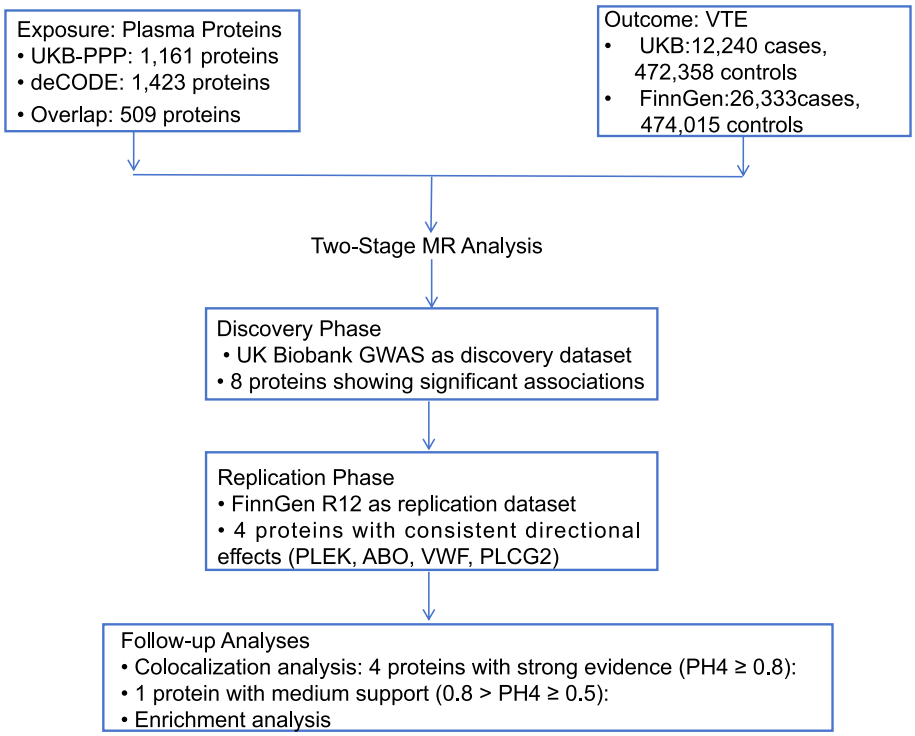

Fig. 1 Overview of the study design in our MR and colocalization study

our identified proteins (PLEK, ABO, VWF, and PLCG2) by cross-referencing them against established drug databases including DrugBank, Open Targets, and ChEMBL to determine whether they are targeted by any approved drugs. Next, we employed the Drug Signature Database (DSigDB), which integrates information from over 19,000 compounds and their molecular targets, to identify potential chemical compounds that might interact with our target proteins. This database encompasses FDA-approved drugs, natural products, and experimental compounds across multiple pharmacological sources. Candidate compounds were then prioritized based on adjusted *p*-values reflecting the statistical significance of their predicted association with our target proteins. Finally, the highest-ranking candidates underwent molecular docking analysis to evaluate binding affinity and predict potential mechanisms of action based on structural interactions.

Results

Associations between plasma proteins and VTE

An overview of the study design is shown in Fig. 1. All analyses were based on summary-level data listed in Table S1. Figure 2 displays the result summary of the analyses of VTE. In the combined analysis of two outcomes data, genetically predicted levels of 8 proteins were significantly associated with the risk of VTE after Bonferroni correction for multiple testing ($p < 3.19 \times 10-5$). Per SD increase in genetically predicted levels of protein, the odds ratio of VTE ranged from 0.98 (95% confidence interval [CI], 0.98–0.99) for PLEK to 1.03 (95% CI, 1.02–1.04) for LRP12 (TableS2). The identified associations were directionally consistent between the discovery and replication studies for 4 proteins (PLEK,ABO,VWF and PLCG2) (Tables S3 and S4). All eight proteins significantly associated with VTE in our primary analysis showed directionally consistent effects when using genetic instruments derived from the UKB-PPP (Olink platform), supporting the robustness of these associations across different proteomics technologies (Table S4). Among identified associations, eight proteins were measured in both outcome datasets and generally showed consistent associations with VTE, despite the proteins being measured on two different profiling platforms. Among MR-identified proteins in relation to VTE, 4 proteins had high support of colocalization analysis (PH4 $\geq$0.8) (Fig. 3 and Table 1), which were LRP12,F11,PLCG2 and ABO. PLEK had medium support of colocalization analysis (0.8 $\geq$ PH4 $\geq$0.5) (Table 1). Sum of single effects (SuSiE) identified strong colocalization support for the associations for PROC (TableS5).

To investigate potential reverse causation, we conducted bidirectional MR analysis examining the effect of genetic liability to VTE on levels of our eight identified

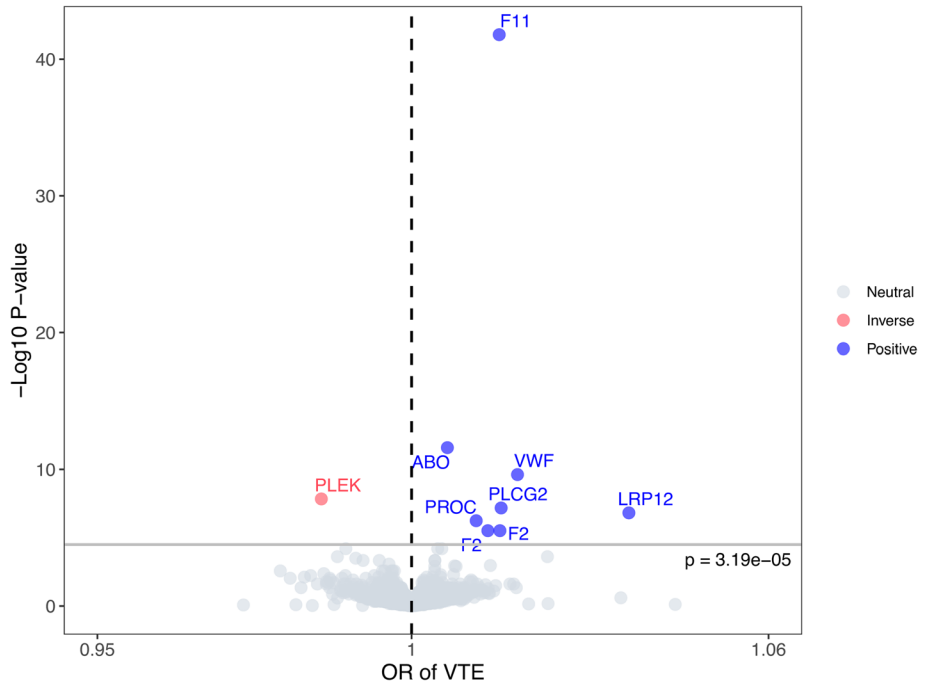

Fig. 2 Result summary of MR on the associations between plasma proteins and the risk of VTE. Plot displaying the odds ratios and 95% confidence intervals for the eight plasma proteins significantly associated with VTE risk. The x-axis represents the odds ratio per standard deviation increase in genetically predicted protein levels, with values below 1.0 indicating protective effects and values above 1.0 indicating increased risk

proteins. Our reverse MR analysis revealed no significant causal effects of VTE genetic liability on any of the eight proteins after correction for multiple testing (all $p >$ 0.00625), suggesting that the protein-VTE associations we identified are unlikely to be explained by reverse causation. The complete results of reverse MR analysis are presented in Table S6.

Exploring the biological significance by enrichment analysis

GO and KEGG analyses provided comprehensive insights into the biological functions, metabolic pathways, and protein interactions of four key proteins: PLEK, ABO, VWF, and PLCG2. GO annotation analysis revealed distinct enrichment patterns across three categories. In biological processes (BP), the proteins showed significant enrichment in blood coagulation, hemostasis, wound healing, and regulation of body fluid levels. Cellular component (CC) analysis identified enrichment in specific structures, including ruffle membrane, Golgi lumen, leading edge membrane, and endoplasmic reticulum lumen. The molecular function (MF) category demonstrated enrichment in various enzymatic activities, particularly serine-type endopeptidase, serine-type peptidase, serine hydrolase, and endopeptidase activities, as well as heparin binding (Fig. 4A). KEGG pathway enrichment analysis highlighted the involvement of these proteins in several critical pathways, including complement and coagulation cascades, platelet activation, phospholipase D signaling, and neutrophil extracellular trap formation (Fig. 4A, Table 2). We constructed an interaction network centered on PLEK, ABO, VWF, and PLCG2 (Fig. 4B), which revealed that these proteins primarily participate in pathways related to blood coagulation and hemostasis.

Candidate drug prediction

To identify potentially effective therapeutic interventions, we utilized the Drug Signature Database (DSigDB) for computational drug prediction analysis. Our investigation revealed that PLEK, ABO, VWF, and PLCG2 currently lack corresponding pharmaceutical interventions, highlighting these targets as promising candidates for future drug development efforts. Through analysis based on adjusted *p*-values, we identified several potential chemical compounds of interest. Notably, valine,folic acid,ibrutinib and simvastatin as the most statistically significant drug candidates, demonstrating strong associations with PLEK, ABO, VWF, and PLCG2 respectively. Structural analysis of the molecular docking results provided insights beyond simple binding affinities. For each protein-drug pair, we examined the binding site location relative to known functional domains and the potential conformational changes induced by ligand binding. The

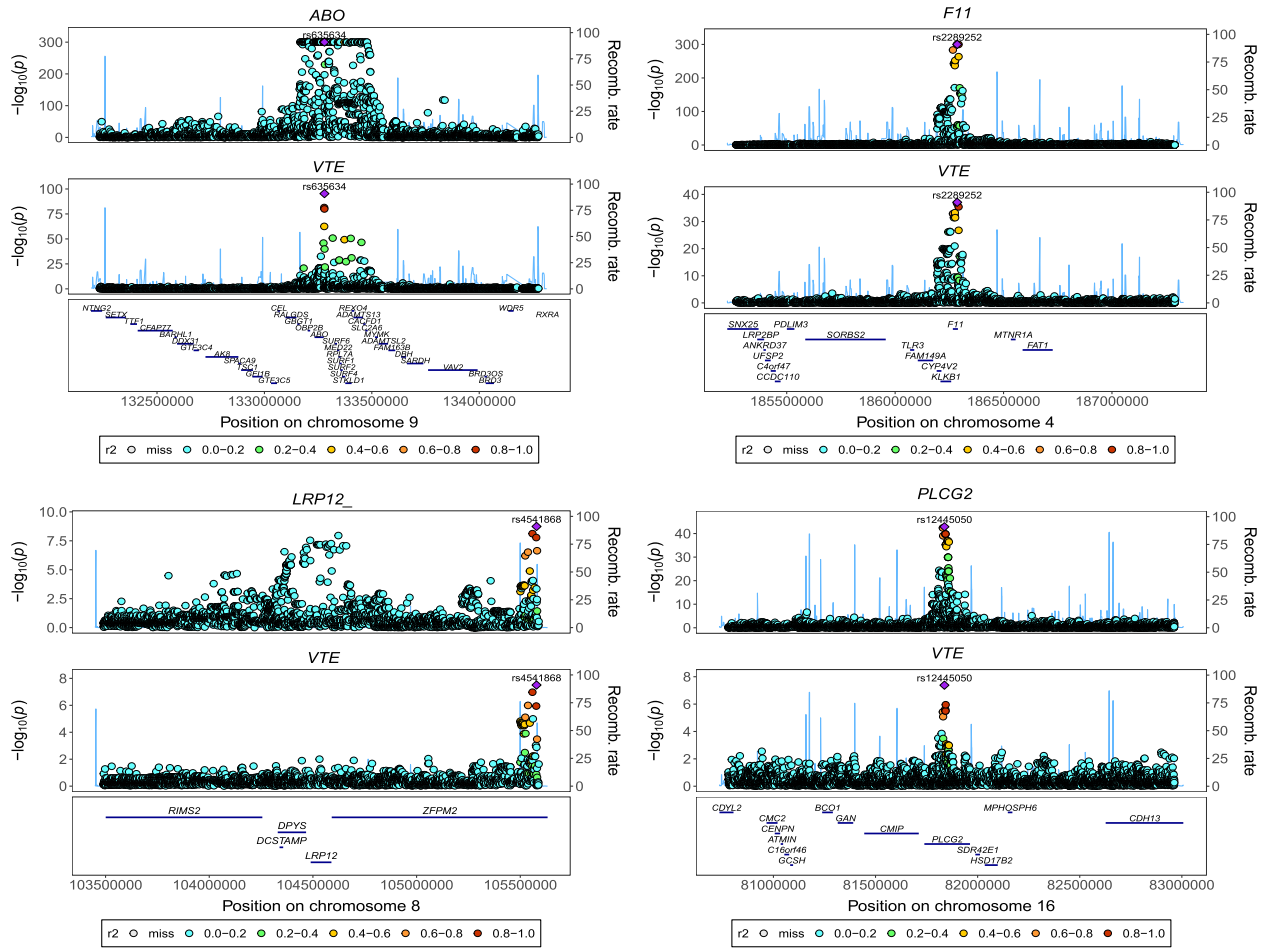

Fig. 3 The results of colocalization analysis between candidate genes and VTE. Visualization of colocalization analysis results showing posterior probabilities for the five hypotheses (H0-H4) tested for each protein-VTE association. Strong evidence of colocalization (PH4 ≥ 0.8) was observed for LRP12, F11, PLCG2, and ABO, indicating shared causal genetic variants between protein levels and VTE risk. PLEK showed moderate colocalization evidence (0.5 < PH4 < 0.8)

Table 1 Colocalization analysis for plasma proteins and VTE

Protein	PP.H0.abf	PP.H1.abf	PP.H2.abf	PP.H3.abf	PP.H4.abf
LRP12	0	0.01	0	0	0.99
VWF	0	0	0	1	0
F2	0	0	0	1	0
F11	0	0	0	0	1
PLCG2	0	0	0	0.01	0.99
PLEK	0	0.11	0	0.27	0.62
PROC	0	0.02	0	0.72	0.26
ABO	0	0	0	0	1

PLEK-valine interaction occurred at a site distinct from the protein's catalytic domain, suggesting an allosteric modulation mechanism that could enhance protein activity. In contrast, the interactions between risk-associated proteins (LRP12, F11, ABO) and their respective compounds involved binding at or near functional domains, indicating potential inhibitory effects. These structural insights align with the directionality of our MR findings, where therapeutic intervention would ideally enhance protective factors (PLEK) while inhibiting risk factors (LRP12, F11, ABO). The binding poses observed for ibrutinib with PLCG2 resembled those seen with known kinase inhibitors, further supporting a potential inhibitory effect consistent with reducing VTE risk.

Molecular docking

In order to assess the affinity of drug candidates (valine,folic acid,ibrutinib and simvastatin) for their targets (PLEK, ABO, VWF, and PLCG2) and from this to understand the drug target's druggability, molecular docking was performed in this study. The binding poses and interactions of 4 drug candidates with 4 protein were

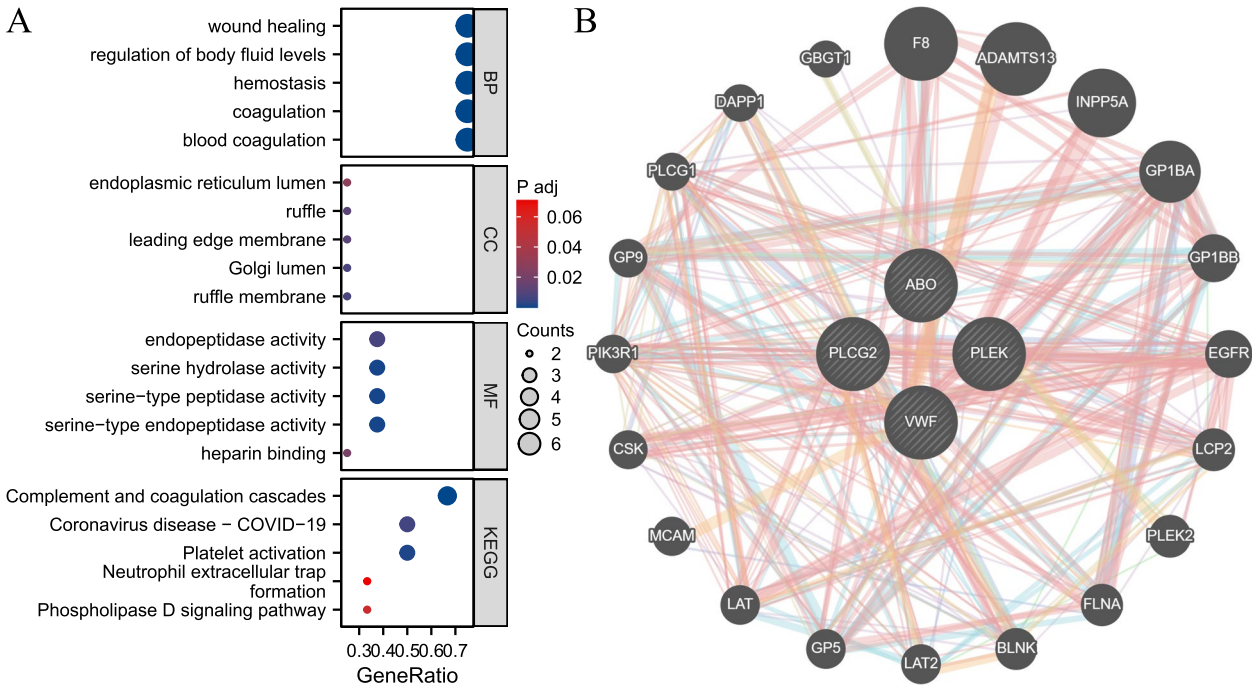

Fig. 4 Enrichment Analysis. **A** GO and KEGG pathway enrichment analysis results for PLEK, ABO, VWF, and PLCG2, showing significant enrichment in biological processes related to blood coagulation, hemostasis, and wound healing, as well as in pathways including complement and coagulation cascades and platelet activation. The color intensity corresponds to statistical significance, and circle size indicates the number of genes involved. **B** Protein–protein interaction network centered on the four key proteins, illustrating their functional connections

Table 2 Enrichment Analysis

Ontology	ID	Description	GeneRatio	BgRatio	*p*value	p.adjust
BP	GO:0007596	blood coagulation	6/8	221/18800	6.77e-11	1.13e-08
BP	GO:0050817	coagulation	6/8	226/18800	7.75e-11	1.13e-08
BP	GO:0007599	hemostasis	6/8	227/18800	7.96e-11	1.13e-08
BP	GO:0050878	regulation of body fluid levels	6/8	382/18800	1.83e-09	1.94e-07
BP	GO:0042060	wound healing	6/8	429/18800	3.67e-09	3.12e-07
CC	GO:0032587	ruffle membrane	2/8	97/19594	0.0007	0.0084
CC	GO:0005796	Golgi lumen	2/8	104/19594	0.0008	0.0084
CC	GO:0031256	leading edge membrane	2/8	175/19594	0.0021	0.0121
CC	GO:0001726	ruffle	2/8	177/19594	0.0022	0.0121
CC	GO:0005788	endoplasmic reticulum lumen	2/8	311/19594	0.0066	0.0286
MF	GO:0004252	serine-type endopeptidase activity	3/8	174/18410	4.49e-05	0.0010
MF	GO:0008236	serine-type peptidase activity	3/8	191/18410	5.92e-05	0.0010
MF	GO:0017171	serine hydrolase activity	3/8	195/18410	6.3e-05	0.0010
MF	GO:0004175	endopeptidase activity	3/8	432/18410	0.0007	0.0077
MF	GO:0008201	heparin binding	2/8	168/18410	0.0022	0.0210
KEGG	hsa04610	Complement and coagulation cascades	4/6	85/8164	1.62e-07	7.6e-06
KEGG	hsa04611	Platelet activation	3/6	124/8164	6.62e-05	0.0016
KEGG	hsa05171	Coronavirus disease—COVID-19	3/6	232/8164	0.0004	0.0067
KEGG	hsa04072	Phospholipase D signaling pathway	2/6	148/8164	0.0047	0.0548
KEGG	hsa04613	Neutrophil extracellular trap formation	2/6	190/8164	0.0076	0.0714

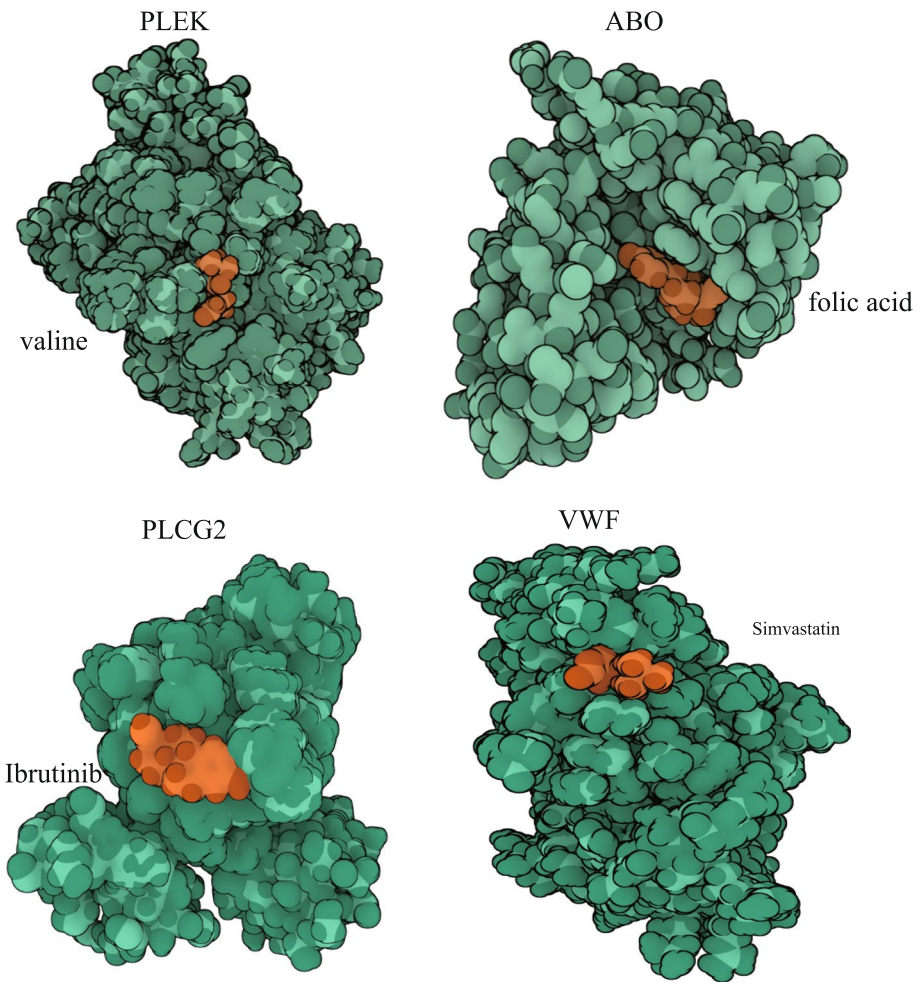

Fig. 5 Molecular docking analysis of candidate genes with potential drug candidates. Computational modeling of binding interactions between identified proteins and candidate compounds. Panel displays the binding poses of valine with PLEK, folic acid with ABO, ibrutinib with PLCG2, and simvastatin with VWF. Hydrogen bonds and key electrostatic interactions are indicated, with binding energies ranging from −7.41 to −32.057 kcal/mol. Valine showed the strongest binding affinity to PLEK (−32.057 kcal/mol), suggesting a potential mechanism for modulating its protective effect against VTE

obtained with Autodock Vina v.1.2.2 and binding energy for each interaction was generated (Fig. 5). Each medication candidate connects to its protein target via visible hydrogen bonds and strong electrostatic interactions. In addition, the binding pocket of each target was successfully occupied by drug candidates. Valine and curcumin exhibited the lowest binding energy (− 32.057 kcal/mol), indicating extremely stable binding.

Discussion

In this comprehensive study combining proteome-wide Mendelian randomization, colocalization analyses, and molecular docking approaches, we identified several plasma proteins causally associated with VTE risk and evaluated potential therapeutic compounds. Our findings provide novel insights into the molecular mechanisms underlying VTE and suggest promising therapeutic targets through a systematic, multi-layered analytical approach. While our Mendelian randomization approach suggests potential causal relationships between the identified proteins and VTE, we must emphasize that MR provides evidence for causality rather than definitive proof. The observed associations represent genetic predispositions that may influence VTE risk, but environmental factors, gene-environment interactions, and complex biological pathways not captured in our analysis could modify these relationships in clinical settings.

Our MR analyses identified eight proteins significantly associated with VTE risk, with four proteins (LRP12, F11, PLCG2, and ABO) showing strong evidence of genetic

colocalization. The identification of ABO among these proteins is particularly noteworthy, as it aligns with previous studies demonstrating the role of ABO blood groups in thrombotic risk [1]. The consistency between our findings and established knowledge not only validates our methodological approach but also strengthens the credibility of our novel discoveries. Furthermore, the identification of F11 as a risk factor aligns with current understanding of the coagulation cascade, providing additional validation of our analytical framework [6].

The discovery of PLEK as a protective factor against VTE (OR: 0.98, 95% CI: 0.98–0.99) represents a particularly interesting and novel finding. PLEK, a protein involved in platelet signaling and cytoskeleton reorganization [22], has not been previously identified as a therapeutic target for VTE. Our colocalization analysis showed moderate support for this association, suggesting a potential causal relationship that warrants further investigation. The protective effect of PLEK might be mediated through its role in platelet function regulation, potentially offering a new therapeutic strategy for VTE prevention.

LRP12's strong association with increased VTE risk (OR: 1.03, 95% CI: 1.02–1.04) represents another notable finding. While the biological mechanism underlying this association requires further investigation, the strong genetic colocalization evidence suggests a genuine causal relationship. This discovery highlights the value of unbiased, proteome-wide approaches in identifying novel disease mechanisms [9]. To contextualize our findings, we compared the effect sizes (odds ratios) of our identified protein targets with established VTE therapeutic targets. For example, our top protein candidate LRP12 showed an odds ratio of 1.03, which is comparable to the effect size observed for genetic variants targeting Factor XI, a clinically validated target currently in late-stage clinical trials. This comparison provides perspective on the potential clinical relevance of our findings while acknowledging that genetic effect sizes do not directly translate to drug efficacy.

The functional enrichment analyses provided comprehensive insights into the biological mechanisms underlying these associations. Our identified proteins showed significant enrichment in blood coagulation, hemostasis, and wound healing pathways, consistent with their role in VTE pathogenesis. The GO analysis revealed enrichment in specific cellular components and molecular functions, suggesting potential mechanisms of action. Particularly noteworthy was the enrichment in serine-type endopeptidase activities and heparin binding, both of which are relevant to thrombosis [23]. The KEGG pathway analysis further illuminated the biological context of our findings, highlighting the involvement of these proteins in complement and coagulation cascades, platelet activation, and phospholipase D signaling. The construction of an interaction network centered on PLEK, ABO, VWF, and PLCG2 revealed complex interconnections between these proteins, suggesting potential synergistic effects in VTE pathogenesis. This network analysis also identified several hub proteins that might serve as additional therapeutic targets [7]. Our molecular docking analyses yielded several promising drug candidates, with valine showing particularly strong binding affinity to its target protein (binding energy −32.057 kcal/mol). The identification of existing drugs like simvastatin as potential therapeutic agents is especially interesting, as it suggests possibilities for drug repurposing [24]. This approach could significantly accelerate the development of new VTE treatments by bypassing several stages of the traditional drug development pipeline. The strong binding affinities observed between ibrutinib and PLCG2, as well as folic acid and ABO, suggest multiple promising therapeutic strategies. It is important to emphasize that our findings represent early-stage discovery rather than clinically validated targets. The pathway from genetic association to effective therapeutic intervention is complex and requires multiple validation steps. While our results provide a promising starting point, substantial additional research—including mechanistic studies, in vitro and in vivo validation, and eventually clinical trials—will be necessary before these protein targets could be translated into clinical applications.

For proteins showing protective associations with VTE (such as PLEK), we hypothesize that identified compounds (valine in this case) may serve as activators or expression enhancers, while for risk-associated proteins (LRP12, F11, ABO), the identified compounds would ideally function as inhibitors. Our structural analysis of the PLEK-valine complex suggests binding at an allosteric site that potentially enhances protein activity, consistent with PLEK's observed protective effect against VTE. However, we emphasize that these computational predictions require experimental validation to confirm both the direction and magnitude of functional effects.

The computational drug prediction analysis revealed that several of our identified proteins (PLEK, ABO, VWF, and PLCG2) currently lack corresponding pharmaceutical interventions, highlighting them as promising candidates for future drug development efforts. The identification of both novel compounds and existing drugs provides multiple avenues for future therapeutic development, potentially leading to more effective and targeted treatments for VTE [25].

Several limitations of our study should be acknowledged. First, while MR analysis helps establish causality, it relies on several assumptions that cannot be fully

verified [10]. These include the assumptions that genetic variants are valid instrumental variables and that there are no alternative pathways through which these variants might affect VTE risk. Second, our findings are based primarily on populations of European ancestry, potentially limiting their generalizability to other populations. Third, the molecular docking analyses, while informative, require experimental validation through in vitro and in vivo studies. Fourth, while technical differences between Olink and SomaScan platforms may introduce variability in protein measurements, the directional consistency of our findings for proteins measured on both platforms strengthens confidence in our results. Nevertheless, we cannot exclude the possibility that platform-specific measurement characteristics might influence effect sizes for some proteins. While MR minimizes confounding by environmental factors, residual confounding from unmeasured variables, such as medication use or comorbidities, could still influence our findings if these factors are associated with VTE risk through pleiotropic pathways. Our study was conducted exclusively in individuals of European ancestry to minimize population stratification, which may limit the generalizability of our findings to other populations. Future MR studies incorporating multi-ancestry GWAS data are needed to confirm the robustness of these associations across diverse genetic backgrounds. We acknowledge that our molecular docking results represent purely computational predictions that require experimental validation. While these in silico analyses suggest potential interactions between candidate drugs and target proteins, we recognize that target engagement and therapeutic relevance can only be definitively established through functional assays such as platelet aggregation tests, thrombin generation assays, or appropriate animal models. The absence of such experimental validation is a limitation of the current study.

Our findings have important implications for future research and clinical practice. The identification of novel protein targets, particularly PLEK and LRP12, opens new avenues for therapeutic intervention. The potential repurposing of existing drugs like simvastatin for VTE prevention could provide more immediate clinical applications, while the strong binding affinity of valine suggests promising directions for new drug development. Furthermore, the identification of multiple protein targets suggests the possibility of combination therapies targeting different pathways simultaneously [26].

The integration of genetic and proteomic data in our study provides a framework for future investigations into disease mechanisms and therapeutic targets. Our findings suggest that a systems biology approach, combining multiple levels of biological information, can yield valuable insights into complex diseases like VTE. This approach could be extended to other cardiovascular diseases and therapeutic areas [12].

Conclusion

In conclusion, this proteome-wide Mendelian randomization study has identified eight proteins potentially causally associated with VTE risk. These findings contribute to our understanding of VTE pathophysiology and highlight promising directions for future research. Based on our results, we propose a research roadmap that begins with prioritization of PLEK, ABO, VWF, and PLCG2 for in-depth functional validation studies, including mechanistic experiments to elucidate their roles in thrombosis and hemostasis. This should be followed by development and testing of selective inhibitors or activators for these protein targets in pre-clinical models, investigation of whether existing approved drugs that modulate these proteins could be repurposed for VTE prevention or treatment, and design of proof-of-concept clinical studies to assess the safety and efficacy of targeting these proteins in high-risk populations. Our computational analyses provide a foundation for these future investigations, but we acknowledge that substantial additional research is required before clinical translation becomes feasible. The proteins identified in this study represent potential therapeutic targets rather than clinically validated ones, and their ultimate utility will depend on rigorous experimental validation and clinical assessment.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12959-025-00733-9.

Table S1. Data sources for studied phenotypes.

Table S2. Associations of plasma proteins with the risk of VTE in the meta-analysis.

Table S3. Associations of plasma proteins with the risk of VTE in the UKB.

Table S4. Associations of plasma proteins with the risk of VTE in the FinnGen.

TableS5. Sum of single effects analysis for plasma proteins and VTE.

Table S6. Results of reverse Mendelian randomization analysis examining the causal effect of VTE genetic liability on protein levels.

Acknowledgements
Not applicable.

Clinical trial number
Not applicable.

Authors' contributions
All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to

submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work.

Funding
This work was supported by the National Natural Science Foundation of China (Grant No. 82470510) and the Science and Technology Innovation Action Plan of Shanghai (Project No. 24SF1902200).

Data availability
Correspondence and requests for materials should be addressed to Min-yi Yin.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Patients gave consent for publication.

Competing interests
The authors declare no competing interests.

Received: 15 February 2025 Accepted: 6 May 2025
Published online: 26 May 2025

References

1. Di Nisio M, van Es N, Büller HR. Deep vein thrombosis and pulmonary embolism. Lancet (London, England). 2016;388:3060–73. https://doi.org/10.1016/S0140-6736(16)30514-1.
2. Sharathkumar AA, Biss T, Kulkarni K, Ahuja S, Regan M, Male C, et al. Epidemiology and outcomes of clinically unsuspected venous thromboembolism in children: A systematic review. Journal of Thrombosis and Haemostasis : Jth. 2020;18:1100–12. https://doi.org/10.1111/jth.14739.
3. Maughan BC, Marin M, Han J, Gibbins KJ, Brixey AG, Caughey AB, et al. Venous thromboembolism during pregnancy and the postpartum period: risk factors, diagnostic testing, and treatment. Obstet Gynecol Surv. 2022;77:433–44. https://doi.org/10.1097/OGX.0000000000001043.
4. Fujieda Y, Amengual O. New insights into the pathogenic mechanisms and treatment of arterial thrombosis in antiphospholipid syndrome. Eur J Rheumatol. 2021;8:93–9. https://doi.org/10.5152/eurjrheum.2020.20058.
5. Crous-Bou M, Harrington LB, Kabrhel C. Environmental and genetic risk factors associated with venous thromboembolism. Semin Thromb Hemost. 2016;42:808–20.
6. Klarin D, Busenkell E, Judy R, Lynch J, Levin M, Haessler J, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. Nat Genet. 2019;51:1574–9. https://doi.org/10.1038/s41588-019-0519-3.
7. Geyer PE, Holdt LM, Teupser D, Mann M. Revisiting biomarker discovery by plasma proteomics. Mol Syst Biol. 2017;13:942. https://doi.org/10.15252/msb.20156297.
8. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1–22. https://doi.org/10.1093/ije/dyg070.
9. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558:73–9. https://doi.org/10.1038/s41586-018-0175-2.
10. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. Bmj (Clinical Research Ed). 2018;362:k601. https://doi.org/10.1136/bmj.k601.
11. Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, et al. A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics (Oxford, England). 2018;34:2538–45. https://doi.org/10.1093/bioinformatics/bty147.
12. Sun BB, Chiou J, Traylor M, Benner C, Hsu Y, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023;622:329–38. https://doi.org/10.1038/s41586-023-06592-6.
13. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet. 2021;53:1712–21. https://doi.org/10.1038/s41588-021-00978-w.
14. Dönertaş HM, Fabian DK, Valenzuela MF, Partridge L, Thornton JM. Common genetic associations between age-related diseases. Nature Aging. 2021;1:400–12. https://doi.org/10.1038/s43587-021-00051-5.
15. Yuan S, Xu F, Li X, Chen J, Zheng J, Mantzoros CS, et al. Plasma proteins and onset of type 2 diabetes and diabetic complications: Proteome-wide Mendelian randomization and colocalization analyses. Cell Rep Med. 2023;4: 101174. https://doi.org/10.1016/j.xcrm.2023.101174.
16. Brion MA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42:1497–501. https://doi.org/10.1093/ije/dyt179.
17. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10: e1004383. https://doi.org/10.1371/journal.pgen.1004383.
18. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol. 2008;9(Suppl 1):S4. https://doi.org/10.1186/gb-2008-9-s1-s4.
19. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38:W214–20. https://doi.org/10.1093/nar/gkq537.
20. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res. 2021;49:D1388–95. https://doi.org/10.1093/nar/gkaa971.
21. Morris GM, Huey R, Olson AJ. Using AutoDock for ligand-receptor docking. Curr Protocols Bioinform Chapter. 2008;8:8–14. https://doi.org/10.1002/0471250953.bi0814s24.
22. Mackman N, Bergmeier W, Stouffer GA, Weitz JI. Therapeutic strategies for thrombosis: new targets and approaches. Nat Rev Drug Discovery. 2020;19:333–52. https://doi.org/10.1038/s41573-020-0061-0.
23. Montoro-García S, Schindewolf M, Stanford S, Larsen OH, Thiele T. The role of platelets in venous thromboembolism. Semin Thromb Hemost. 2016;42:242–51. https://doi.org/10.1055/s-0035-1570079.
24. Schmidt AF, Swerdlow DI, Holmes MV, Patel RS, Fairhurst-Hunter Z, Lyall DM, et al. PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2017;5:97–105. https://doi.org/10.1016/S2213-8587(16)30396-5.
25. Suzuki K. The multi-functional serpin, protein C inhibitor: beyond thrombosis and hemostasis. J Thrombosis Haemostasis h. 2008;6:2017–26. https://doi.org/10.1111/j.1538-7836.2008.03181.x.
26. Wolberg AS, Rosendaal FR, Weitz JI, Jaffer IH, Agnelli G, Baglin T, et al. Venous thrombosis. Nat Rev Dis Primers. 2015;1:15006. https://doi.org/10.1038/nrdp.2015.6.

Publisher's Note