Genetic determinants of normal variation in coagulation factor (F) IX levels: genome-wide scan and examination of the FIX structural gene

M. Khachidze,* A. Buil, K. R. Viel,* S. Porter, D. Warren,§ D. K. Machia, # J. M. Soria, J. C. Souto, A. Ameri, M. Lathrop,§§ J. Blangero,§ J. Fontcuberta, S. T. Warren, L. Almasy§ and T. E. Howard* *Department of Pathology and Laboratory-Medicine, Emory University, Atlanta, GA, USA; Department of Hematology, Unitat d Hemostàsia i Trombosi, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; Geospiza Inc., Seattle, WA; §Department of Genetics, Southwest Foundation, San Antonio, TX; –Department of Pediatrics, Medical College of Georgia, Augusta, GA, USA; **Centre National de Genotypage, Evry, France; and Department of Human Genetics, Emory University, Atlanta, GA, USA


Summary. Background: High-normal and elevated plasma FIX activity (FIX:C) levels are associated with increased risk for venous- and possibly arterial-thrombosis. Objective: Because the broad normal range for FIX:C involves a substantial unknown genetic component, we sought to identify quantitative-trait loci (QTLs) for this medically important hemostasis trait. Methods: We performed a genome-wide screen and a resequencing-based variation scan of the known functional regions of every distinct FIX gene (F9) in the genetic analysis of idiopathic thrombophilia project (GAIT), a collection of 398 Spanish-Caucasians from 21 pedigrees. Results: We found no evidence for linkage (LOD scores <1.5) despite genotyping more than 540 uniformly-spaced microsatellites. We identified 27 candidate F9 polymorphisms, including three in cis-elements responsible for the increase in FIX:C that occurs with aging, but found no significant genotype-specific differences in mean FIX:C levels (P-values ≥ 0.11) despite evaluating every polymorphism in GAIT by marginal multivariable measured-genotype association analysis. Conclusions: The heritable component of interindividual FIX:C variability likely involves a collection of QTLs with modest effects that may reside in genes other than F9. Nevertheless, because the alleles of these 27 polymorphisms exhibited a low overall degree of linkage disequilibrium, we are currently defining their haplotypes to interrogate several highly-conserved non-exonic sequences and other F9 segments not examined here.

Keywords: association, linkage, linkage disequilibrium, quantitative-trait loci, single-nucleotide polymorphism, thrombosis.

Introduction

Thromboembolic disorders are a major cause of mortality and morbidity worldwide. The ability to identify accurately individuals at risk would advance management of these conditions through more effective prophylactic use of antithrombotic agents. The pathogenesis of thrombosis is complex, involving multiple genetic and environmental factors. While notable prothrombotic gene variants have been discovered [1,2], family studies indicate additional heritable factors remain unidentified [3,4].

Individuals with elevated levels of coagulation factor (F) IX, the plasma glycoprotein whose activity (FIX:C) is deficient in patients with hemophilia B, have long been suspected of being prothrombotic [5,6]. Recent studies have established that FIX levels constitute a quantitative-trait which, across its greater than 3-fold normal-range [6–9], correlates with the risk of venous and possibly arterial thrombosis [7,10,11]. Based on a study of 398 Caucasian-subjects from 21 extended Spanish pedigrees, known as the Genetic Analysis of Idiopathic Thrombophilia Project (GAIT), Souto et al. [12] estimated 39% heritability for FIX:C.

To identify the quantitative-trait loci (QTLs) underlying FIX:C variability, we examined the study-subjects in GAIT via a comprehensive strategy that included both genome-wide and F9 screens. We conducted the genome-wide screen first because this approach, which is based on linkage analysis, does not require that candidate genes be specified a priori. Next, we identified every polymorphism in known functional regions of F9, by resequencing all potentially distinct alleles in GAIT, and subsequently quantified their allelic contributions to FIX:C using...
association analysis. Finally, we examined F9 in 51 healthy unrelated individuals from three ethnic groups as a first step towards determining whether FIX QTLs may differ by ethnicity.

Materials and methods

Materials

All F9 amplicons were generated by polymerase chain reaction (PCR) using HotStarTaq-Master Mix (Qiagen, Valencia, CA, USA) and purified with AmpPure (Agencourt Biosciences; Beverly, MA, USA). Oligonucleotides for PCR, sequencing and amplicon-length polymorphism (ALP) analyses were made by Invitrogen (Carlsbad, CA, USA); all sequences were based on the genomic-DNA in GenBank accession K02402 and are available upon request. BigDye-terminator sequencing kits were purchased from ABI (Foster City, CA, USA). CleanSEQ was obtained from Agencourt. SeqMan (v5.03) was purchased from DNASTAR (Madison, WI, USA). Phred, a gift from the University of Washington, was used within the Finch Sequencing System, purchased from Geospiza (Seattle, WA, USA).

Subjects and phenotype

Two groups of human subjects, HS-1 and HS-2, were investigated in this institutional review board (IRB) approved study. HS-1 represents the 398 Spanish-Caucasian (SC) GAIT subjects (216 females, 182 males), aged from <1 to 88 years, who were used to identify FIX QTLs genome-wide and within F9. The recruitment, sampling, and phenotyping methods used for FIX:C (Fig. 1) and other measured-variables used in GAIT have been extensively described previously [3,12]. We set any FIX:C beyond four SD from the mean to missing.

The individuals of HS-2 (n = 229) were used in the F9 variation scan and comprised three subgroups (HS-2a, HS-2b, and HS-2c). HS-2a, which is also a subgroup of HS-1, contained all of the 100 GAIT founders who were able to be enrolled (44 females, 56 males). However, the total number of founders for the 21 GAIT pedigrees was 143 (55 females, 88 males). As any of the 54 potentially distinct F9 alleles from the 43 non-enrolled founders, which could be segregating in these families, would not be scanned if only the 100 founders in HS-2a were studied, we included HS-2b, a second GAIT subgroup with 78 non-founders (58 females, 20 males), for variation discovery. By also scanning the 136 X-chromosomes from these 78 descendents of the 43 non-enrolled founders, we ensured that all distinct alleles from non-sampled founders were examined and that all polymorphisms in known functional regions of the F9 from every GAIT subject were identified. HS-2c contained 51 unrelated non-thrombotic males from different ethnic groups as follows: 41 Caucasian-Americans (CA), six African-Americans (AA), two Japanese-Americans (JA) and two African-Pygmies. Archived DNA samples from these subjects, whose enrollment, including informed consent and biological sampling, were for an unrelated IRB-approved study [13], were the only specimens used.

Microsatellite genotypes and genotyping

The genome-wide screen, as previously described [3,12], was performed with 6.2 cM average coverage by genotyping every GAIT subject (HS-1) at 563 loci comprising 556 uniformly spaced microsatellites, six single-nucleotide polymorphisms (SNPs) and the ABO blood-group locus (Fig. 2). All microsatellites were genotyped by ALP-analysis using previously described assays [3]. Average heterozygosity was 0.79 for the 563 markers overall and 0.75 for the 23 X-linked microsatellites, four of which were located less than 10 Mb from F9. We also included a triallelic cryptic-dinucleotide repeat (cDR) in the F9 3'-UTR, a putative cis-element involved in the normal aging-related increase in F9 expression (Fig. 3) [14], and 12 microsatellites in or near candidate genes encoding functionally and/or structurally related proteins: prothrombin, FVII, FX, protein (PC, PS, FVIII, FV, thrombomodulin, and NADPH-dehydrogenase-quinone-1 (NQO1).

Linkage analysis

Genotypic data for all microsatellites were evaluated for violations of Mendelian inheritance with the program INFER in PEDSYS [15]. Discrepancies were checked for mistyping, and either corrected or excluded from further analysis. Standard multipoint-variance-component linkage methods (SOLAR) were used for the genome-wide screen [16]. Because multipoint linkage analysis currently cannot be performed on X-linked microsatellites, the 23 X-chr markers were analyzed with the two-point-variance-component method. A difficulty for quantitative-trait linkage analyses on the X-chr is that the covariance structure is not the same for male–male, female–female, or opposite-sex pairs because of the phenomena of X-inactivation and dosage-compensation. Therefore, linkage analyses on the X-chr were conducted while allowing for different genetic- and environmental-variances in males and females [17] to account for differences in covariance by sex because of dosage-compensation and X-inactivation.

![Fig. 1. Relationship between factor (F) IX activity (FIX:C) and age. Points represent the mean FIX:C level of each Genetic Analysis of Idiopathic Thrombophilia (GAIT) subject.](image-url)
Variance-component methods are vulnerable to deviations from normality in trait distribution, particularly to high levels of kurtosis [18]. The FIX.C distribution in GAIT exhibited a kurtosis of 0.03. Statistical-genetic theory demonstrates that this level will not affect the distribution of LOD scores and that the standard nominal $P$-values for LOD scores are appropriate for the FIX.C linkage screen [19]. Allele frequencies were estimated in GAIT, and the marker maps for multipoint analyses were obtained from ABI and the Marshfield Clinic (Marshfield, WI, USA). Because 12 GAIT families were ascertained through thrombophilic probands, all analyses included an ascertainment correction achieved by conditioning the likelihood of these pedigrees on the likelihoods of their respective probands [20].

Discovery and genotyping of F9 polymorphisms

F9 sequence variations were discovered and genotyped by directly sequencing nine amplicons, designated 1–9, that contained all known functional regions of this locus (Fig. 3B). These amplicons, located between sense-strand nucleotides 1,965–35,850 in GenBank sequence K02402 [21], were generated using genomic-DNA from individuals in HS-2 and the remaining GAIT subjects in HS-1. In order to define precisely the F9 regions investigated (Table 1), the 5′- and 3′-most nucleotide of each amplicon are designated with the number of the corresponding base in K02402 (Fig. 3B). All amplifications were performed using reaction conditions and thermal-cycling parameters that are available upon request. A water negative-control was included for each amplicon, in every run, and all were validated by 1% agarose-gel electrophoresis. All amplicons were purified using AmpPure, as described by the manufacturer, and subjected to cycle-sequencing using reaction conditions and thermal-cycling parameters that are available upon request. All sequencing-products were purified using CleanSEQ, as described by the manufacturer, and electrophoresed on either an ABI Prism 377 or 3100 instrument. ABI1-files were uploaded into the Finch server and assessed for overall quality; any sequence with an average Phred quality score (Q) less than 30 [22,23] was repeated. The sequence of each amplicon was determined from both strands for variation discovery and one strand for polymorphism genotyping.

All F9 variants (Fig. 3A) were identified by manually reviewing trace files in SeqMan multiple-sequence alignments. To minimize false-positives, all poor quality data (average $Q < 30$) were first removed. To be considered a true sequence variation, the allele(s) had to be consistent in both the forward- and reverse-strand reactions. Taq-induced errors are unlikely under the conditions used as the amount of genomic-DNA in each reaction provided several thousand haploid genome equivalents as template and amplicons were subjected to direct sequencing. However, before a previously unknown F9 variation was designated as naturally occurring, the minor allele had to be present in at least two subjects in HS-2 or be confirmed by directly-sequencing a second preparation of this amplicon derived from an independent PCR that used the same genomic-DNA as template. To confirm the sequencing-based genotypes of 5:32197:64 in the cDR (Fig. 3), fluorescently-labeled amplicons were generated, electrophoresed on an ABI Prism 3100 instrument, and evaluated by ALP-analysis using GeneScan (v3.7).

Linkage disequilibrium (LD)

We determined the pairwise LD between alleles of the 23 F9 polymorphisms that were informative in at least the SC and/or CA subjects (Fig. 4). Allele frequencies for these polymorphisms were estimated in 148 unrelated Caucasian individuals including the 100 SC subjects in HS-2a, seven additional GAIT males whose genotypes were inferred from family information, and the 41 CA males in HS-2c (Table 2). Weighted $D$ and $r^2$ were calculated for all consecutive SNP pairs using maximum-likelihood haplotype frequencies estimated by the expectation-maximization method implemented in GENECOUNTING [24]. The small number of JA and individuals of African-descent (AD) precluded accurate frequency estimates for F9 haplotypes found only in the non-Caucasian populations sampled, that is, those defined by SNPs monomorphic in Caucasians (G13516A, G30249A, A30555G, and G31093A). Thus we could neither investigate the possibility for ethnic differences in LD patterns across F9 nor determine the degree of LD within polymorphism pairs involving these four SNPs.

Marginal measured-genotype association analysis

Blinded genotypic data from the entire GAIT cohort for the 17 F9 polymorphisms that were informative in SC subjects (Table 2) were analyzed with INFER for violations of Mendelian inheritance. All disparities were checked for mistyping and either corrected or excluded from further analysis. Marginal association analysis was performed, using the measured-genotype approach [20], to test for genotype-specific differences in mean FIX.C while allowing for non-independence among family members. As males have just one allele, we used a dosage-compensation model in which hemizygous males had a gene effect equivalent to females homozygous for the same allele.
Endogenous, genetic and environmental covariates

GAIT data for age, sex, smoking status, oral contraceptive (OC) use, ABO-genotype, body mass index (BMI), diabetes-mellitus status, and plasma levels of total cholesterol (TC), high-density lipoprotein (HDL), low-density lipoprotein (LDL), very low-density lipoprotein (VLDL), triglycerides (TG), lipoprotein(a), von Willebrand factor antigen (VWF:Ag) and FVIII:C were subjected to both univariate and multivariate analyses with SOLAR [16]. We used indicator variables to represent ABO-genotypes, with the OO-genotype being the reference level, and modeled separate mean FVIII:C levels for each.

Results

FIX:C is correlated with age in univariate analysis

Because FIX:C was one of more than 25 hemostasis traits assayed in GAIT, duplicate measures were available for only 342 subjects; genomic-DNA from one was not available for study. In the remaining 341 individuals, FIX:C ranged from 42 to 237 IU dL\(^{-1}\) (mean \(\frac{1}{4} 115\) IU dL\(^{-1}\); SD \(\frac{1}{4} 26\) IU dL\(^{-1}\)). Data from one male, whose FIX:C (237 IU dL\(^{-1}\)) was greater than 4 SD above the mean (219 IU dL\(^{-1}\)) was excluded from further analysis. FIX:C in these 340 subjects (185 females, 155 males), which ranged from 42 to 191 IU dL\(^{-1}\) and demonstrated the greater than threefold concentration range reported for other non-hemophilic populations [6–9], are presented in Fig. 1 as a function of age. When evaluated independently from other measured covariates, we observed significant relationships between FIX:C and both age (\(P \frac{1}{4} 0.0000021\)) and age\(^2\) (\(P \frac{1}{4} 0.016\), but not for age\(^3\) (\(P \frac{1}{4} 0.582\) or sex (\(P \frac{1}{4} 0.237\)), confirming the widely reported observation that FIX:C increases with age in univariate analysis [8,9,14].

No major autosomal FIX determinants identified

We genotyped all GAIT subjects for 540 autosomal polymorphisms, including 506 uniformly spaced microsatellites and 34 random markers located in or near obvious candidate genes, and applied multipoint-variance-component methods to identify regions on chromosomes 1–22 in linkage with FIX:C (Fig. 2). As no LOD scores greater than 1.5 were observed, no regions met criteria suggestive for linkage. This finding indicates there may be no major autosomal determinants of FIX variability, at least in the SC population sampled in GAIT. However, despite the use of a high-density map providing 6.2 cm average genome coverage, two chr6 regions were not adequately screened. Referred to as 20 cm gaps, these regions span the 7.6 Mb chr6p21.2–12.3 segment flanked by D6S1610 and D6S452, and the 29Mb chr6q15–22.31 segment flanked by D6S462 and D6S287, which contain 96 and 104 genes, respectively (available upon request). Because of the small number of subjects, the power to detect QTLs with moderate effects is only modest.

No X-linked FIX determinants identified

Seventeen of the 23 X-linked microsatellites genotyped in GAIT were uniformly distributed across the chr; six other markers (DXS1227, DXS8043, DXS52, DXS1073, DXS1253E, and cDR) were selected for their locations in or near either F9 or the gene encoding FVIII (F8), the procofactor for catalytically active FIX. Analogous to the autosomes, we found no X-chr regions that met genome-wide criteria suggestive for linkage; LOD scores for these 23 markers were all less than 1.0, including 0.35 for the cDR and 0.76, the highest, for DXS986. As FIX levels are reported to increase with age in humans [8,9,14], we modeled age as a covariate. However, the linkage signals remained unchanged after including age, age\(^2\) and age · sex. Despite using equally informative X-linked and autosomal-microsatellites, X-linked loci including F9 may contain determinants of FIX variance because X-chr markers can currently only be evaluated with less powerful two-point linkage methods. Considered together, these results suggest that a collection of QTLs with modest effects is most likely responsible for the heritable variance in FIX:C.

Multiple novel F9 polymorphisms

Because association analysis is statistically superior to linkage analysis for detecting QTLs within candidate genes, we identi-
Table 2 Characteristics, status, and allelic associations of F9 polymorphisms with FIX.C

<table>
<thead>
<tr>
<th>F9 polymorphisms*</th>
<th>MAF</th>
<th>SC</th>
<th>CA</th>
<th>AD</th>
<th>JA</th>
<th>T</th>
<th>F9 variation databases</th>
<th>FIX-C association–P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TU</td>
<td>GenBank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HBMD</td>
<td>VDR</td>
</tr>
<tr>
<td>G-816A</td>
<td>4656574</td>
<td>1.4</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>G-793A</td>
<td>48397547</td>
<td>33.6</td>
<td>49.0</td>
<td>13.0</td>
<td>NP</td>
<td>47.0</td>
<td>+</td>
<td>F9-000716</td>
</tr>
<tr>
<td>C-698T</td>
<td>48397548</td>
<td>33.6</td>
<td>49.0</td>
<td>13.0</td>
<td>NP</td>
<td>47.0</td>
<td>+</td>
<td>F9-000811</td>
</tr>
<tr>
<td>G-653C</td>
<td>46532015</td>
<td>0.7</td>
<td>2.0</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>A-634G</td>
<td>46532016</td>
<td>0.7</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>T-609C</td>
<td>46532017</td>
<td>0.7</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>T-561G</td>
<td>48397549</td>
<td>8.4</td>
<td>15</td>
<td>50.0</td>
<td>NP</td>
<td>20.0</td>
<td>+</td>
<td>F9-000948</td>
</tr>
<tr>
<td>A192G</td>
<td>48397550</td>
<td>0.7</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>+</td>
<td>NF</td>
</tr>
<tr>
<td>C10118G</td>
<td>46532020</td>
<td>NP</td>
<td>29.0</td>
<td>25.0</td>
<td>50.0</td>
<td>29.0</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>A10365G</td>
<td>48397558</td>
<td>NP</td>
<td>2.0</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>NF</td>
<td>F9-01902</td>
</tr>
<tr>
<td>A10512G</td>
<td>48397559</td>
<td>NP</td>
<td>5.0</td>
<td>NP</td>
<td>NP</td>
<td>4.0</td>
<td>+</td>
<td>F9-012049</td>
</tr>
<tr>
<td>A11113G</td>
<td>48397551</td>
<td>28.0</td>
<td>32.0</td>
<td>13.0</td>
<td>NP</td>
<td>27.0</td>
<td>NF</td>
<td>F9-012651</td>
</tr>
<tr>
<td>T11269A</td>
<td>48397552</td>
<td>5.6</td>
<td>7.0</td>
<td>NP</td>
<td>NP</td>
<td>6.0</td>
<td>NF</td>
<td>F9-012806</td>
</tr>
<tr>
<td>C13162T</td>
<td>46532021</td>
<td>NP</td>
<td>2.0</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>G13516A</td>
<td>46532022</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>50.0</td>
<td>2.0</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>A20002C</td>
<td>48397553</td>
<td>28.0</td>
<td>NP</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>+</td>
<td>F9-021554</td>
</tr>
<tr>
<td>A20422G</td>
<td>48397554</td>
<td>27.3</td>
<td>32.0</td>
<td>25.0</td>
<td>NP</td>
<td>29.0</td>
<td>+</td>
<td>F9-021975</td>
</tr>
<tr>
<td>G30249A</td>
<td>46532023</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>C30249T</td>
<td>46532024</td>
<td>NP</td>
<td>2.0</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>A30555G</td>
<td>48397560</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>NF</td>
<td>F9-032113</td>
</tr>
<tr>
<td>G301093A</td>
<td>48397561</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>+</td>
<td>NF</td>
</tr>
<tr>
<td>G30205A</td>
<td>48397555</td>
<td>17.5</td>
<td>24.0</td>
<td>25.0</td>
<td>NP</td>
<td>24.0</td>
<td>NF</td>
<td>F9-033614</td>
</tr>
<tr>
<td>A32168T</td>
<td>46532018</td>
<td>2.8</td>
<td>2.0</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>532197/64</td>
<td>48397557</td>
<td>25.0/10.0</td>
<td>25.0/10.0</td>
<td>13.0/0.0</td>
<td>22.0/8.0</td>
<td>NF</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>T32770C</td>
<td>48397562</td>
<td>NP</td>
<td>2.0</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>+</td>
<td>NF</td>
</tr>
<tr>
<td>A32781G</td>
<td>46532019</td>
<td>NP</td>
<td>2.0</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>NF</td>
<td>NF</td>
</tr>
<tr>
<td>C32847T</td>
<td>48397556</td>
<td>1.4</td>
<td>2.0</td>
<td>NP</td>
<td>NP</td>
<td>2.0</td>
<td>NF</td>
<td>rs428199</td>
</tr>
</tbody>
</table>

FIX, factor IX; F9, FIX gene; FIX:C, FIX activity; MAF, minor-allele frequencies; TU, transcription unit; SC, Spanish-Caucasian; CA, Caucasian-American; AD, African-descent; JA, Japanese-American; T, total; HBMD, hemophilia-B mutation database; VDR, variation discovery resources (UW-FHCRC); dbsNP, database of single-nucleotide polymorphisms (NCBI); NP, not polymorphic; NT, not tested; NF, not found; NA, not applicable.

*The alleles, position in TU (start-site: nucleotide 1), and GenBank number (dbsNP) are listed for each polymorphism.

MAF estimated in unrelated subjects: HS-2a (SC); HS-2c: either its ethnic subgroups (CA, AD, and JA) or overall (T).

Designates databases with F9 polymorphisms: HBMD (+) designates known polymorphisms;

I denotes that A20422G is listed as 2042G>A and 147Ala>Thr; VDR (polymorphisms designated by gene and unique identifier); and dbsNP (polymorphisms designated by unique identifier).

-P-values from marginal measured-genotype association analysis.

We identified all polymorphisms in known functional regions of F9 by sequencing two subgroups of GAIT subjects, HS-2a and HS-2b. Although HS-2a contained all enrolled founders, 43 GAIT founders could not be enrolled for reasons previously described [3,12]. Consequently, any distinct F9 allele passed-down by these non-enrolled founders would be missed if the variation scan were limited to HS-2a. By including the descendents of these 43 non-enrolled founders (HS-2b), we ensured that all naturally occurring polymorphisms in the known functional regions of every founding F9 allele segregated in GAIT would be identified. By doing so, we maximized the probability of discovering FIX QTLs in the structural locus by association analysis, as these sequence variations would likely contain the functional subset of GAIT F9 polymorphisms.

We resequenced the known functional F9 regions including all exons, 1 kb of the contiguous promoter region, 50 bp of each intronic junction, and 127 bp of 3′-genomic-DNA (Table 1) [5,6,14,25]. We directly sequenced nine 500–3000 bp F9 amplicons generated from each person in HS-2a and HS-2b (Fig. 3). Using RepeatMasker (http://repeatmasker.org) to screen the 38059 bp of K02402 for repeats and low complexity sequences [26] in Repbase Update [27], we found that repetitive elements comprise 33% (12665 bp) of the structural locus and that 40.2% of all non-repetitive F9 sequences were covered in our resequencing scan. As the first step towards determining whether heritable FIX determinants in the SC population are conserved across ethnic groups, we resequenced the same F9 regions in a group containing 51 healthy unrelated male subjects (HS-2c) of different ethnicity, including: 41 CA, 6 AA, 2 JA, and 2 AP. The eight people of AD, including 6 AA and 2 AP, comprised the second largest ethnic group examined in this study.

We identified 27 polymorphisms distributed throughout F9 including seven in the promoter, two in coding sequences, 12 in introns, two in the 3′-UTR, and three in contiguous 3′-genomic-DNA (Table 2). Eleven of these polymorphisms, including four in the promoter (G)816A, G)653C, A)634G, and T)609C), five in introns (C10118G, C13162T, G13516A, G30249A, and

2006 International Society on Thrombosis and Haemostasis
from ethnicity, the minor-allele frequencies (MAFs) of these variations ranged from 0.7-47% (Table 2).

Some MAFs vary with ethnicity

While MAF for a few polymorphisms, e.g. C10118G, were similar across the ethnicities studied, several were not (Table 2). For example, while G13516A was not polymorphic in GAIT and demonstrated an overall MAF of 2% in HS-2c, the minor-allele was observed in one of two JA subjects. Eleven other SNPs were defined analogously by a single individual from one ethnic group, including eight restricted to CA and three to individuals of AD. Furthermore, four SNPs that were variable in SC subjects were monomorphic in CA. Likewise, while the MAF of C10118G was 29% in CA, as well as 25% and 50% in individuals of AD and JA, respectively, the minor-allele was absent in GAIT. However, the small numbers of individuals from other ethnic groups are insufficient for accurate frequency comparisons.

Patchy LD across F9

Because of their location in F9, each of the 27 polymorphisms may be functional and represent potential QTLs. To determine if these polymorphisms should be evaluated separately for association with FIX-C, we analyzed their pairwise LD. Because only ten non-Caucasian X-chromosomes were examined, AD (n ¼ 8) and JA (n ¼ 2), our estimated allele frequencies were based only on the 148 unrelated Caucasians

2006 International Society on Thrombosis and Haemostasis
in HS-2a, HS-2c, and the seven additional GAIT subjects described in Methods. Using these estimates, we found low overall LD across F9 in Caucasian populations (Fig. 2). In fact, of the 23 sites polymorphic in either SC or CA, 12 were not in LD with any other $(D_e + r^2 < 0.2)$. Nevertheless, the alleles in two pairs of polymorphisms (G-793AJc-698T, A10365G/C13162T) exhibited complete LD $(D_e \neq r^2 \neq 1)$ in the analyzed sample for both $D_e$ and $r^2$; alleles in four other pairs (A20002CA20422G, A11113GlcA20422G, A11113GlcA20002C, and G32056A15:32197:6/4) demonstrated moderate $(0.4 \leq LD < 0.6)$ to high $(0.6 \leq LD < 1)$ LD.

No major FIX determinants in F9

Because of the weak overall pattern of LD across F9, we genotyped every GAIT subject for the subset of polymorphisms variable in the SC population, including 17 (16 SNPs and 5:32197:6/4) whose minor-alleles were found in at least one person in HS-2a or HS-2b (Table 2). We then evaluated the relationship between the genotypes of each F9 polymorphism and FIX:C independently by marginal association analysis using the measured-genotype approach [29]. After allowing for non-independence among family members, we found no significant genotype-specific differences in mean FIX:C ($P$-values $> 0.11$) (Table 2). These results suggest that naturally occurring variations in known functional F9 regions are not likely to be determinants of FIX:C variability, at least in the SC population.

FIX:C is not associated with age in multivariate analysis

The identification of common variants in both age-related, cis-acting regulatory elements, without a clear association with FIX:C, led us to reexamine the relationship of FIX with age in the context of other endogenous, genetic and environmental variables, for which data was available (Table 3). After separately evaluating the relationship of FIX:C with each covariate independently, we found the following to have at least suggestive associations ($P \leq 0.10$): age, smoking status, BMI, ABO-genotype, and TC, LDL, HDL, VLDL, TG, VWF:Ag, and FVIII:C levels. Because the correlation coefficients between TC and LDL, and TG and VLDL were both greater than 0.90 ($P < 0.05$), we evaluated FIX:C by multivariate analysis using a model that included only TC and TG, together with age, sex, smoking status, BMI, ABO-genotype, and levels of HDL, VWF:Ag and FVIII:C. Despite not finding an association between sex and FIX:C, we choose to include sex in the model for final analysis (Table 3). While three endogenous parameters (TG, FVIII:C, BMI) and one environmental factor (smoking) were correlated significantly with FIX:C ($P$-values $< 0.005$), age was not.

### Discussion

FIX:C, a quantitative-trait exhibiting a broad normal range in non-hemophilic populations, represents a risk factor for venous thromboembolism [7,10] and possibly arterial thrombosis [11]. Despite mounting evidence that a substantial heritable component contributes to interindividual variance in FIX [4,12] and pleiotropically influences thrombosis risk [3], the underlying QTLs have not been identified. We used genome-wide and candidate-gene strategies to identify determinants of FIX variability. The genome-wide linkage screen was performed initially because FIX is a complex trait, involving multiple genetic and environmental factors, and this approach does not require candidate genes to be specified a priori. This strategy’s usefulness has been demonstrated by numerous studies that have identified heritable determinants of continuous traits in several physiologic systems and their correlated complex diseases [30–34].

We conducted the genome-wide screen in GAIT because, in addition to exhibiting a high FIX:C heritability, these families represented the general SC population sampled. This is the first reported use of a genome-wide approach to identify FIX QTLs. No autosomal regions met criteria for linkage with FIX:C (all LOD scores $< 1.5$), despite the use of multipoint variance-component methods and 563 markers that provided 6.2 cm genome coverage (Fig. 2). These results suggest that the 39% heritability found in GAIT [12] is likely because of a collection of QTLs with modest effects. Although none of the 204 genes contained in the two 20 cm gaps on chr6 are obvious candidates, genotyping additional microsatellites in these regions could help to determine if they contain determinants of FIX variance. Similarly, no single region of the X-chr met genome-wide criteria for linkage. However, male hemizygosity precludes the use of multipoint linkage analysis for X-linked markers, including those that were near or in F9. Therefore, it is not possible to exclude formally the possibility of an X-chr locus with a modest effect on FIX:C.

Given markers that are themselves functional or in high LD with an un-typed functional site, association analysis may be

---

**Table 3 Multicovariate analysis of FIX:C levels in the Genetic Analysis of Idiopathic Thrombophilia Project (GAIT)**

<table>
<thead>
<tr>
<th>Covariate</th>
<th>$\beta$-coefficient</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>8.963 - 10$^{-12}$</td>
<td>0.3728</td>
</tr>
<tr>
<td>Sex</td>
<td>0.295</td>
<td>0.9193</td>
</tr>
<tr>
<td>Age x Sex</td>
<td>9.764 - 10$^{-12}$</td>
<td>0.4169</td>
</tr>
<tr>
<td>Age$^2$</td>
<td>5.671 - 10$^{-14}$</td>
<td>0.8914</td>
</tr>
<tr>
<td>Age$^2$ x Sex</td>
<td>2.566 - 10$^{-13}$</td>
<td>0.6303</td>
</tr>
<tr>
<td>Smoking (yes/no)</td>
<td>6.972</td>
<td>0.0021</td>
</tr>
<tr>
<td>AA</td>
<td>3.110</td>
<td>0.4271</td>
</tr>
<tr>
<td>AO</td>
<td>0.361</td>
<td>0.8953</td>
</tr>
<tr>
<td>AB</td>
<td>1.109</td>
<td>0.8864</td>
</tr>
<tr>
<td>BO</td>
<td>5.394</td>
<td>0.1955</td>
</tr>
<tr>
<td>TC (mM)</td>
<td>1.008</td>
<td>0.4519</td>
</tr>
<tr>
<td>HDL (mM)</td>
<td>5.792</td>
<td>0.1146</td>
</tr>
<tr>
<td>TG (mM)</td>
<td>15.212</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>BMI (kg m$^{-2}$)</td>
<td>1.752</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>FVIII:C (IU dL$^{-1}$)</td>
<td>0.137</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>VWF:Ag (%)</td>
<td>4.695 - 10$^{-12}$</td>
<td>0.2639</td>
</tr>
</tbody>
</table>

TC, total cholesterol; HDL, high-density lipoprotein; TG, triglycerides; BMI, body mass index; FVIII:C, factor VIII activity; VWF:Ag, von Willebrand factor antigen.
able to detect loci with smaller effects. This is supported by findings from a similar study to identify variance determinants of FVIII:C [35]. Although FVIII:C represents one of the most heritable coagulation phenotypes in GAIT, we found no evidence for linkage with microsatellites in or near F8. Using association analysis, however, we found that D1241E, a conservative non-hemophilic B-domain substitution, represents a FVIII QTL. After identifying all naturally occurring F9 polymorphisms in the known functional regions (Table 1) of every distinct allele segregated in GAIT, we used association analysis to evaluate the relationship between the structural locus and FIX:C. Because the pattern of LD across F9 was weak overall (Fig. 4), we examined each polymorphism (Fig. 3) separately for potential contributions to FIX:C variability by marginal measured-genotype association analysis. However, no significant differences in mean FIX:C levels were found (P-values ≤ 0.11). While these findings suggest that the structural locus is not a determinant of FIX:C variance, our study was confined to known functional regions that comprise only 40% of all non-repetitive F9 sequences (Fig. 3). Using PhastCons (http://www.genome.ucsc.edu/cgi-bin/hgTrackUi?hgsql/41178675&c=chrX&BphastConsElements) [36–38], we identified 17 non-exonic F9 segments that are highly conserved across vertebrates (not shown). These segments, ranging from 58–449 bp and containing 15 SNPs from dbSNP, displayed conservation scores greater than the mean score for all exonic sequence. None of these candidate cis-acting regulatory elements, which may contain determinants of FIX:C variance, were examined here. To interrogate these segments by association analysis, we are currently defining naturally occurring allelic combinations of the F9 polymorphisms identified here. By taking advantage of the greater overall degree of LD provided by haplotypes, this approach should increase the power of the GAIT sample to detect functional variants in these regions.

Limiting factors of the present genome-wide and candidate-gene studies may be considered in the following categories: sample size, phenotype (FIX:C), and ethnicity. To overcome limitations related to the current sample size, a larger family study of the SC population (GAIT-II) is underway. FIX:C is measured under conditions which may not recapitulate precisely those in vivo and requires interactions between additional complex hemostasis traits, all of which, except FIX, are derived from numerous unrelated healthy donors. As FIX:Ag measurements require a single interaction between FIX and reagent antibodies, this complimentary phenotype is being used, along with FIX:C, in GAIT-II. This will allow use of FIX:C and FIX:Ag independently and/or in combination, by dividing FIX:C by FIX:Ag, as a parameter of specific activity. While our findings demonstrate that no single gene, including F9, is likely to be a major FIX variance determinant in the SC population sampled by GAIT, it is not clear whether this applies to non-Caucasian populations or even non-Spanish subgroups of Caucasians, such as CA. To determine whether there are ethnic differences in the types and/or frequencies of F9 polymorphisms, we examined the same gene regions in a group of 51 unrelated healthy male subjects (HS-2c) including 41 CA, two JA, and eight of AD (Table 2). Although 14 polymorphisms were restricted to Caucasians, such MAF deviations are likely because of the small number of non-Caucasian F9 alleles sampled, that is, eight of AD and two from JA. While the MAF of some polymorphisms were similar across ethnic groups, there were several notable exceptions. Of the 27 variations identified, ten SNPs were not variable in SC subjects, including C10118G whose MAF ranged from 0.25 to 0.5% in HS-2c. Moreover, while the MAF of G13516A was 50% in the two JA individuals, it was not polymorphic in other ethnic groups. Furthermore, G30249A, A30555G, and G31093A were restricted to individuals of AD. As the results from this study may apply only to the SC subjects sampled in GAIT, similar family studies should be conducted in other ethnic populations.

Conclusions

Our findings suggest that no single gene, including F9, is likely to be a major determinant of FIX:C. We hypothesize that a collection of QTLs with modest effect sizes and distinct genomic locations underlie the variance in this phenotype. We observed a significant univariate correlation between FIX:C and age, consistent with previous studies. However, as this correlation canceled out in multivariate analysis, further investigation will be required to elucidate the relationship between age and F9 expression. Finally, in vitro functional investigations may be necessary to overcome the limitations of the current study.

Author contributions


Acknowledgements

We thank all patients and other individuals for their participation in this study. We also thank Pete Lollar for mentorship and early support, Ming Shen for technical help, Bruce Evatt and Connie Miller for helpful discussions regarding the manuscript, Russ Howard and Peter Hay for editing the manuscript, and Mary Howard for her enthusiastic support. This work was supported by: NIH-R01-HL-70751 (L. A.), NIH-K08-HL-71130 and -R01-HL-68016 (T. E. H.), FIS-02/0375, SAF2002-03449, FIS-RECAVA-C03/01, FIS-99/3048 and -01/A046 (J. M. S. and A. B.).

2006 International Society on Thrombosis and Haemostasis
References


8 Sweeney JD, Hoernig LA. Age-dependent ect on the level of factor IX. Am J Clin Pathol 1993; 99: 687–8.


