HHS Public Access

Author manuscript

Thromb Haemost. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Thromb Haemost. 2017 April 03; 117(4): 758–768. doi:10.1160/TH16-08-0652.

Identification of unique venous thromboembolism-susceptibility variants in African-Americans

John A. Heit[1,2], **Sebastian M. Armasu**[3], **Bryan M. McCauley**[3], **Iftikhar J. Kullo**[1], **Hugues Sicotte**[3], **Jyotishman Pathak**[4], **Christopher G. Chute**[5], **Omri Gottesman**[6], **Erwin P. Bottinger**[6], **Joshua C. Denny**[7,8], **Dan M. Roden**[9], **Rongling Li**[10], **Marylyn D. Ritchie**[11], and **Mariza de Andrade**[3]

[1]Department of Cardiovascular Diseases, Mayo Clinic, Rochester, Minnesota, USA

[2]Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

[3]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

[4]Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medical College, Cornell University, New York, New York, , USA

[5]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University, Baltimore, Maryland, USA

[6]The Institute for Personalized Medicine, Mt. Sinai School of Medicine, New York, New York, USA

[7]Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, USA

[8]Department of Medicine, Vanderbilt University, Nashville, Tennessee, USA

[9]Department of Pharmacology, Vanderbilt University, Nashville, Tennessee, USA

[10]The Office of Population Genomics, National Human Genome Research Institute, Bethesda, Marlyland, USA

Correspondence to: John A. Heit, MD, Stabile 6-Hematology Research, Mayo Clinic, 200 First Street, SW, Rochester, MN 55905, USA, Tel.: +1 507 284 4634, Fax: +1 507 266 9302, heit.john@mayo.edu.

Supplementary Material to this article is available online at www.thrombosis-online.com.

Author contributions

John A. Heit, M. D. formulated the hypothesis, designed the study, collected the data, participated in the analyses and wrote the manuscript. Sebastian M. Armasu, M. S. performed the statistical analyses and participated in writing the manuscript. Bryan M. McCauley, M. S. performed the statistical analyses and participated in writing the manuscript. Iftikhar J. Kullo, M. D. participated in the study design, data collection, interpretation of the analyses and manuscript preparation. Jyotishman Pathak, Ph.D. participated in the study design, data collection, interpretation of the analyses and manuscript preparation. Christopher G. Chute, M. D., Dr.P.H. participated in the study design, data collection, interpretation of the analyses and manuscript preparation. Omri Gottesman, M. D. participated in the study design, data collection, interpretation of the analyses and manuscript preparation. Erwin P. Bottinger, M. D. participated in the study design, data collection, interpretation of the analyses and manuscript preparation. Joshua C. Denny, M. D., M. S. participated in the study design, data collection, interpretation of the analyses and manuscript preparation. Dan M. Roden, M. D. participated in the study design, data collection, interpretation of the analyses and manuscript preparation. Rongling Li, Ph.D. participated in the study design, data collection, interpretation of the analyses and manuscript preparation. Marylyn D. Ritchie, Ph.D. participated in the study design, data collection, interpretation of the analyses and manuscript preparation. Mariza de Andrade, Ph.D. participated in the study design and data collection, performed the statistical analyses, and wrote the manuscript.

Conflicts of interest

None declared.

[11]Center for Systems Genomics, The Huck Institutes of the Life Sciences, Eberly College of Science, Pennsylvania State University, University Park, Pennsylvania, USA

Summary

To identify novel single nucleotide polymorphisms (SNPs) associated with venous thromboembolism (VTE) in African-Americans (AAs), we performed a genome-wide association study (GWAS) of VTE in AAs using the Electronic Medical Records and Genomics (eMERGE) Network, comprised of seven sites each with DNA biobanks (total ~39,200 unique DNA samples) with genome-wide SNP data (imputed to 1000 Genomes Project cosmopolitan reference panel) and linked to electronic health records (EHRs). Using a validated EHR-driven phenotype extraction algorithm, we identified VTE cases and controls and tested for an association between each SNP and VTE using unconditional logistic regression, adjusted for age, sex, stroke, site-platform combination and sickle cell risk genotype. Among 393 AA VTE cases and 4,941 AA controls, three intragenic SNPs reached genome-wide significance: *LEMD3* rs138916004 (OR=3.2; p=1.3E-08), *LY86* rs3804476 (OR=1.8; p=2E-08) and *LOC100130298* rs142143628 (OR=4.5; p=4.4E-08); all three SNPs validated using internal cross-validation, parametric bootstrap and meta-analysis methods. *LEMD3* rs138916004 and *LOC100130298* rs142143628 are only present in Africans (1000G data). *LEMD3* showed a significant differential expression in both NCBI Gene Expression Omnibus (GEO) and the Mayo Clinic gene expression data, *LOC100130298* showed a significant differential expression only in the GEO expression data, and *LY86* showed a significant differential expression only in the Mayo expression data. *LEMD3* encodes for an antagonist of TGF-β-induced cell proliferation arrest. *LY86* encodes for MD-1 which down-regulates the pro-inflammatory response to lipopolysaccharide; *LY86* variation was previously associated with VTE in white women; *LOC100130298* is a non-coding RNA gene with unknown regulatory activity in gene expression and epigenetics.

Keywords

Genetic variation; venous thromboembolism; African Americans; association analyses

Introduction

The estimated annual incidence rates of VTE among people of European ancestry range from 104 to 183 per 100,000 person-years - rates that are similar to that of stroke (1–3). Overall VTE incidence may be higher in African-American (AA) populations (4). Studies of twins and families show that VTE is highly heritable and follows a complex mode of inheritance, involving interaction with clinical risk factors (5–7). Several genetic variants have been associated with VTE in whites (8) and AAs (9–11) including a small AA GWAS (12). However, common VTE genetic risk factors in whites are rare in AAs (13, 14), raising the possibility of as yet, undiscovered genetic variation associated with VTE in individuals with African ancestry. To test this hypothesis, we performed a larger GWAS of VTE in AAs from the Electronic Medical Records and Genomics (eMERGE) Network (15). The eMERGE network is a National Human Genome Research Institute-funded consortium engaged in the development of methods and best practices for using the electronic medical record as a tool for genomic research (16, 17).

Methods

Study design and population

We report a genome-wide investigation of genetic susceptibility variants for VTE in AAs. The study included 402 VTE case subjects and 5,078 control subjects from seven eMERGE member sites. Adult subjects only (age ≥ 18 years old per NIH-Exclusion of Children) participated in the AAs VTE study. Each study within each site was approved by the Institutional Board Review of its respective institution and all participants provided informed consent.

VTE phenotyping and the eMERGE Network

Using previously-identified Olmsted County, MN residents with objectively-diagnosed VTE from 1996 to 2005 (24), we derived and validated an electronic health record (EHR)-driven VTE phenotype extraction algorithm that leverages structured data (ICD-9-CM codes) and unstructured data (EHR clinical notes) via natural language processing (see Suppl. Material for the code, available online at www.thrombosis-online.com). Extraction algorithm operating characteristics for VTE were 100 % and 94 % positive and negative predictive value for cases and controls, respectively, at Mayo Clinic (Rochester, MN, USA). Patients with cancer ICD-9-CM codes were excluded. We used the same VTE algorithm in seven eMERGE sites on a total of 39,200 unique participants with genotype data, each one with genotypes linked to the EHR. The eMERGE phenotypes are derived and validated from an EHR-driven phenotype extraction algorithm that leverages structured data (ICD-9-CM codes) and unstructured data (EHR clinical notes) via natural language processing (16–18). Each eMERGE site specified the phenotypes of interest and created the HER-driven phenotype extraction algorithm that is validated by comparing with a random sample of EHR in cases and controls (19). Children (age < 18) and adults (age ≥ 18) participate in the eMERGE. The phenotypes are then merged with the genotyped data after quality control procedures for a GWA for a pre-specified phenotype (20, 21). To assure that EHR provides the correct assessment of the phenotype, genotype-phenotype association analyses across multiple disease phenotypes using EHR were performed and validated (22), identifying new genetic risk factors (22, 23). More details about eMERGE can be found at its website (https://emerge.mc.vanderbilt.edu/). While all eMERGE cohort members with a cancer ICD-9 code were excluded, the lack of a validated NLP algorithm and poor specificity of ICD-9 codes precluded categorising VTE cases as idiopathic or secondary or as deep-vein thrombosis (DVT) or pulmonary embolism (PE) with and without DVT. Thus, VTE was defined as either DVT or PE with or without DVT.

Control samples

Control samples were selected from each site using the EHR taking into account that no controls had an ICD-9 code related to VTE (see Suppl. Material for the Mayo Clinic VTE algorithm, available online at www.thrombosis-online.com). In this way each case is matched to its corresponding site control to avoid site bias.

Genotyping, imputation and quality control (QC)

Within each eMERGE site, DNA samples were genotyped with one or more high-density SNP arrays and were imputed for SNPs available in the 1000 Genome Project cosmopolitan reference dataset (1,092 individuals from the African (AFR), European (EUR) and Asian (ASN) ancestry panels) from the March 2012 data release. All samples are checked for relatedness between the sites as part of the QC. QC was performed similarly for each eMERGE site. First, SNP call rate was performed using all samples, where monomorphic SNPs and SNPs with poor quality were removed (SNP call rate below 95 %). Second, sample call rate was performed using the SNPs with call rate above 95 %, and samples with call rate below 95 % were removed. Study participants were selected for imputation using the eMERGE network recommended sample and SNP call rate, above 95 %, depending on the genotyping platform. The imputation was performed for each site separately due to the heterogeneity of platform and site for all autosomes using SHA-PEIT2 and IMPUTE2 software (25, 26). We used the cosmopolitan panel rather than the AFR panel because larger reference panels have been shown to increase imputation accuracy (26, 27). Imputed data for all sites were merged based on the intersection of successfully imputed SNPs among them, to about 38 million SNPs. Sites with less than four VTE cases and/or controls were eliminated. Additionally, only adult age individuals were used for the association analysis. All SNPs with high imputation quality (R-squared quality metric or information content value between 0.8 and 1.2, provided by PLINK (28) software as INFO metric in the dosage analysis) and minor allele frequency (MAF) higher than 1 % were assessed in the association with VTE risk. The INFO metric is based on the ratio of empirical and expected variance in dosage. Values closer to 1 indicate better expected quality of imputation. Since the SNPs were imputed, we used the dosage as our genotype information and avoided using the “most probable genotype” due to a potential error in predicting the corrected genotype. Therefore we did not calculate the HWE for these SNPs. The participants used in the analyses are unrelated (PIHAT < 0.125).

Population stratification

To test the African-ancestry of the study participants, we performed a principal components (PC) analysis on 400 VTE cases, 5,065 controls and 209 unrelated individuals from HapMap phase II populations (Yorubans [YRI]; European-Americans from the CEPH collection [CEU]; Han Chinese from Beijing [CHB]; and Japanese from Tokyo [JPT]) using observed genotype and Eigensoft software (29, 30). The principal components plots provided a clear representation of the genetic structure for the study participants and the HapMap reference populations. Samples that were outside ± 2 SD (standard deviation) from the mean of the PC1 and PC2 were eliminated from the analysis.

Association analysis

After quality control, the sample size analysed was of 393 VTE cases and 4,941 controls. The primary outcome was VTE status, a binary measure. To identify covariates that differ between VTE cases and controls to include in the model (i. e. potential confounders), potential covariates were examined for association with VTE status within a stepwise logistic regression model, resulting in the inclusion of age, sex, study-genotyping platform

combination, stroke, coronary heart disease and the first two principal components. Among the covariates of study site, genotyping platform and study-genotyping platform combination, the latter had a stronger univariate association with VTE status and was selected for inclusion in the stepwise model selection with the other covariates. The first two principal components explained the majority (52 %) of the total variation. Sex and the two principal components were not selected by the stepwise model but they were included as adjustment in order to control for sex and also correct for potential residual population stratification. Association analysis of imputed SNPs with VTE risk was performed using logistic regression analysis under the assumption of additive allele effects, adjusted for site-platform combination, age, sex, stroke, and the first two principal components of ancestry, with PLINK software. We also adjusted for the sickle cell disease risk variant (HBB rs77121243 T allele) as sickle cell trait/disease is more prevalent in individuals of African ancestry and is an independent risk factor for VTE (10). We applied a conventional statistical threshold of $p < 5.0xE-08$ to declare genome-wide significance (31).

Complementary approaches for confirmation of the findings

Since no other AA VTE study population of sufficient size was available for a replication study, we employed three validation strategies for the genome-wide significant SNPs. The first strategy was an internal cross-validation where the data were randomly divided into 1000 replicates of training (80 % of the samples) and testing (20 % of the samples) sets with a balanced number of VTE cases and controls within each site-genotyping platform combination. Samples from site-platform combinations with low numbers of VTE cases were excluded in order to have an acceptable number of cases and controls in each site-platform strata at each replication. For each replication, we performed an association analysis in both training and testing sets for every genome-wide significant SNP identified in the pooled analysis, using logistic regression adjusted for site-platform combination, age, sex, stroke, sickle cell risk variant (HBB rs77121243 T allele) and principal components 1 and 2. For each genome-wide significant SNP, we obtained a distribution for the odds ratio over the 1000 replicates in both training and testing sets.

The second strategy was a meta-analysis approach due to the multi-site aspect of the study. For each site, the association of SNP with VTE risk was assessed using logistic regression analysis, adjusting for genotyping platform, age, sex, stroke, sickle cell risk variant (HBB rs77121243 T allele) and principal components 1 and 2. The results at each site were meta-analysed via a random effect model based on the inverse-variance weighting. Samples from sites with low numbers of VTE cases were excluded. Heterogeneity of odds ratios in the SNP associations across studies was tested via Wolf's test as implemented in R library 'rmeta' (https://cran.r-project.org/package=rmeta).

The third validation strategy was a parametric bootstrapping approach for testing the genetic association of the genome-wide significant SNPs identified in the pooled analysis (32). We obtained parameter estimates from the original data by fitting a null-hypothesis model comprising only the covariates and calculated the fitted value for each individual. Next, a bootstrap sample of individuals was selected with replacement. This process was repeated to generate 5000 bootstrap samplings under the null hypothesis of no SNP effect. Each of these

5000 datasets was analysed using logistic regression to obtain an empirical distribution of the p-value for the SNP under the null hypothesis by fitting and comparing the null versus alternative-hypothesis models via the likelihood ratio test. The observed p-value for the SNP from the original data was then compared to the empirical null distribution. Under the parametric bootstrap, the p-value for the SNP was obtained as the fraction of empirical p-values that were smaller than or equal to the observed p-value. All the validation analyses were performed using R version 2.15 (33).

Whole blood gene expression

Whole blood gene expression profiles were available for adult Mayo Clinic patients of European ancestry with objectively-confirmed VTE (n=53) and controls (n=25), as previously reported (34). Briefly, cases and controls were recruited over the two-year period, 2009–2010, as part of the Centers for Disease Control (CDC) Thrombosis and Hemostasis Network. Patients with at least one VTE, defined as either PE or DVT of the leg or arm, with the first event occurring at age 18 years or older, and who were, at the time of enrollment, greater than 10 weeks from their most recent VTE, were approached for participation. Patients with known antiphospholipid syndrome, active or prior cancer (excluding skin cancer), infection within the past two weeks of enrollment or currently pregnant were excluded. Individuals with no prior history of VTE or known inherited clotting disorder and similar in age, sex and race to the VTE case were approached to participate as controls. Blood was collected into PAXgene RNA tubes and stored according to the manufacturer's instructions. De-identified samples were shipped to the CDC Division of Blood Disorders' Molecular and Hemostasis Laboratories for analysis. Total RNA was isolated using the PAXgene Blood RNA kit (PreAnalytiX; Qiagen, Valencia, CA, USA). cRNA samples were hybridised to Illumina HT-12 V4 Beadchips to assay whole genome gene expression, as previously described (34). The quality of the gene expression data was assessed for all Mayo Clinic samples via box plots, MA plots, average bias plots and sample call rate density plots to view experimental artifacts such as batch effects (35). Data were normalised on the $\log_2$ scale via quantile normalisation. The effect of the normalisation on the data was assessed via numerical measures such as stress and dfbeta, which are measures of the magnitude of change due to normalization (36). Criteria for exclusion were median stress >1 (no samples excluded) and median dfbeta >1.5; based on these critiera, 17 samples (5 cases and 12 controls) were excluded. Per-probe batch effects and unwanted variation remaining after normalisation were removed using the RUV-4 algorithm which utilises factor analysis of control genes (37). Differential expression was performed for 48 VTE cases and 13 controls using 'limma' package in Bioconductor (38). A secondary VTE was defined as a VTE occurring in a patient with a clear transient acquired risk factor for VTE, i. e. VTE occurring within three months after trauma, hospitalisation, prolonged immobilisation, or surgery and the post-operative setting; or in patients taking oral contraceptives or hormone replacement therapy; or during pregnancy or the postpartum period. Idiopathic VTE were defined as VTE occurring in the absence of any of these transient risk factors (34). Of the 48 VTE cases, 18 had ≥1 secondary VTE with no idiopathic VTE (mean ± SD age=54.1 ± 12.5 years; 39 % female), 17 had a single idiopathic VTE with or without additional secondary VTE (mean ± SD age=58.7 ± 12.6 years; 50 % female), and 13 had ≥2 idiopathic VTE (mean ± SD

age=46.5 ± 11.9 years; 62 % female). The mean ± SD control age was 46.5 ± years and 62 % were female.

Results

The distribution of VTE cases and controls by study site, genotyping platform and study-genotyping platform combination is shown in Suppl. Table 1 (available online at www.thrombosis-online.com). Of the nine study site-genotyping platform combinations, fourteen AA patients from three site-platforms were excluded due to a low number of cases and/or controls per site, and one non-adult AA control also was excluded. In the initial population stratification analysis (Suppl. Figure 1A, available online at www.thrombosis-online.com), 118 patients (7 cases and 111 controls) exceeded the PC1 × PC2 distribution by more than ± 2 SD from the mean and were excluded. These two PCs explained 52 % of the total variance, and the other PC contributions were incremental (data not shown). In an additional graphic representation of the first two PCs of ancestry, stratified by site-platform (Suppl. Figure 1B, available online at www.thrombosis-online.com), 13 outliers (all controls) also were excluded, leaving a total of 393 VTE cases and 4,941 controls for genome-wide association analyses from seven sites with genotyping data (Suppl. Figure 1C, available online at www.thrombosis-online.com). By using the Wald test in the univariate logistic regression model, we observed significant differences between VTE cases and controls by age (cases were older than controls, p < 1E-04), by stroke and coronary heart disease (cases had a higher prevalence of stroke [p < 1E-04] and coronary heart disease [p < 2E-03] than controls) as well as by study site, genotyping platform, and site-genotyping platform combination, with p < 1E-04 (Table 1).

After applying the SNP imputation quality filter (INFO metric) and MAF>0.01, 14,074,516 SNPs were tested for association with VTE. The sickle cell mutation (*HBB* [Hemoglobin Beta] rs77121243, known also as rs334) was significantly associated with VTE (odds ratio [OR]=1.51; 95 % confidence interval [CI]: 1.11, 2.06; p=0.009; minor allele A; MAF=0.07) and was included as a covariate in all the association analyses. Adjusting for age, sex, stroke, site-platform, sickle cell genotype (*HBB* rs77121243 T allele), and principal components 1 and 2, three intronic SNPs reached genome-wide significance: *LEMD3* rs138916004 (OR=3.2; 95 % CI: 2.1, 4.7; p=1.3E-08; MAF=0.02) on chromosome 12q14, *LY86* rs3804476 (OR=1.8; 95 % CI: 1.5, 2.3; p=2E-08; MAF=0.13) on chromosome 6p25.1, and *LOC100130298* rs142143628 (OR=4.5; 95 % CI: 2.8, 8.8; p=4.4E-08; MAF=0.01) on chromosome 8q12.2 (Figure 1 and Suppl. Table 2, available online at www.thrombosis-online.com). The genomic inflation factor was 1.005, suggesting no evidence of population stratification in our study (Figure 2). Among SNPs previously associated with VTE in whites, *F5* rs6025 (Factor V Leiden), *ABO* rs8176746 (ABO blood type A) and *ABO* rs8176719 (ABO blood type non-O) were significantly associated with VTE in AAs (Table 2); SNPs within *F2* rs1799963 (prothrombin G20210A), *F11* (procoagulant factor XI) and *FGG* (fibrinogen gamma) were not associated.

To confirm the parameter estimates and significance of the genome-wide significant SNPs using complementary statistical tools, sites with a small number of VTE cases and controls were excluded. The complementary approaches employed for the validation of these SNPs

may be affected by the reduced sample size. On average, the internal cross-validation confirmed similar magnitudes for the ORs in the training and testing sets for *LEMD3* rs138916004, *LY86* rs3804476 and *LOC100130298* rs142143628 (Suppl. Figure 2A and B, available online at www.thrombosis-online.com). The meta-analysis showed ORs and 95 % CIs comparable to those obtained in the pooled analysis (Suppl. Table 3, and Suppl. Figures 3A–C, available online at www.thrombosis-online.com). The parametric bootstrapping estimated an empirical p-value of <0.0002 and 95 % confidence upper bound of <0.0006 for all three loci.

We performed whole blood gene expression profiles on these three genes using our Mayo 53 VTE and 13 control subjects using a version of the t-test adjusted by multiple comparisons available in the 'limma' package (38). Two probes for the *LEMD3* gene, ILMN_1727361 and ILMN_2183938, had lower expression for cases than controls; however, only the first probe is highly significant (p = 7.5E-04) compared to the second probe (p = 0.107) (Figure 3 A and B). One probe for the *LY86* gene, ILMN_1807825, showed no significance in the expression (p = 0.805) (Figure 3 C). For the *LOC100130298* gene, probe ILMN_3190482 had a lower expression for cases than controls (p = 0.069) (Figure 3 D) and ILMN_3270853 showed no significance in the expression (p = 0.845) (Figure 3 E). We repeated the same analyses for age and gender adjusted expression probes and the results did not change (data not shown).

Discussion

In a GWAS of VTE in AAs, we identified the novel intronic SNPs *LEMD3* rs138916004, *LY86* rs3804476, and *LOC100130298* rs142143628 as associated with significantly increased odds of VTE. These three intronic SNPs show unique genetic background specific to African ancestry. *LEMD3* rs138916004 and *LOC100130298* rs142143628 show no genetic variation in European and Asian 1000 Genomes Project Phase 3 populations but have G (minor) and T (minor) allele frequencies of 2 % and 1 %, respectively, in the African population. *LY86* rs3804476 has G (minor) allele frequency of 44 % in European, 28 % in Asian, and 8 % in African populations. (Ensembl Genome Browser; www.ensembl.org). Due to the lack of a comparably large AA VTE case/control population available for a replication study, we employed complementary statistical approaches for the confirmation of the main findings. These statistical approaches were meant to be complementary and not to replace the parametric methods. We applied three different approaches to validate our results, including meta-analysis. The meta-analysis was performed on data from two eMERGE sites, Mount Sinai and Vanderbilt, which provided the majority of the AA samples. The three top SNPs results of these two sites are very homogeneous, with similar OR and direction, which indicates that each site replicates each other (50). Furthermore when the two sites are analysed together, the level of significance attains the whole genome wide significance of 5.0E-08 (Suppl. Figure 3, available online at www.thrombosis-online.com).

To further confirm the association of these three genes with VTE, we correlated whole blood gene expression with VTE using 48 VTE cases and 13 controls from Mayo Clinic as previously described (34). Two of the top genes, *LEMD3* and *LOC100130298* showed

significantly lower and higher gene expression in cases than controls, respectively. However, these results from European ancestry patients may not be representative of the AA population. We also reviewed GTEx data that consists of about 85 % Caucasians but unfortunately only the most significant eQTL were reported. Since our four topmost significant SNPs were not included in the GTEx SNP data, we could not draw any conclusion regarding the association between SNP and gene expression. We also investigated the GEO (GSE19151) expression data that contains 70 VTE cases and 63 controls from multiple ancestries, and only two probes were available in *LY86* and *LEMD3*; each showed significant differential expression (adjusted p-values of 3.57E-06 and 1.4E-02, respectively; Suppl. Figure 5A and B, available online at www.thrombosis-online.com), validating our findings for *LY86* and *LEMD3* gene in VTE whites.

LEMD3 encodes for LEM (LAP2, Emerin, and MAN1) domain-containing protein 3 (also known as MAN1), a 910 amino acid (82.3 kDa) integral membrane protein within the inner nuclear membrane of the nuclear envelope (39). MAN1 interacts with mediators of transforming growth factor (TGF)-β to down regulate the activation of TGF-β target genes (40, 41). Dysregulation of TGF-β signalling has been associated with cancer and cardiovascular, fibrotic, and skeletal diseases (42). Plasma levels of TGF-β1 and TGF-β2 isoforms are significantly lower in patients with recurrent VTE (43). Potential mechanisms of TGF-β signalling dysregulation in the causative pathway to VTE include suppressed expression of heme oxygenase-1 (44, 45) and promotion of monocyte adhesion, migration and chemotaxis, platelet activation and thrombogenesis by TGF-β-induced protein (TGFBIp/β ig-h3) (46).

LY86 encodes for MD-1, a member of the MD-2-related lipid-recognition protein family. MD-1 regulates the cell-surface expression of RP105 (CD180), which is a homolog to toll-like receptor 4 (TLR4). The RP105/MD-1 complex is widely expressed on antigen-presenting cells and down regulates the pro-inflammatory response to the gram negative bacterial cell wall endotoxin, lipopolysaccharide (LPS), by inhibiting LPS-induced TLR4 signalling. VTE is associated with recent urinary tract infection (UTI) (47), and UTI is most commonly caused by gram-negative bacteria and much more frequent in women. Two *LY86* SNPs (rs1073897 and rs9328375) have been associated with VTE in white women (1). However, these two SNPs were not associated with VTE in AAs. Furthermore the *LY86* rs3804476 was not associated with VTE in whites with OR from 0.94 to 1.13 with p-values from 0.105 to 0.607, and MAF of 0.4375 in whites compared to 0.1313 in AAs. Of note, *LEMD3* and *LY86* are plausible as components of the innate immunity system and activation of this system has been associated with VTE (48). We speculate that unique variation in these two genes may, in part, explain the increased incidence of VTE among AAs after exposure to such VTE risk factors as surgery, acute medical illness and trauma (4).

LOC100130298 is at 8q12.2 on the reverse strand. This long non-coding RNA gene is expressed at low level with unknown *in vivo* function; no phenotype has been associated with this gene. The intron 1 SNP rs142143628 is in linkage disequilibrium (**LD**) with two other SNPs (rs191573294 and rs187648811) in the 1000G African population (Suppl. Figure 4, available online at www.thrombosis-online.com) but not in Europeans. This index SNP

overlaps a CTCF transcription factor binding site (motifs for FoxP3, Meg2 and NF-Y) in the intron of *LOC100130928*. It is located in a region with an enhancer histone mark in cell line K562, and the chromatin is seen to be open in Th2 cells. The index SNP is in a promoter chromatin region in several cell lines (HSMM [skeletal muscle myoblasts], GM12878 [B-lymphocyte], K562[leukaemia]). In the ENCODE Roadmap, this index SNP is in an annotated transcription start site region for multiple cell lines; the details of which nucleotide is affected in the motif are shown in Suppl. Table 4 (available online at www.thrombosis-online.com). Examination of the pattern of conservation in the UCSC genome browser indicates that this SNP lies at the 5' far end of a conserved block in mammals (not conserved in vertebrates). Because of this conservation pattern, it is more likely that the relevant motif is NFY_known2 (Suppl. Table 4, available online at www.thrombosis-online.com); rs187648811 is highly conserved and alters 12 different motifs.

Among SNPs previously associated with VTE in whites, *F5* rs6025 (Factor V Leiden), *ABO* rs8176746 (ABO blood type A) and *ABO* rs8176719 (ABO blood type non-O) were significantly associated with VTE in AAs but not at the GWA significance level (Table 2). In a previous GWAS of AAs with idiopathic VTE (n=146 cases), three SNPs (rs73692310 near *IBFBP2*; rs58952918; rs28496996) reached genome-wide significance, and four SNPs (rs62322307 near *ATOH1*; rs2144940, rs25676617 and rs1998081, all near *THBD* [thrombomodulin]) were marginally significant. Of these SNPs that were tested in a replication study (n=94 cases), *THBD* rs2144940 (OR=1.89; p=0.02) and *THBD* rs1998081 (OR=1.94; p=0.02) were declared to be associated with idiopathic VTE in AAs (12). We were unable to replicate any of the original seven SNPs that were significantly or marginally associated with idiopathic VTE, although we emphasize that our study population was not limited to idiopathic VTE (Suppl. Table 5, available online at www.thrombosis-online.com). In this study, the authors did not adjust for sickle cell anaemia disease or its corresponding SNP, which is an important risk factor for VTE in AAs (10). In our study, *THBD* rs2144940 (OR=0.98; p=0.82) and *THBD* rs1998081 (OR=0.95; p=0.59) were not associated with VTE in AAs.

Since our eMERGE VTE NPL algorithm could not distinguish between idiopathic and non-cancer secondary VTE, we used our previously published VTE GWA data in whites to investigate whether genetic variation associated with VTE varies in idiopathic (n=548) and secondary VTE (n=722) (49, 50). We performed separate GWA analyses in these two sets of VTE cases compared to 1,302 controls using logistic regression. The results were very similar, implying that genetic VTE risk factors do not vary among idiopathic and secondary VTE cases (see Suppl. Figure 6A–C, available online at www.thrombosis-online.com).

In conclusion, unique genetic variation appears to be associated with VTE in African-Americans. These findings require confirmation in a larger replication study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support:

This article was partially supported by grants from the National Institutes of Health, National Heart, Lung and Blood Institute (HL66216 and HL83141 to JAH), the National Human Genome Research Institute (HG04735 to JAH, HG06379 to IJK and CGC), and by Mayo Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

We thank DeLaine Anderson for her technical assistance with the manuscript.

References

1. Heit JA, Cunningham JM, Petterson TM, et al. 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]
2. Huang W, Goldberg RJ, Anderson FA, et al. Secular trends in occurrence of acute venous thromboembolism: the Worcester VTE study (1985–2009). Am J Med. 2014; 127:829–839. e5. [PubMed: 24813864]
3. Koton S, Schneider AL, Rosamond WD, et al. Stroke incidence and mortality trends in US communities, 1987 to 2011. J Am Med Assoc. 2014; 312:259–268.
4. White RH, Keenan CR. Effects of race and ethnicity on the incidence of venous thromboembolism. Thromb Res. 2009; 123(Suppl 4):S11–S17. [PubMed: 19303496]
5. Larsen TB, Sorensen HT, Skytthe A, et al. Major genetic susceptibility for venous thromboembolism in men: a study of Danish twins. Epidemiology. 2003; 14:328–332. [PubMed: 12859034]
6. Heit JA, Phelps MA, Ward SA, et al. Familial segregation of venous thromboembolism. J Thromb Haemost. 2004; 2:731–736. [PubMed: 15099278]
7. Zöller B, Ohlsson H, Sundquist J, et al. Familial risk of venous thromboembolism in first-, second- and third-degree relatives: a nationwide family study in Sweden. Thromb Haemost. 2013; 109:458–463. [PubMed: 23348971]
8. Germain M, Chasman DI, de Haan H, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. Amer J Hum Gen. 2015; 96:532–542.
9. Dilley A, Austin H, Hooper WC, et al. Prevalence of the prothrombin 20210 G-to-A variant in blacks: infants, patients with venous thrombosis, patients with myocardial infarction, and control subjects. J Lab Clin Med. 1998; 132:452–455. [PubMed: 9851733]
10. Austin H, Key NS, Benson JM, et al. Sickle cell trait and the risk of venous thromboembolism among blacks. Blood. 2007; 110:908–912. [PubMed: 17409269]
11. Austin H, De Staercke C, Lally C, et al. New gene variants associated with venous thrombosis: a replication study in White and Black Americans. J Thromb Haemost. 2011; 9:489–495. [PubMed: 21232005]
12. Hernandez W, Gamazon ER, Smithberger E, et al. Novel genetic predictors of venous thromboembolism risk in African Americans. Blood. 2016; 127:1923–1929. [PubMed: 26888256]
13. Rees DC, Cox M, Clegg JB. World distribution of factor V Leiden. Lancet. 1995; 346:1133–1134. [PubMed: 7475606]
14. Dilley A, Austin H, Hooper WC, et al. Relation of three genetic traits to venous thrombosis in an African-American population. Am J Epidemiol. 1998; 147:30–35. [PubMed: 9440395]
15. Genome-Wide Studies in Biorepositories with Electronic Medical Record Data (U01). 2007. [cited; National Human Genome Research Institute RFA Announcement]. Available from: http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-07-005.html
16. McCarty CA, Chisholm RL, Chute CG, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genomics. 2011; 4:13. [PubMed: 21269473]

17. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med. 2013; 15:761–771. [PubMed: 23743551]
18. Pathak J, Wang J, Kashyap S, et al. Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. J Am Med Inform Assoc. 2011; 18:376–386. [PubMed: 21597104]
19. McPeek Hinz, ER., Bastarache, L., Denny, JC. A natural language processing algorithm to define a venous thromboembolism phenotype. Association of Moving Image Archivists Conference; 2013; November 2013; Richmond, VA. 2013. p. 975-983.
20. Turner S, Armstrong LL, Bradford Y, et al. Quality control procedures for genome-wide association studies. Curr Protoc Hum Genet. 2011 Chapter 1: Unit1 19.
21. Zuvich RL, Armstrong LL, Bielinski SJ, et al. Pitfalls of merging GWAS data: lessons learned in the eMERGE network and quality control procedures to maintain high data quality. Genetic Epidemiol. 2011; 35:887–898.
22. Denny JC, Crawford DC, Ritchie MD, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. Amer J Hum Gen. 2011; 89:529–542.
23. Ritchie MD, Denny JC, Crawford DC, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Amer J Hum Gen. 2010; 86:560–572.
24. Heit JA. Epidemiology of venous thromboembolism. Nat Rev Cardiol. 2015; 12:464–474. [PubMed: 26076949]
25. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda). 2011; 1:457–470. [PubMed: 22384356]
26. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013; 10:5–6. [PubMed: 23269371]
27. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. [PubMed: 19543373]
28. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Amer J Hum Gen. 2007; 81:559–575.
29. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2:e190. [PubMed: 17194218]
30. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]
31. Panagiotou OA, Ioannidis JP. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. Intern J Epidemiol. 2012; 41:273–286.
32. Buzkova, Lumley, Rice. Permutation and parametric bootstrap tests for gene-gene and gene-environment interactions. Ann Hum Gen. 2011; 75:36–45.
33. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2013. 2014
34. Lewis DA, Suchindran S, Beckman MG, et al. Whole blood gene expression profiles distinguish clinical phenotypes of venous thromboembolism. Thromb Res. 2015; 135:659–665. [PubMed: 25684211]
35. Cunningham JM, Oberg AL, Borralho PM, et al. Evaluation of a new high-dimensional miRNA profiling platform. BMC Med Genomics. 2009; 2:57. [PubMed: 19712457]
36. Mahoney DW, Therneau TM, Anderson SK, et al. Quality assessment metrics for whole genome gene expression profiling of paraffin embedded samples. BMC Res Notes. 2013; 6:33. [PubMed: 23360712]
37. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. Biostatistics. 2012; 13:539–552. [PubMed: 22101192]
38. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004; 3 Article3.

39. Lin F, Blake DL, Callebaut I, et al. MAN1, an inner nuclear membrane protein that shares the LEM domain with lamina-associated polypeptide 2 and emerin. J Biol Chem. 2000; 275:4840–4847. [PubMed: 10671519]
40. Lin F, Morrison JM, Wu W, et al. MAN1, an integral protein of the inner nuclear membrane, binds Smad2 and Smad3 and antagonizes transforming growth factor-beta signaling. Human Mol Gen. 2005; 14:437–445.
41. Bourgeois B, Gilquin B, Tellier-Lebegue C, et al. Inhibition of TGF-beta signaling at the nuclear envelope: characterisation of interactions between MAN1, Smad2 and Smad3, and PPM1A. Sci Signal. 2013; 6:ra49. [PubMed: 23779087]
42. Horbelt D, Denkis A, Knaus P. A portrait of Transforming Growth Factor beta superfamily signalling: Background matters. Int J Biochem Cell Biol. 2012 Mar.44:469–474. [PubMed: 22226817]
43. Ashfaque M, Sundquist K, Wang X, et al. Transforming growth factor (TGF)-levels and unprovoked recurrent venous thromboembolism. J Thromb Thrombolysis. 2014; 38:348–354. [PubMed: 24402195]
44. Tracz MJ, Juncos JP, Grande JP, et al. Induction of heme oxygenase-1 is a beneficial response in a murine model of venous thrombosis. Am J Pathol. 2008; 173:1882–1890. [PubMed: 18988794]
45. Mustafa S, Weltermann A, Fritsche R, et al. Genetic variation in heme oxygenase 1 (HMOX1) and the risk of recurrent venous thromboembolism. J Vasc Surg. 2008; 47:566–570. [PubMed: 18201862]
46. Kim HJ, Kim PK, Bae SM, et al. Transforming growth factor-beta-induced protein (TGFBIp/beta ig-h3) activates platelets and promotes thrombogenesis. Blood. 2009; 114:5206–5215. [PubMed: 19738031]
47. Smeeth L, Cook C, Thomas S, et al. Risk of deep vein thrombosis and pulmonary embolism after acute infection in a community setting. Lancet. 2006; 367:1075–1079. [PubMed: 16581406]
48. Reitsma P, Rosendaal F. Activation of innate immunity in patients with venous thrombosis: the Leiden Thrombophilia Study. J Thromb Haemost. 2004; 2:619–622. [PubMed: 15102017]
49. Heit JA, Armasu SM, Asmann YW, et al. 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]
50. Heller R, Bogomolov M, Benjamini Y. Deciding whether follow-up studies have replicated findings in a preliminary large-scale omics study. Proc Nat Acad Sci USA. 2014; 111:16262–16267. [PubMed: 25368172]

What is known about this topic?

- Twins and family studies show that venous thromboembolism (VTE) is highly heritable.
- Several genetic variants have been associated with VTE in whites and African-Americans (AAs).
- Common VTE genetic risk factors in whites are rare in AAs, raising the possibility of as yet, undiscovered VTE genetic variation in AAs.

What does this paper add?

- In a genome-wide association study, unique variations in 3 genes (*LEMD3, LY86, LOC100130298*) were significantly associated with VTE in AAs.
- Two genes (*LEMD3, LY86*) showed significant differential whole blood mRNA expression in VTE cases compared to controls.

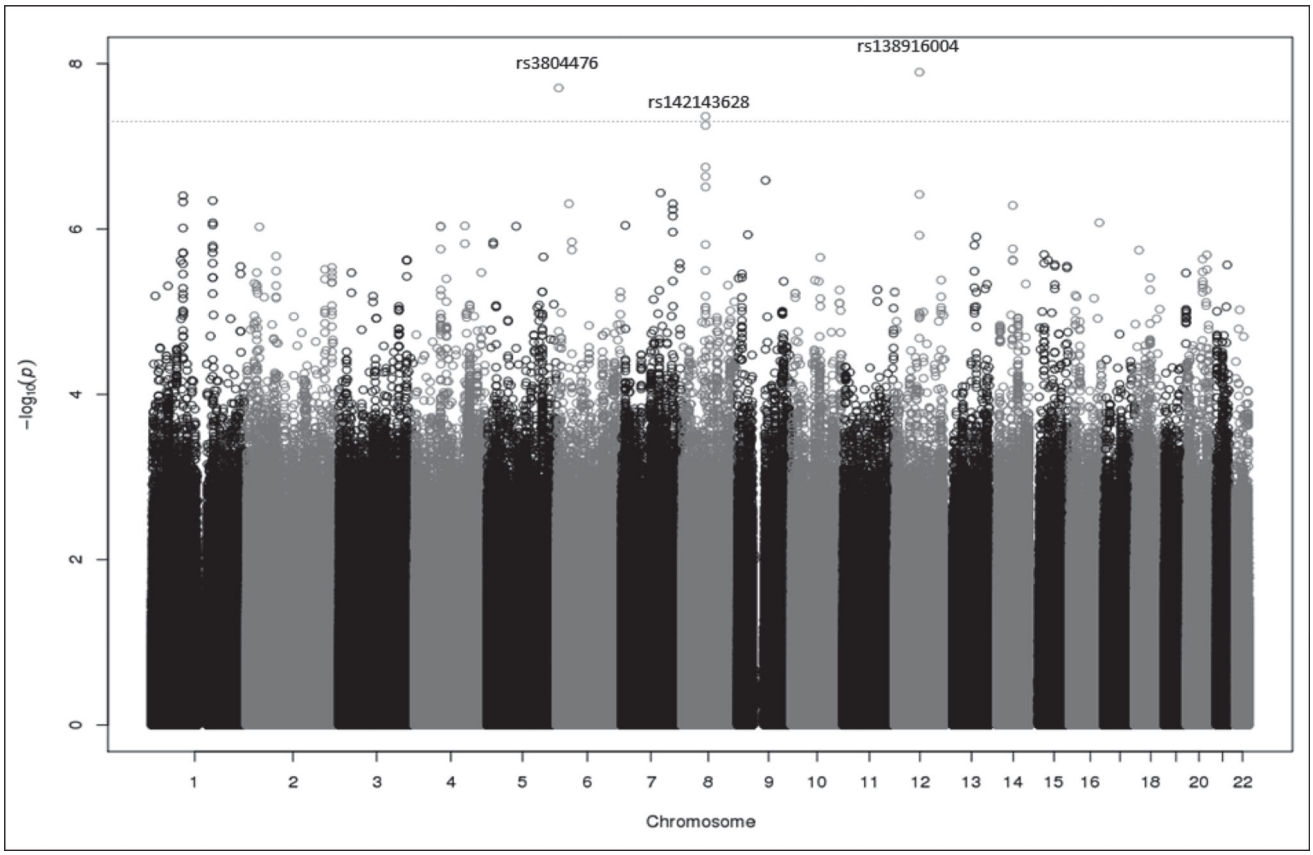

Figure 1. Manhattan plot of GWA of VTE among African-Americans using 1000G

Imputed data adjusted for site-platform, age, sex, stroke, PC1, PC2 and the sickle cell anaemia risk variant, keeping only high quality imputed SNPs (0.8 < INFO < 1.2) with MAF > 0.01.

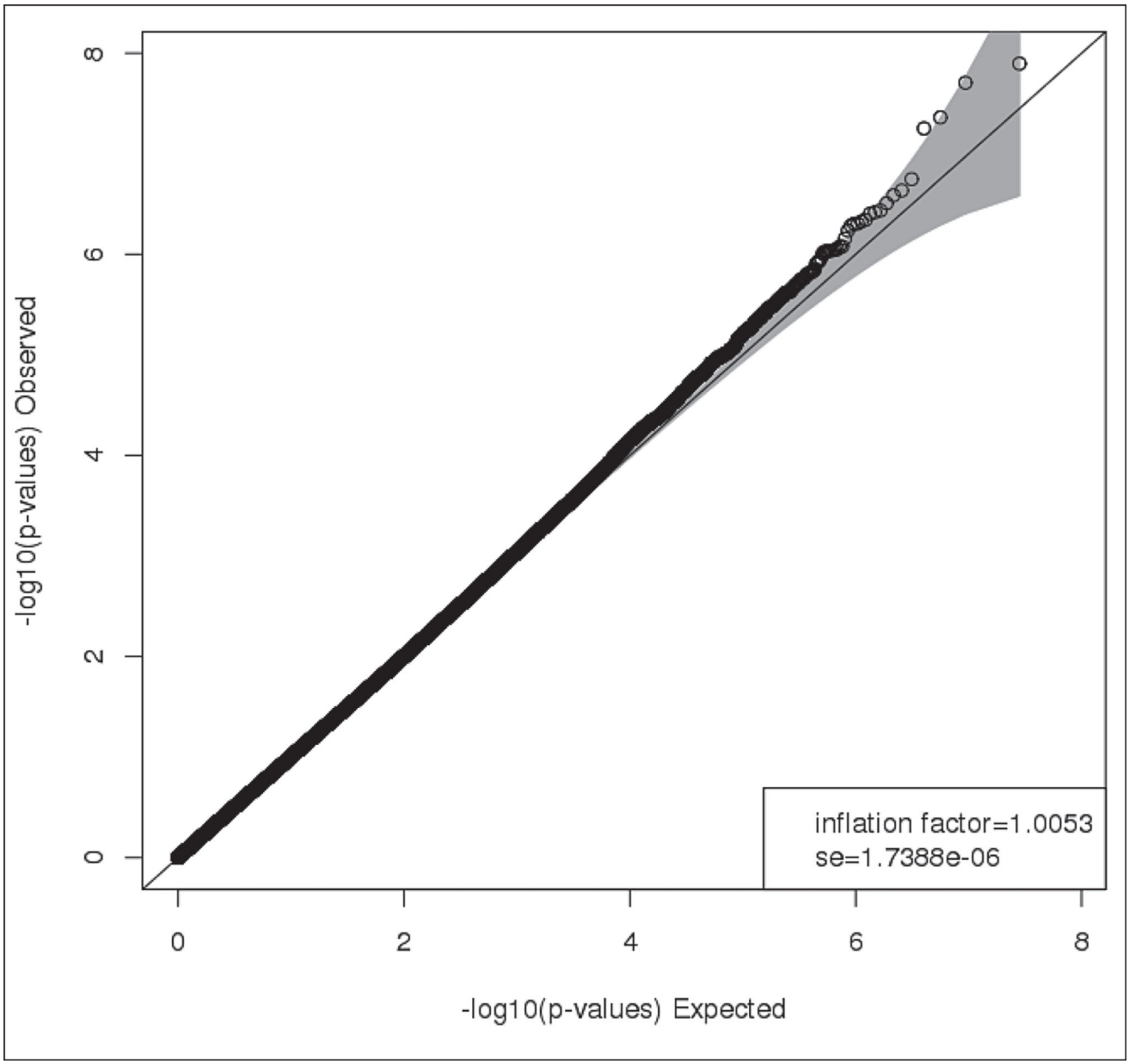

Figure 2.

Q-Q plot of the p-values of the GWA of VTE among African-Americans results from Figure 1, including the inflation factor of 1.0053 and SE of 1.7388E-06.

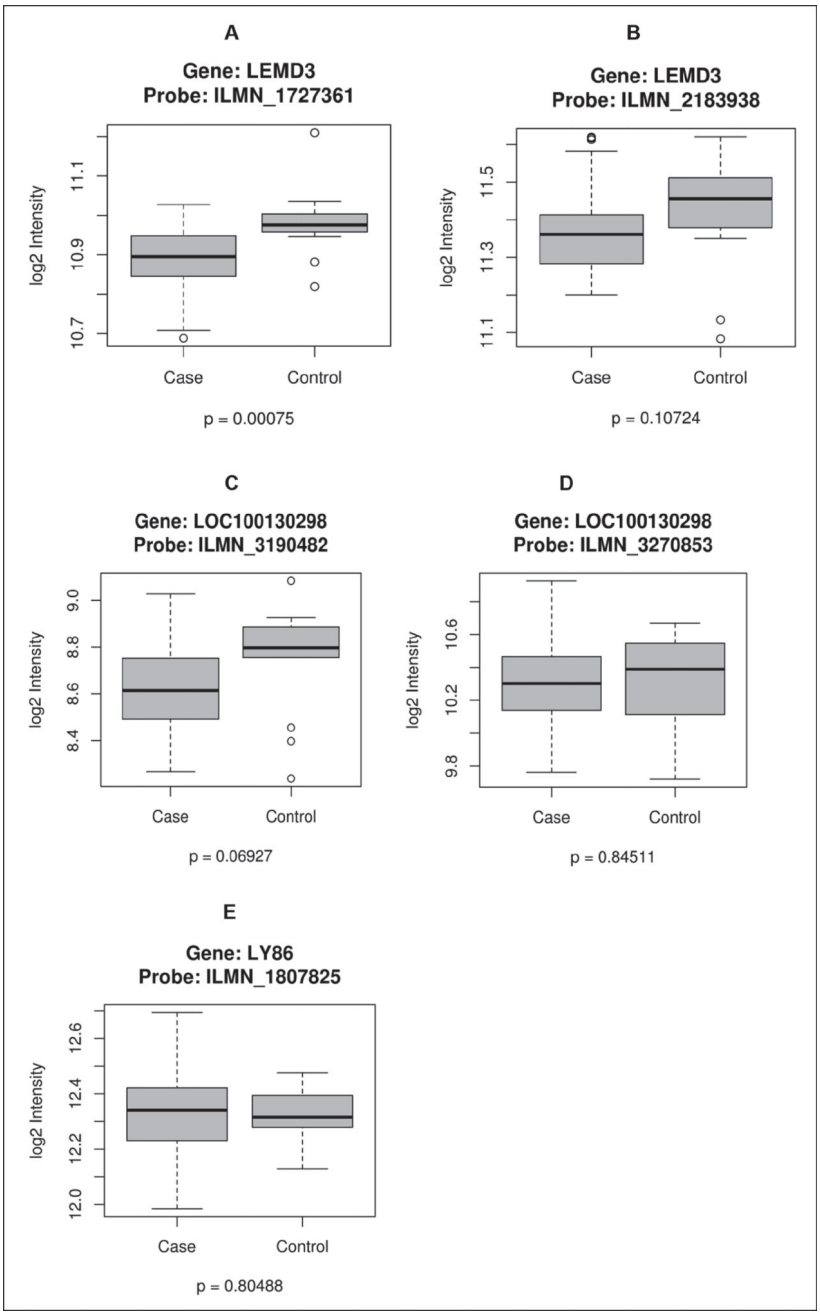

Figure 3.

Boxplots of the gene expression probes in *LEMD3* (A, B), *LOC100130298* (C, D), and *LY86* (E) genes.

Table 1

Study population demographic and clinical characteristics after quality control and population stratification exclusions.

Characteristic	Cases (n=393)	Controls (n=4941)
Age, years[†]; mean ± SD	56 ± 14	50 ± 16
Sex; % female	68.7	64.8
Stroke[†]; n (%)	197(50.1)	537(10.9)
Coronary heart disease[‡]; n (%)	30(7.6)	208(4.2)
Study Site[†]; n (%)		
Group Health Cooperatives (GHC)	6(1.5)	60(1.2)
Mt. Sinai	286(72.8)	3291(66.6)
Northwestern	8(2.0)	4(0.1)
Vanderbilt	93(23.7)	1586(32.1)
Genotyping Platform[‡]; n (%)		
Affymetrix 6.0	74(18.8)	592(12.0)
Illumina 1M	98(24.9)	1299(26.3)
Illumina 660	9(2.3)	79(1.6)
Illumina Omni	212(53.9)	2971(60.1)
Site Genotyping Platform[†]; n (%)		
GHC Illumina 660	6(1.5)	60(1.2)
Mt. Sinai Affymetrix 6.0	74(18.8)	592(12.0)
Mt. Sinai Illumina Omni	212(53.9)	2699(54.6)
Northwestern Illumina 1M	8(2.0)	4(0.1)
Vanderbilt Illumina 1M	90(22.9)	1295(26.2)
Vanderbilt Illumina 660/Omni	3(0.8)	291(5.9)

[†] $p<0.0001$;

[‡] $p<0.002$.

Table 2

Association of single nucleotide polymorphisms (previously associated with venous thromboembolism in whites) with venous thromboembolism in African-Americans.

Gene	SNP	Chromo-some	Base Pair Position	MAF	Minor Allele	OR (95% CI)	P-value
F5	rs4524	1	169511755	0.182	C	0.83(0.67,1.03)	0.08
F5	**rs6025**	**1**	**169519049**	**0.004**	**T**	**5.00(2.02,11.03)**	**0.0002**
SERPINC1	rs2227589	1	173886216	0.0605	T	0.81(0.58,1.14)	0.22
RGS7	rs670659	1	241161775	0.2086	T	0.94(0.77,1.15)	0.54
PROC	rs1158867	2	128177377	0.2964	T	0.91(0.76,1.08)	0.29
PROS1	rs121918472	3	93598150	1.00E-04	G	NA	NA
KNG1	rs710446	3	186459927	0.4964	T	1.14(0.98,1.34)	0.10
FGG	rs2066865	4	155525276	0.3028	A	1.08(0.92,1.27)	0.33
CYP4V2	rs13146272	4	187120211	0.3986	C	0.91(0.77,1.07)	0.24
F11	rs3822057	4	187188152	0.4354	A	1.01(0.87,1.18)	0.86
F11	rs2036914	4	187192481	0.3589	T	0.95(0.81,1.12)	0.55
F11	rs4253417	4	187199005	0.1545	C	1.06(0.85,1.31)	0.63
F11	rs2289252	4	187207381	0.2569	T	1.05(0.88,1.25)	0.59
STAB2	rs159381	5	58171932	0.3087	A	1.06(0.90,1.25)	0.48
HIVEP1	rs169713	6	11920517	0.4249	T	1.03(0.88,1.2)	0.71
STXBP5	rs1039084	6	147635413	0.4405	G	0.86(0.74,1.01)	0.06
ABO	rs8176747	9	136131315	0.3401	G	1.16(0.95,1.43)	0.14
ABO	**rs8176746**	**9**	**136131322**	**0.1638**	**T**	**1.33(1.09,1.62)**	**0.005**
ABO	**rs8176719**	**9**	**136132908**	**0.3047**	**TC**	**1.30(1.11,1.53)**	**0.002**
ABO	rs2519093	9	136141870	0.1122	T	1.04(0.82,1.32)	0.76
ABO	rs495828	9	136154867	0.1338	T	0.99(0.79,1.24)	0.92
TSPAN15	rs78707713	10	71245276	0.0236	C	1.02(0.59,1.76)	0.95
HBB*	**rs77121243**	**11**	**5248232**	**0.072**	**A**	**1.51(1.11,2.06)**	**0.009**
F2	rs1799963	11	46761055	0.0021	A	1.23(0.08,6.75)	0.84
VWF	rs1063856	12	6153534	0.4239	T	1.01(0.86,1.18)	0.91

Gene	SNP	Chromo-some	Base Pair Position	MAF	Minor Allele	OR (95% CI)	P-value
STAB2	rs4981021	12	104149999	0.1323	T	1.11(0.89,1.4)	0.35
TC2N	rs1884841	14	92309229	0.4827	G	0.95(0.81,1.11)	0.54
SLC44A2	rs2288904	19	10742170	0.0603	A	1.04(0.75,1.45)	0.82
GP6	rs1613662	19	55536595	0.2377	G	0.95(0.79,1.14)	0.57
PROCR	rs867186	20	33764554	0.0936	G	0.83(0.63,1.1)	0.20
PROCR	rs6087685	20	33777612	0.4053	G	1.16(0.98,1.38)	0.09

Note: the result for HBB* rs77121243 (sickle cell risk variant) is from the analysis adjusted for site-platform combination, age, sex, stroke and principal components 1 and 2; the results for all the other SNPs are from the analysis adjusted for site-platform combination, age, sex, stroke, sickle cell risk variant (HBB rs77121243 T allele) and principal components 1 and 2. The SNPs are ordered by chromosome and base-pair position. CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; SNP, single nucleotide polymorphism.