

# A sequence variation scan of the coagulation factor VIII (FVIII) structural gene and associations with plasma FVIII activity levels

Kevin R. Viel,<sup>1,2</sup> Deepa K. Machiah,<sup>1</sup> Diane M. Warren,<sup>3</sup> Manana Khachidze,<sup>1</sup> Alfonso Buil,<sup>4</sup> Karl Fernstrom,<sup>1</sup> Juan C. Souto,<sup>4</sup> Juan M. Peralta,<sup>3,5</sup> Todd Smith,<sup>6</sup> John Blangero,<sup>3</sup> Sandra Porter,<sup>6</sup> Stephen T. Warren,<sup>7</sup> Jordi Fontcuberta,<sup>4</sup> Jose M. Soria,<sup>4</sup> W. Dana Flanders,<sup>2</sup> Laura Almasy,<sup>3</sup> and Tom E. Howard<sup>1</sup>

<sup>1</sup>Department of Pathology and Laboratory Medicine and <sup>2</sup>Department of Epidemiology, Emory University, Atlanta, GA; <sup>3</sup>Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX; <sup>4</sup>Unitat d Hemostàsia i Trombosi, Departament d Hematologia, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; <sup>5</sup>Centro de Investigacion en Biologia Celular y Molecular, Universidad de Costa Rica, San José, Costa Rica; <sup>6</sup>Geospiza, Seattle, WA; and <sup>7</sup>Department of Human Genetics, Emory University, Atlanta, GA

**Plasma factor VIII coagulant activity (FVIII:C) level is a highly heritable quantitative trait that is strongly correlated with thrombosis risk. Polymorphisms within only 1 gene, the ABO blood-group locus, have been unequivocally demonstrated to contribute to the broad population variability observed for this trait. Because less than 2.5% of the structural FVIII gene (*F8*) has been examined previously, we resequenced all known functional regions in 222 potentially distinct alleles from 137 unrelated nonhemophilic individuals representing 7 racial groups. Eight-**

**teen of the 47 variants identified, including 17 single-nucleotide polymorphisms (SNPs), were previously unknown. As the degree of linkage disequilibrium across *F8* was weak overall, we used measured-genotype association analysis to evaluate the influence of each polymorphism on the FVIII:C levels in 398 subjects from 21 pedigrees known as the Genetic Analysis of Idiopathic Thrombophilia project (GAIT). Our results suggested that 92714C>G, a nonsynonymous SNP encoding the B-domain substitution D1241E, was significantly associated with FVIII:C**

**level. After accounting for important covariates, including age and ABO genotype, the association persisted with each C-allele additively increasing the FVIII:C level by 14.3 IU dL<sup>-1</sup> ( $P = .016$ ). Nevertheless, because the alleles of 56010G>A, a SNP within the 3' splice junction of intron 7, are strongly associated with 92714C>G in GAIT, additional studies are required to determine whether D1241E is itself a functional variant. (Blood. 2007;109:3713-3724)**

© 2007 by The American Society of Hematology

## Introduction

Factor VIII (FVIII) circulates, bound to von Willebrand factor (VWF), as an inactive heterodimer with domain configuration A1-A2-B (heavy chain) and A3-C1-C2 (light chain). Activated FVIII (FVIIIa), which is released from VWF as a heterotrimer with A1, A2, and A3-C1-C2 subunits, binds the protease FIXa to form intrinsic Xase whose sole function is to catalyze cleavage activation of FX. Intrinsic Xase is regulated by several processes including FVIIIa inactivation through A2-subunit dissociation and cleavage by activated protein C (aPC). Although FVIII coagulant activity (FVIII:C) is deficient in hemophilia-A patients, individuals with elevated levels have an increased risk for both venous<sup>1-3</sup> and arterial thrombosis.<sup>2,4</sup>

FVIII levels in nonhemophilic populations are broadly variable, spanning a greater than 5-fold range.<sup>5,6</sup> Age,<sup>7-12</sup> sex,<sup>7,9</sup> oral contraception (OC),<sup>10,13</sup> smoking,<sup>7,14,15</sup> body-mass index (BMI),<sup>7,12,16,17</sup> diabetes mellitus (DM),<sup>7,16</sup> ABO blood type,<sup>17-22</sup> and plasma levels of total cholesterol (TC),<sup>16,17</sup> low-density lipoprotein (LDL),<sup>16,17,23</sup> triglyceride (TG),<sup>16,17</sup> VWF,<sup>24-29</sup> and FIX coagulant activity (FIX:C)<sup>10,24,30</sup> all have ample evidence supporting their role as correlates of FVIII:C. Family studies, which have yielded heritability estimates ranging between 40% and 57% and 20% and 61%, respectively, for the levels of FVIII:C<sup>21,31</sup> and FVIII antigen (FVIII:

Ag),<sup>32-34</sup> also demonstrate a substantial genetic contribution to FVIII variance. Furthermore, the finding by Souto et al<sup>35</sup> of a strong genetic correlation between FVIII:C level and thrombosis ( $\rho_g = 0.689$ ;  $P < .001$ ) suggests that a subset of polymorphisms underlying FVIII variability pleiotropically influence thrombosis risk. Supporting this, Soria et al<sup>36</sup> discovered a region on chromosome 18 (chr18) that contributes to the variance in both FVIII:C and aPC resistance (aPC-R), an established pathophysiologically related thrombosis risk factor. However, this putative FVIII quantitative-trait locus (QTL) and 2 other potential determinants localized to regions on chr5 and chr11 in the first reported genome-wide screen for which FVIII:C was the primary phenotype<sup>37</sup> have yet to be confirmed. In a family-based study of the ABO structural gene, the only well-established FVIII QTL,<sup>17-22</sup> Souto et al<sup>22</sup> applied combined tests for linkage and association to demonstrate that the polymorphisms encoding allelic glycosyltransferases responsible for ABO blood-group antigens also directly influence FVIII:C levels and represent the only known FVIII quantitative-trait nucleotides (QTNs).

While approximately 1000 loss-of-function FVIII gene (*F8*) alleles underlie the heterogeneity observed for FVIII:C in hemophilia A,<sup>38</sup> only 1 *F8*-based QTN contributing to the FVIII

Submitted June 9, 2006; accepted November 3, 2006. Prepublished online as *Blood* First Edition Paper, January 5, 2007; DOI 10.1182/blood-2006-06-026104.

A portion of this work was presented initially at the 45th annual meeting of the American Society of Hematology, San Diego, CA, December 8, 2003.<sup>39</sup>

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2007 by The American Society of Hematology

variability in nonhemophilic individuals has been identified<sup>39</sup> and confirmed<sup>40</sup> despite numerous studies.<sup>25,36,37,39,41-45</sup> The X-linked *F8*<sup>46</sup> is thought to be less variable than other genes,<sup>6</sup> in part because no polymorphisms have been found in prior investigations of this candidate FVIII determinant.<sup>25,41-43,45</sup> Mansvelt et al<sup>41</sup> identified no promoter variants in subjects with high FVIII levels. Regulatory regions required for *F8* expression in vivo remain poorly characterized despite findings from in vitro assays demonstrating maximal transcription with less than 500 bp of the 5' genomic sequence.<sup>47,48</sup> Morange et al<sup>25</sup> also found no variants in *F8* regions encoding residues implicated in FVIII clearance by the lipoprotein receptor-related protein (LRP). Furthermore, studies of idiopathic thrombophilia subjects, with or without aPC-R, have not identified polymorphisms within *F8* regions encoding aPC-cleavage sites,<sup>42,43,45</sup> in contrast to the aPC-R variants frequently found within homologous *F5* gene regions in thrombosis patients.<sup>49,50</sup> Finally, while a subset of thrombophilic individuals with high prothrombin levels carry the prothrombin gene 3' untranslated region (UTR) mutation 20210G>A,<sup>51,52</sup> Mansvelt et al<sup>41</sup> found no polymorphisms within this region of *F8* in thrombosis patients with elevated FVIII levels. However, these studies scanned only small segments of *F8* and often used screening methods less sensitive than DNA sequencing.<sup>41</sup>

Since loss-of-function hemophilic variants have been identified within every exon and junctional intronic sequence, multiple *F8* regions are likely essential for FVIII expression and/or function in vivo.<sup>38</sup> Future attempts to identify FVIII QTNs in *F8* should examine every functional region by direct sequencing in a large sample of unrelated healthy subjects. While 98 predominantly unknown *F8* polymorphisms were recently identified in a scan that met most of these criteria, subsequent analyses for associations with FVIII:C were not possible, as no phenotypic data were available for these subjects.<sup>53</sup> To determine whether *F8* is a determinant of FVIII variability, we resequenced all known functional regions within a collection of 222 X chromosomes from 137 unrelated subjects. Next we genotyped the polymorphisms identified in the 398 subjects of the Genetic Analysis of Idiopathic Thrombophilia project (GAIT).<sup>31</sup> Finally, we used association analysis to determine if GAIT subjects have genotype-specific differences in mean FVIII:C levels after adjusting for relevant covariates.

## Patients, materials, and methods

### Reagents and instruments

HotStarTaq-Master Mix was from QIAGEN (Valencia, CA); AmpPure and CleanSEQ were from Agencourt Biosciences (Beverly, MA); polymerase chain reaction (PCR) and DNA-sequencing oligonucleotides (oligos) were from Invitrogen (Carlsbad, CA); and BigDye-Terminator (v3.1 or v1.1) Cycle-Sequencing Kits and ABI Prism-3100 and -3700 automated sequencers were from ABI (Foster City, CA). SeqMan multiple-sequence alignment program was from DNASTAR (Madison, WI); PHRED (v020425.c)<sup>54,55</sup> was a gift from the University of Washington; and FINCH DNA-Sequencing System (v.2.09) was from Geospiza (Seattle, WA). Reaction conditions and thermal-cycling parameters for all PCR and DNA sequencing are available upon request.

### Subjects and phenotypes

(1) GAIT study subjects. All procedures were approved by the institutional review board (IRB) of the Hospital de la Santa Creu i Sant Pau (Barcelona, Spain). Adult subjects gave informed consent for themselves and their minor children. Coded archived DNA samples

from these subjects as well as previously obtained data for measured variables, including FVIII:C levels, were used. (2) Emory study subjects. All studies performed on these DNA specimens were approved by the IRB of Emory University (Atlanta, GA). Archived coded DNA samples from these subjects, which were collected for an unrelated study, were the only biologic specimens used. (3) Coriell study subjects. These subjects represent a collection from the National Institute of General Medical Sciences (NIGMS) of commercially available genomic-DNA samples from unrelated healthy subjects and different ethnic backgrounds that we obtained through the Coriell Cell Repository.

Studies on samples from the 3 groups of subjects referred to as GAIT, Coriell, and Emory were approved by the Emory University IRB. GAIT is composed of 398 Spanish white (SW) individuals from 21 extended pedigrees.<sup>31</sup> The recruitment, sampling, and phenotypes measured in GAIT, including FVIII:C levels and additional measured variables, have been extensively described.<sup>31,35</sup> Coriell designates a collection of genomic DNAs from NIGMS human cell panels representing 45 unrelated healthy subjects, including 10 female African American (AA), 3 female and 7 male Chinese (Ch), 5 female and 2 male Japanese (J), 2 female and 3 male Mexican Indian (MI), 3 female and 2 male South American Andean (SAA), and 5 female and 5 male non-J/non-Ch Southeast Asian (SEA) individuals, which were obtained from Coriell (Camden, NJ); catalog numbers available upon request. Emory designates genomic DNAs from 24 unrelated nonthrombotic male subjects, a subset of an archived collection originally obtained for an unrelated study,<sup>56</sup> of which 18 are white American (WA) and 6 AA. The variation discovery group (VDG), which contained 137 unrelated subjects (85 female, 52 male) from 7 racial groups, included all Coriell and Emory individuals and a subset of 68 GAIT founders (58 female, 10 male) with FVIII:C levels at the extremes. In the VDG, "white" designates both SW and WA subjects.

### F8 reference sequence

We used the May 2004 human genome (hg) assembly to download sequence data for a 286-kb stretch of DNA from Xq28.1 that contained the entire *F8* (~186 kb) plus 50-kb segments of contiguous DNA flanking both the 5' and 3' ends.<sup>57</sup> Referred to as hg17, this was the first version (National Center for Biotechnology Information [NCBI] Build-35) containing all 26 exons of *F8*; specifically, it included exons 21 and 22, which were missing from earlier releases. This sequence is contained in GenBank accession NG\_005114 and can be used to cross-reference oligos (Table 1) and polymorphisms (Table 2). To define polymorphisms with respect to the *F8* transcription unit (TU), we used the reverse complement of these sequences, assigned 1 to the start site mapped by Mansvelt et al<sup>41</sup> and -1 to the base immediately 5' to it (Figure 1). In this report, we use "hg17" to indicate that the nucleotide (nt) numbering follows this convention.

### F8 variation scan

To identify potential FVIII determinants, we scanned the known functional regions of *F8* (including all exons, 50 to 100 bp of each junctional intronic region, approximately 1.2 kb of contiguous promoter sequence, and approximately 300 bp of flanking 3' genomic DNA) by directly sequencing 500- to 600-bp amplicons that, where necessary to cover extended regions, were overlapping (Table 1). Table S1 (available on the *Blood* website; see the Supplemental Materials link at the top of the online article) lists the lengths and hg17 nt boundaries of these amplicons and amino acids encoded by the exonic regions examined. Based on the July 2003 hg assembly, we initially generated 39 amplicons. Due to the presence of gaps, amplicons 29 and 30 (designated 29A and 30A) lacked exons 21 and 22 (Figure 1). Although absent from this scan, we included amplicons 29B and 30B, containing exons 21 and 22, respectively, in the genotyping phase described below under "Genotyping." All amplicons, which were generated using genomic DNAs from VDG subjects, and their hg17 nt positions are listed in Table 1.

Agencourt Biosciences performed all initial amplification and cycle-sequencing reactions for the variation scan, in which each amplicon was examined on both strands. Water-negative controls for each amplicon were included and evaluated identically. Agencourt used ABI-3700 sequencers

**Table 1. F8 regions resequenced to identify and genotype common polymorphisms**

Amplicon*	Forward primer	Reverse primer	Nucleotides†
01	5' -TTATCAAAGGGGCTTCTTGC	5' -CATGCCCTTTCTCCTGACC	-1214 to -573
02	5' -AGCAAGTGTGAGGTCCAGG	5' -TGAAGTAGCAAAAAGGAGGC	-683 to -56
03	5' -CTTCTCCAATCCCTCTCCTCC	5' -CAGAAATGTTTCTTTGGGGC	-168 to 441
04	5' -CTTCAAATTTGCTCCTTGC	5' -AGACCAAGCAGAGGAAGACG	23012 to 23436
05	5' -AATCTTGCCCTCAGAGCAACC	5' -GAAAAGCAATTCCTAGGGGG	25450 to 26066
06	5' -GGGCAACAGAGTGAGACTCC	5' -TTCTGGAACCTCAGCTCCTCC	29413 to 30016
07	5' -GGAGACCTGACATCAAAGCC	5' -AACCCATCTCCTTTCATTCC	35266 to 35608
08	5' -TAAGGTGTGAGCACACTGGG	5' -CGATGAGTTCTGTCTGAGCC	37821 to 38404
09	5' -ATGGTGATTGGTGACCTTGG	5' -GGAAACTAGGGGATCTTGGC	52986 to 53544
10	5' -GTCCTGCTCCTGCTTTCACC	5' -TACCCTTGCCATTTGATTCC	55802 to 56428
11	5' -CTGCTGAAGAGGAGGACTGG	5' -ATGTCCATTGGAGACAAGGC	56244 to 56855
12	5' -GATTGTGGTATCTGCAGGGG	5' -CAACAGCTGGAGAAAAGACC	61345 to 61753
13	5' -TGACACTTTCACAGTCAACCG	5' -CAGCAGGCACGTTTACTACG	65333 to 65920
14	5' -CAGTCACCTCTTGTCTCCGG	5' -GGGAATTAAGAGGAGAGGG	68456 to 69067
15	5' -CCTGGGAATAAGATAATGGGC	5' -AAATGCTGGTGAGGATGTGG	74699 to 75338
16	5' -ACAGCAGCAATGCAAAAACC	5' -TCTATTGCTCCAGGTGATGG	90867 to 91468
17	5' -ATGCTCTTGCAGCAGAGTCC	5' -AACAAAGCAGGTCCATGAGC	91365 to 91942
18	5' -TTGGCAAAAAGTCATCTCCC	5' -CTAATTGCTTTGAGCTGGGG	91756 to 92379
19	5' -CCACCAGATGCACAAAATCC	5' -TTTGCTTGGTTTGATTTCCC	92247 to 92850
20	5' -GAAGGTCATATGAGGGGGC	5' -ATGACTGCTTCTTTGGACCC	92700 to 93290
21	5' -TCTGACCAGGTCCTTATCC	5' -CATGATTGCTTTTCAACAGCG	93200 to 93816
22	5' -ATTGGATCCTCTGTCTGGG	5' -TGTCCCTGATTCTCTTACCC	93674 to 94323
23	5' -ATGCAAAATGCTTCTCAGGC	5' -AAAAGCTTGTTCAAAATAAATGG	116050 to 116647
24	5' -TCTGTACCCTTCTTCCAGGG	5' -TTTATGGCAGTCCAACTGTC	117559 to 118125
25	5' -TATTTTGGAAAGTGGGAGG	5' -CGAATCTCTTGTATCTGAGC	118087 to 118699
26	5' -TTGATGAGACAAAAGCTGG	5' -AGAGCATGGAGCTTGTCTGC	118318 to 118933
27	5' -AAGCACTTTGCATTTGAGGG	5' -TGGAGATCTTCGAGCTTTACC	120414 to 120947
28	5' -GGACCCAGTTTCTTTCAGC	5' -AGTGGGAAGTGGAGAGGAGG	121066 to 121510
29B	5' -GAATTTAATCTCTGATTCTCTAC	5' -GAGTGAATGTGATACATTTCCC	122740 to 122902
30B	5' -TAAAAATAGGTTAAAATAAAGTG	5' -TTTAAATGACTAATTACATACCA	126453 to 126668
29A	5' -TCAGGGTTGGTTACTGGAGC	5' -ACACTACCATTGGTCTTGGGG	158245 to 158735
30A	5' -AGTCAGTGGCCGTGTTATGG	5' -GTCCCTAGCTCTTGTTCCTCC	158526 to 159105
31	5' -TGGGCAGATAGGGATAGTGG	5' -TTTGTGCGTTTCTCAACAGC	158833 to 159409
32	5' -TTCCCACTTCTCTTGGTGC	5' -TGGGCATTTAGGTTGACTCC	159304 to 159934
33	5' -TCATGCCACTTACTCCAGC	5' -CTGCCATAAACCAACTTCC	160633 to 161150
34	5' -GGGTGACAGAGCAAGACTCC	5' -AAAAGCTTGGGAATCAAGG	161982 to 162549
35	5' -AGATGTCCCAGATGCGTAGG	5' -GCTTTCATGCAGGTTTCTCC	184818 to 185411
36	5' -TATTTCTGCAGCTGCTCCC	5' -CTTTCAACAATTCGATCTCTCC	185329 to 185941
37	5' -GAGGGGCACATTTATCTCC	5' -TCATAGTGAAGGGGTCAGGC	185733 to 186382
38	5' -CACCACACAATAGGATCCCC	5' -GTCAATGGGAAAAGAAATGCC	186289 to 186832
39	5' -CAATCCACAATGATGCAGG	5' -AGTGCCAGGATTACAGGCAT	186639 to 187259

\*To include all known functional F8 regions in the variation scan, 41 distinct amplicons of the structural locus were PCR amplified from genomic-DNA samples and resequenced directly. The 11 amplicons indicated in italics were directly sequenced to genotype the entire GAIT cohort for the 12 F8 variations that were polymorphic in white individuals.

†Numbering for nucleotides corresponds to the hg17 reference sequence for F8.

and the PHRAP programs PHRED, CONSED, and POLY-PHRED to identify variants.<sup>54,55</sup> The quality of Agencourt's sequence chromatograms, which we used to both detect and genotype polymorphisms, were assessed in-house by (i) uploading chromatograms, including those for a blind genomic-DNA replicate from an Emory subject, into the Finch server to determine average PHRED quality (Q) scores; (ii) manually reviewing SeqMan alignments to validate base calls for minor alleles and strand consistency; (iii) performing agarose-gel electrophoresis and multiple-sequence alignments to validate negative controls for amplification and sequencing reactions, respectively; and (iv) manually reviewing multiple-sequence alignments of an approximately 5% random subset of all chromatograms. Sequences that were either of poor quality (average Q < 30)<sup>54,55</sup> or that yielded inconsistent base calls were repeated in-house. Previously unknown polymorphisms were designated as naturally occurring if minor alleles were present in 2 or more subjects. Polymorphisms with minor alleles found in only 1 subject were considered naturally occurring if confirmed upon sequencing a second amplicon derived from an independent PCR.

**LD**

To evaluate pairwise allelic associations across F8, we calculated *r*<sup>2</sup> and *D'*, 2 commonly used measures of linkage disequilibrium (LD), for the subset of 12 functional-region SNPs that were variable among the unrelated white individuals in the VDG and those GAIT subjects whose parents were not enrolled for study (Table 2). The results were plotted with SOLAR (<http://www.sfbr.org/solar/>), a software package for genetic variance-components analysis (Figure 2).<sup>60</sup> *D'* and *r*<sup>2</sup> were calculated for all SNP pairs.<sup>61</sup>

**Genotyping**

We genotyped remaining GAIT subjects (158 female, 172 male) for the subset of functional-region polymorphisms that were variable in white individuals of the VDG (Table 2) by generating and directly sequencing 1 strand of the 9 amplicons indicated in Table 1. Agencourt performed the initial PCR and sequencing, for which we performed quality control

in-house. To ensure that any polymorphisms identified in exons 21 and 22 could be evaluated by association analysis, 2 additional amplicons, 29B and 30B (Table 1), were generated from every GAIT subject and sequenced on both strands. We made all final genotype calls using PHRED output and a custom-written SAS\_8.2 program (<http://www.SAS.com>) that generates a separate chromatogram for each polymorphism (custom-written program available upon request). By being compact and centering the variant between 2 10-bp flanks, these chromatograms standardized our manual genotype review. We paired

forward- and reverse-strand sequences from the same individual to perform manual reviews but did not reveal the identity further than providing the file names, which contained logistic data such as plate, well address, amplicon, and direction. To resolve missing genotypes, we sequenced amplicons generated in-house on an ABI-3100. Genotypic data were analyzed with INFER, a component program implemented in the PESYS suite ([http://www.sfbr.org/pages/genetics\\_projects.php?p=42](http://www.sfbr.org/pages/genetics_projects.php?p=42)), for violations of Mendelian inheritance; all disparities were reviewed and either corrected or excluded from further analysis.

**Table 2. Characteristics, status, and allelic associations of F8 polymorphisms with FVIII:C levels**

Region, TU*	Protein*	GenBank*	m-AFs, %†								F8 variation databases‡			
			T	W	AA	Ch	SEA	J	MI	SAA	HAMStERS <sup>38,58</sup>	VDR <sup>53</sup>	dbSNP <sup>59</sup>	
<b>Promoter</b>														
-825G>A	—	49785830	0.9	NP	7.7	NP	NP	NP	NP	NP	NP	NF	F8-002289	NF
-824G>A	—	49785831	0.9	NP	7.7	NP	NP	NP	NP	NP	NP	NF	F8-002290	NF
-493G>A	—	49785832	0.5	NP	3.8	NP	NP	NP	NP	NP	NP	NF	F8-002621	rs4898404
-385 A <sub>4</sub> >A <sub>5</sub>	—	49785833	0.9	NP	NP	NP	NP	NP	NP	25.0	NP	NF	NF	NF
-287T>C	—	49785834	0.5	NP	3.8	NP	NP	NP	NP	NP	NP	NF	NF	NF
<b>Intron 2</b>														
25610G>A	—	49785835	0.5	NP	4.5	NP	NP	NP	NP	NP	NP	NF	F8-028722	NF
<b>Intron 3</b>														
25865G>A	—	49785836	0.5	NP	4.2	NP	NP	NP	NP	NP	NP	NF	F8-028977	NF
25885G>C	—	49785837	0.5	NP	4.2	NP	NP	NP	NP	NP	NP	NF	F8-028997	NF
29567C>T	—	49785838	0.5	NP	NP	NP	NP	NP	NP	NP	12.5	NF	NF	NF
<b>Intron 4</b>														
29854T>C	—	49785839	0.5	NP	3.8	NP	NP	NP	NP	NP	NP	NF	NF	NF
<b>Intron 5</b>														
35518C>G	—	49785840	0.9	NP	8.0	NP	NP	NP	NP	NP	NP	NF	NF	NF
<b>Intron 6</b>														
53034A>G§	—	49785841	0.5	0.7	NP	NP	NP	NP	NP	NP	NP	NF	NF	NF
<b>Exon 7</b>														
53206G>T§	W255C	49785842	0.5	NP	NP	NP	6.7	NP	NP	NP	NP	NA	NF	NF
<b>Intron 7</b>														
55938C>A§	—	49785843	0.5	NP	NP	NP	6.7	NP	NP	NP	NP	NF	NF	NF
56010G>A§	—	49785844	9.6	6.4	26.9	7.7	NP	12.5	12.5	25.0	+	F8-059122	rs7058826	
<b>Exon 8</b>														
56113G>A§	A343A	49785845	0.5	NP	3.8	NP	NP	NP	NP	NP	NP	+	F8-059225	rs1800289
<b>Intron 9</b>														
61534T>C	—	49785846	0.5	NP	3.8	NP	NP	NP	NP	NP	NP	NF	F8-064646	rs5986899
<b>Exon 10</b>														
61620G>A	R484H	49785847	0.9	NP	8.3	NP	NP	NP	NP	NP	NP	+	NF	NF
<b>Intron 13</b>														
75215G>A	—	49785848	0.5	NP	3.8	NP	NP	NP	NP	NP	NP	NF	F8-077273	rs5987069
90948A>G	—	49785848	0.5	NP	NP	NP	NP	NP	NP	NP	12.5	NF	NF	NF
<b>Exon 14</b>														
91317A>G	R776G	49785849	0.5	NP	NP	7.7	NP	NP	NP	NP	NP	NF	NF	rs2228152
92555C>T§	H1188H	49785850	0.5	NP	4.3	NP	NP	NP	NP	NP	NP	NF	F8-095667	NF
92714C>G§	D1241E	49785851	14.7	7.7	72.7	7.7	NP	12.5	12.5	25.0	+	F8-095826	rs1800291	
92798A>C§	S1269S	49785852	7.7	6.3	3.8	7.7	20.0	NP	12.5	25.0	+	F8-095910	rs1800292	
92927G>A	K1312K	49785853	0.5	NP	NP	NP	6.7	NP	NP	NP	NP	NF	NF	NF
93401C>G	V1470V	49785854	0.5	NP	4.5	NP	NP	NP	NP	NP	NP	NF	NF	NF
93434G>A	P1481P	49785855	0.9	NP	9.1	NP	NP	NP	NP	NP	NP	NF	NF	NF
<b>Intron 15</b>														
116434C>T	—	49785856	0.5	NP	NP	7.7	NP	NP	NP	NP	NP	NF	NF	NF
<b>Intron 18</b>														
118909T>A§	—	49785857	20.3	14.6	84.6	7.7	14.3	NP	42.9	37.5	+	F8-122021	rs4898352	
<b>Intron 19</b>														
120776T>C§	—	49785858	24.5	17.3	75.0	7.7	16.7	NP	37.5	50.0	+	F8-123888	rs4074307	
<b>Intron 22</b>														
158352C>T	—	49785859	0.5	0.7	NP	NP	NP	NP	NP	NP	NP	NF	NF	NF
158368T>C	—	49785860	0.5	0.7	NP	NP	NP	NP	NP	NP	NP	NF	F8-161500	NF
158635C>T	—	49785861	0.5	NP	3.8	NP	NP	NP	NP	NP	NP	NF	F8-161767	rs5987054
158777A>G	—	49785862	0.5	0.7	NP	NP	NP	NP	NP	NP	NP	NF	NF	NF
158820C>T	—	49785863	0.5	NP	3.8	NP	NP	NP	NP	NP	NP	NF	F8-161952	rs5987053
159087G>A	—	49785864	0.5	0.7	NP	NP	NP	NP	NP	NP	NP	NF	NF	NF

**Table 2. Characteristics, status, and allelic associations of *F8* polymorphisms with FVIII:C levels (continued)**

Region, TU*	Protein*	GenBank*	m-AFs, %†									F8 variation databases‡		
			T	W	AA	Ch	SEA	J	MI	SAA	HAMSTeRS <sup>38,58</sup>	VDR <sup>53</sup>	dbSNP <sup>59</sup>	
<b>Intron 23</b>														
159874G>A§	—	49785865	1.4	1.4	3.8	NP	NP	NP	NP	NP	NP	NF	F8-163006	NF
<b>Intron 24</b>														
162013G>T§	—	49785866	3.3	5.0	NP	NP	NP	NP	NP	NP	NP	NF	F8-165145	NF
<b>Exon 25</b>														
162161A>G§	<i>M2238V</i>	49785867	1.8	NP	15.4	NP	NP	NP	NP	NP	NP	+	F8-165293	rs17051967
<b>Intron 25</b>														
162475T>C§	—	49785868	4.1	5.6	NP	NP	NP	NP	NP	12.5	NP	NF	F8-165607	NF
<b>3' UTR</b>														
185156C>T§	—	49785869	0.5	NP	3.8	NP	NP	NP	NP	NP	NP	NF	F8-188288	rs5986887
186341G>A§	—	49785870	0.5	NP	NP	NP	NP	12.5	NP	NP	NP	NF	NF	NF
186506A>G§	—	49785871	0.5	0.7	NP	NP	NP	NP	NP	NP	NP	NF	NF	NF
186602C>T§	—	49785872	0.5	0.7	NP	NP	NP	NP	NP	NP	NP	NF	NF	NF
186799G>A§	—	49785873	24.8	17.4	76.9	NP	13.3	NP	50.0	50.0	+	F8-189931	rs1050705	
<b>3' DNA</b>														
186987T>G§	—	49785874	0.9	NP	9.1	NP	NP	NP	NP	NP	NP	NF	F8-190119	NF
187064T>C§	—	49785875	1.8	NP	15.4	NP	NP	NP	NP	NP	NP	NF	F8-190196	NF

Except for the mild hemophilic missense mutation W255C, the coding region variants in italics represent ns-SNPs that encode the amino-acid substitutions listed.

T indicates total; W, the 68 SW and 18 WA subjects, which were considered together; HAMSTeRS, Hemophilia A Mutation, Structure, Test, and Resource Site; VDR, Variation Discovery Resource (UW-FHCRC); dbSNP (NCBI); —, polymorphisms located outside of the *F8* coding sequence; NP, not polymorphic; NF, not found; NA, not applicable; and +, a previously known polymorphism.

\*Polymorphisms are designated by genic region; nt alleles and position in the TU (start site and adjacent 5' base are designated as nt 1 and –1, respectively); amino acid alleles and position in the mature plasma protein; and GenBank number for dbSNP submission.

†Estimated from subjects in either the total VDG or each of its racial groups separately. Genotypic data were not complete for all variants and resulted in denominators that varied from the maximum number of distinct X-chromosomes (see "Patients, materials, and methods").

‡Public databases with *F8* polymorphisms.

§Twenty-one SNPs contained in the 9 amplicons generated to genotype GAIT subjects for the subset of 12 that were located in a functional-gene region and variable in at least white individuals.

## Covariate selection and genotype association analyses

Measured variables available in GAIT include age, sex, smoking status, ABO genotype, OC status, BMI, DM status, and plasma levels of TC, high-density lipoprotein (HDL), LDL, very-low-density lipoprotein (VLDL), TG, lipoprotein-a, fibrinogen, VWF antigen (VWF:Ag), and FIX:C.<sup>31,35</sup> We used indicator variables to represent ABO genotypes, with OO as the reference level, and modeled separate mean FVIII:C levels for each.<sup>35</sup> In addition, smoking and OC status were represented with dichotomous variables that indicated any use versus the reference level of no use. We selected as covariates all measured variables associated with FVIII:C levels, since these factors could confound the relationship of interest if by chance their data were distributed differentially among the genotypes under evaluation. Since GAIT is a family-based study, we used likelihood-based variance-components analysis, as implemented in SOLAR, to account for the nonindependence between study subjects.<sup>60</sup> Further, we performed bivariate analyses with the endogenous factors and FVIII:C levels to investigate potential common genetic sources of variance.<sup>62</sup>

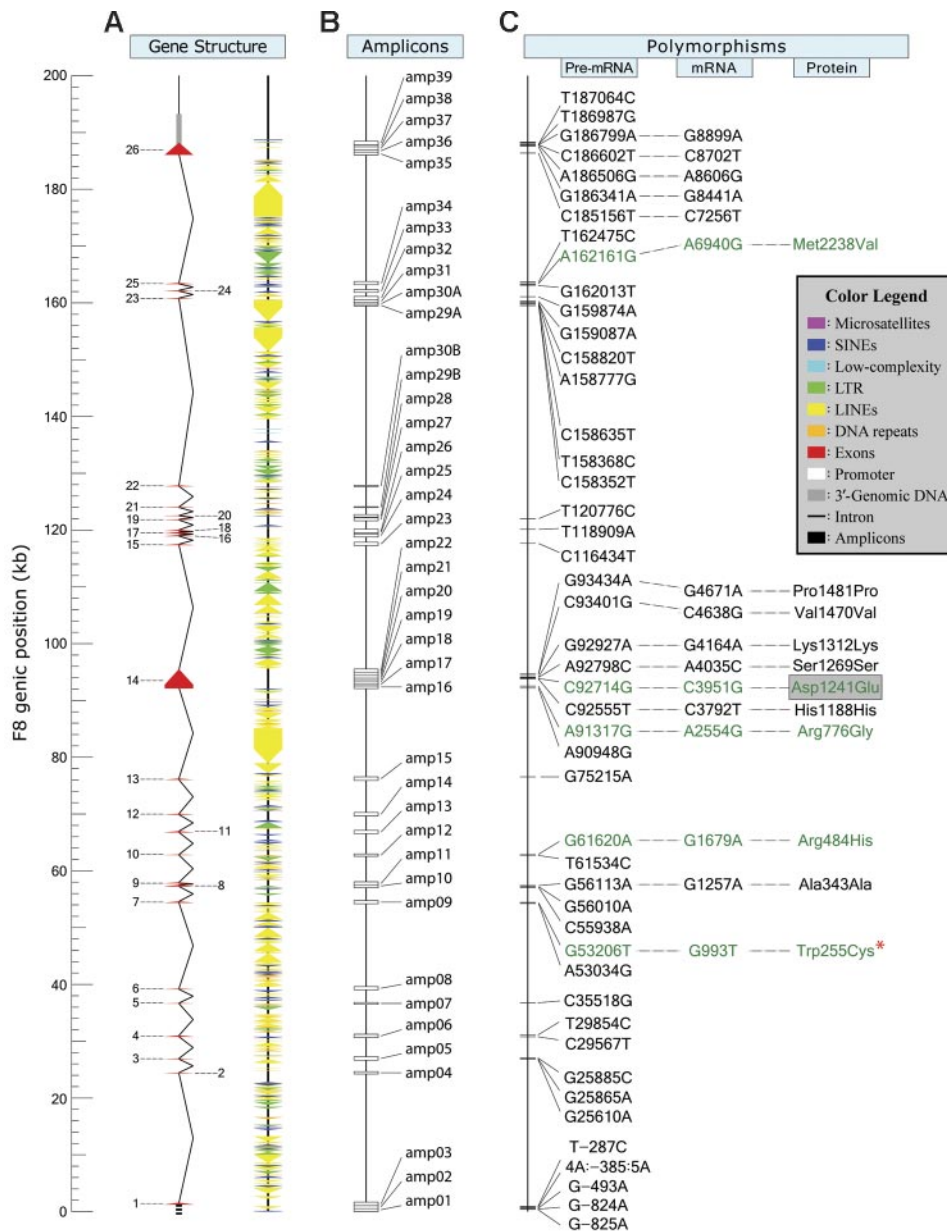
Since all *F8* polymorphisms were biallelic, containing a major (M) and minor (m) allele, with 3 genotypes in females (M/M, M/m, m/m) and 2 in males (M/Y, m/Y) (Table 2), we combined GAIT subjects by genotype into 3 groups (M/M and M/Y; M/m; m/m and m/Y) and used a dosage-compensation model in which hemizygous males had a gene effect equivalent to females homozygous for the same allele. We represented each group with the value of 1, 0, or –1, respectively, using an additive model in which the mean for heterozygous females is halfway between the means for females homozygous for the M and m alleles, and the means for male hemizygotes and females homozygous for the corresponding allele are equivalent. To evaluate the relationship between *F8* and FVIII:C level, we performed marginal measured-genotype association analysis, as implemented in SOLAR.<sup>63</sup> In the first analysis, we tested each polymorphism separately for genotype-specific differences in mean FVIII:C levels using an initial complex model that incorporated as covariates age, age<sup>2</sup>, sex, age × sex, age<sup>2</sup> × sex, ABO

genotype, and smoking and OC status. We reanalyzed all polymorphisms found to be suggestively associated with FVIII:C level ( $P \leq .10$ ) using a more complex model that included as covariates all potentially confounding measured variables. Finally, we performed additional marginal measured-genotype analyses in which paired endogenous factors that shared common genetic influences with FVIII:C level were excluded singly and jointly. This strategy allows for potential individual actions without enforcing the same mechanistic effect on every polymorphism. For these analyses, we used linear splines with knots at 15 and 50 years (ie, approximately corresponding to puberty and menopause) to represent age,<sup>64-66</sup> allowing a more flexible control of confounding than either a linear or a quadratic representation.

## Results

### Multiple unknown *F8* polymorphisms identified

To determine whether *F8* is a determinant of FVIII variability, we first identified candidate functional polymorphisms in 222 potentially distinct alleles contributed by the 137 unrelated VDG subjects. We examined 41 amplicons containing 1195 bp of promoter, 7054 bp of the 7056 bp of total coding sequence, all 1953 bp of untranslated exonic sequence, 4551 bp out of the approximately 5000 bp of total junctional-intronic sequence, approximately 2.3% of all deep intronic sequences (172 921 bp total), and 309 bp of 3'-flanking genomic DNA (Figure 1).<sup>46,57</sup> These amplicons contained 19 157 bp and included every known functional region except 2 bp of exon 22, the 5' splice junction (SJ) of intron 22, and 78% to 90% of the SJ sequences from the 3' ends of introns 20 and 21 and the 5' end of intron 21 (Table 1). By analyzing the approximately 188 kb of genomic DNA between the forward and



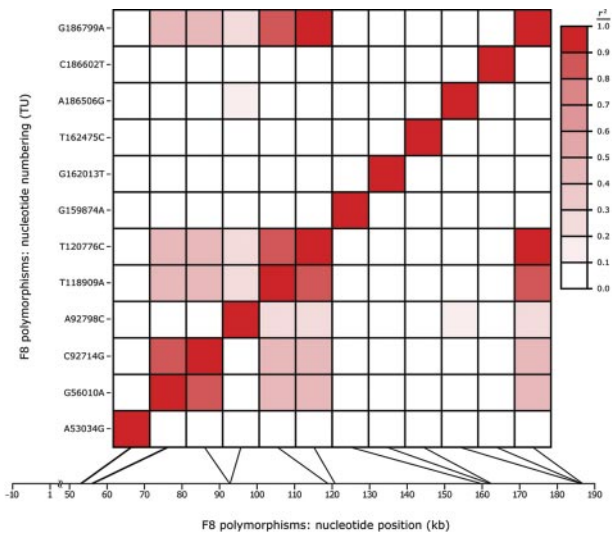
reverse primers of the first and last amplicons, respectively, with RepeatMasker (<http://repeatmasker.org>), we found that approximately 31% of *F8* is composed of nonrepetitive sequence, of which approximately 31% was covered in our scan.

We identified 47 variable sites (Figure 1) including 45 SNPs and a 1-bp insertion/deletion polymorphism (INDEL). The m allele of 53206G>T, which was carried by a female SEA subject and found in only 1 of the 222 X-chromosomes examined, encodes W255C, a missense mutation previously identified in Ch patients with mild hemophilia A.<sup>38</sup> These variants, of which 18 were unknown (SeattleSNPs<sup>53</sup>; The Single Nucleotide Polymorphism Database of Nucleotide Sequence Variation<sup>59</sup>; The Haemophilia A Mutation, Structure, Test and Resource Site<sup>58</sup>; Table 2), were located in all functional genic regions. Four coding sequence polymorphisms were nonsynonymous SNPs (ns-SNPs) encoding the amino-acid substitutions R484H, R776G, D1241E, and M2238V (Figure 1). Although when estimated in the combined VDG (Table 2) m-allele

frequencies (m-AFs) ranged from 0.5% to 24.8%, most *F8* polymorphisms were singletons with infrequent and/or racially restricted minor alleles.

#### m-AFs vary with race

As the first step to investigate whether heritable FVIII determinants may vary across populations, a small number of nonwhite subjects from 6 racial groups were also studied. Specifically, we resequenced the same *F8* regions in 51 individuals (31 female, 20 male) including 16 AA, 10 Ch, 10 SEA, 5 J, 5 MI, and 5 SAA subjects. While the m-AFs of a few polymorphisms (eg, 92798A>C) were similar across different races, most varied substantially. Indeed the m-alleles of 38 polymorphisms were found in only 1 racial group. Only 16 of the 46 polymorphisms were variable in either WA or SW subjects, or both, despite examining 144 white X-chromosomes, approximately 6 times more than the number from African



**Figure 2. Pattern and degree of LD across *F8*.** Allelic associations among the 12 functional region *F8* polymorphisms variable in at least white VDG subjects were estimated pairwise, based on allele frequencies in the GAIT founders, by calculating both  $D'$  and  $r^2$ . While only  $r^2$  values are displayed here, similar estimates of LD were obtained using  $D'$ . The color legend semiquantitatively illustrates  $r^2$ ; white indicates linkage equilibrium ( $r^2 = 0$ ), and red indicates complete LD ( $r^2 = 1$ ). In order to reveal the genic location of each variation with respect to both its position in the transcription unit (y-axis) and physical spacing along the gene, based on hg17 (x-axis), the graph is plotted symmetrically such that the information in analogous positions above and below the diagonal pairwise line-of-identity are identical.

Americans, the second most abundantly sampled racial group studied. Moreover the m-alleles for 30 of the 38 potentially racially restricted polymorphisms were found in a nonwhite racial group. Twenty-two of these nonwhite racially restricted polymorphisms were variable only in individuals of African descent; of these, the m-AF was greater than 5% for 8 and greater than 10% for 2 (162161A>G, 187064T>C). Furthermore, 2 of the 4 ns-SNPs, 61620G>A (R484H) and 162161A>G (M2238V), were AA restricted. Because the m-allele of 91317A>G (R776G) was found in a single Ch individual, 92714C>G (D1241E) is the only white ns-SNP. Although 8 SNPs were found in more than 1 racial group, their m-AFs varied widely. For example, the G-allele of 92714C>G was less frequent in the combined VDG and among individuals from each race separately, except those of African descent; in AA subjects it was actually the M-allele with a frequency of approximately 73% (Table 2). A similar pattern of differences in m-AF was observed for 3 additional SNPs (118909T>A, 120776T>C, and 186799A>G).

#### Low LD across *F8*

To survey LD across *F8* we evaluated the 12 functional-region SNPs that were variable in (at least) the white subset of VDG subjects and/or the GAIT members whose parents were not enrolled in GAIT because the nonwhite populations were insufficiently sampled ( $n \leq 16$ ) to accurately estimate race-specific AF. We found a low overall degree of LD, whether using  $r^2$  (Figure 2) or  $D'$  (not shown), with 6 sites being in linkage equilibrium with all other variants. Nevertheless, 120776T>C and 186799G>A exhibited high LD. The alleles of 92714C>G, which encodes the only white ns-SNP (D1241E), revealed moderate LD ( $0.4 \leq r^2 < 0.5$ ) with 118909A>T, 120776C>T, and 186799A>G and high LD ( $r^2 = 0.83$ ) with 56010G>A (Figure 2).

#### FVIII:C level is broadly variable in GAIT

Genomic DNA from 394 GAIT subjects (213 female, 181 male) with duplicate FVIII:C measures were available for this study. Data from 2 females, with FVIII:C levels greater than 4 standard deviations (SDs) beyond the mean, were excluded from further analysis. FVIII:C levels in the remaining 392 subjects exhibited a mean ( $\pm$  SD) of  $150.7 \text{ IU dL}^{-1}$  ( $\pm 52.2 \text{ IU dL}^{-1}$ ) and a greater than 6-fold concentration range ( $47\text{--}338 \text{ IU dL}^{-1}$ ).

#### *F8* is a modest determinant of FVIII variance

Before investigating the influence of *F8* on FVIII, we attempted to identify variables for which it is necessary to control to avoid a potential bias. Sex, age, smoking status, ABO genotype, DM status, BMI, OC status, and plasma levels of fibrinogen, VWF:Ag, FIX:C, TC, LDL, VLDL, and TG were the measured variables available for this analysis. Although plasma C-reactive protein (CRP) concentration, the most commonly used measure of inflammation, was not assayed in GAIT, fibrinogen level may be a suitable surrogate marker.<sup>67-69</sup> As GAIT subjects are SW, we investigated the relationship between FVIII:C levels and the subset of white *F8* polymorphisms. Although the m-allele for 16 of the 45 SNPs were found in at least 1 white VDG subject (Table 2), 4 (including 158352C>T, 158368T>C, 158777A>G, and 159087G>A) were located more than 500-bp upstream from the 3' SJ of intron 22, a region not known to be essential for *F8* function in vivo (Figure 1). Because LD across *F8* was irregular and low overall (Figure 2), we genotyped all 12 functional-region white SNPs in GAIT and evaluated their influence on FVIII:C levels separately using marginal measured-genotype association analysis. The results from our initial analyses, in which we modeled as covariates only age, age<sup>2</sup>, sex, age  $\times$  sex, age<sup>2</sup>  $\times$  sex, ABO genotype, and smoking and OC status, demonstrated significantly different genotype-specific mean FVIII:C levels ( $P < .05$ ) for both 56010G>A, a noncoding SNP located 27 nucleotides upstream from the 3' SJ of intron 7, and 92714C>G, the only white ns-SNP, which is located in exon 14 and substitutes glutamate for aspartate at residue 1241 in the B domain (Figure 1). Although both SNPs represent potential FVIII QTNs, we are unable to decipher which is the true functional variant from this study because their alleles were strongly associated within GAIT founders (Figure 2).

To explore further the possibility that either 92714C>G or 56010G>A differentially influences FVIII:C level, we performed additional marginal measured-genotype association analyses using more complex models. We chose to present only the results for 92714C>G because in addition to being in high LD with 56010G>A, it encodes the only nonhemophilic amino-acid substitution (D1241E) found in white individuals and had more complete data available (ie, 6 subjects had missing genotypes for 56010G>A). In the final analysis of D1241E (92714C>G) we incorporated as covariates all measured variables associated with FVIII:C level to control for confounding. In a separate bivariate analysis of FVIII:C with both FIX:C and VWF:Ag, we found potential overlap in the sets of (unknown) genes that may affect their trait levels. Thus we present the 4 analyses in Table 3. We excluded FIX:C (model 3), VWF:Ag (model 2), and the pair jointly (model 1), such that they were nested within model 4. Results from model 1 demonstrated a significant association between D1241E and FVIII in which each D-allele additively increased the mean FVIII:C level by  $14.3 \text{ IU dL}^{-1}$  ( $P = .016$ ).

**Table 3. Multivariate analysis of 92714C>G (D1241E) and FVIII:C levels in the GAIT population**

Covariate*	Model 1		Model 2		Model 3		Model 4	
	$\beta$ †	<i>P</i> ‡	$\beta$ †	<i>P</i> ‡	$\beta$ †	<i>P</i> ‡	$\beta$ †	<i>P</i> ‡
Age, y	-4.12	.089	-3.43	.178	-3.75	.065	-3.33	.090
Sex	4.43	.364	6.45	.187	13.72	.001	13.41	.001
Age × sex	-1.33	.729	-4.93	.250	-3.37	.322	-3.73	.257
A15§	3.62	.152	3.00	.258	3.44	.102	3.03	.136
A15 × sex	2.12	.596	5.62	.206	3.88	.272	4.14	.225
A50§	3.48	< .001	3.13	< .001	1.44	.013	1.58	.005
A50 × sex	-2.78	.004	-2.30	.015	-1.48	.058	-1.42	.059
OC status	-23.82	.056	-23.80	.080	-11.22	.298	-15.10	.148
Smoking status	-11.45	.021	-13.48	.008	-5.84	.153	-8.27	.038
AA¶	25.85	.002	27.61	.001	10.36	.135	11.47	.084
AB¶	26.03	.109	35.41	.047	11.23	.444	9.91	.482
AO¶	21.80	< .001	15.78	.006	3.75	.447	2.71	.568
BO¶	26.91	.002	29.53	.001	15.56	.032	16.97	.015
DM status	-6.11	.618	-25.36	.101	-4.01	.746	-9.45	.470
Fibrinogen level, IU dL <sup>-1</sup>	13.56	< .001	9.17	.012	2.55	.400	0.37	.901
BMI, kg m <sup>-2</sup>	0.22	.718	-0.46	.471	0.85	.085	-0.05	.926
VLDL level, mM	45.58	< .001	22.49	.082	21.12	.039	8.60	.398
FIX:C level, IU dL <sup>-1</sup>	N/A	N/A	0.56	< .001	N/A	N/A	0.45	< .001
VWF:Ag level, IU dL <sup>-1</sup>	N/A	N/A	N/A	N/A	0.73	< .001	0.66	< .001
D1241E#	14.33	.016	11.99	.049	7.26	.142	6.10	.199

For model 1, n = 361; for model 2, n = 313; for model 3, n = 307; for model 4, n = 306.

\*Endogenous, environmental, and genetic variables included in the marginal measured-genotype association analysis of the relationship between D1241E and FVIII:C levels in GAIT.

†The estimated effect ( $\beta$ ) of a given covariate or each D-allele of the FVIII D1241E polymorphism (F8 92714C>G) on FVIII:C levels.

‡The *P* value for each  $\beta$ .

§The coefficient for splines with knots at 15 and 50 years.

||Status (yes/no) indicates either any use (smoking and oral contraception) or the presence of diagnostic criteria for DM.

¶The 4 ABO genotypes, other than OO (reference genotype), found in GAIT subjects.

#D1241E genotypes were grouped and represented with indicator variables as follows: 1 (D/D and D/Y); 0 (D/d); and -1 (d/d and d/Y).

Since data were not available for FIX:C and VWF:Ag levels from all GAIT subjects, including them as covariates resulted in smaller sample sizes for the analyses using models 2-4. However, the results of these analyses, when restricted to the 306 subjects with complete data, showed similar trends, suggesting that the effects of adding FIX:C and/or VWF:Ag were not limited to the loss of subjects excluded due to missing data. For instance, when the 306 subjects in model 4 were reanalyzed using model 1, each D-allele was estimated to additively increase the mean FVIII:C level by 11.8 IU dL<sup>-1</sup> (*P* = .042).

As expected, TG and TC levels were nearly colinear with VLDL and LDL levels, respectively, in GAIT (not shown). Furthermore, the Pearson correlation coefficients for these paired physiologically related lipid parameters were high whether we analyzed all GAIT subjects (and did not account for nonindependence between observations) or only the (independent) founders ( $\rho \geq 0.95$ ; *P* < .001). Finally, the "effect" of D1241E on FVIII:C level was found to be similar whether we substituted TG levels for VLDL levels or TC levels for LDL levels as covariates in our analyses (not shown). Since removing both LDL levels and TC levels had a negligible impact upon the estimated effect of D1241E on FVIII (14.0 IU dL<sup>-1</sup>; *P* = .019), these variables do not appear to confound the relationship under study<sup>70</sup> and therefore were not included in our final analysis (Table 3). In contrast, removal of both VLDL and TG levels resulted in a noticeable change (not shown). While the choice between using TG or VLDL levels was arbitrary and the results were comparable, we present the findings obtained with VLDL level as a covariate (Table 3), since fewer subjects had missing data for this measurement.

## Discussion

We performed this study to identify genetic determinants contributing to the broad normal range for FVIII:C level,<sup>5,6</sup> a quantitative trait that influences thrombosis risk,<sup>35</sup> both venous<sup>1-3</sup> and arterial.<sup>2,4</sup> Despite involving a substantial heritable component,<sup>31</sup> functional allelic variants affecting the interindividual variability in this trait have been identified in only 1 gene.<sup>22</sup> While *F8* represents an obvious candidate, the negative findings from previous studies using linkage analysis<sup>36,37,39,71</sup> have led some investigators to conclude that the encoding structural locus is not a FVIII QTL.<sup>71</sup> Linkage studies are in general, however, less powerful than investigations based on association analysis, which usually directly examine potential functional variants, or polymorphisms in high LD with nongenotyped functional sites, especially for X-linked loci like *F8* that presently can be evaluated only by single-marker linkage analysis. Although *F8* has been the focus of numerous candidate gene studies, no polymorphisms were found within the functional regions examined in these investigations.<sup>25,41-43,45</sup> Nevertheless, *F8* should not be excluded as a possible FVIII QTL because less than 25% of all known functional regions, and therefore less than 2.5% of the entire structural locus, were scanned for variants in these studies. Despite the recent discovery of approximately 100 *F8* polymorphisms by SeattleSNPs<sup>53</sup> in a high-throughput resequencing study that examined 71 X-chromosomes from 47 unrelated individuals representing 2 racial groups,<sup>53</sup> these potential FVIII determinants could not be evaluated since no phenotypic data were available.



As our first step to investigate *F8*, we resequenced all known functional regions in a collection of 222 potentially distinct alleles from 137 unrelated nonhemophilic individuals representing 7 racial groups. We identified 47 variants distributed throughout *F8*, including 45 SNPs, a 1-bp INDEL polymorphism, and the mild hemophilic missense mutation W225C, found in a single asymptomatic heterozygous female (Table 2). Despite their location in 1 of the most extensively investigated human loci, 18 variants were previously unknown. Because our study, together with the SeattleSNPs scan,<sup>53</sup> identified 119 different polymorphisms (Table 2), *F8* may be as variable as other human genes. Furthermore, because the 4 SNPs 61620G>A, 91317A>G, 92714C>G, and 162161A>G encode the nonsynonymous substitutions R484H, R776G, D1241E, and M2238V, respectively (Figure 1), wild-type FVIII is a variable protein in nonhemophilic populations and not monomorphic as long thought.<sup>6</sup> Although SeattleSNPs identified 98 polymorphisms,<sup>53</sup> only 27 were located in the regions examined in this study; the larger number of total variants found is likely due to the fact that approximately 40 kb of deep intronic sequence was investigated.

The distribution of FVIII:C levels in GAIT displayed the typical broad variability observed for this trait.<sup>5,6,41</sup> Because the degree of LD across *F8* was found to be weak overall (Figure 2), we separately evaluated the 12 functional-region SNPs variable in white individuals (Table 2) for potential contributions to FVIII variability by measured-genotype association analysis. Preliminary analyses (not shown), in which we accounted for the potential confounders age, sex, ABO genotype, and smoking and OC status as covariates, demonstrated that FVIII:C level was significantly associated ( $P = .007$ ) with only 92714C>G (D1241E).<sup>39</sup> Scanavini et al<sup>40</sup> confirmed this association in a case-control study of women with idiopathic thrombophilia. Specifically, they reported a higher mean FVIII:C level in female subjects with the C/C genotype (D/D phenotype) compared with those with either C/G or G/G genotypes (D/E or E/E phenotypes). Because no other *F8* polymorphisms were (i) found in our preliminary analyses to be significantly associated with FVIII levels or the alleles of D1241E,<sup>40</sup> or (ii) investigated by Scanavini et al,<sup>40</sup> the structurally distinct proteins encoded by this ns-SNP are likely to be functionally distinct and contribute to FVIII:C variability. However, when reanalyzed only among GAIT founders (Figure 2), 92714C>G demonstrated a high LD ( $r^2 = 0.84$ ) with 56010G>A, an SNP that had incomplete genotypic data in the preliminary analysis.<sup>39</sup> Since 56010G>A is located within a potentially functional segment of the 3' SJ of intron 7 (ie, position -27) and was found to have a significant association ( $P < .05$ ) with FVIII:C level when reanalyzed using complete data (not shown), we cannot, based on data from this study alone, unequivocally establish which of these variants, if either (see below), is the true QTN.

To evaluate further the relationship of these 2 SNPs with FVIII, we controlled for all potentially confounding variables available in GAIT. Support for an association of age with FVIII:C level is among the strongest.<sup>7-10,12,72</sup> Results from our analyses of FVIII:C level versus age are consistent with this whether assessed alone (not shown) or with covariates (Table 3). In a large cross-sectional study with a narrower and younger age range than GAIT subjects, Green et al<sup>17</sup> did not observe an association between age and FVIII levels. Because Miller et al<sup>73</sup> also found no association in a study of women of similar age, we used splines at 15 and 50 years to allow a less rigorous imposition on the data across the broad age range in GAIT (2-87 years). This approach allows flexibility in the effects of

age on FVIII:C levels in the young and the old instead of modeling, for instance, a quadratic effect across both, in which the age data for older subjects influences the estimated response for the young and vice versa. The relationship between sex and FVIII:C level remains uncertain, with some studies demonstrating significant associations<sup>7,9</sup> and others not<sup>8,18</sup>; we did not find evidence of a genotype-by-sex interaction. The impact of OC status on FVIII:C level has not been established.<sup>8,10,74,75</sup> Although we found an apparently large effect for OC, comparable to that observed for ABO genotype (Table 3), oral contraceptives were used by only a small number of female GAIT subjects. Two large cross-sectional studies found modestly lower FVIII:C levels in smokers,<sup>7,15</sup> whereas other investigations observed no association with this variable.<sup>8,17</sup> We found that smokers have a lower mean FVIII:C level but could not characterize this relationship further because detailed smoking information is lacking in GAIT. Since FVIII is an acute-phase reactant, its plasma concentration may be transiently affected by various conditions, including inflammation.<sup>69</sup> Although CRP level was not measured, we incorporated the surrogate marker fibrinogen level as a covariate in our analyses (Table 3), similarly to Kreuz et al.<sup>68</sup>

There is strong support for an association of ABO blood type with both FVIII:C level<sup>17-21</sup> and VWF:Ag level.<sup>24-26,28,29</sup> Using linkage analysis, Souto et al<sup>22</sup> demonstrated that the functional ABO polymorphisms responsible for the antigens of this blood group are likely to directly influence the levels of both hemostasis traits. Until recently,<sup>39,40</sup> these polymorphisms were the only known QTNs for FVIII. Although not completely understood, the effect of ABO genotype or phenotype on these 2 proteins may involve the same mechanism. Indeed, in a bivariate analysis of FVIII:C and VWF:Ag levels, we found potential overlap in the sets of genes associated with variability in their plasma levels. While less studied, associations between FIX and FVIII have also been reported.<sup>24</sup> Since neither FIX nor FVIII functions outside of intrinsic Xase, regulatory mechanisms might exist that coordinate their levels to prevent a potential cellular resource misallocation. Consistent with this, our findings from a bivariate analysis demonstrated a potential overlap in the set of genes that affect FIX and FVIII levels. Including factors that share pleiotropic influences with the trait of interest as covariates decreases the power to detect gene variants that influence both phenotypes jointly but may increase power to detect those with unique influences on the focal trait. If *F8* variants are in part responsible for the observed pleiotropy between FVIII and FIX or VWF, then model 1, excluding variables that are genetically correlated with FVIII, provides the most power. In contrast, if unidentified loci contribute to their shared heritability, then model 4 is preferable because it incorporates the known correlates of FVIII that may confound analysis. Without knowing the genetic architecture underlying pleiotropy in these traits, we focus on the results from model 1—the most conservative from the perspective of identifying functional gene variants, the goal of the present study—which demonstrated an additive increase in mean FVIII:C level of 14.3 IU dL<sup>-1</sup> per D-allele (Table 3). Scanavini et al<sup>40</sup> recently reported similar findings in that subjects with the C/C genotype (D/D phenotype) had a 19 IU dL<sup>-1</sup> higher unadjusted mean FVIII:C level than the group of subjects with either a C/G or G/G genotype (D/E or E/E phenotype).

Despite strong allelic association between 56010G>A and 92714C>G, we predict that 92714C>G represents the true QTN because it encodes structurally distinct wild-type FVIII molecules

(D1241E) that represent the only 2 forms of the protein expressed by white individuals, the population sampled in GAIT. Furthermore, while 56010G>A is near the 3' end of intron 7, it is not located in essential cis-elements of the 3' SJ. However, genetic studies in other populations with different patterns of allelic associations and/or in vitro functional assays are necessary to exclude the possibility that 56010G>A, or a nongenotyped functional polymorphism also in high LD with 92714C>G, is the FVIII QTN. Since the B domain undergoes numerous posttranslational modifications that limit FVIII expression,<sup>38,76,77</sup> and yet represents a portion of the protein known to be dispensable for its procoagulant activity once in the circulation,<sup>78-81</sup> we hypothesize that 92714C>G (D1241E) modulates FVIII:C levels by differentially affecting the secretion of this molecule.<sup>82,83</sup> Because D1241E could also influence FVIII clearance and/or activation, elucidating the molecular basis for its effect on FVIII:C level will require in vitro investigations since genetic studies alone cannot differentiate between these biologically plausible alternative mechanisms.

Regardless of whether 92714C>G is the actual QTN or in strong LD with it, we estimate that *F8* accounts for approximately 10% of the total FVIII variability in GAIT (not shown). This is consistent with the lack of evidence for linkage of FVIII:C levels to X-linked microsatellites within or near *F8*, which were genotyped in prior genome-wide screens,<sup>36,37</sup> including 1 performed in GAIT,<sup>39</sup> as we would not expect to detect QTLs with effect sizes in this range. As both 92714C>G and 56010G>A were polymorphic in all racial groups studied except SEA (Table 2), this *F8*-based determinant may influence FVIII variability in these populations. Because GAIT is composed entirely of SW subjects, however, appropriately designed studies in other racial groups are necessary to determine whether our findings apply to nonwhite populations. Such studies are also required to define further the genetic architecture underlying human FVIII variability since 30 of the functional-region *F8* polymorphisms identified, including the 3 additional ns-SNPs R484H, R776G, and M2238V, were not evaluated as potential FVIII determinants because they were not variable in white individuals (Table 2). In summary, it is important to emphasize that we resequenced less than 15% of *F8* and not all GAIT founders were enrolled for study. Therefore, we did not examine entirely every potentially distinct *F8* allele segregated in GAIT. Thus, the possibility that this locus contains additional FVIII determinants, either as common polymorphisms<sup>84</sup> or rare vari-

ants,<sup>85</sup> since both have been shown to contribute substantially to complex trait variability at the population level, cannot be excluded.

## Acknowledgments

We are grateful to all individuals who participated in this study and the physicians who enrolled them. We would also like to thank Cynthia Channell, Ming Shen, and Thuy Tran for excellent technical assistance; Art Thompson, Earl Davie, and Bruce Evatt for helpful discussions and/or reviewing the manuscript; and both Alexander (Sandy) Duncan and Pete Lollar for early support and mentoring.

This work was supported in part by the following grants: NIH-MH-59490 (J.B.); NIH-HL-70751 (L.A.); NIH-HL-71130 and NIH-HL-68016 (T.E.H.), Fondo Investigación Sanitaria (FIS)-02/0375, SAF2002-03449 (FEDER, Spanish Ministry of Science and Technology), FIS-RECAVA C03/01 (Fundació "La Caixa" and Fundació d'Investigació Sant Pau). J.M.S. and A.B. were supported by FIS-99/3048 and FIS-01/A046 (Fondo Investigación Sanitaria; Spanish Ministry of Health).

## Authorship

Contribution: J.C.S., J.B., J.F., J.M.S., L.A., and T.E.H. contributed the study concept and design; J.C.S., J.F., and S.T.W. enrolled the study subjects; K.R.V., D.K.M., J.C.S., J.F., J.M.S., and T.E.H. acquired the data; K.R.V., D.K.M., D.M.W., A.B., K.F., J.P., T.S., J.B., S.P., W.D.F., L.A., and T.E.H. provided data analysis and interpretation; K.R.V., J.M.S., W.D.F., L.A., and T.E.H. drafted the manuscript; and K.R.V., D.K.M., M.K., W.D.F., L.A., and T.E.H. provided critical manuscript revision for important intellectual content.

Conflict-of-interest disclosure: We certify that we have no affiliation with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the manuscript. All financial support for this research project is identified in "Acknowledgments."

K.R.V. and D.K.M. contributed equally to this work.

Correspondence: Tom E. Howard, Southwest Foundation for Biomedical Research, Department of Genetics, Rm 12.306, Slick-Urschel Bldg, San Antonio, TX 78245-0549; e-mail: thoward@sfbgenetics.org.

## References

- Koster T, Blann AD, Briet E, Vandenbroucke JP, Rosendaal FR. Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep-vein thrombosis. *Lancet*. 1995;345:152-155.
- Bank I, Libourel EJ, Middeldorp S, et al. Elevated levels of FVIII:C within families are associated with an increased risk for venous and arterial thrombosis. *J Thromb Haemost*. 2005;3:79-84.
- Kyrle PA, Minar E, Hirschl M, et al. High plasma levels of factor VIII and the risk of recurrent venous thromboembolism. *N Engl J Med*. 2000;343:457-462.
- Cortellaro M, Boschetti C, Cofrancesco E, et al. The PLAT Study: hemostatic function in relation to atherothrombotic ischemic events in vascular disease patients. Principal results. *PLAT Study Group. Progetto Lombardo Atero-Trombosi (PLAT) Study Group. Arterioscler Thromb*. 1992;12:1063-1070.
- Wells PS, Langlois NJ, Webster MA, Jaffey J, Anderson JA. Elevated factor VIII is a risk factor for idiopathic venous thromboembolism in Canada: is it necessary to define a new upper reference range for factor VIII? *Thromb Haemost*. 2005;93:842-846.
- Thompson AR. Structure and function of the factor VIII gene and protein. *Semin Thromb Hemost*. 2003;29:11-22.
- Conlan MG, Folsom AR, Finch A, et al. Associations of factor VIII and von Willebrand factor with age, race, sex, and risk factors for atherosclerosis: The Atherosclerosis Risk in Communities (ARIC) Study. *Thromb Haemost*. 1993;70:380-385.
- Balleisen L, Bailey J, Epping PH, Schulte H, van de Loo J. Epidemiological study on factor VII, factor VIII and fibrinogen in an industrial population, I: baseline data on the relation to age, gender, body-weight, smoking, alcohol, pill-using, and menopause. *Thromb Haemost*. 1985;54:475-479.
- Tracy RP, Bovill EG, Fried LP, et al. The distribution of coagulation factors VII and VIII and fibrinogen in adults over 65 years: results from the Cardiovascular Health Study. *Ann Epidemiol*. 1992;2:509-519.
- Lowe GD, Rumley A, Woodward M, et al. Epidemiology of coagulation factors, inhibitors and activation markers: the Third Glasgow MONICA Survey, I: illustrative reference ranges by age, sex and hormone use. *Br J Haematol*. 1997;97:775-784.
- Lindeberg S, Berntorp E, Nilsson-Ehle P, Terent A, Vessby B. Age relations of cardiovascular risk factors in a traditional Melanesian society: the Kitava Study. *Am J Clin Nutr*. 1997;66:845-852.
- Yarnell JW, Sweetnam PM, Rumley A, Lowe GD. Lifestyle factors and coagulation activation markers: the Caerphilly Study. *Blood Coagul Fibrinolysis*. 2001;12:721-728.
- Schlitt AF, Grandjean P, Donnez J, Lavenne E.

- Large increase in plasmatic 11-dehydro-TxB2 levels due to oral contraceptives. *Contraception*. 1995;51:53-58.
14. Chang SJ, Koh SB, Cha BS, Park JK. Job characteristics and blood coagulation factors in Korean male workers. *J Occup Environ Med*. 2002; 44:997-1002.
  15. Geffken DF, Cushman M, Burke GL, Polak JF, Sakkinen PA, Tracy RP. Association between physical activity and markers of inflammation in a healthy elderly population. *Am J Epidemiol*. 2001; 153:242-250.
  16. Cushman M, Yanez D, Psaty BM, et al. Association of fibrinogen and coagulation factors VII and VIII with cardiovascular risk factors in the elderly: the Cardiovascular Health Study. *Cardiovascular Health Study Investigators*. *Am J Epidemiol*. 1996;143:665-676.
  17. Green D, Ruth KJ, Folsom AR, Liu K. Hemostatic factors in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Arterioscler Thromb*. 1994;14:686-693.
  18. Jeremic M, Weisert O, Gedde-Dahl TW. Factor VIII (AHG) levels in 1016 regular blood donors: the effects of age, sex, and ABO blood groups. *Scand J Clin Lab Invest*. 1976;36:461-466.
  19. McCallum CJ, Peake IR, Newcombe RG, Bloom AL. Factor VIII levels and blood group antigens. *Thromb Haemost*. 1983;50:757.
  20. Mohanty D, Ghosh K, Marwaha N, Kaur S, Chaudhan AP, Das KC. Major blood group antigens: a determinant of factor VIII levels in blood? *Thromb Haemost*. 1984;51:414.
  21. Orstavik KH, Magnus P, Reisner H, Berg K, Graham JB, Nance W. Factor VIII and factor IX in a twin population: evidence for a major effect of ABO locus on factor VIII level. *Am J Hum Genet*. 1985;37:89-101.
  22. Souto JC, Almasy L, Muniz-Diaz E, et al. Functional effects of the ABO locus polymorphism on plasma levels of von Willebrand factor, factor VIII, and activated partial thromboplastin time. *Arterioscler Thromb Vasc Biol*. 2000;20:2024-2028.
  23. Balleisen L, Assmann G, Bailey J, Epping PH, Schulte H, van de Loo J. Epidemiological study on factor VII, factor VIII and fibrinogen in an industrial population. II: baseline data on the relation to blood pressure, blood glucose, uric acid, and lipid fractions. *Thromb Haemost*. 1985;54: 721-723.
  24. Cadroy Y, Daviaud P, Saivin S, Sie P, Boneu B. Distribution of 16 hemostatic laboratory variables assayed in 100 blood donors. *Nouv Rev Fr Hematol*. 1990;32:259-264.
  25. Morange PE, Tregouet DA, Frere C, et al. Biological and genetic factors influencing plasma factor VIII levels in a healthy family population: results from the Stanislas cohort. *Br J Haematol*. 2005; 128:91-99.
  26. Miller CH, Haff E, Platt SJ, et al. Measurement of von Willebrand factor activity: relative effects of ABO blood type and race. *J Thromb Haemost*. 2003;1:2191-2197.
  27. Weiss HJ, Hoyer IW. Von Willebrand factor: dissociation from antihemophilic factor procoagulant activity. *Science*. 1973;182:1149-1151.
  28. Weiss HJ, Sussman II, Hoyer LW. Stabilization of factor VIII in plasma by the von Willebrand factor: studies on posttransfusion and dissociated factor VIII and in patients with von Willebrand's disease. *J Clin Invest*. 1977;60:390-404.
  29. Tuddenham EG, Lane RS, Rotblat F, et al. Response to infusions of polyelectrolyte fractionated human factor VIII concentrate in human haemophilia A and von Willebrand's disease. *Br J Haematol*. 1982;52:259-267.
  30. van Hylckama Vlieg A, van der Linden IK, Bertina RM, Rosendaal FR. High levels of factor IX increase the risk of venous thrombosis. *Blood*. 2000;95:3678-3682.
  31. Souto JC, Almasy L, Borrell M, et al. Genetic determinants of hemostasis phenotypes in Spanish families. *Circulation*. 2000;101:1546-1551.
  32. de Lange M, Snieder H, Ariens RA, Spector TD, Grant PJ. The genetics of haemostasis: a twin study. *Lancet*. 2001;357:101-105.
  33. Kamphuisen PW, Houwing-Duistermaat JJ, van Houwelingen HC, Eikenboom JC, Bertina RM, Rosendaal FR. Familial clustering of factor VIII and von Willebrand factor levels. *Thromb Haemost*. 1998;79:323-327.
  34. Vossen CY, Hasstedt SJ, Rosendaal FR, et al. Heritability of plasma concentrations of clotting factors and measures of a prothrombotic state in a protein C-deficient family. *J Thromb Haemost*. 2004;2:242-247.
  35. Souto JC, Almasy L, Borrell M, et al. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. *Genetic Analysis of Idiopathic Thrombophilia*. *Am J Hum Genet*. 2000;67:1452-1459.
  36. Soria JM, Almasy L, Souto JC, et al. A new locus on chromosome 18 that influences normal variation in activated protein C resistance phenotype and factor VIII activity and its relation to thrombosis susceptibility. *Blood*. 2003;101:163-167.
  37. Berger M, Mattheisen M, Kulle B, et al. High factor VIII levels in venous thromboembolism show linkage to imprinted loci on chromosomes 5 and 11. *Blood*. 2005;105:638-644.
  38. Kembell-Cook G, Tuddenham EG, Wacey AI. The factor VIII structure and mutation resource site: HAMSTeRS version 4. *Nucleic Acids Res*. 1998; 26:216-219.
  39. Machiah DK, Viel K, Almasy L, et al. A common SNP in the factor VIII (f-VIII) gene encodes a conservative aspartate to glutamate substitution (Asp1241Glu) in the B-domain that influences f-VIII activity levels [abstract]. *Blood*. 2003;102: 55a. Abstract 181.
  40. Scanavini D, Legnani C, Lunghi B, Mingozzi F, Palareti G, Bernardi F. The factor VIII D1241E polymorphism is associated with decreased factor VIII activity and not with activated protein C resistance levels. *Thromb Haemost*. 2005;93: 453-456.
  41. Mansvelt EP, Laffan M, McVey JH, Tuddenham EG. Analysis of the F8 gene in individuals with high plasma factor VIII:C levels and associated venous thrombosis. *Thromb Haemost*. 1998;80: 561-565.
  42. Brummer J, Groth J, Flayeh R, Wagener C, Jung R. Absence of mutations at the APC interacting sites of factor VIII in Caucasians. *Thromb Haemost*. 2002;87:170.
  43. Hooper WC, Dillely A, Austin H, et al. Absence of mutations at APC cleavage sites Arg306 in factor V and Arg336, Arg562 in factor VIII in African-Americans. *Thromb Haemost*. 1998;79:236.
  44. Kamphuisen PW, Eikenboom JC, Rosendaal FR, et al. High factor VIII antigen levels increase the risk of venous thrombosis but are not associated with polymorphisms in the von Willebrand factor and factor VIII gene. *Br J Haematol*. 2001;115: 156-158.
  45. Roelse JC, Koopman MM, Buller HR, et al. Absence of mutations at the activated protein C cleavage sites of factor VIII in 125 patients with venous thrombosis. *Br J Haematol*. 1996;92:740-743.
  46. Gitschier J, Wood WI, Goralka TM, et al. Characterization of the human factor VIII gene. *Nature*. 1984;312:326-330.
  47. Figueiredo MS, Brownlee GG. cis-acting elements and transcription factors involved in the promoter activity of the human factor VIII gene. *J Biol Chem*. 1995;270:11828-11838.
  48. McGlynn LK, Mueller CR, Begbie M, Notley CR, Lilliecrap D. Role of the liver-enriched transcription factor hepatocyte nuclear factor 1 in transcriptional regulation of the factor V111 gene. *Mol Cell Biol*. 1996;16:1936-1945.
  49. Bertina RM, Koeleman BP, Koster T, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature*. 1994; 369:64-67.
  50. Williamson D, Brown K, Luddington R, Baglin C, Baglin T. Factor V Cambridge: a new mutation (Arg306—Thr) associated with resistance to activated protein C. *Blood*. 1998;91:1140-1144.
  51. 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;88:3698-3703.
  52. Howard TE, Marusa M, Channell C, Duncan A. A patient homozygous for a mutation in the prothrombin gene 3'-untranslated region associated with massive thrombosis. *Blood Coagul Fibrinolysis*. 1997;8:316-319.
  53. National Heart, Lung, and Blood Institute Program for Genomic Applications. SeattleSNPs. <http://pga.gs.washington.edu>. Accessed November 2004.
  54. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I: accuracy assessment. *Genome Res*. 1998;8:175-185.
  55. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II: error probabilities. *Genome Res*. 1998;8:186-194.
  56. Zerynick C, Torroni A, Sherman SL, Warren ST. Normal variation at the myotonic dystrophy locus in global human populations. *Am J Hum Genet*. 1995;56:123-130.
  57. Karolchik D, Baertsch R, Diekhans M, et al. The UCSC Genome Browser Database. *Nucl Acids Res*. 2003;31:51-54.
  58. Wacey AI, Kembell-Cook G. The Haemophilia A Mutation, Structure, Test and Resource Site. <http://europium.csc.mrc.ac.uk/WebPages/Main/main.htm>. Accessed November 2004.
  59. National Center for Biotechnology Information, National Library of Medicine. The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation. <http://www.ncbi.nlm.nih.gov/SNP/>. Accessed November 2004.
  60. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet*. 1998;62:1198-1211.
  61. Zhao JH. 2LD, GENECOUNTING and HAP: computer programs for linkage disequilibrium analysis. *Bioinformatics*. 2004;20:1325-1326.
  62. Almasy L, Dyer TD, Blangero J. Bivariate quantitative trait linkage analysis: pleiotropy versus coincident linkages. *Genet Epidemiol*. 1997;14:953-958.
  63. Blangero J, Goring HH, Kent JW Jr, et al. Quantitative trait nucleotide analysis using Bayesian model selection. *Hum Biol*. 2005;77:541-559.
  64. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge, NY: Cambridge University Press; 2003.
  65. Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology*. 1995;6:356-365.
  66. Motulsky HJ, Ransnas LA. Fitting curves to data using nonlinear regression: a practical and non-mathematical review. *FASEB J*. 1987;1:365-374.
  67. Folsom AR, Pankow JS, Tracy RP, et al. Association of C-reactive protein with markers of prevalent atherosclerotic disease. *Am J Cardiol*. 2001; 88:112-117.
  68. Kreuz W, Stoll M, Junker R, et al. Familial elevated factor VIII in children with symptomatic venous thrombosis and post-thrombotic syndrome: results of a multicenter study. *Arterioscler Thromb Vasc Biol*. 2006;26:1901-1906.

69. Reitsma PH, Branger J, Van Den Blink B, Weijer S, Van Der Poll T, Meijers JC. Procoagulant protein levels are differentially increased during human endotoxemia. *J Thromb Haemost*. 2003;1:1019-1023.
70. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott-Raven; 1998.
71. De Visser MC, Sandkuijl LA, Lensen RP, Vos HL, Rosendaal FR, Bertina RM. Linkage analysis of factor VIII and von Willebrand factor loci as quantitative trait loci. *J Thromb Haemost*. 2003;1:1771-1776.
72. Lindeberg S, Berntorp E, Carlsson R, Eliasson M, Marckmann P. Haemostatic variables in Pacific Islanders apparently free from stroke and ischaemic heart disease: the Kitava Study. *Thromb Haemost*. 1997;77:94-98.
73. Miller CH, Dille AB, Drews C, Richardson L, Evatt B. Changes in von Willebrand factor and factor VIII levels during the menstrual cycle. *Thromb Haemost*. 2002;87:1082-1083.
74. Meade TW, Brozovic M, Chakraborty R, Howarth DJ, North WR, Stirling Y. An epidemiological study of the haemostatic and other effects of oral contraceptives. *Br J Haematol*. 1976;34:353-364.
75. Kadir RA, Economides DL, Sabin CA, Owens D, Lee CA. Variations in coagulation factors in women: effects of age, ethnicity, menstrual cycle and combined oral contraceptive. *Thromb Haemost*. 1999;82:1456-1461.
76. Toole JJ, Knopf JL, Wozney JM, et al. Molecular cloning of a cDNA encoding human antihemophilic factor. *Nature*. 1984;312:342-347.
77. Vehar GA, Keyt B, Eaton D, et al. Structure of human factor VIII. *Nature*. 1984;312:337-342.
78. Langner KD, Bird RE, McCandliss R, et al. Synthesis of biologically active deletion mutants of human factor VIII:C. *Behring Inst Mitt*. 1988:16-25.
79. Eaton DL, Wood WI, Eaton D, et al. Construction and characterization of an active factor VIII variant lacking the central one-third of the molecule. *Biochemistry*. 1986;25:8343-8347.
80. Toole JJ, Pittman DD, Orr EC, Murtha P, Wasley LC, Kaufman RJ. A large region (approximately equal to 95 kDa) of human factor VIII is dispensable for *in vitro* procoagulant activity. *Proc Natl Acad Sci U S A*. 1986;83:5939-5942.
81. Burke RL, Pacht C, Quiroga M, et al. The functional domains of coagulation factor VIII:C. *J Biol Chem*. 1986;261:12574-12578.
82. Pittman DD, Tomkinson KN, Kaufman RJ. Post-translational requirements for functional factor V and factor VIII secretion in mammalian cells. *J Biol Chem*. 1994;269:17329-17337.
83. Zhang B, Kaufman RJ, Ginsburg D. LMAN1 and MCFD2 form a cargo receptor complex and interact with coagulation factor VIII in the early secretory pathway. *J Biol Chem*. 2005;280:25881-25886.
84. Soria JM, Almasy L, Souto JC, et al. Linkage analysis demonstrates that the prothrombin G20210A mutation jointly influences plasma prothrombin levels and risk of thrombosis. *Blood*. 2000;95:2780-2785.
85. Soria JM, Almasy L, Souto JC, Sabater-Lleal M, Fontcuberta J, Blangero J. The F7 gene and clotting factor VII levels: dissection of a human quantitative trait locus. *Hum Biol*. 2005;77:561-575.