Journal of the American Heart Association

ORIGINAL RESEARCH

Classic Thrombophilias and Thrombotic Risk Among Middle-Aged and Older Adults: A Population-Based Cohort Study

Eric Manderstedt, MSc; Christina Lind-Halldén, PhD; Christer Halldén, PhD; Johan Elf, MD, PhD; Peter J. Svensson, MD, PhD; Björn Dahlbäck, MD, PhD; Gunnar Engström, MD, PhD; Olle Melander, MD, PhD; Aris Baras, MD; Luca A. Lotta, MD; Bengt Zöller, MD, PhD; for the Regeneron Genetics Center*

BACKGROUND: Five classic thrombophilias have been recognized: factor V Leiden (rs6025), the prothrombin G20210A variant (rs1799963), and protein C, protein S, and antithrombin deficiencies. This study aimed to determine the thrombotic risk of classic thrombophilias in a cohort of middle-aged and older adults.

METHODS AND RESULTS: Factor V Leiden, prothrombin G20210A and protein-coding variants in the *PROC* (protein C), *PROS1* (protein S), and *SERPINC1* (antithrombin) anticoagulant genes were determined in 29 387 subjects (born 1923–1950, 60% women) who participated in the Malmö Diet and Cancer study (1991–1996). The Human Gene Mutation Database was used to define 68 disease-causing mutations. Patients were followed up from baseline until the first event of venous thromboembolism (VTE), death, or Dec 31, 2018. Carriership (n=908, 3.1%) for disease-causing mutations in the *PROC*, *PROS1*, and *SERPINC1* genes was associated with incident VTE: Hazard ratio (HR) was 1.6 (95% CI, 1.3–1.9). Variants not in Human Gene Mutation Database were not linked to VTE (HR, 1.1; 95% CI, 0.8–1.5). Heterozygosity for rs6025 and rs1799963 was associated with incident VTE: HR, 1.8 (95% CI, 1.6–2.0) and HR, 1.6 (95% CI, 1.3–2.0), respectively. The HR for carrying 1 classical thrombophilia variant was 1.7 (95% CI, 1.6–1.9). HR was 3.9 (95% CI, 3.1–5.0) for carriers of ≥2 thrombophilia variants.

CONCLUSIONS: The 5 classic thrombophilias are associated with a dose-graded risk of VTE in middle-aged and older adults. Disease-causing variants in the *PROC*, *PROS1*, and *SERPINC1* genes were more common than the rs1799963 variant but the conferred genetic risk was comparable with the rs6025 and rs1799963 variants.

Key Words: epidemiology ■ genetics ■ natural anticoagulants ■ thrombophilia ■ venous thromboembolism

Although new risk variants for venous thromboembolism (VTE) have been discovered, only 5 classic genetic risk factors are widely recognized for thrombophilia.[1–7] The classic thrombotic risk factors are factor V Leiden (rs6025), the prothrombin G20210A (rs1799963) variant, and inherited deficiencies of the natural anticoagulants antithrombin (coded for by *SERPINC1*), protein C (*PROC*), and protein S (*PROS1*).[3–7] The risk of VTE attributed to the common rs6025 and rs1799963 variants has been confirmed in large cohort studies.[8–10] The carrier frequency among White people for the rs6025 variant is around 5% to 10% and around 2% for the rs1799963 variant.[1–7] The relative VTE risk for the rs6025 variant was 2.7 (95% CI, 1.3 to 5.6) and 1.7 (95% CI, 0.9–3.1) for the rs1799963 variant in 2 cohort studies by Ridker et al.[8,9] In a large Danish study the relative risk for the rs6025 variant was 2.2 (95% CI, 2.0–2.5) and 1.5 (95% CI, 1.2–1.9) for the rs1799963 variant.[10] These risk estimates for the common rs6025 and rs1799963 variants are lower than

Correspondence to: Bengt Zöller, Center for Primary Health Care Research, CRC, Building 28, Floor 11, Jan Waldenströms gata 35, Skåne University Hospital, S-205 02 Malmö, Sweden. E-mail: bengt.zoller@med.lu.se

*A complete list of the Regeneron Genetics Center is provided in the Appendix at end of the article.

Supplemental Material for this article is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.121.023018

For Sources of Funding and Disclosures, see pages 10 and 11.

© 2022 The Authors and Regeneron Genetics Center. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- To our knowledge, this is the first population-based genetic study of all 5 classical thrombophilias, factor V Leiden (rs6025), the prothrombin G20210A variant (rs1799963), and *PROC* (protein C), *PROS1* (protein S), and *SERPINC1* (antithrombin) deficiencies, in middle-aged and older adults.
- The 5 classical thrombophilias are associated with a dose-graded risk of incident VTE in middle-aged and older adults.
- The conferred genetic risk of disease-causing variants in the *PROC*, *PROS1*, and *SERPINC1* genes is comparable with the rs6025 and rs1799963 variants.

What Are the Clinical Implications?

- The 5 classical thrombophilias are common risk factors for incident venous thromboembolism in a general population of middle-aged and older individuals.

Nonstandard Abbreviations and Acronyms

ACMG	American College of Medical Genetics and Genomics
HGMD	Human Gene Mutation Database
MDC	Malmö Diet and Cancer cohort
PCA	principal component analysis

initially estimated among selected patients in case-control studies and family studies.[3–7]

Inherited deficiencies of the natural coagulation inhibitors antithrombin, protein C, and protein S are all deemed to be strong but rare (<1%) risk factors for VTE.[1–7] Several hundreds of causative mutations have been described that are responsible for these deficiencies. Due to the rareness of these deficiencies, the estimated VTE risk of inherited deficiencies of the natural anticoagulants antithrombin, protein C, and protein S has been determined mainly in studies of selected thrombosis-prone families. The increased risk of VTE in patients with inherited deficiencies of these 3 anticoagulant proteins was estimated to be 10-fold in these studies.[3–7] However, the MEGA case-control study could not confirm that protein S deficiency was associated with VTE,[11] and deficiencies of antithrombin, protein C, and protein S have also been detected in healthy individuals without personal and/or family history of VTE, thereby questioning their importance.[12–16]

Variants causing such deficiencies (ie, low antigen and/or functional levels) are described in the Human Gene Mutation Database (HGMD).[17] Deficiencies of the 3 natural anticoagulants are either quantitative (type I or III) or qualitative (type II).[3–7] The importance of inherited deficiencies of antithrombin, protein C, and protein S remains to be determined in large population-based cohort studies of the general population. Moreover, classic thrombophilias are believed to be of importance mainly in young people aged <50 years.[18] The importance of classic thrombophilias among the general population of middle-aged and older adults remains to be determined, although VTE is most common among middle-aged and older adults.

In the present study the prevalence and the VTE-associated risk of disease-causing variants in the *SERPINC1*, *PROS1* and *PROC* genes were analyzed in addition to factor V Leiden and the prothrombin G20210A variants in the Malmö Diet and Cancer cohort (MDC).[19] The MDC is a follow-up study of middle-aged and older adults.

METHODS

Study Population

Because of ethical and legal restrictions related to the Swedish Biobanks in Medical Care Act (2002:297) and the Personal Data Act (1998:204), data are available upon request from the data access group of MDC Study by contacting Anders Dahlin (anders.dahlin@med.lu.se). The MDC is a population-based prospective cohort study from the city of Malmö in the south of Sweden. Sample characteristics, data collection, and clinical definitions for the MDC have been described previously.[19] A total of 30 446 individuals, men (n=12 120, born 1923–1945) and women (n=18 326, born 1923–1950), out of an eligible population of ≈74 000 individuals attended a baseline examination between March 1991 and September 1996. DNA testing was available for 29 387 subjects sampled at baseline. DNA was extracted from peripheral blood cells and assigned to batches without regard to disease status or personal identity. The ethics committee at Lund University Lund, Sweden approved the study (LU 51/90) and all participants provided informed written consent.

Clinical Examination

Participants underwent a medical history, physical examination, and laboratory assessment at baseline.[19] Blood pressure was measured using a mercury-column sphygmomanometer after 10 minutes of rest in the supine position. Cigarette smoking was determined by a self-administered questionnaire. Subjects were categorized as current smokers (ie, those who

smoked regularly or occasionally) or non-smokers (ie, former smokers and never smokers). High alcohol consumption was defined as >40 g alcohol per day for men and >30 g per day for women. Weight and height were measured to the nearest 0.1 kg and 0.5 cm, respectively, with subjects wearing light clothing and no shoes. Current body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared (kg/m^2). One primary outcome, VTE, and 2 secondary outcomes, deep venous thrombosis (DVT) and pulmonary embolism (PE) were examined. VTE, DVT, and PE were defined based on the *International Classification of Diseases, Seventh to Tenth Revisions* (*ICD-7* to *ICD-10*) (Table S1). VTE events (main and secondary diagnoses) were identified through linkage of the 10-digit personal identification number of each Swedish citizen with the Swedish National Patient Register. The Swedish National Patient Register had a 100% coverage for inpatients in Malmö from 1970 until end of follow-up (Dec 31, 2018) and for outpatients from 2001. The diagnosis of VTE in the Swedish National Patient Register has been shown to have an accuracy of 95%,[20] whereas the overall validity of the Swedish National Patient Register is 87%.[21] Events were determined both before baseline (prevalent events) and after baseline during follow-up (incident events). Patients with VTE in Malmö University Hospital are diagnosed with objective methods such as phlebography, compression ultrasound, ventilation/perfusion lung scan (V/Q lung scan), computed tomography (CT), or magnetic resonance tomography.[22]

Exome Sequencing and Genetic Analysis

Whole-exome sequencing was performed by Regeneron Genetics Center (Tarrytown, NY).[23] Targeted exonic regions were captured using a slightly modified version of the xGen probe library (Integrated DNA Technologies). Captured DNA was PCR amplified and sequenced with v4 chemistry using 75 bp paired-end reads on the Illumina HiSeq 2500. Whole-exome sequencing was performed such that >85% of targeted bases are covered at a read depth of >20×. We used previously established bioinformatics procedures to process and analyze exome sequence data. ANNOVAR (2019-10-24) was used to aggregate variant annotation, allele frequencies, and in silico predictions of deleteriousness.[24] The rs6025 and the rs1799963 variants were determined from whole-exome sequencing data. The HGMD database (http://www.hgmd.cf.ac.uk/ac/index.php)[17] was used to define high-risk variants, ie, disease causing, in the *SERPINC1*, *PROC*, and *PROS1* genes. In June 2020, the HGMD database contained >289 000 different gene lesions identified in >11 100 genes manually curated from 72 987 articles published in >3100 peer-reviewed journals.[17] There is generally a good agreement between the variant classification in HGMD and the American College of Medical Genetics and Genomics (ACMG) guidelines and the ClinVar database.[25] When variants in HGMD were classified based on the ACMG guidelines, misclassification was observed in only 3.47% (2289/65 896) of variants. The overall concordance between HGMD and the ClinVar database was 97.62% (52 499/53 780) of variants studied.[25] However, we also read the cited articles in the HGMD database and searched PubMed and could confirm that the HGMD variants were correctly curated. All non-synonymous variants present in the HGMD or affecting the same codon as a variant already occurring in the HGMD were defined as high-risk variants. This is in accordance with the PM5 criteria of ACMG (moderate evidence of pathogenicity): Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before.[26] Moreover, we analyzed the twelve-missense mutations occurring at an amino acid that HGMD has reported to be pathogenic with meta-analytic support vector machine.[27] Nine out of 12 (75%) missense variant were pathogenic in meta-analytic support vector machine.[27] We have therefore chosen to designate all as high-risk variants as these results supported the PM5 criteria of ACMG.[26]

Principal component analysis (PCA) was performed as described.[28] The reference genomes were obtained from the 1000 Genomes Project server.[28,29] The PCA was performed with independent (R^2 measure of linkage disequilibrium <0.2) common (minor allele frequency ≥5%) autosomal biallelic variants that were detected in both the reference genomes and the MDC exomes. To avoid extended linkage disequilibrium and high variability regions, such as the major histocompatibility complex, these regions were omitted from the PCA. The principal components were first obtained from the reference genomes and then projected individuals from the MDC onto the principal-component space via PLINK2.[28,30]

Statistical Analysis

R version 4.0.0 was used for all statistical analyses. The contribution to VTE risk was assessed for sex, age, BMI, smoking status, systolic blood pressure, high alcohol consumption, rs6025, rs1799963 and high-risk variants in *PROC*, *PROS1*, and *SERPINC1*. Logistic regression was conducted to calculate crude odds ratios (ORs) for 4 dependent variables (outcomes): all VTE (prevalent and/or incident VTE), prevalent VTE (VTE between 1970 and baseline), incident VTE (VTE after baseline with no prevalent VTE), and recurrent VTE (≥2 incident VTE during follow-up without prevalent VTE) compared with those without a VTE. Prevalent VTE cases were excluded for determination of OR for

incident and recurrent VTE. The Martingale residuals were used for checking linearity of the predictor with logit-transformed outcome for continuous variables (age, BMI and systolic blood pressure). No important deviations from linearity were observed. Prevalent VTE cases were excluded for determination of incident VTE. Incidence rates were calculated as a measure for absolute risk and incidence rate ratios were also presented as a robust measure of relative risk. Cox proportional hazards regression was used to examine the crude and adjusted association between genotype and incident VTE, DVT, and PE. Time was given as the number of years from baseline examination until death, emigration, incident VTE, or end of follow-up, whichever occurred first. Age and sex were included as covariates in the sex- and age-adjusted model. The multivariable model also included BMI, smoking status, blood pressure (systolic), and high alcohol consumption (>30 g/d women, >40 g/d men) as covariates. We also included the top 2 eigenvectors from the PCA analysis as covariates in the Cox proportional hazard regression models to control for population stratification. The fit of the proportional hazards model was checked visually by plotting the incidence rates over time and by calculating Schoenfeld (partial) residuals. Schoenfeld residuals were used as a dependent variable and time as an independent variable, to assess the proportional hazards assumption. No violation against the proportional hazards assumption was observed. Possible interactions between gene variants and age, sex, and risk factors (BMI, smoking status, systolic blood pressure, high alcohol consumption) on VTE was explored by introducing interaction terms in the multivariable models. No interactions were observed. The subjects were categorized according to genotype and Kaplan–Meier plots were calculated for VTE. For curve comparisons, the log-rank test was used.

RESULTS

Cohort Characteristics

A total of 29 387 individuals were available for analysis. The baseline characteristics of the study population are presented in Table 1. A total of 3177 (10.8%) individuals (1869 women, 1308 men) were affected by VTE (prevalent and/or incident VTE). Of these individuals 593 (2.0%) (315 women, 278 men) were affected by VTE between 1970 and baseline, and 2584 (8.8%) individuals were only affected after baseline, ie, incident VTE during follow-up without previous VTE until December 31, 2018. In total, 1491 individuals were affected by recurrent incident VTE during follow-up.

Genetic Findings

The prevalence of heterozygotes for rs6025 was 11.2% (n=3282), whereas it was 1.8% (n=542) for rs1799963. The prevalence of homozygotes for rs6025 was 0.4% (n=119), in comparison with 0.007% (n=2) for rs1799963. Excluding synonymous variants, a total of 171 different coding variants were identified in the *PROC* (55 variants), *PROS1* (69), and *SERPINC1* (47) genes (Tables S2 through S5). In total,

Table 1. Cohort Characteristics for Complete Cohort, No VTE (ie, Individuals With No VTE Between 1970–2018), all VTE (ie, VTE Between 1970–2018), Prevalent VTE (ie, VTE between 1970 and Baseline), Incident VTE During Follow-Up Without Prevalent VTE, and Recurrent VTE During Follow-Up (ie, at Least Two VTE During Follow-Up Without Prevalent VTE)

Independent variables (predictors)	Complete cohort n=29 387	No VTE n=26 210	All VTE n=3177	Prevalent VTE n=593	Incident VTE n=2584	Recurrent VTE n=1491
Female sex (%)	17 687 (60.2%)	15 818 (60.4%)	1869 (58.8%)	315 (53.1%)	1554 (60.1%)	890 (59.7%)
Age at baseline, y (mean±SD)	58.0±7.6	57.8±7.6	59.7±7.4	60.7±7.7	59.4±7.3	59.1±7.2
Body mass index (mean±SD)	25.8±4.0	25.7±4.0	26.8±4.4	27.7±4.7	26.6±4.3	26.8±4.3
Current smoking (%)	7797 (26.5%)	6975 (26.6%)	822 (25.9%)	146 (24.6%)	676 (26.2%)	380 (25.5%)
Systolic blood pressure (mean±SD)	141±20	141±20	142.5±20.1	143.9±20.6	142.2±20.0	141.7±20.2
High alcohol consumption (%)	1177 (4.0%)	1057 (4%)	120 (3.8%)	23 (3.9%)	97 (3.8%)	51 (3.4%)
rs6025 heterozygotes (%)	3282 (11.2%)	2679 (10.2%)	603 (19.0%)	153 (25.8%)	450 (17.4%)	292 (19.6%)
rs6025 homozygotes (%)	119 (0.4%)	75 (0.3%)	44 (1.4%)	16 (2.7%)	28 (1.1%)	21 (1.4%)
rs1799963 heterozygotes (%)	542 (1.8%)	452 (1.7%)	90 (2.8%)	18 (3.0%)	72 (2.8%)	53 (3.6%)
rs1799963 homozygotes (%)	2 (0%)	1 (0%)	1 (0%)	0 (0%)	1 (0%)	1 (0%)
≥1 high-risk variant* (%)	908 (3.1%)	755 (2.9%)	153 (4.8%)	34 (5.7%)	119 (4.6%)	74 (5.0%)
≥1 low-risk variant† (%)	483 (1.6%)	428 (1.6%)	55 (1.7%)	10 (1.7%)	45 (1.7%)	28 (1.9%)

VTE indicates venous thromboembolism.

*High-risk variant=non-synonymous Human Gene Mutation Database variants in the 3 anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see Methods section).

†Low-risk variant=Low-risk non-synonymous non-Human Gene Mutation Database variants in the 3 anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see Methods section).

68 variants were identified as high risk or disease causing (see Methods for definition), whereas 113 (66%) variants were present in the gnomAD database and all were rare (minor allele frequency <0.01 among the non-Finnish European population in gnomAD). Carriership for any of the 68 different high-risk variants in any of the *SERPINC1* (n=14), *PROC* (n=29) or *PROS1* (n=25) genes was 3.1% (n=908). In a few cases (n=12) the obtained HGMD accession number referred to another amino acid shift in the same codon. Among variants known to be associated with deficiency (decreased antigen and/or functional activity), some were associated with type II deficiency (functional deficiency with normal antigen levels): 9 out of 24 (37.5%) *PROC* variants, 6 out of 23 (26%) *PROS1* variants, and 7 out of 9 (78%) *SERPINC1* variants. Three individuals were heterozygotes for Factor V Cambridge (rs118203906). One was affected by incident VTE and 2 were unaffected. No other rare variants associated with VTE in the *F5* or *F2* genes were observed.

The MDC population has only 12% admixture from foreign-born individuals. Among foreign-born individuals only 1% were non-European. To evaluate the population structure and admixture of the MDC population on the basis of the genetic data, we constructed principal components from 3 ancestral super populations (European, East Asian, and African ancestries) from the 1000 Genomes Project and projected MDC study subjects onto the principal-component space. When all results from individuals from the MDC study were plotted, most study participants were found to cluster together (Figure S1). These results confirm that the MDC population is mainly comprised of individuals with complete or partial European White ancestry.

Risk of VTE

The crude ORs for all VTE, prevalent VTE, incident VTE, and recurrent incident VTE were increased for rs6025, rs1799963, and high-risk variants but not low-risk variants (Table 2 and Table S6). Higher ORs for VTE for the rs6025, rs1799963, and high-risk variants were found for those with prevalent VTE compared with incident VTE, although 95% CI overlapped. The genotype ORs were also higher for recurrent incident VTE than for those with only 1 incident VTE event, although 95% CIs overlapped.

During a median follow-up of 23.1 years (interquartile range, 16.8–24.9 years), a total of 2584 incident VTE events occurred (1030 men, 1554 women) among individuals without prevalent VTE (Table 3). The sum of the follow-up time was 587 992.7 years, corresponding to a VTE incidence rate of 4.4 (95% CI, 4.2–4.6) per 1000 person-years. The thrombosis-free survival curves using Kaplan–Meier analysis are presented in Figure for rs6025, rs1799963, and high-risk variants in *PROC*, *PROS1,* and *SERPINC1*. Although there were low numbers of individuals carrying 2 alleles, a clear dose-response is observed in relationship to numbers of affected alleles, see Figure – Panel D where all 5 classic thrombophilic gene variants are included.

Table 2. Crude Odd Ratios for All VTE (ie, VTE Between 1970 and Baseline and/or During Follow-Up), Prevalent VTE (ie, VTE Between 1970 and Baseline), Incident VTE During Follow-Up (Without Prevalent VTE), and Recurrent VTE During Follow-Up (Without Prevalent VTE)

Independent variables (predictors)	All VTE n=3177 OR (95% CI)	*P* value	Prevalent VTE n=593 OR (95% CI)	*P* value	Incident VTE n=2584 OR (95% CI)	*P* value	Recurrent VTE n=1491 OR (95% CI)	*P* value
Female sex	0.94 (0.87–1.04)	0.10	0.74 (0.63–0.88)	4.0e-4	0.99 (0.91–1.08)	0.81	0.97 (0.87–1.08)	0.61
Age at baseline	1.03 (1.03–1.04)	2.1e-38	1.05 (1.04–1.06)	4.5e-20	1.03 (1.02–1.03)	1.1e-24	1.02 (1.01–1.03)	4.2e-10
Body mass index	1.07 (1.06–1.07)	3.2e-49	1.10 (1.08–1.12)	1.0e-29	1.06 (1.05–1.07)	4.0e-29	1.06 (1.05–1.08)	6.0e-25
Current smoking	0.96 (0.88–1.05)	0.37	0.90 (0.74–1.08)	0.28	0.98 (0.89–1.07)	0.62	0.94 (0.84–1.06)	0.33
Systolic blood pressure	1.00 (1.00–1.01)	3.8e-05	1.01 (1.00–1.01)	8.7e-04	1.00 (1.00–1.01)	2.5e-03	1.00 (1.00–1.00)	0.17
High alcohol consumption	0.93 (0.77–1.13)	0.49	0.96 (0.61–1.43)	0.85	0.93 (0.75–1.14)	0.49	0.84 (0.63–1.10)	0.24
rs6025 heterozygotes	2.06 (1.87–2.27)	6.6e-48	3.05 (2.52–3.68)	3.1e-31	1.85 (1.66–2.06)	2.0e-28	2.14 (1.87–2.44)	9.8e-29
rs6025 homozygotes	4.89 (3.34–7.08)	8.7e-17	9.66 (5.41–16.23)	3.9e-16	3.82 (2.43–5.83)	1.7e-09	4.98 (2.99–7.95)	1.0e-10
rs1799963 heterozygotes	1.66 (1.31–2.08)	1.4e-05	1.78 (1.07–2.79)	0.018	1.63 (1.26–2.09)	1.3e-04	2.10 (1.56–2.78)	5.0e-7
≥1 high-risk variant*	1.71 (1.43–2.03)	1.4e-09	2.00 (1.37–2.82)	1.5e-04	1.63 (1.34–1.98)	7.5e-07	1.76 (1.37–2.22)	4.2e-6
≥1 low-risk variant† (%)	1.05 (0.79–1.38)	0.73	1.02 (0.51–1.81)	0.94	1.06 (0.77–1.42)	0.72	1.14 (0.76–1.64)	0.50

For the dependent variables all outcomes were compared with no venous thromboembolism (1970–2018).

Prevalent cases (n=593) were excluded when calculating odds ratio for incident venous thromboembolism and recurrent venous thromboembolism. The independent variables age, body mass index, and systolic blood pressure were included as continuous variables. OR indicates odds ratio; and VTE, venous thromboembolism.

*High-risk variant=High-risk non-synonymous Human Gene Mutation Database variants in the 3 anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see method section).

†Low-risk variant=Low-risk non-synonymous non-Human Gene Mutation Database variants in the 3 anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see Methods section).

Table 3. Hazard Ratios (HRs) For Incident Venous Thromboembolism Adjusted for Either Age, Sex, and Ancestry*, or Multivariable HRs Adjusted for Age, Sex, Body Mass Index, Smoking, High Alcohol Consumption, and Ancestry*

Independent variables (predictors)	Participants at risk	VTE events	Mean age at first VTE	Crude IR	Crude IRR	Age- and sex-adjusted HR		Multivariable HR	
	n	n	Years (SD)	IR (95% CI)	IRR (95% CI)	HR (95% CI)	*P* value	HR (95% CI)	*P* value
Complete cohort	28 794	2584	73.7 (8.6)	4.4 (4.2–4.6)	1	…	…	…	…
Model with factor V Leiden (rs6025)									
Reference no rs6025	25 562	2106	74.1 (8.6)	4.0 (3.9–4.2)	1	1		1	
rs6025 heterozygotes	3129	450	72.3 (8.5)	7.2 (6.6–7.9)	1.8 (1.6–2.0)	1.8 (1.6–2.0)	3.3e-30	1.8 (1.6–2.0)	9.1e-31
rs6025 homozygotes	103	28	69.5 (9.5)	14.5 (9.6–20.9)	3.6 (2.5–5.2)	3.8 (2.6–5.6)	1.6e-12	4.0 (2.7–5.8)	3.9e-13
Model with prothrombin variant (rs1799963)									
Reference no rs1799963	28 268	2511	73.8 (8.6)	4.3 (4.2–4.5)	1	1		1	
rs1799963 heterozygotes	524	72	71.5 (8.1)	6.9 (5.4–8.7)	1.6 (1.3–2.0)	1.6 (1.3–2.1)	4.4e-05	1.6 (1.3–2.0)	6.7e-05
Model with non-synonymous variants in *SERPINC1*, *PROC*, and *PROS1* genes									
Reference no non-synonymous variants	27 447	2420	73.7 (8.6)	4.3 (4.1–4.5)	1	1		1	
1 low-risk variant[*]	469	45	72.4 (8.6)	4.7 (3.5–6.4)	1.1 (0.8–1.5)	1.1 (0.8–1.5)	0.54	1.1 (0.8–1.5)	0.56
1 high-risk variant[†]	868	117	74.0 (8.2)	6.7 (5.5–8.0)	1.6 (1.3–1.85)	1.6 (1.3–1.9)	2.9e-06	1.6 (1.3–1.9)	2.7e-07
Model with all 5 classical thrombophilia variants: rs6025, rs1799963, and high-risk variants in *SERPINC1*, *PROC*, and *PROS1* genes									
No thrombophilia[‡]	24 312	1954	74.2 (8.6)	3.9 (3.7–4.1)	1	1		1	
1 thrombophilia[§] variant	4225	561	72.4 (8.4)	6.6 (6.1–7.2)	1.7 (1.5–1.9)	1.7 (1.6–1.9)	1.1e-28	1.7 (1.6–1.9)	4.3e-29
≥2 thrombophilia[§] variants	257	69	71.1 (9.0)	14.3 (11.1–18.1)	3.7 (2.9–4.6)	3.9 (3.0–4.9)	2.2e-28	3.9 (3.1–5.0)	1.1e-28

Number of participants at risk, number of venous thromboembolism cases during follow-up, mean age at first venous thromboembolism event (SD), incidence rates, and incidence rate ratios are also presented. Prevalent cases of venous thromboembolism were excluded from the study population. HR indicates hazard ratio; IR, incidence rates; IRR, incidence rate ratios; and VTE, venous thromboembolism.

*Ancestry was controlled for by including the top 2 eigenvectors from the principal component analysis as covariates in Cox proportional hazard regression models.

[†]Low-risk variant=Low-risk non-synonymous non-Human Gene Mutation Database variants in the 3 anticoagulant genes *SERPINC1, PROC,* and *PROS1* (see Methods section).

[‡]High-risk variant=High-risk non-synonymous Human Gene Mutation Database variants in the 3 anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see Methods section).

[§]No thrombophilia=No rs6025 allele, no rs1799963 allele, and no high-risk variant[‡].

[‖]Thrombophilia=presence of rs6025 allele, rs1799963 allele, or high-risk variants[‡] in the 3 anticoagulant genes *SERPINC1, PROC,* and *PROS1* (see Methods section).

Participants at risk, VTE events, incidence rates, incidence rate ratios, and hazard ratios are presented in Table 3. The multivariable Cox regression models were adjusted for age, sex, BMI, smoking status, systolic blood pressure, high alcohol consumption, and the top 2 eigenvectors from the PCA analysis. The multivariable HRs were 1.8 (95% CI, 1.6–2.0) for heterozygous rs6025 carriers and 1.6 (95% CI, 1.3–2.0) for heterozygous rs1799963 carriers. The multivariable HR was 1.6 (95% CI, 1.3–1.9) for heterozygous high-risk variant carriers. The HR for LOF variants was 1.6 (95% CI, 1.3–1.9) and the HR for HGMD risk variants excluding LOF was 1.6 (95% CI, 1.3–1.9). Carriers of non-synonymous variants not in HGMD (low-risk variants) had no increased risk of VTE (adjusted HR, 1.1, 95% CI, 0.8–1.5). The HRs were higher for those with more than one thrombophilia allele (Table 3). The multivariable HR was 4.0 (95%CI 2.7–5.8) for homozygous rs6025 carriers. The adjusted HR was 11.3 (95% CI, 1.6–80.8) for homozygous rs1799963 carriers and 5.0 (95% CI, 1.2–20.1) for carriers of >1 high-risk variant allele in the *SERPINC1*, *PROC*, and *PROS1* genes. The 95% CI were wide, and these data are therefore not presented in a table. One of 2 homozygous rs1799963 carriers and 2 of 6 carriers of >1 high-risk variant allele had VTE. The HR for carrying 1 classical thrombophilia variant was 1.7 (95% CI, 1.6–1.9). HR was 3.9 (95% CI, 3.1–5.0) for carriers of ≥2 thrombophilia variants. Thus, the 95% CI did not overlap for those with 1 thrombophilic variant compared with those with ≥2 thrombophilic variants. Adjustment variables had no important effect on genotype risks indicating no confounding from these variables.

The multivariable HRs for VTE were significantly increased for high-risk variants in each of the 3 anticoagulant genes: *SERPINC1* (OR, 1.5; 95% CI, 1.1–1.9), *PROC* (OR, 2.4; 95% CI, 1.5–3.7), and *PROS1* genes (OR, 1.5; 95% CI, 1.1–2.0) (Table S7). Protein S Heerlen and antithrombin Dublin were both

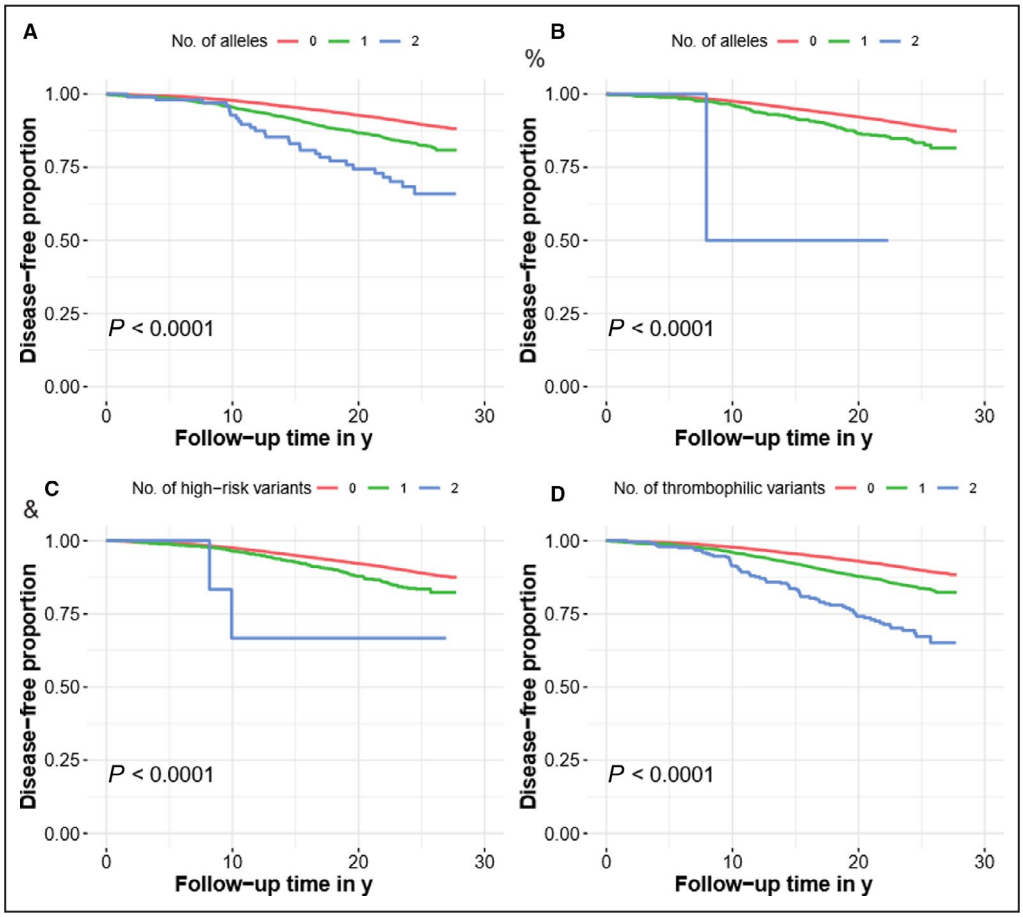

Figure. Kaplan–Meier survival curves showing the proportion of the population remaining free from an venous thromboembolism event stratified on number of factor V Leiden alleles (A), number of rs1799963 alleles (B), number of high-risk variants in *SERPINC1*, *PROC,* or *PROS1* (C) and number of any thrombophilic variants (D).

The log-rank test for curve comparisons, was highly significant for all A through D panels (*P*<0.0001). VTE indicates venous thromboembolism.

significantly associated with incident and/or prevalent VTE (Table S8): protein S Heerlen (OR, 1.93; 95% CI, 1.22–2.94) and antithrombin Dublin (OR, 1.66; 95% CI, 1.21–2.25). Antithrombin Basel was not significantly associated with incident and/or prevalent VTE (OR 1.52; 95% CI, 0.70–2.94). In a sensitivity analysis, high-risk variants of *PROC*, *PROS1*, and *SERPINC1* genes were associated with incident VTE even if these 3 variants were excluded from the analysis (OR, 1.68; 95% CI, 1.30–2.15).

Risk of DVT and PE

Both for heterozygous and homozygous carriers of factor V Leiden (rs6025), the HR for DVT of the legs (HR, 2.3; 95% CI, 2.0–2.7, and HR, 5.4; 95% CI, 3.4–8.6) was higher than for PE (HR, 1.5; 95% CI, 1.2–1.7 and HR, 1.7; 95% CI, 0.8–3.9) (Tables S9 and S10). Otherwise, there were no clear difference for heterozygotes for the rs1799963 variant and the high-risk *PROC*, *PROS1*, and *SERPINC1* variants on DVT of the legs (HR, 1.5; 95% CI, 1.0–2.1 and HR, 1.7; 95% CI, 1.3–2.2) and PE risk (HR, 1.7; 95% CI, 1.2–2.3, and HR, 1.4; 95% CI, 1.0–1.9).

Sensitivity Analysis

The ACMG criteria were determined with Varsome (https://varsome.com/). However, the HR for carrying 1 ACMG positive variant in *SERPINC1*, *PROC*, or *PROS1* was 1.4 (95% CI, 1.1–1.9). HR was 1.4 (95% CI, 0.2–10.0) for carriers of ≥2 thrombophilia variants (Table S11). A sensitivity analysis was also performed with exclusion of malignancy that occurred before cancer. No major differences were observed although the HRs tended to be slightly higher when malignancy was excluded (Table S12). For instance, the multivariable HRs were 1.9 (95% CI, 1.7–2.2) for heterozygous rs6025 carriers and 1.7 (95% CI, 1.3–2.3) for heterozygous rs1799963 carriers (Table S12). In a sensitivity analysis all related people were excluded (ie, up to second-degree cousins). No major difference in effect size was observed (Table S13). The HR for carrying 1 classical thrombophilia variant was 1.7 (95% CI, 1.5–1.9). HR was 4.1 (95% CI, 3.1–5.3) for carriers of ≥2 thrombophilia variants (Table S13).

Additional Analysis

Compound heterozygosity of rs6025 with a high-risk variant in *SERPINC1*, *PROC*, or *PROS1* increased the VTE risk (Table S14). The crude incidence rate was 12.1 (7.6–18.4) per 1000 person-years with an adjusted HR of 2.6 (95% CI, 1.8–3.7). No significant interaction was observed. The HR for the interaction term was 1.02 (not included in the Table). Compound heterozygosity of rs6025 or rs1799963 with a low-risk variant in the *SERPINC1*, *PROC,* or *PROS1* did not significantly increase the VTE risk but the number of individuals were few (Table S15). No significant interaction was observed between genetic risk factors and other included variables in the models (Figure S2).

DISCUSSION

In the present study of elderly and middle-aged adults, classic thrombophilias were found to be associated with a dose-graded risk of incident VTE. Present recommendations[7,18] suggest thrombophilia screening before the age of 50 years among patients with VTE, but classic thrombophilias also inflict an increased risk of VTE in middle-aged and older adults. Although the relative risk is moderate, this corresponds to a high incidence rate of VTE for classic thrombophilias in elderly and middle-aged adults, especially for those with ≥2 risk alleles. To the best of our knowledge, this is the first large population-based cohort study determining the prevalence and VTE risk of disease-causing *PROC*, *PROS1*, and *SERPINC1* variants among middle-aged and older adults. *PROS1*, *PROC*, and *SERPINC1* disease-causing variants are present in the Swedish population with a frequency higher than previously anticipated from population studies based on plasma analysis of protein levels.[12–15] However, a recent gene-based study suggests higher prevalence.[16] Risk estimates were weaker than previous estimates in family studies of thrombosis-prone families.[3–7] Antithrombin type I deficiency was, however, rare, and it was not possible to estimate its VTE risk. Thus, there might exist rare variants that could cause more severe phenotypes, but it was not possible to estimate them in the present study. Most *SERPINC1* variants were heparin-binding site defects (type II antithrombin deficiency), which are known to be associated with lower VTE risk compared with type I deficiency.[3–7] Deficiencies of the 3 natural anticoagulants have previously been considered to be autosomal dominant disorders with varying degrees of penetrance on VTE risk. The present study instead suggests that many variants causing deficiency of the natural anticoagulants are risk variants, such as the rs6025 and rs1799963 variants. This was already suggested in 1995 in families with segregation of both protein S deficiency and rs6025.[31] In these families, protein S deficiency and rs6025 was of equal importance on VTE risk.[31] Still, homozygosity for severe variants in the natural anticoagulant genes (*PROC* and *PROS1*) cause a lethal condition called purpura fulminans.[3–7] Homozygosity for severe antithrombin deficiency has not been described and is probably lethal shortly after conception.

The lower-than-expected VTE risk observed not only for high-risk variants of *SERPINC1*, *PROC*, and

PROS1 genes but also for rs6025 and rs1799963, is partly related to the elderly and middle-age status of the included population. Thus, this study includes an age group with high incidence of VTE, even in the absence of any known genetic predisposition, which attenuates the genetic relative risk. It has been established that the genetic contribution of complex traits such as VTE is age-dependent.[32] In the present study OR was generally higher for prevalent VTE than for incident VTE (Table 2), which may be explained by lower age of the individuals with prevalent VTE (VTE before baseline) than incident VTE (VTE after study baseline during follow-up). The population-based cohort design of the study with unbiased sampling of cases may also decrease the OR/HR because of reduction of selection bias. Compared with only 1 incident VTE event, ORs were higher for recurrent incident VTE suggesting enrichment of thrombophilic variants in patients with recurrent disease. Divergent results exist for the importance of heterozygosity for the rs6025 and rs1799963 variants in recurrent VTE.[3–7]

Protein S Heerlen (Ser501Pro), antithrombin Dublin (Val30Glu) and antithrombin Basel (Pro73Leu) are all individually reported to be associated with VTE.[7,33,34] In the present study protein S Heerlen and antithrombin Dublin were significantly associated with VTE. The OR was not significant for antithrombin Basel. Antithrombin Basel is associated with decreased heparin cofactor activity, ie, type II heparin-binding site defects. Antithrombin Dublin is complex. It has been associated with transient antithrombin deficiency.[33]

The higher than expected prevalence (3.1%) of disease-causing mutations in protein S, protein C, and antithrombin genes based on next-generation sequencing in the Swedish population is in line with the finding of patients deficient in the 3 natural anticoagulants that are present in healthy adults.[12–16] Familial thrombophilia is often the result of >1 segregating in a family.[3,7,31] *PROS1*, *PROC*, or *SERPINC1* variants translating into low protein levels identified in families with a high prevalence of thrombotic events may occur together with other thrombophilias, such as factor V Leiden and the prothrombin G20210A variant, which was also shown in the present study (Table 3).[3,7,35,36] However, although there was a high incidence rate, the HR for combined genetic variants was lower than previously estimated from studies of selected thrombosis-prone families.[3] It is possible that families with high inheritance of VTE have not only 1 or 2 risk variants, but even more variants might be present.

This study has clinical implications. The prevalence of disease-causing variants is higher than expected and targeted sequencing in high-risk individuals for variants in *SERPINC1*, *PROC*, and *PROS1* even among middle-aged and older patients with a VTE might be indicated. Identified variants would then need to be curated and plasma-based studies of unknown non-synonymous mutations might be worthwhile. The finding of enrichment of classical thrombophilia among cases with recurrent incident VTE suggests that classical thrombophilia is involved in recurrent VTE, and further studies could be warranted. The present study may not give a conclusive answer as to whether thromboprophylaxis is warranted in asymptomatic individuals, but the present study at least suggests that recommendations for younger individuals with classical thrombophilia should also be valid for middle-aged and older individuals with classical thrombophilia.

Limitations of the study are that no family data exist, and no plasma samples are available for measurement of the individual proteins, ie, antithrombin, protein C, and protein S. We therefore do not know which of the de novo mutations are associated with a low protein level and/or anticoagulant activity of the respective protein. However, all identified variants were rare, and many had a positive prediction score for pathogenicity (Tables S2 through S4). Still, variants that are absent in the HGMD were not associated with thrombosis. A strength of the study is the identification of high-risk variants using HGMD variants previously associated with deficiencies of the 3 anticoagulant proteins. Another strength is that the prevalence of the factor V Leiden and the prothrombin G20210A thrombophilic variants is similar to the SweGen project (https://swefreq.nbis.se/) and previous studies from Malmö.[16,35,36] A further strength is also the high coverage of VTE diagnosis in the Malmö registers and that VTE diagnosis is confirmed by objective methods in Malmö and Sweden, and the high quality of Swedish registers.[20–22] The present study also confirmed the factor V paradox, ie, rs6025 is associated with a higher risk of DVT than PE, which further strengthens the validity of our data.[7] A minor limitation is that Swedes who had a fatal thrombosis prior to the recruitment (which occurred at a mean age of 58 years) would not have been included in the study cohort. This phenomenon, called depletion of the susceptible, tends to attenuate the HR toward the null. Another limitation is that we were unable to study provoked and unprovoked cases separately. Higher relative genotype risks are likely to be found for unprovoked cases. A limitation is also the classification of novel missense changes at amino acid residues where different missense changes determined to be pathogenic have been seen before in the HGMD database.[26] However, using ACMG criteria gave weaker association with VTE than our definition of high and low risk variants.

Limited information about ethnicity exists. However, only those foreigners or immigrants who could speak Swedish adequately were included in the study.[34] In

total, 88% (not shown) of the total MDC cohort were born in Sweden (around 99% of those born outside Sweden were from European countries).[37] Still, we included the top 2 eigenvectors from the PCA analysis as covariates in Cox proportional hazard regression models to control for population stratification. We therefore believe our results are of general interest and valid for many individuals of White origin in Europe and the United States.[38,39] The Swedish population is closely related to German and British people.[38–40] Moreover, Malmö, and Skåne were an integral part of Denmark until AD 1658.[41] Malmö and several cities in Skåne were important towns in the Hanseatic league.[42] Analysis of genetic differentiation indicating that the population of Sweden's southernmost counties are genetically closer to samples of Northern European ancestry than to the populations of Sweden's northernmost counties.[43,44] Moreover, there is a high degree of correlation in minor allele frequency for variants in genes involved in VTE between gnomAD (non-Finnish Europeans) and SweGen giving further support for the generalizability of our results.[16,45]

In conclusion, the present study shows that classic thrombophilias are important even in elderly and middle-aged people. The relative genetic risk of classic thrombophilia was dose-graded. Moreover, *PROC*, *PROS1*, *SERPINC1* disease causing variants were more common than the rs1799963 variant, but the genetic risk conferred by them was lower than previously anticipated and comparable with the rs6025 and rs1799963 variants. A reservation for certain rare type I deficiency variants causing a severe phenotype should be made. However, these are rare variants, and they were not possible to risk-estimate in the present population-based cohort study.

APPENDIX

Regeneron Genetics Center

RGC Management and Leadership Team

Goncalo Abecasis, PhD, Aris Baras, MD, Michael Cantor, MD, Giovanni Coppola, MD, Aris Economides, PhD, Luca A. Lotta, MD, PhD, John D. Overton, PhD, Jeffrey G. Reid, PhD, Alan Shuldiner, MD

Contribution: All authors contributed to securing funding, study design and oversight. All authors reviewed the final version of the manuscript.

Sequencing and Laboratory Operations

Christina Beechert, Caitlin Forsythe, M.S., Erin D. Fuller, Zhenhua Gu, M.S., Michael Lattari, Alexander Lopez, M.S., John D. Overton, PhD, Thomas D. Schleicher, M.S., Maria Sotiropoulos Padilla, M.S., Louis Widom, Sarah E. Wolf, M.S., Manasi Pradhan, M.S., Kia Manoochehri, Ricardo H. Ulloa.

Contribution: C.B., C.F., A.L., and J.D.O. performed and are responsible for sample genotyping. C.B, C.F., E.D.F., M.L., M.S.P., L.W., S.E.W., A.L., and J.D.O. performed and are responsible for exome sequencing. T.D.S., Z.G., A.L., and J.D.O. conceived and are responsible for laboratory automation. M.P., K.M., R.U., and J.D.O are responsible for sample tracking and the library information management system.

Genome Informatics

Xiaodong Bai, PhD, Suganthi Balasubramanian, PhD, Andrew Blumenfeld, Boris Boutkov, PhD, Gisu Eom, Lukas Habegger, PhD, Alicia Hawes, B.S., Shareef Khalid, Olga Krasheninina, M.S., Rouel Lanche, Adam J. Mansfield, B.A., Evan K. Maxwell, PhD, Mrunali Nafde, Sean O'Keeffe, M.S., Max Orelus, Razvan Panea, PhD, Tommy Polanco, B.A., Ayesha Rasool, M.S., Jeffrey G. Reid, PhD, William Salerno, PhD, Jeffrey C. Staples, PhD

Contribution: X.B., A.H., O.K., A.M., S.O., R.P., T.P., A.R., W.S. and J.G.R. performed and are responsible for the compute logistics, analysis, and infrastructure needed to produce exome and genotype data. G.E., M.O., M.N. and J.G.R. provided compute infrastructure development and operational support. S.B., S.K., and J.G.R. provide variant and gene annotations and their functional interpretation of variants. E.M., J.S., R.L., B.B., A.B., L.H., J.G.R. conceived and are responsible for creating, developing, and deploying analysis platforms and computational methods for analyzing genomic data.

Research Program Management

Marcus B. Jones, PhD, Jason Mighty, PhD, Lyndon J. Mitnaul, PhD

Contribution: All authors contributed to the management and coordination of all research activities, planning, and execution. All authors contributed to the review process for the final version of the manuscript.

ARTICLE INFORMATION

Received July 16, 2021; accepted November 11, 2021.

Affiliations

Department of Environmental Science and Bioscience, Kristianstad University, Kristianstad, Sweden (E.M., C.L., C.H.); ; Department of Clinical Sciences (J.E., P.J.S., G.E., O.M.); and Department of Translational Medicine (B.D.), Lund University, Malmö, Sweden; ; Regeneron Genetics Center, Tarrytown, NY (A.B., L.A.L., ..); and Center for Primary Health Care Research, Lund University and Region Skåne, Malmö, Sweden (B.Z.).

Sources of Funding

This work was supported by a grant awarded to Dr Bengt Zöller by ALF–funding (the Medical Training and Research Agreement—Avtal om Läkarutbildning och Forskning) from Region Skåne, Sparbanken Skåne, and by the Swedish Research Council.

Disclosures

None.

Supplemental Material

Tables S1–S15
Figures S1–S2

REFERENCES

1. Lindström S, Wang LU, Smith EN, Gordon W, van Hylckama Vlieg A, de Andrade M, Brody JA, Pattee JW, Haessler J, Brumpton BM, et al. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood*. 2019;134:1645–1657. doi: 10.1182/blood.2019000435
2. Klarin D, Busenkell E, Judy R, Lynch J, Levin M, Haessler J, Aragam K, Chaffin M, Haas M, Lindström S, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat Genet*. 2019;51:1574–1579. doi: 10.1038/s41588-019-0519-3
3. Zöller B, García de Frutos P, Hillarp A, Dahlbäck B. Thrombophilia as a multigenic disease. *Haematologica*. 1999;84:59–70.
4. Rosendaal FR, Reitsma PH. Genetics of venous thrombosis. *J Thromb Haemost*. 2009;7:301–304. doi: 10.1111/j.1538-7836.2009.03394.x
5. Mannucci PM, Franchini M. Classic thrombophilic gene variants. *Thromb Haemost*. 2015;114:885–889. doi: 10.1160/TH15-02-0141
6. Trégouét DA, Morange PE. What is currently known about the genetics of venous thromboembolism at the dawn of next generation sequencing technologies. *Br J Haematol*. 2018;180:335–345. doi: 10.1111/bjh.15004
7. Zöller B, Svensson PJ, Dahlbäck B, Lind-Hallden C, Hallden C, Elf J. Genetic risk factors for venous thromboembolism. *Expert Rev Hematol*. 2020;13:971–981. doi: 10.1080/17474086.2020.1804354
8. Ridker PM, Hennekens CH, Lindpaintner K, Stampfer MJ, Eisenberg PR, Miletich JP. Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. *N Engl J Med*. 1995;332:912–917. doi: 10.1056/NEJM199504063321403
9. Ridker PM, Hennekens CH, Miletich JP. G20210A mutation in prothrombin gene and risk of myocardial infarction, stroke, and venous thrombosis in a large cohort of US men. *Circulation*. 1999;99:999–1004. doi: 10.1161/01.CIR.99.8.999
10. Sode BF, allin KH, Dahl M, Gyntelberg F, Nordestgaard BG. Risk of venous thromboembolism and myocardial infarction associated with factor V Leiden and prothrombin mutations and blood type. *CMAJ*. 2013;185:E229–E237. doi: 10.1503/cmaj.121636
11. Pintao MC, Ribeiro DD, Bezemer ID, Garcia AA, de Visser MC, Doggen CJ, Lijfering WM, Reitsma PH, Rosendaal FR. Protein S levels and the risk of venous thrombosis: results from the MEGA case-control study. *Blood*. 2013;122:3210–3219. doi: 10.1182/blood-2013-04-499335
12. Miletich J, Sherman L, Broze G Jr. Absence of thrombosis in subjects with heterozygous protein C deficiency. *N Engl J Med*. 1987;317:991–996. doi: 10.1056/NEJM198710153171604
13. Tait RC, Walker ID, Perry DJ, Islam SI, Daly ME, McCall F, Conkie JA, Carrell RW. Prevalence of antithrombin deficiency in the healthy population. *Br J Haematol*. 1994;87:106–114. doi: 10.1111/j.1365-2141.1994.tb04878.x
14. Wells PS, Blajchman MA, Henderson P, Wells MJ, Demers C, Bourque R, McAvoy A. Prevalence of antithrombin deficiency in healthy blood donors: a cross-sectional study. *Am J Hematol*. 1994;45:321–324. doi: 10.1002/ajh.2830450409
15. Tait RC, Walker ID, Reitsma PH, Islam SI, McCall F, Poort SR, Conkie JA, Bertina RM. Prevalence of protein C deficiency in the healthy population. *Thromb Haemost*. 1995;73:87–93. doi: 10.1055/s-0038-1653730
16. Zöller B. Prevalence and in silico analysis of missense mutations in the PROS1 gene in the Swedish population: the SweGen dataset. *Thromb Res*. 2018;168:28–30. doi: 10.1016/j.thromres.2018.06.001
17. Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, Hayden M, Heywood S, Millar DS, Phillips AD, et al. The Human Gene Mutation Database (HGMD): optimizing its use in a clinical diagnostic or research setting. *Hum Genet*. 2020;139:1197–1207. doi: 10.1007/s00439-020-02199-3
18. Connors JM. Thrombophilia testing and venous thrombosis. *N Engl J Med*. 2017;377:1177–1187. doi: 10.1056/NEJMra1700365
19. Melander O, Newton-Cheh C, Almgren P, Hedblad B, Berglund G, Engström G, Persson M, Smith JG, Magnusson M, Christensson A. Novel and conventional biomarkers for prediction of incident cardiovascular events in the community. *JAMA*. 2009;302:49–57. doi: 10.1001/jama.2009.943
20. Rosengren A, Freden M, Hansson P-O, Wilhelmsen L, Wedel H, Eriksson H. Psychosocial factors and venous thromboembolism: a long-term follow-up study of Swedish men. *J Thromb Haemost*. 2008;6:558–564. doi: 10.1111/j.1538-7836.2007.02857.x
21. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, Heurgren M, Olausson PO. External review and validation of the Swedish national inpatient register. *BMC Public Health*. 2011;11:450. doi: 10.1186/1471-2458-11-450
22. Isma N, Svensson PJ, Gottsäter A, Lindblad B. Prospective analysis of risk factors and distribution of venous thromboembolism in the population-based Malmö Thrombophilia Study (MATS). *Thromb Res*. 2009;124:663–666. doi: 10.1016/j.thromres.2009.04.022
23. Williams RC, Knowler WC, Shuldiner AR, Gosalia N, Van Hout C, Center RG, Hanson RL, Bogardus C, Baier LJ. Next generation sequencing and the classical HLA loci in full heritage Pima Indians of Arizona: defining the core HLA variation for North American Paleo-Indians. *Hum Immunol*. 2019;80:955–965. doi: 10.1016/j.humimm.2019.10.002
24. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164. doi: 10.1093/nar/gkq603
25. Park KJ, Lee W, Chun S, Min WK. The frequency of discordant variant classification in the Human Gene Mutation Database: a comparison of the American College of Medical Genetics and Genomics guidelines and ClinVar. *Lab Med*. 2021;52:250–259. doi: 10.1093/labmed/lmaa072
26. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–424. doi: 10.1038/gim.2015.30
27. Kim S, Jhong JH, Lee J, Koo JY. Meta analytic support vector machine for integrating multiple omics data. *BioData Min*. 2017;10:2. doi: 10.1186/s13040-017-0126-8
28. Kim HI, Ye B, Gosalia N, Köroğlu Ç, Hanson RL, Hsueh W-C, Knowler WC, Baier LJ, Bogardus C, Shuldiner AR, et al. Characterization of exome variants and their metabolic impact in 6,716 American Indians from the Southwest US. *Am J Hum Genet*. 2020;107:251–264. doi: 10.1016/j.ajhg.2020.06.009
29. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74. doi: 10.1038/nature15393
30. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. doi: 10.1186/s13742-015-0047-8
31. Zöller B, Berntsdotter A, García de Frutos P, Dahlbäck B. Resistance to activated protein C as an additional genetic risk factor in hereditary deficiency of protein S. *Blood*. 1995;85:3518–3523. doi: 10.1182/blood.V85.12.3518.bloodjournal85123518
32. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science*. 1994;265:2037–2048. doi: 10.1126/science.8091226
33. Navarro-Fernández J, Morena-Barrio M, Padilla J, Miñano A, Bohdan N, Águila S, Martínez-Martínez I, Sevivas T, de Cos C, Fernández-Mosteirín N, et al. Antithrombin Dublin (p.Val30Glu): a relatively common variant with moderate thrombosis risk of causing transient antithrombin deficiency. *Thromb Haemost*. 2016;116:146–154. doi: 10.1160/TH15-11-0871
34. Puurunen M, Salo P, Engelbarth S, Javela K, Perola M. Type II antithrombin deficiency caused by a founder mutation Pro73Leu in the Finnish population: clinical picture. *J Thromb Haemost*. 2013;11:1844–1849. doi: 10.1111/jth.12364
35. Svensson PJ, Zöller B, Mattiasson I, Dahlbäck B. The factor VR506Q mutation causing APC resistance is highly prevalent amongst unselected outpatients with clinically suspected deep venous thrombosis. *J Intern Med*. 1997;241:379–385. doi: 10.1046/j.1365-2796.1997.124140000.x

36. Hillarp A, Zöller B, Svensson PJ, Dahlbäck B. The 20210 A allele of the prothrombin gene is a common risk factor among Swedish outpatients with verified deep venous thrombosis. *Thromb Haemost*. 1997;78:990–992. doi: 10.1055/s-0038-1657674
37. Lahmann PH, Lissner L, Gullberg B, Berglund G. Differences in body fat and central adiposity between Swedes and European immigrants: the Malmö Diet and Cancer Study. *Obes Res*. 2000;8:620–631. doi: 10.1038/oby.2000.80
38. Nelis M, Esko T, Mägi R, Zimprich F, Zimprich A, Toncheva D, Karachanak S, Piskáčková T, Balaščák I, Peltonen L, et al; Genetic structure of Europeans: a view from the North East. Genetic structure of Europeans: a view from the North East. *PLoS One*. 2009;4:e5472. doi: 10.1371/journal.pone.0005472
39. Price AL, Butler J, Patterson N, Capelli C, Pascali VL, Scarnicci F, Ruiz-Linares A, Groop L, Saetta AA, Korkolopoulou P, et al. Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet*. 2008;4:e236. doi: 10.1371/journal.pgen.0030236
40. Zöller B. Nationwide family studies of cardiovascular diseases—clinical and genetic implications of family history. *EMJ Cardiol*. 2013;1:102–113.
41. Neumann J. Great historical events that were significantly affected by the weather: 3, The cold winter of 1657–58, the Swedish army crosses Denmark's Frozen sea areas. *Bull Am Meteor Soc*. 1978;59:1432–1437. doi: 10.1175/1520-0477(1978)059<1432:GHETWS>2.0.CO;2
42. Gaimster D. The Hanseatic cultural signature: exploring globalization on the micro-scale in late medieval northern Europe. *Eur J Archaeol*. 2014;17:60–81. doi: 10.1179/1461957113Y.0000000044
43. Humphreys K, Grankvist A, Leu M, Hall P, Liu J, Ripatti S, Rehnström K, Groop L, Klareskog L, Ding BO, et al. The genetic structure of the Swedish population. *PLoS One*. 2011;6:e22547. doi: 10.1371/journal.pone.0022547
44. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, Viklund J, Kähäri AK, Lundin P, Che H, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. *Eur J Hum Genet*. 2017;25:1253–1260. doi: 10.1038/ejhg.2017.130
45. Manderstedt E, Lind-Halldén C, Svensson P, Zöller B, Halldén C. Next-generation sequencing of 17 genes associated with venous thromboembolism reveals a deficit of non-synonymous variants in procoagulant genes. *Thromb Haemost*. 2019;119:1441–1450. doi: 10.1055/s-0039-1693130

Supplemental Material

Table S1. International Classification of Diseases (ICD) 7th, 8th, 9th and 10th revisions codes used for defining venous thromboembolism (VTE). ICD-9 and ICD-10 codes used for follow-up period and ICD-7, ICD-8, and ICD-9 codes used for VTE before baseline. Deep venous thrombosis (DVT) of the legs is marked with * and Pulmonary embolism (PE) with †.

ICD-10	ICD-9	ICD-8	ICD-7
I26†	325	321	33440
I636	415B†	450†	33450
I676	416W†	451*	463*
I80*	437G	452	464
I81	451*	453	465†
I82	452	671*	466
O222	453	6739†	58300
O223*	639G†		682*
O225	671C		684†
O229	671D*		
O082†	671E*		
O870	671F		
O871*	671X		
O873	673C†		
O879			
O882†			

*Deep venous thrombosis (DVT) of the legs: excluded I800, 451A and 67100.

†Pulmonary embolism (PE).

Table S2. Antithrombin deficiency (ATD). All non-synonymous (ns) and Loss of function mutations (LOF) found in the *SERPINC1* gene among the Malmö diet and cancer (MDC) cohort. Amino acids are numbered according to HGVS nomenclature and not legacy (HGVS-32aa). Variants were annotated with Annovar (2019-10-24).

Chr	Position	Ref	Alt	Codon	Amino acids	HGMD	ATD type	ATD variant name or ref	dbSNP	Meta SVM	Alleles in controls	Alleles cases	gnomAD MAF
1	173917231	G	T	10	T/N	-	-	-	rs61736655	T	7	2	0.000335
1	173914914	A	G	16	V/A	-	-	-	rs531137446	T	3	0	0.000185
1	173914872	A	T	30	V/E	CM900038	?	†#Dublin	rs2227624	T	238	48	0.00305
1	173914861	C	T	34	G/R	-	-	-	rs773254902	T	26	3	0.0000528
1	173914828	G	A	45	R/W	CM128535	I or II	§Caspers 2012	rs768704768	D	2	0	-
1	173914795	G	A	56	R/C	CM890015	II-HBS	†‡Rouen-4	rs28929469	D	4	0	0.0000264
1	173914748	CTTCTG	C	70-71	QK/X	-	-	-	.	.	0	1	-
1	173914743	G	A	73	P/L	CM860004	II-HBS	†‡Basel	rs121909551	T	49	9	0.000888
1	173914725	C	T,G	79	R/P	CM890016*	II-HBS	§Chen 1997	rs121909552	D	7	1	-
1	173914726	G	A	79	R/C	CM860005	II-HBS	†‡Toyama	rs121909547	D	2	0	0.0000264
1	173914695	C	T	89	R/H	-	-	-	rs745583962	T	1	0	0.0000352
1	173914662	T	C	100	D/G	-	-	-	rs369524182	T	10	1	0.000193
1	173914650	T	A	104	D/V	-	-	-	rs200118419	T	1	0	0.0000879
1	173911941	C	A	161	R/L	CM900039*	-	§Gandrille 1990	rs121909563	D	1	0	0.0000264
1	173911888	A	C	179	F/V	CM041434*	-	§ Picard 2006	rs773822689	D	0	1	8.79E-06
1	173911862	A	C	187	N/K	-	-	-	.	T	5	0	-
1	173911848	T	A	192	D/V	-	-	-	.	D	1	0	-
1	173911828	C	T	199	G/R	-	-	-	rs778341415	T	3	1	0.0000176
1	173911824	G	A,T	200	A/V	-	-	-	rs748428859	D	2	0	-
1	173911820	C	A	201	K/N	-	-	-	rs779025291	T	1	0	8.79E-06
1	173911819	G	T	202	L/I	-	-	-	.	D	1	0	-
1	173910837	C	T	227	E/K	CM952085*	-	§Csurgay 1995	.	T	3	1	0.0000155
1	173910822	C	T	232	D/N	-	-	-	.	T	1	0	-
1	173910796	A	C	240	N/K	-	-	-	.	T	2	1	-
1	173910797	T	C	240	N/S	CM153297	II	§Zeng 2015	rs200861147	T	2	0	8.79E-06
1	173910789	T	G	243	T/P	-	-	-	.	D	1	0	-
1	173909900	C	T	269	E/K	CM930054	II-HBS	†‡Truro	rs758087836	T	2	1	0.0000177
1	173909838	C	A	289	K/N	-	-	-	.	T	0	1	-
1	173909833	C	T	291	R/H	-	-	-	rs377588972	T	1	0	8.83E-06
1	173909834	G	A	291	R/C	-	-	-	rs764695432	T	7	0	0.0000706
1	173909831	A	G	292	Y/H	-	-	-	rs769991153	D	4	1	0.0000177

1	173909827	C	T	293	R/Q	CM063129*	-	§Picard 2006	rs572313182	T	1	0	0.0000177
1	173909824	C	T	294	R/H	-	-	-	rs587776397	T	3	0	0.0000441
1	173909819	C	G	296	A/P	CM128544	I or II	§Caspers 2012	rs372820797	T	2	1	0.000124
1	173909683	T	A	341	D/V	-	-	-	.	D	1	0	-
1	173909676	C	A	343	L/F	-	-	-	rs745357314	T	4	0	8.79E-06
1	173909677	A	C	343	L/W	-	-	-	.	D	1	0	,
1	173909645	G	A	354	R/C	-	-	-	.	D	2	0	0.0000155
1	173909638	C	T	356	R/H	-	-	-	rs373515340	T	3	1	0.0000264
1	173909620	C	A	362	S/I	-	-	-	rs762004419	D	1	0	0.0000088
1	173909609	G	C	366	Q/E	-	-	-	rs565091601	T	6	1	0.0000704
1	173909573	G	A	378	P/S	-	-	-	.	T	1	0	-
1	173907474	A	T	398	D/E	-	-	-	rs372611817	T	2	0	0.0000352
1	173904038	C	A	416	A/S	CM910058	II-RS	†‡Cambridge-2	rs121909548	D	30	5	0.00139
1	173903954	T	C	444	I/V	-	-	-	rs777118044	D	6	0	8.79E-06
1	173903942	G	A	448	P/S	-	-	-	rs376029223	T	10	1	0.0000528
1	173903936	T	C	450	N/D	-	-	-	rs747412993	D	2	0	0.0000176
										Number of variant alleles in controls and cases	462	81	

HGMD®=The Human Gene Mutation Database, http://www.hgmd.cf.ac.uk/ac/index.php
Variants found in the HGMD database denoted with HGMD accession number were included among high-risk variants in the statistical analysis.
HGMD accession number refers to a different amino acid substitution at the same codon. Although CM890016 refers to a R/H substitution literature search identified an experimental identified an ATD II-HBS not in HGMD (Chen 1997 see below §).
ATD=Antithrombin deficiency.
Three ATD variants were not classified as type I or II. One ATD II variant was not classified according to subtype.
RS=Reactive site defects, HBS=heparin binding defects, No ATD variants were classified as ATD-II pleiotropic effects.
Chr: chromosome
†=variant present in the Antithrombin mutation database http://www.imperial.ac.uk/immunology-inflammation/research/haematology/haemostasis-and-thrombosis/database/
‡Variant present in Uniprot https://www.uniprot.org/
Predictions score: MetaSVM is an ensemble score using Support Vector Machine (SVM) to integrate nine prediction scores (SIFT, PolyPhen-2, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy and PhyloP) and allele frequencies in 1KG database (1000 Genomes Project). D=damaging, T=tolerant.
gnomAD: Minor allele frequency among European non-Finnish population in gnomAD (https://gnomad.broadinstitute.org/) from https://varsome.com.
dbSNP: The Single Nucleotide Polymorphism Database https://www.ncbi.nlm.nih.gov/snp/
§References all in HGMD except Chen 1997: according to experimental COS-1 study by Chen (1997) Thesis, McMaster University; Caspers (2012) Thromb Haemost 108, 247; Zeng (2015) Thromb Haemost 113, 262; Picard (2006) Hum Mutat 27, 600; Csurgay (1995) SERPINC1 AB, 59; Gandrille (1990) J Biol Chem 265.

Table S3. Protein C deficiency (PCD). All non-synonymous (ns) and Loss of function mutations (LOF) found in the *PROC* gene among the Malmö diet and cancer (MDC) cohort. Amino acids are numbered according to HGVS nomenclature and not legacy (HGVS-42aa). Variants were annotated with Annovar (2019-10-24).

Chr	Position	Ref	Alt	Codon	Amino acids	HGMD	PCD type	PCD variant name or ref	dbSNP	Meta SVM	Alleles in controls	Alleles in cases	gnomAD MAF
2	127419973	G	A	11	V/M	-	-	-	rs368493458	D	2	0	0.0000616
2	127419994	G	A	18	G/S	CM128560	I or II	‡Caspers 2012	rs146793243	D	2	1	0.000114
2	127421301	G	A	30	S/N	-	-	-	-	D	5	1	-
2	127421303	G	A	31	E/K	CM129865	I	‡Pai 2012	rs779710709	D	0	1	-
2	127421304	A	G	31	E/G	-	-	-	-	D	2	0	-
2	127421318	G	A	36	V/M	-	-	-	-	D	2	0	-
2	127421339	G	A	43	A/T	CM004566	II	‡Dodojacek 2000	rs767626189	D	0	1	0.000149
2	127421350	C	G	46	F/L	-	-	-	rs141040323	D	12	1	0.0000529
2	127421369	A	G	53	S/G	CM005569*	-	‡Alhenc-Gelas 2000	-	D	1	1	-
2	127421372	A	T	54	S/C	CM950978	II	‡Hernandez 1995	rs376049280	D	15	0	0.000115
2	127421381	C	T	57	R/W	CM950980	II	‡†Reitsma 1995	rs757583846	D	18	7	0.000044
2	127421393	G	A	61	E/K	-	-	-	-	D	1	0	-
2	127421438	G	A	76	V/M	CM920593	II	†Vermont-1	rs121918149	D	4	1	0.000106
2	127423060	T	G	97	L/V	-	-	-	-	D	1	0	-
2	127423093	C	T	108	H/Y	CM001769* CM950988*	-	‡Millar 2000 ‡Reitsma 1995	-	D	2	0	-
2	127423094	A	G	108	H/R	-	-	-	-	T	1	0	-
2	127423111	G	C	114	G/R	CM950991	I	‡†Lind 1995	rs374476971	D	0	2	0.0000389
2	127423157	G	A	129	R/H	CM950994	II	‡Reitsma 1995	rs746190838	D	3	0	0.0000681
2	127423298	T	C	142	L/P	CM128564	I or II	‡Caspers 2012	rs1018638178	T	3	2	0.0000186
2	127426111	A	T,C	188	K/*	-	-	-	-	-	0	1	-
2	127426114	C	T	189	R/W	CM951003	II	†La Jolla-3	rs146922325	D	5	0	0.000114
2	127426120	GAGA	G	191-192	EK/E	-	-	-	rs199469469	-	1	0	-
2	127426129	C	T	194	R/C	CM951004	II	‡†Reitsma 1995	rs371071104	D	1	0	0.000106
2	127426165	G	A	206	D/N	-	-	-	-	T	1	0	-
2	127426178	C	T	210	P/L	CM951006	I	‡†Reitsma 1995	rs121918145	D	1	0	8.79E-06
2	127426181	G	A	211	R/Q	CM930611	I	‡Poort 1993	rs199469476	D	0	1	8.79E-06
2	127426208	G	A	220	R/Q	CM910314	I	†Vermont-3	rs121918153	T	2	0	0.0000616
2	127427178	C	T	251	A/V	CM951010	I	‡Gandrille 1995	rs568121876	D	1	0	-

2	127428374	C	T	272	R/C	CM910317	I	†‡Reitsma 1991	rs121918154	D	2	0	0.0000353
2	127428441	G	A	294	S/N	CM930616	II	†‡Gandrille 1993	rs200721675	T	8	3	0.0000616
2	127428444	C	T	295	T/I	-	-	-	rs773761677	D	1	0	0.0000176
2	127428449	G	A	297	D/N	CM085656*	-	‡Kim 2008	rs199469471	D	1	0	0.0000264
2	127428485	G	A	309	A/T	CM920595	I	‡†Conard 1992	rs121918146	D	1	0	0.0000264
2	127428545	G	A	329	E/K	-	-	-	rs757023301	T	4	1	8.81E-06
2	127428560	G	A	334	G/S	CM910319	I	‡†Reitsma 1991	rs121918150	D	0	2	-
2	127428579	C	T	340	T/M	CM930620	I	†Vermont-2	rs766261022	D	1	0	0.0000176
2	127428602	C	T	348	R/*	CM870018	I	‡Romeo 1987	rs121918141	-	0	1	0.0000088
2	127428662	G	A	368	V/I	-	-	-	rs752981292	T	2	0	0.0000264
2	127428672	A	G	371	N/S	CM961160*	-	‡Simioni 1996	-	T	1	0	-
2	127428683	G	A	375	E/K	-	-	-	rs368901479	T	1	0	0.0000088
2	127428699	T	C	380	M/T	-	-	-	-	T	7	2	0.0000309
2	127428714	T	C	385	M/T	-	-	-	-	D	1	0	-
2	127428726	G	C	389	G/A	CM001770	I	‡Millar 2000	-	D	1	0	-
2	127428741	G	A	394	R/Q	-	-	-	rs767219916	T	1	0	0.0000177
2	127428750	C	T	397	A/V	CM981643	I	‡Hallam 1998	-	T	1	0	-
2	127428756	A	C	399	E/A	-	-	-	rs201399407	D	10	3	0.0000444
2	127428761	G	A	401	D/N	CM941192	II	†La Jolla-2 Osaka-7 & -8	rs142742242	D	2	2	0.0000445
2	127428782	G	A	408	A/T	-	-	-	rs374259918	T	1	0	0.0000178
2	127428801	G	C	414	W/S	-	-	-	rs768759265	D	4	0	0.0000535
2	127428821	A	G	421	S/G	-	-	-	rs764364405	D	1	0	-
2	127428860	G	A	434	V/I	-	-	-	rs766695272	T	1	0	8.89E-06
2	127428867	C	T	436	T/I	CM951036*	-	‡Reitsma 1995	-	D	1	0	8.88E-06
2	127428868	CAA	C	437	K/X	-	-	-	-	-	1	0	-
2	127428887	G	A	443	D/N	-	-	-	-	D	5	0	-
2	127428923	C	T	455	P/S	-	-	-	-	T	1	0	-
									Number of variant alleles in controls and cases		146	34	

HGMD®=The Human Gene Mutation Database, http://www.hgmd.cf.ac.uk/ac/index.php
Variants found in the HGMD database denoted with HGMD accession number were included among high-risk variants in the statistical analysis.
*HGMD accession number refers to a different amino acid substitution at the same codon.
PCD=Protein C deficiency.
Two PCD variants were not classified as type I or II.
Chr: chromosome
†Variant present in Uniprot https://www.uniprot.org/

Predictions score: MetaSVM is an ensemble score using Support Vector Machine (SVM) to integrate nine prediction scores (SIFT, PolyPhen-2, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy and PhyloP) and allele frequencies in 1KG database (1000 Genomes Project). D=damaging, T=tolerant.
gnomAD: Minor allele frequency among European non-Finnish population in gnomAD (https://gnomad.broadinstitute.org/) from https://varsome.com.
dbSNP: The Single Nucleotide Polymorphism Database https://www.ncbi.nlm.nih.gov/snp/

‡ References all in HGMD: Caspers (2012) Thromb Haemost 108, 247; Pai (2012) Ann Hematol 91, 1471; Dodojacek (2000) Thromb Res 100, 109; Alhenc-Gelas (2000) Thromb Haemost 83, 86; Hernandez (1995) Blood Coagul Fibrinolysis 6, 23; Reitsma (1995) Thromb Haemost 73, 876; Millar (2000) Hum Genet 106, 646; Lind (1995) Thromb Haemost 73, 186; Poort (1993) Blood Coagul Fibrinolysis 4, 273; Gandrille (1995) Blood 86, 2598; Reitsma (1991) Blood 78, 890; Kim (2008) Thromb Res 123, 412; Conard (1992) Lancet 339, 743; Romeo (1987) Proc Natl Acad Sci USA 84, 2829; Simioni (1996) Blood 88, 2101; Millar (2000) Hum Genet 106, 646; Hallam (1998) Clin Genet 54, 231.

Table S4. Protein S deficiency (PSD). All non-synonymous (ns) and Loss of function mutations (LOF) found in the *PROS1* gene among the Malmö diet and cancer (MDC) cohort. Amino acids are numbered according to HGVS nomenclature and not legacy (HGVS-41aa). Variants were annotated with Annovar (2019-10-24).

Chr	Position	Ref	Alt	Codon	Amino acids	HGMD	PSD type	PSD variant name or ref	dbSNP	Meta SVM	Alleles in controls	Alleles in cases	gnomAD MAF
3	93973726	GC	G	8	C/X	CD011648	?	‡Rezende 2001	rs764581697	-	1	0	9.09E-06
3	93973697	A	T	18	V/E	CM011473	I/III	†‡Rezende 2001	-	D	1	1	-
3	93927395	T	G	30	Q/P	-	-	-	-	D	1	0	-
3	93927365	C	A	40	R/L	CM951041	II	†‡Gandrille 1995	rs7614835	D	2	1	0.0000264
3	93927362	C	T	41	R/H	CM951040	II	†‡Gandrille 1995	rs963668412	D	1	0	8.79E-06
3	93927356	T	C	43	N/S	-	-	-	rs748858986	D	2	0	0.0000616
3	93927311	C	T	58	C/Y	-	-	-	-	D	1	0	-
3	93927257	G	A	76	P/L	CD051330*	-	†‡Biguzzi 2005	rs73846070	D	35	1	0.00131
3	93927251	G	A	78	T/M	CM951045	I	†‡Gandrille 1995	rs6122	D	1	2	0.0000616
3	93910699	A	G	89	L/P	-	-	-	-	D	0	1	-
3	93910697	C	T	90	R/H	CM951046	I	†‡Gandrille 1995	rs200886866	D	5	1	0.0000267
3	93910696	G	A	90	R/C	CM011460	I/III	†‡Rezende 2001	rs765935815	D	0	1	0.0000266
3	93910690	A	C	92	F/C	-	-	-	rs772748769	D	0	3	8.85E-06
3	93910682	C	T	95	G/E	CM961169	I	†‡Simmonds 1996	rs144526169	D	45	7	0.000592
3	93910681	C	G	95	G/R	CM971243	III	†‡Gandrille 1997	-	D	2	0	-
3	93910673	G	C	98	T/S	CM005598	I/III	†‡Espinosa-Parrilla 2000	rs142805170	T	4	0	0.000371
3	93910672	T	C	98	T/A	-	-	-	rs747923334	T	18	0	0.0000883
3	93910666	G	T	100	A/E	-	-	-	rs375363379	T	1	0	0.0000353
3	93910664	C	G	101	R/P	-	-	-	-	D	1	0	-
3	93910663	G	A	101	R/C	CM032993	II	†‡Boinot 2003	rs778731080	D	3	0	8.83E-06
3	93910619	C	T	116	A/T	-	-	-	-	D	1	0	-
3	93906132	G	C	120	Q/E	CM109995	I	‡Alhenc-Gelas 2010	-	D	1	0	-
3	93906117	G	C	125	P/A	-	-	-	-	D	1	0	-
3	93906059	G	T,A	144	T/N	CM951047	II	†‡Gandrille 1995	rs146366248	T	12	1	0.000738
3	93906041	T	C	150	Q/R	-	-	-	rs754149240	T	2	0	0.000105
3	93905912	A	G	158	I/T	-	-	-	-	D	1	0	-
3	93905903	C	T	161	C/Y	-	-	-	-	D	1	0	-
3	93905882	T	C	168	N/S	-	-	-	rs144430063	D	3	0	0.000221
3	93905855	T	C	177	N/S	-	-	-	-	D	1	0	-
3	93905822	T	C	188	N/S	-	-	-	rs146070827	T	1	0	8.82E-06
3	93900843	C	T	230	E/K	-	-	-	rs575777099	T	7	1	0.0000439

3	93900833	C	T	233	R/K	-	-	-	rs41267007	T	112	16	0.00368
3	93900830	T	C	234	Y/C	CM103688	II	†‡Fischer 2010	rs387906675	D	1	0	0.0000088
3	93898521	T	G	259	Y/S	-	-	-	-	T	5	0	8.82E-06
3	93898483	C	T	272	G/R	-	-	-	rs41267005	D	13	3	0.000415
3	93898465	C	T	278	D/N	CM163578*	-	‡Alhenc-Gelas 2016	rs750744790	D	29	11	0.0000882
3	93896690	A	C	284	V/G	-	-	-	rs751683365	T	2	1	0.0000177
3	93896633	G	A	303	A/V	-	-	-	rs772677117	D	1	0	8.82E-06
3	93896626	C	A	305	Q/H	-	-	-	-	T	0	1	-
3	93893099	C	T	330	R/Q	-	-	-	rs549405539	D	5	1	0.0000177
3	93893067	C	T	341	A/T	-	-	-	rs189883848	T	35	2	0.000387
3	93893024	C	T	355	R/H	CM041822	III	†‡Okada 2004	rs780863931	D	0	1	8.79E-06
3	93893021	C	A	356	G/V	-	-		-	T	1	0	-
3	93892993	A	C	365	N/K	CM136310	I	‡Tang 2013	rs199469491	T	62	9	0.000616
3	93892964	C	A	375	G/V	-	-		-	D	1	0	-
3	93886471	GCTAATA	GCTAACA,G	394-396	SIS/S	-	-	-	rs368987511	T	4	0	-
3	93886454	T	C	402	E/G	-	-	-	rs775715647	D	1	0	8.79E-06
3	93886368	G	A	431	P/S	-	-	-	rs765473908	D	1	0	0.0000176
3	93884889	G	A	444	P/L	CM128590	I or II	‡Caspers 2012	rs369244777	D	4	1	0.000187
3	93884887	G	A	445	R/C	-	-	-	rs5017719	D	3	0	0.0000444
3	93884857	A	C	455	L/V	-	-	-	-	D	1	0	8.82E-06
3	93884815	C	G	469	E/Q	-	-	-	-	T	2	0	0.0000088
3	93884804	A	AT	472	N/KX	-	-	-	-	-	1	0	-
3	93879306	A	G	501	S/P	CM951058	I/III	†Heerlen	rs121918472	T	107	25	0.00332
3	93879293	C	A	505	G/V	-	-	-	-	T	1	0	-
3	93879279	C	T	510	V/M	CM961187	III	§‡Borgel 1996	rs138925964	T	2	0	0.0000439
3	93879213	T	C	532	T/A	CM961190	I	†‡Borgel 1996	rs371028997	T	2	0	0.0000615
3	93877154	C	T	561	R/Q	-	-	-	rs377174703	T	1	0	0.000106
3	93877128	C	T	570	D/N	-	-	-	rs755684845	T	1	0	-
3	93877089	T	G	583	N/H	CM128596	I or II	‡Caspers 2012	rs139479630	T	2	0	0.000211
3	93877074	T	C	588	T/A	CM011471	?	‡Borgel 2001	rs142846443	T	9	2	0.000264
3	93877029	G	C	603	Q/E	-	-	-	rs751163405	T	1	0	8.79E-06
3	93877019	A	G	606	V/A	-	-	-	rs765135930	T	4	1	0.0000176
3	93876993	C	T	615	V/M	-	-	-	rs368612500	T	2	0	-
3	93874403	C	T	625	V/I	-	-	-	-	T	1	0	-
3	93874387	G	A	630	T/I	CM022828	II	†Minami 2001	rs202190731	T	8	0	0.000153

3	93874315	G	T	654	A/D	-	-	-	-	D	0	1	-
3	93874295	T	C	661	I/V	-	-	-	rs141122478	D	2	0	0.0000155
3	93874254	C	A	674	K/N	-	-	-	rs764034062	T	1	0	8.83E-06
Number of variants in controls and cases											583	94	

HGMD®=The Human Gene Mutation Database, http://www.hgmd.cf.ac.uk/ac/index.php
Variants found in the HGMD database denoted with HGMD accession number were included among high-risk variants in the statistical analysis.
*HGMD accession number refers to a different amino acid substitution at the same codon.
PSD=Protein S deficiency.
Four PSD variants were not classified as type I or II. Four variants were associated with both type I and III deficiency.
Chr: chromosome
†Variant present in Uniprot https://www.uniprot.org/
Predictions score: MetaSVM is an ensemble score using Support Vector Machine (SVM) to integrate nine prediction scores (SIFT, PolyPhen-2, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy and PhyloP) and allele frequencies in 1KG database (1000 Genomes Project). D=damaging, T=tolerant.
gnomAD: Minor allele frequency among European non-Finnish population in gnomAD (https://gnomad.broadinstitute.org/) from https://varsome.com.
dbSNP: The Single Nucleotide Polymorphism Database https://www.ncbi.nlm.nih.gov/snp/
‡ References all in the HGMD: Rezende (2001) PROS1 PC, Gandrille (1995) Blood 85, 130; Biguzzi (2005) Hum Mutat 25, 259; Simmonds (1996) Blood 88, 4195; Gandrille (1997) Thromb Haemost 77, 1201; Espinosa-Parrilla (2000) Hum Mutat 15, 463; Boinot (2003) Blood Coagul Fibrinolysis 14, 191; Alhenc-Gelas (2010) J Thromb Haemost 8, 2718; Fischer (2010) Neonatology 98, 337; Okada (2004) Br J Haematol 126, 219; Tang (2013) Am J Hematol 88, 899; Caspers (2012) Thromb Haemost 108, 247; Borgel (1996) J Lab Clin Med 128, 218; Borgel (2001) PROS1 PC; Minami (2001) Rinsho Ketsueki 42, 610.
§Type III deficiency according to Daneshjou et al Mol Genet Genomic Med. 2016;4:513-20.

Table S5. Summary of the 171 non-synonymous (ns) and Loss of function mutations (LOF) found in the *SERPINC1*, *PROC*, and *PROS1* genes among the entire (n=29,387) Malmö diet and cancer (MDC) cohort.

	SERPINC1 variants n (%)	*PROC* variants n (%)	*PROS1* variants n (%)	All variants n (%)
Number of ns+LOF variants	47 (100%)	55 (100%)	69 (100%)	171 (100%)
High risk, HGMD (same codon)	14 (29.8%)	29 (52.7%)	25 (36.2%)	68 (39.8%)
Non-synonymous variants	46 (97.9%)	53 (96.4%)	66 (95.7%)	165 (96.5%)
Missense variants	46 (97.9%)	51 (92.7%)	66 (95.7%)	163 (95.3%)
Nonsense variants	0 (0%)	2 (3.6%)	0 (0%)	2 (1.2%)
Small deletions	1 (2.1%)	2 (3.6%)	2 (2.9%)	5 (2.9%)
Small insertions	0 (0%)	0 (0%)	1 (1.4%)	1 (0.6%)
LOF	1 (2.1%)	4 (7.3%)	3 (4.3%)	8 (4.7%)
HGMD (same amino acid and same codon)	9 (19.1%)	24 (43.6%)	23 (33.3%)	56 (32.3%)
HGMD (same codon but different amino acid)	5 (10.6%)*	5 (9.1%)	2 (2.9%)	12 (7.0%)
Type I deficiency	0 (0%)	13 (23.6%)	6 (8.7%)	19 (11.1%)
Type II deficiency	7 (14.9%)*	9 (16.4%)	6 (8.7%)	22 (12.9%)
Type III deficiency	NA	NA	3 (4.3%)	3 (1.8%)
Type I/III deficiency	NA	NA	4 (5.8%)	4 (2.3%)
Type of deficiency not known	3 (6.4%)	2 (3.6%)	4 (5.8%)	9 (5.3%)
dbSNP (rsid)	34 (72.3%)	36 (65.5%)	47 (68.1%)	117 (68.4%)
In gnomAD (EU non-Finish)	33 (70.2%)	33 (60%)	47 (68.1%)	113 (66.1%)
Rare variants (MAF<1%)	47 (100%)	55 (100%)	69 (100%)	171 (100%)

EU non-Finnish=European non-Finnish population. Variants were annotated with Annovar (2019-10-24).
*One variant was experimentally confirmed to be associated with antithrombin deficiency: Chen 1997: according to experimental COS-1 study by Chen (1997) Thesis, McMaster University.

Table S6. Contingency tables and odds ratios calculated from contingency tables, for the categorical variables investigated by logistic regression and presented in Table 2. Crude odd ratios (OR) without confidence intervals are shown for all VTE (i.e. VTE between 1970 and baseline and/or during follow-up), prevalent VTE (i.e. VTE between 1970 and baseline), incident VTE during follow-up (without prevalent VTE), and recurrent VTE during follow-up (without prevalent VTE). For the dependent variables all outcomes were compared with no VTE (1970-2018). Prevalent cases (n=593) were excluded when calculating OR for incident VTE and recurrent VTE. The independent variables age, body mass index (BMI), and systolic blood pressure were included as continuous variables in Table 2 and are not displayed here.

	All VTE	No VTE	Odds ratio	Prevalent VTE	No VTE	Odds ratio	Incident VTE	No VTE	Odds ratio	Recurrent VTE	No VTE	Odds ratio
female	1869	15818		315	15818		1554	15818		890	15818	
not female	1308	10392	0.94	278	10392	0.74	1030	10392	0.99	601	10392	0.97
Smoker	822	6975		146	6975		676	6975		380	6975	
not Smoker	2355	19235	0.96	447	19235	0.90	1908	19235	0.98	1111	19235	0.94
High alcohol consumption	120	1057		23	1057		97	1057		51	1057	
no/low alcohol consumption	3057	25153	0.93	570	25153	0.96	2487	25153	0.93	1440	25153	0.84
Heterozygote rs6025	603	2679		153	2679		450	2679		292	2679	
No heterozygote rs6025	2574	23531	2.06	440	23531	3.05	2134	23531	1.85	1199	23531	2.14
Homozygote rs6025	44	75		16	75		28	75		21	75	
No homozygote rs6025	3133	26135	4.89	577	26135	9.66	2556	26135	3.82	1470	26135	4.98
Heterozygote rs1799963	90	452		18	452		72	452		53	452	
No rs1799963	3086	25757	1.66	575	25757	1.78	2511	25757	1.63	1438	25757	2.10
≥ 1 High-risk variant*	153	755		34	755		119	755		74	755	
No high-risk variant	3024	25455	1.71	559	25455	2.05	2465	25455	1.63	1417	25455	1.76
≥ 1 Low-risk variant†	55	428		10	428		45	428		28	428	
No low-risk variant	3122	25782	1.05	582	25782	1.02	2539	25782	1.06	1463	25782	1.14

*High-risk variant = High-risk non-synonymous HGMD variants in the three anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see method section).

†Low-risk variant = Low-risk non-synonymous non-HGMD variants in the three anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see method section).

Table S7. Hazard ratios for venous thromboembolism (VTE) with regard to high-risk variants in *SERPINC1*, *PROC*, and *PROS1*. Each genetic factor tested one-by-one either adjusted for age, sex, and ancestry*, or adjusted for age, sex, BMI, smoking status, systolic blood pressure, high alcohol consumption, and ancestry*. Only cases with 1 high-risk variant were included. Prevalent cases with VTE were excluded.

	Participants	VTE events	VTE per 1000 person years	Age- and sex-adjusted HR		Multivariable HR	
	n	n		HR (95% CI)	P-value	HR (95% CI)	P-value
SERPINC1	387	49	6.3 (4.7-8.3)	1.5 (1.1-1.9)	0.006	1.5 (1.1-1.9)	0.006
PROC	96	20	10.6 (6.5-16.4)	2.3 (1.5-3.6)	0.0002	2.4 (1.5-3.7)	9.5e-05
PROS1	385	48	6.3 (4.7-8.3)	1.5 (1.1-2.0)	0.005	1.5 (1.1-2.0)	0.007

*Ancestry was controlled for by including the top two eigenvectors from the PCA analysis as covariates in Cox proportional hazard regression models.

Table S8. Crude (non-adjusted) odd ratios (OR) for protein S Heerlen (PS Heerlen), antithrombin Dublin (AT Dublin) and antithrombin Basel (AT Basel) in all four venous thromboembolism (VTE) groups compared to the No VTE group (1970-2018).

	All VTE n = 3177 OR (95% CI)	P-value	Prevalent VTE n = 593 OR (95% CI)	P-value	Incident VTE n = 2584 OR (95% CI)	P-value	Recurrent incident VTE n = 1491 OR (95% CI)	P-value
PS Heerlen (n=122)	1.93 (1.22-2.94)	0.0031	3.34 (1.49-6.45)	0.0011	1.61 (0.93-2.62)	0.067	2.31 (1.26-3.91)	0.0033
AT Dublin (n=268)	1.66 (1.21-2.25)	0.0013	1.87 (0.92-3.36)	0.054	1.62 (1.13-2.26)	0.0057	1.55 (0.96-2.38)	0.053
AT Basel (n=55)	1.52 (0.7-2.94)	0.251	1.81 (0.29-5.84)	0.41	1.45 (0.6-3)	0.35	1.08 (0.26-2.93)	0.90
High-risk variant* (n=414)	1.68 (1.3-2.15)	0.00005	1.84 (1.04-2.99)	0.023	1.65 (1.24-2.16)	0.0004	1.80 (1.26-2.49)	7.7e-4

No VTE – Individuals without venous thromboembolism (VTE) between 1970 and 2018; All VTE – Individuals with at least one VTE; Prevalent VTE – Individuals with VTE before baseline, but not after; Incident VTE – Individuals with VTE event after baseline, but not before; Recurrent VTE – Individuals with VTE both before (prevalent) and after baseline (incident).

*High risk variant = non-synonymous HGMD variants in thee three anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* besides protein S Heerlen, antithrombin Dublin and antithrombin Basel (see method section).

Table S9. Hazard ratios (HR) for incident deep venous thrombosis (DVT) of the lower extremities. Prevalent cases with venous thromboembolism (VTE) were excluded. Each genetic factor tested one-by-one either adjusted for age, sex, and ancestry†, or adjusted for age, sex, BMI, smoking status, systolic blood pressure, and high alcohol consumption (>30 g/day for women and >40g/day for men), and ancestry† in the multivariable model. 95% confidence intervals (CI) and p-values are given.

	Participants	DVT	Crude IR	Crude IRR	Age- and sex-adjusted HR		Multivariable HR	
	N	N	IR (95% CI)	IRR (95%CI)	HR (95% CI)	P-value	HR (95% CI)	P-value
Complete cohort	28794	1232	2.1 (2.0-2.2)	1	-	-	-	-
Model with factor V Leiden (rs6025)								
Reference no rs6025	25562	953	1.8 (1.7-1.9)	1	1		1	
rs6025 heterozygotes	3129	261	4.1 (3.6-4.6)	2.3 (2.0-2.6)	2.3 (2.0-2.6)	1.5e-32	2.3 (2.0-2.7)	4.9e-33
rs6025 homozygotes	103	18	8.7 (5.2-13.8)	4.9 (3.1-7.8)	5.2 (3.2-8.2)	5.3e-12	5.4 (3.4-8.6)	1.9e-12
Model with prothrombin variant (rs1799963)								
Reference no rs1799963	28268	1199	2.0 (1.9-2.2)	1	1		1	
rs1799963 heterozygotes	524	32	2.9 (2.0-4.2)	1.4 (1.0-2.1)	1.5 (1.0-2.1)	0.03	1.5 (1.0-2.1)	0.03
rs1799963 homozygotes	2	1	30.0(0.8-167)	14.6 (2.1-104)	16.2 (2.3-115)	0.005	21.2 (3.0-151)	0.002
Model with non-synonymous variants in *SERPINC1*, *PROC*, and *PROS1* genes								
Reference no non-synonymous variants	27462	1152	2.0 (1.9-2.1)	1	1		1	
1 low-risk variant§	472	19	2.1 (1.2-3.2)	1.0 (0.6-1.6)	1.1 (0.7-1.6)	0.75	1.1 (0.7-1.6)	0.77

1 high-risk variant‡	853	58	3.3 (2.5-4.3)	1.6 (1.3-2.1)	1.7 (1.3-2.2)	0.00012	1.7 (1.3-2.2)	0.00012
≥ 2 high-risk variants‡	6	0	NA	NA	NA	NA	NA	NA
Model with all five classical thrombophilia variants: rs6025, rs1799963, and high risk variants in *SERPINC1*, *PROC*, and *PROS1* genes								
No thrombophilia*	24312	880	1.7 (1.6-1.9)	1	1		1	
1 thrombophilia variant	4225	315	3.6 (3.3-4.1)	2.1 (1.8-2.4)	2.1 (1.8-2.4)	1.3e-29	2.1 (1.9-2.4)	6.8e-30
2 or more thrombophilia variants	257	37	7.2 (5.1-9.9)	4.1 (3.0-5.8)	4.4 (3.1-6.1)	1.9e-18	4.4 (3.1-6.1)	1.5e-18

*No rs6025 allele, no rs1799963 allele and no high-risk variant
†Ancestry was controlled for by including the top two eigenvectors from the PCA analysis as covariates in Cox proportional hazard regression models.
‡High-risk variant = High-risk non-synonymous HGMD variants in the three anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see method section).
§Low-risk variant = Low-risk non-synonymous non-HGMD variants in the three anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see method section).

Table S10. Hazard ratios for incident pulmonary embolism (PE). Prevalent cases with venous thromboembolism (VTE) were excluded. Each genetic factor tested one-by-one either adjusted for age, sex, and ancestry†, or adjusted for age, sex, BMI, smoking status, systolic blood pressure, high alcohol consumption (>30 g/day for women and >40g/day for men), and ancestry† in the multivariable model. 95% confidence intervals (CI) and p-values are given.

	Participants	PE	Crude IR	Crude IRR	Age- and sex-adjusted HR		Multivariable HR	
	N	N	IR (95% CI)	IRR (95%CI)	HR (95% CI)	P-value	HR (95% CI)	P-value
Complete cohort	28794	1102	1.8 (1.7-2.0)	1	-	-	-	-
Model with factor V Leiden (rs6025)								
Reference no rs6025	25562	930	1.7 (1.6-1.9)	1	1		1	
rs6025 heterozygotes	3129	166	2.6 (2.2-3.0)	1.5 (1.2-1.7)	1.5 (1.2-1.7)	6.4e-6	1.5 (1.2-1.7)	4.4e-6
rs6025 homozygotes	103	6	2.8 (1.0-6.1)	1.6 (0.7-3.6)	1.7 (0.8-3.8)	0.19	1.7 (0.8-3.9)	0.18
Model with prothrombin variant (rs1799963)								
Reference no rs1799963	28268	1069	1.8 (1.7-1.9)	1	1		1	
rs1799963 heterozygotes	524	32	2.9 (2.0-4.2)	1.6 (1.1-2.3)	1.7 (1.2-2.4)	0.004	1.7 (1.2-2.3)	0.005
rs1799963 homozygotes	2	1	33.0 (0.8-184)	18.2 (2.6-129)	21.8 (3.1-155)	0.002	28.7 (4.0-205)	0.0008
Model with non-synonymous variants in *SERPINC1*, *PROC*, and *PROS1* genes								

Reference no non-synonymous variants	27462	1035	1.8 (1.7-1.9)	1	1		1	
1 low-risk variant§	472	22	2.4 (1.5-3.6)	1.3 (0.9-2.0)	1.3 (0.8-1.9)	0.28	1.3 (0.8-1.9)	0.29
1 high-risk variant‡	853	45	2.6 (1.9-3.4)	1.4 (1.0-1.9)	1.4 (1.0-1.9)	0.036	1.4 (1.0-1.9)	0.036
≥ 2 high-risk variants‡	6	0	NA	NA	NA	NA	NA	NA
Model with all five classical thrombophilia variants: rs6025, rs1799963, and high risk variants in *SERPINC1*, *PROC*, and *PROS1* genes								
No thrombophilia*	24312	863	1.7 (1.6-1.8)	1	1		1	
1 thrombophilia variant	4225	221	2.5 (2.2-2.9)	1.5 (1.3-1.7)	1.5 (1.3-1.7)	2.1e-7	1.5 (1.3-1.7)	1.4e-7
2 or more thrombophilia variants	257	18	3.4 (2.0-5.4)	2.0 (1.3-3.2)	2.1 (1.3-3.3)	0.002	2.1 (1.3-3.3)	0.002

*No rs6025 allele, no rs1799963 allele and no high-risk variant

†Ancestry was controlled for by including the top two eigenvectors from the PCA analysis as covariates in Cox proportional hazard regression models.

‡High-risk variant = High-risk non-synonymous HGMD variants in the three anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see method section).

§Low-risk variant = Low-risk non-synonymous non-HGMD variants in the three anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see method section).

Table S11. Hazard ratios (HRs) for incident venous thromboembolism (VTE) adjusted for either age, sex, and ancestry* or multivariable HRs adjusted for age, sex, body mass index, smoking, high alcohol consumption, and ancestry*. Incidence rates (IR) and incidence rate ratios (IRR) are also presented. Prevalent cases of VTE were excluded. The American College of Medical Genetics and Genomics (ACMG) determined with Varsome (https://varsome.com/) was used to define high-risk variants, i.e., likely pathogenic and pathogenic variants according to ACMG.

	Participants	VTE	Age at VTE event	Crude IR	Crude IRR	Age- and sex-adjusted HR		Multivariable HR	
	N	N	Years (SD)	IR (95% CI)	IRR (95%CI)	HR (95% CI)	P-value	HR (95% CI)	P-value
Model with non-synonymous variants in SERPINC1, PROC, and PROS1 genes									
Reference no high-risk variants	28342	2528	73.9 (8.6)	4.4 (4.2-4.5)	1	1		1	
1 high-risk variant†	446	55	74.9 (7.9)	6.1 (4.6-7.9)	1.4(1.1-1.8)	1.4 (1.1-1.9)	0.0087	1.4 (1.1-1.9)	0.0084
≥ 2 high-risk variants†	6	1	71.3 (NA)	8.0 (0.2-44.4)	1.8 (0.3-13.0)	1.6 (0.2-11.4)	0.63	1.4 (0.2-10.0)	0.72

†High risk variant = variant is likely pathogenic or pathogenic according to criteria from The American College of Medical Genetics and Genomics (ACMG) determined with Varsome (https://varsome.com/).
*Ancestry was controlled for by including the top two eigenvectors from the PCA analysis as covariates in Cox proportional hazard regression models.

Table S12. Hazard ratios (HRs) for incident venous thromboembolism (VTE) adjusted for either age, sex, and ancestry* or multivariable HRs adjusted for age, sex, body mass index, smoking, high alcohol consumption, and ancestry*. Incidence rates (IR) and incidence rate ratios (IRR) are also presented. Prevalent cases of VTE were excluded. VTE Cases with malignancy before VTE were excluded

	Participants	VTE	Age at VTE event	Crude IR	Crude IRR	Age- and sex-adjusted HR		Multivariable HR	
	N	N	Years (SD)	IR (95% CI)	IRR (95%CI)	HR (95% CI)	P-value	HR (95% CI)	P-value
Complete cohort	27720	1510	73.0 (8.9)	2.6 (2.5-2.8)	1	-	-	-	-
Model with factor V Leiden (rs6025)									
Reference no rs6025	24672	1216	73.6 (8.9)	2.4 (2.3-2.5)	1	1		1	
rs6025 heterozygotes	2953	274	70.9 (8.5)	4.6 (4.1-5.2)	1.9 (1.7-2.2)	1.9 (1.7-2.2)	8.8e-23	1.9 (1.7-2.2)	2.9e-23
rs6025 homozygotes	95	20	67.5 (7.1)	11.1 (6.8-17.2)	4.7 (3.0-7.3)	4.9 (3.1-7.6)	2.3e-12	5.1 (3.3-8.0)	5.3e-13
Model with prothrombin variant (rs1799963)									
Reference no rs1799963	27221	1464	73.1 (8.9)	2.6 (2.5-2.7)	1	1		1	
rs1799963 heterozygotes	497	45	69.5 (8.0)	4.5 (3.3-6.0)	1.7 (1.3-2.3)	1.8 (1.3-2.4)	2.0e-04	1.7 (1.3-2.3)	2.8e-04
rs1799963 homozygotes	2	1	69.8 (NA)	33.0 (0.84-184)	12.7 (1.8-89.9)	13.6 (1.9-96.6)	0.0091	18.4 (2.6-131.5)	0.0036

Model with non-synonymous variants in SERPINC1, PROC, and PROS1 genes									
Reference no non-synonymous variants	26445	1403	73.0 (9.0)	2.6 (2.4-2.7)	1	1		1	
1 low-risk variant§	458	30	72.5 (7.6)	3.4 (2.3-4.8)	1.3 (0.9-1.9)	1.3 (0.9-1.9)	0.11	1.3 (0.9-1.9)	0.13
1 high-risk variant‡	809	71	73.6 (8.3)	4.3 (3.4-5.5)	1.7 (1.3-2.1)	1.7 (1.3-2.1)	1.5e-05	1.7 (1.3-2.1)	1.3e-05
≥ 2 high-risk variants‡	6	2	69.1 (3.2)	22.0 (2.7-79.6)	8.6 (2.1-34.3)	8.4 (2.1-33.9)	0.0026	8.6 (2.1-34.4)	0.0024
Model with all five classical thrombophilia variants: rs6025, rs1799963, and high risk variants in SERPINC1, PROC, and PROS1 genes									
No thrombophilia†	23477	1119	73.6 (9.0)	2.3 (2.2-2.4)	1	1		1	
1 thrombophilia variant	4010	346	71.6 (8.5)	4.3 (3.8-4.7)	1.8 (1.6-2.1)	1.8 (1.6-2.1)	2.3e-23	1.9 (1.6-2.1)	7.4e-24
2 or more thrombophilia variants	233	45	67.9 (7.1)	10.2 (7.4-13.6)	4.4 (3.3-5.9)	4.6 (3.4-6.2)	7.0e-24	4.7 (3.5-6.3)	5.4e-24

*Ancestry was controlled for by including the top two eigenvectors from the PCA analysis as covariates in Cox proportional hazard regression models.
†No rs6025 allele, no rs1799963 allele and no high-risk variant.
‡High-risk variant = High-risk non-synonymous HGMD variants in the three anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see method section).
§Low-risk variant = Low-risk non-synonymous non-HGMD variants in the three anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see method section).

Table S13. Hazard ratios (HRs) for incident venous thromboembolism (VTE) adjusted for either age, sex, and ancestry† or multivariable HRs adjusted for age, sex, body mass index, smoking, high alcohol consumption, and ancestry†. Incidence rates (IR) and incidence rate ratios (IRR) are also presented. Prevalent cases of VTE were excluded. First degree, second degree, third degree, and fourth degree relatives were excluded (i.e. up to 2nd degree cousins in the cohort).

	Participants	VTE	Age at VTE event	Crude IR	Crude IRR	Age- and sex-adjusted HR		Multivariable HR	
	N	N	Years (SD)	IR (95% CI)	IRR (95%CI)	HR (95% CI)	P-value	HR (95% CI)	P-value
Complete cohort	24681	2192	73.9 (8.5)	4.3 (4.2-4.5)	1	-	-	-	-
Model with factor V Leiden (rs6025)									
Reference no rs6025	21945	1796	74.2 (8.5)	4.0 (3.8-4.2)	1	1		1	
rs6025 heterozygotes	2653	372	72.6 (8.4)	7.0 (6.3-7.8)	1.8 (1.6-2.0)	1.8 (1.6-2.0)	9.2e-24	1.8 (1.6-2.0)	2.2e-24
rs6025 homozygotes	83	24	69.7 (8.9)	15.4 (9.9-22.9)	3.9 (2.6-5.8)	4.2 (2.8-6.3)	2.8e-12	4.4 (3.0-6.6)	5.1e-13
Model with prothrombin variant (rs1799963)									
Reference no rs1799963	24207	2126	74.0 (8.6)	4.3 (4.1-4.5)	1	1		1	
rs1799963 heterozygotes	472	65	71.9 (8.0)	7.0 (5.4-8.9)	1.6 (1.3-2.1)	1.7 (1.3-2.1)	5.4e-05	1.6 (1.3-2.1)	8.6e-05
rs1799963 homozygotes	2	1	69.8 (NA)	33.0 (0.84-184)	7.7 (1.1-54.6)	8.7 (1.2-62.4)	0.03	11.5 (1.6-82.1)	0.01
Model with non-synonymous variants in SERPINC1, PROC, and PROS1 genes									

Reference no non-synonymous variants	23521	2055	73.9 (8.6)	4.3 (4.1-4.5)	1	1		1	
1 low-risk variant	415	35	72.4 (7.0)	4.3 (3.0-6.0)	1.1 (0.8-1.5)	1.0 (0.8-1.4)	0.99	1.0 (0.7-1.4)	0.98
1 high-risk variant	741	98	74.0 (8.3)	6.6 (5.3-8.0)	1.7 (1.4-2.1)	1.5 (1.3-1.9)	2.2e-05	1.6 (1.3-1.9)	1.8e-05
≥ 2 high-risk variants	4	1	71.3 (NA)	14.9 (0.4-83.1)	3.8 (0.5-27.3)	3.6 (0.5-25.7)	0.20	3.5 (0.5-24.8)	0.21
Model with all five classical thrombophilia variants: rs6025, rs1799963, and high risk variants in SERPINC1, PROC, and PROS1 genes									
No thrombophilia*	20845	1665	74.3 (8.6)	3.9 (3.7-4.1)	1	1		1	
1 thrombophilia variant	3620	467	72.8 (8.2)	6.4 (5.9-7.1)	1.7 (1.5-1.8)	1.7 (1.5-1.8)	1.6e-22	1.7 (1.5-1.9)	4.3e-23
2 or more thrombophilia variants	216	60	71.3 (9.1)	14.9 (11.4-19.2)	3.8 (3.0-5.0)	4.1 (3.1-5.3)	1.6e-26	4.1 (3.1-5.3)	1.4e-26

High risk variant = non-synonymous variants in the three anticoagulant genes SERPINC1, PROC, and PROS1 (see method section).
*No rs6025 allele, no rs1799963 allele and no high-risk variant.
†Ancestry was controlled for by including the top two eigenvectors from the PCA analysis as covariates in Cox proportional hazard regression models.

Table S14. Hazard ratios (HRs) for incident venous thromboembolism (VTE) adjusted for either age, sex, and ancestry* or multivariable HRs adjusted for age, sex, body mass index, smoking, high alcohol consumption, and ancestry*. Incidence rates (IR) and incidence rate ratios (IRR) are also presented. Prevalent cases of VTE were excluded.

	Participants	VTE	Age at VTE event	Crude IR	Crude IRR	Age- and sex-adjusted HR		Multivariable HR	
	N	N	Years (SD)	IR (95% CI)	IRR (95%CI)	HR (95% CI)	P-value	HR (95% CI)	P-value
Reference No rs6025 + HGMD	28699	2562	73.9 (8.6)	4.4 (4.2-4.6)	1	1		1	
rs6025 + HGMD	95	22	73.2 (8.6)	12.1 (7.6-18.4)	2.8 (1.8-4.2)	2.6 (1.8-3.8)	3.8e-06	2.6 (1.8-3.7)	6.8e-06

*Ancestry was controlled for by including the top two eigenvectors from the PCA analysis as covariates in Cox proportional hazard regression models.

Table S15. Hazard ratios (HRs) for incident venous thromboembolism (VTE) adjusted for either age, sex, and ancestry*, or multivariable HRs adjusted for age, sex, body mass index, smoking, high alcohol consumption, and ancestry*. Incidence rates (IR) and incidence rate ratios (IRR) are also presented. Prevalent cases of VTE were excluded.

	Participants	VTE	Age at VTE event	Crude IR	Crude IRR	Age- and sex-adjusted HR		Multivariable HR	
	N	N	Years (SD)	IR (95% CI)	IRR (95%CI)	HR (95% CI)	P-value	HR (95% CI)	P-value
Reference no rs6025 + low-risk variant†	28759	2578	73.9 (8.6)	4.4 (4.2-4.6)	1	1		1	
rs6025 + low-risk variant†	35	6	75.2 (5.2)	8.5 (3.1-18.4)	1.9 (0.9-4.3)	2.0 (0.9-4.5)	0.081	2.0 (0.9-4.5)	0.088
Model with prothrombin variant (rs1799963)									
Reference no rs1799963 + low-risk variant†	28791	2583	73.9 (8.6)	4.4 (4.2-4.6)	1	1		1	
rs1799963 + low-risk variant†	3	1	66.6 (NA)	15.8 (0.4-87.9)	3.6 (0.5-25.5)	3.5 (0.5-25.0)	0.21	3.3 (0.5-23.4)	0.23

*Ancestry was controlled for by including the top two eigenvectors from the PCA analysis as covariates in Cox proportional hazard regression models.

†Low-risk variant = Low-risk non-synonymous non-HGMD variants in the three anticoagulant genes *SERPINC1*, *PROC*, and *PROS1* (see method section).

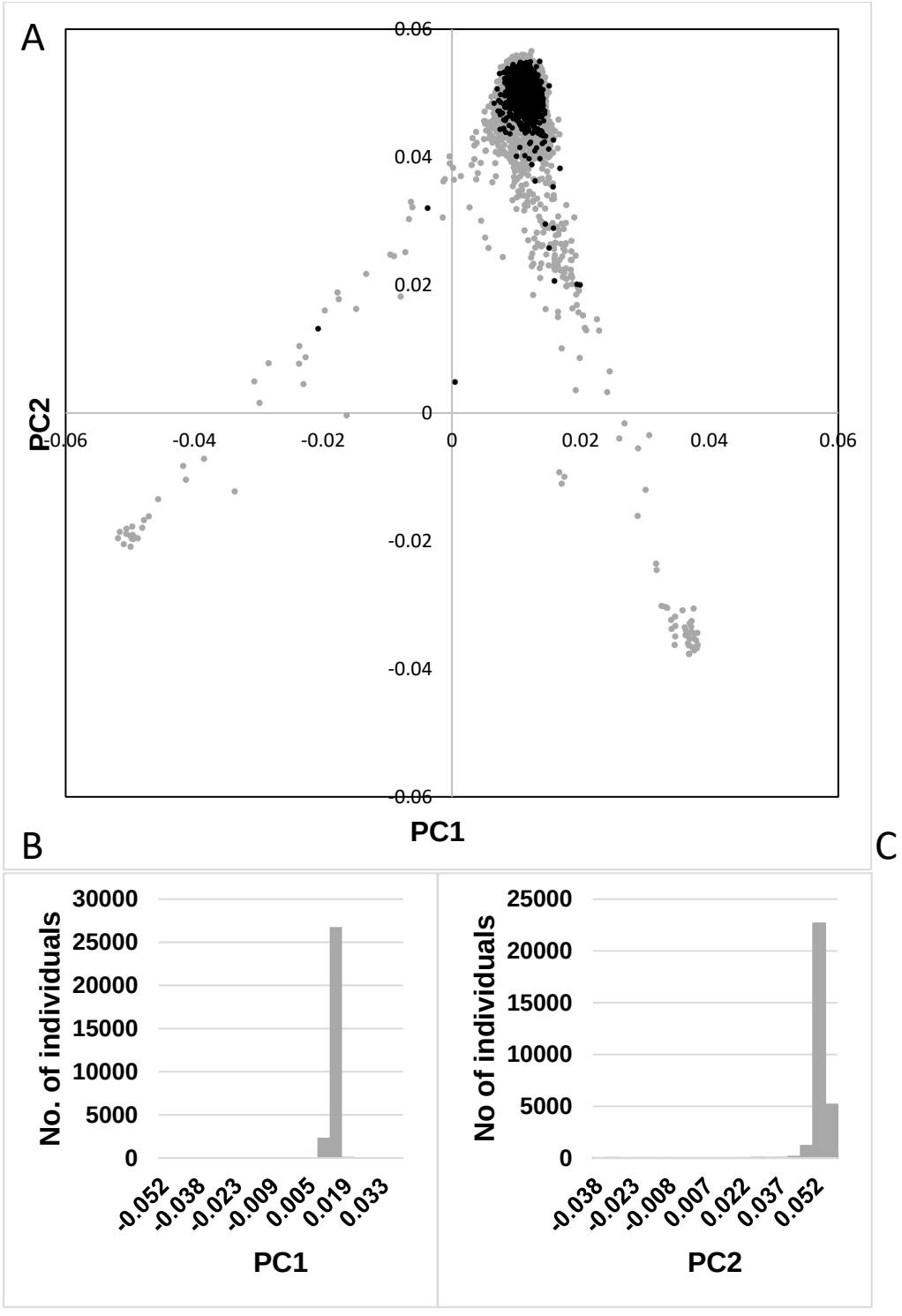

Figure S1. Principal-component analysis (PCA) was performed as described to show the population structure of the Malmö diet and cancer (MDC) cohort. The reference genomes were obtained from 1000 Genomes Project server. The principal components were first obtained from the reference genomes and then projected on individuals from MDC. A) The distribution of MDC individuals displayed on the two most informative principal components in 1000 Genomes, grey dots indicate individuals without a VTE event, black dots individuals with a VTE event. B) Histogram of number of individuals along PC1. C) Histograms of number of individuals along PC2

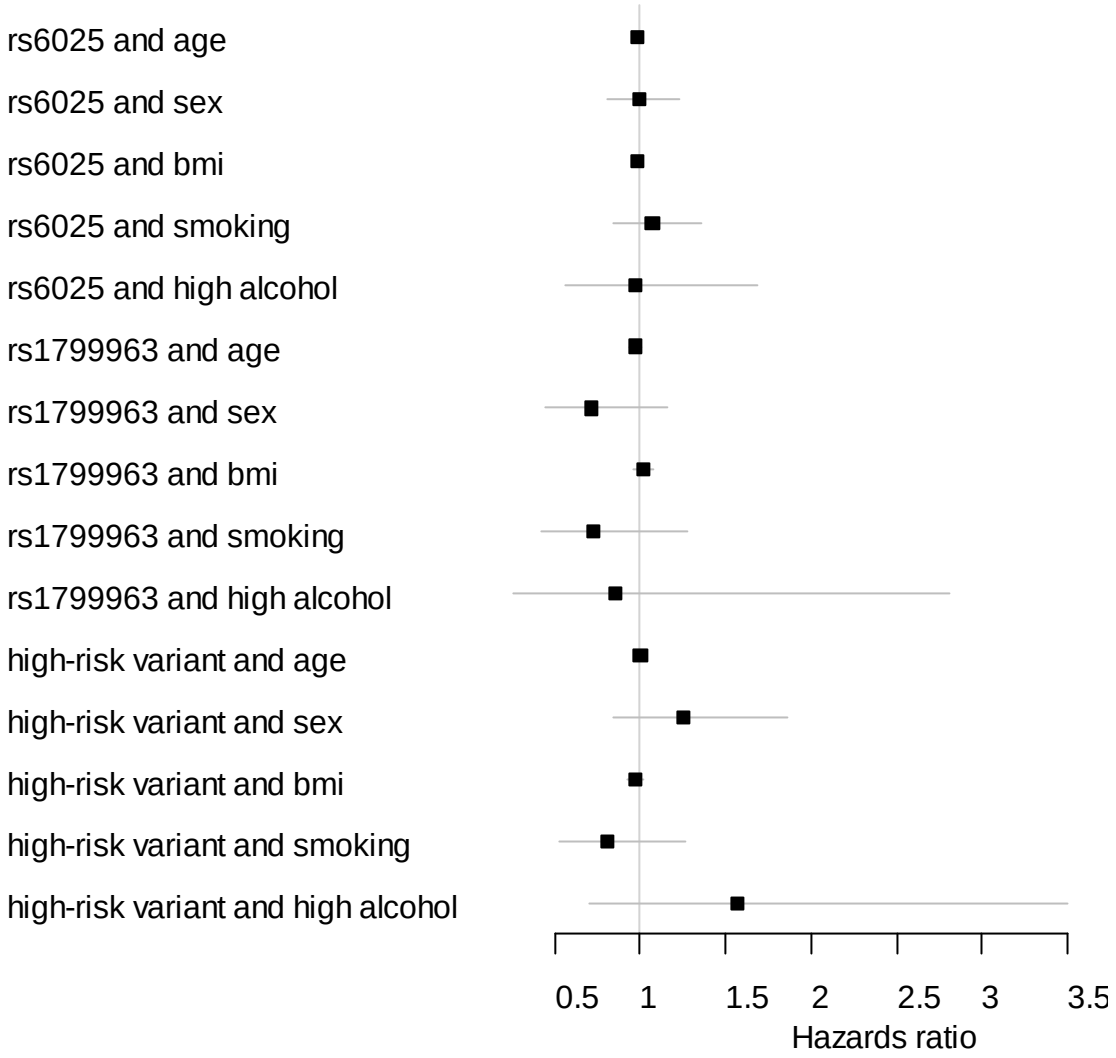

Figure S2. Forest plot showing hazard ratios over interactions between genetic effects and co-variates. No significant interactions were detected.