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

Kira C. Taylor[1,*], Leslie A. Lange[2], Delilah Zabaneh[3], Ethan Lange[2], Brendan J. Keating[5], Weihong Tang[6], Nicholas L. Smith[7], Joseph A. Delaney[8], Meena Kumari[9], Aroon Hingorani[10], Kari E. North[1], Mika Kivimaki[9], Russell P. Tracy[11], Christopher J. O'Donnell[12], Aaron R. Folsom[6], David Green[13], Steve E. Humphries[4] and Alexander P. Reiner[7]

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

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 Whitehall II study. We confirmed four previously reported FVII-associated loci (*GCKR*, *MS4A6A*, *F7* and *PROCR*) in CARe EA samples. In AAs, the *F7* and *PROCR* 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 9199669800; Email: kirat@email.unc.edu

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 50 K 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 CHS and lower 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 3898 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, *MS4A2-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/BUD13/APOA5* region on chromosome 11q23, 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/BUD13/APOA5* region on chromosome 11q23 (rs2266788) and the *HNF4A* non-synonymous SNP (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' ($< 5 \times 10^{-8}$). 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 *PRKCQ*), 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

Table 1. Characteristics of the 16352 EA and 3948 AA study participants[a]

| | ARIC | | CARDIA | | CHS | | FHS[b] |
	AA	EA	AA	EA	AA	EA	EA
N	2768	9123	463	753	717	3873	2603
Female, %	63	54	60	53	63	56	55
Current smoker, %	29	24	33	23	16	11	18[c]
Diabetes, %	13	5	2	3	8	8	7
Age, years	53 ± 6	54 ± 6	30 ± 4	31 ± 3	73 ± 6	73 ± 6	54 ± 10
Body mass index, kg/m^2	29.6 ± 6.1	26.9 ± 4.8	28.2 ± 6.8	25.2 ± 5.0	28.5 ± 5.6	26.4 ± 4.5	27.4 ± 4.9
Median Factor VII activity (%, IQR)	115 (97, 133)	116 (100, 135)	85 (70, 99)	81 (68, 96)	110 (95, 127)	122 (106, 142)	99 (89, 110)[d]

[a]Mean $\pm$ SD for continuous variables, % for categorical variables. Only includes individuals with both FVII and IBC data.
[b]Offspring cohort at visit 5.
[c]Smoked in last year (other cohorts have a variable 'current smoker'. This was not available for visit 5 in FHS).
[d]FVII antigen activity.

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 *PRKCQ* SNP (rs602419) had a minor allele frequency of only 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 *PRKCQ* 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 *PRKCQ* 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*, *PRKCQ*, *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 *GKCR* 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

Table 2. Loci associated with FVII in EAs

Region (Chr)	No. of SNPs in region, $P < 1.0 \times 10^{-6}$	Most significant[a]	Position[b]	Non-coded allele	Coded (effect) allele	N	I/G[c]	MAF	Beta(SE)[d]	P-value[e]	Closest genes; function of index SNP	Directions of effect[f]
13q34[g]	40	rs561241	112808035	C	T	16324	I/G[h]	0.12	0.88 (0.01)	6.78×10^{-485}	*F7, F10*; promoter region	+ + + +
20q11[g]	113	rs867186	33228215	G	A	16263	G	0.09	−0.26 (0.02)	1.84×10^{-36}	*PROCR*; non-synonymous	− − − −
2p23[g]	94	rs1260326	27584444	C	T	16309	G	0.41	0.13 (0.01)	1.70×10^{-28}	*GCKR*; non-synonymous	+ + + +
11q12[g]	105	rs17529983	59683615	G	A	16324	I	0.42	0.09 (0.01)	1.55×10^{-11}	*MS4A2; MS4A64*; intergenic	− − − −
11q23	13	rs6589566	116157633	G	A	16324	I	0.07	−0.12 (0.02)	2.82×10^{-7}	*ZNF259; BUD13; APOA5*; intronic	− + − −
7q	1	rs7777102	72695953	G	A	16324	I	0.17	0.08 (0.02)	1.24×10^{-6}	*MLXIPL; VPS37D*; intergenic	+ − + +
20q12–13	1	rs1800961	42475778	C	T	13702	G	0.03	−0.17 (0.03)	1.62×10^{-6}	*HNF4A*; non-synonymous	− − − −

[a]Most significant SNP or a proxy in high LD ($R^2 > 0.95$, D' = 1.0).
[b]Positions given throughout the manuscript are from NCBI Build 36.
[c]Imputed (I) or genotyped (G).
[d]Effect estimates were meta-analyzed for ARIC, CARDIA and CHS (FHS excluded because it measured antigen levels, not activity).
[e]Sample sized-based meta-analysis P-value (for ARIC, CARDIA, CHS and FHS).
[f]Order shown is ARIC, CARDIA, CHS, FHS.
[g]Previously associated with FVII levels in the literature (including Smith et al., 2010, which included ARIC, CHS and FHS).
[h]Genotyped in ARIC and CARDIA but imputed in CHS and FHS.

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 *ZNF259–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 in 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

Table 3. Replication of novel FVII-associated loci in EAs in the Whitehall II Study

Region (Chr)	Genes in region	SNP	Position	Coding allele	CARe EA beta (SE)	CARe EA P-value	WhII, beta (SE)[a]	WhII P-value	Meta-analysis P-value[b]	Evidence for replication
11q23	*ZNF259, BUD13, APOA5*	rs2266788[c] (rs6589566)	116165896	A	−0.12 (0.02)	3.6×10^{-7}	−2.6 (0.9)	0.0039	4.0×10^{-9}	Definite
7q	*MLXIPL, VPS37D*	rs7777102	72695953	A	0.08 (0.02)	1.24×10^{-6}	n/a[d]	n/a	n/a	n/a
20q12–13	*HNF4A*	rs1800961	42475778	T	−0.17 (0.03)	1.5×10^{-7}	−3.3 (1.4)	0.017	8.8×10^{-8}	Definite

WhII, Whitehall II Study.
[a]The Whitehall II study utilized a different transformation for FVII levels. The direction of effect and the *P*-values were used to determine whether an association replicated.
[b]Meta-analysis *P*-values were calculated using *P*-values and sample sizes, taking into account the direction of effect.
[c]Proxy for rs6589566; $R^2 = 0.92$, $D' = 1.0$ in HapMap CEU.
[d]Not available in the Whitehall II study; no proxy was available.

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 *PRKCQ* 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-synonymous SNP in *MXLIPL* 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 in comparison to GWAS. When comparing 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 13q34 *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 *PROCR* 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.

Table 4. Loci associated with FVII in AAs

Region (Chr)	No. significant	Most significant	Position	Non-coding allele	Coding (effect) Allele	N	I/G	MAF	Beta (SE)	P-value	Closest genes; function/location of most significant SNP	Directions of effect[a]
13q34	24	rs561241	112808035	C	T	3855	I/G[b]	0.13	0.57 (0.03)	1.02×10^{-64}	*F7, F10*; promoter region	+++
10p15[c]	2	rs602419	6513222	G	T	3898	I	0.32	−0.13 (0.03)	3.58×10^{-7}	*PRKCQ*; intronic	−+−
20q11	1	rs7265317[d]	33232184	C	T	3877	G	0.08	−0.21 (0.04)	7.29×10^{-7}	*PROCR*; downstream	−−−

[a]Order: ARIC, CARDIA, CHS.
[b]Genotyped in ARIC and CARDIA; imputed in CHS.
[c]Not previously associated with FVII levels in the literature.
[d]In LD with rs867186 in EAs ($R^2 = 0.99$, $D' = 1.0$ in ARIC EAs) and AAs ($R^2 = 0.52$, $D' = 0.78$ in ARIC 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

Region (Chr)	Index SNP	N (EA)	MAF[a] (EA)	Beta (SE) (EA)	P-value (EA)	N (AA)	MAF (AA)	Beta (SE) (AA)	P-value (AA)	Genes in region
SNP identified in both EAs and AAs										
13q34	rs561241	16324	0.12	0.88 (0.01)	6.78×10^{-485}	3855	0.13	0.57 (0.03)	1.02×10^{-64}	*F7, F10*
SNPs identified in EAs										
20q11	rs11167260[c] (rs867186)	16324	0.10	-0.26 (0.02)	2.16×10^{-38}	3898	0.08	0.22 (0.05)	3.07×10^{-6}	*PROCR*
2p23	rs1260326	13691	0.42	0.09 (0.01)	1.55×10^{-11}	3880	0.15	0.04 (0.05)	0.460	*GCKR*
11q12	rs17529983	16324	0.42	-0.12 (0.02)	2.82×10^{-7}	3898	0.09	0.05 (0.10)	0.558	*MS4A2*
11q23	rs6589566	16324	0.07	0.08 (0.02)	1.24×10^{-6}	3898	0.01	n/a	n/a	*ZNF259, BUD13, APOA5*
7q	rs7777102[d]	16324	0.17	-0.17 (0.03)	1.62×10^{-6}	n/a	n/a	$-0.13(0.16)$	0.390	*MLXIPL, VPS37D*
20q12–13	rs1800961	13702	0.03	-0.26 (0.02)	2.16×10^{-38}	3898	0.006	0.22 (0.05)	3.07×10^{-6}	*HNF4A*
SNPs identified in AAs										
10p15	rs602419	13706	0.02	-0.04 (0.05)	0.387	3898	0.32	-0.13 (0.03)	3.60×10^{-7}	*PRKCQ*
20q11	rs7265317	15826	0.10	-0.26 (0.02)	3.94×10^{-37}	3877	0.08	-0.21 (0.04)	7.29×10^{-7}	*PROCR*

[a]Minor allele frequency.
[b]Order is ARIC, CARDIA, CHS, FHS for EAs, and ARIC, CARDIA and CHS for AAs.
[c]Proxy for rs867186 ($R^2 = 1.0$, $D' = 1.0$ in ARIC EAs) used here because rs867186 was only available in CARDIA and CHS for AAs.
[d]No information for rs7777102 in AAs, and no proxies were available.

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

Region	Gene (index SNP)	Beta (SE) for index SNP in CARe AAs	WHI beta (SE) for index SNP	WHI P-value for index SNP	Number of SNPs in gene region[a]	Threshold P-value using Bonferroni correction (SNP)	Lowest WHI P-value in gene region[a]	Evidence for replication
13q34	*F7*, rs561241	0.57 (0.03)	0.76 (0.08)	$<1 \times 10^{-10}$	70	7.1×10^{-4}	$<1 \times 10^{-10}$	Definite
10p15	*PRKCQ*, rs602419	-0.13 (0.03)	-0.04 (0.05)	0.40	143	3.5×10^{-4}	0.0095, rs944715	Weak[b]
20q11	*PROCR*, rs867186[c]	-0.20 (0.04)	-0.36 (0.08)	9.7×10^{-6}	4	1.3×10^{-2}	9.7×10^{-6}, rs867186	Definite

[a]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.
[b]Criterion for 'weak': same direction of effect for index SNP.
[c]rs7265317 was not available in the WHI data set.

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 recombined 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.

Table 7. Fine mapping of the *F7/F10* region in EAs and AAs

	Cumulative % variation explained by genotype(s) (adj. R^2)				Iterative meta-analysis results (*P*-value for SNP adjusted for previous SNPs in table)	Minor allele	MAF	Beta (SE) for minor allele in model including listed SNPs in ARIC	LD with rs561241 in ARIC (R^2)	Gene
	ARIC	CARDIA	CHS	FHS						
EAs										
Only rs561241	14.75	14.53	12.94	8.37	$<1 \times 10^{-500}$	G	0.13	-0.42 (0.08)	–	*F7*
Plus rs555212[a]	16.78	16.42	15.92	8.99	8.45×10^{-86}	A	0.22	0.34 (0.03)	0.04	*F7*
Plus rs3093265	17.68	16.32	16.34	9.40	1.47×10^{-28}	C	0.04	-0.33 (0.03)	0.01	*F7*
Plus rs3211727	17.97	16.67	16.70	9.39	6.39×10^{-13}	A	0.22	-0.17 (0.03)	0.04	*F10*
Plus rs493833	18.22	17.41	17.39	9.37	3.72×10^{-07}	G	0.14	-0.40 (0.08)	0.90	*F7*
AAs										
Only rs561241	7.65	7.03	6.64		2.16×10^{-66}	G	0.12	-0.47 (0.05)	–	*F7*
Plus rs493833	8.16	7.92	7.87		1.31×10^{-08}	G	0.32	-0.12 (0.03)	0.16	*F7*
Plus rs3093230[a]	9.10	NA	9.80		2.37×10^{-09}[b]	A	0.07	0.26 (0.06)	0.01	*F7*

[a]rs555212 and rs3093230 are in LD with one another in EAs ($R^2 = 0.98$, $D' = 0.99$ in ARIC EAs).
[b]Only includes data from ARIC and CHS.

Haplotype analysis was conducted using the R-based software package HAPLO.STATS 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 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 SAS v. 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

REFERENCES

1. Lippi, G., Favaloro, E.J., Franchini, M. and Guidi, G.C. (2009) Milestones and perspectives in coagulation and hemostasis. *Semin. Thromb. Hemost.*, **35**, 9–22.
2. Meade, T.W., Mellows, S., Brozovic, M., Miller, G.J., Chakrabarti, R.R., North, W.R., Haines, A.P., Stirling, Y., Imeson, J.D. and Thompson, S.G. (1986) Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. *Lancet*, **2**, 533–537.
3. Wu, K.K., Folsom, A.R., Heiss, G., Davis, C.E., Conlan, M.G. and Barnes, R. (1992) Association of coagulation factors and inhibitors with carotid artery atherosclerosis. Early results of the Atherosclerosis Risk in Communities (ARIC) Study. *Ann. Epidemiol.*, **2**, 471–480.
4. Junker, R., Heinrich, J., Schulte, H., van de Loo, J. and Assmann, G. (1997) Coagulation factor VII and the risk of coronary heart disease in healthy men. *Arterioscler. Thromb. Vasc. Biol.*, **17**, 1539–1544.
5. Reiner, A.P., Carlson, C.S., Rieder, M.J., Siscovick, D.S., Liu, K., Chandler, W.L., Green, D., Schwartz, S.M. and Nickerson, D.A. (2007) Coagulation factor VII gene haplotypes, obesity-related traits, and cardiovascular risk in young women. *J. Thromb. Haemost.*, **5**, 42–49.
6. Karatela, R.A. and Sainani, G.S. (2010) Interrelationships of Factor VII activity and plasma leptin with insulin resistance in coronary heart disease. *Atherosclerosis*, **209**, 235–240.
7. Carew, J.A., Basso, F., Miller, G.J., Hawe, E., Jackson, A.A., Humphries, S.E. and Bauer, K.A. (2003) A functional haplotype in the 5' flanking region of the factor VII gene is associated with an increased risk of coronary heart disease. *J. Thromb. Haemost.*, **1**, 2179–2185.
8. Lindman, A.S., Pedersen, J.I., Arnesen, H., Hjerkinn, E.M. and Seljeflot, I. (2005) Association between the factor VII haplotype containing the -402A allele and myocardial infarction in a population of elderly men at high risk for coronary heart disease. *Thromb. Haemost.*, **94**, 226–227.
9. Bozzini, C., Girelli, D., Bernardi, F., Ferraresi, P., Olivieri, O., Pinotti, M., Martinelli, N., Manzato, F., Friso, S., Villa, G. *et al.* (2004) Influence of polymorphisms in the factor VII gene promoter on activated factor VII levels and on the risk of myocardial infarction in advanced coronary atherosclerosis. *Thromb. Haemost.*, **92**, 541–549.
10. Ye, Z., Liu, E.H., Higgins, J.P., Keavney, B.D., Lowe, G.D., Collins, R. and Danesh, J. (2006) Seven haemostatic gene polymorphisms in coronary disease: meta-analysis of 66,155 cases and 91,307 controls. *Lancet*, **367**, 651–658.
11. Souto, J.C., Almasy, L., Muniz-Diaz, E., Soria, J. M., Borrell, M., Bayen, L., Mateo, J., Madoz, P., Stone, W., Blangero, J. *et al.* (2000) 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.*, **20**, 2024–2028.
12. de Lange, M., Snieder, H., Ariens, R.A., Spector, T.D. and Grant, P.J. (2001) The genetics of haemostasis: a twin study. *Lancet*, **357**, 101–105.
13. Soria, J.M., Almasy, L., Souto, J.C., Sabater-Lleal, M., Fontcuberta, J. and Blangero, J. (2005) The F7 gene and clotting factor VII levels: dissection of a human quantitative trait locus. *Hum. Biol.*, **77**, 561–575.
14. Smith, N.L., Chen, M.H., Dehghan, A., Strachan, D.P., Basu, S., Soranzo, N., Hayward, C., Rudan, I., Sabater-Lleal, M., Bis, J.C. *et al.* (2010) Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation*, **121**, 1382–1392.
15. Ken-Dror, G., Drenos, F., Humphries, S.E., Talmud, P.J., Hingorani, A.D., Kivimaki, M., Kumari, M., Bauer, K.A., Morrissey, J.H. and Ireland, H.A. (2010) Haplotype and genotype effects of the F7 gene on circulating factor VII, coagulation activation markers and incident coronary heart disease in UK men. *J. Thromb. Haemost.*, **8**, 2394–2403.
16. Sabater-Lleal, M., Chillon, M., Howard, T.E., Gil, E., Almasy, L., Blangero, J., Fontcuberta, J. and Soria, J.M. (2007) Functional analysis of the genetic variability in the F7 gene promoter. *Atherosclerosis*, **195**, 262–268.
17. Funk, M., Endler, G., Lalouschek, W., Hsieh, K., Schillinger, M., Lang, W. and Mannhalter, C. (2006) Factor VII gene haplotypes and risk of ischemic stroke. *Clin. Chem.*, **52**, 1190–1192.
18. Zakai, N., Lange, L., Longstreth, W.T. Jr, O'Meara, E.S., Kelley, J.L., Fornage, M., Nickerson, D., Cushman, M. and Reiner, A.P. (2011) Association of coagulation-related and inflammation-related genes and factor VIIc levels with stroke: the Cardiovascular Health Study. *J. Thromb. Haemost.*, **9**, 267–274.
19. Maguire, J.M., Thakkinstian, A., Sturm, J., Levi, C., Lincz, L., Parsons, M., Whyte, S. and Attia, J. (2008) Polymorphisms in platelet glycoprotein 1balpha and factor VII and risk of ischemic stroke: a meta-analysis. *Stroke*, **39**, 1710–1716.
20. Green, D., Foiles, N., Chan, C., Kang, J., Schreiner, P.J. and Liu, K. (2010) An association between clotting factor VII and carotid intima-media thickness: the CARDIA study. *Stroke*, **41**, 1417–1422.
21. Ireland, H.A., Cooper, J.A., Drenos, F., Acharya, J., Mitchell, J.P., Bauer, K.A., Morrissey, J.H., Esnouf, M.P. and Humphries, S.E. (2009) FVII, FVIIa, and downstream markers of extrinsic pathway activation differ by EPCR Ser219Gly variant in healthy men. *Arterioscler. Thromb. Vasc. Biol.*, **29**, 1968–1974.
22. Reiner, A.P., Carty, C.L., Jenny, N.S., Nievergelt, C., Cushman, M., Stearns-Kurosawa, D.J., Kurosawa, S., Kuller, L.H. and Lange, L.A. (2008) PROC, PROCR and PROS1 polymorphisms, plasma anticoagulant phenotypes, and risk of cardiovascular disease and mortality in older adults: the Cardiovascular Health Study. *J. Thromb. Haemost.*, **6**, 1625–1632.
23. Nayak, R.C., Sen, P., Ghosh, S., Gopalakrishnan, R., Esmon, C.T., Pendurthi, U.R. and Rao, L.V. (2009) Endothelial cell protein C receptor cellular localization and trafficking: potential functional implications. *Blood*, **114**, 1974–1986.
24. Biguzzi, E., Merati, G., Liaw, P.C., Bucciarelli, P., Oganesyan, N., Qu, D., Gu, J.M., Fetiveau, R., Esmon, C.T., Mannucci, P.M. *et al.* (2001) A 23bp insertion in the endothelial protein C receptor (EPCR) gene impairs EPCR function. *Thromb. Haemost.*, **86**, 945–948.
25. Ireland, H., Konstantoulas, C.J., Cooper, J.A., Hawe, E., Humphries, S.E., Mather, H., Goodall, A.H., Hogwood, J., Juhan-Vague, I., Yudkin, J.S. *et al.* (2005) EPCR Ser219Gly: elevated sEPCR, prothrombin F1+2, risk for coronary heart disease, and increased sEPCR shedding *in vitro*. *Atherosclerosis*, **183**, 283–292.
26. Mitropoulos, K.A., Miller, G.J., Reeves, B.E., Wilkes, H.C. and Cruickshank, J.K. (1989) Factor VII coagulant activity is strongly associated with the plasma concentration of large lipoprotein particles in middle-aged men. *Atherosclerosis*, **76**, 203–208.
27. Moyer, M.P., Tracy, R.P., Tracy, P.B., van't Veer, C., Sparks, C.E. and Mann, K.G. (1998) Plasma lipoproteins support prothrombinase and other procoagulant enzymatic complexes. *Arterioscler. Thromb. Vasc. Biol.*, **18**, 458–465.
28. Jackson, A.A., Cronin, K.R., Zachariah, R. and Carew, J.A. (2007) CCAAT/enhancer-binding protein-beta participates in insulin-responsive expression of the factor VII gene. *J. Biol. Chem.*, **282**, 31156–31165.
29. Wallace, C., Newhouse, S.J., Braund, P., Zhang, F., Tobin, M., Falchi, M., Ahmadi, K., Dobson, R.J., Marcano, A.C., Hajat, C. *et al.* (2008) Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am. J. Hum. Genet.*, **82**, 139–149.
30. Smith, E.N., Chen, W., Kahonen, M., Kettunen, J., Lehtimaki, T., Peltonen, L., Raitakari, O.T., Salem, R.M., Schork, N.J., Shaw, M. *et al.* (2010) Longitudinal genome-wide association of cardiovascular disease risk factors in the bogalusa heart study. *PLoS Genet.*, **6**, pii: e1001094.
31. Waterworth, D.M., Ricketts, S.L., Song, K., Chen, L., Zhao, J.H., Ripatti, S., Aulchenko, Y.S., Zhang, W., Yuan, X., Lim, N. *et al.* (2010) Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.*, **30**, 2264–2276.
32. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T. *et al.* (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.*, **41**, 56–65.
33. Kooner, J.S., Chambers, J.C., Aguilar-Salinas, C.A., Hinds, D.A., Hyde, C.L., Warnes, G.R., Gomez Perez, F.J., Frazer, K.A., Elliott, P., Scott, J.

et al. (2008) Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat. Genet.*, **40**, 149–151.

34. Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath, S.C., Timpson, N.J., Najjar, S.S., Stringham, H.M. *et al.* (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.*, **40**, 161–169.

35. Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G., Nerup, J., Nierras, C. *et al.* (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.*, **41**, 703–707.

36. Cooper, J.D., Smyth, D.J., Smiles, A.M., Plagnol, V., Walker, N.M., Allen, J.E., Downes, K., Barrett, J.C., Healy, B.C., Mychaleckyj, J.C. *et al.* (2008) Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat. Genet.*, **40**, 1399–1401.

37. Stahl, E.A., Raychaudhuri, S., Remmers, E.F., Xie, G., Eyre, S., Thomson, B.P., Li, Y., Kurreeman, F.A., Zhernakova, A., Hinks, A. *et al.* (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.*, **42**, 508–514.

38. Raychaudhuri, S., Remmers, E.F., Lee, A.T., Hackett, R., Guiducci, C., Burtt, N.P., Gianniny, L., Korman, B.D., Padyukov, L., Kurreeman, F.A. *et al.* (2008) Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat. Genet.*, **40**, 1216–1223.

39. Kottgen, A., Pattaro, C., Boger, C.A., Fuchsberger, C., Olden, M., Glazer, N.L., Parsa, A., Gao, X., Yang, Q., Smith, A.V. *et al.* (2010) New loci associated with kidney function and chronic kidney disease. *Nat. Genet.*, **42**, 376–384.

40. Saxena, R., Hivert, M.F., Langenberg, C., Tanaka, T., Pankow, J.S., Vollenweider, P., Lyssenko, V., Bouatia-Naji, N., Dupuis, J., Jackson, A.U. *et al.* (2010) Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.*, **42**, 142–148.

41. Musunuru, K., Lettre, G., Young, T., Farlow, D.N., Pirruccello, J.P., Ejebe, K.G., Keating, B.J., Yang, Q., Chen, M.H., Lapchyk, N. *et al.* (2010) Candidate gene association resource (CARe): design, methods, and proof of concept. *Circ. Cardiovasc. Genet.*, **3**, 267–275.

42. Green, D., Ruth, K.J., Folsom, A.R. and Liu, K. (1994) Hemostatic factors in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Arterioscler. Thromb.*, **14**, 686–693.

43. Keating, B.J., Tischfield, S., Murray, S.S., Bhangale, T., Price, T.S., Glessner, J.T., Galver, L., Barrett, J.C., Grant, S.F., Farlow, D.N. *et al.* (2008) Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS ONE*, **3**, e3583.

44. Li, Y. and Abecasis, G. (2006) Mach 1.0: rapid haplotype reconstruction and missing genotype inference. *Am. J. Hum. Genet.*, **S79**, 2290.

45. Li, Y., Willer, C., Sanna, S. and Abecasis, G. (2009) Genotype imputation. *Annu. Rev. Genomics Hum. Genet.*, **10**, 387–406.

46. Chen, M-.H. and Yang, Q. (2010) GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics*, **26**, 580–581.

47. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a toolset for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 551–575.

48. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.

49. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.

50. Lo, K.S., Wilson, J.G., Lange, L.A., Folsom, A.R., Galarneau, G., Ganesh, S.K., Grant, S.K., Keating, B.J., McCarroll, S.A., Mohler, E.R. 3rd *et al.* (2011) Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Hum. Genet.*, **129**, 307–317.

51. Talmud, P.J., Drenos, F., Shah, S., Shah, T., Palmen, J., Verzilli, C., Gaunt, T.R., Pallas, J., Lovering, R., Li, K. *et al.* (2009) Gene-centric association signals for lipids and apolipoproteins identified via the Human CVD BeadChip. *Am. J. Hum. Genet.*, **85**, 628–642.

52. Brozovic, M., Stirling, Y., Harricks, C., North, W.R. and Meade, T.W. (1974) Factor VII in an industrial population. *Br. J. Haematol.*, **28**, 381–391.

53. Aulchenko, Y.S., Struchalin, M.V. and van Duijn, C.M. (2010) ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*, **11**, 134.