Published in final edited form as:
Genet Epidemiol. 2013 July ; 37(5): 512–521. doi:10.1002/gepi.21731.

A genome-wide association study for venous thromboembolism: the extended Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium

Weihong Tang, MD, PhD[#1], **Martina Teichert, PharmD, PhD**[#2,3], **Daniel I. Chasman, PhD**[#4,5], **John A. Heit, MD**[#6], **Pierre-Emmanuel Morange, MD, PhD**[#7], **Guo Li, MS**[#8], **Nathan Pankratz, PhD**[9], **Frank W. Leebeek, MD, PhD**[10], **Guillaume Paré, MD, MSc**[11], **Mariza de Andrade, PhD**[12], **Christophe Tzourio, MD, PhD**[13], **Bruce M. Psaty, MD, PhD**[14,15], **Saonli Basu, PhD**[16], **Rikje Ruiter, MD, PhD**[2], **Lynda Rose, MS**[4], **Sebastian M. Armasu, MS**[12], **Thomas Lumley, PhD**[17], **Susan R. Heckbert, MD, PhD**[18,15,33], **André G. Uitterlinden, MD, PhD**[2,19,20], **Mark Lathrop, PhD**[21], **Kenneth M. Rice, PhD**[22], **Mary Cushman, MD, MSc**[23], **Albert Hofman, MD, PhD**[2], **Jean-Charles Lambert, PhD**[24], **Nicole L. Glazer, PhD**[25], **James S. Pankow, PhD**[1], **Jacqueline C. Witteman, MD, PhD**[2], **Philippe Amouyel, MD, PhD**[24,26], **Joshua C. Bis, PhD**[8], **Edwin G. Bovill, MD**[27], **Xiaoxiao Kong, PhD**[16], **Russell P. Tracy, PhD**[28], **Eric Boerwinkle, PhD**[29], **Jerome I. Rotter, MD**[30], **David-Alexandre Trégouët, PhD**[31], **Daan W. Loth, PhD**[2,32], **Bruno H. Ch. Stricker, MB, PhD**[#2,20,32], **Paul M Ridker, MD, MPH**[#4,5], **Aaron R. Folsom, MD**[#1], and **Nicholas L. Smith, PhD**[#18,33]

[1]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA [2]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands [3]Scientific Institute Dutch Pharmacists, The Hague, The Netherlands [4]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA [5]Harvard Medical School, Boston, MA, USA [6]Division of Cardiovascular Diseases, Department of Medicine, Mayo Clinic, Rochester, MN, USA [7]INSERM UMR_S 1062; Université de la Méditerranée, Marseille, France [8]Department of Medicine, University of Washington, Seattle, WA, USA [9]Department of Lab Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA [10]Department of Haematology, Erasmus Medical Center, Rotterdam, The Netherlands [11]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, L8S4K1, Canada [12]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA [13]University of Bordeaux, U708, F-33000, Bordeaux, France [14]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA [15]Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA [16]Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA [17]Department of Statistics, University of Aukland, Aukland, NZ [18]Department of Epidemiology, University of Washington, Seattle, WA, USA [19]Netherlands Consortium for Healthy Aging, Rotterdam, The Netherlands [20]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands [21]Commissariat à l'Energie Atomique, Institut de Génomique, Centre

Corresponding author: Weihong Tang, MD, PhD, Division of Epidemiology and Community Health, University of Minnesota, 1300 South Second Street, WBOB 300, Minneapolis, MN 55454, USA. Fax: (612) 624-0315, Phone: (612) 626-9140, tang0097@umn.edu.

The authors declare no conflict of interest.

National de Génotypage, F-91057, Evry, France [22]Department of Biostatistics, University of Washington, Seattle, WA, USA [23]Departments of Medicine and Pathology, University of Vermont, Burlington, Vermont, USA [24]INSERM, UMR_S 744, Institut Pasteur de Lille; Université de Lille Nord de France, F-59019, Lille, France [25]Department of Medicine, Boston University, Boston, MA, USA [26]Centre Hospitalier Régional Universitaire de Lille, F-59037, Lille, France [27]Department of Pathology, University of Vermont, Burlington, VT, USA [28]Departments of Biochemistry and Pathology, University of Vermont, Burlington, VT, USA [29]Human Genetics Center and Human Genome Sequencing Center, University of Texas Health Science Center at Houston, Houston, Texas, USA [30]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA [31]INSERM, UMR_S 937, ICAN Institute for Cardiometabolism and Nutrition, University Pierre & Marie Curie, Paris 06, F-75013, Paris, France [32]Drug Safety Unit, Inspectorate for Health Care, The Hague, the Netherlands [33]Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA.

[#] These authors contributed equally to this work.

Abstract

Venous thromboembolism (VTE) is a common, heritable disease resulting in high rates of hospitalization and mortality. Yet few associations between VTE and genetic variants, all in the coagulation pathway, have been established. To identify additional genetic determinants of VTE, we conducted a 2-stage genome-wide association study (GWAS) among individuals of European ancestry in the extended CHARGE VTE consortium. The discovery GWAS comprised 1,618 incident VTE cases out of 44,499 participants from six community-based studies. Genotypes for genome-wide single-nucleotide polymorphisms (SNPs) were imputed to ~2.5 million SNPs in HapMap and association with VTE assessed using study-design appropriate regression methods. Meta-analysis of these results identified two known loci, in *F5* and *ABO*. Top 1,047 tag SNPs ($p \leq 0.0016$) from the discovery GWAS were tested for association in an additional 3,231 cases and 3,536 controls from three case-control studies. In the combined data from these two stages, additional genome-wide significant associations were observed on 4q35 at *F11* (top SNP rs4253399, intronic to *F11*) and on 4q28 at *FGG* (rs6536024, 9.7 kb from *FGG*) ($p < 5.0 \times 10^{-13}$ for both). The associations at the *FGG* locus were not completely explained by previously reported variants. Loci at or near *SUSD1* and *OTUD7A* showed borderline yet novel associations ($p < 5.0 \times 10^{-6}$) and constitute new candidate genes. In conclusion, this large GWAS replicated key genetic associations in *F5* and *ABO*, and confirmed the importance of *F11* and *FGG* loci for VTE. Future studies are warranted to better characterize the associations with *F11* and *FGG* and to replicate the new candidate associations.

Keywords

venous thrombosis; genetics; genome-wide association; genetic epidemiology

INTRODUCTION

Venous thromboembolism (VTE), manifesting as either deep venous thromboembolism (DVT) or pulmonary embolism (PE), is a common and clinically important cardiovascular condition with a high mortality rate. Each year in the United States, approximately 2 million adults develop DVT [Hirsh and Hoak 1996], and an estimated 600,000 PE hospitalizations and 60,000 deaths occur [Anderson et al., 1991; Goldhaber 1994; Hirsh and Hoak 1996]. VTE constitutes the third most common life-threatening cardiovascular condition after coronary heart disease (CHD) and stroke [Goldhaber 1994]. Twin and family studies suggest that VTE is highly heritable, with heritability estimates ranging from 0.5 to 0.6 [Heit et al., 2004; Larsen et al., 2003; Souto et al., 2000]. A growing number of genetic variants, most in the coagulation and fibrinolysis pathways, have been consistently linked to VTE [Anderson and Spencer 2003; Bezemer et al., 2008; Blondon et al., 2011; Cushman 2007; Dennis et al., 2012; Emmerich et al., 2001; Germain et al., 2011; Heit et al., 2012; Lowe 2006; Rosendaal 1999a; Rosendaal 1999b; Smith et al., 2007; Tregouet et al., 2009]. New variants and candidate genes may yet be identified by genome-wide investigations. In this study, we conducted a two-stage investigation of 2.5 million single nucleotide polymorphisms (SNPs) and their associations with risk of VTE among 9 epidemiologic studies that included 4,849 cases of European ancestry.

MATERIALS AND METHODS

Design

We conducted a two-stage GWAS that included discovery in 6 studies (CHARGE VTE), examination of genome-wide significant and high-signal findings from the discovery analysis in 3 independent populations, and testing of the high-signal markers in all 9 studies. Each study was approved by the Institutional Review Board of its respective institution and all participants provided informed consent.

Discovery: Study Population

The study sample for the discovery stage comprised 1,618 incident VTE cases among 44,499 participants from five cohort studies and one case-control study. The cohort studies were the Longitudinal Investigation of Thromboembolism Etiology (LITE) [Cushman 2007], which consisted two population-based cohorts [Cushman 2007] (the Atherosclerosis Risk in Communities (ARIC) Study [1989] and the Cardiovascular Health Study (CHS) [Fried et al., 1991; Tell et al., 1993]), the Rotterdam Studies (RSI and RSII) [Hofman et al., 2009], and the Women's Genome Health Study (WGHS) [Ridker et al., 2008]. The case-control study was the Heart and Vascular Health (HVH) Study of Venous Thrombosis [Smith et al., 2007].

All analyses in this study were restricted to participants of European ancestry (Table 1).

VTE Diagnosis

All VTE events were objectively-diagnosed DVT or PE that were confirmed by venous or pulmonary imaging, pathology examination of thrombus removed at surgery, or autopsy

(~5% of HVH cases were by physician diagnosis only). Among all events, each study indentified idiopathic VTE events as those that had no obvious precipitants, such as major surgery, major trauma, marked immobility, or cancer.

Further details on the study design and VTE ascertainment and classification are provided in Supplementary Methods and Supplementary Table 1.

Discovery: Genotyping, Imputation and Quality Control (QC)

GWAS—Each study performed genotyping using genome-wide SNP arrays, and conducted QC and imputation of the data to ~2.5 million autosomal SNPs identified in the HapMap II sample of European ancestry (CEU). The SNP arrays were from Affymetrix or Illumina. Before imputation, SNPs were filtered based on sample and SNP call rates, minor allele frequencies (MAF), and Hardy-Weinberg equilibrium p-values. The imputation was performed using MACH [Li et al., 2009b; Li et al., 2010] in ARIC, RSI, RSII, and WGHS, and BIMBAM [Servin and Stephens 2007] in CHS and HVH. Details of the GWAS arrays and genotyping, QC, and imputation are provided in Supplementary Methods and Supplementary Table 1.

Discovery: Statistical Methods

Study-Specific Analysis—Each study conducted regression of VTE on genotype for each SNP using additive genetic models. For cohort studies, Cox proportional-hazards regression analyses were performed using allele dosages and times-to-event calculated from cohort-entry to the time of first VTE diagnosis. Censoring occurred at time of death from other causes, loss to follow-up, or the end of follow-up (2005 in ARIC, 2001 in CHS, 2009 in RS, and 2010 in WGHS). Analyses were adjusted for age, sex, and field center in ARIC, age, sex, field center, and CHD status in CHS, age and sex in RS, and age and a single eigenvector of ancestry in WGHS. Association analyses were conducted using the ProbABEL package in ARIC, CHS, RS, and WGHS [Aulchenko et al., 2007]. In the HVH case-control study, logistic regression implemented in R was used to assess the association between allele dosage of each SNP and incident VTE while adjusting for study design variables age, age by decade, hypertension status, menopausal status, and index year. Since HVH uses incidence density sampling for the controls, under the assumption of proportional hazards, which holds under the null hypothesis of independence between genotype and VTE outcome, estimates of ORs in this study also estimate hazard ratios.

When associations were identified for loci previously shown to be associated with VTE, we conducted secondary analyses that excluded carriers of known variants or conditioned on the known variants to see if new, independent signals remained. These included the FV Leiden causal variant (rs6025) and known variants in *ABO* and *FGG*. rs6025 was genotyped separately from the GWAS in individual cohorts. Linkage disequilibrium (LD) statistics (r^2) were estimated from the HapMap CEU sample.

In secondary analyses, data were re-analyzed while restricting events to those that were idiopathic.

Meta-analysis—A fixed-effects meta-analysis with inverse-variance weighting was performed to combine regression parameter (ie, beta: log hazard ratio or OR) and standard error from the individual studies [Benjamin et al., 2009]. The following exclusion criteria were applied to SNPs in each individual study before meta-analysis: 1) beta estimates > 5 or < −5; 2) invalid standard error (NA or zero value); 3) imputation quality score (RSQ) ≤ 0.3; and 4) post-meta-analysis population-size-weighted MAF < 0.024 (based on MAF*2* total events <75) to remove SNPs for which standard error estimates are unreliable in survival analysis [Strawderman 1997]. Prior to the meta-analysis, genomic control inflation factors (λ) from individual GWAS were calculated and corrected for (when λ was greater than 1.0), to account for population stratification. The λ from individual GWAS was around 1.00 (range: 0.94-1.06, Supplementary Table 1), suggesting negligible population stratification in the individual studies. The meta-analysis yielded estimates of log risk ratio, standard error, and p-value, along with an overall λ across all SNPs. The meta-analysis was performed using METAL [Willer et al., 2010]. We used a conventional *a priori* threshold of genome-wide significance of $p< 5.0 \times 10^{-8}$.

Investigation of GWAS Findings at the Second Stage

The investigation of top GWAS hits for VTE at the second stage was performed in three case-control studies: the Mayo Clinic VTE Study (Mayo, US) [Heit et al., 2011], the MARseille THrombosis Association VTE Study (MARTHA, France) [Germain et al., 2011], and a French case-control study on early-onset VT (EOVT) [Tregouet et al., 2009]. The investigation in MARTHA and EOVT was by *in silico* analysis of existing data while that in the Mayo study combined *de novo* genotyping with *in silico* analysis.

SNP selection—We identified SNPs in the discovery set that failed to reach genome-wide significance but that nonetheless had p-values suggestive of an association with VTE. We first used a p-value threshold of 0.001 to select the top tag SNPs and functional SNPs. These were subsequently thinned to an "independent" set; independence being defined as an r^2<0.80 between the top SNP and adjacent SNPs from the same chromosome. This filter was only applied to non-functional (intronic and intergenic) SNPs. A total of 939 tag SNPs were selected. To fill the remaining capacity for the genotyping platform, the same tagging procedure was applied to additional SNPs (with p>0.001) in order of their p-value. In total, 1,047 SNPs with p≤0.0016 were selected, of which 318 were available *in silico* from the Mayo GWAS data and the remaining 729 were genotyped. The final set of 1,047 SNPs was also submitted to the MARTHA and EOVT studies, where results were obtained using *in silico* analysis.

De Novo Genotyping in Mayo—Genotyping was performed in 2,572 samples using the Illumina Veracode platform by the Mayo Genotyping lab. Of the 729 genotyped SNPs, 30 were removed due to no call, missing call rate > 20% or being monomorphic. Forty-six samples were removed due to ≥90% missing SNP calls.

Analysis of Genetic Associations with VTE—In Mayo, PLINK [Purcell et al., 2007] was used to perform logistic regression to evaluate the association of the replication SNPs with VTE adjusting for age and sex. In MARTHA and EOVT, logistic regression was

conducted and adjusted for the first four principal components of ancestry, with PLINK [Purcell et al., 2007] being used for genotyped SNPs and Mach2dat [Li et al., 2009b; Li et al., 2010] for imputed ones.

Finally, for the 1,047 SNPs examined at the second stage, an inverse-variance weighted meta-analysis was performed to combine association parameter from the discovery and second-stage studies. To allow for the two-stage testing of the 1,047 SNPs in the meta-analysis, we adjusted the significance threshold downward by a factor of two ($p<2.5\times10^{-8}$). Heterogeneity of the association parameters across studies was assessed by Cochran's Q-test[Willer et al., 2010].

RESULTS

GWAS Discovery

Manhattan and quantile-quantile (Q-Q) plots of p-value from the meta-analysis of the discovery cohorts are shown in Figure 1 and Supplementary Figure 1, respectively. The genomic inflation λ was 1.0009, consistent with negligible population stratification. A total of 34 SNPs from two genomic regions exceeded the genome-wide significance threshold of 5×10^{-8}: the *F5*/*NME7* region at chromosome 1q24.2 (19 SNPs, spanning 379 kb) and the *ABO* region at chromosome 9q34.1-q34.2 (15 SNPs, spanning 18 kb). Regional association plots are shown in Figures 2 and 3.

The signals at the *F5/NME7* region were led by an intronic variant rs2420370 and a coding variant rs6427196 in the 3' untranslated region of *F5* (RR= 1.77 and 1.82 associated with each copy of the minor G allele, Table 2). The FV Leiden variant (rs6025) was excluded from the GWAS meta-analysis because of small MAF (<0.024) or poor imputation quality. However, the significant signals at the *F5* region disappeared after excluding approximately 5% of participants who were carriers of the FV Leiden variant that was genotyped separately from the GWAS (smallest p>0.01, Supplementary Figure 2). The signals at the *ABO* region were led by rs687621 (RR= 1.37, Table 2), an intronic variant in the *ABO* gene and a strong tag for the O group: r^2=0.67 with the O functional variant (rs8176719). Adjustment for rs514659, one of the top SNPs for VTE and the strongest tags for O group (r^2=0.71), removed the associations for the remaining significant SNPs (smallest p > 0.01, Supplementary Figure 3), consistent with the signals at this region being due to the effect of the O blood group. The association of FV Leiden and *ABO* O variants with VTE is established and therefore replication was not pursued [Anderson and Spencer 2003; Bezemer et al., 2008; Blondon et al., 2011; Cushman 2007; Dennis et al., ; Emmerich et al., 2001; Germain et al., ; Heit et al., 2012; Lowe 2006; Rosendaal 1999a; Rosendaal 1999b; Tregouet et al., 2009].

The analyses for idiopathic VTE yielded similar results (Supplementary Table 2).

Meta-analysis of discovery and second-stage data for new discovery

In the meta-analysis that pooled data for the 1,047 replication SNPs from the discovery and second-stage studies, two additional loci emerged at $p< 2.5\times10^{-8}$ at chromosomes 4q28 (seven SNPs in or around *FGG*) and 4q35 (two SNPs in or around *F11*). The signals at the

FGG region were led by rs6536024 (RR=0.80, 95% confidence interval 0.76-0.85, associated with each copy of minor allele), which is 9.7 kb downstream from *FGG*; it was also the most significant SNP in the CHARGE cohorts (RR=0.79, 95% confidence interval 0.73-0.87). One of the significant SNPs (rs1049636, $p=2.54\times10^{-8}$ in all studies) is located at the 3' untranslated coding region of *FGG*. The signals at the *F11* region were led by rs4253399 (RR=1.24, 95% confidence interval 1.17-1.31), an intronic variant to *F11* (Table 2). Regional association plots at the *FGG* and *F11* regions are shown in Figures 4 and 5.

Using summary relative risks and allele frequencies from the meta-analysis, we estimated sibling relative risks (λ_s) attributable to each of the lead variants in the *ABO*, *F5*, *F11*, and *FGG* regions [Risch and Merikangas 1996]. Assuming a multiplicative model within and between variants, these four variants together account for a λ_s of 1.11, below the sibling relative risk estimate of 2.45 based on a nationwide register of hospitalizations in Sweden [Zoller et al., 2011]. These results indicate that the established common variants included on the GWAS arrays are largely insufficient to explain the heritability of VTE in populations of European ancestry.

Additionally, five high-signal variants were identified in the second-stage meta-analysis at $p<1.0\times10^{-5}$ and marked four novel chromosomal regions (Table 2). Two SNPs are from *sushi domain containing 1* (*SUSD1*); one is from *ovarian tumor* (*OTU*) *domain containing 7A* (*OTUD7A*); one is an intergenic SNP at chromosome 3p26 and about 90 kb from *contactin-6* (*CNTN6*); and one is an intergenic SNP on chromosome 5q13.3 that is about 6.1 kb from *synaptic vesicle glycoprotein 2C* (*SV2C*) (Table 2). Regional association plots at the four loci are shown in Supplementary Figures 4-7.

Test of homogeneity across studies—For the genetic associations presented in Table 2, no significant heterogeneity was detected among the CHARGE cohorts ($p>0.01$). When the CHARGE results were compared with the second-stage data, only the *ABO* variant rs687621 showed a significant difference ($p=1.5\times10^{-5}$), with the risk estimates being higher in the second-stage samples, especially in the two French studies.

DISCUSSION

We provided the results of a discovery GWAS for VTE including a total of 1,618 VTE cases from 44,499 participants of European origin in six population-based studies. We further investigated the associations for the top 1,047 SNPs identified from the GWAS in 3,231 VTE cases and 3,536 controls from one US and two French case-control studies. The discovery GWAS confirmed the well-known associations of the FV Leiden variant and the ABO O blood group [Rosendaal and Reitsma 2009; Tregouet et al., 2009]. The combined analysis of the discovery and the second stage data identified additional genome-wide significant signals at the *FGG* and *F11* regions, in which variants were associated with 20% reduced and 24% increased risk of VTE, respectively. Associations at these two loci have been previously reported but not well characterized. Four subthreshold new associations were within or close to the *SUSD1*, *OTUD7A*, *CNTN6*, and *SV2C* genes, identifying new candidate loci for VTE and coagulation pathways.

Only three GWAS studies have previously evaluated VTE, mainly involving hospital-based, case-control samples [Germain et al., 2011; Heit et al., 2012; Tregouet et al., 2009]. These three studies, two in French and one in the Mayo clinic populations, also provided replication data here. In the first French and the Mayo GWAS, the only genome-wide significant variants were those that were in high LD with the FV Leiden or in the ABO region [Heit et al., 2012; Tregouet et al., 2009]. The second French GWAS included 1,542 VTE cases and 1,110 controls and pooled data with those from the first French GWAS [Germain et al., 2011]. In the pooled analysis of the two French GWAS, variants in the *F11* and *FGG* loci also reached genome-wide significance, with the top SNPs being two intergenic variants rs3756008 and rs7659024 that are close to *F11* (<2 kb) and *FGG* (<5 kb), respectively.

FGG and *F11*

The *FGG* locus identified in this study may or may not reflect a different loci than those reported by the French GWAS [Germain et al., 2011] and other candidate gene studies [Carter et al., 2000; Grunbacher et al., 2007; Nowak-Gottl et al., 2009; Smith et al., 2007; Uitte de Willige et al., 2005; Uitte de Willige et al., 2009]. In our CHARGE discovery GWAS, the top *FGG* SNP rs6536024 is in modest LD (r^2=0.18-0.26) with three tightly linked *FGG* SNPs (rs7659024, rs2066865, rs2066854, r^2=1.0 among them) and a FGA polymorphism rs6050 (Thr312Ala, r^2= 0.88 with rs2066865) identified in these studies [Carter et al., 2000; Germain et al., 2011; Grunbacher et al., 2007; Smith et al., 2007; Uitte de Willige et al., 2005; Uitte de Willige et al., 2009]. Of note, the minor allele of rs2066865 has been associated with reduced γ' fibrinogen levels which increased VTE risk [Uitte de Willige et al., 2005]. These reported SNPs showed a weaker association with VTE in the CHARGE cohorts than the top CHARGE SNP: rs7659024 (RR=1.16, $p=1.2\times10^{-3}$), rs2066865 (RR=1.16, $p=1.3\times10^{-3}$), and rs6050 (RR=1.17, $p=7.3\times10^{-4}$) vs. rs6536024 (RR=0.79, $p=4.0\times10^{-7}$). In CHARGE, rs6536024 was also in LD with another *FGG* SNP rs1049636 (r^2=0.53; RR=0.86, $p=9.6\times10^{-4}$ for rs1049636 in CHARGE) that tagged a different haplotype than rs2066865 and was associated with increased γ' fibrinogen levels and lower risk of VTE in one of the candidate gene studies [Uitte de Willige et al., 2009]. In CHARGE, the signal for rs6536024 was modestly reduced but not completely abolished after adjustment for rs7659024 ($p=4.1\times10^{-4}$ for rs6536024) or rs1049636 ($p=7.6\times10^{-4}$). Moreover, two FGG haplotypes defined by rs1049636 and two other SNPs rs2066861 and rs2066860 were associated with DVT in a German family study of pediatric DVT [Nowak-Gottl et al., 2009]. Notably, rs2066861 is tightly linked with rs2066865 (r^2=1.0) while rs2066860 was not associated with VTE in CHARGE (p=0.42). Therefore, we postulate that the *FGG* locus identified in CHARGE either reflects a different locus than those tagged by rs7659024 or rs1049636, or it is more tightly linked with the causal variant among the haplotypes analyzed in the CHARGE meta-analysis.

The top *F11* SNP rs4253399 identified in this study is in high LD with the top variant rs3756008 from the pooled analysis of the two French GWAS (r^2=0.91) [Germain et al., 2011], indicating that they likely reflect the same underlying locus. Another two intronic variants in *F11*, rs2036914 and rs2289252, were reported to be associated with the risk of VTE in candidate gene studies [Li et al., 2009a; Smith et al., 2007]. More interestingly,

these two variants were found to confer independent risk for VTE in one study [Li et al., 2009a]. These two SNPs are closely correlated with our top *F11* variant rs4253399: r^2=0.78 and 0.53 with rs2289252 and rs2036914, respectively. Therefore, it is difficult to disentangle the independence of VTE risk associated with the three variants based on the current data.

Subthreshold Associations—Of the four candidate loci that showed subthreshold associations with VTE in this study, the most biologically compelling is *SUSD1*, which encodes the sushi domain-containing protein 1 precursor. A typical sushi domain contains approximately 60 amino acids and four cysteines [Ichinose et al., 1990] and is believed to serve as a common motif in protein-protein interaction [Komaromi et al., 2011; Wei et al., 2001]. This domain has been found in a number of proteins including factor XIII [Ichinose et al., 1990; Ichinose et al., 1986; Komaromi et al., 2011]. The B-subunit of FXIII serves as inhibitor of, and carrier for the potentially active A subunit and consists of 10 sushi domains, which appeared to be involved in interaction and complex formation with FXIII-A [Komaromi et al., 2011]. A His95Arg substitution in the sushi domains of FXIII-B was associated with increased risk of VTE in two independent studies [Komanasin et al., 2005]. Therefore, we hypothesize that SUSD1 protein might interact with factors of the coagulation pathway via the sushi domains to increase the risk of VTE. These suggestive associations for VTE, if confirmed in other studies, raise the possibility of new candidate genes and pathways for venous thrombosis.

Strengths and Limitations—The CHARGE GWAS represents a large population-based GWAS study for VTE. While all of the genome-wide significant loci have been previously reported, the population-based samples in CHARGE provide direct estimates of the VTE risk these loci confer in the general population. In addition, our study adds to the literature by confirming the importance of the *F11* and *FGG* loci, which emerged in a few recent studies and were not well characterized for VTE. The SNP arrays used in current GWAS studies, including this one, are designed to capture common genetic variations based on the HapMap data and thus may miss some additional common variants as well as low frequency and rare variants. Imputation to the HapMap data provides an uniform panel for pooling data of different SNP arrays but might be subject to measurement errors. To reduce the influence of measurement errors, all SNPs with low imputation quality were excluded from the meta-analysis. The residual measurement error introduced by imputation is not correlated with the disease outcome and thus does not bias the results. It does, however, reduces statistical power compared to measuring the full genotypes. The variants identified in *FGG* and *F11* in this and the previously published studies are either intronic or intergenic and do not appear to influence protein structure and function. It is possible that the responsible variants in the *FGG* and *F11* loci are rare, functional variants that were not included on the genome array or in the HapMap SNP panel and thus not imputed in our samples. For example, the recently completed 1000 Genomes Project Data increased the number of variants in LD with published GWAS signals by more than twofold when compared with the HapMap data [Abecasis et al., 2012]. As such, the confirmation of *FGG* and *F11* loci as important genes for VTE should stimulate finemapping studies, such as those utilizing next-generation sequencing technologies, to identify the underlying, responsible variants in these loci.

In summary, in this large GWAS, we demonstrated at genome-wide significance level key genetic associations for VTE in the general population, including with the *F5*, *ABO*, *FGG*, and *F11* loci. Our study also identified additional candidate loci that are not known to affect coagulation or thrombosis. Future studies are warranted to identify the causal variants in *F11* and *FGG* regions and to replicate the new candidate associations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

ARIC and CHS: We would like to thank the University of Minnesota Supercomputing Institute for use of the blade supercomputers. The authors thank the staff and participants of the ARIC and CHS studies for their important contributions. A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm.

RS: The authors thank Pascal Arp, Mila Jhamai, Dr. Michael Moorhouse, Marijn Verkerk and Sander Bervoets for their help in creating the GWAS database, Najaf Amin and Abbas Dehghan for their help in the x-chromosome analysis.

HVH and WGHS: We thank the study research team and the women who have generously participated in the studies.

The Mayo Clinic VTE study: We thank Cynthia E. Regnier, R.N., Jennifer L. Alkhamis, Lonnie M. Heimer, Rene M. Weatherly, Roger A. Mueller, David Tines, Ailing Xue, M.D., Ross A. Miller, Joseph J. Larson, Elysia N. Jeavons and Ann F. Beauseigneur for their excellent technical assistance.

Funding Sources

The ARIC Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The LITE was funded by grant R01-HL59367. Part of the work was supported by grant R01-HL095603.

The CHS research was supported by NHLBI contracts N01-HC-85239, N01-HC-85079 through N01-HC-85086; N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, HHSN268201200036C and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from NINDS. Additional support was provided through AG-023629, AG-15928, AG-20098, and AG-027058 from the NIA. See also http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping was supported in part by National Center of Advancing Translational Technologies CTSI grant UL1TR000124 and National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center. Additional support was also provided by the Cedars-Sinai Board of Governors' Chair in Medical Genetics (JIR).

The HVH research reported in this article was supported by the National Health Lung and Blood Institute grants HL043201, HL060739, HL068986, HL073410, HL074745, HL085251, and HL95080; and by a grant from the Leducq Foundation, Paris, France for the development of Transatlantic Networks of Excellence in Cardiovascular Research.

RS was supported from grants from the Netherlands Organization of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012); the Research Institute for Diseases in the Elderly (014-93-015; RIDE2); the Netherlands Genomics Initiative (NGI) NWO (050-060-810); Erasmus Medical Center, Erasmus University, Rotterdam, the Netherlands; Organization for the Health Research and Development (ZonMW); the Research Institute for Diseases in the Elderly (RIDE); the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam.

WGHS was supported by grants from the National Heart, Lung, and Blood Institute and the National Cancer Institute (Bethesda, MD, USA; grants HL080467, HL043851 and CA047988), the Donald W Reynolds Foundation (Las Vegas, NV, USA), the Doris Duke Charitable Foundation (New York, NY, USA), and the Foundation Leducq (Paris, France), with collaborative scientific and genotyping support provided by Amgen Inc (Thousand Oaks, CA, USA).

The Mayo Clinic VTE study was funded, in part, by grants from the National Institutes of Health, National Heart, Lung and Blood Institute (HL66216, HL83141), National Human Genome Research Institute (HG04735), and National Cancer Institute (CA92153), U.S. Public Health Service; Centers for Disease Control and Prevention (DD000235); and by Mayo Foundation.

MARTHA and EOVT studies: The MARTHA study was supported by a grant from the Program Hospitalier de la Recherche Clinique. Statistical analyses benefit from the C2BIG computing centre funded by the Fondation pour la Recherche Médicale and La Région Ile de France. The 3C Study is conducted under a partnership agreement between Inserm, the Victor Segalen –Bordeaux II University and Sanofi-Synthélabo. The Fondation pour la Recherche Médicale funded the preparation and first phase of the study. The 3C-Study is also supported by the Caisse Nationale Maladie des Travailleurs Salariés, Direction Générale de la Santé, Mutuelle Generale de l'Education Nationale, the Institut de la Longévité, Agence Française de Séurité Sanitaire des Produits de Santé, the Regional Governments of Aquitaine, Bourgogne and Languedoc-Roussillon and, the Fondation de France, the Ministry of Research-Inserm Programme ‘Cohorts and collection of biological material’. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol. 1989; 129:687–702. [PubMed: 2646917]

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]

Anderson FA Jr. Spencer FA. Risk factors for venous thromboembolism. Circulation. 2003; 107:I9–16. [PubMed: 12814980]

Anderson FA Jr. Wheeler HB, Goldberg RJ, Hosmer DW, Patwardhan NA, Jovanovic B, Forcier A, Dalen JE. A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism. The Worcester DVT Study. Arch Intern Med. 1991; 151:933–938. [PubMed: 2025141]

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007; 23:1294–1296. [PubMed: 17384015]

Benjamin EJ, Rice KM, Arking DE, Pfeufer A, van Noord C, Smith AV, Schnabel RB, Bis JC, Boerwinkle E, Sinner MF. Variants in ZFHX3 are associated with atrial fibrillation in individuals of European ancestry. Nat Genet. 2009; 41:879–881. others. [PubMed: 19597492]

Bezemer ID, Bare LA, Doggen CJ, Arellano AR, Tong C, Rowland CM, Catanese J, Young BA, Reitsma PH, Devlin JJ. Gene variants associated with deep vein thrombosis. JAMA. 2008; 299:1306–1314. others. [PubMed: 18349091]

Blondon M, Hwang M, Smith NJ. Genotyping in prothrombotic states: implications. Curr Cardiovasc Risk Rep. 2011; 5:525–532.

Carter AM, Catto AJ, Kohler HP, Ariens RA, Stickland MH, Grant PJ. alpha-fibrinogen Thr312Ala polymorphism and venous thromboembolism. Blood. 2000; 96:1177–1179. [PubMed: 10910940]

Cushman M. Epidemiology and risk factors for venous thrombosis. Semin Hematol. 2007; 44:62–69. [PubMed: 17433897]

Dennis J, Johnson CY, Adediran AS, de Andrade M, Heit JA, Morange PE, Tregouet DA, Gagnon F. The endothelial protein C receptor (PROCR) Ser219Gly variant and risk of common thrombotic disorders: a HuGE review and meta-analysis of evidence from observational studies. Blood. 2012; 119:2392–2400. [PubMed: 22251481]

Emmerich J, Rosendaal FR, Cattaneo M, Margaglione M, De Stefano V, Cumming T, Arruda V, Hillarp A, Reny JL. Combined effect of factor V Leiden and prothrombin 20210A on the risk of venous thromboembolism--pooled analysis of 8 case-control studies including 2310 cases and 3204 controls. Study Group for Pooled-Analysis in Venous Thromboembolism. Thromb Haemost. 2001; 86:809–816. [PubMed: 11583312]

Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A. The Cardiovascular Health Study: design and rationale. Ann Epidemiol. 1991; 1:263–276. others. [PubMed: 1669507]

Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, Cohen W, Oudot-Mellakh T, Antoni G, Alessi MC. Genetics of venous thrombosis: insights from a new genome wide association study. PLoS One. 2011; 6:e25581. others. [PubMed: 21980494]

Goldhaber, SZ. Epidemiology of pulmonary embolism and deep vein thrombosis.. In: Bloom, AL.; Forbes, CD.; Thomas, DP.; Tuddenham, EGD., editors. Haemostasis and Thrombosis. 3rd Edition. Churchill Livingstone; New York: 1994. p. 1327-1333.

Grunbacher G, Weger W, Marx-Neuhold E, Pilger E, Koppel H, Wascher T, Marz W, Renner W. The fibrinogen gamma (FGG) 10034C>T polymorphism is associated with venous thrombosis. Thromb Res. 2007; 121:33–36. [PubMed: 17445871]

Heit JA, Armasu SM, Asmann YW, Cunningham JM, Matsumoto ME, Petterson TM, M DEA. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. J Thromb Haemost. 2012; 10:1521–1531. [PubMed: 22672568]

Heit JA, Cunningham JM, Petterson TM, Armasu SM, Rider DN, M DEA. Genetic variation within the anticoagulant, procoagulant, fibrinolytic and innate immunity pathways as risk factors for venous thromboembolism. J Thromb Haemost. 2011; 9:1133–1142. [PubMed: 21463476]

Heit JA, Phelps MA, Ward SA, Slusser JP, Petterson TM, De Andrade M. Familial segregation of venous thromboembolism. J Thromb Haemost. 2004; 2:731–736. [PubMed: 15099278]

Hirsh J, Hoak J. Management of deep vein thrombosis and pulmonary embolism. A statement for healthcare professionals. Council on Thrombosis (in consultation with the Council on Cardiovascular Radiology), American Heart Association. Circulation. 1996; 93:2212–2245. [PubMed: 8925592]

Hofman A, Breteler MM, van Duijn CM, Janssen HL, Krestin GP, Kuipers EJ, Stricker BH, Tiemeier H, Uitterlinden AG, Vingerling JR. The Rotterdam Study: 2010 objectives and design update. Eur J Epidemiol. 2009; 24:553–572. others. [PubMed: 19728115]

Ichinose A, Bottenus RE, Davie EW. Structure of transglutaminases. J Biol Chem. 1990; 265:13411–13414. [PubMed: 1974250]

Ichinose A, McMullen BA, Fujikawa K, Davie EW. Amino acid sequence of the b subunit of human factor XIII, a protein composed of ten repetitive segments. Biochemistry. 1986; 25:4633–4638. [PubMed: 3021194]

Komanasin N, Catto AJ, Futers TS, van Hylckama Vlieg A, Rosendaal FR, Ariens RA. A novel polymorphism in the factor XIII B-subunit (His95Arg): relationship to subunit dissociation and venous thrombosis. J Thromb Haemost. 2005; 3:2487–2496. [PubMed: 16241947]

Komaromi I, Bagoly Z, Muszbek L. Factor XIII: novel structural and functional aspects. J Thromb Haemost. 2011; 9:9–20. [PubMed: 20880254]

Larsen TB, Sorensen HT, Skytthe A, Johnsen SP, Vaupel JW, Christensen K. Major genetic susceptibility for venous thromboembolism in men: a study of Danish twins. Epidemiology. 2003; 14:328–332. [PubMed: 12859034]

Li Y, Bezemer ID, Rowland CM, Tong CH, Arellano AR, Catanese JJ, Devlin JJ, Reitsma PH, Bare LA, Rosendaal FR. Genetic variants associated with deep vein thrombosis: the F11 locus. J Thromb Haemost. 2009a; 7:1802–1808. [PubMed: 19583818]

Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet. 2009b; 10:387–406. [PubMed: 19715440]

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010; 34:816–834. [PubMed: 21058334]

Lowe GD. Can haematological tests predict cardiovascular risk? The 2005 Kettle Lecture. Br J Haematol. 2006; 133:232–250. [PubMed: 16643425]

Nowak-Gottl U, Weiler H, Hernandez I, Thedieck S, Seehafer T, Schulte T, Stoll M. Fibrinogen alpha and gamma genes and factor VLeiden in children with thromboembolism: results from 2 family-based association studies. Blood. 2009; 114:1947–1953. [PubMed: 19515723]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. others. [PubMed: 17701901]

Ridker PM, Chasman DI, Zee RY, Parker A, Rose L, Cook NR, Buring JE. Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. Clin Chem. 2008; 54:249–255. [PubMed: 18070814]

Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996; 273:1516–1517. [PubMed: 8801636]

Rosendaal FR. Risk factors for venous thrombotic disease. Thromb Haemost. 1999a; 82:610–619. [PubMed: 10605758]

Rosendaal FR. Venous thrombosis: a multicausal disease. Lancet. 1999b; 353:1167–1173. [PubMed: 10209995]

Rosendaal FR, Reitsma PH. Genetics of venous thrombosis. J Thromb Haemost 7 Suppl. 2009; 1:301–304.

Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. PLoS Genet. 2007; 3:e114. [PubMed: 17676998]

Smith NL, Hindorff LA, Heckbert SR, Lemaitre RN, Marciante KD, Rice K, Lumley T, Bis JC, Wiggins KL, Rosendaal FR. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. JAMA. 2007; 297:489–498. others. [PubMed: 17284699]

Souto JC, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria JM, Coll I, Felices R, Stone W, Fontcuberta J. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Genetic Analysis of Idiopathic Thrombophilia. Am J Hum Genet. 2000; 67:1452–1459. others. [PubMed: 11038326]

Strawderman RL. An asymptotic analysis of the logrank test. Lifetime Data Anal. 1997; 3:225–249. [PubMed: 9384654]

Tell GS, Fried LP, Hermanson B, Manolio TA, Newman AB, Borhani NO. Recruitment of adults 65 years and older as participants in the Cardiovascular Health Study. Ann Epidemiol. 1993; 3:358–366. [PubMed: 8275211]

Tregouet DA, Heath S, Saut N, Biron-Andreani C, Schved JF, Pernod G, Galan P, Drouet L, Zelenika D, Juhan-Vague I. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. Blood. 2009; 113:5298–5303. others. [PubMed: 19278955]

Uitte de Willige S, de Visser MC, Houwing-Duistermaat JJ, Rosendaal FR, Vos HL, Bertina RM. Genetic variation in the fibrinogen gamma gene increases the risk for deep venous thrombosis by reducing plasma fibrinogen gamma' levels. Blood. 2005; 106:4176–4183. [PubMed: 16144795]

Uitte de Willige S, Pyle ME, Vos HL, de Visser MC, Lally C, Dowling NF, Hooper WC, Bertina RM, Austin H. Fibrinogen gamma gene 3′-end polymorphisms and risk of venous thromboembolism in the African-American and Caucasian population. Thromb Haemost. 2009; 101:1078–1084. [PubMed: 19492150]

Wei X, Orchardson M, Gracie JA, Leung BP, Gao B, Guan H, Niedbala W, Paterson GK, McInnes IB, Liew FY. The Sushi domain of soluble IL-15 receptor alpha is essential for binding IL-15 and inhibiting inflammatory and allogenic responses in vitro and in vivo. J Immunol. 2001; 167:277–282. [PubMed: 11418660]

Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]

Zoller B, Li X, Sundquist J, Sundquist K. Age- and gender-specific familial risks for venous thromboembolism: a nationwide epidemiological study based on hospitalizations in Sweden. Circulation. 2011; 124:1012–1020. [PubMed: 21824919]

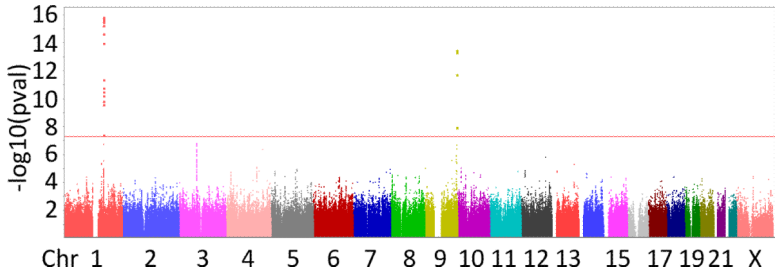

Figure 1.

Manhattan plot showing the genome-wide $-\log_{10}$ p-values against physical position for total incident VTE in CHARGE.

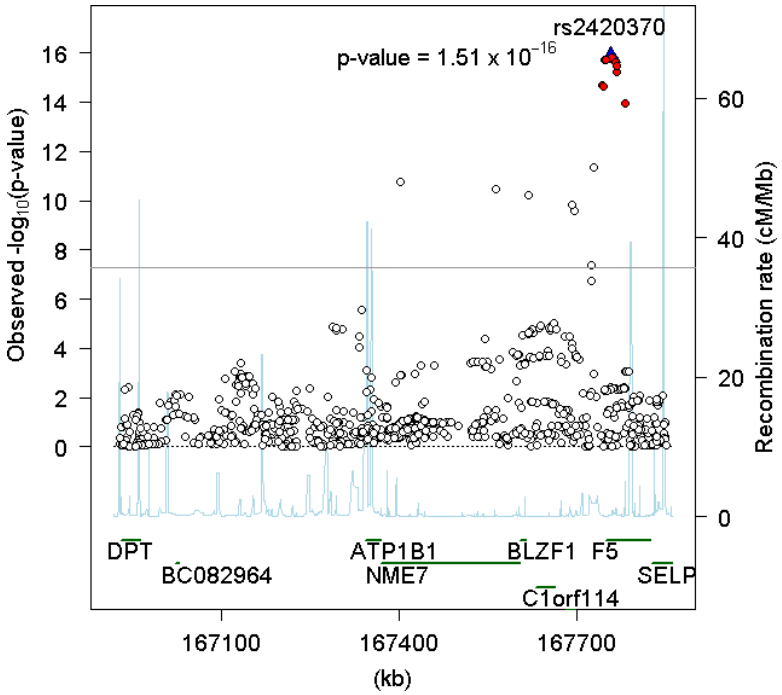

Figure 2.

Regional association plot for total incident VTE at the *F5* region on chromosome 1 in CHARGE. The horizontal line indicates the genome-wide significance threshold of $p=5.0\times10^{-8}$. The top SNP is shown by blue triangle. The color of the remaining SNPs reflects the r^2 with the top SNP based on the HapMap CEU data with the following color scheme: $r^2 \geq 0.8$ – red, $0.5 \leq r^2 < 0.8$ - orange, $0.2 \leq r^2 < 0.5$ - yellow, and $r^2 < 0.2$ - white. The light blue line represents the recombination rate (the y axis at right side) based on the data from the HapMap CEU population. Gene annotations are shown under the x axis.

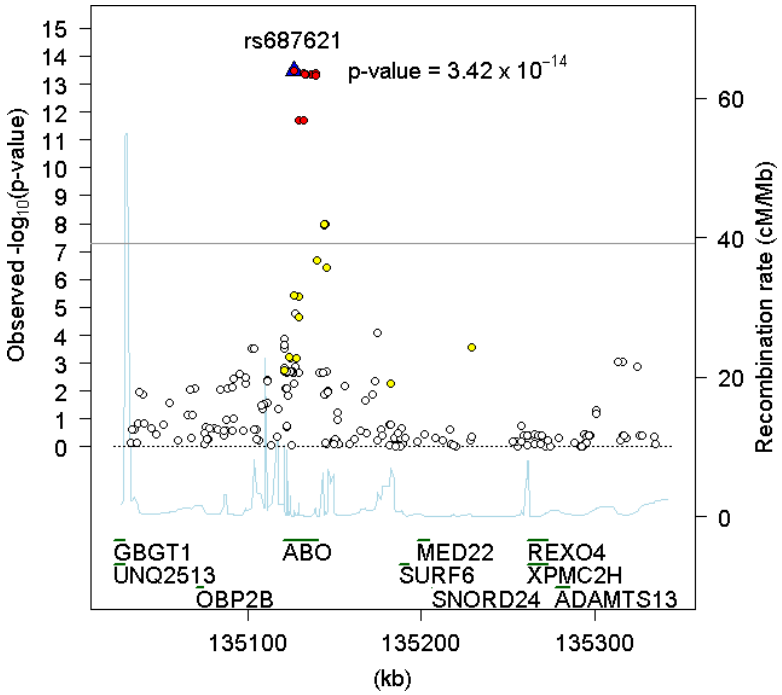

Figure 3.

Regional association plot for total incident VTE at the *ABO* region on chromosome 9 in CHARGE.

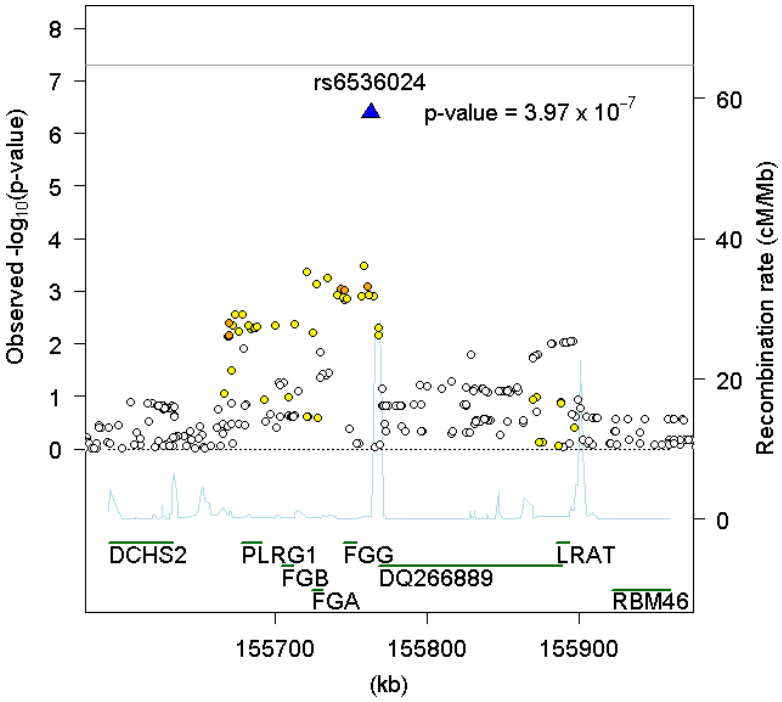

Figure 4.

Regional association plot for total incident VTE at the *FGG* region on chromosome 4 in CHARGE.

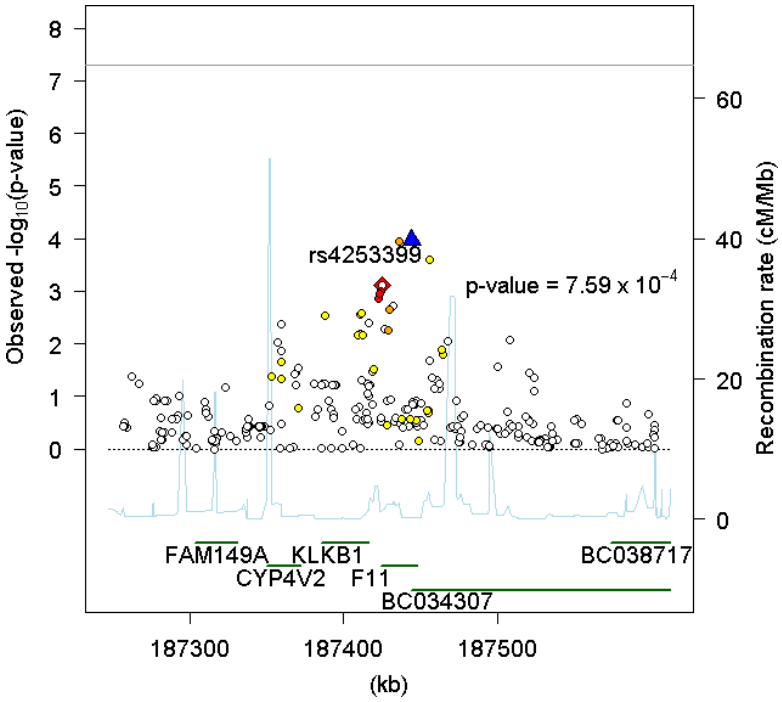

Figure 5.
Regional association plot for total incident VTE at the *F11* region on chromosome 4 in CHARGE.

Table 1

Characteristics of study designs and participants: mean (SD), percentage, or count

Characteristics	GWAS Discovery Study						Second Stage Study		
	ARIC	CHS	HVH Case/Control	RSI	RSII	WGHS	Mayo Case/Control	MARTHA Case/Control	EOVT Case/Control
Location	Four US States	Four US States	Washington State, US	Rotterdam, NL	Rotterdam, NL	Massachusetts, US	Minnesota, US	Marseille, FR	Four French Cities
Study design	Population-based cohort	Population-based cohort	Population-based case-control, incident VTE	Population-based cohort	Population-based cohort	Population-based cohort, Vitamin E and aspirin trial	Hospital-based case-control	Hospital-based case-control	Hospital-based case-control, age of first VTE ≤50 years
Exclusions	History of VTE at baseline	History of VTE at baseline	None	None	None	History of CVD, cancer, or other major chronic illness	Acquired risk factors for VTE	Strong genetic risk factors for VTE	Strong genetic and acquired risk factors for VTE
Case-control match	NA	NA	Frequency matched on age, index year, hypertension status	NA	NA	NA	Age, sex, state of residence and CVD	None	None
Participants, n	9,185	3,197	656/709	5,942	2,155	22,655	1,270/1,302	1,542/1,006	419/1,228
Mean follow-up, yrs	16.1 (3.3)	10.2 (3.0)	NA	11.1 (4.8)	7.3 (1.5)	12.8 (1.9)	NA	NA	NA
Incident VTE events, n	248	119	656	177	19	399	1270	NA*	NA*
PE	104	36	315	92	8	178	563	330 / NA	NA
VT only	187	98	341	85	11	326	707	1212 / NA	NA
Incident idiopathic VTE	89	46	385	94	NA	192	814	NA	NA
Baseline characteristics									
Mean age, yrs	54.3(5.7)	72.4(5.4)	64.7(12.1)/67.7(9.3)	69.4(9.1)	64.8(8.0)	54.7(7.1)	54.4(16.2)/55.6(15.8)	47.1(15.2)/NA	36.2(9.4) / 50.2(6.3)
Female, %	52.6	59.8	100/100	59.4	54.4	100	49.9/51.8	66.0/69.0	70.0/45.0
Mean BMI, kg/m^2	27.0(4.8)	26.2(4.4)	31.8(9.4)/29.0(6.6)	26.3(3.7)	27.3(4.2)	25.9(5.0)	30.3(7.0)/29.9(6.9)	25.2 / NA	NA
Diabetes, %	8.7	12.3	7.8/10.7	6.2†	10.1‡	2.5	NA	NA	NA
Current smoker, %	24.7	11.2	7.8/7.0	17.6	12.6	11.6	9.5/9.8	24.1 / NA	NA
Current HRT, %	20.7	13.3	30.2/31.4	3.1	4.9	43.9	NA	2.0/ NA	NA
Factor V Leiden									
Total N (carrier§, %)	8,650(5.7)	2,034(5.6)	656(14.3)/709(4.1)	5,830(4.8)	2,104(3.7)	1,182(5.1)	238(18.7)/76(5.8)	1,542(22.0) / NA	NA

VTE=venous thromboembolism, CVD=cardiovascular disease, NA=not applicable or not available.

*Including some recurrent events

†Treatment for diabetes or reported diagnosis by physicians

‡Fasting serum glucose >126mg/dl

§Minor allele homozygotes and heterozygotes for the FV Leiden (rs6025)

Table 2

Top SNP associations for total incident VTE in CHARGE and the second-stage studies (Mayo/MARTHA/EOVT).

Region	Top SNP	Position	Gene/var	A1/A2	Imput	CHARGE			Mayo/MARTHA/EOVT		All Studies	
						AFA1	RR (95% CI)	p	RR (95% CI)	p	RR (95% CI)	p
Significant at p-value < 5.0×10^{-8} in CHARGE meta-analysis												
1q24.2	rs6427196	167747847	F5/utr	G/C	0.57-1.00	0.93	1.82 (1.58 2.10)	1.97×10^{-16}	2.31 (2.04 2.62)	2.56×10^{-38}	2.07 (1.89 2.28)	4.47×10^{-51}
9q34.1-.2	rs687621	135126886	ABO/intr	A/G	0.93-0.99	0.65	1.37 (1.26 1.49)	3.42×10^{-14}	1.75 (1.62 1.89)	1.20×10^{-44}	1.55 (1.47 1.64)	1.55×10^{-52}
Significant at p-value < 2.5×10^{-8} in meta-analysis of all studies												
4q35	rs4253399	187425088	F11/intr	T/G	0.97-1.00	0.61	1.15 (1.06 1.24)	7.59×10^{-4}	1.32 (1.23 1.43)	2.07×10^{-13}	1.24 (1.17 1.31)	2.78×10^{-14}
4q28	rs6536024*	155762816	Interg	C/T	0.81-0.99	0.58	0.79 (0.73 0.87)	4.04×10^{-7}	0.81 (0.75 0.87)	5.59×10^{-8}	0.80 (0.76 0.85)	1.75×10^{-13}
Subthreshold at p-value<1.0×10^{-5} in meta-analysis of all studies												
3p26	rs6764623†	1021038	Interg	A/C	0.32-1.00	0.74	1.23 (1.11 1.38)	9.56×10^{-5}	1.14 (1.05 1.24)	0.002	1.18 (1.10 1.26)	1.57×10^{-6}
9q31.3-33.1	rs4979078	113862912	SUSD1/intr	T/C	0.84-1.00	0.86	1.31 (1.17 1.47)	2.46×10^{-6}	1.11 (1.00 1.24)	0.047	1.21 (1.11 1.30)	3.06×10^{-6}
15q13.3	rs7164569	29581216	OTUD7A/cs	A/G	0.58-0.93	0.64	0.84 (0.76 0.92)	3.54×10^{-4}	0.88 (0.82 0.95)	0.002	0.87 (0.81 0.92)	3.27×10^{-6}
5q13.3	rs3733860‡	75658560	Interg	C/A	0.73-1.00	0.87	1.22 (1.09 1.37)	6.27×10^{-4}	1.17 (1.05 1.30)	0.003	1.19 (1.10 1.29)	8.06×10^{-6}

var=variant of the top SNP, A1=allele 1 (major allele in CHARGE), A2=allele 2 (minor allele in CHARGE), Imput=ratio of observed to expected variance as a measure of imputation quality (values are the range within studies), AFA1=average allele frequency for A1 in CHARGE, RR=risk of VTE associated with 1 allele increase in the minor allele (A2), 95% CI=95% confidence interval, utr=in coding region but untranslated, intr=intron, Interg=intergenic, cs= coding-synonymous

*9.7 kb from the *FGG* gene

†90 kb from the *CNTN6* gene

‡6.1 kb to the *SV2C* gene.