Journal of Thrombosis and Haemostasis, **11**: 1474–1484 DOI: 10.1111/jth.12313

ORIGINAL ARTICLE

Genome-wide linkage scan in affected sibling pairs identifies novel susceptibility region for venous thromboembolism: Genetics In Familial Thrombosis study

M.C.H. DE VISSER,* R. VAN MINKELEN,* V. VAN MARION,* M. DEN HEIJER,†‡ J. EIKENBOOM,* H.L. VOS,* P.E. SLAGBOOM,§ J.J. HOUWING-DUISTERMAAT,¶ F.R. ROSENDAAL*‡ and R.M. BERTINA*

**Einthoven Laboratory for Experimental Vascular Medicine, Department of Thrombosis and Hemostasis, Leiden University Medical Center, Leiden; †Department of Endocrinology, VU University Medical Center, Amsterdam; ‡Department of Clinical Epidemiology; §Department of Molecular Epidemiology; and ¶Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, the Netherlands*

To cite this article: de Visser MCH, van Minkelen R, van Marion V, den Heijer M, Eikenboom J, Vos HL, Slagboom PE, Houwing-Duistermaat JJ, Rosendaal FR, Bertina RM. Genome-wide linkage scan in affected sibling pairs identifies novel susceptibility region for venous thromboembolism: Genetics in Familial Thrombosis study. *J Thromb Haemost* 2013; **11**: 1474–84.

Summary. *Background:* Venous thromboembolism (VTE) is a multicausal disorder involving environmental and genetic risk factors. In many thrombophilic families the clustering of thrombotic events cannot be explained by known genetic risk factors, indicating that some remain to be discovered. *Objectives:* We aimed to identify novel thrombosis susceptibility alleles in a large panel of small thrombophilic families: the Genetics In Familial Thrombosis (GIFT) study. *Patients/Methods:* In the GIFT study, 201 families were recruited consisting of 438 siblings with an objectively confirmed VTE at a young age. Multipoint linkage analysis (402 SSR markers) and fine mapping were performed, followed by genotyping of tagging SNPs in positional candidate genes. *Results:* Established genetic risk factors such as factor V Leiden, ABO blood group non-O, prothrombin 20210A, fibrinogen gamma 10034T and deficiencies of antithrombin, protein C and protein S were more frequent in GIFT patients than in unselected VTE patients. Linkage supported the presence of novel thrombosis susceptibility loci on 7p21.3–22.2 (LOD score = 3.23) and Xq24–27.3 (LOD score = 1.95). Simulation analysis showed that the chr7 signal was genome-wide statistically significant ($P = 0.022$). Tagging SNPs ($n = 157$) in eight positional candidate genes (LOD drop 1.5 regions) were genotyped in GIFT patients and 332 healthy controls. Five chr7 SNPs associated with VTE. SNP *THSD7A* rs2074597 was responsible for part of the chr7 signal. *Conclusions:* The GIFT panel is rich in established genetic risk factors for VTE, but genetic factors remain unidentified in many families. Genome-wide linkage failed to identify the previously established genetic risk factors for VTE, but identified a novel VTE susceptibility locus on chr7.

Keywords: genetic linkage, siblings, thromboembolism, thrombophilia, thrombosis.

Correspondence: Marieke C.H. de Visser, Einthoven Laboratory for Experimental Vascular Medicine, Department of Thrombosis and Hemostasis, C2-R, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, the Netherlands.
Tel.: +31 71 5261893; fax: +31 71 5266755.
E-mail: mdvisser@lumc.nl

Received 15 October 2012
Manuscript handled by: M. Cushman
Final decision: D. Lane, 29 May 2013

Introduction

Venous thromboembolism (VTE) is a common disorder, with an annual incidence of about 1 per 100 000 in childhood, rising to nearly 1 per 100 in old age. The most frequent clinical manifestations are deep vein thrombosis (DVT) of the leg and pulmonary embolism (PE). Rarely, thrombus formation occurs at other locations (the upper extremities, liver, cerebral sinus, retina and mesentery). Major outcomes of VTE are death, recurrence, post-thrombotic syndrome and major bleeding due to anticoagulant treatment.

VTE is a multicausal disease involving multiple environmental and genetic risk factors [1]. Family and twin studies estimated the heritability of VTE to be about 50–60% [2–4], indicating that genetics play an important role. Two different genetic models for the molecular basis of common diseases exist. One model assumes that disease susceptibility results from the combined action of

many common disease-causing variants of minor effect ('common disease, common variant'). Genome-wide association studies (GWAS) are based on this hypothesis. The second model ('rare disease, rare variant') assumes the presence of rare, but severe, single gene mutations that are often private to a single family. The established genetic risk factors for VTE fit with both hypotheses [5]. Inherited deficiencies of the natural anticoagulants antithrombin, protein C and protein S are severe risk factors (relative risk for heterozygotes about 10) caused by rare and heterogeneous 'loss of function' mutations. In contrast, the factor (F) V Leiden, prothrombin 20210A and fibrinogen gamma (FGG) 10034T variants are more common and have a lower associated risk. The most common established genetic risk factor for VTE is ABO blood group non-O, which is present in more than 50% of individuals in many populations. In addition to these well-established genetic risk factors, several common low-risk variants have been discovered, which are all located in or near coagulation genes [6–8].

The established genetic risk factors for VTE do not fully explain the clustering of VTE in thrombophilia families and family history remains a risk indicator for VTE, regardless of the presence of known genetic factors [9,10]. Furthermore, several risk-enhancing intermediate phenotypes, mostly high levels of procoagulant factors, have been identified, of which the genetic determinants remain largely unknown. These data indicate that there is a substantial unidentified genetic contribution to VTE. As most of the hemostasis-related genes have been investigated extensively, missing susceptibility variants may be identified by hypothesis-free genome-wide approaches.

Based on the presence of both common and rare variants among the established thrombosis risk factors, we decided to conduct a genome-wide linkage study. Where previous linkage studies on VTE used one or more extended thrombosis families [2,11], we used a large panel of small thrombosis families, namely affected siblings with VTE at a young age, which were collected in the Genetics In Familial Thrombosis (GIFT) study. A linkage study looks for gene variants (detected by simple sequence repeats [SSRs] or single nucleotide polymorphisms [SNPs]) that are more often shared among affected siblings than expected, but does not require that different families share the same gene variant, as is the case for a GWAS. Therefore, it is a hypothesis-free approach that allows us to probe for both relatively common unique (FV Leiden like) variants as well as more heterogeneous loss-of-function variants.

In this paper we first describe the recruitment of a large panel of small thrombophilic families. Second, we show that established genetic risk factors are, as expected, more prevalent in these families than in unselected VTE patients. However, we also show that genetic factors are still missing in a large number of families. Third, genome-wide linkage with SSRs was performed to search for these missing factors. Two promising linkage regions were identified, which were narrowed by fine mapping with additional SSRs. SNPs in plausible candidate genes were tested for association with VTE and SNPs that were significantly associated with VTE were subsequently genotyped in a population-based case–control study on venous thrombosis.

Subjects and methods

Recruitment of GIFT population

All young VTE patients (≤ 45 years at the time of the thrombotic event) referred to one of the 29 participating Dutch anticoagulation clinics between 1 January 2001 and 1 January 2005 were contacted (n = 6483; Fig. 1). The (first or recurrent) event could have been a DVT (in a leg or arm), a PE or a rare presentation of venous thrombosis. Of the 4225 patients who responded positively, 298 patients with one or more siblings with VTE were eligible. After exclusion of 48 patients, 250 index patients entered the study together with their 293 affected siblings.

All participants completed a questionnaire about their VTE event(s) and the presence of provoking environmental factors. As provoking factors we considered surgery, immobilization (plaster cast, hospitalization without surgery, extended bed rest at home for more than 4 days), pregnancy and the post-partum period, all present in the 3 months preceding the VTE, malignancies (diagnosed within 5 years before or within 6 months after the event) and use of oral contraceptives or hormone replacement therapy at the time of the VTE.

DNA and plasma (snap-frozen) were collected and stored at −80 °C using standardized procedures. Parents were asked to participate by donating buccal swab DNA. When parents were unavailable, unaffected siblings were asked to participate.

Information on the tests that had been performed to diagnose the VTE event(s) was obtained by requesting discharge letters and radiology reports from general practitioners and hospitals. The provided information was reviewed independently by two physicians using a standardized approach (details available on request). At least one VTE diagnosis was objectively confirmed in 494 of the 543 subjects, resulting in 209 sibships with at least two siblings with a confirmed VTE.

Genotype results were used to verify familial relationships using the software program GRR (Graphical Representation of Relationships)[12], resulting in exclusion of 15 monozygotic twins and identification of two three-generation families.

Eventually 201 families, consisting of 203 sibships, were included in the GIFT study. DNA samples are available for all 438 affected siblings (201 index cases and 237 non-index cases), 227 parents and 118 unaffected siblings.

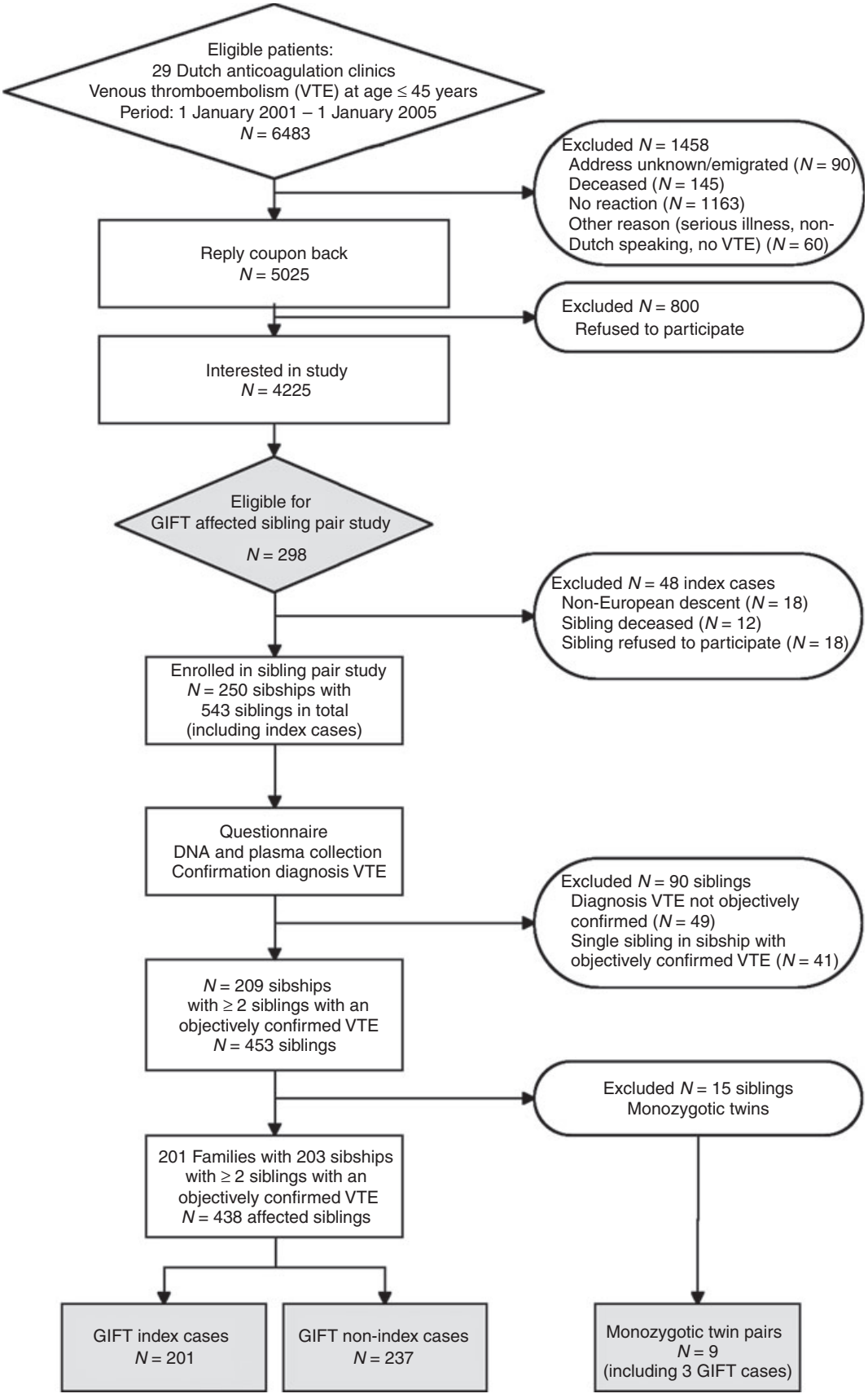

Fig. 1. Flowchart of patient inclusion.

Plasma and serum samples are available for 414 affected siblings (189 index and 225 non-index patients). At the time of venipuncture, 69 (36.5%) index patients and 73 (32.4%) non-index patients were using vitamin K antagonists (VKAs).

The study was approved by the Medical Ethics Committee of the Leiden University Medical Center (Leiden, the Netherlands). All participants gave written informed consent.

Established genetic risk factors

Factor V Leiden (rs6025), prothrombin 20210G>A (rs1799963), FGG 10034C>T (rs2066865) and rs8176719 [261G/delG] in the ABO blood group gene, discriminating blood group O and non-O, were genotyped using a 5′-nuclease/TaqMan assay (Applied Biosystems, Foster City, CA, USA).

Antithrombin activity and antigen levels of protein C and total and free protein S were measured in plasma and criteria for establishing a deficiency in both VKA-users and non-users were applied (see Table S1). Antithrombin activity was measured by chromogenic assay (Coamatic® Antithrombin, Chromogenix-Instrumentation Laboratory, Milan, Italy) on a STA-R coagulation analyzer (Diagnostica Stago, Asnières-sur-Seine, France), according to the manufacturer's protocol. The results were the mean of two measurements (1:40 and 1:80 dilutions). Protein C antigen (Ag) level and total protein S Ag level were measured by ELISA. All antibodies were obtained from Dako A/S (Glostrup, Denmark). Three different independently diluted plasma samples were tested. Free protein S Ag was measured by an enzyme-linked ligandsorbent assay (ELSA) according to Giri *et al.* [13] with some modifications as described before [14]. The results are expressed in U dL^{-1}, where 1 U dL^{-1} refers to the concentration of the relevant protein (antithrombin, protein C and protein S) in pooled normal plasma.

Genome-wide linkage scan and fine mapping

A genome-wide linkage scan was performed in all affected siblings using 402 SSR markers of the ABI Prism Linkage Mapping Set MD-10 ($n = 380$) or MD-5 ($n = 22$) and an ABI Prism DNA Analyzer 3700 or 3730 (Applied Biosystems). Average marker spacing was 9.3 centiMorgan (cM) and average heterozygosity was 0.79 (range 0.58–0.91). During fine mapping of interesting regions, affected siblings, parents and unaffected siblings were genotyped for 19 additional SSR markers of the ABI Prism Linkage Mapping Set HD-5 (12 chromosome X markers, six chromosome seven markers and D1S452 on chromosome 1 near *F5,* the gene coding for coagulation FV), resulting in an average marker spacing of about 4 cM.

Genotypes were analyzed using Genemapper Version 3.0 (Applied Biosystems) and independently checked by two operators. As part of quality control 5% of the samples were genotyped in duplicate. We considered the average discordance rate of 0.2% (range = 0–0.5%) between monozygous twin pairs ($n = 9$) to be an approximation of the genotyping error rate.

The position of the markers is given in deCODE cM, estimated via locally weighted linear regression (lo[w]ess) from the physical map positions of Build 35.1 and from published deCODE and Marshfield genetic map positions.

Average success rate per marker was 94.5% for the initial markers and 98.2% for the fine mapping markers. The average success rate per sample was 97.6% in the initial scan and 98.6% during fine mapping.

Linkage analysis

Non-parametric linkage (NPL) analysis based on the S_{all} statistic [15] was performed with MERLIN for all autosomes and with MINX for the X–chromosome [16]. Mendelian inconsistencies were identified with PEDSTATS [17], unlikely double recombinants were identified with MERLIN, and erroneous genotypes were removed with PEDWIPE [16]. The genome-wide significance level was estimated by performing a simulation analysis in 10 000 random datasets, which were generated with MERLIN using the same marker allele frequencies, missing data, marker spacing and family structures as used in the actual analyses. Subsequently, each dataset was analyzed with MERLIN. The probability (with 95% confidence interval (CI)) of observing a linkage signal equal to or higher than our maximum LOD score was calculated (i.e. genome-wide *P*–value = $n/10\,000$, in which n is the total number of simulations with a signal equal to or higher than the observed LOD score).

Following the power calculations of Risch [18] the power of our study with 273 sibling pairs to find a locus with a lambda sib of 1.75 is 87% for detecting a lod score exceeding 3.3% and 97% for a lod score exceeding 2.2.

Selection of candidate genes in linkage regions

Candidate genes were selected from the LOD-1.5 support intervals. For the chromosome 7p linkage peak (LOD score = 3.23), this interval is 15.3 cM wide, stretching from 3.2 to 11.6 Mb (NCBI database build 36.1). It contains about 80 (predicted) genes. For the chromosome Xq linkage peak (LOD score = 1.95), this interval is 24.4 cM wide, ranging from 120.7 to 140.1 Mb, and contains about 170 (predicted) genes. Eight candidate genes were selected that were hypothesized to potentially influence the risk of VTE. Selection criteria were based on a previously reported role in coagulation (*RAC1* and *F9*),

glycosylation (*C1GALT1* and *HS6ST2*) or platelet or endothelial related functions (*COL28A1*, *THSD7A* and *CD40LG*). *NXPH1* was selected because it is present at the location of the maximum LOD score. Only for *F9* and its protein factor IX was the association with thrombosis previously extensively investigated.

Selection and genotyping of SNPs in candidate genes

The online Genome Variation Server (GVS; http://gvs.gs.washington.edu/GVS/) was used to graphically display the linkage disequilibrium (LD) structure of the SNPs present in the eight genes and their flanking regions (dbSNP build 129 database). Genotypic data from Caucasians (CEU) from the HapMap project (http://hapmap.ncbi.nlm.nih.gov/) and from SeattleSNPs (http://pga.gs.washington.edu; *RAC1* and *F9*) were used. Haplotypes were identified for each gene using $r^2 > 0.8$ as the criterion for high correlation between two SNPs. In total, 157 haplotype tagging SNPs (htSNPs) were selected and genotyped by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry, using the Sequenom MassARRAY® iPLEX™ Platform (Sequenom, San Diego, CA, USA; Tables S2 and S3). About 7% of the samples were genotyped in duplicate as part of genotyping quality control. In total, 10 SNPs failed to genotype correctly. For the remaining 147 SNPs, the average success rate per SNP and per sample was > 97%. For all except eight SNPs the distribution of genotypes among control subjects (for X chromosome SNPs only in women) was in Hardy-Weinberg equilibrium. An extra check of the genotypic data of these SNPs did not indicate a systematic error in the iPLEX™ assays used.

Healthy subjects

The control group of healthy subjects was recruited through a general practice in The Hague (the Netherlands) and has previously been used in a case–control study on the causes of recurrent VTE [19,20]. Individuals (n = 2812), aged 20–90 years, were approached to take part in a health survey on risk factors of cardiovascular disease. In total, 532 individuals agreed to take part in the study. For the current study, all available DNA samples of Caucasian individuals without a history of VTE and cardiovascular disease were used (n = 332; 130 men, 202 women).

Leiden Thrombophilia Study

SNPs that appeared to be associated with VTE were genotyped in the Leiden Thrombophilia Study (LETS), a Dutch population-based case–control study on venous thrombosis, including 474 consecutive patients aged 18–70 years with a first deepvein thrombosis and 474 age- and sex-matched healthy controls [21].

Statistical analyses

To investigate whether an SNP was associated with VTE, we used a Cochran-Armitage trend test to test for differences in genotype distributions between the affected siblings and healthy controls. When computing the variance of the statistic, we took into account that genotypes of siblings are correlated [22,23]. An association given linkage analysis using an in-house written score statistic [24] was performed to test whether an SNP was in LD with a putative disease locus. In this test statistic the pairs that share two alleles identical by descent (IBD) are assigned more weight. Hence, when the IBD distribution explains part of the variance of the SNP genotypes (i.e. when the effect comes from the sibling pairs with IBD 2), this test is more powerful than the Cochran-Armitage trend test.

In the LETS study a Cochran-Armitage trend test was performed and allelic odds ratios (ORs) were calculated.

Results

VTE patients with one or more siblings who also had experienced a VTE were identified among almost 6500 consecutive young (aged 45 years or less) VTE patients (Fig. 1). Nine monozygous twin pairs with VTE were identified. Eventually, 201 families were included in the GIFT study, consisting of 203 sibships with two (n = 177), three (n = 20) or four (n = 6) siblings, each with at least one objectively confirmed VTE. Of these sibships, 41% consisted of only women, 13% consisted of only men and 46% of the sibships were mixed. General characteristics of the 438 GIFT siblings are summarized in Table 1. In 101 (50.2%) families at least one parent had a history of VTE.

Established genetic risk factors

Table 2 shows the frequencies of well-established genetic risk factors for VTE in GIFT. Factor V Leiden (36.8%) and ABO blood group non-O (82.6%) were much more prevalent in the 201 GIFT index patients than among 474 unselected VTE patients of the LETS study [25,26]. Prothrombin 20210A (7.0%), homozygosity for FGG 10034T (15.4%) and deficiencies of antithrombin (3.7%), protein C (5.3%) and protein S (7.9%) were also more frequent in GIFT than in unselected VTE patients [27–29].

In 165 out of 201 index patients all seven well-established genetic factors could be determined. In 7.3% (12/165) no established genetic risk factor was present. In 39.4% (n = 65), 43.0% (n = 71), 9.7% (n = 16) and 0.6% (n = 1) of 165 index patients one, two, three or four established genetic risk factors were present, respectively. When ABO blood group, which is often disregarded, was excluded, 38.8% (n = 64) of 165 index cases have no known defect and 49.7% (n = 82) have one established risk factor, whereas only 10.3% (n = 17) carry two factors and 1.2% (n = 2) carry three factors.

Table 1 Characteristics of 438 GIFT siblings

Women	279 (63.7)
First VTE	
Age at onset, years	
Men	37.2 (8.9)
Women	32.0 (9.0)
Type of VTE	
Deep vein thrombosis (DVT, leg or arm)	286 (65.3)
Pulmonary embolism (PE)	94 (21.5)
DVT + PE	51 (11.6)
Other presentations*	7 (1.6)
Provoked	310 (70.8)
Men	54 (34.0)
Women	256 (91.8)
Surgery, immobilization	123 (28.1)
Malignancy	9 (2.1)
Pregnancy or puerperium (% women)	51 (18.3)
Oral contraceptive use (% women)	187 (67.0)
Recurrent VTE	155 (35.4)
Men	67 (42.1)
Women	88 (31.5)
Age at study entry, years	
Men	43.2 (7.9)
Women	41.2 (8.1)

Continuous variables are denoted as mean (standard deviation); categorical variables are denoted as number (%).
*Sinus thrombosis, portal vein thrombosis and mesenteric vein thrombosis.

Table 2 Established genetic risk factors for VTE in GIFT index patients, control subjects and replication study

			Replication study [25–29]	
Classical genetic risk factor	% GIFT index patients (n = 201)	% controls (n = 332)	% LETS patients (n = 474)	% LETS controls (n = 474)
Factor V Leiden	36.8	6.7	19.5	3.0
Prothrombin 20210A	7.0	3.9	6.2	2.3
Antithrombin deficiency	3.7*	n/a	1.1–4.2†	0.2–1.9
Protein C deficiency	5.3*	n/a	2.7–4.6†	0.4–1.5
Protein S deficiency	7.9‡	n/a	1.2	0.7
ABO blood group non-O	82.6	60.3	70.9	57.1
FGG 10034T (homozygous)	15.4	n/a	12.2	6.0

n/a, not available; FGG, fibrinogen gamma.
*n = 189; no plasma available for 12 index patients.
†Frequency ranges due to different criteria used: single measurement or two measurements (antithrombin); single measurement, two measurements or a single measurement and the presence of a mutation (protein C).
‡n = 165; no plasma available for 12 index patients, while 24 index women who were pregnant or using oral contraceptives were excluded.

Within families, we looked at sharing of FV Leiden and prothrombin 20210A by affected siblings. In 24 (29.3%) of the 82 families with FV Leiden, the mutation was not present in all affected siblings. In seven (36.8%) of 19 families with prothrombin 20210A, not all affected siblings carried this mutation. The same observations were made for phenotypic deficiencies of protein C, protein S and antithrombin.

Initial genome-wide linkage scan and fine mapping

The initial linkage scan for VTE in 273 affected sibling pairs yielded two linkage signals with an LOD score above 1.5 (Table 3), namely on chromosome 7p (LOD score=2.28; $P = 0.0006$) and chromosome Xq (LOD score=1.63; $P = 0.003$). After fine mapping, the LOD score on chromosome 7 increased to 3.23 ($P = 0.00004$; Table 3), which was genome-wide statistically significant, because an LOD score of 3.23 or higher was observed only in 217/10 000 simulations ($P = 0.022$; 95% CI, 0.019–0.025). The maximum LOD score on chromosome X also increased (1.95; $P = 0.0014$) but did not reach genome-wide significance ($P = 0.345$; 95% CI, 0.336-0.354).

Factor V Leiden was highly prevalent in the GIFT siblings (Table 2). However, linkage analysis, including genotype data of affected siblings, parents and unaffected siblings for the FV Leiden SNP and closest repeat marker D1S452, showed an LOD score of only 0.45 ($P = 0.08$) at the location of the FV gene (*F5*). Also, no promising linkage signals were found at the locations of the other well-established genetic risk factors for VTE, which is not unexpected given the poor sharing of these traits within affected sibships.

Candidate genes

Eight plausible candidate genes were selected under the linkage peaks: *RAC1*, *C1GALT1*, *COL28A1*, *NXPH1*, *THSD7A* on chromosome 7 (Fig. 2) and *HS6ST2*, *CD40LG* and *F9* on chromosome X (Fig. 3). All 438 GIFT siblings and 332 healthy controls (Table 2) were genotyped for 157 htSNPs in these candidate genes and an association analysis was performed (Tables S2 and S3). Tagging SNPs associated with VTE were located on chromosome 7 in *RAC1* (rs836480, $P = 0.033$), *COL28A1* (rs17519199, $P = 0.031$), *NXPH1* (rs4141173, $P = 0.033$) and *THSD7A* (rs40244, $P = 0.046$; rs2074597, $P = 0.005$; Table 4). The overall sex-adjusted association test did not show association with VTE for SNPs on chromosome X.

Table 3 Chromosomal regions with an LOD score > 1.5 in initial genome-wide scan

	Initial scan		After fine mapping	
Chromosome	Position (cM)	LOD score	Position (cM)	LOD score
7p	14.75	2.28	15.62	3.23
Xq	134.2	1.63	140.59	1.95

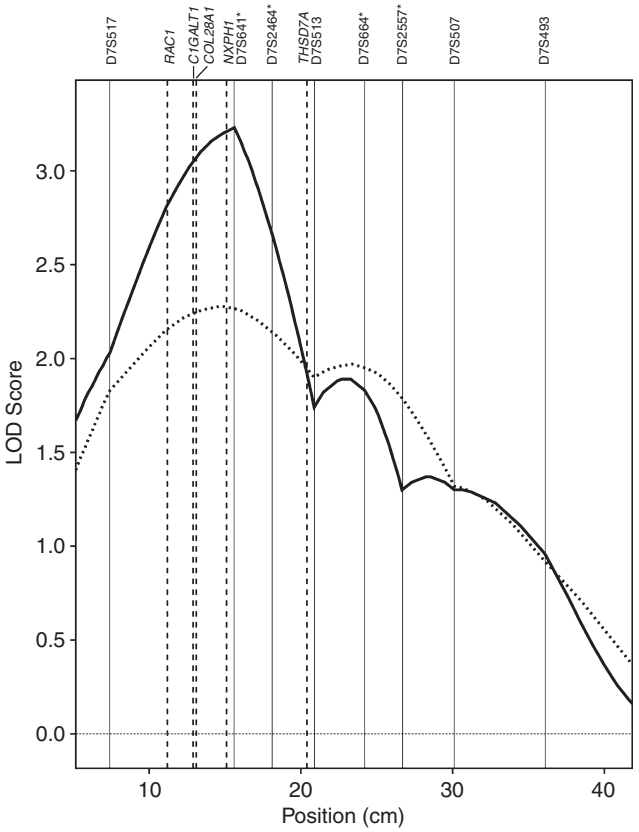

Fig. 2. Linkage analysis of chromosomal region 7p. Dashed line represents the results of initial linkage analysis. Solid line represents the results after fine mapping. Fine mapping markers are indicated with an asterisk.

Association given linkage analysis

To take the linkage signal on chromosome 7p into account, an association given linkage analysis was performed. For *THSD7A* rs2074597, the most significant SNP, the association became even more significant when linkage was taken into account: $P = 0.005$ for the association analysis vs. $P = 5.807\text{E-}05$ for the association given linkage analysis. This indicates that rs2074597 contributes to the linkage signal on chromosome 7p.

Leiden Thrombophilia Study

The five significant ($P < 0.05$) SNPs on chromosome 7 were genotyped in the LETS study (Table 4). For all SNPs, except for *THSD7A* rs40244, the direction of the effect of the minor allele was the same as observed in GIFT. However, none of the SNPs showed a significant association with venous thrombosis in this study. When the association analysis was performed in a selection of young LETS individuals ($\leq$ 45 years) similar results were obtained.

Discussion

In this report, we describe the first genome-wide linkage scan in affected sibling pairs with VTE. In total, 201 families with multiple siblings with at least one objectively confirmed VTE were included in GIFT. This large panel of small thrombophilia families is rich in previously established genetic risk factors for VTE. As genetic risk factors remain unidentified in a large percentage of families, the GIFT study offers an excellent opportunity to search for these missing genetic variants. Through genome-wide linkage analysis one novel susceptibility region for VTE was identified on chromosome 7p. Haplotype tagging SNP *THSD7A* rs2074597 explains part of the chromosomal 7p linkage peak.

To exclude the contribution of the strong non-genetic risk factor advanced age, patients were selected from a population of young VTE patients. The age limit of 45 years was chosen because the majority of patients from thrombophilic families develop their thrombosis before this age [30,31]. Only 6% of 5025 consecutive young VTE patients reported at least one sibling with VTE. In half of these sibships at least one parent had a history of VTE, demonstrating that our selection method indeed had resulted in the recruitment of true thrombophilia families.

Up till 1984, the year that the youngest GIFT sibling was born, about 10–13 twins were born per 1000 pregnancies in the Netherlands, of which 55–60% were monozygotic [32]. In GIFT, nine monozygotic (two male and

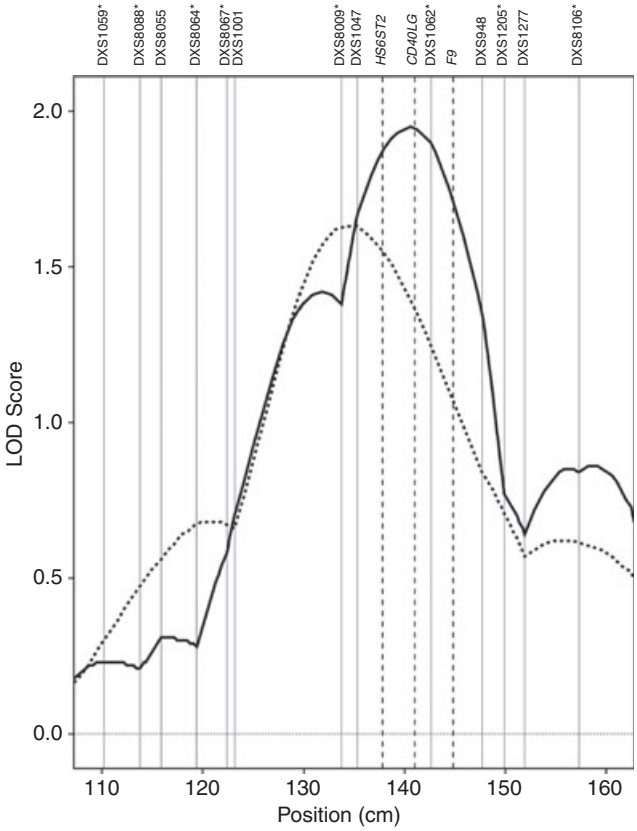

Fig. 3. Linkage analysis of chromosomal region Xq. Dashed line represents the results of initial linkage analysis. Solid line represents the results after fine mapping. Fine mapping markers are indicated with an asterisk.

seven female) and two dizygotic twin pairs were identified. This high frequency of identical twins (82% of all twins) points to an important contribution of genetic factors to the development of their VTE. Only in two monozygotic twin pairs were two established genetic factors (including blood group) found. Of the seven female twin pairs, at least one twin used oral contraceptives at the time of her first VTE. Three female twin pairs without any known classical genetic risk factor developed their first VTE while using oral contraceptives. A further search for genetic variants that interact with oral contraceptive use seems therefore warranted.

In almost half of GIFT index patients less than two of the seven established genetic risk factors (including ABO blood group non-O) were found, whereas it is generally assumed that familial thrombophilia is a multigenic disorder, in which at least two genetic defects segregate in the family [33–35]. When ABO blood group was excluded, almost 90% of index patients carried less than two classical defects. Currently, ABO blood group typing is not included in the panel of tests used to identify those considered at risk of VTE, whereas non-O blood group is a well-established and common risk factor for VTE (prevalence 57.1% in the Dutch population)[36], and synergistic effects on risk with FV Leiden have been reported [26,37]. The observed frequency of non-O in GIFT (82.6%) even exceeds the already high frequency in unselected patients (70.9%) [26], indicating that it plays a major role in familial thrombophilia and most likely in general VTE as well.

The high prevalence of FV Leiden in GIFT (36.8%) resembles incidences of 40–60% that have previously been reported in small panels of large thrombophilic families [9,38]. Despite this high frequency only a weak linkage signal was found at the location of *F5* (1q24.1, LOD score = 0.45). This can be explained by the observation that FV Leiden was not always present in all affected siblings within a sibship. Furthermore, a number of sibling pairs have inherited different FV Leiden alleles because they have two heterozygous parents or one homozygous parent. This observation suggests that novel genetic risk factors that are as common as FV Leiden will not be detected in GIFT using a genome-wide linkage scan. ABO blood group non-O (82.6% in GIFT) is also too common to be detected as a risk factor in a genome-wide linkage scan, which might explain why only a weak linkage signal (LOD = 0.27) was found at the location of the *ABO* locus (9q34.2). Whereas FV Leiden and ABO blood group non-O are too common to be detected via linkage analysis, deficiencies of antithrombin, protein C and protein S are still too rare in our study population to be detected via genome-wide linkage.

Table 4 SNPs associated at $P < 0.05$ in the GIFT study and their replication in the Leiden Thrombophilia Study (LETS)

					GIFT study				Leiden Thrombophilia Study			
Gene	SNP	Type	Major/ minor allele	MAF Hapmap CEU	MAF Affected sibs ($n = 438$)	MAF Controls ($n = 332$)	*P* value	OR (95% CI)	MAF Cases ($n = 471$)	MAF Controls ($n = 471$)	*P* value	OR (95% CI)
RAC1	rs836480	Intronic	G/T	0.473	0.424	0.486	0.033	0.78 (0.61–0.99)	0.462	0.485	0.303	0.91 (0.76–1.09)
COL28A1	rs17519199	Coding synonymous	T/C	0.058	0.068	0.039	0.031	1.79 (1.05–3.06)	0.057	0.050	0.478	1.16 (0.78–1.73)
NXPH1	rs4141173	Intronic	T/C	0.324	0.387	0.327	0.033	1.30 (1.01–1.66)	0.382	0.358	0.310	1.11 (0.92–1.33)
THSD7A	rs40244	Intronic	G/A	0.270	0.261	0.314	0.046	0.78 (0.60–0.99)	0.293	0.269	0.244	1.13 (0.92–1.38)
THSD7A	rs2074597	Intronic	G/T	0.181	0.211	0.151	0.005	1.57 (1.14–2.17)	0.201	0.187	0.453	1.09 (0.87–1.37)

MAF, minor allele frequency; OR, indicates allelic odds ratio for the minor allele.
P-value: Cochran-Armitage trend test; for calculation of *P*-value and 95% CI in GIFT study correlation of sibling genotypes was taken into account [22].

The initial genome-wide linkage scan in GIFT yielded two linkage signals with an LOD score higher than 1.5, namely at chromosomal regions 7p21.3–22.2 and Xq24–27.3. Fine mapping strengthened the support for linkage in both regions, but only the chromosome 7p linkage signal was genome-wide significant. In the two linkage regions no associations ($P<10^{-5}$) between SNPs and thrombosis-related traits are currently listed in the catalog of published genome-wide association studies [39].

Eight candidate genes were selected from the two identified thrombosis susceptibility regions. Genotyping of haplotype tagging SNPs in the 438 GIFT siblings and 332 healthy controls revealed five chromosome 7 SNPs that were associated with VTE. Association given linkage analysis indicated that *THSD7A* rs2074597 contributes to the linkage signal on chromosome 7p, because evidence for association increased when the linkage peak was taken into account. *THSD7A* (thrombospondin, type 1, domain containing 7A) is a large gene (458 kb) encoding for a protein involved in endothelial cell migration and angiogenesis and as such it might be involved in maintaining the integrity of the endothelial lining of the vessels. The functional variant(s) may be identified via resequencing of the linkage region. We are not aware of a VTE study with the same design as GIFT (young patients with a family history of VTE) that can be used for a true replication. Therefore we determined the five SNPs in the LETS study, consisting of consecutive patients with a first deep vein thrombosis. In this study no statistically significant results were obtained for the five SNPs, indicating that those SNPs are not relevant in consecutive patients with a first deep vein thrombosis. However, for four SNPs the direction of the effect of the minor allele was the same as in GIFT.

Genome-wide linkage was used before in the Vermont [11] and GAIT (Genetic Analysis of Idiopathic Thrombophilia) [2] studies. These previous linkage scans for VTE did not include the X chromosome and did not show evidence for linkage at the chromosome 7p region.

In this study we made two important observations. First, we showed that even among nuclear families, the prothrombotic genetic factors may differ from one individual to another. This suggests that within thrombophilic families there are multiple prothrombotic factors to 'choose' from. This can be understood if we realize that all thrombotic risk factors are relatively weak, necessitating the presence of multiple (genetic) factors to precipitate a VTE [1]. However, the chance that, for example, the same two risk alleles in the parents will be inherited by both siblings in a sibship of two is only 6.2%. Therefore it is likely that affected siblings are present in families with an 'excess' of genetic risk factors or with shared environmental risk factors (e.g. use of the contraceptive pill). Secondly, we showed that the thrombotic risk factors that will be found within a particular study will depend strongly on the design of that study. We did not

detect *F5* and *ABO* with genome-wide linkage analysis. However, with a GWAS approach *F5* and *ABO* would definitely have been detected in GIFT. It is obvious then that complementary study designs will be needed to complete the list of relevant prothrombotic variants and mutations at our disposal. In the three recent GWAS for VTE only variants in known thrombosis genes (*F5*, *ABO*, *FGG*, *F11*) reached genome-wide significance [40–42]. This suggests that the as yet unidentified genetic variants are either rare or are common, but with small effect sizes. The advent of rapid, low-cost whole genome sequencing will make it possible to further unravel the role of rare variants.

Addendum

M.C.H. de Visser, H.L. Vos, P.E. Slagboom, J.J. Houwing-Duistermaat, F.R. Rosendaal and R.M. Bertina: study concept and design. M.C.H. de Visser, R. van Minkelen, V. van Marion, J. Eikenboom and M. den Heijer: acquisition of data. M.C.H. de Visser, R. van Minkelen, H.L. Vos, J.J. Houwing-Duistermaat and R.M. Bertina: analysis and interpretation of data. M.C.H. de Visser, R. van Minkelen, H.L. Vos and R.M. Bertina: drafting of the manuscript. All authors approved the final manuscript.

Acknowledgements

This study was financially supported by the Netherlands Organisation for Scientific Research (NWO, grant 912-02-036), the Netherlands Heart Foundation (NHF grant 2005T55: Dr E. Dekker fellowship to MdV) and the Centre for Medical Systems Biology (CMSB), a centre of excellence approved by the Netherlands Genomics Initiative (NGI) and NWO. The LETS study was originally supported by NHF grant 89.063. We express our gratitude to all patients who participated in our study and thank the directors of the 29 anticoagulation clinics who made the recruitment of patients possible.

Disclosure of Conflicts of Interests

The authors state that they have no conflict of interests.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Criteria for deficiencies of antithrombin, protein C and protein S.
Table S2. Association analysis of SNPs in chromosomal region 7p.
Table S3. Association analysis of SNPs in chromosomal region Xp.

References

1 Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet* 1999; **353**: 1167–73.
2 Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, Coll I, Felices R, Stone W, Fontecuberta J, Blangero J. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Genetic Analysis of Idiopathic Thrombophilia. *Am J Hum Genet* 2000; **67**: 1452–9.
3 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; **14**: 328–32.
4 Heit JA, Phelps MA, Ward SA, Slusser JP, Petterson TM, De Andrade M. Familial segregation of venous thromboembolism. *J Thromb Haemost* 2004; **2**: 731–6.
5 Rosendaal FR, Reitsma PH. Genetics of venous thrombosis. *J Thromb Haemost* 2009; **7**(Suppl. 1): 301–4.
6 Smith NL, Hindorff LA, Heckbert SR, Lemaitre RN, Marciante KD, Rice K, Lumley T, Bis JC, Wiggins KL, Rosendaal FR, Psaty BM. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *JAMA* 2007; **297**: 489–98.
7 Bezemer ID, Bare LA, Doggen CJM, Arellano AR, Tong C, Rowland CM, Catanese J, Young BA, Reitsma PH, Devlin JJ, Rosendaal FR. Gene variants associated with deep vein thrombosis. *JAMA* 2008; **299**: 1306–14.
8 Bezemer ID, Bare LA, Arellano AR, Reitsma PH, Rosendaal FR. Updated analysis of gene variants associated with deep vein thrombosis. *JAMA* 2010; **303**: 421–2.
9 Bertina RM. Genetic approach to thrombophilia. *Thromb Haemost* 2001; **86**: 92–103.
10 Bezemer ID, van der Meer FJM, Eikenboom JC, Rosendaal FR, Doggen CJM. The value of family history as a risk indicator for venous thrombosis. *Arch Intern Med* 2009; **169**: 610–5.
11 Hasstedt SJ, Scott BT, Callas PW, Vossen CY, Rosendaal FR, Long GL, Bovill EG. Genome scan of venous thrombosis in a pedigree with protein C deficiency. *J Thromb Haemost* 2004; **2**: 868–73.
12 Abecasis GR, Cherny SS, Cookson WOC, Cardon LR. GRR: graphical representation of relationship errors. *Bioinformatics* 2001; **17**: 742–3.
13 Giri TK, Hillarp A, Hardig Y, Zöller B, Dahlbäck B. *Thromb Haemost* 1998; **79**: 767–72.
14 van Vliet HAAM, Bertina RM, Dahm AEA, Rosendaal FR, Rosing J, Sandset PM, Helmerhorst FM. *J Thromb Haemost* 2008; **6**: 346–51.
15 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–63.
16 Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
17 Wigginton JE, Abecasis GR. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 2005; **21**: 3445–7.
18 Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 1990; **46**: 229–41.
19 den Heijer M, Blom HJ, Gerrits WB, Rosendaal FR, Haak HL, Wijermans PW, Bos GM. Is hyperhomocysteinaemia a risk factor for recurrent venous thrombosis? *Lancet* 1995; **345**: 882–5.
20 Keijzer MBAJ, den Heijer M, Blom HJ, Bos GMJ, Willems HPJ, Gerrits WBJ, Rosendaal FR. Interaction between hyperhomocysteinemia, mutated methylenetetrahydrofolatereductase (MTHFR) and inherited thrombophilic factors in recurrent venous thrombosis. *Thromb Haemost* 2002; **88**: 723–8.

21 van der Meer FJM, Koster T, Vandenbroucke JP, Briët E, Rosendaal FR. The Leiden thrombophilia study (LETS). *Thromb Haemost* 1997; **78**: 631–5.
22 Slager SL, Schaid DJ. Evaluation of candidate genes in case-control studies: a statistical method to account for related subjects. *Am J Hum Genet* 2001; **68**: 1457–62.
23 Uh HW, Deelen J, Beekman M, Helmer Q, Rivadeneira F, Hottenga JJ, Boomsma DI, Hofman A, Uitterlinden AG, Slagboom PE, Böhringer S, Houwing-Duistermaat JJ. How to deal with the early GWAS data when imputing and combining different arrays is necessary. *Eur J Hum Genet* 2012; **20**: 572–6.
24 Houwing-Duistermaat JJ, Uh HW, van Houwelingen HC. A new score statistic to test for association given linkage in affected sibling pair-control designs. *BMC Proc* 2007; **1**(Suppl. 1): S39.
25 Rosendaal FR, Koster T, Vandenbroucke JP, Reitsma PH. High risk of thrombosis in patients homozygous for factor V Leiden (activated protein C resistance). *Blood* 1995; **85**: 1504–8.
26 Morelli VM, de Visser MCH, Vos HL, Bertina RM, Rosendaal FR. ABO blood group genotypes and the risk of venous thrombosis: effect of factor V Leiden. *J Thromb Haemost* 2005; **3**: 183–5.
27 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; **88**: 3698–703.
28 Uitte de Willige S, De Visser MCH, Houwing-Duistermaat JJ, Rosendaal FR, Vos HL, Bertina RM. Genetic variation in the fibrinogen gamma gene increases the risk for deep venous thrombosis by reducing plasma fibrinogen gamma' levels. *Blood* 2005; **106**:4176–83.
29 Koster T, Rosendaal FR, Briët E, van der Meer FJ, Colly LP, Trienekens PH, Poort SR, Reitsma PH, Vandenbroucke JP. Protein C deficiency in a controlled series of unselected outpatients: an infrequent but clear risk factor for venous thrombosis (Leiden Thrombophilia Study). *Blood* 1995; **85**: 2756–61.
30 Lensen RPM, Rosendaal FR, Koster T, Allaart CF, de Ronde H, Vandenbroucke JP, Reitsma PH, Bertina RM. Apparent different thrombotic tendency in patients with factor V Leiden and protein C deficiency due to selection of patients. *Blood* 1996; **88**: 4205–8.
31 Vossen CY, Conard J, Fontcuberta J, Makris M, van der Meer FJM, Pabinger I, Palareti G, Preston FE, Scharrer I, Souto JC, Svensson P, Walker ID, Rosendaal FR. Risk of a first venous thrombotic event in carriers of a familial thrombophilic defect. The European Prospective Cohort on Thrombophilia (EPCOT). *J Thromb Haemost* 2005; **3**: 459–64.
32 Statistics Netherlands. http://statline.cbs.nl. Accessed 5 March 2013.
33 Koeleman BPC, Reitsma PH, Bertina RM. Familial thrombophilia: a complex genetic disorder. *Semin Hematol* 1997; **34**: 256–64.
34 Bovill EG, Hasstedt SJ, Leppert MF, Long GL. Hereditary thrombophilia as a model for multigenic disease. *Thromb Haemost* 1999; **82**: 662–6.
35 Miletich JP. Thrombophilia as a multigenic disorder. *Semin Thromb Hemost* 1998; **24**(Suppl. 1): 13–20.
36 Wu O, Bayoumi N, Vickers MA, Clark P. ABO(H) blood groups and vascular disease: a systematic review and meta-analysis. *J Thromb Haemost* 2008; **6**: 62–9.
37 Ohira T, Cushman M, Tsai MY, Zhang Y, Heckbert SR, Zakai NA, Rosamond WD, Folsom AR. ABO blood group, other risk factors and incidence of venous thromboembolism: the Longitudinal Investigation of Thromboembolism Etiology (LITE). *J Thromb Haemost* 2007; **5**: 1455–61.
38 Lane DA, Mannucci PM, Bauer KA, Bertina RM, Bochkov NP, Boulyjenkov V, Chandy M, Dahlbäck B, Ginter EK, Miletich JP, Rosendaal FR, Seligsohn U. Inherited Thrombophilia: part 1. *Thromb Haemost* 1996; **76**: 651–62.
39 Hindorff LA, MacArthur J(European Bioinformatics Institute), Wise A, Junkins HA, Hall PN, Klemm AK, Manolio TA. A catalog of published genome-wide association studies. www.genome.gov/gwastudies. Accessed 5 march 2013.
40 Trégouët DA, Heath S, Saut N, Biron-Andreani C, Schved JF, Pernod G, Galan P, Drouet L, Zelenika D, Juhan-Vague I, Alessi MC, Tiret L, Lathrop M, Emmerich J, Morange PE. 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; **113**: 5298–303.
41 Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, Cohen W, Oudot-Mellakh T, Antoni G, Alessi MC, Zelenika D, Cambien F, Tiret L, Bertrand M, Dupuy AM, Letenneur L, Lathrop M, Emmerich J, Amouyel P, Trégouët DA, *et al.* Genetics of venous thrombosis: insights from a new genome wide association study. *PLoS ONE* 2011; **6**: e25581.
42 Heit JA, Armasu SM, Asmann YW, Cunningham JM, Matsumoto ME, Petterson TM, De Andrade M. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *J Thromb Haemost* 2012; **10**: 1521–31.