

From the Department of Medicine Solna
Karolinska Institutet, Stockholm, Sweden

**EXPLORATION OF
NOVEL MECHANISMS AND BIOMARKERS FOR
VENOUS THROMBOEMBOLISM**

- A GENETICS AND PROTEOMICS STUDY

Maria Bruzelius



**Karolinska
Institutet**

Stockholm 2015

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by US-AB 2015

© Maria Bruzelius, 2015

ISBN 978-91-7676-089-5

**Exploration of novel mechanisms and biomarkers for
venous thromboembolism
- a genetics and proteomics study**

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Maria Bruzelius

Principal supervisor:

Professor Jacob Odeberg
Royal Institute of Technology (KTH)
School of Biotechnology
Science for Life Laboratory Stockholm
Division of Proteomics

Karolinska Institutet
Department of Medicine Solna
Cardiovascular Medicine Unit

Co-supervisors:

Professor Anders Hamsten
Karolinska Institutet
Department of Medicine Solna
Cardiovascular Medicine Unit

Doctor Rona J. Strawbridge
Karolinska Institutet
Department of Medicine Solna
Cardiovascular Medicine Unit

Opponent:

Professor John-Bjarne Hansen
University of Tromsø, Norway
Department of Clinical Medicine
Hematological Research Group

Examination Board:

Professor Eva Hellström-Lindberg
Karolinska Institutet
Department of Medicine Huddinge, H7

Docent Masood Kamali-Moghaddam
Uppsala University
Department of Immunology, Genetics and
Pathology

Docent Pelle Lindqvist
Karolinska Institutet
Department of Clinical Science,
Intervention and Technology, H9

*“There are two ways to be fooled.
One is to believe what isn't true,
the other is to refuse to accept what is true”*

Søren Kierkegaard, Danish philosopher (1813-1855)

To Patrik, Petter and Axel

ABSTRACT

Background: Venous thromboembolism (VTE) contributes to a large health burden and incidence increases exponentially with age. ~25% patients will experience a recurrent event and there is a 2-fold risk for death in the following years. Risk prediction remains a challenge.

Aims to: Investigate a presumed overlap between cardiovascular disease and VTE. Expand current knowledge of established pathways in VTE risk by combining clinical and genetic epidemiology. Apply affinity proteomics to identify novel plasma susceptibility biomarkers.

Methods and results: 39 single nucleotide polymorphisms (SNPs) associated with coronary artery disease (CAD) and 18 risk VTE-SNPs from candidate gene approach studies or genome-wide association studies (GWAS) were identified from a literature search. The SNPs were genotyped in 2,835 women from the ThromboEmbolism Hormone study (TEHS), a Swedish nationwide case-control study in women (2002-2009). Association was assessed with logistic regression. Clinical and genetic predictors that contributed significantly to the fit of the logistic regression model were included in the prediction models. The genetic predictors that contributed to first VTE were assessed in a cohort of 1010 women with VTE from TEHS, followed up until a recurrence, death or November 2011 (TEHS follow-up). The SNP r579459 in the *ABO* locus was the only CAD-SNP that was significant associated with VTE (OR 1.57 (95% CI: 1.39-1.78, $p= 6.4 \cdot 10^{-13}$)). Seven VTE risk-SNPs (*F5* rs6025, *F2* rs1799963, *ABO* rs514659, *FGG* rs2066865, *F11* rs2289252, *PROC* rs1799810 and *KNG1* rs710446) with 4 SNP-SNP interactions contributed to the genetic risk score for VTE with an AUC of 0.66 (95% CI: 0.64-0.68). After adding clinical risk factors the AUC attained 0.84 (95% CI: 0.82-0.85). The goodness of fit of the genetic and combined scores improved when significant SNP-SNP interaction terms were included. In TEHS follow-up study, the overall recurrence rate was 20 per 1000 person-years (95% CI; 16-24). Carriers of the risk alleles of *F5* rs6025 (FVL) (HR=1.7 (95% CI; 1.1-2.6)) and *F11* rs2289252 (HR=1.8 (95% CI; 1.1-3.0)) had significantly higher rates of recurrence compared to non-carriers. The cumulative recurrence was 2.5-fold larger in carriers of both *F5* rs6025 and *F11* rs2289252 than in non-carriers at 5 years follow-up.

Plasma samples from 88 VTE cases and 85 controls, part of the Venous thromboembolism Biomarker Study (VEBIOS), were screened against 408 candidate proteins targeted by 755 antibodies using multiplex bead arrays. Proteins that significantly associated with VTE after Bonferroni correction were tested for replication in plasma samples of 580 cases and 589 controls from the FARIVE study. In VEBIOS, plasma levels of four proteins, HIVEP1, VWF, GPX3 and PDGFB were significantly associated with VTE after Bonferroni correction. VWF and PDGFB successfully replicated in FARIVE with increased plasma levels in VTE cases compared to controls as initially observed in VEBIOS.

Conclusion: *ABO* locus was the only shared genetic risk factor between CVD and VTE. A limited set of genetic predictors improved prediction of incident VTE in women at high risk. Our data indicated that interactions among SNPs increase the goodness of fit when predicting VTE. A combination of genotypes i.e. *F11* rs2289252 and FVL, may be of potential clinical relevance for risk prediction for recurrence. Affinity plasma proteomics proved to be a valuable research strategy to discover novel biomarkers.

LIST OF SCIENTIFIC PAPERS

- I. **Bruzelius M**, Strawbridge RJ, Trégouët DA, Wiggins KL, Gertow K, Sabater-Lleal M, Ohrvik J, Bergendal A, Silveira A, Sundstrom A, Kieler H, Syvanen AC, Smith NL, Morange PE, Odeberg J, Hamsten A. Influence of coronary artery disease-associated genetic variants on risk of venous thromboembolism. *Thrombosis research*. 2014;134(2):426-32.

- II. **Bruzelius M**, Bottai M, Sabater-Lleal M, Strawbridge RJ, Bergendal A, Silveira A, Sundstrom A, Kieler H, Hamsten A, Odeberg J. Predicting venous thrombosis in women using a combination of genetic markers and clinical risk factors. *Journal of thrombosis and haemostasis*. 2015;13:219-27.

- III. **Bruzelius M**, Ljungqvist M, Bottai M, Bergendal A, Strawbridge RJ, Holmstrom M, Silveira A, Kieler H, Hamsten A, Larfars G, Odeberg J. *F11* is associated with recurrent VTE in women: A prospective cohort study. *Thrombosis and Haemostasis*. 2015 Oct 1;115(2). Epub ahead of print.

- IV. **Bruzelius M**, Iglesias MJ, Hong MG, Souto JC, Franberg M, Holmstrom M, Strawbridge RJ, Hamsten A, Uhlén M, Silveira A, Soria JM, Smadja DM, Butler L, Schwenk JM, Trégouët DA, Morange PE, Odeberg J. Identification of novel plasma biomarkers for venous thromboembolism: the VEREMA affinity proteomics study. *Manuscript*. 2015.

TABLE OF CONTENTS

1	INTRODUCTION.....	7
2	BACKGROUND.....	8
2.1	Epidemiology.....	8
2.2	Etiology.....	8
2.2.1	Heritability.....	9
2.2.2	Gender	11
2.2.3	Cardiovascular disease and VTE.....	12
2.3	Identification of plasma biomarkers	12
2.3.1	Definition of a biomarker.....	12
2.3.2	Plasma biomarkers in clinical practice for VTE management	12
2.3.3	Technology platforms for Biomarker Discovery studies.....	13
3	AIMS	15
4	MATERIAL AND METHODS.....	16
4.1	Studies.....	16
4.1.1	The ThromboEmbolism Hormone Study (TEHS).....	16
4.1.2	TEHS follow-up study	16
4.1.3	Venous thromboembolism biomarker study (VEBIOS).....	16
4.1.4	FARIVE.....	17
4.2	Ethics permission.....	17
4.3	Clinical variables	18
4.4	Genotyping	18
4.5	Affinity proteomics	18
4.5.1	Antibody reagents	18
4.5.2	Methodology	19
4.6	Statistical methods.....	20
5	RESULTS and DISCUSSION.....	22
5.1	PAPER I.....	22
5.1.1	Results	22
5.1.2	Discussion	23
5.2	PAPER II	25
5.2.1	Results	25
5.2.2	Discussion	27
5.3	PAPER III.....	29
5.3.1	Results	29
5.3.2	Discussion	31
5.4	PAPER IV.....	33
5.4.1	Results	33
5.4.2	Discussion	35
6	CONCLUSION	38
7	FUTURE PERSPECTIVES.....	39
8	ACKNOWLEDGEMENTS.....	40

9	REFERENCES.....	43
---	-----------------	----

LIST OF ABBREVIATIONS

APC	Activated Protein C
AUC	Area under the receiver-operating characteristics curve
BMI	Body mass index (kg/m ²)
CAD	Coronary artery disease
CI	Confidence interval
CVD	Cardiovascular disease
DAG	Direct acyclic graph
DNA	Deoxyribonucleic acid – the genetic code
DVT	Deep vein thrombosis
FVL	Factor V Leiden, <i>F5</i> rs6025
<i>F2</i> (G20210A)	Polymorphism in prothrombin gene, <i>F2</i> rs1799963
FVIII	Coagulation factor VIII
FXI	Coagulation factor XI
ELISA	Enzyme-Linked Immunosorbent Assay
GPX3	Glutathione Peroxidase 3
GWAS	Genome-wide association study
HIVEP1	Human Immuno-deficiency Virus type I Enhancer binding Protein 1
HPA	Human Proteome Atlas
HR	Hazard Ratio
LD	Linkage Disequilibrium
MEGA	Multiple Environmental and Genetic Assessment of Risk Factors for Venous Thrombosis study
MS	Mass Spectrometric analysis
OR	Odds ratio
PDGFB	Platelet-derived growth factor beta
PE	Pulmonary embolism
SNP	Single nucleotide polymorphism
TEHS	Thromboembolism Hormone Study
VEBIOS	Venous thromboembolism Biomarker Study
VTE	Venous thromboembolism
VWF	Von Willebrand factor

1 INTRODUCTION

Venous thromboembolism (VTE) comprises of deep vein thrombosis (DVT) and pulmonary embolism (PE) and is the third most common cardiovascular disease (CVD) after myocardial infarction and stroke. Even if VTE is responsible for a large global public health burden, the public awareness is strikingly low for this potentially fatal but preventable disease (1-3).

About 160 years ago in Berlin, Rudolf Virchow set the stage for modern ideas for VTE. He found evidence that blood clots (thrombi) form in the veins and could move from their site (emboli) to the lungs. He proposed three principal causes of clot formation; injury to the vessel wall, impaired blood flow (stasis) and hypercoagulability. This concept, known as the Virchow's triad still holds today (4).

Several milestones have been reached in understanding of the complex interactions between genetic, environmental and acquired risk factors for this disease (5). New effective anticoagulant treatments have been introduced during the first decade of the 21st century, though still with an increased risk of bleeding complications as drawback (6).

The development of high throughput technologies for profiling of protein composition in plasma allows for exploration of the proteome in plasma samples of large cohorts and the discovery of new tentative biomarkers for VTE (7). Risk prediction remains a challenge for both first VTE and recurrence, thus identification of biomarkers specific for VTE could open up for new possibilities in clinical management as well as provide leads to explore the pathophysiology of VTE (8).

The focus of this thesis is on the application of genetic and proteomic approaches in an attempt to gain a further understanding of the complexity of VTE to potentially contribute to improved risk prediction.

2 BACKGROUND

2.1 EPIDEMIOLOGY

Venous thromboembolism (VTE) has an incidence of 1-2/1000 person years that exponentially increases with age and is higher among Caucasians compared to Blacks and Asians (2, 9, 10). Around 25% of the patients will suffer from a recurrent VTE within 5 years after discontinuation of anticoagulant treatment with the highest recurrence rate in the first half year after diagnosis and in the younger population with unprovoked VTE (9, 11, 12). Around 10 to 20 per cent will face an impaired quality of life after a first DVT because of post thrombotic syndrome in the lower limbs that leads to chronic swollenness and ulceration (2). In a Norwegian population cohort the case-fatality rate was 13% within the first year from diagnosis (10), and in the follow-up of the Dutch ‘Multiple Environmental and Genetic Assessment of Risk Factors for Venous Thrombosis study’ (MEGA) an overall shorter life expectancy with 5 years and a two-fold increased risk for death up to 8 years was reported (13).

2.2 ETIOLOGY

VTE is multifactorial disease with interacting genetic, acquired and environmental risk factors (14). According to Lijfering *et al.* description in the Figure 1, all individuals hold a risk to develop a VTE, a ‘thrombosis potential’ that increases with age. Whether an individual reaches the ‘thrombosis threshold’ (diagnosed with VTE) is decided by several interacting factors (15). An individual’s ‘thrombosis potential’ is influenced of more persistent risk factors over time i.e. genetic and acquired risk factors (Figure 1). The latter defined as obesity, malignancy, rheumatic systemic disorders and other chronic conditions. Environmental risk factors are often regarded as transient factors but crucial in causing an individual to reach above the ‘thrombosis threshold’ (Figure 1). Among the environmental risk factors are surgery, immobilisation (i.e. cast or orthosis, hospitalisation), long distance flight, pregnancy and the exposure to oestrogens from combined contraceptives and hormonal replacement therapy. Even if the environmental risk factors have a large contribution to risk, no such provoking factors can be identified in one third of the patients diagnosed with incident VTE (5, 15).

An example of how risk factors may interact is through the casual pathways of activated protein C (APC)-resistance. High body mass index (BMI) leads to Factor VIII (FVIII)-related increased APC-resistance and hypercoagulability, an effect that is amplified in carriers of

Factor V Leiden (FVL) and ABO blood group non-O (16). These interacting factors influence the ‘thrombosis potential’ in an individual (Figure 1)

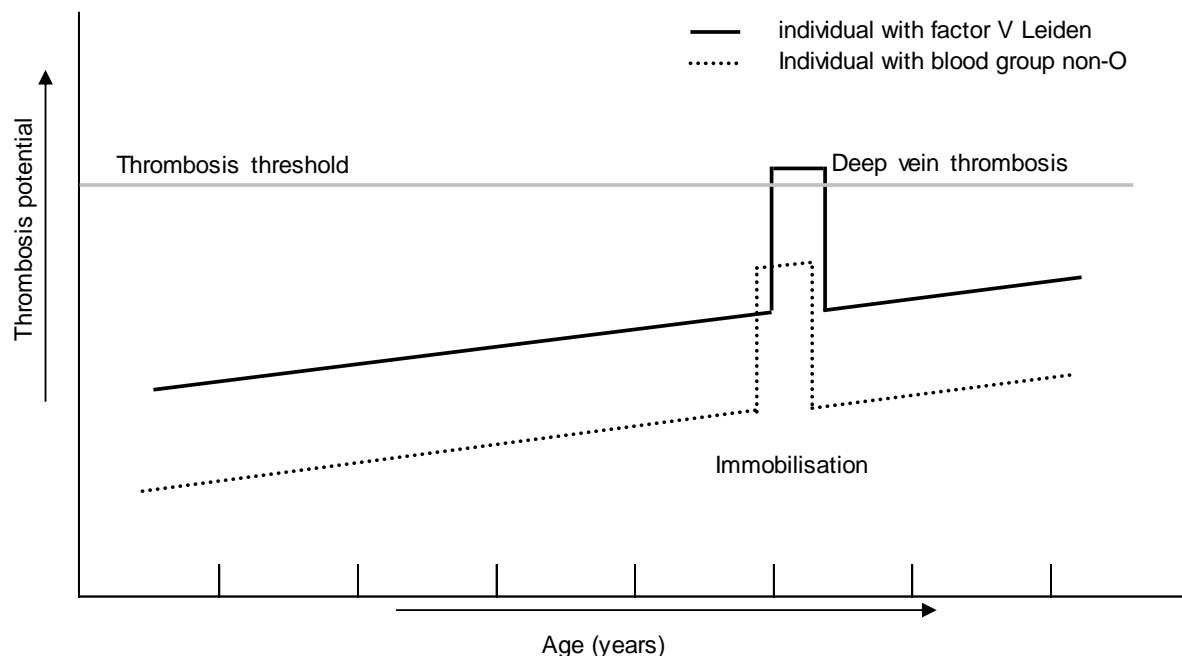


Figure 1 shows the thrombosis potential of an individual with factor V Leiden and of an individual with blood group non-O. Both experience a short period of immobilization at a similar age. This leads to an excess of the thrombosis threshold level in the individual with factor V Leiden which consequently leads to deep vein thrombosis, while the individual with blood group non-O does not develop thrombosis. *Br J Haematol.* 2010;149:824-33. Figure is reused with the kind permission from W.M. Lijfering

2.2.1 Heritability

In both case-control studies and nation-wide family studies, VTE has been observed to cluster in families (17-20), suggesting that there is an important underlying genetic component. Family and twin studies of thrombophilia have estimated the heritability for VTE to be approximately 60 % (21-23). However, from the French genome wide association study (GWAS) on unrelated individuals, Germain *et al* reported that common genetic variants explain about 35% of the genetic variance underlying VTE susceptibility, whereas only ~5% could be attributed to established VTE loci (24).

2.2.1.1 Family history of VTE

Family history of VTE is a well-established risk factor for incident VTE, where the risk is more pronounced in families where more than one relative have been affected with VTE and if the thrombosis was unprovoked and/or occurred at a younger age (25, 26). Family history

of VTE was only modestly associated with recurrence according Zöller *et al.* in the Swedish Multigeneration Study (27).

2.2.1.2 Genetic risk factors

Known strong genetic factors for VTE include deficiency of inhibitors in the coagulation system; Antithrombin, Protein C and its co-factor Protein S (28-30). These risk factors are very rare in the population, but highly penetrant in some families with a diversity of different mutations described (14). Among those with moderate risk, the polymorphisms in coagulation Factor V, FVL (~4 fold risk) and the prothrombin (*F2 G20210A*) (~2 fold risk) are common in the Caucasian population, ~7-10% and ~3-5%, respectively (Figure 2) (14, 15, 31, 32). FVL cause an impaired activated Protein C inhibition of thrombin generation, i.e. APC resistance (31). *G20210A* increases synthesis of Prothrombin, and thereby also thrombin generation (32). Carriership of the genetic variant for ABO blood group non-O is not a part of inherited thrombophilia screening in clinical practice. Still it is a very common (~40%) and well-established genetic risk factor for VTE (~1.5-2 fold risk), explained largely by higher levels of Factor VIII (FVIII) and von Willebrand factor (VWF) (Figure 2) (33-35). According to Heit *et al.* the risk genotypes for FVL, *F2 G20210A* and ABO blood group non-O, contribute to ~40% risk of the incident VTE (35).

In clinical practice, current tests for inherited thrombophilia after first VTE do not reduce the incidence of recurrent events (36, 37). This is also reflected by the fact that in three different prediction scores for recurrence, HER DOO2 (HER for women; hyperpigmentation, oedema, or redness; and DOO2; D dimer, Obesity, Old age and 2 is too (also) continue anticoagulant treatment), Vienna Prediction model and DASH (D dimer, Age, Sex (male) and Hormonal therapy) neither FVL or the *G20210A* polymorphism in Prothrombin gene alone, nor inherited thrombophilia, are important predictors (38-40).

Through GWAS, new susceptibility genetic variants/loci have emerged for VTE (Figure 2). These are described as single nucleotide polymorphisms (SNPs) and are common in the general population (minor allele frequency larger than 1 %). The corresponding risk estimates are relative low (less than 1.5 fold) compared to previous detected genetic variants (Figure 2) (41). Most of these new variants are associated with intermediate traits for VTE, i.e. *STXBP* is associated with VWF level and *KNG1* is associated with activated partial thromboplastin time and Factor XI (FXI) level (42, 43). However, a recent meta-analysis with ~66,000 individuals by Germain *et al.* has identified two new risk VTE-loci; *TSPAN15* and *SLC44A2*,

for which the functional link to thrombosis formation is still not established (44). It remains an open question how to best translate this new genetic information into clinical practice.

With the exception from van Hylckama Vlieg *et al.*'s recent publication, there are few studies on the association between common genetic variants and recurrent event of VTE (45).

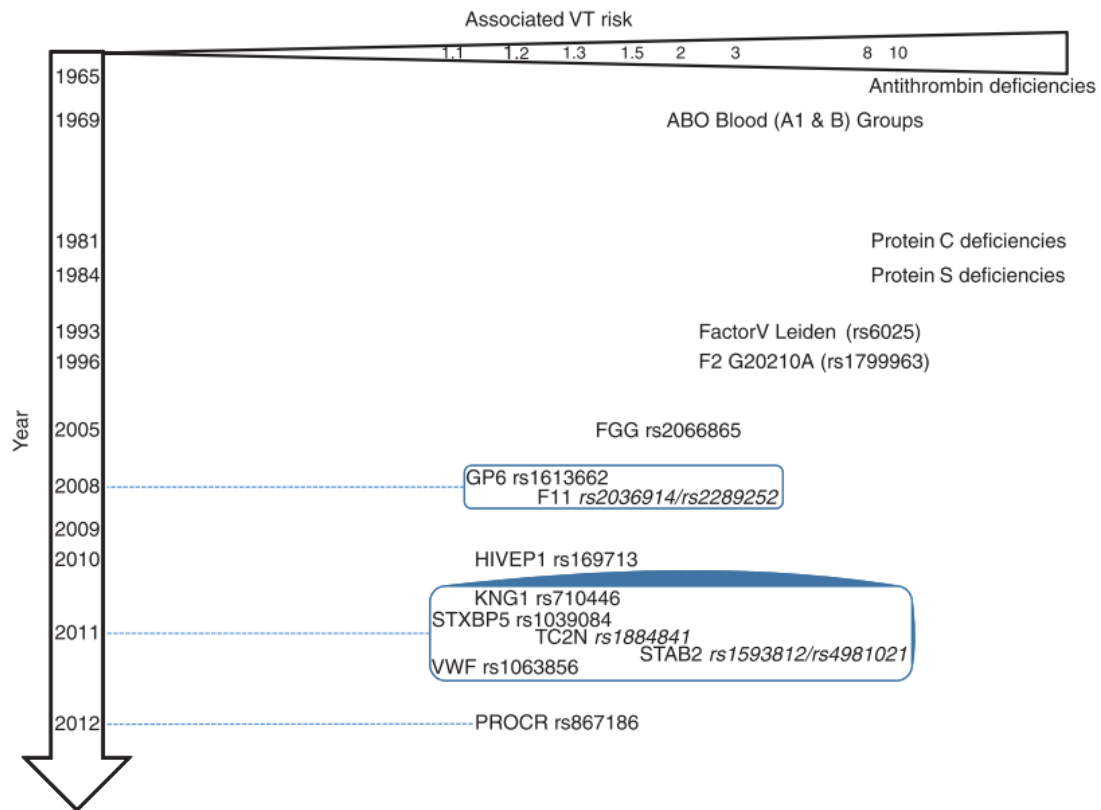


Figure 2: Historical perspectives of known genetic risk factors for VTE. *J Thromb Haemost.* 2011; 11 Suppl. 1:111-2. Figure reused with the kind permission from P.E Morange.

2.2.2 Gender

The incidence for VTE is reported to be higher in women during child-bearing age attributed to pregnancy and the influence of oestrogen containing contraceptives. However, when female specific risk factors are taken into account, the risk is twofold higher for incident VTE in men compared to women in all age groups and pronounced in the youngest group (46). This also holds for recurrent VTE, making it plausible that there is a sex-specific difference in genetic risk (9, 12, 47).

2.2.3 Cardiovascular disease and VTE

There is an increased risk of CVD amongst patients previously diagnosed with VTE suggesting that VTE is a predictor of subsequent arterial thrombosis (13, 48, 49). In a Norwegian population cohort, the Tromsø study, 64% of arterial thrombotic events among the VTE patients were attributable to the incident VTE (49). However, the relationship between VTE and subsequent CVD is more likely to be explained by common etiological factors than to be causal according to more recent study (50). Increasing age and obesity are the only cardiovascular risk factors with robust influences on VTE risk in prospective population studies, whereas for smoking, hyperlipidaemia, hypertension and diabetes mellitus the observed associations are inconsistent (51-53). Furthermore, arterial and venous thrombosis share common mechanisms i.e. activation of blood coagulation (54-56), hypofibrinolysis (57, 58) and inflammation (59).

2.3 IDENTIFICATION OF PLASMA BIOMARKERS

2.3.1 Definition of a biomarker

According to the National Institutes of Health Biomarkers Definitions Working group from 1998, a biomarker is defined as; ‘A characteristics that is objectively measured and evaluated as an indicator of normal biological process, pathological process, or pharmacological response to a therapeutic intervention’ (60).

Biomarkers are usefully categorised depending on intended use to; a) diagnostic, which incorporates detection of disease, b) prognostic, that predict the course of a disease (i.e. recurrence, progression and survival) and c) predictive, that allows for prediction of the response of a treatment (7, 60).

2.3.2 Plasma biomarkers in clinical practice for VTE management

2.3.2.1 D-dimer

The only plasma biomarker for VTE that is in clinical use for diagnosis or risk prediction is D-dimer, a split product from the cross-linked fibrin clot. Elevated D-dimer is observed in increased thrombin generation as a result of hypercoagulability and associated with both first and recurrent VTE (61). In Well’s diagnostic score, D-dimer has a high negative prediction value to rule out diagnosis for those with low suspicion of acute thrombosis (62). However, D-dimer has very low specificity and furthermore it is increased in many other conditions apart from VTE, i.e. cancer, pregnancy, inflammation and older age. A biomarker with high positive predictive value for VTE, equivalent to troponin for acute myocardial injury, i.e. myocardial infarction, is still missing.

Risk scores to predict recurrence of VTE that incorporate D-dimer level together with clinical data have been proposed (38-40). Still, these models does not satisfactory discriminate those who will have a recurrent VTE from those who will not. Furthermore, the value of a negative D-dimer test, sampled one month after stopped anticoagulant treatment of an unprovoked VTE, as an indicator of a low recurrence risk has been challenged (63, 64). A recent cohort study reported that even if a D-dimer test was negative twice (during anticoagulant treatment and one month after stopped treatment), the risk for recurrence was not low enough in men to discontinue anticoagulant treatment according to clinical practice (64).

2.3.2.2 *Factor VIII*

High FVIII levels in plasma is a well-established risk factor for first VTE (65). For recurrence, FVIII levels have currently been found to have a dose-dependent association with risk in the large MEGA follow-up study (66). Subjects within the highest category of FVIII levels (> 200 IU/dL) had a threefold risk for recurrence of VTE compared to those in the lowest range (\leq 100 IU/dL). The performance of the risk prediction score for recurrence DASH (D-dimer level, Age, Sex and Hormone use), improved with FVIII, even when D-dimer was left out from the score (DASH AUC=0.64 and 'FASH' AUC=0.67, respectively) (40, 66). It remains to be further evaluated if FVIII level can be implemented in clinical practice.

There is a need to improve both the precision and the discrimination power in the prediction scores for recurrence to better identify high risk individuals in whom the benefits of long-term anticoagulant treatment overrule the yearly risk of serious bleeding associated with this treatment (8, 67).

2.3.3 **Technology platforms for Biomarker Discovery studies**

The main high throughput technologies for discovery of putative plasma biomarkers are based on mass spectrometric analysis (MS) and affinity proteomics. For affinity proteomics there are different platforms for planar arrays or bead arrays that can be used for multiplexed protein profiling in plasma (68). These arrays are built with single-capture antibodies (single binders), sandwich immunoassays (dual binders) or antigen- or peptide-capture arrays (69, 70).

2.3.3.1 *'Shotgun' (untargeted) Mass spectrometry*

Shotgun mass spectrometry is based on the conversion of intact proteins in a sample into peptide fragments. These peptides are identified by their mass-to charge (m/z ratio), and

thereafter the protein is identified by studying a database of peptides. Experiments involve the comparison of results from two small groups of samples with identification of many intensity peaks that represent peptide levels differences (69, 71). The method has the potential to detect up to 5,000 different proteins, but with limitations in the number of samples that can be analysed in parallel, and thereby a high false discovery is expected (71). Extensive sample preparation is frequently necessary i.e. protein depletion or purification of samples. Many proteins are labile, and degradation during sample preparation or storage can impair proper protein identification (7, 69, 71).

2.3.3.2 Affinity Proteomics based on suspension bead arrays

Affinity Proteomics with suspension bead arrays has successfully been used to identify biomarkers in various diseases, i.e. gastrointestinal cancer, osteoporosis and amyotrophic lateral sclerosis (68, 72, 73). These suspension bead arrays are built with single binders such as antibodies from the Human Proteome Atlas (HPA) (74). According to the latest release of the Atlas (version 13), 24,000 polyclonal antibodies are available corresponding to 17,000 protein-coding genes in human (74, 75). With this multiplexed technology platform, it is possible to generate protein profiles of 384 protein targets per array in 384 samples in parallel. This method has the advantage of not involving any sample purification, fractionation or depletion steps (68, 76, 77). However, there is a vast dynamic range of concentration of different proteins in plasma, why single binder arrays are vulnerable to potential off-target interactions, requiring further verification of identified tentative targets (77). Targeted MS is an important tool to further confirm candidate proteins and holds also the potential for quantification with dual binder assays (ELISA) (78, 79).

3 AIMS

The main objectives of this thesis were to:

- Expand current knowledge of established pathways of importance in VTE risk by combining clinical and genetic epidemiology
- Apply affinity proteomics to identify novel susceptibility plasma biomarkers that may improve prediction tools for VTE-risk and provide new leads to further understand mechanisms contributing to the pathogenesis of VTE

More specifically, the three first papers focused on investigating the genetics of VTE in Swedish young and middle-age women with the aim to:

- Investigate a presumed overlap between CVD and VTE
- Evaluate the contribution of potentially important predictors in a combined genetic and clinical risk score for first VTE
- Explore if important genetic predictors for first VTE also were associated with recurrence

The fourth paper was a proteomic biomarker discovery study aiming to identify novel plasma biomarkers for VTE risk.

4 MATERIAL AND METHODS

4.1 STUDIES

4.1.1 The ThromboEmbolism Hormone Study (TEHS)

TEHS is a nation-wide case-control study of VTE in women initiated by the Medical Products Agency (MPA) and conducted through the Centre for Pharmacoepidemiology at Karolinska Institutet. TEHS, previously well described by Bergendal *et al* (80), was designed to investigate how environmental and genetic factors affect the risk of VTE, with a focus on women taking different combined hormonal contraceptives or menopausal hormone therapy. A total of 1,433 cases and 1,402 controls, aged between 18 and 65, were recruited between 2003 and 2009. Cases included both out- and in-patients recruited from 43 hospitals in Sweden, who had a first-time episode of DVT in lower limbs and/or PE. All cases of VTE were objectively confirmed with established diagnostic imaging methods. Exclusion criteria were previous VTE, current or recent (< 5 years) malignancy and pregnancy. Female controls were selected randomly from the population register held by the National Board of Taxation and matched to birth year of the cases. Information was obtained through telephone interviews, precluding participation of non-Swedish-speaking women. All participants donated 5 mL of whole blood and the DNA was prepared using QIAGEN FlexiGene DNA kit.

4.1.2 TEHS follow-up study

TEHS follow-up study is a prospective cohort study on recurrent event from the cases in TEHS. The follow-up study aims to identify important risk factors for a recurrent event in young and middle aged women, previously described in detail by Ljungqvist *et al* (81). Briefly, in 2011, all cases included in TEHS who were still living in Sweden were followed up by a questionnaire. Information about recurrent events was obtained from the questionnaire and from data recorded in the Swedish Patient Register (82). All events in the register and/or stated in the questionnaires were verified by a review of medical records. The recurrent event had to be diagnosed by an objective radiological method and regarded by the treating physician as an indication for resuming anticoagulant treatment for at least three months. From the Swedish Cause of Death Register, data was obtained for the causes of death of deceased participants in TEHS (83).

4.1.3 Venous thromboembolism biomarker study (VEBIOS)

VEBIOS was initiated in 2010 at Centre for Hematology, Karolinska University Hospital in collaboration with Karolinska Institutet and Science for Life Laboratory, KTH Royal Institute

of Technology, Stockholm with the aim of identifying and translating biomarkers for VTE using resources available through the HPA (74). VEBIOS, in detailed described in Paper IV, is an on-going case-control study since January 2011 and conducted by the Coagulation Unit at Karolinska University Hospital with recruitment of cases from three regional hospitals. Eligible cases are patients from 18 to 70 years of age with first VTE, which have been referred from the Emergency clinics after diagnosis for a follow up at the respective outpatient clinics for Thrombosis and Haemostasis. Cases have been treated with anticoagulants during six to twelve months with any type of anticoagulant medication. Cases are identified by the physicians at the out clinics at respective hospital. Patients found with severe thrombophilia are not eligible. Controls are matched to cases for age and gender and continuously recruited from the population living in Stockholm County, based on a randomised selection using the Swedish Tax Agency register (84). Exclusion criteria for participants were personal history of VTE, pregnancy during the last three months before the index date or an active cancer within the last five years. Blood sampling is performed at the Coagulation unit by a research nurse and from each participant both plasma and whole blood are stored in a biobank. Plasma samples have been centrifuged at 2000g for 15 minutes at room temperature and aliquots were snap frozen and stored at -80°C . The cases have been sampled one to six months after discontinuation of anticoagulant.

4.1.4 FARIVE

The FARIVE study was used in Paper IV as a replication study to VEBIOS and is a French multicentre case-control study carried out between 2003-2009 with the aim to study the interaction of environmental, genetic and biologic risk factors for first VTE and risk of recurrence. The study has previously been well described and involves consecutive inpatients or outpatients from 18 years of age, confirmed by diagnostic imaging and treated for a first episode of DVT and/or PE (85). The controls were age- and sex matched and consisted of in- and outpatients, free of personal history of venous and arterial thrombotic disease. Exclusion criteria for all eligible participants were diagnosis of cancer, short life expectancy owing to other causes, renal - or liver failure. The replication study was based on a subset from FARIVE (86).

4.2 ETHICS PERMISSION

Informed written consent was obtained from all participants in accordance with the Declaration of Helsinki. The TEHS, TEHS follow-up study, VEBIOS and FARIVE were approved by the regional research ethics committees (KI 01-255), (KI 2010/1200-31/1), (KI 2010/636-31/4) and (Paris Broussais-HEGP 2002-034), respectively.

4.3 CLINICAL VARIABLES

Clinical data was obtained from TEHS, TEHS follow up study (Paper I-III) and FARIVE (Paper IV) (80, 81, 85). Clinical variables are described in detail in the respective publication and manuscript (80, 81, 85). The clinical variables used from TEHS in Paper I-III included localisation of thrombosis (PE and DVT), and self-reported information from questionnaires. This included weight, height, BMI, cardiovascular risk factors, immobilisation and hormonal use during the 3-month period preceding diagnosis for cases and index date for controls and family history of VTE.

In VEBIOS the information was obtained from a questionnaire asked to be filled in by all participants covering information about established risk factors for VTE and CVD and in detailed described in Paper IV. The biometry, i.e. lengths, weight and BMI was measured by the research nurse the day of blood sampling.

4.4 GENOTYPING

Genotyping was performed by using the Illumina GoldenGate platform, read and called by using the Illumina BeadXpress and Illumina GenomeStudio 2011.1 software at the SNP&SEQ Technology Platform, Uppsala, Sweden. SNPs in *F5* (rs6025), *F13A1* (rs5985), and *F2* (rs1799963) were genotyped by Pyrosequencing™ technology (ISO standard 2004) at the Royal Institute of Technology, Stockholm, Sweden (87). Complementary genotyping was done with Taqman® SNP genotyping assays (Applied Biosystems, CA, USA) and analysed using StepOne software v2.1. Standard quality control procedures were performed with SNP exclusion due to call rate (<95%) and/or deviation from Hardy-Weinberg equilibrium ($p < 0.005$). Subjects with a low call rate (<90%) were excluded. Genotyping was performed on plates containing cases and controls and was completed in one stage each for the Illumina and Taqman® platforms.

4.5 AFFINITY PROTEOMICS

4.5.1 Antibody reagents

Antibodies to the selected protein targets were obtained from the polyclonal antibody reagents resource generated by HPA (74). The HPA produces polyclonal antibodies towards recombinantly expressed human protein targets based on in-silico selection of antigen regions, so called protein epitope signature tags (PrEST), which consist of the most unique 50-150 amino acid regions of each protein as compared to all other human proteins (70). The antibody reagents selected for the plasma profiling in paper IV were checked according to following quality criteria; (a) concentration higher than 0.05 mg/mL and (b) clear indication

of the specific binding to the candidate target protein or protein fragment. The latter was based on confirmation of antibody binding by planar arrays with PrEST fragments, determination of antibody specificity with western blot and finally immunohistochemical or immunofluorescence stained tissue arrays that had been annotated by trained pathologists (74).

4.5.2 Methodology

4.5.2.1 Suspension Bead Array Generation

Protein plasma profiles were generated using suspension bead arrays in which HPA antibodies were coupled to color-coded magnetic beads as illustrated in Figure 3 and previously described (68, 77). The coupling efficiency of each antibody to its corresponding bead ID was confirmed by the incubation with a labelled anti-rabbit antibody and represented as a median of fluorescence intensity (MFI). Equal amounts of each bead ID were combined to generate the 384 antibody-plex suspension bead array (68, 76, 77).

4.5.2.2 Sample preparation

Thawed plasma samples were transferred to microtiter plates in a final format of 384 well plate, randomized according to pre-set criteria in which matching of cases and controls as well as confounders as age, gender and centre were taking in account. The samples in the well plates were then labeled with biotin and after dilution heat-treated for 30 minutes at 56 °C (68, 76, 77).

4.5.2.3 Protein Profiling

The suspension bead array was dispensed into a 384 well plate and the samples were added for overnight incubation, allowing for antibodies binding to the targeted proteins in plasma. After washing a detection solution of R-Phycoerythrin-labeled streptavidin (SAPE) was added and incubated for 30 minutes. SAPE enables the captured proteins to be measured in parallel as a relative level of immunocapture reactions on beads by fluorescence in MFI using assay laser detection FlexMap3D instrument (Luminex) (Figure 3) (68, 76, 77).

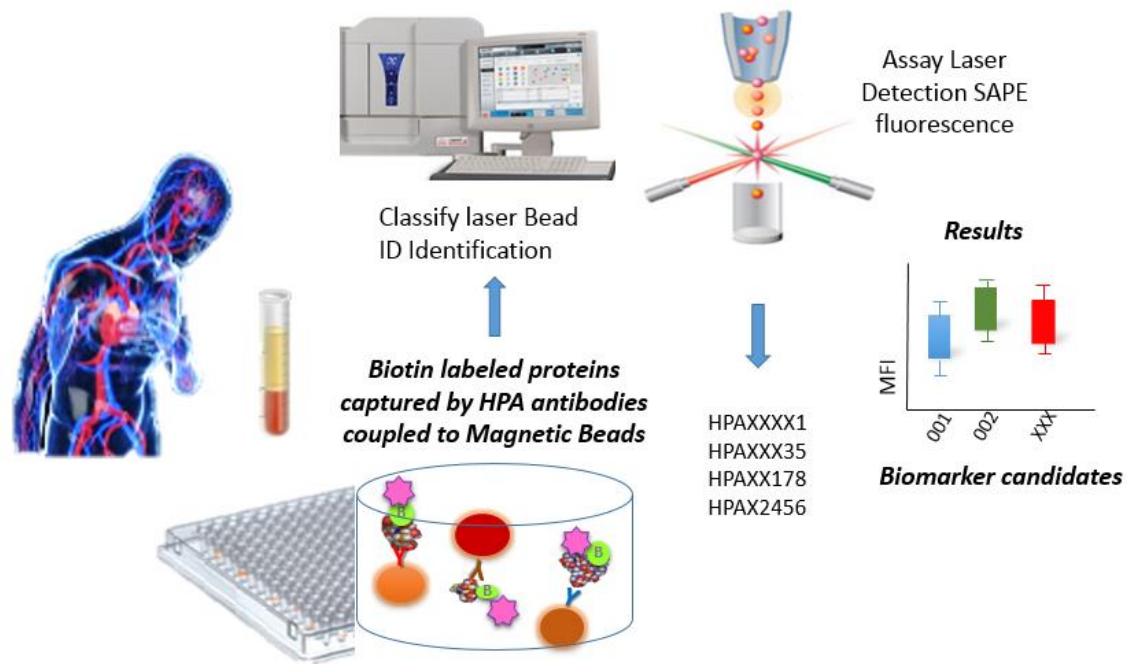


Figure 3: Methodological overview of Affinity plasma proteomics using suspension bead array

4.6 STATISTICAL METHODS

In the univariate analysis for genetic variants and VTE-risk in Paper I-II, unconditional logistic regression was applied to calculate the odds ratio (OR) with 95% confidence interval (95% CI) assuming an additive genetic model with adjustment to age. Significance level was set to p-value < 0.05. Bonferroni correction was used in Paper I to correct for multiple testing.

The predictive models in Paper II were based on logistic regression. For the clinical model all the variables were statistically significant at the 0.05-level and kept in the model. The genetic variables were removed one at a time, according to a backward stepwise selection process, using a p-value of 0.05 as cut-off limit for inclusion. The sum of the regression coefficients was used as a risk score to calculate the area under the receiver-operating characteristics (ROC) curve (AUC). The prediction model was calibrated and cross-validated to assess overfitting and over-optimism. Goodness of fit of the combined model was assessed by inspecting the observed proportion of cases by classes of risk and the Hosmer-Lemeshow was performed to test for lack of fit of the combined model. The AUC is known to be insensitive to differences between prediction models therefore clinical risk reclassification table was calculated to compare the performance between different scores (88).

Recurrence rate in Paper III was calculated as the number of events over the accumulated patient time. Cox-proportional hazard model was used to evaluate risk between groups and hazard ratio (HR) with 95% CI were reported. Crude HR's were assessed in univariate analysis assuming a dominant model for the genetic variables. Cumulative recurrence was calculated with the Kaplan-Meier estimator by subgroups and tested for differences with the log-rank test.

In Paper IV, linear regression was applied to test the association of proteomic data and VTE, estimating the mean difference for the targeted protein levels in plasma with the adjustment for age, gender and additionally by centre in FARIVE. Log-transformation was applied to remove any skewness in the proteomic data distribution. Significance level was set to p-value < 0.05, Bonferroni-corrected.

In Paper I, the univariate analysis was estimated by PLINK and the *ABO*-locus' haplotype analysis was carried out using THESIAS software (89, 90). In Paper II-III, all analyses were performed in Stata version 13. In Paper IV, the analyses were performed using the R statistical computing software (91).

5 RESULTS AND DISCUSSION

5.1 PAPER I

In the first paper, we studied the presumed overlap between arterial CVD and VTE by exploring if common genetic variants associated with CVD-risk were also found associated with VTE.

We identified from a literature search 39 SNPs in 32 loci associated with CAD from GWAS. These were genotyped in TEHS along with genotyping for ABO-blood group and common genetic variants for VTE. The association of SNPs with VTE was assessed by logistic regression.

5.1.1 Results

Only rs579459 in the *ABO* locus demonstrated a significant association with VTE (OR 1.57 (95% CI: 1.39-1.78, p-value 6.4×10^{-13})). Interestingly, rs579459 was found to be in perfect linkage disequilibrium (LD) with a SNP in ABO intron 1 (rs2519093) that had previously been reported independently associated with VTE from ABO blood group genotype. Our haplotype analysis of the ABO SNPs revealed rs579459 to be a perfect proxy for ABO blood group subtype A1 (Table 1). A tentative association between ANRIL and VTE in the discovery analysis in TEHS failed to replicate in a meta-analysis of 4 independent cohorts (total n=7181).

Table 1: Haplotype analysis of the ABO-locus in chromosome 9q34

ABO Haplotype	A1 vs B	O1 vs O1v/O2	A1 vs A2	O vs non-O	intron 1	Frequency	
	rs8176749	rs512770	rs8176704	rs514659	rs579459	contr.	cases
O1	C	G	G	A	T	0.43	0.35
A1	C	G	G	C	C	0.22	0.30
A2	C	G	A	C	T	0.09	0.07
O1v/O2	C	A	G	A	C	0.01	0.01
O1v/O2	C	A	G	A	T	0.16	0.14
B	T	G	G	C	T	0.09	0.11

Where the effect allele is marked in bold and contr. is controls

ABO Haplotype	All		O & A2 vs. A1 & B	
	OR [95% CI]	p-value	OR [95% CI]	p-value
O1	Reference		reference	
A1	1,69 [1,46 - 1,94]	5×10^{-13}	1,66 [1,45 - 1,89]	3×10^{-14}
A2	1,03 [0,84 - 1,27]	0,76	1	
O1v/O2	1,32 [0,81 - 2,14]	0,27	1	
O1v/O2	1,03 [0,87 - 1,21]	0,73	1	
B	1,57 [1,30 - 1,90]	4×10^{-6}	1,54 [1,28 - 1,85]	4×10^{-6}

Where the OR is odds ratio and [95% CI] is 95% confidence interval

5.1.2 Discussion

There is methodological aspects that may have influenced our results. The common CAD risk SNPs that emerged from GWAS have all weak effect estimates (<1.5). Although TEHS was composed of ~2,800 women, the power was insufficient for a discovery analysis where significance of association was corrected for multiple testing.

In our statistical association test we adjusted for variables which were not true confounders defined to be associated with the exposure (genetic variants) and affect the outcome (VTE) well illustrated in direct acyclic graphs (DAG) (Figure 4 and 5) (92). This introduced residual confounding and gave an incorrect picture of plausible association of VTE which in our study was *ANRIL*. The importance to use confounders stringently has to be stressed in genetic studies, as most genetic and clinical variables could impossible be associated with the genetic variants as an exposure (15).

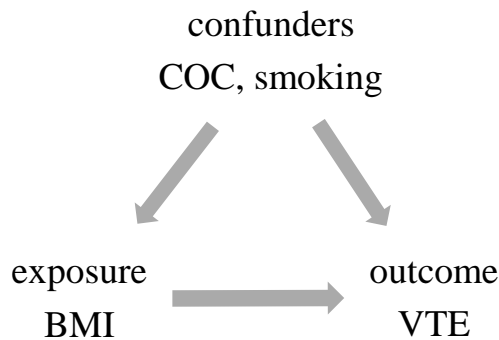


Figure 4. Direct acyclic graphs (DAG) with examples of confounders for VTE. COC is combined contraceptives with oestrogens

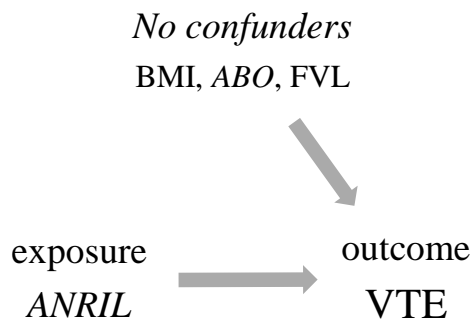


Figure 5. DAGs illustrating that genetic and clinical variables are not associated with the exposure *ANRIL* and therefore no confounders

A genetic variable may act as an effect modifier in the causal pathway, i.e. carriers for FVL have much higher risk for VTE compared to non-carrier if they are exposed to combined oral contraceptive with oestrogens, OR 20.6 compared to OR 3.4, well demonstrated by Bergendal *et al.* in TEHS (93).

However, it has to be stressed there that are some situations when a genetic variable may be a true confounder (Figure 4). This was the case for ABO blood group non-O tagging SNP (rs514659) that was used as confounder. The association with *ABO* rs579459 in intron 1 and VTE declined from an OR of 1.57 to 1.24 when rs514659 was introduced as confounder in the association analysis. This led us to further question if the genetic variant in *ABO* intron 1 really was independent from ABO blood group genotype as previously suggested by Heit *et al* (35, 94). We found that haplotype LD plots may not be sufficient to claim a genetic variant's independent association with a disease. When we constructed haplotypes from four ABO blood group tagging SNPs (O1, A1, A2, B, O1v/O2) and rs579459, we found that *ABO* rs579459 in intron 1 was a perfect proxy for ABO blood group haplotype A1 (Table 1).

Apart from the *ABO* locus, we did not find support for other genetic risk factors shared between CVD and VTE in Swedish women. ABO blood group has repetitively been found associated with both CVD and VTE. This is to a large extent explained by that levels of FVIII and VWF varies between blood groups (34, 95, 96).

5.2 PAPER II

In the second paper we evaluated a genetic risk score, in isolation and when combined with clinical risk factors.

From a literature search, 18 VTE-risk SNPs were identified from candidate gene approach studies or GWAS, and genotyped in DNA samples from 2,835 women from TEHS. Clinical and genetic (SNPs) predictors that contributed significantly to the fit of the logistic regression model were included in the prediction models. Risk scores were evaluated by calculating the AUC.

5.2.1 Results

All clinical risk predictors were significant with the highest risk estimate for immobilisation. In the genetic risk prediction model, 7 SNPs, *F5* rs6025 (FVL), *F2* rs1799963 (*F2* (G20210A), *ABO* rs514659, *FGG* rs2066865, *F11* rs2289252, *PROC* rs1799810 and *KN1* rs710446, contributed significantly to the risk for VTE and remained in the model. Interactions were found between *ABO* rs514659 and *F5* rs6025 and between *ABO* rs514659 and *F11* rs2289252. Furthermore, we found a significant interaction between one or two alleles of *KN1* rs710446 and *F5* rs6025. There were no major differences in risk estimates in the combined model except for family risk of VTE, where the risk estimate declined slightly (Table 2). The clinical risk score attained an AUC of 0.80 (95% CI: 0.79-0.83) and the genetic risk score reached an AUC of 0.66 (95% CI: 0.64-0.68). The combined score attained an AUC of 0.84 (95% CI: 0.82-0.85) with a significant improvement of risk prediction over the clinical risk score (p-value < 0.001) (Figure 6). For comparison, we investigated how the genetic risk score by de Haan *et al.* based on 5 SNPs performed in TEHS (97). This score attained an AUC of 0.65 (95% CI: 0.63-0.67) compared with the AUC of 0.69 attained in the de Haan *et al.* study. The AUC for the TEHS genetic risk score was significantly larger (p-value=0.03) than de Haan *et al.*'s. The AUC is known to be insensitive subtle to differences between prediction models and therefore we proceeded to calculate clinical risk reclassification tables. In the high-risk group (70-100% risk), 51 of 815 were wrongly assigned to a lower than observed risk category using de Haan *et al.*'s score combined with TEHS clinical predictors, while the TEHS combined score classified all high-risk individuals correctly. If only FVL rs6025 and F2 1799963 were used in the combined risk score, 80/815 participants were misclassified in the high-risk group (70-100%). The clinical risk score misclassified 174/815 in the high risk group.

Table 2: Association and significance from the clinical, genetic and combined prediction models in TEHS including SNP-SNP interactions

Variable	Clinical	Genetic	Combined
<i>Clinical:</i>			
Height (cm)	1.04***		1.04***
BMI (kg/m ²)	1.11***		1.12***
Immobilization†	16.5***		18.9***
Long distance flight†	1.6***		1.6***
Hormonal use†‡	4.3***		4.1***
Family history of VTE	2.4***		2.0***
<i>Genetic:</i>			
<i>F5</i> rs6025		6.8***	7.7***
<i>F2</i> rs1799963		1.94**	1.74*
<i>ABO</i> rs514659		1.70***	1.67**
<i>F11</i> rs2289252		1.37***	1.29*
<i>PROC</i> rs1799810		1.14*	1.10
<i>KNG1</i> rs710446		1.56**	1.50*
<i>FGG</i> rs2066865		1.37***	1.43***
<i>SNP interaction:</i>			
<i>F5</i> rs6025_rs6025		0.39**	0.36**
<i>F5</i> rs6025_ <i>ABO</i> rs514659		1.63*	1.65*
<i>ABO</i> rs514659_ <i>F11</i> rs2289252		0.82*	0.85
<i>KNG1</i> rs710446_rs710446		0.84*	0.86

Where: * p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001, BMI, body mass index; †, within three month from diagnosis or index date; ‡, oestrogen containing contraceptive and menopausal replacement therapy(oral or patch)

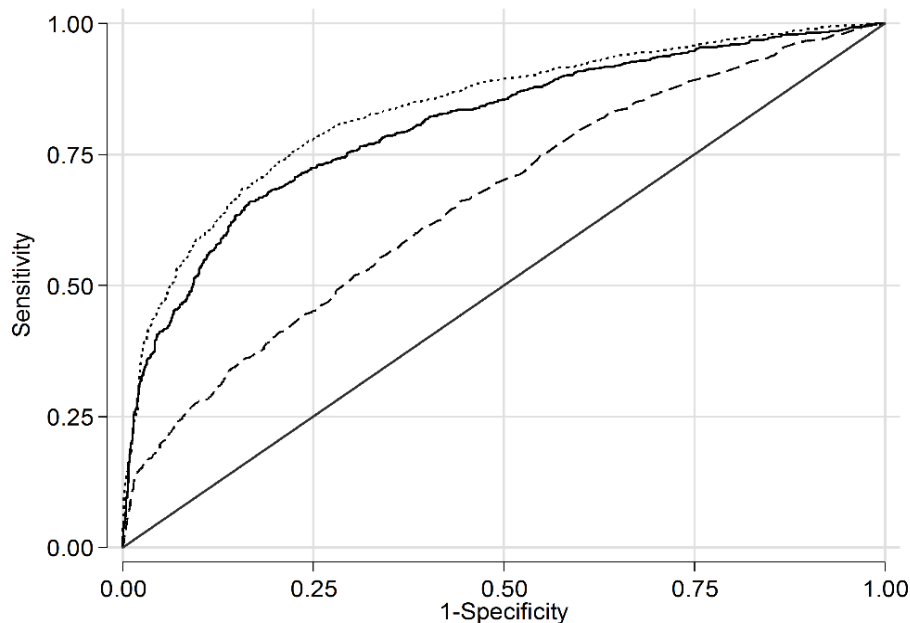


Figure 6: ROC curves for the clinical (solid line), genetic (dashed) and combined (dotted) model in TEHS. Clinical risk score, AUC 0.80 (0.79-0.82); Genetic risk score, AUC 0.66 (0.64-0.68) and Combined score, AUC 0.84 (0.82-0.86)

5.2.2 Discussion

It is recommended to first evaluate existing prediction models before suggesting a new one (98). In TEHS, we evaluated the previously published genetic risk score from de Haan *et al.* with 5 SNPs used as a single predictor with very similar performance in AUC compared to TEHS own genetic risk score. However, it is well known that the AUC is insensitive for differences in prediction models (88, 99). We continued by comparing TEHS clinical-genetic combined model with TEHS clinical models combined with de Haan's 5 SNPs or FVL and F2 (G20210A) (today in clinical use). The clinical risk reclassification tables clearly demonstrated the difference and the risk for misclassification to a lower than observed risk category depending of what prediction models that are used (On-line table 4 of the paper II).

In TEHS, we found that a limited set of genetic predictors including SNP-SNP interactions improved prediction for VTE independently from family history of VTE. This is in contrast to family history of VTE that has been suggested to be more useful in risk assessment (17). To continue, family history of VTE has previously been reported in TEHS, compatible with Bezmer *et al.* and Suchon *et al.*, with low positive predictive value with the established clinical genetic risk factors used in thrombophilia screening (17, 100, 101). This has a clinical implication, as there is today no recommendation for thrombophilia screening in anticipation of risk exposure in individuals with no personal history of VTE or known thrombophilia defect in the family. For instance, as many as 50 percent of the cases in TEHS who had undergone surgery did receive thrombosis prophylaxis with low-molecular heparin but still developed VTE (102), indicating the need to improve current risk assessment practice, i.e. to identify those individuals in whom increased doses or longer duration of thrombosis prophylaxis treatment is warranted.

It can be argued that the need of risk prediction models is of less importance in women compared to men, as women have half the risk for VTE when female specific risk factors are taken into account (46). However the use of combined contraceptives with oestrogens are common among young and middle aged women (in TEHS 33% of the women under 50 years of the cases) and increases the risk for VTE on average 5-fold (93). Even if the incidence is overall low in younger age (1 per 10,000 person years) it is the underlying risk in combination with transient provoking factors for each individual (thrombosis potential) that decide if this individual will suffer from VTE (reach thrombosis threshold), well-illustrated by Lijfering *et al* (Figure 1) (10, 15). We find that our combined risk score may have a clinical implication in women, i.e. prior to high-risk situations or decision on the recommendation of the use of combined contraceptives, but needs to be further evaluated in

other cohorts and improved. It has to be noted that TEHS does not include pregnant women thus our prediction score cannot be applied for this condition without modification. However, a prediction score for thrombosis prophylaxis during pregnancy is already clinically implemented in Sweden (103, 104).

5.3 PAPER III

The objective in the third paper was to explore if the genetic predictors for first event VTE in TEHS were also associated with the risk of recurrence.

In TEHS follow up study, 1,010 women with first VTE were available for analysis of the recurrence rate with respect to the seven genetic predictors identified in paper II; *F5* rs6025 (FVL), *F2* rs1799963 (G20210A), *ABO* rs514659 (non-O ABO blood group), *FGG* rs2066865, *F11* rs2289252, *PROC* rs1799810 and *KNGL1* rs710446.

5.3.1 Results

During a mean period of 5 years, there were 101 recurrent VTE with an estimated rate of 20 per 1,000 person-years (95% CI: 16-24). The recurrence rate was highest in women with an unprovoked first event and obesity. Carrier of the risk alleles of *F5* rs6025 (FVL, HR 1.7 (95% CI: 1.1-2.6)) and *F11* rs2289252 (HR 1.8 (95% CI: 1.1-3.0)), had a significant higher recurrence rate compared to non-carriers of the risk allele. The cumulative recurrence was 2.5-fold larger in carriers of both *F5* rs6025 and *F11* rs2289252 (14.6% (95% CI: 9.9-21.3%)) than in non-carriers (5.9% (95% CI: 3.4-10.3%)) at 5 years follow-up (Figure 7). In subgroups analysis among women with first unprovoked VTE, the cumulative recurrence was even larger in carriers for both risk alleles compared to the non-carriers (21.9% (95% CI: 13.7-34.0%) versus 7.5% (95% CI: 3.4%-16.0%)) (Figure 8). The log rank tests demonstrated the statistical differences between the Kaplan Meier curves in both analysis.

We followed up the predictive ability of the 7 SNPs of our own genetic risk score from Paper II by comparing women grouped into three groups according to the number of SNPs for which they carried the risk allele(s): 0-2, 3-4, and 5-7 SNPs. The cumulative recurrence 5 years into the study was 12.8% for carriers of 5-7 SNPs compared to 8.9% for carriers of 3-4 SNPs and 4.6% for the group carrying the risk allele(s) of 0-2 SNPs. There was not statistically significant difference between the three groups.

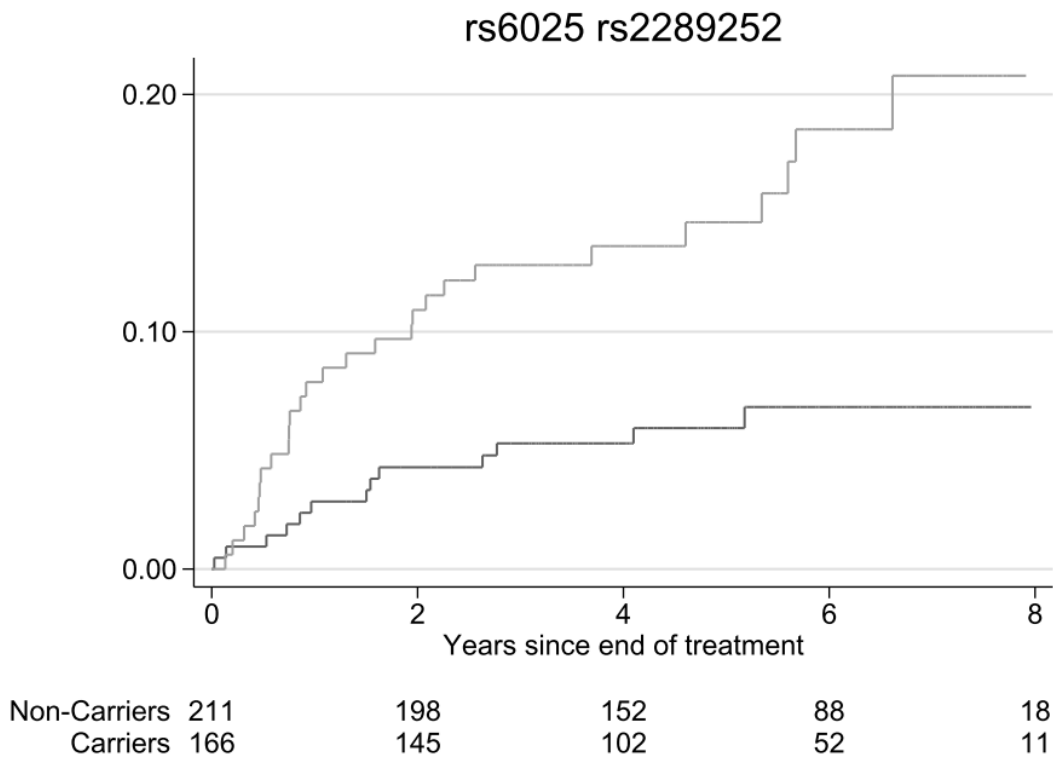


Figure 7: Cumulative recurrence comparing carriers of both risk alleles versus non-carriers for F5 rs6025 and F11 rs228925

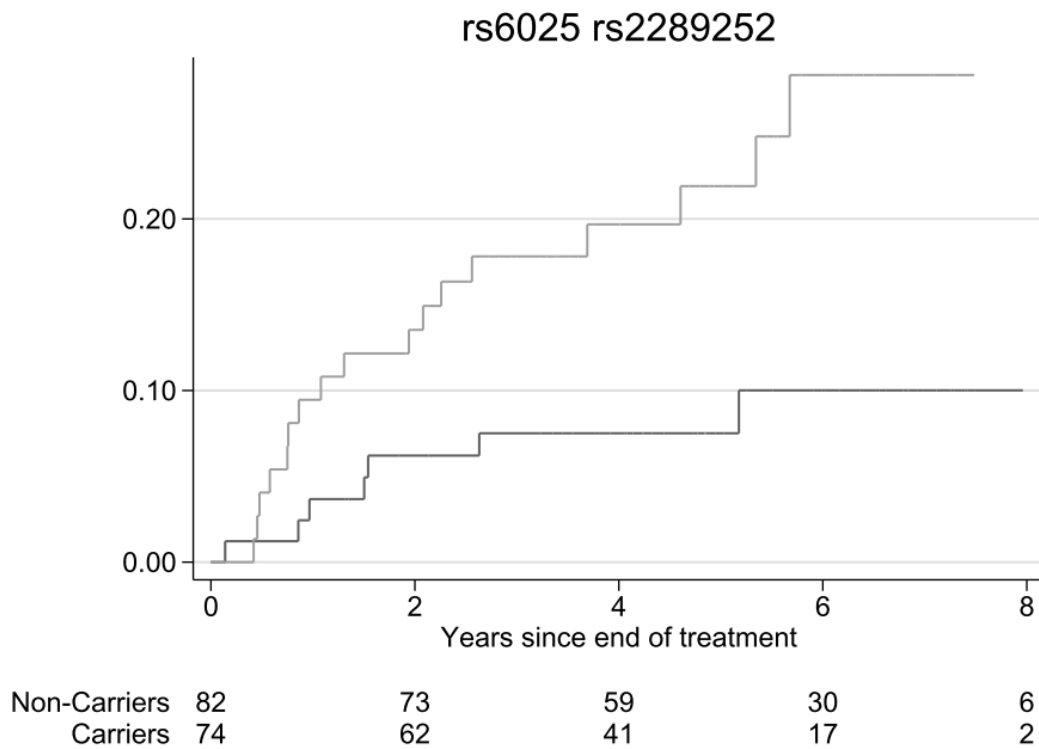


Figure 8: Subgroup analysis of the women with unprovoked fist VTE; cumulative recurrence comparing carriers of both risk alleles versus non-carriers for F5 rs6025 and F11 rs2289252

5.3.2 Discussion

For clinicians, prediction of VTE recurrence remains a challenge. Even if risk prediction models, i.e. HER DOO2, Vienna Prediction Model and DASH, have been suggested, the lack of precision gives the models limited power to identify high risk individuals who would benefit for extended treatment with anticoagulants (38-40). Furthermore, the models include few clinical variables together with D-dimer levels, and screening for thrombophilia was not shown to add any predictive value (38-40).

The TEHS follow up study had few recurrent events over the follow-up period of 5 years compared to other studies which most probably could be explained by the study design of TEHS that comprised young and middle age women with both provoked and unprovoked first VTE (81). This left us with a reduced power to analyse whether the genetic markers (SNPs) which predicted first VTE in TEHS (Paper II) also were associated with recurrence. In our association test we therefore chose to use a dominant model instead of the more commonly used additive model for the SNPs (carrier=1 non-carrier=0 of a risk allele in a SNP). We expected only a minor influence on the results as the risk for recurrence is relatively small on the log scale for a carrier compared to a non-carriers of a risk allele.

Our finding that *F11* rs2289252 was associated on recurrence is interesting as there is growing evidence that FXI is a key player in the pathogenesis of thrombosis (105-109). It has also gained attention as an anticoagulant drug target (110, 111). The *F11* rs2289252 association with first time VTE has partly been explained by increased levels of FXI (112, 113). In the MEGA follow-up study, an association with recurrence was also reported for *F11* but with two other SNPs; rs2036914 and rs3822057. However, these two SNPs were not independently associated with VTE (LD $r^2 = 0.91$ in the 1,000 Genomes study) (114). We continued to follow up the finding of *F11* rs2036914 from MEGA follow up in relation to rs2289252. In the 1,000 Genomes study the LD was only $r^2=0.46$ between the SNPs (114). However, Germain *et al.* has clearly demonstrated an at-risk haplotype, derived from these two SNPs, with as high frequency as ~0.40 that increased the risk with an OR of 1.40 (24).

We observed that 5 years into the follow up study in TEHS, women who carried risk alleles for both *F11* rs2289252 and *F5* rs6025 had a slightly higher cumulative recurrence compared to those who carried risk alleles for ≥ 5 of the 7 SNPs which predicted first VTE in TEHS. Based on this, we believe there is a risk of misclassification of individuals at high risk to a lower risk-group if a sum of alleles/genetic variants is applied in a prediction model.

Much research remains to be conducted regarding risk prediction of reoccurrence of VTE. Analyses in collaboration between research groups would provide more statistical power to find the optimal set of genetic markers to integrate with clinical predictors and plasma biomarkers.

In the genetic risk scores for first VTE, *F11* rs2036914 was included in de Haan *et al.*'s 5 SNPs score whereas *F11* rs2289252 was added in TEHS own score (Paper II) (97). It is tempting to believe that the haplotype of *F11* rs2289252 and rs2036914 would be more useful in risk assessment of both incident VTE and recurrence than to use of only one of these SNPs in *F11* (24).

5.4 PAPER IV

Our aim in the fourth paper was to apply a plasma proteomics approach for the discovery of novel biomarkers for VTE that could improve current prediction tools and to identify new leads to explore the pathophysiology involved in VTE.

In the discovery experiment, plasma samples of 88 cases and 85 controls from participants of VEBIOS were screened for 408 proteins targeted by 755 HPA antibodies using multiplex bead arrays. The selection of targets was based on a) evidence on association with VTE or intermediate traits and b) HPA antibodies' availability and predefined quality criteria. Proteins that significantly associated with VTE after Bonferroni correction were tested for replication in citrated plasma samples of 580 cases and 589 controls from FARIVE.

5.4.1 Results

In VEBIOS, plasma levels of four proteins, Human Immuno-deficiency Virus type I Enhancer binding Protein 1 (HIVEP1), VWF, Glutathione Peroxidase 3 (GPX3) and Platelet-derived growth factor beta (PDGFB) targeted by HPA050724, HPA002082, HPA059686 and HPA011972 antibodies, respectively, were found significantly associated with VTE after Bonferroni correction, with a good concordance between measurements in citrate and EDTA plasma. Even though the direction of associations of GPX3 and HIVEP1 with VTE observed in FARIVE paralleled those observed in VEBIOS, they did not reach statistical significance. Furthermore, in FARIVE a trend for a genotype-specific (*HIVEP1* rs169713) association was found between HIVEP1 protein levels and VTE risk. VWF (Figure 9) and PDGFB (Figure 10) successfully replicated in FARIVE with significantly higher signals in plasma from VTE cases compared to controls. These two potential biomarkers were modestly correlated ($\rho \sim 0.30$) with each other and in a joint model independently associated with VTE risk (FARIVE; $p < 0.001$ and $p = 0.002$ for VWF and PDGFB, respectively).

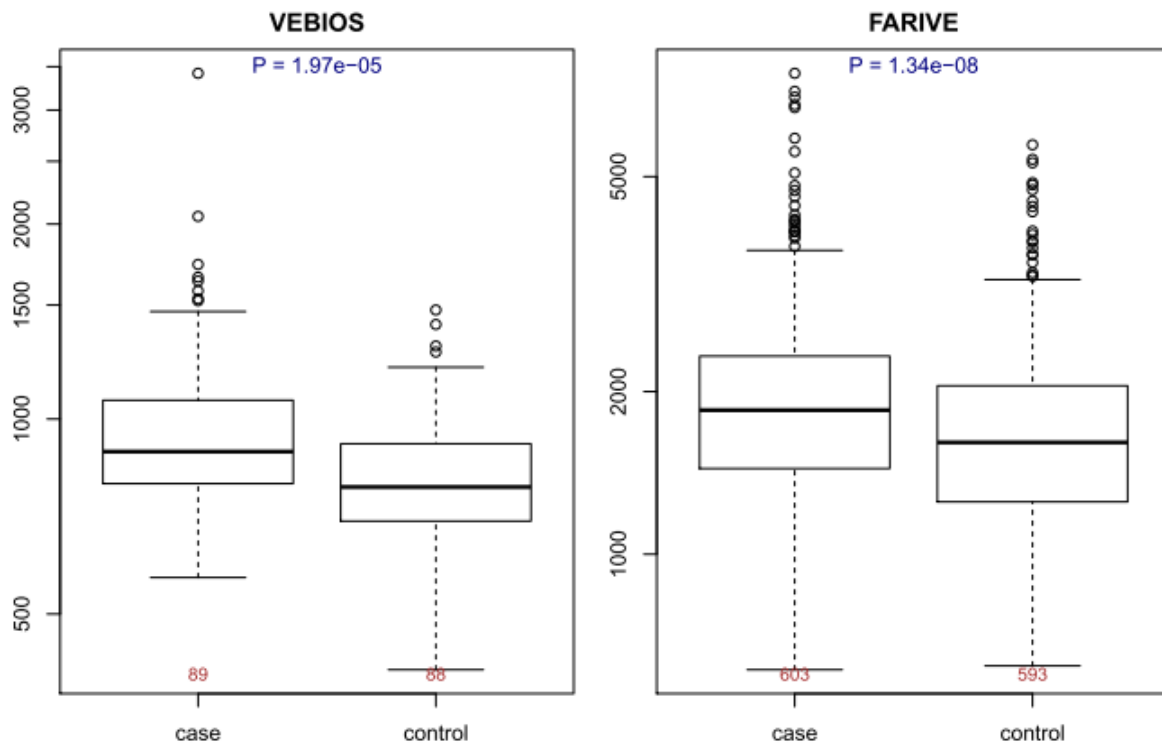


Figure 9: Boxplot of VWF (targeted by HPA002082) in VEBIOS and FARIVE where X-axis is disease status and Y-axis is normalized median fluorescent intensity

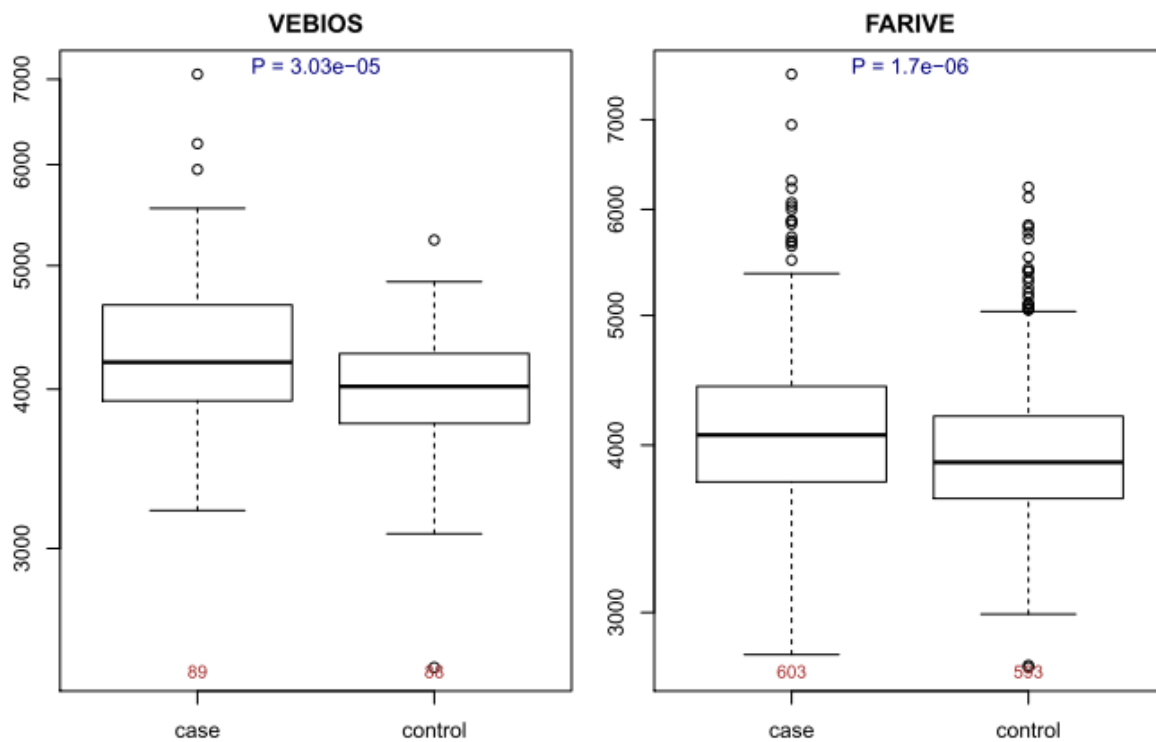


Figure 10: Boxplot of PDGFB (HPA011972) in VEBIOS and FARIVE where X-axis is disease status and Y-axis is normalized median fluorescent intensity

5.4.2 Discussion

We identified one novel candidate biomarker, PDGFB, for VTE risk. The significant difference in VWF levels between VTE patients and controls, served as a proof of principle for our study design and methodological approach.

The study included plasma sample sets from a discovery study - VEBIOS and an independent replication study - FARIVE. Both studies have the advantage of being well phenotyped and they are similar in study design with the exception of recruitment of controls. Most importantly the protocols for collection, handling and storage of blood samples applied were comparable, however the sampling time from diagnosis differ for the cases; in VEBIOS it was one month after stopped treatment (at least 7 months from diagnosis) and in FARIVE it was during the acute phase (within the first weeks from diagnosis) under influence of anticoagulant treatment. This introduces selection bias that has to be taken into account when interpreting the results. With the knowledge that proteomics analysis is multidimensional and prone to many technical and methodological pitfalls, adopting a stringent statistical approach was necessary.

PDGFB is a protein expressed in the endothelium. The increased plasma levels could reflect increased protein synthesis and secretion in VTE patients compared to controls, or could reflect tissue leakage under endothelial dysfunction. The association was independent of sampling time from diagnosis and of anticoagulant treatment. However, at this point we do not know if PDGFB could be causal to VTE or reactive (a consequence of the disease) (Figure 11) (115). To further elucidate PDGFBs role as a biomarker, both functional and clinical studies (preferable in prospective cohorts) are required.

Our most significant finding in the discovery phase was increased levels of HIVEP1 in VTE patients compared to controls. This finding was not replicated in FARIVE although a tendency of higher levels was found for the patients with VTE compared to controls. In addition, a trend in FARIVE for an association between HIVEP1 plasma levels and VT risk was observed, dependent on the VTE-associated rs169713 polymorphism previously identified from GWAS (116). The biological function of HIVEP1 is unknown which is why we cannot rule out at this stage that sampling time or anticoagulant treatment may have affected the protein levels in the VTE patients in FARIVE (Figure 11). Our preliminary results observed for HIVEP1 require further validation in additional materials with both DNA and plasma samples but suggests that findings from GWAS studies can be translated into proteomics studies.

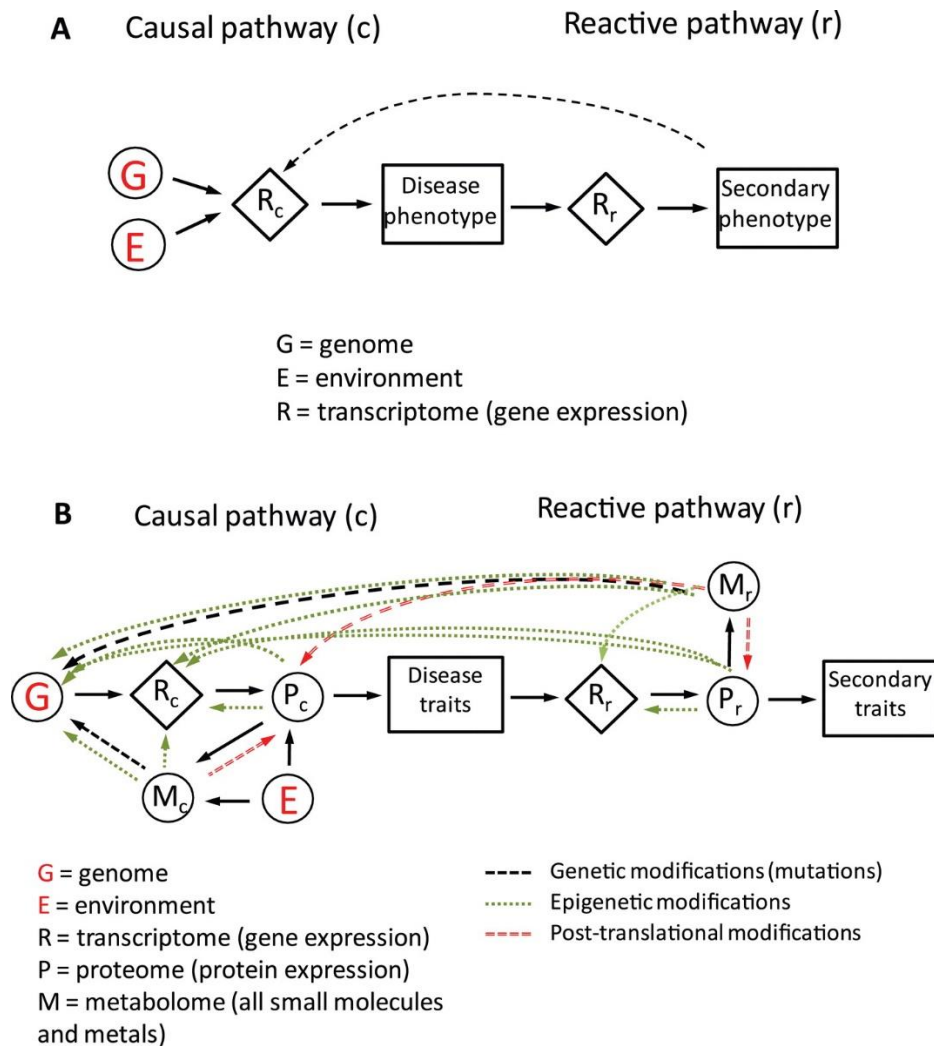


Figure 11. Pathways showing influences of genetic (G) and environmental (E) factors on chronic diseases. A) Model of causal and reactive pathways due to effects of G and E on gene expression (R), as reported by Schadt et al. (Schadt et al. 2005). B) Model of causal and reactive pathways from (A) extended to include the proteome (P) and metabolome (M). *Biomarkers* 2012, 17, 483-489. Copyright © 2012 Informa UK, Ltd. Figure reused with the kind permission from. S.M Rappaport.

Furthermore, our putative biomarker has to survive many steps before application in clinics. Work has begun to verify the identified protein. Even if we could replicate PDGFB in an independent sample set, verification is necessary because single binder assays are vulnerable to off-target binding of other proteins in plasma (77). Initially we are using targeted MS with immunoprecipitation. In parallel, dual binders enzyme-linked immunosorbent assay (ELISA)

is developed to enable quantification of the protein in plasma (78, 79). This enables validation of PDGFB with estimation of the sensitivity and specificity for disease status. However, PDGFB's predictive value is dependent on the prevalence of VTE in the cohort PDGFB will be applied on. Only a biomarker (or a panel of biomarkers) with high positive predictive value (PPV) or negative predictive value (NPV) will finally be considered for prediction in clinical practice (7, 60, 117, 118).

6 CONCLUSION

Conclusions from the genetic studies in TEHS in Paper I-III:

- Apart from the *ABO* locus, a presumed overlap in genetic risk between arterial CVD and VTE could not be established when studying common genetics variants of CAD and the association on VTE risk.
- A limited set of genetic predictors improved prediction of VTE in women at high risk, independently of family history of VTE. Our data indicate that interactions among SNPs increase the goodness of fit when predicting VTE.
- For the first time, *F11* rs2289252 was found to be associated with risk of recurrent VTE in women. A combination of genotypes of *F11* rs2289252 and FVL SNPs may be of potential clinical relevance for risk prediction.

Conclusions from the proteomic study in VEBIOS and FRAIVE in Paper IV:

- The platelet-derived growth factor beta (PDGFB) was identified as a candidate biomarker for VTE-risk.
- Affinity plasma proteomics proved to be a valuable research strategy to discover novel biomarkers and opened new promising perspectives for VTE and its clinical outcomes.

7 FUTURE PERSPECTIVES

We are in the ‘omics’ era with powerful high throughput methods. During the last 10 years a dozen susceptibility genetic markers have been generated in the field of VTE, however with insufficient predictive value as single markers (119). To incorporate genetic markers in prediction scores for both incident VTE and recurrence have been suggested (97, 120). However, I am convinced that we have not used the full potential of the information that genetic markers can provide us with in combination with clinical predictors and plasma biomarkers. This demands more advanced models that allow for biological interaction and synergetic effects to be incorporated to increase accuracy and discrimination power. For clinical application any prediction models have to be user friendly and the panel of SNPs available as routine detection assays at a reasonable cost from a health economic perspective.

In terms of plasma proteomics, I am cautiously optimistic. The discovery of tentative biomarkers for VTE reflects the interplay between genetic and environmental factors (Figure 11) (115). However, to take a biomarker from discovery to clinical application is a major task, requiring access to large patient cohorts and the expertise of epidemiologists, biostatisticians, bioinformaticians, engineers, basic scientists and clinicians in the VTE field.

8 ACKNOWLEDGEMENTS

This thesis has not been possible without the support and encouragement from colleagues, friends and family over the last years.

I want to express my sincere gratitude to;

My main supervisor Jacob Odeberg. Thank you for being an inspiring supervisor and a true friend. Thank you for generously sharing your research experience with me and involving me in your proteomic project. Thank you for believing in my capacity and always caring for me.

My co-supervisor Anders Hamsten. Thank you for introducing me to cardiovascular research and helping me in the design of our study VEBIOS. Thank you for supporting me during my PhD work and generously presenting me to your international network.

My co-supervisor Rona Strawbridge. Thank you for introducing me to cardiovascular genetics and sharing your knowledge. Thank you for introducing me to scientific writing and teaching me how to structure a manuscript.

Angela Silveira, thank you for your enormous kindness and stepping in when Anders did not have the possibility to supervise because of other commitments. I do appreciate all your guidance and encouragement during my PhD work as well as planning for my thesis defense.

To continue, three persons who truly contributed with their excellence to my thesis;

Matteo Bottai, a role model for us all at KI. Thank you for making a difference, our many biosatisfical sessions gave fruit. I am so proud of what we accomplished!

Pierre Morange, your scientific contribution to VTE genetics has indeed inspired and helped me during my PhD. I treasure our many discussions and your encouragement in my research.

David Trégouët, I am very thankful that you draw my attention to the ABO-locus. Thank you for all great ideas and constructive problem solving in our proteomic study, I am amazed.

My thesis have not been possible without;

Maria Jesus Inglesias, thank you for learning me all about Affinity Proteomics and very daringly letting me assist you in the lab. Thank you for doing such a great work in our VEREMA project.

Maria Ljungqvist, I truly enjoyed to work with you in your TEHS follow-up study. It is so well done and I wish you all the best for you thesis.

Mun-Gwan Hong, biostatistician in VEREMA. Thank you for all your hard work and taking your time to explain and share your experience from previous proteomic studies.

Anna Fahlén, thank you for being a structured and efficient research nurse. You came as a blessing to help us with VEBIOS when everything was in chaos.

Mattias Frånberg, thank you for building the VEBIOS data base and always being so helpful.

Annica Bergendal and Helle Kieler. I feel very fortunate to have worked with TEHS but also to have had your stringent epidemiological support when writing the articles.

Eva-Lotta Hempel, thank you for including many of the cases in VEBIOS. You are a great clinician and so fun to work with.

Lynn Butler, thank you for adding another dimension to the proteomics project with your research skills as well as fulfilling Jacob's life.

I would never have been able to finish my thesis without:

Research school for clinicians in Epidemiology with teacher principal Ola Olén. Tekla, Mia, Katarina and Alexandra, thank you for sharing the Research school time with me. I would never imagine that epidemiology could give so many good laughs.

And support from my employer;

Margareta Holmström, thank you for providing me with the great possibility to work part time during my PhD work. Thank you for all your help with VEBIOS. You have been a true support.

Viktoria Hjalmar, you employed me 2006 and was for long time my boss, thank you for being such a wise and good person. You are great manager.

My colleagues at the Hematology centre, but in particular

Anna Ågren, thank you for always trying to do your best with our working schedule. Thank you for being a real friend with many debriefing talks about work, research and life.

Malin Hultcrantz, my roommate. Your positive energy and passion for research and life inspired me. You will make it anywhere – best of luck in New York.

Ammar Majeed, thank you for being a great colleague and researcher.

Petra Elfvinge, Malin Svennefalk, Elsa Olsson and Doris Näslin, you are the foundation of our Coagulation Unit. You are all so great, so great that the patients never want to leave us.

And finally and most important of all,

My wonderful parents, you are so full of joy and curiosity about the world and the human kind, thank you for your endless love and care.

Patrik, my love, thank you for sharing your life with me and for coping with my constant working. I spent far too much time with my computer instead of enjoying life with you and our gorgeous sons, Petter and Axel. However, very soon, it's time to go skiing!

This thesis was supported with grants from Stockholm County Council.

TEHS was supported by unrestricted grants from Janssen-Cilag, Novartis, Organon, Schering, Wyeth, AFA Insurance, Center for Gender Medicine Karolinska Institutet and the Medical Products Agency.

The KTH Center for Applied Proteomics funded by the Erling-Persson Family Foundation is acknowledged for financial support.

9 REFERENCES

1. Wendelboe AM, McCumber M, Hylek EM, Buller H, Weitz JI, Raskob G. Global public awareness of venous thromboembolism. *Journal of thrombosis and haemostasis : JTH*. 2015 Aug;13(8):1365-71.
2. Thrombosis: a major contributor to the global disease burden. *Journal of thrombosis and haemostasis : JTH*. 2014 Oct;12(10):1580-90.
3. Jha AK, Larizgoitia I, Audera-Lopez C, Prasopa-Plaizier N, Waters H, Bates DW. The global burden of unsafe medical care: analytic modelling of observational studies. *BMJ Qual Saf*. 2013 Oct;22(10):809-15.
4. Owen CA. *A History of Blood Coagulation*. 1 ed. Rochester, Minnesota, U.S: Mayo Foundation for Medical Education and Research; 2001. p. 169-77.
5. Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet*. 1999 Apr 3;353(9159):1167-73.
6. Bacchus F, Schulman S. Clinical experience with the new oral anticoagulants for treatment of venous thromboembolism. *Arteriosclerosis, thrombosis, and vascular biology*. 2015 Mar;35(3):513-9.
7. Frantzi M, Bhat A, Latosinska A. Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development. *Clin Transl Med*. 2014;3(1):7.
8. Spronk HM, Cannegieter S, Morange P, Hackeng T, Huisman M, Nagler M, et al. Theme 2: Epidemiology, Biomarkers, and Imaging of Venous Thromboembolism (and postthrombotic syndrome). *Thrombosis research*. 2015 Sep;136 Suppl 1:S8-S12.
9. Martinez C, Cohen AT, Bamber L, Rietbrock S. Epidemiology of first and recurrent venous thromboembolism: A population-based cohort study in patients without active cancer. *Thrombosis and Haemostasis*. 2014 Apr 3;112(2).
10. Naess IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerstrom J. Incidence and mortality of venous thrombosis: a population-based study. *J Thromb Haemost*. 2007 Apr;5(4):692-9.
11. Christiansen SC, Lijfering WM, Helmerhorst FM, Rosendaal FR, Cannegieter SC. Sex difference in risk of recurrent venous thrombosis and the risk profile for a second event. *Journal of thrombosis and haemostasis : JTH*. 2010 Oct;8(10):2159-68.
12. Douketis J, Tosetto A, Marcucci M, Baglin T, Cosmi B, Cushman M, et al. Risk of recurrence after venous thromboembolism in men and women: patient level meta-analysis. *Bmj*. 2011;342:d813.
13. Flinterman LE, van Hylckama Vlieg A, Cannegieter SC, Rosendaal FR. Long-term survival in a large cohort of patients with venous thrombosis: incidence and predictors. *PLoS Med*. 2012 Jan;9(1):e1001155.
14. Rosendaal FR, Reitsma PH. Genetics of venous thrombosis. *J Thromb Haemost*. 2009 Jul;7 Suppl 1:301-4.
15. Lijfering WM, Rosendaal FR, Cannegieter SC. Risk factors for venous thrombosis - current understanding from an epidemiological point of view. *Br J Haematol*. 2010 Jun;149(6):824-33.

16. Christiansen SC, Lijfering WM, Naess IA, Hammerstrom J, van Hylckama Vlieg A, Rosendaal FR, et al. The relationship between body mass index, activated protein C resistance and risk of venous thrombosis. *Journal of thrombosis and haemostasis : JTH*. 2012 Sep;10(9):1761-7.
17. Bezemer ID, van der Meer FJ, Eikenboom JC, Rosendaal FR, Doggen CJ. The value of family history as a risk indicator for venous thrombosis. *Arch Intern Med*. 2009 Mar 23;169(6):610-5.
18. Mili FD, Hooper WC, Lally C, Austin H. The impact of co-morbid conditions on family history of venous thromboembolism in Whites and Blacks. *Thromb Res*. 2011 Apr;127(4):309-16.
19. Zoller B, Li X, Sundquist J, Sundquist K. Age- and gender-specific familial risks for venous thromboembolism: a nationwide epidemiological study based on hospitalizations in Sweden. *Circulation*. 2011 Aug 30;124(9):1012-20.
20. Sorensen HT, Riis AH, Diaz LJ, Andersen EW, Baron JA, Andersen PK. Familial risk of venous thromboembolism: a nationwide cohort study. *J Thromb Haemost*. 2011 Feb;9(2):320-4.
21. Heit JA, Phelps MA, Ward SA, Slusser JP, Petterson TM, De Andrade M. Familial segregation of venous thromboembolism. *J Thromb Haemost*. 2004 May;2(5):731-6.
22. Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, et al. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. *Genetic Analysis of Idiopathic Thrombophilia*. *Am J Hum Genet*. 2000 Dec;67(6):1452-9.
23. Larsen TB, Sorensen HT, Skytthe A, Johnsen SP, Vaupel JW, Christensen K. Major genetic susceptibility for venous thromboembolism in men: a study of Danish twins. *Epidemiology*. 2003 May;14(3):328-32.
24. Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, et al. Genetics of venous thrombosis: insights from a new genome wide association study. *PLoS One*. 2011;6(9):e25581.
25. Couturaud F, Leroyer C, Tromeur C, Julian JA, Kahn SR, Ginsberg JS, et al. Factors that predict thrombosis in relatives of patients with venous thromboembolism. *Blood*. 2014 Sep 25;124(13):2124-30.
26. Cohen W, Castelli C, Suchon P, Bouvet S, Aillaud MF, Brunet D, et al. Risk assessment of venous thrombosis in families with known hereditary thrombophilia: the MARseilles-NImes prediction model. *Journal of thrombosis and haemostasis : JTH*. 2014 Feb;12(2):138-46.
27. Zoller B, Ohlsson H, Sundquist J, Sundquist K. Family history of venous thromboembolism (VTE) and risk of recurrent hospitalization for VTE: a nationwide family study in Sweden. *Journal of thrombosis and haemostasis : JTH*. 2014 Mar;12(3):306-12.
28. Egeberg O. On the natural blood coagulation inhibitor system. Investigations of inhibitor factors based on antithrombin deficient blood. *Thromb Diath Haemorrh*. 1965 Nov 15;14(3-4):473-89.
29. Griffin JH, Evatt B, Zimmerman TS, Kleiss AJ, Wideman C. Deficiency of protein C in congenital thrombotic disease. *The Journal of clinical investigation*. 1981 Nov;68(5):1370-3.
30. Schwarz HP, Fischer M, Hopmeier P, Batard MA, Griffin JH. Plasma protein S deficiency in familial thrombotic disease. *Blood*. 1984 Dec;64(6):1297-300.

31. Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature*. 1994 May 5;369(6475):64-7.
32. Poort SR, Rosendaal FR, Reitsma PH, Bertina RM. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood*. 1996 Nov 15;88(10):3698-703.
33. Morelli VM, de Visser MC, van Tilburg NH, Vos HL, Eikenboom JC, Rosendaal FR, et al. ABO blood group genotypes, plasma von Willebrand factor levels and loading of von Willebrand factor with A and B antigens. *Thrombosis and Haemostasis*. 2007 Apr;97(4):534-41.
34. Dentali F, Sironi AP, Ageno W, Turato S, Bonfanti C, Frattini F, et al. Non-O blood type is the commonest genetic risk factor for VTE: results from a meta-analysis of the literature. *Semin Thromb Hemost*. 2012 Jul;38(5):535-48.
35. Heit JA, Cunningham JM, Petterson TM, Armasu SM, Rider DN, M DEA. Genetic variation within the anticoagulant, procoagulant, fibrinolytic and innate immunity pathways as risk factors for venous thromboembolism. *Journal of thrombosis and haemostasis : JTH*. 2011 Jun;9(6):1133-42.
36. Coppens M, Reijnders JH, Middeldorp S, Doggen CJ, Rosendaal FR. Testing for inherited thrombophilia does not reduce the recurrence of venous thrombosis. *Journal of thrombosis and haemostasis : JTH*. 2008 Sep;6(9):1474-7.
37. Kyrle PA, Rosendaal FR, Eichinger S. Risk assessment for recurrent venous thrombosis. *Lancet*. 2010 Dec 11;376(9757):2032-9.
38. Eichinger S, Heinze G, Jandek LM, Kyrle PA. Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model. *Circulation*. 2010 Apr 13;121(14):1630-6.
39. Rodger MA, Kahn SR, Wells PS, Anderson DA, Chagnon I, Le Gal G, et al. Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. 2008 Aug 26;179(5):417-26.
40. Tosetto A, Iorio A, Marcucci M, Baglin T, Cushman M, Eichinger S, et al. Predicting disease recurrence in patients with previous unprovoked venous thromboembolism: a proposed prediction score (DASH). *Journal of thrombosis and haemostasis : JTH*. 2012 Jun;10(6):1019-25.
41. Morange PE, Tregouet DA. Current knowledge on the genetics of incident venous thrombosis. *Journal of thrombosis and haemostasis : JTH*. 2013 Jun;11 Suppl 1:111-21.
42. Antoni G, Oudot-Mellakh T, Dimitromanolakis A, Germain M, Cohen W, Wells P, et al. Combined analysis of three genome-wide association studies on vWF and FVIII plasma levels. *BMC Med Genet*. 2011;12:102.
43. Sabater-Lleal M, Martinez-Perez A, Buil A, Folkersen L, Souto JC, Bruzelius M, et al. A genome-wide association study identifies KNG1 as a genetic determinant of plasma factor XI Level and activated partial thromboplastin time. *Arteriosclerosis, thrombosis, and vascular biology*. 2012 Aug;32(8):2008-16.

44. Germain M, Chasman DI, de Haan H, Tang W, Lindstrom S, Weng LC, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *American journal of human genetics*. 2015 Apr 2;96(4):532-42.
45. Vlieg A, Flinterman LE, Bare LA, Cannegieter SC, Reitsma PH, Arellano AR, et al. Genetic variations associated with recurrent venous thrombosis. *Circulation Cardiovascular genetics*. 2014 Dec;7(6):806-13.
46. Roach RE, Lijfering WM, Rosendaal FR, Cannegieter SC, le Cessie S. Sex difference in risk of second but not of first venous thrombosis: paradox explained. *Circulation*. 2014 Jan 7;129(1):51-6.
47. Roach RE, Lijfering WM, Tait CR, Baglin T, Kyrle PA, Cannegieter SC, et al. Sex difference in the risk of recurrent venous thrombosis: a detailed analysis in four European cohorts. *Journal of thrombosis and haemostasis : JTH*. 2015 Aug 13.
48. Becattini C, Vedovati MC, Ageno W, Dentali F, Agnelli G. Incidence of arterial cardiovascular events after venous thromboembolism: a systematic review and a meta-analysis. *J Thromb Haemost*. 2010 May;8(5):891-7.
49. Lind C, Flinterman LE, Enga KF, Severinsen MT, Kristensen SR, Braekkan SK, et al. Impact of incident venous thromboembolism on risk of arterial thrombotic diseases. *Circulation*. 2014 Feb 25;129(8):855-63.
50. Roach RE, Lijfering WM, Flinterman LE, Rosendaal FR, Cannegieter SC. Increased risk of CVD after VT is determined by common etiologic factors. *Blood*. 2013 Jun 13;121(24):4948-54.
51. Glynn RJ, Rosner B. Comparison of risk factors for the competing risks of coronary heart disease, stroke, and venous thromboembolism. *Am J Epidemiol*. 2005 Nov 15;162(10):975-82.
52. Holst AG, Jensen G, Prescott E. Risk factors for venous thromboembolism: results from the Copenhagen City Heart Study. *Circulation*. 2010 May 4;121(17):1896-903.
53. Braekkan SK, Hald EM, Mathiesen EB, Njolstad I, Wilsgaard T, Rosendaal FR, et al. Competing risk of atherosclerotic risk factors for arterial and venous thrombosis in a general population: the Tromso study. *Arterioscler Thromb Vasc Biol*. 2012 Feb;32(2):487-91.
54. Tsai AW, Cushman M, Rosamond WD, Heckbert SR, Polak JF, Folsom AR. Cardiovascular risk factors and venous thromboembolism incidence: the longitudinal investigation of thromboembolism etiology. *Arch Intern Med*. 2002 May 27;162(10):1182-9.
55. Bank I, van de Poel MH, Coppens M, Hamulyak K, Prins MH, van der Meer J, et al. Absolute annual incidences of first events of venous thromboembolism and arterial vascular events in individuals with elevated FVIII:c. A prospective family cohort study. *Thrombosis and Haemostasis*. 2007 Nov;98(5):1040-4.
56. Mulder R, van Schouwenburg IM, Mahmoodi BK, Veeger NJ, Mulder AB, Middeldorp S, et al. Associations between high factor VIII and low free protein S levels with traditional arterial thrombotic risk factors and their risk on arterial thrombosis: results from a retrospective family cohort study. *Thrombosis research*. 2010 Oct;126(4):e249-54.
57. Meltzer ME, Doggen CJ, de Groot PG, Rosendaal FR, Lisman T. The impact of the fibrinolytic system on the risk of venous and arterial thrombosis. *Seminars in thrombosis and hemostasis*. 2009 Jul;35(5):468-77.

58. Meltzer ME, Lisman T, de Groot PG, Meijers JC, le Cessie S, Doggen CJ, et al. Venous thrombosis risk associated with plasma hypofibrinolysis is explained by elevated plasma levels of TAFI and PAI-1. *Blood*. 2010 Jul 8;116(1):113-21.
59. Cushman M, O'Meara ES, Folsom AR, Heckbert SR. Coagulation factors IX through XIII and the risk of future venous thrombosis: the Longitudinal Investigation of Thromboembolism Etiology. *Blood*. 2009 Oct 1;114(14):2878-83.
60. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001 Mar;69(3):89-95.
61. van Hylckama Vlieg A, Baglin CA, Luddington R, MacDonald S, Rosendaal FR, Baglin TP. The risk of a first and a recurrent venous thrombosis associated with an elevated D-dimer level and an elevated thrombin potential: results of the THE-VTE study. *Journal of thrombosis and haemostasis : JTH*. 2015 Sep;13(9):1642-52.
62. Wells PS, Forgie MA, Simms M, Greene A, Touchie D, Lewis G, et al. The outpatient bleeding risk index: validation of a tool for predicting bleeding rates in patients treated for deep venous thrombosis and pulmonary embolism. *Archives of internal medicine*. 2003 Apr 28;163(8):917-20.
63. Kearon C, Akl EA. Duration of anticoagulant therapy for deep vein thrombosis and pulmonary embolism. *Blood*. 2014 Mar 20;123(12):1794-801.
64. Kearon C, Spencer FA, O'Keefe D, Parpia S, Schulman S, Baglin T, et al. D-dimer testing to select patients with a first unprovoked venous thromboembolism who can stop anticoagulant therapy: a cohort study. *Annals of Internal Medicine*. 2015 Jan 6;162(1):27-34.
65. Jenkins PV, Rawley O, Smith OP, O'Donnell JS. Elevated factor VIII levels and risk of venous thrombosis. *British journal of haematology*. 2012 Jun;157(6):653-63.
66. Timp JF, Lijfering WM, Flinterman LE, van Hylckama Vlieg A, le Cessie S, Rosendaal FR, et al. Predictive value of factor VIII levels for recurrent venous thrombosis: results from the MEGA follow-up study. *Journal of thrombosis and haemostasis : JTH*. 2015 Aug 13.
67. Ost D, Tepper J, Mihara H, Lander O, Heinzer R, Fein A. Duration of anticoagulation following venous thromboembolism: a meta-analysis. *JAMA*. 2005 Aug 10;294(6):706-15.
68. Drobin K, Nilsson P, Schwenk JM. Highly multiplexed antibody suspension bead arrays for plasma protein profiling. *Methods in molecular biology*. 2013;1023:137-45.
69. Kingsmore SF. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nat Rev Drug Discov*. 2006 Apr;5(4):310-20.
70. Ayoglu B, Haggmark A, Neiman M, Igel U, Uhlen M, Schwenk JM, et al. Systematic antibody and antigen-based proteomic profiling with microarrays. *Expert Rev Mol Diagn*. 2011 Mar;11(2):219-34.
71. Pirmoradian M, Budamgunta H, Chingin K, Zhang B, Astorga-Wells J, Zubarev RA. Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol Cell Proteomics*. 2013 Nov;12(11):3330-8.
72. Qundos U, Drobin K, Mattsson C, Hong MG, Sjoberg R, Forsstrom B, et al. Affinity proteomics discovers decreased levels of AMFR in plasma from Osteoporosis patients. *Proteomics Clin Appl*. 2015 Feb 16.
73. Haggmark A, Mikus M, Mohsenchian A, Hong MG, Forsstrom B, Gajewska B, et al. Plasma profiling reveals three proteins associated to amyotrophic lateral sclerosis. *Ann Clin Transl Neurol*. 2014 Aug;1(8):544-53.

74. The Human Protein Atlas. <http://www.proteinatlas.org/>.
75. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015 Jan 23;347(6220):1260419.
76. Bystrom S, Ayoglu B, Haggmark A, Mitsios N, Hong MG, Drobin K, et al. Affinity proteomic profiling of plasma, cerebrospinal fluid, and brain tissue within multiple sclerosis. *J Proteome Res*. 2014 Nov 7;13(11):4607-19.
77. Neiman M, Fredolini C, Johansson H, Lehtio J, Nygren PA, Uhlen M, et al. Selectivity analysis of single binder assays used in plasma protein profiling. *Proteomics*. 2013 Dec;13(23-24):3406-10.
78. Schneck NA, Lowenthal M, Phinney K, Lee SB. Current trends in magnetic particle enrichment for mass spectrometry-based analysis of cardiovascular protein biomarkers. *Nanomedicine (Lond)*. 2015 Feb;10(3):433-46.
79. Lin D, Alborn WE, Slebos RJ, Liebler DC. Comparison of protein immunoprecipitation-multiple reaction monitoring with ELISA for assay of biomarker candidates in plasma. *J Proteome Res*. 2013 Dec 6;12(12):5996-6003.
80. Bergendal A, Bremme K, Hedenmalm K, Larfars G, Odeberg J, Persson I, et al. Risk factors for venous thromboembolism in pre- and postmenopausal women. *Thrombosis research*. 2012 Oct;130(4):596-601.
81. Ljungqvist M, Sonnevi K, Bergendal A, Holmstrom M, Kieler H, Larfars G. Risk factors for recurrent venous thromboembolism in young and middle-aged women. *Thrombosis research*. 2014 Feb 26.
82. National Inpatient Register [Patientregistret] Socialstyrelsen [National Board of Health and Welfare]. <http://www.socialstyrelsen.se/register/halsodataregister/patientregistret>.
83. Cause of Death Register [Dödsorsaksregistret]. Socialstyrelsen [National Board of Health and Welfare]. <http://www.socialstyrelsen.se/register/dodsorsaksregistret>.
84. Swedish Tax Agency register [Skatteverket]. <https://www.skatteverket.se/>.
85. Zhu T, Carcaillon L, Martinez I, Cambou JP, Kyndt X, Guillot K, et al. Association of influenza vaccination with reduced risk of venous thromboembolism. *Thrombosis and Haemostasis*. 2009 Dec;102(6):1259-64.
86. Tregouet DA, Heath S, Saut N, Biron-Andreani C, Schved JF, Pernod G, et al. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. *Blood*. 2009 May 21;113(21):5298-303.
87. Holmberg K, Persson ML, Uhlen M, Odeberg J. Pyrosequencing analysis of thrombosis-associated risk markers. *Clinical chemistry*. 2005 Aug;51(8):1549-52.
88. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007 Feb 20;115(7):928-35.
89. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*. 2007 Sep;81(3):559-75.
90. Tregouet DA, Garelle V. A new JAVA interface implementation of THESIAS: testing haplotype effects in association studies. *Bioinformatics*. 2007 Apr 15;23(8):1038-9.

91. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics *Journal of Computational and Graphical Statistics*. 1996;5 (3):299-314
92. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999 Jan;10(1):37-48.
93. Bergendal A, Persson I, Odeberg J, Sundstrom A, Holmstrom M, Schulman S, et al. Association of venous thromboembolism with hormonal contraception and thrombophilic genotypes. *Obstet Gynecol*. 2014 Sep;124(3):600-9.
94. Heit JA, Armasu SM, Asmann YW, Cunningham JM, Matsumoto ME, Petterson TM, et al. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *Journal of thrombosis and haemostasis : JTH*. 2012 Aug;10(8):1521-31.
95. Zakai NA, Judd SE, Alexander K, McClure LA, Kissela BM, Howard G, et al. ABO blood type and stroke risk: the REasons for Geographic And Racial Differences in Stroke Study. *Journal of thrombosis and haemostasis : JTH*. 2014 Apr;12(4):564-70.
96. He M, Wolpin B, Rexrode K, Manson JE, Rimm E, Hu FB, et al. ABO blood group and risk of coronary heart disease in two prospective cohort studies. *Arteriosclerosis, thrombosis, and vascular biology*. 2012 Sep;32(9):2314-20.
97. de Haan HG, Bezemer ID, Doggen CJ, Le Cessie S, Reitsma PH, Arellano AR, et al. Multiple SNP testing improves risk prediction of first venous thrombosis. *Blood*. 2012 May 14.
98. Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. *Journal of thrombosis and haemostasis : JTH*. 2013 Jun;11 Suppl 1:129-41.
99. Cook NR. Clinically relevant measures of fit? A note of caution. *American journal of epidemiology*. 2012 Sep 15;176(6):488-91.
100. Sonnevi K, Bergendal A, Adami J, Larfars G, Kieler H. Self-reported family history in estimating the risk of hormone, surgery and cast related VTE in women. *Thrombosis research*. 2013 Aug;132(2):164-9.
101. Suchon P, Al Frouh F, Henneuse A, Ibrahim M, Brunet D, Barthet MC, et al. Risk factors for venous thromboembolism in women under combined oral contraceptive. The PIL Genetic Risk Monitoring (PILGRIM) Study. *Thrombosis and Haemostasis*. 2015 Aug 20;114(6).
102. Kieler H, Persson I, Sundstrom A, Hedenmalm K, Odeberg J, Bergendal A. [Birth control pills, surgery and casting leads to high risk of venous thrombosis in women. Better prophylaxis is needed for surgical procedures, as shown in case-control study]. *Lakartidningen*. 2013 Dec 4-17;110(49-50):2233-7.
103. Lindqvist PG, Kublikas M, Dahlback B. Individual risk assessment of thrombosis in pregnancy. *Acta Obstet Gynecol Scand*. 2002 May;81(5):412-6.
104. Lindqvist PG, Hellgren M, Nord E, Nyman M. [Thrombosis prophylaxis decisions in pregnancy--simpler with risk scores]. *Lakartidningen*. 2006 May 3-9;103(18):1429-32.
105. Doggen CJ, Rosendaal FR, Meijers JC. Levels of intrinsic coagulation factors and the risk of myocardial infarction among men: Opposite and synergistic effects of factors XI and XII. *Blood*. 2006 Dec 15;108(13):4045-51.

106. Gailani D, Smith SB. Structural and functional features of factor XI. *J Thromb Haemost.* 2009 Jul;7 Suppl 1:75-8.
107. Meijers JC, Tekelenburg WL, Bouma BN, Bertina RM, Rosendaal FR. High levels of coagulation factor XI as a risk factor for venous thrombosis. *N Engl J Med.* 2000 Mar 9;342(10):696-701.
108. Tucker EI, Marzec UM, White TC, Hurst S, Rugonyi S, McCarty OJ, et al. Prevention of vascular graft occlusion and thrombus-associated thrombin generation by inhibition of factor XI. *Blood.* 2009 Jan 22;113(4):936-44.
109. Yang DT, Flanders MM, Kim H, Rodgers GM. Elevated factor XI activity levels are associated with an increased odds ratio for cerebrovascular events. *Am J Clin Pathol.* 2006 Sep;126(3):411-5.
110. Fjellstrom O, Akkaya S, Beisel HG, Eriksson PO, Erixon K, Gustafsson D, et al. Creating Novel Activated Factor XI Inhibitors through Fragment Based Lead Generation and Structure Aided Drug Design. *PLoS One.* 2015;10(1):e0113705.
111. Buller HR, Bethune C, Bhanot S, Gailani D, Monia BP, Raskob GE, et al. Factor XI antisense oligonucleotide for prevention of venous thrombosis. *The New England journal of medicine.* 2015 Jan 15;372(3):232-40.
112. Li Y, Bezemer ID, Rowland CM, Tong CH, Arellano AR, Catanese JJ, et al. Genetic variants associated with deep vein thrombosis: the F11 locus. *J Thromb Haemost.* 2009 Nov;7(11):1802-8.
113. Folsom AR, Tang W, Roetker NS, Heckbert SR, Cushman M, Pankow JS. Prospective study of circulating factor XI and incident venous thromboembolism: The Longitudinal Investigation of Thromboembolism Etiology (LITE). *American journal of hematology.* 2015 Aug 11.
114. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 2008 Dec 15;24(24):2938-9.
115. Rappaport SM. Biomarkers intersect with the exposome. *Biomarkers.* 2012 Sep;17(6):483-9.
116. Morange PE, Bezemer I, Saut N, Bare L, Burgos G, Brocheton J, et al. A follow-up study of a genome-wide association scan identifies a susceptibility locus for venous thrombosis on chromosome 6p24.1. *Am J Hum Genet.* 2010 Apr 9;86(4):592-5.
117. WHO International Programme on Chemical Safety Biomarkers in Risk Assessment: Validity and Validation. 2001.: <http://www.inchem.org/documents/ehc/ehc/ehc222.htm>.
118. Rothman KJ. *Epidemiology - An introduction.* 1 ed. New York: Oxford University Press; 2002. p. 201-3.
119. Morange PE, Suchon P, Tregouet DA. Genetics of Venous Thrombosis: update in 2015. *Thrombosis and Haemostasis.* 2015 Sep 10;114(1).
120. van Hylckama Vlieg A, Flinterman LE, Bare LA, Cannegieter SC, Reitsma PH, Arellano AR, et al. Genetic variations associated with recurrent venous thrombosis. *Circulation Cardiovascular genetics.* 2014 Dec;7(6):806-13.