Genetic variation and susceptibility to venous thrombosis : Etiology and risk assessment

Bezemer, I.D.

General Discussion

Chapter 8

GENERAL DISCUSSION

This thesis describes an effort to gain more insight into the genetic factors that influence venous thrombosis susceptibility. In this discussion section we focus on the implications of our research and answer the following questions: "Have we learned more about the etiology of the disease under study?" and "What is the clinical importance of the findings in terms of prediction and prevention?".

ETIOLOGY

Several study designs can be applied to study the association between genetics and disease [15]. The classic genetic association study evaluates the association between variants in a candidate gene and the disease of interest in an observational study of unrelated individuals. Because the studied individuals are unrelated, they are representative of a general population but the regions of LD are consequently small. The success of the candidate gene approach depends on having correctly predicted the identity of the risk-affecting gene, based on prior knowledge. A potential source of confounding is population stratification or admixture. A spurious association between an allele and disease may be observed when the study population includes ethnic groups that differ in both the specific allele frequency and disease frequency. In a genetic linkage study, genetic variants throughout the genome are tested for segregation with disease in a family. Because regions of LD in a family are large, a genetic linkage study needs fewer markers, and there can be no spurious associations due to admixture. A disadvantage of the linkage study design is that the results may not be applicable to the general population when the genetic variants identified are specific to the family studied. In addition, statistical power is limited by the number of families and affected family members. The modern genome-wide association study unites the advantages of both approaches: due to advanced technology, the number of variants tested can be large (starting with about 100,000 variants in 2004, and up to one million on current chips),

and since the variants are distributed throughout the genome, a candidate gene is no longer needed. For etiologic research, this hypothesis-free approach is attractive because it may lead to novel etiologic pathways involved in common diseases. Testing many variants simultaneously, however, increases the likelihood of false-positive findings. In the genome-wide association study as well as in the classic designs, the analysis is restricted to relatively common variants (minor allele frequency ≥0.01); rare variants will not be detected or their disease associations will not reach statistical significance. The SNP study described in Chapter 5 was designed as a genome-wide association study, although with a relatively small number of SNPs and therefore not with full genome coverage. While there was no exact hypothesis about the identity of the potential disease genes, we focused on potentially causal SNPs. The SNPs selected were mainly missense, located in coding regions. The chance of false-positive findings was reduced by including two replication steps in which only those SNPs associated in the previous steps were tested, and by calculating the false discovery rate in the second replication step.

Eighteen SNPs of 19,682 tested were associated with venous thrombosis in the LETS and MEGA-1 studies. In validating these associations in MEGA-2, priority was given to the most likely candidates. Seven of nine SNP associations replicated and for four of these we have a coagulation-cascade hypothesis. The nine remaining SNPs from LETS and MEGA-1 were included in Chapter 7. Two of these were linked to the factor V Leiden mutation and one was linked to another SNP (rs4524) in *F5*. The other six SNPs were genotyped in MEGA-2 and described in Chapter 7, but only rs4525 in *F5* replicated. Population stratification is not likely to have confounded our findings. In LETS, no information on ethnicity was collected but MEGA participants were recruited from the same population as LETS albeit 10 years later. Ninety percent of MEGA participants had both parents born in northwestern Europe; restricting the analyses to this 90% of MEGA did not modify our results.

Once an association is confirmed, the question remains whether the observed association is causal. Specific to genetic studies is the issue of linkage: an

associated SNP might be just a marker for the true causal variant in the same gene, in another gene or within a regulatory region. The analyses described in Chapter 5 and 6 aimed to determine whether the original SNPs in *CYP4V2* and *F9*, both identified in the SNP association study, are linked to stronger variants that are thus more likely to be causal. For *CYP4V2* we found additional linked variants in *KLKB1* and *F11* and each might be involved in the etiology of venous thrombosis; in *F9* we did not find variants that were more strongly associated than *F9* Malmö.

To answer the question about etiology: some SNP associations confirmed our current knowledge about coagulation and others might lead to new knowledge. The newly identified SNPs presented in this thesis are listed in the Table.

Potentially novel thrombosis-susceptibility genes are *CYP4V2*, *RGS7*, *NR1I2* and *NAT8B*. The function of the protein encoded by *CYP4V2* is not known, but since it is a member of the cytochrome P450 family 4, it is hypothesized to be involved in lipid metabolism [175]. Dyslipoproteinemia is an established risk factor for atherosclerosis and arterial thrombosis, and might also be involved in venous thrombosis [208-211]. The gene product of *NR1I2* is known as the pregnane X receptor and regulates the expression of p450 enzymes, especially cytochrome P450-3A expression in response to a wide variety of xenobiotics and has a critical role in mediating drug-drug interactions [212]. In this context, we might speculate that *NR1I2* and *CYP4V2* are involved in the same pathway leading to venous thrombosis; however, it is not known if the pregnane X receptor also regulates the expression of *CYP4V2*.
RGS7 encodes regulator of G protein signaling 7 and is expressed primarily in brain tissue [213]. Its potential functional role in venous thrombosis is not known. It may also well be that the CYP4V2 variant is only associated with thrombotic risk due to its association to prekallikrein and FXI variants.

Table. Summary of newly identified SNPs presented in this thesis.

SNPs		Location in gene	Amino acid change	(Proposed) protein function
Potentially novel thrombosis-susceptibility genes				
rs13146272	*CYP4V2*	exon 6	Gln259Lys	lipid metabolism
rs670659	*RGS7*	intron 3		brain signaling
rs1523127	*NR1I2*	5′ UTR		drug-drug interactions
rs2001490	*NAT8B*	exon 1	Gly112Ala	blood pressure
Coagulation related genes				
rs1613662	*GP6*	exon 5	Pro219Ser	platelet activation
rs3087505	*KLKB1*	3′UTR		FXI activation
Apparent candidate genes				
rs2227589	*SERPINC1*	intron 1		coagulation cascade
rs6048	*F9*	exon 6	Thr194Ala	coagulation cascade
rs4524	*F5*	exon 13	Lys858Arg	coagulation cascade
rs3756008	*F11*	5′UTR		coagulation cascade
rs2036914	*F11*	intron 2		coagulation cascade
rs4253418	*F11*	intron 7		coagulation cascade

Abbreviations: UTR = Untranslated region

NAT8B is a duplicate gene of *NAT8*, which is expressed in kidney and liver and has been suggested to be involved in the regulation of blood pressure and kidney function [214]. *NAT8B* is expected to have similar function. Regulation of blood pressure has been reported to be linked to the expression of factor VIII [215], which might be the intermediate pathway through which *NAT8B* is associated with risk of venous thrombosis. However, *NAT8B* seems to be inactive in humans due to two stop codons [216]. Whether *NAT8B* is polymorphic and active in some individuals is not clear. Alternatively, the association with venous thrombosis may be due to LD with, or regulation of transcription of *NAT8* or another gene.

The SNPs in *GP6* and *KLKB1* are located in genes that are linked to coagulation, but not much is known about their role in venous thrombosis specifically. The SNP in *GP6* is well known to be associated with GPVI expression, platelet aggregation and thrombin generation [217]. Some evidence exists about an association between the *GP6* SNP and arterial thrombosis [218-222]. However, nothing is known about the relation between the intermediate, GPVI, and venous thrombosis. We only know that aspirin treatment, which inhibits platelet function in general, is weakly protective [223,224]. Since GPVI is a platelet collagen receptor, further research should be directed towards situations in which collagen exposure plays a role, such as surgery-related venous thrombosis. Prekallikrein, the protein encoded by *KLKB1*, is part of the contact system of blood coagulation and in complex with factor XII and high-molecular-weight kininogen it activates factor XI. Deficiency of prekallikrein does not lead to bleeding symptoms but, on the contrary, might lead to hypercoagulability due to impaired fibrinolysis [225]. Data on the association between prekallikrein levels and venous thrombosis are limited but so far do not suggest an effect on venous thrombosis risk [176].

Finally, the SNPs in *SERPINC1* (antithrombin), *F5*, *F9* and *F11* are located in apparent candidate genes with known function in the etiology of venous thrombosis. We do not yet know *how* these SNPs (or linked causal variants) affect risk of venous thrombosis. Sequence alterations might influence transcription efficiency or splicing, and amino acid substitutions might alter protein structure or the ability to be activated or inactivated by other proteins. The biological mechanism underlying an association between a genetic variant and disease can never be fully understood using epidemiologic methods alone. Associations between genetic variants and intermediate phenotypes that are in turn associated with disease provide biological support for the association, such as the SNPs in *F11*. Or they might point out that maybe the obvious biologic pathway is not the pathway through which the genetic variant influences disease risk, which might be the case for *F9* Malmö. Full understanding of the biology underlying an association between a genetic variant and disease requires fundamental laboratory research.

RISK PREDICTION AND PREVENTION

For risk prediction, we do not need to fully understand the biology of a disease, although prediction will be more accurate for a causal variant than when using a surrogate. Recent developments in genetic research have raised hopes for disease prediction based on genetic profiling. However, as the SNP associations identified in genome-wide association studies are generally weak and as common complex diseases are caused by the simultaneous action of many genes and environmental triggers, the predictive value of genetic profiling remains unclear.

Chapter 2 evaluates the value of family history as a disease predictor in venous thrombosis. When we compared the family history (positive or negative) with the presence of classical genetic risk factors, we saw only modest correlation. Apart from the conclusion that more genetic risk factors for venous thrombosis are to be identified, these results show that disease prediction must be based on multiple factors, both genetic and acquired. The family history allows us to get an idea of thrombosis susceptibility without the need to know the underlying genetic factors.

Another advantage of family history as a predictor in clinical practice is that assessment is easy and cheap through an interview. The family history could guide decisions on prophylaxis in high-risk situations, such as immobilization due to a plaster cast, or avoiding high-risk situations when possible, such as oral contraceptive use. A disadvantage, in terms of prediction accuracy, is that the family history is a surrogate measure for the real thrombosis susceptibility. It measures susceptibility of the relative, which is just a marker of an individual's own susceptibility as the individual may or may not have inherited the susceptibility alleles. In addition, the amount of information depends on family size.

In Chapter 7 we explored whether we could discriminate patients with venous thrombosis from control subjects, based on the genetic profile. This

profile was made of genetic variants described in Chapter 3 and 5. The genetic profiles in patients and control subjects demonstrated a graded effect of the number of genetic risk variants on the risk of venous thrombosis in the total study population as well as in high-risk groups. These high-risk groups are the first and probably only feasible target for genetic profiling in order to predict risk of venous thrombosis. In recent studies of genetic profiling, a comparison is often made between the discriminative accuracy of environmental triggers and the genetic profile in the total study population [200-202]. For venous thrombosis, we should rather use the environmental triggers to identify high-risk situations and then apply targeted genetic profiling instead of comparing the predictive accuracies in a general population. This is because in a general population the incidence of venous thrombosis and the prevalence of high-risk profiles are presumably too low to make genetic profiling cost-effective. As suggested earlier for the family history, the genetic profile can guide decisions on prophylaxis in or avoidance of high-risk situations. The advantage of the genetic profile over the family history is that no longer a surrogate (the relative) is measured. Therefore, the genetic profile should in theory perform better than the family history in discrimination of patients from control subjects when the genetic profile includes all or sufficient genetic predictors.

To answer the question about the clinical use of our findings: the work presented in this thesis has added to the knowledge of genetic risk factors for venous thrombosis, but risk prediction based on genetic profiling is not yet feasible. More genetic variants should be added to the genetic profile to improve discrimination between those who will and those who will not develop venous thrombosis. Eventually, the genetic profile may allow us to stratify individuals according to their susceptibility and target prevention of venous thrombosis. Until that time, the family history and triggers from the environment remain most useful in clinical practice.