The plasma proteome and risk of future venous thromboembolism – results from the HUNT study.

Sigrid K. Brækkan[1,2], Asbjørn L. Onsaker[2], Therese H. Nøst[3,4], Weihong Tang[5], Kristian D. Hindberg[1], Vania M. Morelli[1,2], Weihua Guan[6], Christian Jonasson[7], Aaron R. Folsom[5], Kristian Hveem[3,7], John-Bjarne Hansen[1,2]

1. Thrombosis Research Center (TREC), Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway

2. Thrombosis Research group (TREC), Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway

3. HUNT Center for Molecular and Clinical Epidemiology, Norwegian University of Science and Technology, Trondheim, Norway

4. Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway

5. Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN 55454, USA

6. Division of Biostatistics, University of Minnesota School of Public Health, Minneapolis, MN, 55455, USA

7. HUNT research Center, Norwegian University of Science and Technology, Levanger, Norway

Short title: Plasma proteome and VTE

Word count abstract: 250

Word count text: 3507

Number of tables: 3 (+2 Suppl)

Number of figures: 3

Number of references: 68

Correspondence to: Sigrid K. Brækkan, Thrombosis Research Group (TREC), Department of Clinical Medicine, UiT – the Arctic University of Norway, 9037 Norway. E-mail: Sigrid.brakkan@uit.no, Phone: +4777625106.

Abstract

Background: This study aimed to identify novel plasma proteins associated with first-lifetime venous thromboembolism (VTE) and molecular pathways involved in VTE pathogenesis.

Methods: A case-cohort comprising incident VTE-cases (n=294) and a randomly sampled age- and sex-weighted subcohort (n=1066) was derived from the Trøndelag Health Study (HUNT3, n=50800). Blood samples were collected and stored at cohort inclusion (2006-2008), and participants were followed up to five years. Proteome-wide analyses was performed using the 7k SomaScan® proteomics platform, and weighted Cox-regression models adjusted for age, sex and sample batch were conducted, with the Bonferroni method applied to account for multiple testing. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were applied on the top-ranked 200 proteins associated with VTE.

Results: Out of 7288 human proteins, seven proteins were significantly associated with higher VTE risk with p-value $<6.9 \times 10^{-6}$ (hazard ratios per 1 standard deviation increase in protein levels ranging from 1.39 to 1.86). Except for Coagulation factor VIII and Tumor necrosis factor soluble receptor II, these proteins were novel associations and included Collagen alpha-3(VI):BPTI/Kunitz inhibitor, Histo-blood group ABO system transferase, Peroxidasin, Human epididymis protein 4 and Regulator of G Protein Signaling 3. KEGG analyses of the top-ranked 200 proteins revealed significant pathway enrichment of nine proteins in the complement (mainly lectin pathway) and coagulation (mainly intrinsic pathway) cascades.

Conclusions: Our proteome-wide analysis led to novel discovered of five protein candidates associated with five-year risk of future VTE. KEGG analyses supported an interplay between the complement and coagulation pathways in the pathogenesis of VTE.

Keywords: Proteomics, Plasma, Aptamer, Venous thromboembolism, Deep vein thrombosis, Pulmonary embolism

Introduction

Venous thromboembolism (VTE) is a frequent and potentially life-threatening disease, affecting more than 1 in 12 individuals during their lifetime.[1,2] The disease burden of VTE is substantial due to a high frequency of severe complications,[3-5] as well as high recurrence- and mortality rates.[6,7] Around 40% of all VTE events occur in the absence of known provoking factors,[8-10] and the pathophysiological mechanisms and molecular pathways involved in initiation of venous thrombus formation are not completely understood [11]. Hence, there is a need to discover novel predictive biomarkers and molecular mechanisms of VTE to improve risk prediction and pursue targeted prevention and treatment to reduce the individual suffering and socioeconomic burden of VTE in society.[12]

A systematic profiling of the plasma proteome can provide new insight into molecular pathways involved in the pathogenesis of VTE and reveal potential predictive biomarkers or drug targets. Three previous studies have applied proteomic-techniques for the detection of blood biomarkers for first-lifetime VTE, but these studies were restricted to a limited number of proteins,[13-15] a low number of participants,[14,15] or blood samples collected after the VTE.[14] Proteomic techniques have evolved rapidly during recent years, currently facilitating high quality measurements of thousands of proteins from small sample volumes. The SomaScan® aptamer-based proteomics platform currently enables measurement of >7000 proteins in plasma.[16] As the previous proteomic studies on VTE were restricted to up to 500 proteins, a larger proteome-wide discovery has the potential to identify new biomarkers of VTE risk. In a population-based case-cohort study, we decided to apply an unsupervised approach by using aptamer-based proteomics on blood samples collected at cohort inclusion (i.e., before VTE), with the aims to (i) identify plasma proteins associated with 5-year risk of future VTE, and (ii) assess potential pathways involved in the pathogenesis of VTE by performing pathway enrichment analyses.

Methods

Study population – the HUNT3 Survey

The study population was recruited from the third survey of the Trøndelag Health study (HUNT), a population-based cohort of inhabitants of the (former) Nord-Trøndelag county.[17] The third survey was conducted in 2006-2008, and all inhabitants aged 20 years or older were invited. In total, 50800 individuals participated (54% of those invited). All participants provided written informed consent. Each participant was followed for up to five years from the inclusion date, and all first-lifetime VTE events occurring among the participants during the study period were recorded.

Baseline measurements in HUNT3

Baseline information was collected at inclusion in the HUNT3 survey by physical examination, blood samples and validated self-administered questionnaires. Non-fasting blood samples were collected into 10 ml vacutainer tubes (Becton, Dickinson, NJ) containing EDTA as anticoagulant. The tubes were preserved at 4 °C immediately after sampling, but not centrifuged until the samples arrived the biobank. The collected samples were temperature-monitored at the screening sites and during the transport to HUNT biobank in the evening. Next morning, EDTA samples were centrifuged at 2500g for 15 minutes at 6 °C, before fractionation and dispersion into 2D Matrix tubes (Matrix™ 2D, ThermoFisher, USA). Fractionation was performed below 10 °C using the RTS Blood Fractionation System (RTS Life Science, Manchester, UK). Aliquots of EDTA plasma were immediately frozen at –80 °C and stored in the HUNT biobank. A detailed description of the data collection at HUNT3 and storage in the HUNT biobank can be found in the Data Resource Profile publication.[18] The HUNT biobank is certified according to the ISO standard 9001.

Body mass index was calculated in kg/m^2 based on height and weight measured at the physical examination (participants wearing light clothing and no shoes). Information on daily smoking (yes/no), history of cardiovascular disease (CVD; included angina, stroke and myocardial infarction) and history of diabetes was obtained from the questionnaire. Renal function was assessed as estimated glomerular

filtration rate (eGFR) based on the Chronic Kidney Disease Epidemiology (CKD-EPI) collaboration equations.[19]

Outcome registration and validation of venous thromboembolism

All potential VTE cases during follow-up were identified by searching the hospital discharge diagnosis registry and autopsy registry at the hospitals covering the catchment area of the HUNT participants (Levanger hospital, Namsos hospital, and St. Olavs hospital) using a broad search with relevant International Classification of Diseases revision 10 (ICD-10) codes (I26.0, I26.9, I67.6, I80.0-I80.3, I80.8, I80.9, I81, I82.0-I82.9, O22.3, O22.5, O87.1, O87.3) in the years 2006-2019. The medical record of each potential VTE case was thoroughly reviewed, and VTE events were adjudicated and recorded when signs and symptoms of lower extremity DVT or PE were objectively confirmed by radiological procedures (ultrasound, venography, computed tomography pulmonary angiography [CTPA], ventilation-perfusion scan) or autopsy. Cases of concomitantly confirmed DVT and PE were classified as PE. Information on clinical risk factors and potential provoking factors in the three months preceding the event was extracted from the medical records using a standardized form.

Study design

This project was named the PRoteomic profiling and Etiological understanding of VENous Thromboembolism (PREVENT). As depicted in Figure 1, we used a case-cohort design which included all incident VTE cases during the five first years of follow up, and a randomly sampled age- and sex-weighted subcohort of individuals from the HUNT3 cohort (source cohort, n= 50800). Due to a concomitant collaboration with a project on abdominal aortic aneurism (AAA), which would use the same subcohort, 100 participants with prevalent AAA before baseline inclusion were excluded from the source cohort. In addition, participants with a history of VTE (n=622) or cancer (n=1908) prior to baseline inclusion in HUNT3, as well as participants who did not have plasma samples available (n=1657), were excluded. Participants with VTE, AAA and cancer before inclusion in HUNT3 were

identified by relevant ICD-10 codes (Supplementary table 1) derived from the hospital discharge registries at the three Nord-Trøndelag hospitals in the period 1999-2008.

Among the 46513 source cohort members eligible for our study, 294 had an incident VTE during the first five years of follow-up. We included all VTE cases, and a subcohort (n=1085) randomly sampled from the source cohort based on the sex- and age-distribution of the VTE cases in 5-year age groups. Due to the sampling nature of the subcohort, 8 VTE cases were also included in the subcohort. In total, 19 subcohort samples did not pass all the quality steps before the proteomics measurements, and the final subcohort population therefore consisted of 1066 individuals.

Proteomics measurements

In 2022, samples from the cases and subcohort participants were retrieved from the HUNT biobank, thawed, aliquoted, and shipped frozen on dry ice to SomaLogic Inc. (Boulder Colorado, USA) for analysis. The SomaScan® 7k aptamer-based platform was used for protein quantification.[20] Each SOMAmer® has a high affinity towards a specific protein, and is synthesized with a fluorophore, photocleavable linker and biotin.[16] Plasma and streptavidin bead-coupled aptamers were incubated together, and after multiple washing steps to eliminate unbound aptamers and proteins, the bead-bound proteins were biotinylated and photocleaved off the bead, releasing aptamer-protein complexes back into solution. Biotinylated proteins were recaptured on new streptavidin beads, and the aptamers were eluted from the proteins. Fluorophores were measured following hybridization to custom arrays of aptamer complementary oligonucleotides, and the resulting raw intensities were reported in relative fluorescence units. Reference samples included on each plate were used to process the raw intensities by hybridization normalization, median signal normalization, plate-scale normalization and inter-plate calibration following SomaLogic normalization procedures.

Statistical analyses – proteome wide association with VTE

Data processing and statistical analyses were carried out using R (version 4.1.3, R Core Team, 2022) and the 'SomaDataIO' package.[21] Each protein measurement derived from the SomaScan® data was $\log_2$ transformed and standardized to a mean of zero and a standard deviation (SD) of 1. Person-time of follow-up was calculated from the date of inclusion in HUNT3 to the date of an incident VTE, migration, death or end of follow-up (i.e., five years after inclusion date), whichever came first. For proteome-wide analyses, weighted Cox proportional hazards regression models adjusted for age, sex and sample batch were used to test for associations between each protein and VTE (one model per protein) using the 'cchs' (Cox model for case-cohort data with stratified subcohort-selection) package in R.[22] The hazard ratios (HRs) per one SD increase in normalized protein values were presented in a volcano plot. The Bonferroni method was applied to account for multiple testing, yielding a significance threshold of $p<6.9x10^{-6}$ (0.05/7288) for the proteome-wide analyses. For the proteins that reached statistical significance, HRs of VTE per one SD increase in normalized protein values were additionally estimated after adjustment for BMI (Model 2: age, sex, sample batch and BMI) and eGFR (Model 3: age, sex, sample batch and eGFR), respectively.

To further explore pathways involved in the pathogenesis of VTE, we performed Kyoto encyclopedia of Genes and Genomes (KEGG) analysis,[23] using the application http://bioinformatics.sdstate.edu/go/. We included the top-ranked 200 proteins associated with VTE, ranked by the lowest p-value in the proteome-wide analyses. The analysis was restricted to human pathways with a minimum of five proteins, and a false discovery rate (FDR)<0.05 was considered significant.

Results

Characteristics of the study population

The mean age of the cases and subcohort members was 66 and 64 years, and the proportion of women was 51% and 45%, respectively (Table 1). The mean BMI was 28.4 kg/m^2 among cases and 27.6 kg/m^2 in the subcohort, while the proportion of individuals with a history of arterial cardiovascular disease

was 7% in cases and 6% in the subcohort. Among the VTE cases, 38% were DVTs while 62% were PEs, and 39% were unprovoked. In total, 19% of the VTEs were related to active cancer.

Proteome-wide analyses

Among 7288 human proteins measured, proteome-wide analyses identified seven proteins that were significantly associated with higher VTE risk with a p-value $<6.9x10^{-6}$ (Figure 2). These were Coagulation factor VIII (FVIII), Regulator Of G Protein Signaling 3 (RGS3), Collagen alpha-3(VI):BPTI/Kunitz inhibitor, Histo-blood group ABO system transferase (BGAT), Peroxidasin (PXDN), Tumor necrosis factor soluble receptor II (TNF sR-II) and Human epididymis protein 4 (HE4). The HRs with 95% confidence intervals per 1 SD in log_2-transformed protein levels are presented in Table 2. The strongest association with VTE was observed for FVIII (HR per SD: 1.86, 95% C 1.59-2.18), while HRs for the other proteins were within the range of 1.39-1.44 (Table 2). Further adjustments for BMI (Model 2) and eGFR (Model 3) had negligible impact on the results.

An overview of the top-ranked 200 proteins associated with VTE in our study (ranked by p-value in ascending order) is presented in Supplementary table 2. Among these proteins, 54 were associated with lower risk of VTE (HR<1), while 146 were associated with higher risk of VTE (HR>1), and the p-value cut-off for the top-ranked 200 proteins was p<0.0062.

Pathway enrichment analyses

Among the top-ranked 200 proteins associated with VTE, we identified four KEGG pathways displaying over-representation of proteins (genes) (Table 3). Of these pathways, the *complement and coagulation cascade* pathways showed the largest fold enrichment, with 9 genes contributing to a 14-fold pathway enrichment (FDR for enrichment: $3.3x10^{-6}$). These genes were: *F9, VWF, C4BPA, F10, MASP1, VSIG4, MBL2, F8* and *CD55* (corresponding to the proteins: Coagulation factor IX, von Willebrand factor, Complement component 4 binding protein alpha, Coagulation factor X, Mannan-binding lectin serine protease 1, Complement receptor of immunoglobulin superfamily, Mannose-binding lectin,

Coagulation factor FVIII, and Complement decay-accelerating factor). The KEGG diagram of the complement and the coagulation cascades, highlighting the proteins that were enriched in those who developed VTE, compared to the subcohort, is presented in Figure 3. Of note, the enriched proteins were mainly located in the intrinsic pathway of the coagulation system and the lectin pathway of the complement system (Figure 3).

Discussion

Among 7288 proteins measured, seven were significantly associated with five-year risk of VTE after Bonferroni correction for multiple testing. Of these, FVIII is a well-established risk factor for VTE[24-29] and tumor necrosis factor receptor 2 (TNFR2) was found in one previous proteomics study on VTE,[13] whereas the other five protein candidates were considered novel discoveries with a potential role in thrombogenesis. KEGG pathway analyses of the top-ranked 200 proteins showed significant enrichment of proteins in the complement and coagulation cascades in relation to VTE, and the enriched proteins were mainly located in the intrinsic pathway of the coagulation system and the lectin pathway of the complement system. Our findings reinforce the involvement of the complement system in the pathogenesis of VTE and provide new information on protein candidates that may have the potential to be utilized as predictive biomarkers or targets for intervention.

Only a few previous studies have applied proteomics-techniques to identify proteins associated with first-lifetime VTE in a population-based setting.[13-15] In a discovery study of 88 VTE cases and 85 controls, targeting 408 candidate proteins, significant associations were found for platelet-derived growth factor β (PDGFB) and von Willebrand Factor (vWF), which remained after replication in 580 cases and 589 controls.[14] Of note, the blood samples were collected after the VTE event, entailing a potential risk of reverse causation.[14] In a nested case-control study of 80 VTE patients and 88 controls, with blood samples collected at cohort inclusion (i.e., before VTE), 501 proteins were measured using an untargeted mass-spectrometry approach.[15] Of these, 46 proteins were associated

with VTE at *p*-values below 0.05,[15] and some of these proteins were further evaluated in a larger study using enzyme immunoassays (EIAs) for protein assessment.[30,31] In a cohort of 11803 participants, with 353 VTE cases during a median of 6.6 years of follow-up, 257 proteins were measured using the Olink platform.[13] At an FDR<0.1, 21 proteins were significantly associated with VTE.[13] Of note, only one of the protein candidates from the previous studies (i.e., TNFR2 [13]), overlapped with the seven top-ranked proteins in our study. The low degree of overlap with previous studies can be explained by several factors. First, a much larger number of proteins was assessed in our study than in previous studies (i.e., 7000 vs. 500). Second, use of the Bonferroni threshold, which is more conservative than the FDR when adjusting for multiple testing, limited the number of significant proteins to those with a very low p-value ($p<0.0000069$). Third, differences in specificity and targets between different proteomic platforms could lead to differences in protein detection. For some proteins, the SomaScan® technology may have limited specificity, as a single aptamer may not be sensitive for changes in protein structure (e.g., posttranslational modifications, genetic polymorphisms or degradation). Low specificity that influence protein quantification will lead to dilution of the results (i.e., underestimation of hazard ratios) and thereby potentially missed associations.[32] Of note, the SomaScan® measurements showed poor agreement with some of the standard laboratory measurements for VTE biomarkers, such as D-dimer and fibrinogen,[33] which explains why these markers were not associated with VTE in our study. Even though low specificity for certain proteins may lead to missed associations, it is not regarded as a major concern for untargeted discovery of novel protein associations in large screening studies (such as ours), as the aptamers in general are considered to measure their target proteins robustly and accurately based on verification from genetic associations[34,35] or mass-spectrometry.[20]

The number one top-ranked protein associated with VTE in our study was FVIII, which has a known procoagulant function and is an established risk factor for VTE.[24-29] This finding serves as a valuable confirmation and positive control for the case-cohort approach and sample storage in HUNT, although it does not verify the accuracy of all aptamers. To our knowledge, our study is the first to show an association between increased endogenous levels of collagen alpha-3(VI):BPTI/Kunitz

inhibitor and risk of future VTE. Of note, the drug aprotinin, a small bovine pancreatic trypsin inhibitor (BPTI), is a broad acting serine protease inhibitor that has been used to prevent bleeding during major surgery.[36] *In vitro* studies have suggested that aprotinin has antifibrinolytic effect by inhibition of plasmin and anticoagulant effect by inhibition of kallikrein.[36] Although the Kunitz-type domain from the alpha-3 chain of human type IV collagen could be expected to have similar function as aprotinin, the net hemostatic balance by endogenous levels of this protein *in vivo* has not been previously addressed.

Interestingly, we found an association between elevated BGAT levels and VTE. BGAT refers to the A- and B-glycosyltransferases, which catalyze the addition N-Acetylgalactosamin (A antigen) and D-Galactose (B antigen) to the H-antigen in individuals with blood type A or B. Individuals with blood type O lack BGAT.[37] Numerous studies have reported on the association between ABO blood type and VTE,[38-40] and the risk further varies with ABO haplotypes.[38] The carbohydrate structures that comprise the ABO blood group system (A, B and H antigens) are expressed on a wide range of different cell types, including platelets and endothelial cells,[41,42] and affect the quantity and function of several proteins (ABO-related proteins) involved in hemostasis, including von Willebrand factor (VWF) and FVIII,[43] and inflammatory pathways.[44,45] Although the main hypothesis is that the increased VTE risk observed in individuals with non-O blood types is mediated through glycosylation by BGAT with subsequent altered function and quantity of proteins implicated in the pathogenesis of VTE,[43] further investigations are needed to disentangle the role of ABO haplotypes, BGAT levels and ABO-related proteins in VTE.

One previous study has reported an association between TNFR2 and VTE.[13] TNFR2 is expressed on hematopoietic cells and endothelial cells, and the prothrombotic effects of TNF-α have been shown to require TNFR2.[46] It has been suggested that circulating soluble TNF receptors may act as both antagonist (competing with cell-surface receptors) and agonist (protecting TNF-α from degradation) of TNF-α activity, and that the biological activity may vary with plasma concentration.[47] Elevated levels of TNF sR-II are found in patients with cancer, HIV and autoimmune diseases,[48-50] which are all conditions associated with increased risk of VTE.[51,52] HE4, also known as WAP four-disulfide core

domain protein 2, is upregulated in many cancer types[53] and is currently used as a diagnostic biomarker of ovarian cancer. In a study of 208 patients with epithelial ovarian cancer, median HE4 levels were somewhat higher (793 vs. 628 pmol/L) in patients who developed VTE (n=31), but the difference was not statistically significant.[54] HE4 has also been associated with incident cardiovascular disease,[55] as well as severity and prognosis of heart failure.[56]

The proteins RGS3, and PXDN were also associated with VTE with p-values below the Bonferroni threshold, but to our knowledge, the role of these proteins in thrombosis and hemostasis has not been investigated. RGS3 is a GTPase-activating protein that inhibits G-protein-mediated signal transduction. In an experimental study, ectopic expression of R4 subfamily members, including RSG3, was associated with marked attenuation of protease-activated receptor-1 (PAR1) signaling (which is the predominant mediator of thrombin signaling in most cell types.[57] PXDN (also known as vascular peroxidase 1) is a heme peroxidase which serves as a key enzyme for cross-linking collagen IV during basement membrane formation [58]. Furthermore, the protein is proposed to be involved in host defense and innate immunity, endothelial dysfunction and oxidative stress.[58]

The KEGG pathway analysis revealed enrichment of proteins in the coagulation system (F8, F9, F10 and vWF) and the complement system (C4BPA, MASP1, VSIG4, MBL2, and CD55). In vitro and animal model studies support a role of the intrinsic coagulation system in venous thrombus formation.[59] Furthermore, growing evidence suggests that the innate immune system is involved in the pathogenesis of thrombosis[60] and that there are several points of crosstalk between the complement and coagulation system.[61] During the recent years, population-based studies have reported an association between several individual components of the complement system and risk of future VTE, including complement components C3 and C5,[62,63] MBL,[64] MASP-2,[65] and the terminal complement complex (TCC).[66] Interestingly, increasing levels of C1 inhibitor, a serine protease that inhibits both the classical- and the lectin pathway of complement, was associated with lower risk of VTE[31] and inhibited ex vivo thrombin generation triggered by the intrinsic coagulation system.[31] Furthermore, C1 inhibitor deficient mice had increased venous thrombus formation compared to wild

type mice.[67] Our findings of protein enrichment particularly in the intrinsic coagulation cascade and the lectin pathway of the complement system further reinforce the hypothesis of an interplay between these two pathways in the pathogenesis of VTE.

Due to our discovery framework, our results on candidate biomarkers cannot be directly implemented into clinical practice. Validation of our findings in an external cohort, preferably with quantitative techniques to measure proteins (e.g. ELISA), will be an important next step to verify our results. In addition, Mendelian Randomization analyses can be applied to assess whether protein candidates are likely to be causally involved in the pathogenesis of VTE or rather markers of disease risk.[68] Finally, the clinical utility of the protein candidates for VTE risk stratification, individually or added to already established risk assessment models for VTE, should be explored in studies using an appropriate prediction modelling framework.

Strengths of the current study include recruitment of participants from a population-based cohort with high-quality blood samples and thorough detection and validation of VTE events, clinically relevant follow-up time for risk assessment (5 years), the large number of proteins screened, and the clear temporal sequence between exposure (protein measurements) and outcome (VTE). As described above, a limitation related to the SomaScan® technology is the potential lack of specificity for some proteins. Even though this is considered a larger concern in hypothesis driven research than in untargeted approaches like ours, it may lead to missed discoveries in large screening studies. Although we applied the strict Bonferroni threshold to correct for multiple testing, chance findings cannot be ruled out.

In conclusion, our proteome-wide screening of more than 7000 proteins in plasma led to novel discovery of five proteins (Collagen alpha-3(VI):BPTI/Kunitz inhibitor, RGS3, BGAT, PXDN and HE4) associated with 5-year risk of future VTE (significant at the Bonferroni threshold). Furthermore, pathway analyses of the top-ranked 200 proteins revealed significant enrichment of nine proteins in the complement and coagulation cascades, further reinforcing the hypothesis of an interplay between these pathways in the pathogenesis of VTE.

Funding

The Thrombosis Research Center (TREC) was supported by an independent grant from Stiftelsen Kristian Gerhard Jebsen (SKGJ-MED-012) during the establishment of this project. The SOMAscan lab work was partly supported by the grant R01HL059367 and R01HL155209 from the National Heart, Lung, and Blood Institute.

Acknowledgements

We acknowledge SomaLogic Operating Co., Inc. as the provider of the proteomic data measured using the modified aptamer-based SomaScan® Assay. SomaScan®, SOMAmer® and SomaSignal™ are trademarks of SomaLogic Operating Co., Inc.

The Trøndelag Health Study (HUNT) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology NTNU), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health.

Disclosure of Interest

The authors report no disclosures of interest

Data availability statement

Access to HUNT data can be obtained by application to the HUNT administration https://www.ntnu.edu/hunt/research.

Author contributions

Conception and design: JBH. Data collection: JBH, KH, CJ, ARF, WT. Statistical analysis: THN. Interpretation of data: JBH, SKB, KH, ARF, CJ, THN, WT, WG, ALO, KDH, VMM. Draft of manuscript: SKB. Critical revision of manuscript: JBH, KH, ARF, CJ, THN, WT, WG, ALO, KDH, VMM. Review and approval of final version: All co-authors.

References

1. Arnesen CAL, Veres K, Horvath-Puho E*, et al.* Estimated lifetime risk of venous thromboembolism in men and women in a Danish nationwide cohort: impact of competing risk of death. *Eur J Epidemiol* 2022;**37**:195-203. doi: 10.1007/s10654-021-00813-w
2. Bell EJ, Lutsey PL, Basu S*, et al.* Lifetime Risk of Venous Thromboembolism in Two Cohort Studies. *Am J Med* 2016;**129**:339 e319-326. doi: 10.1016/j.amjmed.2015.10.014
3. Kahn SR. The post-thrombotic syndrome. *Hematology Am Soc Hematol Educ Program* 2010;**2010**:216-220. doi: 10.1182/asheducation-2010.1.216
4. Klok FA, van der Hulle T, den Exter PL*, et al.* The post-PE syndrome: a new concept for chronic complications of pulmonary embolism. *Blood Rev* 2014;**28**:221-226. doi: 10.1016/j.blre.2014.07.003
5. Jorgensen H, Horvath-Puho E, Laugesen K*, et al.* Risk of a permanent work-related disability pension after incident venous thromboembolism in Denmark: A population-based cohort study. *PLoS Med* 2021;**18**:e1003770. doi: 10.1371/journal.pmed.1003770
6. Arshad N, Bjori E, Hindberg K*, et al.* Recurrence and mortality after first venous thromboembolism in a large population-based cohort. *J Thromb Haemost* 2017;**15**:295-303. doi: 10.1111/jth.13587
7. Prandoni P, Noventa F, Ghirarduzzi A*, et al.* The risk of recurrent venous thromboembolism after discontinuing anticoagulation in patients with acute proximal deep vein thrombosis or pulmonary embolism. A prospective cohort study in 1,626 patients. *Haematologica* 2007;**92**:199-205. doi:
8. Cushman M, Tsai AW, White RH*, et al.* Deep vein thrombosis and pulmonary embolism in two cohorts: the longitudinal investigation of thromboembolism etiology. *Am J Med* 2004;**117**:19-25. doi:
9. Arshad N, Isaksen T, Hansen JB, Braekkan SK. Time trends in incidence rates of venous thromboembolism in a large cohort recruited from the general population. *Eur J Epidemiol* 2017;**32**:299-305. doi: 10.1007/s10654-017-0238-y
10. Jorgensen CT, Tavoly M, Pettersen HH*, et al.* The venous thrombosis registry in Ostfold Hospital (TROLL registry) - design and cohort description. *Res Pract Thromb Haemost* 2022;**6**. doi: 10.1002/rth2.12770
11. Wolberg AS, Rosendaal FR, Weitz JI*, et al.* Venous thrombosis. *Nat Rev Dis Primers* 2015;**1**:15006. doi: 10.1038/nrdp.2015.6
12. Cushman M, Barnes GD, Creager MA*, et al.* Venous Thromboembolism Research Priorities: A Scientific Statement From the American Heart Association and the International Society on Thrombosis and Haemostasis. *Circulation* 2020;**142**:e85-e94. doi: 10.1161/CIR.0000000000000818
13. Yuan S, Titova OE, Zhang K*, et al.* Plasma protein and venous thromboembolism: prospective cohort and mendelian randomisation analyses. *Br J Haematol* 2023;**201**:783-792. doi: 10.1111/bjh.18679
14. Bruzelius M, Iglesias MJ, Hong MG*, et al.* PDGFB, a new candidate plasma biomarker for venous thromboembolism: results from the VEREMA affinity proteomics study. *Blood* 2016;**128**:e59-e66. doi: 10.1182/blood-2016-05-711846
15. Jensen SB, Hindberg K, Solomon T*, 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**:1763-1774. doi: 10.1111/jth.14220
16. Candia J, Daya GN, Tanaka T, Ferrucci L, Walker KA. Assessment of variability in the plasma 7k SomaScan proteomics assay. *Sci Rep* 2022;**12**:17147. doi: 10.1038/s41598-022-22116-0
17. Asvold BO, Langhammer A, Rehn TA*, et al.* Cohort Profile Update: The HUNT Study, Norway. *Int J Epidemiol* 2023;**52**:e80-e91. doi: 10.1093/ije/dyac095
18. Naess M, Kvaloy K, Sorgjerd EP*, et al.* Data Resource Profile: The HUNT Biobank. *Int J Epidemiol* 2024;**53**. doi: 10.1093/ije/dyae073
19. Levey AS, Stevens LA, Schmid CH*, et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;**150**:604-612. doi: 10.7326/0003-4819-150-9-200905050-00006

20. Gold L, Ayers D, Bertino J*, et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 2010;**5**:e15004. doi: 10.1371/journal.pone.0015004
21. Stu Field (2022).SomaDataIO: SomaData Input/Output Package. R package version 5.3.0.9000. www.somalogic.com. In.
22. Jones E. cchs: An R Package for Stratified Case-Cohort Studies. *R Journal* 2018;**10**:484-494. doi:
23. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**:27-30. doi: 10.1093/nar/28.1.27
24. Rietveld IM, Lijfering WM, le Cessie S*, et al.* High levels of coagulation factors and venous thrombosis risk: strongest association for factor VIII and von Willebrand factor. *J Thromb Haemost* 2019;**17**:99-109. doi: 10.1111/jth.14343
25. Koster T, Blann AD, Briet E, Vandenbroucke JP, Rosendaal FR. Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep-vein thrombosis. *Lancet* 1995;**345**:152-155. doi: S0140-6736(95)90166-3 [pii]
26. Kraaijenhagen RA, in't Anker PS, Koopman MM*, et al.* High plasma concentration of factor VIIIc is a major risk factor for venous thromboembolism. *Thromb Haemost* 2000;**83**:5-9. doi: 00010005 [pii]
27. 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**:2424-2429. doi: 10.1182/blood-2006-06-028241
28. Evensen LH, Folsom AR, Pankow JS*, et al.* Hemostatic factors, inflammatory markers, and risk of incident venous thromboembolism: The Multi-Ethnic Study of Atherosclerosis. *J Thromb Haemost* 2021. doi: 10.1111/jth.15315
29. Hansen ES, Edvardsen MS, Aukrust P*, et al.* Combined effect of high factor VIII levels and high mean platelet volume on the risk of future incident venous thromboembolism. *J Thromb Haemost* 2023. doi: 10.1016/j.jtha.2023.06.022
30. 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**:523-535. doi: 10.1111/joim.13502
31. Grover SP, Snir O, Hindberg K*, et al.* High plasma levels of C1-inhibitor are associated with lower risk of future venous thromboembolism. *J Thromb Haemost* 2023;**21**:1849-1860. doi: 10.1016/j.jtha.2023.03.024
32. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BMJ* 2010;**340**:c2289. doi: 10.1136/bmj.c2289
33. Faquih T, Mook-Kanamori DO, Rosendaal FR*, et al.* Agreement of aptamer proteomics with standard methods for measuring venous thrombosis biomarkers. *Res Pract Thromb Haemost* 2021;**5**:e12526. doi: 10.1002/rth2.12526
34. Ferkingstad E, Sulem P, Atlason BA*, et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet* 2021;**53**:1712-1721. doi: 10.1038/s41588-021-00978-w
35. Eldjarn GH, Ferkingstad E, Lund SH*, et al.* Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* 2023;**622**:348-358. doi: 10.1038/s41586-023-06563-x
36. Lisman T, Adelmeijer J, Huskens D, Meijers JCM. Aprotinin Inhibits Thrombin Generation by Inhibition of the Intrinsic Pathway, but is not a Direct Thrombin Inhibitor. *TH Open* 2021;**5**:e363-e375. doi: 10.1055/s-0041-1735154
37. Yamamoto F, Clausen H, White T, Marken J, Hakomori S. Molecular genetic basis of the histo-blood group ABO system. *Nature* 1990;**345**:229-233. doi: 10.1038/345229a0
38. Goumidi L, Thibord F, Wiggins KL*, et al.* Association between ABO haplotypes and the risk of venous thrombosis: impact on disease risk estimation. *Blood* 2021;**137**:2394-2402. doi: 10.1182/blood.2020008997

39. Ohira T, Cushman M, Tsai MY*, et al.* ABO blood group, other risk factors and incidence of venous thromboembolism: the Longitudinal Investigation of Thromboembolism Etiology (LITE). *J Thromb Haemost* 2007;**5**:1455-1461. doi: JTH02579 [pii]

10.1111/j.1538-7836.2007.02579.x
40. Wolpin BM, Kabrhel C, Varraso R*, et al.* Prospective study of ABO blood type and the risk of pulmonary embolism in two large cohort studies. *Thromb Haemost* 2010;**104**:962-971. doi: 10-05-0312 [pii]

10.1160/TH10-05-0312
41. Anstee DJ. The relationship between blood groups and disease. *Blood* 2010;**115**:4635-4643. doi: 10.1182/blood-2010-01-261859
42. Oriol R, Mollicone R, Coullin P, Dalix AM, Candelier JJ. Genetic regulation of the expression of ABH and Lewis antigens in tissues. *APMIS Suppl* 1992;**27**:28-38. doi:
43. Ward SE, O'Sullivan JM, O'Donnell JS. The relationship between ABO blood group, von Willebrand factor, and primary hemostasis. *Blood* 2020;**136**:2864-2874. doi: 10.1182/blood.2020005843
44. Emilsson V, Gudmundsdottir V, Gudjonsson A*, et al.* Coding and regulatory variants are associated with serum protein levels and disease. *Nat Commun* 2022;**13**:481. doi: 10.1038/s41467-022-28081-6
45. Gudjonsson A, Gudmundsdottir V, Axelsson GT*, et al.* A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nat Commun* 2022;**13**:480. doi: 10.1038/s41467-021-27850-z
46. Pircher J, Merkle M, Wornle M*, et al.* Prothrombotic effects of tumor necrosis factor alpha in vivo are amplified by the absence of TNF-alpha receptor subtype 1 and require TNF-alpha receptor subtype 2. *Arthritis Res Ther* 2012;**14**:R225. doi: 10.1186/ar4064
47. Aderka D, Engelmann H, Maor Y, Brakebusch C, Wallach D. Stabilization of the bioactivity of tumor necrosis factor by its soluble receptors. *J Exp Med* 1992;**175**:323-329. doi: 10.1084/jem.175.2.323
48. Cope AP, Aderka D, Doherty M*, et al.* Increased levels of soluble tumor necrosis factor receptors in the sera and synovial fluid of patients with rheumatic diseases. *Arthritis Rheum* 1992;**35**:1160-1169. doi: 10.1002/art.1780351008
49. Hober D, Benyoucef S, Delannoy AS*, et al.* High plasma level of soluble tumor necrosis factor receptor type II (sTNFRII) in asymptomatic HIV-1-infected patients. *Infection* 1996;**24**:213-217. doi: 10.1007/BF01781095
50. Aderka D, Englemann H, Hornik V*, et al.* Increased serum levels of soluble receptors for tumor necrosis factor in cancer patients. *Cancer Res* 1991;**51**:5602-5607. doi:
51. Zoller B, Li X, Sundquist J, Sundquist K. Risk of pulmonary embolism in patients with autoimmune disorders: a nationwide follow-up study from Sweden. *Lancet* 2012;**379**:244-249. doi: 10.1016/S0140-6736(11)61306-8

S0140-6736(11)61306-8 [pii]
52. Blix K, Gran OV, Severinsen MT*, et al.* Impact of time since diagnosis and mortality rate on cancer-associated venous thromboembolism: the Scandinavian Thrombosis and Cancer (STAC) cohort. *J Thromb Haemost* 2018;**16**:1327-1335. doi: 10.1111/jth.14130
53. Hertlein L, Stieber P, Kirschenhofer A*, et al.* Human epididymis protein 4 (HE4) in benign and malignant diseases. *Clin Chem Lab Med* 2012;**50**:2181-2188. doi: 10.1515/cclm-2012-0097
54. Wang Y, Zhou H, Zhong G*, et al.* Development and Validation of a Nomogram to Predict the Probability of Venous Thromboembolism in Patients with Epithelial Ovarian Cancer. *Clin Appl Thromb Hemost* 2022;**28**:10760296221095558. doi: 10.1177/10760296221095558
55. Lind L, Zanetti D, Ingelsson M*, et al.* Large-Scale Plasma Protein Profiling of Incident Myocardial Infarction, Ischemic Stroke, and Heart Failure. *J Am Heart Assoc* 2021;**10**:e023330. doi: 10.1161/JAHA.121.023330

56. de Boer RA, Cao Q, Postmus D*, et al.* The WAP four-disulfide core domain protein HE4: a novel biomarker for heart failure. *JACC Heart Fail* 2013;**1**:164-169. doi: 10.1016/j.jchf.2012.11.005
57. Chen B, Siderovski DP, Neubig RR, Lawson MA, Trejo J. Regulation of protease-activated receptor 1 signaling by the adaptor protein complex 2 and R4 subfamily of regulator of G protein signaling proteins. *J Biol Chem* 2014;**289**:1580-1591. doi: 10.1074/jbc.M113.528273
58. Cheng G, Shi R. Mammalian peroxidasin (PXDN): From physiology to pathology. *Free Radic Biol Med* 2022;**182**:100-107. doi: 10.1016/j.freeradbiomed.2022.02.026
59. Grover SP, Mackman N. Intrinsic Pathway of Coagulation and Thrombosis. *Arterioscler Thromb Vasc Biol* 2019;**39**:331-338. doi: 10.1161/ATVBAHA.118.312130
60. Engelmann B, Massberg S. Thrombosis as an intravascular effector of innate immunity. *Nat Rev Immunol* 2013;**13**:34-45. doi: 10.1038/nri3345
61. Foley JH, Conway EM. Cross Talk Pathways Between Coagulation and Inflammation. *Circ Res* 2016;**118**:1392-1408. doi: 10.1161/CIRCRESAHA.116.306853
62. Norgaard I, Nielsen SF, Nordestgaard BG. Complement C3 and High Risk of Venous Thromboembolism: 80517 Individuals from the Copenhagen General Population Study. *Clin Chem* 2016;**62**:525-534. doi: 10.1373/clinchem.2015.251314
63. Skjeflo EW, Braekkan SK, Ludviksen JK*, et al.* Elevated plasma concentration of complement factor C5 is associated with risk of future venous thromboembolism. *Blood* 2021. doi: 10.1182/blood.2021010822
64. Liang RA, Hoiland, II, Ueland T*, et al.* Plasma levels of mannose-binding lectin and future risk of venous thromboembolism. *J Thromb Haemost* 2019;**17**:1661-1669. doi: 10.1111/jth.14539
65. Damoah CE, Snir O, Hindberg K*, et al.* High Levels of Complement Activating Enzyme MASP-2 Are Associated With the Risk of Future Incident Venous Thromboembolism. *Arterioscler Thromb Vasc Biol* 2022;**42**:1186-1197. doi: 10.1161/ATVBAHA.122.317746
66. Hoiland, II, Liang RA, Braekkan SK*, et al.* Complement activation assessed by the plasma terminal complement complex and future risk of venous thromboembolism. *J Thromb Haemost* 2019;**17**:934-943. doi: 10.1111/jth.14438
67. Grover SP, Kawano T, Wan J*, et al.* C1 inhibitor deficiency enhances contact pathway-mediated activation of coagulation and venous thrombosis. *Blood* 2023;**141**:2390-2401. doi: 10.1182/blood.2022018849
68. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;**23**:R89-98. doi: 10.1093/hmg/ddu328

Table 1. Baseline characteristics of the study population.

	Subcohort (n=1066)	**Incident VTE cases (n=294)**
Baseline characteristics		
Age (years)	64 ± 13	66 ± 15
Sex (% women)	475 (45%)	151 (51%)
Body mass index (kg/m^2)	27.6 ± 4.1	28.4 ± 4.6
History of CVD (%)*	59 (6%)	20 (7%)
Diabetes (%)	68 (6%)	11 (4%)
Smoking (%)**	181 (17%)	52 (18%)
eGFR (ml/min/1.73m^2)***	88 (± 18)	84 (± 20)
VTE characteristics		
Deep vein thrombosis (%)	NA	113 (38%)
Pulmonary embolism (%)	NA	181 (62%)
Cancer-related VTE (%)	NA	56 (19%)
Unprovoked VTE (%)	NA	115 (39%)

Values are means ± standard deviations, or numbers with percentages in brackets
*CVD: cardiovascular disease (history of myocardial infarction, stroke or angina pectoris)
**Daily smokers
***Estimated glomerular filtration rate based on EPI-CKD equation

Table 2. Hazard ratios (HR) of VTE with 95% confidence interval (CI) per one standard deviation increase in log2-transformed protein values for each of the seven proteins that were significantly associated with VTE at Bonferroni level ($p<6.9x10^{-6}$).

Protein name	Entrez gene symbol	UniProt	Model 1 HR (95% CI)	P-value discovery*	Model 2 HR (95% CI)	Model 3 HR (95% CI)
Coagulation factor VIII (FVIII)	*F8*	P00451	1.86 (1.59-2.18)	$1.6x10^{-14}$	1.84 (1.57-2.16)	1.82 (1.55-2.13)
Regulator Of G Protein Signaling 3 (RGS3)	*RGS3*	P49796	1.39 (1.22-1.58)	$3.0x10^{-7}$	1.37 (1.20-1.56)	1.41 (1.24-1.59)
Collagen alpha-3(VI):BPTI/Kunitz inhibitor	*COL6A3*	P12111	1.40 (1.23-1.60)	$7.9x10^{-7}$	1.37 (1.19-1.58)	1.37 (1.17-1.61)
Histo-blood group ABO system transferase (BGAT)	*ABO*	P16442	1.40 (1.22-1.61)	$2.1x10^{-6}$	1.40 (1.22-1.61)	1.39 (1.21-1.60)
Peroxidasin (PXDN)	*PXDN*	Q92626	1.40 (1.22-1.62)	$3.0x10^{-6}$	1.37 (1.18-1.59)	1.36 (1.15-1.63)
Tumor necrosis factor soluble receptor II (TNF sR-II)	*TNFRSF1B*	P20333	1.41 (1.21-1.63)	$6.1x10^{-6}$	1.37 (1.18-1.60)	1.33 (1.13-1.58)
Human epididymis protein 4 (HE4)	*WFDC2*	Q14508	1.44 (1.23-1.69)	$6.8x10^{-6}$	1.51 (1.28-1.77)	1.36 (1.14-1.61)

Model 1: age, sex and sample batch

Model 2: Model 1+ body mass index (BMI)

Model 3: Model 1 + glomerular filtration rate (eGFR)

*p-value corresponding to Model 1 that was used for protein discovery

Table 3: Results of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis.

FDR p-value	Number of genes	Pathway genes	Fold enrichment	Pathway	Genes
$3.3x10^{-6}$	9	84	14.04	Complement and coagulation cascades	*F9 VWF C4BPA F10 MASP1 VSIG4 MBL2 F8 CD55*
$2.3x10^{-2}$	9	294	4.01	MAPK signaling pathway	*TNFRSF1A ANGPT2 EFNA2 MAP2K2 EPHA2 KIT CACNA2D3 IL1RAP EFNA4*
$2.3\ x10^{-2}$	9	294	4.01	Cytokine-cytokine receptor interaction	*TNFRSF1B RELT TNFRSF1A LIFR TNFRSF19 MSTN CCL19 LEP IL1RAP*
$2.3\ x10^{-2}$	10	354	3.70	PI3K-Akt signaling pathway	*ANGPT2 EFNA2 VWF FN1 SPP1 MAP2K2 EPHA2 KIT COL6A3 EFNA4*

*FDR: False discovery rate

Figure 1. Overview of study design for the PRoteomic profiling and Etiological understanding of VENous Thromboembolism (PREVENT) project. We performed a case-cohort including all VTE cases (n=294) occurring during the first five years of follow-up after the survey of the Trøndelag Health Study (HUNT3) and a randomly sampled age- and sex-weighted subcohort (n=1066). In total 7288 proteins were measured in blood samples collected at cohort inclusion. Figure created with BioRender.com.

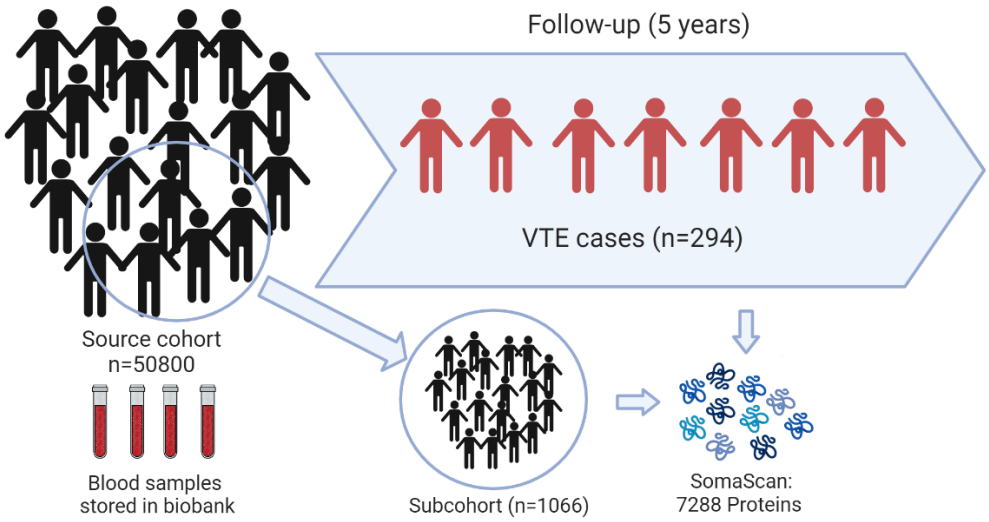

Figure 2. Volcano plot of 7288 plasma proteins and their association with VTE in the case-cohort. The age, sex and sample batch-adjusted hazard ratio (HR) for venous thromboembolism (VTE) is plotted against -log10(p-value). Seven proteins (red dots) were associated with VTE at Bonferroni level ($p<6.9x10-6$) and are indicated by their protein names. BGAT: Histo-blood group ABO system transferase; HE4: Human epididymis protein 4; RGS3: Regulator of G Protein Signaling 3; TNF sR-II: Tumor necrosis factor soluble receptor II.

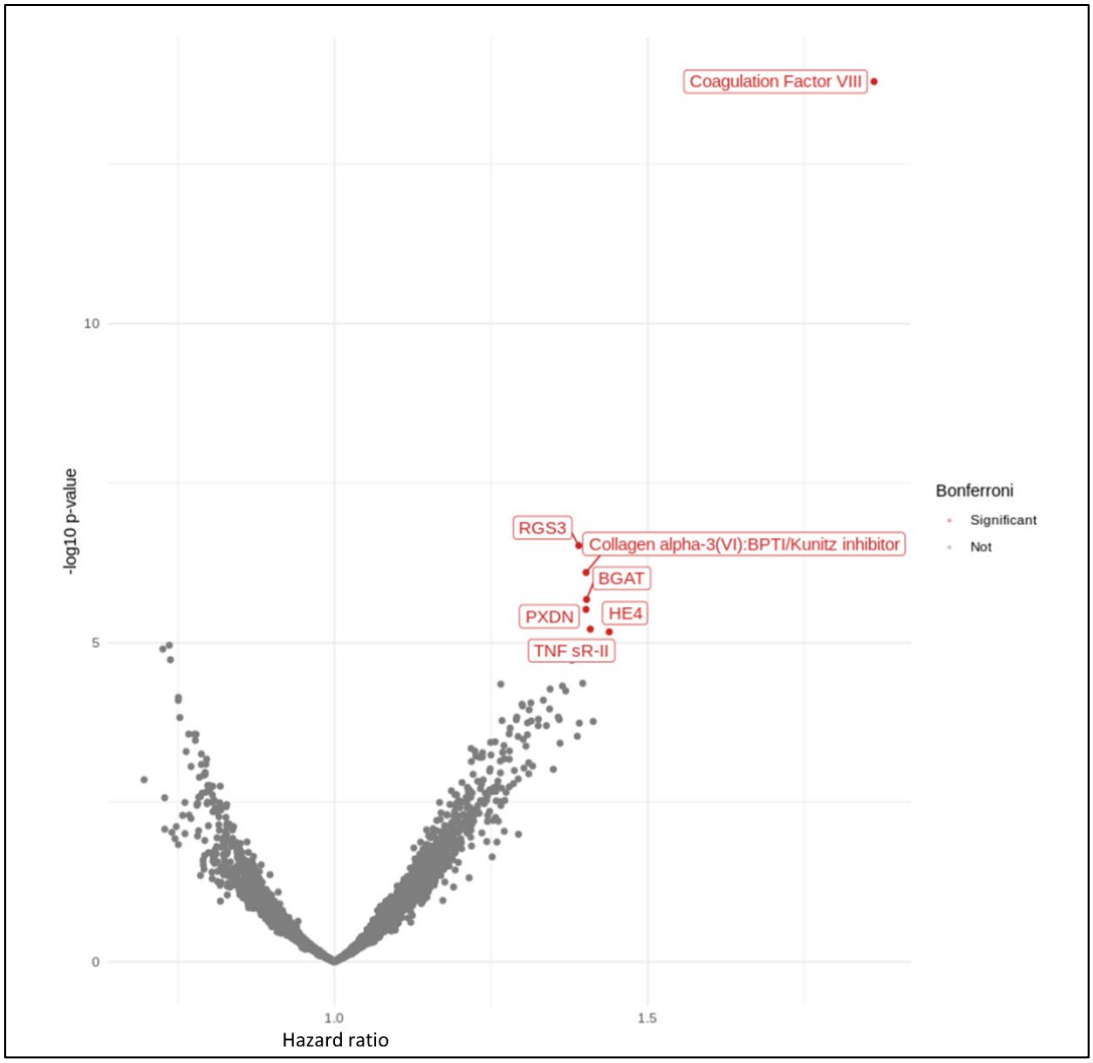

Figure 3. Depiction of the complement and coagulation cascade from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Significant pathway enrichment of nine proteins associated with in venous thromboembolism are indicated in red.