A gene-centric association scan for Coagulation Factor VII levels in European and African Americans: the Candidate Gene Association Resource (CARe) Consortium


1Department of Epidemiology, University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, NC 27514, USA, 2Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27514, USA, 3University College London Genetics Institute, Department of Genetics Environment and Evolution and 4Centre for Cardiovascular Genetics, BHF Laboratories, Institute of Cardiovascular Science, University College London, London WC1E 6JJ, UK, 5Center for Applied Genomics, Abramson Research Center, The Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA, 6Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN 55454, USA, 7Department of Epidemiology, University of Washington School of Public Health, Seattle, WA 98195, USA, 8Department of Pharmaceutical Outcomes and Pharmacy, University of Florida College of Pharmacy, Gainesville, FL 32610, USA, 9Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK, 10Genetic Epidemiology Group, Department of Epidemiology and Public Health, University College London, London WC1E 6BT, UK, 11Department of Pathology and Laboratory Medicine, University of Vermont College of Medicine, Burlington, VT 05405, USA, 12National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20824, USA and 13Division of Hematology/Oncology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA

1Department of Epidemiology, University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, NC 27514, USA, 2Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27514, USA, 3University College London Genetics Institute, Department of Genetics Environment and Evolution and 4Centre for Cardiovascular Genetics, BHF Laboratories, Institute of Cardiovascular Science, University College London, London WC1E 6JJ, UK, 5Center for Applied Genomics, Abramson Research Center, The Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA, 6Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN 55454, USA, 7Department of Epidemiology, University of Washington School of Public Health, Seattle, WA 98195, USA, 8Department of Pharmaceutical Outcomes and Pharmacy, University of Florida College of Pharmacy, Gainesville, FL 32610, USA, 9Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK, 10Genetic Epidemiology Group, Department of Epidemiology and Public Health, University College London, London WC1E 6BT, UK, 11Department of Pathology and Laboratory Medicine, University of Vermont College of Medicine, Burlington, VT 05405, USA, 12National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20824, USA and 13Division of Hematology/Oncology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA

Received February 3, 2011; Revised May 13, 2011; Accepted June 6, 2011

Polymorphisms in several distinct genomic regions, including the F7 gene, were recently associated with factor VII (FVII) levels in European Americans (EAs). The genetic determinants of FVII in African Americans (AAs) are unknown. We used a 50 000 single nucleotide polymorphism (SNP) gene-centric array having dense coverage of over 2 000 candidate genes for cardiovascular disease (CVD) pathways in a community-based sample of 16 324 EA and 3898 AA participants from the Candidate Gene Association Resource (CARe) consortium. Our aim was the discovery of new genomic loci and more detailed characterization of existing loci associated with FVII levels. In EAs, we identified three new loci associated with FVII, of which APOA5 on chromosome 11q23 and HNF4A on chromosome 20q12–13 were replicated in a sample of 4289 participants from the Candidate Gene Association Resource (CARe) consortium. Our aim was the discovery of new genomic loci and more detailed characterization of existing loci associated with FVII levels. In EAs, we identified three new loci associated with FVII, of which APOA5 on chromosome 11q23 and HNF4A on chromosome 20q12–13 were replicated in a sample of 4289 participants from the Whitehall II study. We confirmed four previously reported FVII-associated loci (GCKR, MS4A6A, F7 and PROCRI) in CARe EA samples. In AAs, the F7 and PROCRI regions were significantly associated with FVII. Several of the FVII-associated regions are known to be associated with lipids and other cardiovascular-related traits. At the F7 locus, there was evidence of at least five independently associated

∗To whom correspondence should be addressed at: Department of Epidemiology, UNC Gillings School of Global Public Health, 137 E Franklin St., Suite 306, Chapel Hill, NC 27514, USA. Tel: +1 4048081718; Fax: +1 9199696980; Email: kirat@email.unc.edu

© The Author 2011. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
SNPs in EAs and three independent signals in AAs. Though the variance in FVII explained by the existing loci is substantial (20% in EA and 10% in AA), larger sample sizes and investigation of lower frequency variants may be required to identify additional FVII-associated loci in EAs and AAs and further clarify the relationship between FVII and other CVD risk factors.

INTRODUCTION

Coagulation Factor VII (FVII) is a vitamin K-dependent serine protease that is activated in response to vascular or tissue injury. Activated FVII (FVIIa) interacts with tissue factor to initiate coagulation (1). The association of FVII with cardiovascular disease (CVD) events is uncertain, with some positive and negative findings (2–6). Commonly, FVII had an association in unadjusted analyses, but this was often removed by adjusting for known CVD risk factors, especially lipids. Polymorphisms in the FVII gene (F7) have been associated with increased risk of coronary heart disease and stroke in some studies but not all (7–10).

FVII levels are highly heritable (estimates range from 53 to 63%) and polymorphisms in F7 account for approximately a third of the variance of FVII levels in European Americans (EAs) (11–13). A recent genome-wide association study (GWAS) in EAs confirmed that polymorphisms in the F7 gene region on chromosome 13q34 are strongly associated with FVII levels and also identified novel associations in four additional genomic regions: GCKR, ADH4, MS4A6A and PROCR (14). Whether any of these or other genetic loci are associated with FVII levels in African Americans (AAs) is unknown.

The purpose of this study was to (i) perform more detailed characterization of FVII-associated loci and fine-mapping of previously discovered associations in EAs; and (ii) identify genes associated with FVII levels in AAs. We analyzed samples from a consortium of community-based prospective cohort studies [the Candidate Gene Association Resource (CARe)] with EA and AA participants genotyped using the custom 50K single nucleotide polymorphism (SNP) cardiovascular gene-centric ITMAT Broad-CARE (IBC) array. The IBC array has dense coverage of over 2000 CVD candidate genes and includes SNPs selected to ensure that common patterns of genetic variation in populations of European and African descent are well-captured.

RESULTS

Participant characteristics

The characteristics of our EA and AA study populations are reported in Table 1. Prior to exclusion of individuals due to low genotyping call rates, there were a total of 16,352 EAs and 3948 AAs. The mean age at FVII measurement ranged from 30 [Coronary Artery Risk in Young Adults (CARDIA)] to 73 [Cardiovascular Health Study (CHS)] years. Mean FVII levels were higher in the studies which sampled older individuals. The prevalence of cardiovascular risk factors tended to be higher among AAs than EAs. FVII levels did not consistently differ by race, with higher levels found among AAs compared with EAs in CARDIA. After filtering individuals with less than 90% genotype call rate, our final meta-analysis sample sizes were 16,324 EA and 3,898 AA subjects.

FVII association results in EAs

In EAs, 7 discrete regions spanning 12 candidate genes on the IBC array contained SNPs significantly associated with FVII levels (Table 2). Four of the FVII association signals in EAs correspond to loci previously reported in a GWAS of three of the same cohorts (GCKR on chromosome 2p23, MS4A42-MS4A6A on chromosome 11q12, F7–F10 on chromosome 13q34 and PROCR on chromosome 20q11.2) (14). The addition of the CARDIA cohort to the three cohorts used in Smith et al. [Atherosclerosis Risk in Communities (ARIC), CHS and Framingham Heart Study (FHS)] strengthened these associations. A fifth FVII-associated locus reported by Smith et al., which spans two alcohol dehydrogenase (ADH) genes, ADH4 and ADH5, was not represented on the IBC array.

Three of the genomic regions significantly associated with FVII in EAs have not previously been reported, including eight SNPs in or near the ZNF259/BUD31/APOA5 region on chromosome 1q23, one SNP upstream of MLXIPL on chromosome 7q11 and a non-synonymous SNP in HNF4A on chromosome 20q12. To further investigate the three newly associated FVII loci, we examined the results in 4289 participants from the Whitehall II study (Table 3). For the top SNP in the ZNF259/BUD31/APOA5 region on chromosome 1q23 (rs2266788) and the non-synonymous SNP in HNF4A (rs1800961), the magnitudes and directions of association were similar in the Whitehall II study, replicating the CARe results. Upon meta-analysis of the CARe and the Whitehall II results, both SNPs reached or closely approached the conventional level of ‘genome-wide significance’ (< 5 × 10^-8). The chromosome 7q11 SNP upstream of MLXIPL (rs7777102) was not available in the Whitehall II data set, and therefore could not be validated.

FVII association results in AAs

In AA participants, three distinct regions were associated with FVII (Table 4). Two of these regions were concordantly associated with FVII in EAs (F7–F10, PROCR), while one region appeared to be specific to AAs (two SNPs at 10p15 near PRKCOQ), although the direction of the association was similar to what was observed in EAs. Of the five remaining loci associated with FVII in EAs, GCKR, MS4A2 and HNF4A showed a similar direction of association in AAs (Table 5). In AAs, effect estimates were smaller and P-values were larger for the top SNP associations noted in EAs. There were substantial observed allele frequency...
differences between the two populations for several of the top SNPs that would result in differences in statistical power to detect an effect when present. Notably, the PRKCCQ SNP (rs602419) had a minor allele frequency of 0.02 in EAs, compared with 0.32 in AAs. In addition, observed linkage disequilibrium (LD) patterns differed between EAs and AAs in the region of F7 and F10 genes, as shown in Supplementary Material, Figures S1 and S2, respectively, with the extent of LD considerably lower among AAs.

To further investigate the F7, PROCR and newly associated PRKCCQ region in AA, we assessed the association of the top SNPs in 918 AA participants from the Women’s Health Initiative (WHI) (Table 6). The magnitude and direction of associations for the top F7 and PROCR SNPs were similar between WHI and CARe, replicating the CARe AA results. While the novel PRKCCQ SNP was not significantly associated with FVII levels in the WHI sample, the direction of association was the same.

Fine-mapping of FVII-associated loci in EAs and AAs

Regional LD association plots for array-wide significant loci in EAs and AAs are shown in Supplementary Material, Figures S3–12. Based on assessment of regional LD association patterns and results from the conditional regression analyses, there was little evidence supporting multiple independent association signals within the PROCR, PRKCCQ, HNF4A, MS4A2, APOA5, MLXIPL regions, and moderate evidence for the GCKR region. In contrast, there was evidence for more than one independent association at an array-wide significant level in the F7 gene region in both EAs and AAs (Supplementary Material, Table S1).

The regional plots for F7 (Supplementary Material, Figures S3 and S4) demonstrate multiple significant loci in the F7–F10 region for EA and AA. Step-wise regression analyses revealed five independent signals in EAs (rs561241, rs555212, rs3093265, rs3211727, rs493833) and three independent signals in AAs (rs561241, rs493833, rs3093230) (Table 7). All signals identified were in the F7 gene, except for one in the F10 gene in EAs. In EA, the minor allele of SNPs rs555212 was associated with higher FVII levels, whereas the minor alleles of SNPs rs561241, rs3093265, rs3211727 and rs493833 were associated with lower FVII levels. Together, these SNPs explained 18% of the variation of normalized FVII levels in EAs and 9% of the variation in AAs.

Observed LD patterns differed between EAs and AAs in the region of the F7 and F10 genes, as shown in Supplementary Material, Figures S1 and S2, with the extent of LD considerably lower among AAs. The lower extent of LD between SNPs in the F7, along with the results of conditional association analysis, could be used to fine-map the FVII association signals in this region. For example, the most significant SNP (rs561241) in EAs, located within the F7 promoter region, is in strong LD ($r^2 > 0.9$) with several other SNPs that span the entire F7 gene region, including rs1755685, rs6039, rs7981123 and rs493833. In AAs, the same SNP rs561241 showed LD ($r^2 > 0.6$) with only one other SNP, rs1755685, also in the promoter region. These findings suggest that the main signal associated with lower FVII levels is localized to the F7 promoter region. Similarly, the promoter SNP associated with higher FVII levels in EA (rs555212) is in strong LD with at least two other SNPs (rs3093229 and rs3093230) located in the promoter region ($r^2 > 0.98$). In AA, there is less LD between rs555212, rs3093229 and rs3093230 (pairwise $r^2$ ranging between 0.4 and 0.5). Of these three SNPs, rs3093230 is most strongly associated with higher FVII in AA.

Haplotype-based analysis confirmed the individual F7 SNP association results (Supplementary Material, Table S2). There were three prominent F7 haplotypes in EAs. The haplotype tagged by rs555212 was associated with higher FVII levels, while the haplotype tagged by rs561241 was associated with lower FVII levels. In AA, there was greater haplotype diversity, with five haplotypes having frequencies of at least 5%. There are two haplotypes tagged by the minor allele of rs555212; only the haplotype that additionally contains the rs3093230 minor allele is associated with higher FVII levels in AA.

Total variance explained by known and novel loci

When SNPs representing the five independent signals for the F7/F10 region, the two potentially independent signals from the GCKR region and the other loci identified in Tables 2 and 3 were all included simultaneously in a linear regression model for the ARIC study, they explained a total of 20.2% of the variance in age- and sex-adjusted normalized FVII levels.
levels in EAs, and 10.2% in AAs (adjusted $R^2 = 0.202$ and 0.102, respectively). Most of this variance is due to variation in the $F7$ gene, with the five independent regions in the $F7$ gene explaining 18% of the variance, and the remaining loci from Table 2 explaining the remaining 2% in ARIC EAs.

Comparison of genetic association results for FVII activity versus FVII antigen assays

For each genomic region, associations of the top SNPs with FVII antigen levels closely mirrored those with FVII activity levels. Effect estimates and standard errors were similar, validating previous studies that have meta-analyzed activity and antigen levels (Supplementary Material, Table S3).

DISCUSSION

By using a custom 50 K SNP genotyping array with dense, multi-population tag SNP coverage of >2,000 genes in CVD-related pathways including coagulation, we further characterized regions of the genome associated with FVII levels in EAs and AAs. Our findings can be summarized as follows: (i) the chromosome 13q34 $F7$–$F10$ gene region contained multiple variants associated with higher and lower FVII in both EAs and AAs. (ii) The lower extent of LD among AA enabled finer localization of the strongest $F7$ signals. (iii) We identified two novel genomic regions significantly associated with FVII in EAs, the ZNF239–$BUD13$–$APOA5$ gene region on chromosome 11q23 and chromosome 20q12 HNF4A gene (4). Of other candidate regions associated with FVII in EAs, we found evidence for replication of the chromosome 20 $PROCR$ SNP rs867186 (Ser219Gly) and GCKR associations with higher FVII levels in AAs; on the other hand, FVII association with polymorphisms within the MS4A2, $APOA5$ and HNF4A gene regions were not evident in AAs.

Polymorphisms in the $F7$ gene could be grouped into three major haplotypes in European descent populations (5). Compared with the most common or “wild-type” haplotype, the haplotype tagged by the minor allele of rs555212 (and rs3093230) was associated with higher FVII levels while the haplotype tagged by the minor allele at rs561241 was associated with lower levels. These results confirm previous findings from smaller European population studies that SNPs tagging these haplotypes constitute the major determinants of FVII activity in EAs (5,15). Because of the large sample size of CARe and the dense SNP coverage of the IBC genotyping array, we were able to identify at least three additional variants within the $F7$–$F10$ genes independently associated FVII levels. Two of these are located within intron 5 of the $F7$ gene (rs493833 and rs3093265) and one within intron 1 of the $F10$ gene (rs3093268). The strong LD among the promoter variants in Europeans has previously hampered detailed characterization of the functional variants responsible for reduced FVII levels. Functional analysis of the $F7$ promoter suggested that the –323ins10 and –122C (rs561241) variants are strongly associated with decreased promoter activity. It has remained uncertain whether the –402A variant (rs510317) or other upstream variants are responsible for the increased F7
promoter activity (16). In our analysis, AA showed greater F7 haplotype diversity, which suggested further localization of the higher-expression F7 promoter region to rs3093230 (or an untyped variant in LD).

Characterization of common cis-acting F7 gene variants may have relevance for athero-thrombotic disease risk. The F7 polymorphisms associated with higher FVII levels have been associated with increased risk of stroke and myocardial infarction (MI) in young women (5,17,18). The relationship of haplotype C FVII-lowering polymorphisms to risk of clinical CVD has been equivocal, with published meta-analyses of MI (10) and stroke (19) showing no significant association. Since our results suggest independent associations of several additional F7 polymorphisms on FVII level, evaluation of these variants in the context of clinical CVD risk, or extent of subclinical atherosclerosis in young adults (20) both in EA and AA populations, may be warranted.

Polymorphisms in LD with the PROCR Ser219Gly variant (rs867186) have been strongly associated with higher soluble endothelial protein C receptor (EPCR) levels, as well as with higher levels of clotting FVII, FVIIa, protein C antigen and several downstream markers of activated coagulation in the extrinsic pathway (21–22) in whites. Binding of FVII/FVIIa to soluble and/or membrane-associated EPCR and subsequent endocytosis of the receptor-ligand complexes may be important for regulating the circulating concentrations of FVII and protein C (23). These findings have potential implications for the role of the PROCR Ser219Gly dimorphism in risk of clinical thrombotic disease (24,25).

FVII levels are correlated with other CVD risk factors, such as lipid levels, body mass index and insulin resistance (6,26). The causal nature and mechanisms of these associations remain to be elucidated. There is a fairly well-described connection between high-fat diet, triglycerides and FVII activity. FVII and other vitamin K-dependent coagulation proteins bind to triglyceride-rich lipid particles (27). Moreover, expression of the F7 gene can be modulated by glucose and insulin levels due to a specific promoter element (28). The association of FVII levels with polymorphisms of the GCKR, APOA5, MLXIPL and HNF4A genes suggests an additional level of shared genetic regulation between FVII synthesis, glucose metabolism and low-grade inflammation. The APOA5 SNP rs6589566 that was discovered in our study was also associated with low density lipoprotein levels in a recent genome-wide association scan (29). Other SNPs in LD with rs6589566 (e.g. rs964184, rs1558861, rs4938303, rs12280753) have been associated with high density lipoprotein (HDL-C) levels and triglyceride levels (30–34).

While the same SNPs were not identified in this study, variants in or near PRKCO have been found to be associated at a genome-wide significance level with type I diabetes (35–36) and rheumatoid arthritis (37–38). One non-syonymous SNP in MLXIPL was also significantly associated with plasma triglycerides in another recent genome-wide scan (33), although LD information was not available between our SNP (rs7777102) and theirs (rs3812316).

The GCKR gene product inhibits glucokinase in liver and pancreatic islet cells and is considered a susceptibility gene candidate for a form of maturity-onset diabetes of the young. Polymorphisms of GCKR have been associated with a number of metabolic and CVD traits, including HDL, triglycerides, glucose, insulin resistance, renal function, C-reactive protein and protein C levels. GWASs have shown that the GCKR SNP identified in our study, rs1260326, is also associated with triglyceride levels (32), chronic kidney disease (39) and 2 h post-load glucose levels (40).

This study has several notable strengths, including a large sample size for both EA and AA participants. The IBC custom chip enables a more comprehensive study of lower frequency variants than the CARe EA custom chip. However, the replication of novel FVII-associated loci in EAs in the Whitehall II Study might have been compromised due to the small sample size and difference in minor allele frequency or effect LD patterns, lower statistical power in AAs due to a smaller sample size for both EA and AA participants. The IBC custom chip enables a more comprehensive study of lower frequency variants. The chromosome 20q12–13 region contains multiple variants independently associated with higher and lower FVII in both EAs and AAs. These findings between populations, it is important to note differences in findings for individual loci may be due to different LD patterns, lower statistical power in AAs due to a smaller sample size, difference in minor allele frequency or effect modification by genetic or environmental factors.

In summary, we have confirmed and refined known loci, and identified several new genomic loci associated with FVII levels, including APOA5 and HNF4A in EAs. The chromosome 13q14 F7–F10 gene region contains multiple variants independently associated with higher and lower FVII in both EAs and AAs. Of other candidate regions associated with FVII in EA, we found evidence for replication of the chromosome 20 PROCRA gene Ser219Gly and GCKR associations with higher FVII levels in AAs; on the other hand, FVII association with polymorphisms within the MS4A2, APOA5 and HNF4A gene regions were not evident in AAs. Larger samples and investigation of lower frequency variants may be required to identify additional FVII-associated loci in AAs.
MATERIALS AND METHODS

CARe discovery population

The purpose of the CARe (Candidate Gene Association Resource) consortium is to bring together deeply phenotyped prospective cohort studies to increase power for genetic association scans of CVD and other disorders (41). Cohorts included in this analysis of FVII levels are ARIC, CARDIA, the CHS and the FHS. Further details of the participating CARe studies are reported in the Supplementary Materials.

To be included in the current analysis, the participant must have provided informed consent for genetic testing and had FVII antigen levels or activity levels measured. Exclusion criteria included use of the anticoagulant warfarin. In ARIC, CHS and CARDIA (year 7 exam), FVII activity levels were measured in citrated plasma by assaying clotting time using FVII-deficient plasma (see (14), for ARIC and CHS; see (42) for CARDIA). In FHS and CARDIA (year 20 exam), FVII antigen levels were measured in ethylenediaminetetraacetic acid plasma by Enzyme-linked immunosorbent assay (Diagnostica Stago) (14,20). As CARe study participants were examined at multiple time points, covariate values that were included in the analyses were selected from the visit closest to the time of FVII measurement. After excluding study participants based on low genotyping quality control (see further information below), there were a total of 16 352 EA and 3948 AA CARe study participants available for the analysis.

Genotyping

Both AAs and EAs in CARe were genotyped using the 50 000 SNP, cardiovascular gene-centric IBC SNP array, described by Keating et al. (43). The IBC SNP array includes 49 320 SNPs selected across ≏2100 candidate loci for CVD. Of the five genomic regions previously reported to be associated with FVII in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium (GCKR, ADH4, MS4A6A, F7 and PROCR), all except ADH4 are represented on the IBC array. Genotyping was done at the Broad Institute (Cambridge, MA, USA). Criteria for DNA sample exclusion based on genotype data included sex mismatch, discordance among duplicate samples or sample call rate <95%. For each set of duplicates or monozygotic twins, data from the sample with the highest genotyping call rate were retained. SNPs were excluded when monomorphic, the call rate was <95% or Hardy-Weinberg equilibrium (HWE) was $P < 10^{-5}$ in EAs. Given the genetic admixture in AAs, there was no HWE filter used for AAs. After these exclusions were applied, data remained on 47 539 working SNPs.

Additional SNPs were imputed from HapMap Phases 1 and 2 using MACH (44–45), as described in detail under Supplemental Materials. The total number of SNPs was ≏250 000 after adding the imputed SNPs. We reported the genotyped results from the raw genotype data when a SNP was both genotyped and imputed.
Table 5. Comparison of top SNPs from Tables 2 and 3 in EAs and AAs

<table>
<thead>
<tr>
<th>Region (Chr)</th>
<th>Index SNP</th>
<th>N (EA)</th>
<th>MAF*(EA)</th>
<th>Beta (SE) (EA)</th>
<th>P-value (EA)</th>
<th>N (AA)</th>
<th>MAF (AA)</th>
<th>Beta (SE) (AA)</th>
<th>P-value (AA)</th>
<th>Genes in region</th>
</tr>
</thead>
<tbody>
<tr>
<td>13q34</td>
<td>rs561241</td>
<td>16324</td>
<td>0.12</td>
<td>0.88 (0.01)</td>
<td>6.78 × 10⁻⁴⁸⁵</td>
<td>3855</td>
<td>0.13</td>
<td>0.57 (0.03)</td>
<td>1.02 × 10⁻⁶⁴</td>
<td>F7, F10</td>
</tr>
<tr>
<td>20q11</td>
<td>rs11167260</td>
<td>16324</td>
<td>0.10</td>
<td>-0.26 (0.02)</td>
<td>2.16 × 10⁻³⁸</td>
<td>3898</td>
<td>0.08</td>
<td>0.22 (0.05)</td>
<td>3.07 × 10⁻⁷</td>
<td>PROC PRCR</td>
</tr>
<tr>
<td>2p3</td>
<td>rs1260326</td>
<td>13691</td>
<td>0.42</td>
<td>0.09 (0.01)</td>
<td>1.55 × 10⁻¹¹</td>
<td>3880</td>
<td>0.15</td>
<td>0.04 (0.05)</td>
<td>0.460</td>
<td>GCKR</td>
</tr>
<tr>
<td>11q12</td>
<td>rs17529983</td>
<td>16324</td>
<td>0.42</td>
<td>-0.12 (0.02)</td>
<td>2.82 × 10⁻⁷</td>
<td>3898</td>
<td>0.09</td>
<td>0.05 (0.10)</td>
<td>0.558</td>
<td>MSA42 APOE</td>
</tr>
<tr>
<td>11q23</td>
<td>rs6589566</td>
<td>16324</td>
<td>0.07</td>
<td>0.06 (0.02)</td>
<td>1.24 × 10⁻⁶</td>
<td>3898</td>
<td>0.01</td>
<td>n/a</td>
<td>n/a</td>
<td>ZNF259 BUD13</td>
</tr>
<tr>
<td>7q</td>
<td>rs7777102</td>
<td>16324</td>
<td>0.17</td>
<td>-0.17 (0.03)</td>
<td>1.62 × 10⁻⁶</td>
<td>n/a</td>
<td>n/a</td>
<td>-0.13(0.16)</td>
<td>0.390</td>
<td>MLX1PL VPS37D</td>
</tr>
<tr>
<td>20q12–13</td>
<td>rs1800961</td>
<td>13702</td>
<td>0.03</td>
<td>-0.26 (0.02)</td>
<td>2.16 × 10⁻³⁸</td>
<td>3898</td>
<td>0.006</td>
<td>0.22 (0.05)</td>
<td>3.07 × 10⁻⁷</td>
<td>HNF4A</td>
</tr>
<tr>
<td>10p15</td>
<td>rs602419</td>
<td>13706</td>
<td>0.02</td>
<td>-0.04 (0.05)</td>
<td>0.387</td>
<td>3898</td>
<td>0.32</td>
<td>-0.13 (0.03)</td>
<td>3.60 × 10⁻⁷</td>
<td>PRKCIq</td>
</tr>
<tr>
<td>20q11</td>
<td>rs7265317</td>
<td>15826</td>
<td>0.10</td>
<td>-0.26 (0.02)</td>
<td>3.94 × 10⁻³⁷</td>
<td>3877</td>
<td>0.08</td>
<td>-0.21 (0.04)</td>
<td>7.29 × 10⁻⁷</td>
<td>PROC ROR1</td>
</tr>
</tbody>
</table>

*Minor allele frequency.

Table 6. Replication of array-wide significant regions identified in AAs in WHI-SHARE AAs (N = 918)

<table>
<thead>
<tr>
<th>Region</th>
<th>Gene (index SNP)</th>
<th>Beta (SE) for index SNP in CARE AAs</th>
<th>WHI beta (SE) for index SNP</th>
<th>WHI P-value for index SNP</th>
<th>Number of SNPs in gene region in WHI</th>
<th>Threshold P-value using Bonferroni correction (SNP)</th>
<th>Lowest WHI P-value in gene region</th>
<th>Evidence for replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>13q34</td>
<td>F7, rs561241</td>
<td>0.57 (0.03)</td>
<td>0.76 (0.08)</td>
<td>&lt;1 × 10⁻¹⁰</td>
<td>70</td>
<td>7.1 × 10⁻⁴</td>
<td>&lt;1 × 10⁻¹⁰</td>
<td>Definite</td>
</tr>
<tr>
<td>10p15</td>
<td>PRKCIq, rs602419</td>
<td>-0.13 (0.03)</td>
<td>-0.04 (0.05)</td>
<td>0.40</td>
<td>143</td>
<td>3.5 × 10⁻⁴</td>
<td>0.0095, rs944715</td>
<td>Weak</td>
</tr>
<tr>
<td>20q11</td>
<td>PROC ROR1, rs67186</td>
<td>-0.20 (0.04)</td>
<td>-0.36 (0.08)</td>
<td>9.7 × 10⁻⁶</td>
<td>4</td>
<td>1.3 × 10⁻²</td>
<td>9.7 × 10⁻⁶, rs67186</td>
<td>Definite</td>
</tr>
</tbody>
</table>

*Gene region defined by the start of the first gene (in a multiple gene region) to the last position of the last gene as defined by the USC human genome browser, build hg18.

Data analysis

Initial IBC-array-wide association analyses were conducted at the Broad Institute, and subsequent analyses were conducted at UNC-Chapel Hill. Family structure (where applicable) was taken into account during the association analysis, using a linear mixed-effects model (46). To generate a normally distributed outcome variable for use in association analyses, we stratified the data by sex, race and cohort, and modeled FVII levels as a function of age and study site (where applicable). We inverse normally transformed the residuals, and recomposed the data from race and sex strata within each cohort before conducting the association analyses. For both genotyped and imputed SNPs, we assumed an additive genetic model. Consistent with this assumption, we used dosage values (i.e. a value between 0.0 and 2.0 calculated as the expected number of copies of the reference allele based on the posterior probability of each of the three possible genotypes) in the regression model implemented in PLINK (for cohorts with unrelated individuals) or the linear mixed-effects model implemented in genome-wide association study in families (for cohorts with related individuals) (46,47). To control for potential population stratification, we included the top 10 principal components, calculated in Eigenstrat, as covariates in the association analysis (48). Principal components were calculated using the cleaned CARe IBC genotype data and the HapMap populations (CEU, YRI, CHB + JPT) as reference (seed) populations.

After obtaining cohort- and race-specific results, we used the METAL software to meta-analyze the data within race (49). We used a fixed-effect Z-score approach when combining results from studies that analyzed FVII activity and antigen levels. Meta-analyses were based on P-values and weighted by sample sizes. We also calculated summary effect estimates and standard errors by meta-analyzing the three studies which measured FVII activity levels. As a supplementary analysis, we compared effect estimates based on FVII activity versus FVII antigen levels.

For the IBC array, the effective number of independent tests was calculated to be 26 482 for AA and 20 544 for EA, based on LD between markers on the array (50). To maintain an overall type 1 error rate of 5%, a statistical threshold of $\alpha = 2 \times 10^{-6}$ (0.05/25000) was thus used to declare array-wide (experiment-wide) significance.
Haplotype analysis was conducted using the R-based software package HAPLOSTATS for the known haplotypes in the F7 gene represented by four SNPs (rs561241, rs2274030, rs3093265 and rs6046) that tag common patterns of LD in Caucasians (5), and an additional SNP that defined another haplotype in AAs in our data (rs3093230). The association of each haplotype with FVII levels, relative to the most common referent haplotype, was estimated using generalized linear models (haplo.glm). The EA and AA samples had the common referent haplotype, was estimated using generalized linear models (haplo.glm). The EA and AA samples had the same referent haplotype. Haplotype analyses were restricted to the ARIC study (the largest sample) in EAs and AAs.

### Conditional analyses

For genomic regions which initially revealed more than one significant SNP in the association scan, we conducted conditional regression analyses to estimate the number of independently associated SNPs. Conditional analyses were performed iteratively in a forward stepwise manner, with a meta-analysis done at each stage. We began by conditioning on the most significant SNP in the region (or a closely linked proxy) by including it as a covariate in the linear regression model, in each population. We determined whether any other SNPs in that region remained significant (defined as $P < 2 \times 10^{-6}$) after meta-analysis. If some results were still significant, we took the most significant of those SNPs, and in the next round, conditioned on the top SNPs from both rounds. This continued until there were no longer any additional significant SNPs. To estimate the cumulative proportion of variance in the sex- and age-adjusted, normalized FVII levels explained by the associated SNPs, we obtained the sum of the partial $R^2$ for each of the associated SNPs when all were simultaneously included in the regression model, adjusting for 10 PCs. All conditional analyses were conducted in SAV 9.2 (Cary, NC, USA).

### EA replication cohort: Whitehall II Study

Replication analyses of novel FVII association findings for EAs were performed in the Whitehall II study. Details of this cohort study are provided in ref. (51). Briefly, 10,308 UK participants (70% men) were recruited between 1985 and 1988 of whom >6156 participants had blood samples for DNA collected in 2002–2004. After filtering based on call rates and removal of ancestry outliers based on principal component analysis, custom genotyping for the IBC array (48,032 SNPs) was available on 4289 individual of white ancestry. Fasting blood samples were measured for FVII activity in 1991–1993 by the method of Brozovic et al. (52). Data analysis of inverse normalized FVII activity levels was performed similarly to CARe using PLINK.

### AA replication cohort: WHI

For replication of association findings in AAs, we used genotyped and imputed SNPs available through a GWAS performed in the AA participants from the WHI. Of the 8421 AA women in the WHI-SHARE project who were genotyped using the Affymetrix 6.0 chip, 984 of these women also had FVII antigen and activity levels measured. Following the primary data analysis plan, FVII association analyses were performed, using PLINK for genotyped SNPs, and ProbAbel for imputed SNPs (53).

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

**Conflict of Interest statement.** None declared.

### FUNDING

This work was supported by the National Heart, Lung and Blood Institute of the National Institutes of Health (#5T32HL007055-34); the Heart Lung and Blood Institute of the National Institutes of Health (HL36310, HL71862-06); the National Institutes of Health (HHSN268200625226C, N01HC65226-6-0-0); the National Institute of Aging of the National Institutes of Health (AG032136, AG034454,


