Novel genetic predictors of venous thromboembolism risk in African Americans

Wenndy Hernandez,1 Eric R. Gamazon,2,3 Erin Smithberger,1 Travis J. O’Brien,4 Arthur F. Harralson,5,6 Matthew Tuck,7 April Barbour,5 Rick A. Kittles,8 Larisa H. Cavallari,9 and Minoli A. Perera1

1Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL; 2Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN; 3Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; 4Department of Pharmacology and Physiology and 5Department of Medicine, George Washington University, Washington, DC; 6Bernard J. Dunn School of Pharmacy, Shenandoah University, Winchester, VA; 7Veterans Affairs Medical Center, Washington, DC; 8Division of Urology, Department of Surgery, University of Arizona College of Medicine, Tucson, AZ; and 9Center for Pharmacogenomics, Department of Pharmacotherapy and Translational Research, College of Pharmacy, University of Florida, Gainesville, FL

Key Points
- Our study has identified common genetic risk factors for VTE among AAs.
- These risk factors are associated with decreased thrombomodulin gene expression, suggesting a mechanistic link.

Venous thromboembolism (VTE), which encompasses deep vein thrombosis (DVT) and pulmonary embolism (PE), is a significant health problem in the United States, resulting in up to 600,000 new cases annually.1 In the United States, African Americans (AAs) exhibit the highest incidence of DVT and mortality rates of PE, having a 30% to 60% higher incidence compared with other ethnicities. The mechanisms underlying population differences in the risk of VTE are poorly understood. We conducted the first genome-wide association study in AAs, comprising 578 subjects, followed by replication of highly significant findings in an independent cohort of 159 AA subjects. Logistic regression was used to estimate the association between genetic variants and VTE risk. Through bioinformatics analysis of the top signals, we identified expression quantitative trait loci (eQTLs) in whole blood and investigated the messenger RNA expression differences in VTE cases and controls. We identified and replicated single-nucleotide polymorphisms on chromosome 20 (rs2144940, rs2567617, and rs1998081) that increased risk of VTE by 2.3-fold ($P < 6 \times 10^{-7}$). These risk variants were found in higher frequency among populations of African descent (>20%) compared with other ethnic groups (<10%). We demonstrate that SNPs on chromosome 20 are cis-eQTLs for thrombomodulin (THBD), and the expression of THBD is lower among VTE cases compared with controls ($P = 9.87 \times 10^{-6}$). We have identified novel polymorphisms associated with increased risk of VTE in AAs. These polymorphisms are predominantly found among populations of African descent and are associated with THBD gene expression. Our findings provide new molecular insight into a mechanism regulating VTE susceptibility and identify common genetic variants that increase the risk of VTE in AAs, a population disproportionately affected by this disease. (Blood. 2016;127(15):1923-1929)

Introduction

Venous thromboembolism (VTE), which encompasses deep vein thrombosis (DVT) and pulmonary embolism (PE), is a significant health problem in the United States, resulting in up to 600,000 new cases annually.1 In the United States, African Americans (AAs) exhibit the highest incidence of DVT and mortality rates of PE, having a 30% to 60% higher incidence compared with populations of European ancestry (EA) and a 74% higher incidence compared with Asian and Pacific Islander populations.2,3

A complex interplay between genetic and environmental risk factors results in VTE.4-7 Among these factors are deficiencies in protein C, protein S, and antithrombin, as well as elevated factor VIII and factor XI.3,5 Twin and family studies among populations of EA suggest that genetic factors explain up to 60% of VTE heritability.6,8 Genomic-wide association studies (GWASs) in populations of EA have confirmed the 2 well-established risk variants, factor V Leiden (rs6025) and prothrombin G20210A (rs1799963), and have identified several single-nucleotide polymorphisms (SNPs) in the ABO blood group gene (ABO) as susceptibility loci.10-12 However, these variants are found in higher frequencies among individuals of EA compared with AAs, particularly rs6025 and rs1799963 which are nearly absent in AAs. These studies suggest that genetic variation outside the well-established findings in EA populations may contribute to VTE risk in populations of African ancestry. To identify novel VTE susceptibility loci among AAs, we conducted a 2-stage analysis that included a GWAS in a discovery cohort, followed by examination of the most significant SNPs in an independent replication cohort. In addition, we validated the role of these variants in gene regulation through the integration of transcriptomic data sets.

Materials and methods

Subjects

Participants in the discovery and replication cohorts were unrelated, self-described as AA, and over the age of 18 years. Study participants provided a DNA sample (whole-blood, saliva, or mouthwash sample). Data collected on potential risk factors for VTE included age, height, weight, ethnicity, and sex. The research protocol was approved by the local institutional review boards, and there is an Inside Blood Commentary on this article in this issue.

The online version of this article contains a data supplement.


The online version of this article contains a data supplement.

From www.bloodjournal.org by guest on August 30, 2017. For personal use only.
study participants gave written informed consent. Cases had a documented history of VTE defined as proximal DVT or PE and without strong known risk factors, including prolonged hospitalization, surgery, active cancer or history of malignancy <5 years, pregnancy or puerperium, oral contraceptive use, menopausal replacement therapy, or protein C/S deficiency. VTE (DVT and/or PE) was diagnosed by physicians using different methods, including venous examination of the lower extremities using duplex ultrasound, spiral computed tomography, computed tomography pulmonary angiogram, or ventilation-perfusion scan. All cases were out-patients placed on warfarin as previously described.13 Control subjects were outpatients free of VTE. Control subjects with cancer, liver, or kidney disease/failure, arterial thrombotic disease, or risk factors for VTE (as described for cases) were excluded.

**Discovery study population**

The discovery cohort consisted of 146 VTE cases and 432 controls. VTE cases and a subset of the controls (n = 88) were obtained from two International Warfarin Pharmacogenetics Consortium (IWPC) sites: The University of Chicago and the University of Illinois at Chicago. Additional controls (n = 344) were obtained from the DC Prostate Cancer Study (DCPC) recruited at the Division of Urology at Howard University Hospital in Washington, DC.

**Replication study population**

For the replication cohort, 94 VTE cases and 65 controls were recruited from The University of Chicago, the University of Illinois at Chicago, The George Washington University Medical Faculty Associates, and the Veterans Affairs Hospital in Washington, DC. The replication cohort was independent of the discovery cohort.

**Genotyping**

For the discovery cohort, IWPC and DCPC subjects were genotyped using the Illumina 610 Quad BeadChip and the Illumina Infinium HumanM1-Duo, respectively. Genotyping procedures for each data set have been previously described.13,14 For the replication cohort, genomic DNA was isolated from either whole-blood or buccal samples as described previously.15 Genotyping was conducted using the TaqMan allelic discrimination assay according to manufacturer’s instructions. To assess genotyping reproducibility, replicate samples were included, and the concordance was >98% for each SNP. The TaqMan assay for SNP rs62322307 failed these quality control measures and consequently was not analyzed.

**Quality control**

Because different genotyping platforms were used to generate the discovery cohort, each data set (IWPC and DCPC) underwent rigorous quality control filters individually and then as a merged data set. SNPs were excluded based on a genotyping rate <95%, a minor allele frequency of <0.03, and failed Hardy-Weinberg equilibrium (HWE) tests P < .0001. SNPs on the X and Y chromosomes and mitochondrial SNPs were also excluded. Because different platforms may have distinct biases that could influence results, we took further measures to filter out errors arising from merging the IWPC and DCPC data sets, including SNPs that were not present in both data sets, SNPs that were A/T or C/G SNPs to eliminate flip-strand issues, or SNPs that were significantly (P < .05) missing between cases and controls. In addition, we performed a pseudo-GWAS between IWPC controls and DCPC controls and removed SNPs with an association P value of <10−3 (925 SNPs). After exclusion criteria, the final number of genome-wide genotyped SNPs for the discovery cohort was 514,419. Genome-wide genotype data were used to validate sex, as well as identity by descent. No sample had a call rate of <95%, missingness >0.10, sex misspecification, or identity by descent >0.125. For the replication cohort, SNPs were also excluded when the genotyping rate was <95% or failed HWE test P values were <.00001. All quality control procedures were conducted using PLINK.16

**Global ancestry**

Potential population stratification was examined by principal component analysis conducted through genome-wide complex trait analysis17 using a linkage disequilibrium (LD)-pruned (r2 >0.2) set of 149,606 markers (supplemental Figure 1, available on the Blood Web site). Percentage West African ancestry was determined for each individual using ancestry-informative markers for European and West African ancestry.18 Individual ancestry estimates were obtained using the Bayesian Markov Chain Monte Carlo method implemented in the program STRUCTURE 2.3.19

**Imputation**

Genotypes were phased using SHAPEIT and imputed with IMPUTE2 using reference files from the ‘1000 Genomes haplotypes–Phase I integrated variant set release (v3) in NCBI build 37 (hg19) coordinates.’20–22 SNPs were excluded if the minor allele frequency was <0.03, the imputation quality was <0.6, and the HWE P value was <.00001, resulting in 10 690 342 SNPs for analysis.

**Statistical analysis**

A quantile-quantile plot of expected and observed P values revealed no evidence for systematic genotype calling error, and the genomic inflation factor (based on median χ2) was1.0047, indicating sufficient control for possible population stratification (supplemental Figure 2). Covariates and the first 10 principal components were tested as single covariates for association with VTE risk using IBM SPSS Statistics version 19.0.0 package (SPSS, Chicago, IL). Sex was tested only among the IWPC subjects due to the DCPC data set comprising only males. The association of each SNP with risk of VTE for the discovery cohort was conducted using SNPTEST23 v2.5 and for the replication cohort using PLINK,16 adjusting for age. A P value <5.0 × 10−8 was considered significant. In the replication cohort, independent SNPs with a highly suggestive association to VTE risk (P < 5.0 × 10−5) were genotyped. The significance threshold was set at P < .016 (0.05/3 SNPs); SNP rs1998081 was genotyped to confirm LD with rs2144940 and rs2567617. Results from the discovery and replication cohorts were meta-analyzed using the software METAL.24 Gene region plots of top SNPs were generated with LocusZoom.25

**Thrombomodulin gene expression**

We used the Genotype-Tissue Expression (GTEx) Portal (http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi) to retrieve precomputed significant cis and trans expression quantitative trait loci (eQTLs) from whole-blood tissue tested in 338 samples.26 To examine whether thrombomodulin (THBD) is differentially expressed between VTE patients and healthy controls, we used whole-blood gene expression data from the Gene Expression Omnibus (GEO), accession number GSE1915127. The microarray data set, consisting of 70 adults with ±1 prior VTE on warfarin and 63 healthy controls, was analyzed using an independent sample Student t test and the Benjamin-Hochberg false discovery rate correction for multiple testing (P < .05).28

**Results**

Demographic and clinical characteristics for the discovery and replication cohorts are provided in Table 1. Only mean age was statistically significant between VTE cases and controls in both the discovery and replication cohorts (54.9 ± 16.9 and 58.6 ± 16.0, P < .001; and 59.3 ± 10.9 and 63.0 ± 16.4, P = .04; respectively [Table 1]), therefore, all analyses were adjusted for age. Utilizing healthy controls recruited from the DCPC resulted in an overrepresentation of males among controls in the discovery cohort, therefore the association between sex and risk of VTE excluded DCPC controls (Table 1). After excluding DCPC controls, sex was not found to be significantly associated with risk of VTE in the remaining discovery cohort, and this lack of association was also observed in the replication cohort. Participants in the discovery cohort clustered between the HapMap CEU (northern and western European ancestry) and YRI (African ancestry) samples, as expected (supplemental Figure 1). Only 1 sample deviated from the expected clustering and was excluded from the analysis.
The first 10 principle components were not associated with disease status.

**Association of genetic variants with VTE**

In the discovery cohort, we identified 7 SNPs that increased the risk of VTE by 2.18- to 3.04-fold. Among these, SNP rs73692310 on chromosome 7 (odds ratio [OR], 3.04; 95% confidence interval [CI], 1.8-4.3; P = 1.73 × 10−3) and SNPs rs85952918 (OR, 2.48; 95% CI, 1.6-2.9; P = 4.01 × 10−7) and rs28496996 (OR, 2.44; 95% CI, 1.6-3.6; P = 1.07 × 10−8) on chromosome 18 reached genome-wide significance. On chromosome 20, SNPs rs2144940 (OR, 2.18; 95% CI, 1.6-2.9; P = 3.52 × 10−5), rs2567617 (OR, 2.17; 95% CI, 1.6-2.9; P = 4.01 × 10−7), and rs1998081 (OR, 2.28; 95% CI, 1.6-3.1; P = 5.17 × 10−7), as well as SNP rs62322307 on chromosome 4 (OR, 2.79; 95% CI, 1.8-4.3; P = 2.25 × 10−7) were strongly suggestive of association to VTE risk (Table 2). These risk alleles were found either almost exclusively or in higher frequency among populations of African descent (Table 3). All SNPs are intergenic; however, rs73692310 is ~50 kb from IGFBP3; rs2144940, rs2567617, and rs1998081 are located between THBD and CD93; and the closest gene to rs62322307 is ATOH1 (Figure 1).

To validate the association between SNPs and VTE risk in the discovery cohort, we sought to replicate these findings in an independent AA cohort. LD between SNPs was obtained from the “1000 Genomes among Americans of African Ancestry in SW USA.” For SNPs in LD (coefficient of determination r² ≥ 0.8), we chose to genotype the SNP with the lowest P value for each pair. To validate LD with SNP rs2144940, an exception was made for rs1998081. SNP rs62322307 did not pass genotyping quality control measures and was therefore not tested. The replication study confirmed a significant association with increased risk of VTE for rs2144940 and rs1998081 (Table 2). Carriers of the minor allele (C) for rs2144940 and the minor allele (T) for rs1998081 had an increased risk of VTE (OR, 1.89; 95% CI, 1.1-3.3; P = .02 and OR, 1.94; 95% CI, 1.1-3.5; P = .02, respectively) compared with noncarriers (Table 2). In the meta-analysis of the results, rs2144940 and rs1998081 reached genome-wide significance (1.88 × 10−8 and 4.62 × 10−8, respectively; Table 2). In our replication cohort, we were able to confirm LD between rs2144940 and rs1998081 (r² = 0.82). Together, these data support the association of SNPs rs2144940, rs2567617, and rs1998081 with increased risk of VTE among AAs.

We also compared previously identified VTE risk alleles in populations of EA with our discovery cohort (Table 4). SNPs rs6025 (factor V) and rs1799963 (coagulation factor II [F2]) were monomorphic; however, we replicated 3 previously identified ABO SNPs associated with risk of VTE in our discovery cohort, although not at genome-wide significance (P > .002; Table 4). More recently, a very large meta-analysis among populations of EA identified rs78707713 (TSPAN15) and rs2288904 (SLC44A2) as susceptibility loci for VTE (Table 4). However, rs78707713 is found in very low frequency outside of EA populations, and rs2288904 was not associated with VTE risk in our discovery cohort (Table 4).

**Effect of VTE-associated variants on gene expression**

To identify a plausible biological function for our top SNP associations with VTE risk, we used the GTEx Portal, which provides information on correlations between tissue-specific gene expression levels and

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>Nearby genes</th>
<th>MA</th>
<th>IM</th>
<th>MAF cases (Discovery cohort)</th>
<th>MAF cases (Replication cohort)</th>
<th>OR (95% CI) Discovery cohort</th>
<th>OR (95% CI) Replication cohort</th>
<th>OR (95% CI) Meta-analysis</th>
<th>P Discovery cohort</th>
<th>P Replication cohort</th>
<th>P Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>rs62322307</td>
<td>ATOH1</td>
<td>T</td>
<td>0.94</td>
<td>0.15</td>
<td>0.06</td>
<td>2.79 (1.8-4.3)</td>
<td>2.25 × 10⁻⁷</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>7</td>
<td>rs73692310</td>
<td>IGFBP3</td>
<td>T</td>
<td>0.78</td>
<td>0.15</td>
<td>0.05</td>
<td>3.04 (2.0-4.7)</td>
<td>1.73 × 10⁻²</td>
<td>0.09</td>
<td>0.07</td>
<td>1.27 (0.4-2.7)</td>
<td>.60</td>
</tr>
<tr>
<td>18</td>
<td>rs58952918</td>
<td>NA</td>
<td>A</td>
<td>0.90</td>
<td>0.17</td>
<td>0.08</td>
<td>2.48 (1.7-3.7)</td>
<td>1.07 × 10⁻⁸</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>18</td>
<td>rs26496996</td>
<td>NA</td>
<td>G</td>
<td>0.90</td>
<td>0.17</td>
<td>0.08</td>
<td>2.44 (1.6-3.6)</td>
<td>1.13 × 10⁻⁸</td>
<td>0.13</td>
<td>0.11</td>
<td>1.34 (0.6-2.6)</td>
<td>.45</td>
</tr>
<tr>
<td>20</td>
<td>rs2144940</td>
<td>THBD, CD93</td>
<td>C</td>
<td>0.97</td>
<td>0.31</td>
<td>0.17</td>
<td>2.18 (1.6-2.9)</td>
<td>3.52 × 10⁻⁷</td>
<td>0.35</td>
<td>0.21</td>
<td>1.89 (1.1-3.3)</td>
<td>.016</td>
</tr>
<tr>
<td>20</td>
<td>rs2567617</td>
<td>THBD, CD93</td>
<td>G</td>
<td>0.98</td>
<td>0.31</td>
<td>0.17</td>
<td>2.17 (1.6-2.9)</td>
<td>4.01 × 10⁻⁷</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>20</td>
<td>rs1998081</td>
<td>THBD, CD93</td>
<td>NA</td>
<td>0.27</td>
<td>0.14</td>
<td>0.22</td>
<td>2.28 (1.6-3.1)</td>
<td>5.17 × 10⁻⁷</td>
<td>0.30</td>
<td>0.18</td>
<td>1.94 (1.1-3.5)</td>
<td>.016</td>
</tr>
</tbody>
</table>

SNP rs1998081 was genotyped.

Chr, chromosome; IM, infometric (takes a value between 0 and 1, whereby values near 1 indicate that an SNP has been imputed with high certainty); MA, minor allele; MAF, minor allele frequency; NA, not applicable; ND, not determined due to failed assay; NT, not tested due to high LD.

**Table 2. Top allelic associations to VTE risk among AAs**

*Age adjusted.*
genetic variation. We found that rs1998081, rs2567617, and rs2144940 genotypes are associated with differential \textit{THBD} gene expression in whole blood ($P < 1.3 	imes 10^{-7}$, $P = 4.8 	imes 10^{-6}$, and $P = 4.6 	imes 10^{-6}$, respectively; Figure 2). This information suggests that SNPs rs1998081, rs2567617, and rs2144940 are \textit{cis}-eQTLs (map within 500 kb of the transcription start site) for \textit{THBD}, a candidate gene in the coagulation pathway. Furthermore, the lower \textit{THBD} gene expression levels observed in the presence of the minor alleles are in accordance with its association with increased risk of VTE. To examine whether \textit{THBD} gene expression levels vary between VTE cases and controls, we conducted a whole-genome differential expression analysis of data obtained from the GEO. We found mean \textit{THBD} gene expression levels to be significantly lower in VTE cases (7.15 ± 0.59) compared with controls (7.61 ± 0.54; unadjusted $P = 8.10 \times 10^{-6}$; adjusted for multiple testing $P = 4.31 \times 10^{-5}$; Figure 3), and the variance ($r^2$) explained by \textit{THBD} expression levels to be 14%.

Discussion

Our study is the first to investigate genetic variation associated with risk of VTE at a genome-wide level, which allowed us to identify regulatory variants and \textit{THBD} gene expression predictors of VTE affecting AAs specifically. In the United States, VTE remains associated with significant morbidity and mortality and disproportionately affects AAs. Our GWAS identified 3 novel SNPs located on chromosome 20 (rs2144940 [OR, 2.18; 95% CI, 1.6-2.9; $P = 3.52 \times 10^{-7}$], rs2567617 [OR, 2.17; 95% CI, 1.6-2.9; $P = 4.01 \times 10^{-7}$], and rs1998081 [OR, 2.28; 95% CI, 1.6-3.1; $P = 5.17 \times 10^{-7}$]) associated with increased risk of VTE among AAs, which were validated in an independent cohort (Table 2). These risk alleles are found in much higher frequency among populations of African descent (~20%-30%) compared with European and Asian populations (8% and 5%, respectively) (Table 3). Through bioinformatics analyses of whole-blood transcriptome data, we determined that rs2144940, rs2567617, and rs1998081 are significant \textit{cis}-eQTLs for \textit{THBD}, and the minor alleles are associated with decreased \textit{THBD} gene expression (Figure 2). In addition, when comparing VTE cases to controls, \textit{THBD} gene expression was found to be significantly lower in cases vs controls ($P = 8.0 \times 10^{-6}$; Figure 3). \textit{THBD} plays a pivotal role in the regulation of coagulation. \textit{THBD} is an endothelial glycoprotein that binds to thrombin and thus dramatically suppresses the amount of thrombin available for clot formation. \textit{THBD} acts as an intrinsic anticoagulant by forming a 1:1 complex with the coagulation factor thrombin (F2) and altering F2 specificity for several substrates, ultimately acting as an antithrombotic factor. In addition, the F2:THBD complex activates

Table 3. Minor allele frequency distribution among populations from the 1000 Genomes

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNP</th>
<th>Discovery cohort</th>
<th>ASW</th>
<th>AFR</th>
<th>ASN</th>
<th>EUR</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>rs62322307 (T)</td>
<td>0.08</td>
<td>0.11</td>
<td>0.10</td>
<td>0.0</td>
<td>0.03</td>
</tr>
<tr>
<td>7</td>
<td>rs73692310 (T)</td>
<td>0.08</td>
<td>0.06</td>
<td>0.07</td>
<td>0.0</td>
<td>0.001</td>
</tr>
<tr>
<td>18</td>
<td>rs58652918 (A)</td>
<td>0.10</td>
<td>0.10</td>
<td>0.11</td>
<td>0.0</td>
<td>0.003</td>
</tr>
<tr>
<td>18</td>
<td>rs28496996 (G)</td>
<td>0.11</td>
<td>0.12</td>
<td>0.11</td>
<td>0.0</td>
<td>0.003</td>
</tr>
<tr>
<td>20</td>
<td>rs2144940 (C)</td>
<td>0.21</td>
<td>0.24</td>
<td>0.27</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>20</td>
<td>rs2567617 (G)</td>
<td>0.21</td>
<td>0.24</td>
<td>0.27</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>20</td>
<td>rs1998081 (T)</td>
<td>0.17</td>
<td>0.21</td>
<td>0.24</td>
<td>0.05</td>
<td>0.08</td>
</tr>
</tbody>
</table>

AFR, African; ASN, Asian; ASW, Americans of African Ancestry in SW USA; EUR, European.

Figure 1. Locus-specific plots of genotyped and imputed SNP results for top loci. Position on chromosome 7 (A), chromosome 18 (B), chromosome 20 (C), and chromosome 4 (D). Purple diamond indicates the most significant finding in the region. Colors on the marks represent LD ($r^2$) with the top SNP. Chromosomal positions and linkage LD are based on hg19/1000 Genomes Mar 2012 AFR (http://csg.sph.umich.edu/locuszoom/).
protein C, leading to degradation of factors V and VIII.30 Consequently, *THBD* is an important candidate gene in VTE risk. However the specific *THBD* SNPs associated with VTE risk have not been definitively identified.32,33 The functional implication that the minor alleles of rs2144940 (C), rs2567614 (G), and rs1998081 (T) are associated with lower *THBD* gene expression, combined with lower *THBD* gene expression in VTE cases compared with controls, supports our findings that carriers of the minor alleles for rs2144940, rs2567617, and rs1998081 have an increased risk of VTE (Figure 2; Table 2).

Recently, a study among African-Caribbean DVT patients found thrombin levels to be significantly higher compared with DVT patients of EA and healthy African-Caribbean control subjects.34 It is possible that a combination of higher levels of thrombin and genetic polymorphisms that reduce *THBD* gene expression may significantly increase an individual’s risk for VTE. Furthermore, a previous study has suggested that chromosome 20 may harbor common variants yet to be identified that could contribute ~7% of the total genetic variance underlying VTE susceptibility.10 Our results help place these previous findings in the context of a specific gene and regulatory SNPs that affect the expression of this gene.

GTEx data, which consist of samples from mostly (85%) EA identified rs2424508 as highly associated with VTE expression (P = 1.0 x 10^{-6}; data not shown). In individuals of EA, rs2424508 is in strong LD with rs2144940, rs2567617, and rs1998081 (r^2 > 0.8), but not in populations of African ancestry. This may explain the association to *THBD* gene expression observed in GTEx data and the lower association to VTE risk found in our discovery cohort for rs2424508 (OR, 1.93; 95% CI, 1.4-2.7; P = .0002). Nonetheless, it demonstrates the potential of rs2144940, rs2567617, and rs1998081 to affect *THBD* gene expression in other populations, albeit with a smaller effect, given the low minor allele frequency in populations outside of Africa. According to GEO (GSE19151) expression data, we found *THBD* gene expression to be lower among VTE cases compared with controls. The data were collected from cases during warfarin treatment; therefore, the effect of warfarin on *THBD* gene expression cannot be assessed independently of case status. SNPs rs2144940, rs2567617, and rs1998081 are located in close proximity to CD93, which has been implicated in acute myocardial infarction.35 Based on the GSE19151 data set, CD93 and *THBD* gene expression levels are highly correlated (β = 0.42, P < .001).

The 2 well-established VTE risk alleles, rs6025 (factor V Leiden) and rs1799963 (prothrombin G20210A), which are used clinically, are rare in AAs and are therefore of limited clinical utility in this population.36-38 As expected, we did not observe these risk alleles in our discovery cohort; yet, among EA populations, these risk alleles continue to be highly associated with VTE susceptibility (Table 4). Since the early 1960s, the ABO blood group has been recognized as a risk factor for VTE, with the non-O blood types having a higher risk of VTE compared with O blood type.19 In the United States, AAs exhibit a higher incidence of VTE and mortality rates from the disease compared with populations of European and Asian ancestry.34,40 Nonetheless, AAs have a higher percentage of O blood type, which is in the opposite direction of what would be expected.41 Several GWASs have identified polymorphisms in the *ABO* gene to be associated with risk of VTE among populations of EA.42-45 Although not at GWAS

<table>
<thead>
<tr>
<th>Chr</th>
<th>SNP</th>
<th>Allele</th>
<th>AF</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>AF</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs6025 (F5)</td>
<td>T</td>
<td>0.0</td>
<td>NA</td>
<td>NA</td>
<td>0.06</td>
<td>2.56</td>
<td>1.40 x 10^{-12}</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>rs1799963 (F2)</td>
<td>A</td>
<td>0.0</td>
<td>NA</td>
<td>NA</td>
<td>0.02</td>
<td>1.69</td>
<td>3.00 x 10^{-2}</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>rs567621 (ABO)</td>
<td>G</td>
<td>0.43</td>
<td>1.55 (1.2-2.0)</td>
<td>0.002</td>
<td>0.43</td>
<td>1.84</td>
<td>6.69 x 10^{-22}</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>rs565922 (ABO)</td>
<td>C</td>
<td>0.36</td>
<td>1.52 (1.2-2.0)</td>
<td>0.002</td>
<td>0.43</td>
<td>1.85</td>
<td>1.84 x 10^{-22}</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>rs657152 (ABO)</td>
<td>A</td>
<td>0.44</td>
<td>1.39 (1.1-1.8)</td>
<td>0.03</td>
<td>0.45</td>
<td>1.7</td>
<td>2.00 x 10^{-17}</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>rs17997713 (TSPAN15)</td>
<td>T</td>
<td>0.971</td>
<td>NA</td>
<td>NA</td>
<td>0.08</td>
<td>1.28</td>
<td>5.74 x 10^{-11}</td>
<td>29</td>
</tr>
<tr>
<td>9</td>
<td>rs2286904 (SLC4A4A)</td>
<td>G</td>
<td>0.93</td>
<td>0.90 (0.5-1.5)</td>
<td>0.75</td>
<td>0.79</td>
<td>1.19</td>
<td>1.07 x 10^{-9}</td>
<td>29</td>
</tr>
</tbody>
</table>

**Table 4. Comparison of previously identified VTE risk alleles between the discovery cohort and populations of EA**

**Figure 2. THBD gene expression levels.** THBD gene expression levels by rs1998081 (A), rs2567617 (B), and rs2424500 (C) genotypes in whole blood. Data were obtained from the GTEx Analysis V6 for 338 individuals. The box and whisker plot shows the correlation of genotype with THBD gene expression. The minor allele for each SNP represents the risk allele in our study, and increasing copies of the risk allele are associated with decreased THBD gene expression in whole blood. The x-axis of each plot corresponds to the 3 observed genotypes for each SNP. The y-axis represents the normalized gene expression. The SNPs are located on chromosome 20 and are in high LD with each other (r^2 > 0.8). The P value refers to the allelic association to THBD gene expression levels.
that our novel SNPs are also associated with decreased THBD gene expression. Taken together, our findings support a novel role for SNPs rs2144940, rs2567617, and rs1998081 as VTE risk alleles among AAs. In addition, we further validate that rs6025 (factor V Leiden) and rs1799963 (prothrombin G20210A) are extremely rare in AAs and are therefore of limited clinical utility in risk assessment in this population as the current standard of care, and demonstrate the limited role of ABO variants in VTE risk in AAs. Thus, our findings may help better understand the etiology of VTE among AAs and highlight the importance of conducting population-specific research in precision medicine. These results demonstrate the unique discoveries that are possible through ethnic-specific genomic studies due to differences in LD structure and allele frequencies between populations.

Acknowledgments

The authors thank Dan Nicolae, Department of Medicine, Section of Genetic Medicine, for his valuable input in the analysis process; and the RIKEN research institute for their ongoing collaboration that provided our high-quality genome-wide genotyping.

This study was supported (in part) by research funding from the National Collaborative on Aging Faculty Awards Program (T.J.O., A.F.H., and M.T.); the American Heart Association Midwest Affiliate Grant-In-Aid (10GRNT3750024) (L.H.C.); the National Institutes of Health National Heart, Lung, and Blood Institute grants K23 HL089808-01A2 and R21 HL106097-01A1 and National Institute on Minority Health and Health Disparities grant 1R01MD009217-01 (M.A.P.); the University of Chicago Cardiovascular Sciences Training grant 5T32 HL007381 (W.H.); CA157823, and National Institutes of Health National Institute of Mental Health grants R01 MH101820 and R01 MH090937 (E.R.G.).

Authorship

Contribution: W.H. analyzed and interpreted the data and cowrote the paper; E.R.G. provided analytical support and analyzed the GEO gene expression data; E.S., A.B., and R.A.K. were responsible for sample processing, DNA extraction, and/or genotyping; R.A.K., T.J.O., A.F.H., M.T., and L.H.C. provided patient samples and clinical information; and M.A.P. contributed to the design of the study, data analysis, data interpretation, and cowrote the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Minoli A. Perera, Section of Genetic Medicine, Department of Medicine, University of Chicago, 900 E. 57th St, Room 3220B, Chicago IL 60637; e-mail: mperera@bsd.uchicago.edu.

References


Novel genetic predictors of venous thromboembolism risk in African Americans