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

Bezemer, I.D.

Introduction

Chapter 1

INTRODUCTION

Venous thrombosis is the result of excessive blood coagulation in veins, most frequently in the deep veins of the leg. The thrombus impairs or obstructs the blood flow which leads to swelling of the affected limb, pain and red discoloration. Besides painful and disabling, venous thrombosis can be life-threatening when part of the thrombus breaks off. Via the inferior caval vein the thrombus travels to the lungs where it may obstruct branches of the lung artery; a pulmonary embolism. Venous thrombosis occurs in about one in every 1000 persons per year [1]. About two thirds of patients have deep vein thrombosis of the leg and one third has pulmonary embolism with or without concurrent deep vein thrombosis of the leg [1-3].

The traditional model for classification of the pathology of venous thrombosis is known as 'Virchow's triad', after the work by 19th century pathologist Rudolf Virchow [4,5]. According to the triad, venous thrombosis occurs as a result of (1) alterations in blood flow, (2) endothelial injury, or (3) alterations in blood constitution. Another classification of causes is often made into (1) genetic causes and (2) acquired (or environmental) causes: some individuals have a strong genetic predisposition to develop venous thrombosis while others experience venous thrombosis only after environmental triggers such as long-term immobilization, oral contraceptive use or advanced age [6]. In most cases, venous thrombosis presumably occurs after an environmental trigger on a background of increased susceptibility. Figure 1 (from [6]) illustrates how these risk factors might relate to the occurrence of venous thrombosis in an individual. The factor V Leiden mutation here represents genetic predisposition, which remains constant during life. Depending on the genetic make-up of an individual, the "genetic predisposition level" will be higher or lower. Combined with increasing risk of venous thrombosis with advancing age and risk due to the presence of possible environmental triggers, an individual's "thrombosis potential" might exceed the threshold for developing venous thrombosis.

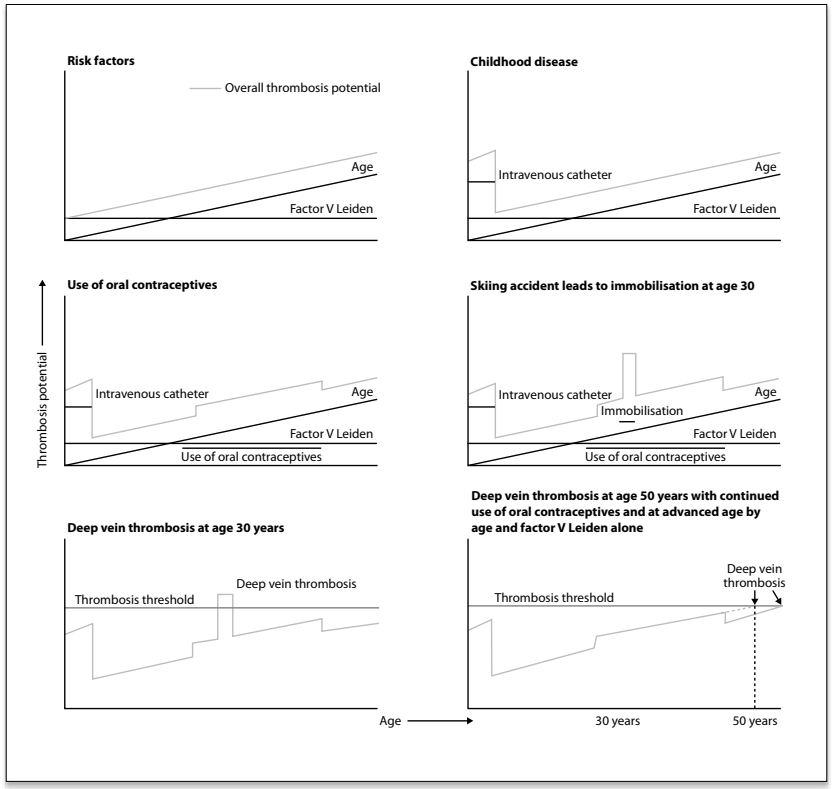

Figure 1: Models of thrombosis risk. In each panel, the figure shows the thrombosis (black) potential of each risk factor present during an individual's life and the resultant thrombosis potential (red). (Reprinted from The Lancet 6)

Genetic risk factors are generally classified as 'alterations in blood constitution', although they could also involve the vessel wall or stasis. A range of pro- and anticoagulant proteins act in the process that leads to thrombus formation; genetic alterations in any of these proteins may alter their level or function. For example, deficiencies of the anticoagulant proteins antithrombin, protein C and protein S increase the risk of venous thrombosis, as do increased levels of the procoagulant proteins fibrinogen, prothrombin, factor VIII, factor IX and factor XI [7].

The search for genes involved in common diseases was facilitated by the rapid evolution of the field of genetics in the last decade. With the completion of the Human Genome Project in 2003, a reference sequence of the 3 billion base pairs in the human genome has become publicly available (www.ornl.gov/sci/techresources/Human_Genome/home.shtml). In 2002, The International HapMap Project was initiated with the goal to make a catalogue of genetic variation between individuals [8]. The HapMap database, together with advances in genotyping technology, made it possible to perform genome-wide association studies for common diseases like type 2 diabetes [9-11] and breast cancer [12-14]. A genome-wide association study measures genetic variants throughout the genome and links them to the occurrence of disease [15].

A main source of genetic variation between individuals are single nucleotide polymorphisms (SNPs). In a SNP, two (for some SNPs three) allele options exist at one nucleotide position, for example an A-allele or a G-allele. The allele that is most prevalent in a population is called the major allele; the other is the minor allele. About 10 million "common" SNPs (minor allele frequency ≥0.01) exist in the human genome [16]; on average one in every 300 bases is a common SNP. SNPs that are located on the same chromosome are usually inherited together. This association between adjacent SNPs is known as linkage disequilibrium (LD), as opposed to the random combination (equilibrium) of different chromosomes during meiosis. Because of LD, it is possible to statistically predict the status of one SNP by genotyping a nearby SNP. SNPs that are genotyped to predict the status of other SNPs are known as "tag SNPs". However, even for SNPs on the same chromosome, loss of LD occurs. During meiosis, two homologous chromosomes often exchange sections. This may result in recombination of DNA stretches of the two chromosomes, which decreases the degree of linkage between the nucleotides in a population. The larger the distance between two nucleotides, the higher the likelihood of recombination between the two positions and the lower the degree of LD. The principle of LD is the rationale for haplotype analysis, and the construction of the HapMap and similar databases. Knowledge of LD patterns in a population allows to efficiently capture variation in a genetic

region by genotyping tag SNPs, instead of genotyping all known variants. These tag SNPs can be studied for association with disease.

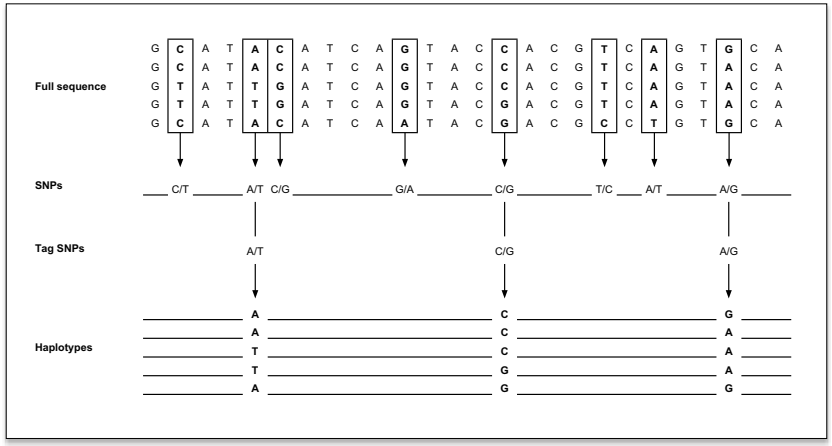

Figure 2: From full nucleotide sequence to tag SNPs and haplotypes. From 5 individuals the full sequence of a stretch of DNA (e.g. part of a gene) is shown. Most of the DNA sequence is identical between these individuals, but some positions vary: the SNPs. For some of these SNPs the allele is predicted from other SNPs because of LD. For example, the three SNPs on the left only occur in combinations CAC or TTG. With the allele of one SNP the other two are known as well. SNPs 4, 6 and 7 are also in perfect LD; their alleles are predicted from combining SNPs 5 and 8 (GG for the minor alleles, any other combination of SNP 5 and 8 for the major allele). A haplotype is the specific sequence of a DNA stretch and is identified by the tag SNPs. In the figure, each individual carries another haplotype.

STUDY POPULATIONS

Two case-control studies with similar design were used for the analysis of SNPs and the risk of venous thrombosis: the Leiden Thrombophilia Study and the Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis. The Northwick Park Heart Study-II was used for the analysis of the *F9* Malmö SNP and activation of coagulation factor IX in chapter 6.

Leiden Thrombophilia Study (LETS)

The LETS was designed to identify genetic risk factors for venous thrombosis. Four hundred seventy four patients with a first venous thrombosis of the leg between January 1988 and December 1992 were recruited from the anticoagulation clinics of Leiden, Amsterdam and Rotterdam. Patients with malignant disorders, known to strongly increase the risk of venous thrombosis, were excluded. For each patient an unrelated control subject was recruited, matched on age and sex and having no history of venous thrombosis and no malignancy. Participants filled in a standard questionnaire on potential risk factors for venous thrombosis, and a blood sample was taken at least three months after discontinuation of anticoagulant therapy. Details of the LETS have been described previously [17,18].

Multiple Environmental en Genetic Assessment of risk factors for venous thrombosis (MEGA study)

The MEGA study was set up a decade after the LETS, and aimed to study combinations of genetic and environmental risk factors. The MEGA study included 4930 patients with a first venous thrombosis of the leg or arm or pulmonary embolism who attended the anticoagulation clinics of Leiden, Amsterdam, Rotterdam, Den Haag, Utrecht or Amersfoort between March 1999 and August 2004. Patients were asked to bring their partners as control subject. In addition, a population-based control group, frequency-matched to the patients on age and sex, was recruited by random digit dialing between January 2002 and December 2004. Overall 6287 control subjects were included in the MEGA study. Participants filled in a standard questionnaire on potential risk factors for venous thrombosis. A blood sample was taken at least three months after discontinuation of anticoagulant therapy from patients who were diagnosed before June 1, 2002, and their partners, and from the random population controls. Participants who refused to or were unable to provide a blood sample were offered the option of providing a buccal swab sample. Patients who were diagnosed from June 1, 2002, onwards and their partners received a cotton swab along with their questionnaire for collecting buccal cells. Details of the MEGA study have been described previously [19,20].

Northwick Park Heart Study-II (NPHS-II)
The NPHS-II is a cohort study of middle-aged men and was set up to study the association between coagulation and coronary heart disease. In total 2951 men aged 50-61 years registered with nine general medical practices in England and Scotland were included in the study. Exclusion criteria were a history of unstable angina or myocardial infarction; regular anti-platelet or anticoagulant therapy; cerebrovascular disease; malignancy; conditions exposing staff to risk or precluding informed consent. Details of the NPHS-II have been described previously [21].

OUTLINE OF THIS THESIS

The aim of the research presented in this thesis is to identify and evaluate common genetic variants that contribute to genetic susceptibility to venous thrombosis. We also studied the clinical importance of genetic variants in prediction of venous thrombosis.

Chapter 2 addresses the question whether the classical parameter to determine genetic predisposition, family history, remains of use now genetic predisposition can be measured at the molecular level.
In **Chapter 3** the recent history of genetic research in venous thrombosis is summarized; it describes genetic variants that were identified in the last decade and aims to put the findings in perspective. One of the genetic variants of which its involvement in the etiology of venous thrombosis has been long debated is the MTHFR 677 C>T SNP. **Chapter 4** describes the association between this SNP and venous thrombosis in the MEGA study.
Chapter 5 reports of a large-scale SNP association study in the LETS and MEGA studies. The study aimed to identify new genetic variants that contribute to the risk of venous thrombosis. Because the study design does not address the question of whether a single SNP association is causal, the chapter zooms in on the most strongly associated SNP in the *CYP4V2* gene. In the region of this SNP, fine-scale genotyping was performed in order

to determine whether the association was due to linkage to another, more strongly associated and potentially causal SNP. The study also aimed to gain insight into the mechanistic aspect of the association. On one of the other SNPs from the SNP association study, we reported in more detail in **Chapter 6**. The SNP in the *F9* gene, also known as *F9* Malmö, was studied for linkage with nearby SNPs, and for association with coagulation factor IX levels (LETS and MEGA) and factor IX activation (NPHS-II).
Finally, in **Chapter 7**, we incorporated knowledge from previous and present research into a model that aims to predict the risk of venous thrombosis based on associated SNPs. We aimed to explore to what extent we can predict venous thrombosis using the currently known thrombosis-associated SNPs.