

THROMBOSIS AND HEMOSTASIS

Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism

Sara Lindström,^{1,2} Lu Wang,³ Erin N. Smith,^{4,5} William Gordon,¹ Astrid van Hylckama Vlieg,⁶ Mariza de Andrade,⁷ Jennifer A. Brody,⁸ Jack W. Pattee,⁹ Jeffrey Haessler,² Ben M. Brumpton,¹⁰⁻¹² Daniel I. Chasman,^{13,14} Pierre Suchon,^{15,16} Ming-Huei Chen,^{17,18} Constance Turman,¹⁹ Marine Germain,²⁰ Kerri L. Wiggins,⁸ James MacDonald,³ Sigrid K. Braekkan,^{5,21} Sebastian M. Armasu,⁷ Nathan Pankratz,²² Rebecca D. Jackson,²³ Jonas B. Nielsen,²⁴ Franco Giulianini,¹³ Marja K. Puurunen,¹⁸ Manal Ibrahim,¹⁵ Susan R. Heckbert,^{1,25} Scott M. Damrauer,^{26,27} Pradeep Natarajan,²⁸⁻³¹ Derek Klarin,^{28,32,33} The Million Veteran Program, Paul S. de Vries,³⁴ Maria Sabater-Lleal,^{35,36} Jennifer E. Huffman,³⁷ The CHARGE Hemostasis Working Group, Theo K. Bammler,³ Kelly A. Frazer,^{4,5,38} Bryan M. McCauley,⁷ Kent Taylor,³⁹ James S. Pankow,⁴⁰ Alexander P. Reiner,^{1,2} Maiken E. Gabrielsen,¹⁰ Jean-François Deleuze,^{41,42} Chris J. O'Donnell,^{17,18,43} Jihye Kim,¹⁹ Barbara McKnight,^{2,44} Peter Kraft,¹⁹ John-Bjarne Hansen,^{5,21} Frits R. Rosendaal,⁶ John A. Heit,⁷ Bruce M. Psaty,^{25,45} Weihong Tang,⁴⁰ Charles Kooperberg,² Kristian Hveem,¹⁰ Paul M. Ridker,^{13,14} Pierre-Emmanuel Morange,^{15,16,46} Andrew D. Johnson,^{17,18} Christopher Kabrhel,⁴⁷⁻⁴⁹ David-Alexandre Tréguët,²⁰ and Nicholas L. Smith,^{1,25,50} on behalf of the INVENT Consortium

¹Department of Epidemiology, University of Washington, Seattle, WA; ²Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; ³Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA; ⁴Department of Pediatrics and Rady Children's Hospital, University of California San Diego, La Jolla, CA; ⁵K.G. Jebsen Thrombosis Research and Expertise Center, Department of Clinical Medicine, UiT—The Arctic University of Norway, Tromsø, Norway; ⁶Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; ⁷Department of Health Sciences Research, Mayo Clinic, Rochester, MN; ⁸Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA; ⁹Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN; ¹⁰K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; ¹¹Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; ¹²Clinic of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; ¹³Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA; ¹⁴Department of Medicine, Harvard Medical School, Boston, MA; ¹⁵Laboratory of Haematology, La Timone Hospital, Marseille, France; ¹⁶Center for Cardiovascular and Nutrition research (C2VN), Université Aix-Marseille, Institut National de la Recherche Agronomique (INRA), INSERM, Marseille, France; ¹⁷Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Framingham, MA; ¹⁸The Framingham Heart Study, Framingham, MA; ¹⁹Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA; ²⁰INSERM UMR_S 1219, Bordeaux Population Health Research Center, University of Bordeaux, Bordeaux, France; ²¹Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway; ²²Department of Laboratory Medicine and Pathology, School of Medicine, University of Minnesota, Minneapolis, MN; ²³Division of Endocrinology, Diabetes, and Metabolism, The Ohio State University, Columbus OH; ²⁴Division of Cardiology, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI; ²⁵Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle, WA; ²⁶Department of Surgery, Corporal Michael Crescenz VA Medical Center, Philadelphia, PA; ²⁷Department of Surgery, Perleman School of Medicine, University of Pennsylvania, Philadelphia, PA; ²⁸Boston VA Healthcare System, Boston, MA; ²⁹Cardiovascular Research Center and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; ³⁰Department of Medicine, Harvard Medical School, Boston, MA; ³¹Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; ³²Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA; ³³Department of Surgery, Massachusetts General Hospital, Boston, MA; ³⁴Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; ³⁵Unit of Genomics of Complex Diseases, Institut de Recerca de l'Hospital de Sant Pau, IIB-Sant Pau, Barcelona, Spain; ³⁶Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden; ³⁷Center for Population Genomics, MAVERIC, VA Boston Healthcare System, Boston, MA; ³⁸Institute of Genomic Medicine, University of California San Diego, La Jolla, CA; ³⁹Los Angeles Biomedical Research Institute and Department of Pediatrics, Harbor—University of California Los Angeles Medical Center, Torrance CA; ⁴⁰Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN; ⁴¹Centre National de Recherche en Génétique Humaine, Direction de la Recherche Fondamentale, Le Commissariat à l'énergie atomique et aux énergies alternatives, Evry, France; ⁴²The Centre d'Etude du Polymorphisme Humain (CEPH), Fondation Jean Dausset, Paris, France; ⁴³Million Veteran Program, Veteran's Administration, Boston, MA; ⁴⁴Department of Biostatistics, University of Washington, Seattle, WA; ⁴⁵Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA; ⁴⁶Centre de Ressources Biologiques Assistance Publique—Hôpital de Marseille, HemoVasc, Marseille, France; ⁴⁷Center for Vascular Emergencies, Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA; ⁴⁸Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; ⁴⁹Department of Emergency Medicine, Harvard Medical School, Boston, MA; and ⁵⁰Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research and Development, Seattle, WA

KEY POINTS

- We identified new genetic loci contributing to the risk of venous thromboembolism, some of which are outside known coagulation pathways.
- We provide evidence that blood traits may contribute to the underlying biology of venous thromboembolism risk.

Venous thromboembolism (VTE) is a significant contributor to morbidity and mortality. To advance our understanding of the biology contributing to VTE, we conducted a genome-wide association study (GWAS) of VTE and a transcriptome-wide association study (TWAS) based on imputed gene expression from whole blood and liver. We meta-analyzed GWAS data from 18 studies for 30 234 VTE cases and 172 122 controls and assessed the association between 12 923 718 genetic variants and VTE. We generated variant prediction scores of gene expression from whole blood and liver tissue and assessed them for association with VTE. Mendelian randomization analyses were conducted for traits genetically associated with novel VTE loci. We identified 34 independent genetic signals for VTE

risk from GWAS meta-analysis, of which 14 are newly reported associations. This included 11 newly associated genetic loci (C1orf198, PLEK, OSMR-AS1, NUGGC/SCARA5, GRK5, MPHOSPH9, ARID4A, PLCG2, SMG6, EIF5A, and STX10) of which 6 replicated, and 3 new independent signals in 3 known genes. Further, TWAS identified 5 additional genetic loci with imputed gene expression levels differing between cases and controls in whole blood (SH2B3, SPSB1, RP11-747H7.3, RP4-737E23.2) and in liver (ERAP1). At some GWAS loci, we found suggestive evidence that the VTE association signal for novel and previously known regions colocalized with expression quantitative trait locus signals. Mendelian randomization analyses suggested that blood traits may contribute to the underlying risk of VTE. To conclude, we identified 16 novel susceptibility loci for VTE; for some loci, the association signals are likely mediated through gene expression of nearby genes. (*Blood*. 2019;134(19):1645-1657)

Introduction

Venous thromboembolism (VTE) contributes significantly to morbidity in adults, occurring at rates of 1 to 2 events per 1000 person-years.¹ Despite improved prophylaxis in high-risk patients, neither the incidence of VTE nor mortality among VTE patients with pulmonary embolism (PE) has decreased in recent decades.^{1,2}

Many of the disease-causing mechanisms underlying VTE remain to be fully characterized, including those involving genetics. It has been >50 years since the first article linking genetic variation to the risk of VTE was published.³ With the availability of large population-based data sets and new technologies, including genome-wide association studies (GWASs), there has been a rapid accumulation of information to link genetic variation to VTE risk.⁴⁻⁹ To advance our understanding of the biology of VTE, which may lead to better disease risk prediction and prevention, we meta-analyzed VTE GWAS data from 18 studies. We also used tissue-specific gene expression data to impute predicted expression based on GWAS variants. With these imputed expressions, we conducted a transcriptome-wide association study (TWAS) of VTE.¹⁰ Additionally, we conducted colocalization analyses to assess the probability that the expression of genes near identified GWAS signals mediates the VTE associations observed in our GWASs.¹¹ Lastly, we applied Mendelian randomization (MR) analysis to better understand the role of traits that are genetically related to novel VTE loci in the development of VTE.

Methods

Setting, design, and participants

The setting for this multi-ancestry GWAS meta-analysis of VTE is the International Network Against Venous Thrombosis (INVENT) Consortium.⁷ The meta-analysis is based on prospective cohort and case-control data from 18 studies. Supplemental Table 1 (available on the *Blood* Web site) provides study-specific information for all included studies. Participants were adult men and women of European ancestry (EA) or African American ancestry (AA). Cases experienced a VTE, either deep vein thrombosis or PE. Supplemental Table 1 provides details about case definitions across the 18 studies. Supplemental Table 1 also provides information on the selection of controls for case-control and cohort analyses, when applicable. All participants provided informed consent for use of their genetic and health information for analysis.

Study-specific analyses, meta-analyses, and replication

Before transancestry meta-analysis, additional quality control of the study-specific GWAS summary statistics was conducted according to previously published protocols.¹²

Summary descriptions of study-specific genotyping arrays, quality control filtering, and imputation methods are shown in supplemental Table 1. Variant-specific analyses of the association between allele dosage and VTE risk were performed separately in each study using an additive genetic model and either logistic or Cox-proportional regression analyses adjusted for study-specific covariates and participant relatedness, when applicable (supplemental Table 1). Chromosome X variants located in the pseudo-autosomal regions were excluded. For each study, we also excluded variants with imputation quality score $r^2 < 0.3$ (< 0.7 for eMERGE),¹³ as well as variants with a minor allele count < 20 in either case or control counts. For the EA autosomal data, we also excluded variants with minor allele frequencies (MAFs) that differed between our data sets and the 1000 Genomes Project or the Haplotype Reference Consortium according to the following formula: $(p_1 - p_0)^2 / ((p_1 + p_0)(2 - p_1 - p_0)) > 0.007$,¹⁴ where p_0 and p_1 are the MAFs in the 1000 Genomes Project (or Haplotype Reference Consortium) and our data sets, respectively.

For autosomal chromosomes, meta-analysis was conducted using a fixed-effects inverse-variance weighting as implemented in METAL.¹⁵ For the X chromosome, we conducted sex-specific meta-analysis using fixed-effects inverse-variance weighting and then used 2 approaches to combine the results from men and women: fixed-effects inverse-variance weighted meta-analysis and sample-size weighted meta-analysis based on P values and effect direction.^{16,17} Across all GWAS analyses, $P < 5 \times 10^{-8}$ was considered genome-wide significant.^{18,19} We only included variants that were observed in ≥ 3 studies. We also conducted sensitivity analyses stratified by ancestry: EA and AA.

To search for additional independent genome-wide significant variants at a particular locus, we conducted conditional analyses using GCTA COJO software.^{20,21} For these analyses, a set of 9330 individuals from 4 studies with genotypes from the Illumina OmniExpress array imputed to 1000 Genomes Project was used to create a linkage disequilibrium (LD) reference panel.²² The conditional analyses were conducted using the EA results only.

For novel variants meeting genome-wide significance, we extracted information from published GWASs on 10 hemostasis

traits to characterize associations with intermediate phenotypes: fibrinogen,²³ fibrin D-dimer,²⁴ coagulation factor VII,²⁵ coagulation factor VIII (FVIII),²⁶ coagulation factor XI,²⁷ and von Willebrand factor (VWF),²⁶ tissue plasminogen activator,²⁸ plasminogen-activator inhibitor-1,²⁹ activated partial thromboplastin time (aPTT),³⁰ and prothrombin time.³⁰

Genome-wide significant findings were then replicated using VTE GWAS data from the Department of Veterans Affairs Million Veteran Program (MVP), which included 11 029 cases and 235 202 controls.³¹ For replication, we considered a $P < .05$ divided by the number of variants tested as statistically significant. In addition, the direction of the association was required to be the same in the discovery and replication data.

Previously reported variants associated with VTE

As a follow-up to the agnostic GWASs, we assessed the association of common (MAF > 0.01) candidate VTE variants previously reported with high-quality evidence in the literature but not reaching genome-wide significance in the GWASs. The level of significance was set as 0.05 divided by the number of tests.

Heritability and genetic risk score

We calculated the heritability on the observed scale due to genotyped variants as previously defined using nested case-control data on 6 184 VTE cases and 19 337 controls in UK Biobank.³² For these analyses, we only included genotyped variants and excluded those with MAF < 0.01 and Hardy-Weinberg equilibrium $P < .01$. We additionally excluded subjects so that no pair of individuals had a relatedness coefficient > 0.025, as calculated by the GCTA software.³³ Analyses were adjusted for age, sex, array, and top 10 principal components. We calculated the heritability on the liability scale assuming a disease prevalence of 0.5%.³⁴

We also calculated a genetic risk score (GRS) to summarize the risk for each individual according to the number of risk alleles carried.³⁵ The GRS was calculated in 6 573 VTE cases and 20 515 controls in UK Biobank; UK Biobank was selected for the GRS because the study was the largest and we had easy access to individual-level data, which are required for the GRS. Therefore, the GRS analysis was restricted to the EA individuals. To minimize overfitting, we reran the EA meta-analysis for the included variants excluding this set of UK Biobank cases and controls to obtain log odds ratio (β) estimates. For each individual j , the GRS was calculated as $GRS_j = \sum_{i=1}^k \beta_i G_{ij}$, where k is the number of selected variants, β_i is the log odds ratio for variant i based on the aforementioned meta-analysis, and G_{ij} is the number of risk alleles of variant i that individual j is carrying. The GRS was then standardized to have a mean of 0 and a variance of 1. We assessed the association between the GRS and VTE risk using logistic regression, adjusting for sex and age at the time of interview.

In silico bioinformatics, eQTL expression, TWASs, and colocalization analyses

We used HaploReg (version 4.1) to describe any predicted functional aspects of the lead variants in the newly identified GWAS regions.³⁶ We conducted a TWAS of VTE using 4 gene expression datasets: 1 from liver tissue (Genotype-Tissue

Expression [GTEx] version 7, $N = 153$) and 3 from whole blood (GTEx version 7, $N = 369$; Young Finns Study [YFS], $N = 1414$; and the Netherlands Twin Registry [NTR], $N = 1247$).^{10,37,38} Gene expression prediction models for YFS and NTR were downloaded from <http://gusevlab.org/projects/fusion/>. Using the 2 GTEx data sets as training data, we used the FUSION pipeline to construct gene expression prediction models based on each gene's local variants (as described previously).¹⁰ We only considered *cis* variants, those located within 500 kb of the gene of interest. For the YFS and NTR data sets, gene expression models were created using 5 approaches: single best expression quantitative trait locus (eQTL), elastic net, lasso, Bayesian sparse linear mixed model, and best linear unbiased predictor.¹⁰ For the 2 GTEx data sets, we limited the models to the single best eQTL, elastic net, and lasso for computational feasibility. We only included genes whose expression showed evidence of a significant heritable component ($P < .01$). In total, we successfully obtained gene expression prediction models for 15 982 genes across the 4 data sets; some genes are counted multiple times because we successfully generated gene expression imputation models for that gene in multiple reference data sets. These variant-based gene expression prediction models were then assessed in the EA VTE GWAS to estimate the association between imputed gene expression and VTE.¹⁰ We only included VTE results from the EA analysis because the method relies on LD patterns generated from the 1000 Genomes Project EA population. We considered associations with $P < 3.13 \times 10^{-6}$ (0.05/15 982 genes) statistically significant.

We also conducted colocalization analysis using the coloc pipeline.¹¹ Details are provided in the supplemental Methods. Briefly, coloc is a Bayesian approach to estimate the probability that 2 observed association signals (here GWAS and eQTL) in a given locus are consistent with a shared causal variant. Coloc calculates the posterior probability for 5 scenarios, 1 of which represents a shared causal variant modifying gene expression and phenotype. A high posterior probability for this scenario indicates a shared eQTL-VTE GWAS signal, suggesting that the VTE association is mediated by gene expression of the gene of interest.

MR analyses of selected traits and VTE

We reviewed the published literature to search for other traits associated with the newly identified VTE genetic loci, and these associations are reported. Then, to better understand the association of these traits with VTE, we conducted MR analysis of the associated traits and VTE. Identified associated traits needed to have GWAS summary results available to conduct the MR analysis. We obtained the effect allele, trait-specific association estimates, and standard errors for variants from the trait's GWAS summary results. Because we obtained GWAS summary results for the traits from UK Biobank, we reran the VTE GWAS excluding UK Biobank to avoid sample overlap. We then extracted VTE-specific effect estimates and P values from the EA VTE GWAS for each of the variants. We conducted 2-sample MR analyses to estimate the association between traits and VTE using summary genetic association statistics, as described previously.^{39,40} We also conducted sensitivity analysis using MR Egger and weighted median regression.^{41,42}

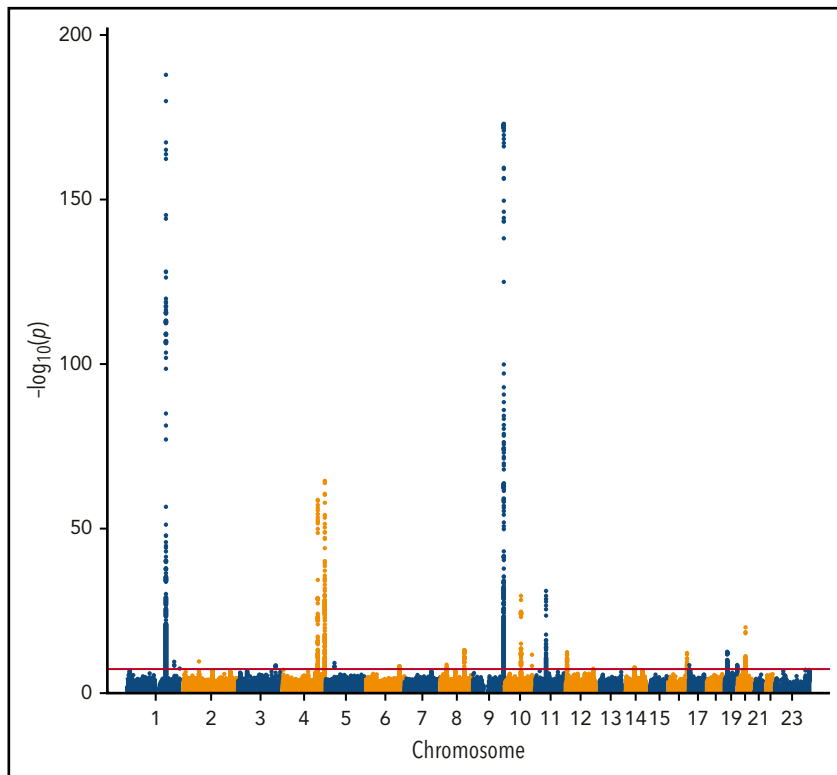


Figure 1. Manhattan plot of the association of 12 923 718 variants by chromosome and their association with VTE.

Results

GWAS genome-wide significant associations

We performed a transancestry meta-analysis based on 18 studies, which included 30 234 VTE cases (29 435 EA and 799 AA), 172 122 controls (157 769 EA and 14 353 AA), and 12 923 718 variants across 23 chromosomes. The genomic inflation factor for the meta-analysis was 1.08 (supplemental Figure 1), an elevation likely due to a polygenic signal (the intercept from LD score regression was 1.03).⁴³ We identified 33 independent signals across 24 genetic loci that reached genome-wide significance (Figure 1; Table 1). In addition, we performed ancestry-stratified analyses (supplemental Table 2) and identified 1 additional locus at 17p13.1, which was genome-wide significant in the EA meta-analysis but not in the transancestry meta-analysis (Table 1). Across the 34 independent genome-wide significant signals, 14 are newly reported associations; 11 are in newly associated loci, and 3 are in previously characterized genes. In “Novel VTE-associated Loci,” we describe the newly associated loci, the details of which are found in Table 1; supplemental Table 2 provides details of ancestry-stratified results. Supplemental Tables 3 and 4 provide ancestry-stratified results for all of the variants with P values $< 1.0 \times 10^{-6}$. Supplemental Table 5 presents replication results in MVP; given 11 novel associations, the threshold for replication was $0.05/11 = 0.0045$. Supplemental Table 6 presents associations of the novel top variants with the 10 hemostasis traits. Variant-specific regional plots are provided in supplemental Figure 2, and forest plots are shown in supplemental Figure 3.

Novel VTE-associated loci

Chr1.q42.2 We identified a novel association with an intronic variant (rs145470028) in open reading frame *C1orf198*

($P = 3.0 \times 10^{-8}$; odds ratio [OR], 0.78). The variant T allele was uncommon (MAF 0.02 in EA and absent in AA) and is located in an enhancer region marked with H3K4me1 and H3K27ac in liver tissue. The association did not replicate ($P = .6$) in the MVP data set and was not associated with any of the 10 hemostasis phenotypes ($P > .05$ for all associations).

Chr2.p14 We identified a novel association with rs1867312 ($P = 2.5 \times 10^{-10}$; OR, 1.06), intronic to *PLEK* and located in multiple histone marks in blood. The association was replicated ($P = 1.9 \times 10^{-4}$) and the minor allele (C) was associated with higher plasma FVII. Another *PLEK* variant (rs13011075) in moderate LD with rs1867312 ($r^2 = 0.3$) has been associated with mean corpuscular volume but not with our VTE variant.⁴⁴ Variant rs13011075 showed a nominal association with VTE (OR, 1.04; $P = 1.2 \times 10^{-4}$). The transcribed protein of *PLEK*, pleckstrin, is found in platelets and is involved with platelet biology.^{45,46}

Chr5.p13.1 We identified a common intronic variant (rs4869589) in *OSMR-AS1* ($P = 7.6 \times 10^{-10}$; OR, 1.07), the antisense gene to *OSMR*, the oncostatin M receptor. This variant is located in a region with enhancer histone marks H3K4me1 and H3K27ac in liver tissue. The association was replicated ($P = 2.4 \times 10^{-6}$) and was not associated with any of the 10 hemostasis traits. Variants in *OSMR* have been associated with amyloidosis, but there are no reported associations of its antisense gene with human traits.⁴⁷

Chr8.p21.1 We identified a common intronic variant (rs12675621) in *NUGGC* ($P = 2.4 \times 10^{-9}$; OR, 1.07); *NUGGC* (nuclear GTPase, germinal center associated) is not known to be associated with hemostasis or thrombosis phenotypes. The association was not replicated ($P = .62$), and its proxy's minor allele (rs1716157;

$r^2 = 0.97$, $D' = 0.98$) was associated with greater prothrombin time. The variant is 48 kb upstream from *SCARA5* (scavenger receptor class A member 5), in which the rs2726953 (intronic) variant has been associated with plasma levels of VWF but not previously with VTE.^{48,49} The LD between rs12675621 and rs2726953 is limited ($r^2 = 0.04$, $D' = 0.20$). The rs2726953 variant was associated with VTE ($P = 2.6 \times 10^{-5}$; OR, 1.05) in the current meta-analysis and appears to be an independent signal in regional association plots (supplemental Figure S2i).

Chr10.q26.11 We identified a novel association with the rs10886430 variant, intronic to *GRK5* ($P = 2.2 \times 10^{-12}$; OR, 1.12), a G protein-coupled receptor for kinase 5. The association replicated ($P = 6.7 \times 10^{-11}$), and the variant minor allele (G) was associated with higher D-dimer levels. The same variant has previously been associated with higher mean platelet volume ($P = 9.0 \times 10^{-18}$), wider mean platelet distribution width ($P = 3.1 \times 10^{-13}$), and lower platelet count ($P = 6.1 \times 10^{-9}$) (supplemental Table 7) and is located in multiple histone markers in liver tissue.⁴⁴

Chr12.q24.31 A novel association was found for a common intronic variant rs2851436 in *MPHOSPH9* (M-phase phosphoprotein 9; $P = 3.9 \times 10^{-8}$; OR, 0.94). This was the only trans-ancestry genome-wide significant signal that did not reach genome-wide significance in the EA population alone ($P = 2.1 \times 10^{-7}$, supplemental Table 2). The association replicated ($P = 1.0 \times 10^{-6}$), and the minor allele of its proxy (rs1716157; $r^2 = 1.0$, $D' = 1.0$) was associated with higher plasma tissue plasminogen activator levels. rs2851436 was associated with smaller platelet mass ($P = 2.4 \times 10^{-10}$) and lower platelet count ($P = 8.5 \times 10^{-10}$), as reported for variants in LD ($r^2 = 0.9$) with rs2851436; genetic variants in this region have previously been associated with hemoglobin levels.^{44,50} Colocalization analyses with eQTL data in blood and liver tissue revealed multiple candidate genes through which the GWAS signal could be mediated (supplemental Table 8). These include *ABCB9*, *ARL6IP4*, *MPHOSPH9*, and *SBNO1*, with the highest posterior probability ($PP = 0.96$) for *MPHOSPH9* (supplemental Table 9).

Chr14.q23.1 A novel association was found for a common intragenic variant, rs11158204 ($P = 1.5 \times 10^{-8}$; OR, 0.94), which was 4 kb downstream from *ARID4A* (AT-rich interaction domain 4A). The association fell short of replication at the Bonferroni-correction level ($P = 2.01 \times 10^{-2}$; effect direction same) and it (or its proxy rs12433823; $r^2 = 0.98$, $D' = 1.0$) was associated with lower fibrinogen, lower plasminogen-activator inhibitor 1, and aPTT plasma levels. This region has been associated with red blood cell distribution width ($P = 1.1 \times 10^{-11}$) and our lead variant is in moderate LD ($r^2 = 0.5$) with variants associated with coronary heart disease in a large GWAS meta-analysis.^{44,51} Colocalization analysis suggests that the observed VTE association may be mediated by variants that predict gene expression of *ARID4A* ($PP = 0.72$) and/or *RP11-349A22* ($PP = 0.69$) in blood and/or liver.

Chr16.q23.3 A relatively common intronic variant in *PLCG2*, rs12445050, was associated with the risk of VTE ($P = 6.7 \times 10^{-13}$; OR, 1.11). The association replicated ($P = 2.9 \times 10^{-15}$) and was associated with aPTT. Similar to the *GRK5* locus, the variant found to be associated with VTE has also been associated with larger mean platelet volume ($P = 7.2 \times 10^{-24}$).⁴⁴ The gene

product is a signaling enzyme, phospholipase $C\gamma_2$, and it has been shown to have a role in platelet activation biology.⁵²

Chr17.p13.3 rs1048483 ($P = 3.4 \times 10^{-9}$; OR, 1.06) is intronic to *SMG6* (nonsense-mediated messenger RNA decay factor) and is located in a conserved region and in enhancer histone mark H3K27ac in liver tissue. *SMG6* codes for a housekeeping protein. The association replicated ($P = 2.6 \times 10^{-5}$) in MVP and was not associated with any of the 10 hemostasis phenotypes. Variants in and close to *SMG6* have been associated with coronary heart disease (rs170041, $r^2 = 0.32$), as well as with platelet mass and count (rs4455005; $r^2 = 0.59$ between lead variants); rs1048483 was associated with lower platelet mass ($P = 2.0 \times 10^{-7}$) and smaller platelet count ($P = 4.9 \times 10^{-6}$) in UK Biobank data.^{44,51,53}

Chr17.p13.1 rs12450494 is intragenic and 2.4 kb upstream from *EIF5A* (eukaryotic translation initiation factor 5A) and was only associated with VTE in EA-only meta-analysis ($P = 2.5 \times 10^{-8}$; OR, 1.06). The association did not replicate ($P = .58$) and was not associated with any of the 10 hemostasis phenotypes. Colocalization analyses provided evidence that the rs12450494 VTE signal is mediated through *EIF5A* expression in liver and blood ($PP = 0.99$ for blood and liver), as well as for *GPS2* (G protein pathway suppressor 2) expression ($PP = 0.99$), a neighboring gene for which rs12450494 is situated downstream (supplemental Table 9).

Chr19.p13.2 rs7508633, an intronic variant in *STX10*, was associated with VTE risk ($P = 2.0 \times 10^{-9}$; OR, 0.94). The variant is located in a DNase I hypersensitive site and enhancer histone marks found in liver tissue and multiple blood-derived cells. The association did not replicate in MVP ($P = .12$; effect direction the same) and was not associated with the hemostasis traits. *STX10* is a SNARE protein; another SNARE protein (*STX2*) and a SNARE binding protein (*STXBP5*) have been associated with plasma levels of VWF and with VTE.^{48,49}

Novel VTE-associated variants at established loci We identified genome-wide significant associations at the *F5*, *C4BPA*, *FGG*, *F11*, *STXBP5*, *ZFPM2*, *ABO*, *TSPAN15*, *F2*, *VWF*, *SLC44A2*, *GP6*, *PROCR*, and *F8* loci (Table 1), all of which are genes known to be associated with VTE.^{4,5,7-9,49,54-62} For all loci with the exception of *C4BPA*, *F11*, and *VWF*, we identified only previously known signals at the established loci after conditional analyses. We did not find evidence for an association between the recently reported *VWF* variant rs35335161 with VTE ($P = .88$).⁶³ At the *C4BPA*, *F11*, and *VWF* loci, we also identified new independent variants associated with VTE in conditional analyses. The genome-wide significant variant rs2842700 in *C4BPA* was in modest LD ($r^2 = 0.01$; $D' = 1.0$) with the previously reported rs3813948 variant in the *C4BPA/C4BPB* locus.⁵⁶ The rs3813948 variant was not associated with VTE in our study ($P = .38$). No other independent variant at this locus was identified. For the *F11* locus, we identified 3 genome-wide significant variants (rs2289252, rs2036914, and rs4253421); all remained significant after conditioning on the 2 previously reported variants at the *F11* locus (rs2289252 and rs2036914).^{55,58,59} rs4253421 is intronic and was previously reported as being associated with *F11*, but not with VTE, phenotypes.²⁷ rs4253421 is in strong LD with rs1593 ($r^2 = 0.84$; $D' = 0.93$), which has been associated with stroke.⁶⁴ For the *VWF* locus, we observed a second genome-wide significant signal for rs216311

Table 1. Independent genome-wide significant loci associated with VTE from transancestry meta-analyses

Chr:Position	Variant identifier: effective/reference allele (amino acid substitution)	MAF: EA	MAF: AA	Transancestry P value	Transancestry OR (95% CI)	Locus (location)	Status
1:169519049	rs6025: T/C (Arg>Gln)	0.03	NA	1.4E-188	2.39 (2.25-2.53)	F5 (exon)	Known
1:169511755	rs4524: C/T (Lys>Arg)	0.26	0.19	3.1E-29	0.88 (0.86-0.90)	F5 (exon)	Known
1:207282149	rs2842700: A/C	0.06	0.28	3.1E-10	1.14 (1.09-1.18)	C4BPA (intron)	New variant
1:230986407	rs145470028: T/G	0.02	NA	3.0E-08	0.78 (0.72-0.85)	C1orf198 (intron)	New locus: not replicated
2:68619981	rs1867312: C/A	0.43	0.45	2.5E-10	1.06 (1.04-1.09)	PLEK (intron)	New locus: replicated
4:155525695	rs2066864: A/G	0.24	0.41	2.0E-59	1.20 (1.17-1.22)	FGG (intron)	Known
4:187207381	rs2289252: T/C	0.40	0.29	3.0E-65	1.19 (1.16-1.21)	F11 (intron)	Known
4:187192481	rs2036914: T/C	0.47	0.35	1.6E-54	0.86 (0.84-0.87)	F11 (intron)	Known
4:187204937	rs4253421: A/G	0.12	0.04	8.1E-41	0.80 (0.78-0.83)	F11 (intron)	New variant
5:38707871	rs4869589: T/G	0.26	0.27	7.6E-10	1.07 (1.05-1.10)	OSMR-AS1 (intron)	New locus: replicated
6:147701133	rs9373523: T/G	0.42	0.28	5.4E-09	0.94 (0.93-0.96)	STXBP5 (intron)	Known
8:27898452	rs12675621: G/A	0.29	0.35	2.4E-09	1.07 (1.04-1.09)	NUGGC (intron), SCARAS5 (upstream)	New locus: not replicated
8:106590705	rs4541868: A/C	0.25	0.39	7.1E-14	0.92 (0.90-0.94)	ZFPF2 (intron)	Known
9:136131188	rs8176749: T/C (Leu>Leu)	0.08	0.16	2.3E-30	1.23 (1.19-1.28)	ABO (exon)	Known
9:136137106	rs687289: A/G	0.35	0.41	1.3E-174	1.34 (1.31-1.36)	ABO (intron)	Known
9:136141870	rs2519093: T/C	0.21	0.11	3.8E-169	1.40 (1.37-1.43)	ABO (intron)	Known
9:136154168	rs579459: C/T	0.23	0.14	3.9E-145	1.36 (1.33-1.39)	ABO (intron)	Known
10:71245276	rs78707713: C/T	0.12	0.02	2.6E-30	0.83 (0.81-0.86)	TSPAN15 (intron)	Known
10:121010256	rs10886430: G/A	0.13	0.03	2.2E-12	1.12 (1.09-1.16)	GRK5 (intron)	New locus: replicated
11:465933311	rs191945075: A/G	0.01	NA	9.5E-32	1.86 (1.68-2.07)	LRP4 (intron), F2 (downstream)	Known
11:46760756	rs3136516: G/A	0.47	0.12	1.0E-15	1.08 (1.06-1.11)	F2 (intron)	Known

Bold text indicates novel associations.

CI, confidence interval; OR, odds ratio.

*For rs216311, the EA meta-analysis P value was 5.3E-09 (OR 0.94) after conditioning on rs1558519, the strongest signal in WWF.

†For rs12450494, the EA-only P value was 2.5E-08 (OR 1.06), which qualified it for genome-wide significance.

‡For rs143478537, meta-analysis was by done using a sample-size weighted model in EA only so no odds ratio and MAF in AA was estimated.

Table 1. (continued)

Chr:Position	Variant identifier: effective/reference allele (amino acid substitution)	MAF: EA	MAF: AA	Transancestry P value	Transancestry OR (95% CI)	Locus (location)	Status
12:6153738	rs1558519: G/A	0.35	0.44	3.7E-13	1.08 (1.06-1.10)	VWF (intron)	Known
12:6128443	rs216311: T/C (Thr>Ala)	0.39	0.12	1.8E-02*	0.98 (0.96-1.00)	VWF (exon)	New variant
12:123667354	rs2851436: G/T	0.21	0.27	3.9E-08	0.94 (0.91-0.96)	MPHOSPH9 (intrinsic)	New locus: replicated
14:58844526	rs11158204: T/C	0.25	0.25	1.5E-08	0.94 (0.92-0.96)	Intragenic, ARID4A (downstream)	New locus: not replicated
16:81870969	rs12445050: T/C	0.14	0.03	6.7E-13	1.11 (1.08-1.14)	PLCG2 (intron)	New locus: replicated
17:1966457	rs1048483: T/C	0.49	0.30	3.4E-09	1.06 (1.04-1.08)	SMG6 (intron)	New locus: replicated
17:7207887	rs12450494: A/T	0.47	0.18	1.0E-07†	1.06 (1.03-1.08)	Intergenic, EIF5A (upstream)	New locus: not replicated
19:10740871	rs4548995: G/C	0.22	0.16	3.7E-19	0.90 (0.87-0.92)	SCL44A2 (intron)	Known
19:13258290	rs7508633: G/A	0.45	0.32	2.0E-09	1.06 (1.04-1.09)	STX10 (intron)	New locus: not replicated
19:55511873	rs1671135: G/C	0.21	0.23	2.9E-09	0.93 (0.90-0.95)	NLRP2 (intron), GP6 (downstream)	Known
20:33745676	rs6088735: T/C	0.23	0.32	3.0E-10	1.08 (1.05-1.10)	PROCR (intron)	Known
20:33764554	rs867186: G/A (Ser>Gly)	0.10	0.09	6.8E-09	1.11 (1.07-1.14)	PROCR (exon)	Known
X:154424170	rs143478537: G/C	0.14	NA	3.1E-08‡	1.08 (1.05-1.11)	F8 (upstream)	Known

Bold text indicates novel associations.

CI, confidence interval; OR, odds ratio.

*For rs216311, the EA meta-analysis P value was 5.3E-09 (OR 0.94) after conditioning on rs1558519, the strongest signal in VWF.

†For rs12450494, the EA-only P value was 2.5E-08 (OR 1.06), which qualified it for genome-wide significance.

‡For rs143478537, meta-analysis was by done using a sample-size weighted model in EA only so no odds ratio and MAF in AA was estimated.

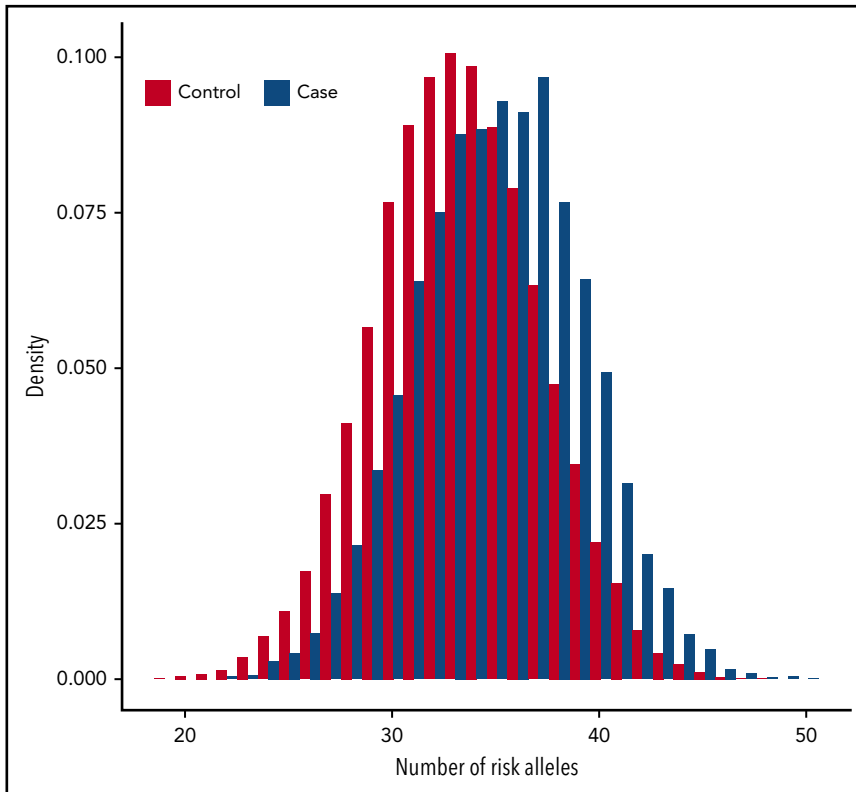


Figure 2. GRS distribution for VTE cases and controls.

($P = 5.9 \times 10^{-9}$) after conditioning on the strongest signal (rs1558519, $P = 1.8 \times 10^{-13}$), which is in LD with rs1063856 ($r^2 = 0.99$; $D' = 1.0$), a missense variant (Thr > Ala) previously associated with levels of VWF and FVIII, as well as with VTE.^{48,49} Rs216311 leads to a change in protein transcription (Thr > Ala) and represents a new VTE association within VWF. However, rs216311 was not associated with plasma VWF levels ($P = .4$) in a recently published GWAS, but it was associated with plasma levels of FVIII ($P = .03$), suggesting that it may influence the ability of the VWF molecule to carry FVIII.²⁶

At the *ABO*, *TSPAN15*, *SLC44A2*, and *PROCR* loci, we found evidence of colocalization between GWAS and eQTL signals (supplemental Table 9). For *ABO*, the GWAS signal showed high posterior probability ($PP = 0.93$) of colocalizing with eQTL signals for *ABO* expression in liver. For the *TSPAN15* region, we observed evidence of colocalization ($PP = 0.88$) with the eQTL signal for *TSPAN15* expression in blood. We observed evidence of colocalization between the GWAS signal in the *SLC44A2* region and eQTL signals for *SLC44A2* expression ($PP = 0.99$) and *ILF3-AS1* expression ($PP = 0.98$), all in blood. For the *PROCR* region, we observed the strongest posterior probability ($PP = 0.93$) of colocalization between the GWAS signal and eQTL for *PROCR* expression in blood. Evidence of colocalization of an eQTL and GWAS signal was not found for *F5*, *FGG*, *F11*, *STXBP5*, *ZFPM2*, *F2*, *VWF*, *GP6*, or *F8* loci.

Previously identified variants not reaching genome-wide significance

We tested 9 variants in genes that have previously been reported to be associated with VTE (supplemental Table 10) but that did not reach genome-wide significance in our primary analyses:

SERPINC1, *KNG1*, *F13A1*, *HIVEP1*, *PAI-1*, *FGA*, *TCN2*, *F9*, and *THBD*.⁶⁵⁻⁷³ The threshold for significance was set to 5.6×10^{-3} (0.05/9), and 3 variants reached the significance threshold: *KNG1* rs1648711 (OR, 1.05 for minor allele; $P = 4.5 \times 10^{-4}$), *TCN2* rs1884841 (OR, 1.05; $P = 8.8 \times 10^{-6}$), and *F9* rs6048 (OR, 0.94 for minor allele; $P = 7.8 \times 10^{-8}$). A fourth variant (*FGA* rs6050) reached significance in the GWAS but was no longer significant ($P = .49$) in analyses conditioned on *FGA* rs2066864.

Previously known loci and variants

In addition to the novel associations, single variant and conditional meta-analysis provided evidence of association for known variants (Table 1; supplemental Results): *F5* rs6025 and rs4524^{54,55}; *FGG* rs2066864⁵⁷; *STXBP5* rs1039084 (strong LD with rs9390460)⁴⁹; *ZFPM2* rs4541868 (exact proxy for rs4541868)^{7,8}; *ABO* variants rs8176749, rs687289, rs2519093, and rs579459^{5,60}; *TSPAN15* rs7870713⁷; *F2* variants rs191945075 (LD with rs1799963) and rs3136516⁸; *SLC44A2* rs4548995 (strong LD with rs2288904)⁷; *NLRP2* rs1671135 (strong LD with *GP6* rs1613662)⁴; *PROCR* variants rs867186 and rs6088735^{61,62}; and *F8* rs143478537 (moderate LD with rs114209171).⁹

Heritability and genetic risk score

The heritability on the observed scale due to genotyped variants was 0.23 (standard error 0.02). Transforming this to the liability scale assuming a disease prevalence of 0.5% resulted in a heritability of 0.15 (standard error 0.01).

In total, the GRS included 37 variants (supplemental Table 11). Figure 2 shows the distribution of risk alleles for cases and controls. Relative to individuals with scores in the interquartile range (25% to 75% percentile), those in all other categories had

VTE risks that were statistically different (Table 2). At the extremes, individuals at and below the bottom fifth percentile had a 50% lower risk for VTE (OR, 0.51; 95% confidence interval [CI], 0.43-0.61) and those at or above the top fifth percentile had a 3.2-fold greater risk for VTE (OR: 3.19; 95% CI: 2.89-3.52) relative to half of the population in the middle of the range.

Novel associations from transcriptome-wide association analyses of VTE

Transcriptome-wide association analyses identified 4 genes not previously associated with VTE: *SPSB1*, *ERAP1*, *RP11-747H7.3*, and *RP4-737E23.2*, and replicated *SH2B3* (Table 3). *SPSB1* genetic variation has previously been associated with eosinophil counts, as well as with neutrophil and eosinophil percentages of granulocytes.⁴⁴ *ERAP1* has been associated with Behcet's syndrome, psoriasis, and ankylosing spondylitis, inflammatory and autoimmune conditions that are purported risk factors for VTE.⁷⁴⁻⁷⁹ *RP11-747H7.3* and *RP4-737E23.2* are long noncoding RNAs. *RP11-747H7.3* genetic variation has been associated with mean platelet volume.⁴⁴ We replicated associations with *SH2B3*, previously hypothesized to be associated with VTE.⁸⁰ *SH2B3* plays a critical role in hematopoiesis, and *SH2B3* has been associated with platelet density and count, lymphocyte count, reticulocyte count, hemoglobin concentration, hematocrit, leukocyte count, erythrocyte count, granulocyte count, eosinophil and neutrophil counts, monocyte counts,⁴⁴ coronary artery disease,⁵¹ stroke,⁸¹ and autoimmune diseases.⁸²

MR of blood cell phenotypes and VTE

In our GWAS and TWAS findings, we observed that many of the novel loci associated with VTE have also been associated with blood cell traits. Therefore, we conducted MR analysis of VTE and 7 blood cell traits: lymphocyte counts, red blood cell volume and distribution width, and platelet volume, mass, distribution width, and count. We extracted the genome-wide significant variants associated with each of the blood cell traits from the report by Astle and colleagues, conducted using UK Biobank data.⁴⁴ For VTE GWAS summary data, we excluded UK Biobank from our VTE meta-analysis to avoid sample overlap. We observed high heterogeneity in the inverse-variance-weighted MR analysis, which can imply pleiotropy; therefore, we performed MR Egger as a sensitivity analysis. Across the MR approaches, there was some evidence that blood traits, particularly red blood cell and platelet distribution widths, were causally associated with VTE risk (supplemental Table 12). Although MR Egger regression can provide unbiased estimates in the presence of directional pleiotropy, it relies on the InSIDE assumption, that the instrument strength is independent of the direct effect. Therefore, we conducted an additional sensitivity analysis using the weighted-median MR, which is not as sensitive as MR Egger regression to violation of the InSIDE assumption but requires that at least 50% of the genetic variants are valid instrumental variables.⁴² In contrast to the MR Egger regression, we observed no evidence that platelet mass and count are associated with VTE; however, we observed suggested associations for platelet and red cell distribution width (both $P < .05$). We note that the estimates for platelet and red blood cell distribution width were consistent across MR methods, adding support for the association. Although platelet distribution width was associated with an increased risk for VTE, red blood cell distribution width was associated with a decreased risk, an unexpected result given previous observations studies.⁸³⁻⁸⁵ However, given the

Table 2. Genetic risk score

GRS percentile	OR (95% CI)
≤Fifth	0.51 (0.43- 0.61)
5-15th	0.63 (0.56-0.72)
15-25th	0.73 (0.65-0.81)
25-75th	1.00 (reference)
75-85th	1.39 (1.27-1.52)
85-95th	1.97 (1.81-2.14)
≥95th	3.19 (2.89-3.52)

inconsistent results from the range of MR approaches that we applied, our results should be interpreted with caution.

Discussion

In transancestry GWAS meta-analysis data from 18 studies, combined with TWASs and colocalization analyses, we identified 16 loci newly associated or confirmatory with the VTE phenotype: 11 from GWAS meta-analyses and 5 from TWAS analyses. Among the 11 novel GWAS findings, 6 replicated in an independent data set; 3 of 11 had strong evidence of colocalized GWAS genetic and transcription signals associated with the VTE phenotype. The 37-variant GRS identified risk across the range of scores. MR investigations suggested that blood traits may contribute to the underlying biology of VTE risk.

Biologic insights

A theme from the discovery was that many newly associated loci were also associated with blood traits, particularly platelets. *GP6*,⁴ *ZFPM2*,⁸⁶ *VWF*, and *ABO* loci have been associated with platelet biology and with VTE risk. Discovery findings produced a list of new candidates, including *PLEK*, *GRK5*, *MPHOSPH9*, *PLCG2*, *SMG6*, and *SH2B3*. Further, *PLEK*, *MPHOSPH9*, and *ARID4A* were associated with red blood cell traits. MR analyses suggested that platelet and red blood cell traits have an etiologic role in VTE; however, the inconsistency in MR results across various statistical approaches indicates that the relationship between blood traits and VTE is likely complex and that more research is needed to clarify the role of individual blood cell traits. Although the role of coagulation in VTE is very well characterized, the roles of platelets and red blood cells in thrombosis are less well characterized, and research continues to provide new insights into biologic mechanisms.^{87,88}

Most of the 11 newly identified signals were intragenic, and the highest-signal variants were intronic. For 3 of the loci (12.q24.31 [*MPHOSPH9*], 14.q23.1 [*ARID4A*], and 17.p13.1 [*EIF5A*]), we found evidence that the identified variants were influencing VTE risk through gene expression in blood and liver tissue. The role of gene-specific expression in VTE risk was further supported by our agnostic TWAS analysis identifying 4 novel candidate genes whose imputed gene expression was associated with VTE risk. In addition, several of the lead variants in the newly identified loci were located in regulatory regions, as identified in blood and liver tissue.

Table 3. TWAS findings meeting genome-wide significance

Tissue	Gene	Chr	Heritability of the gene	No. of SNPs*	Best GWAS <i>P</i> value†	TWAS Z-score	TWAS <i>P</i> value	PP‡
GTEx blood	<i>SPSB1</i>	1	0.25	3	1.99×10^{-7}	5.20	2.03×10^{-7}	0.997
GTEx liver	<i>ERAP1</i>	5	0.50	11	7.82×10^{-6}	-4.90	9.47×10^{-7}	0.349
YFS blood	<i>SH2B3</i>	12	0.046	22	2.87×10^{-6}	5.15	2.67×10^{-7}	0.158
GTEx blood	<i>RP11-747H7.3</i>	14	0.07	13	5.96×10^{-8}	4.88	1.07×10^{-6}	0.945
GTEx blood	<i>RP4-737E23.2</i>	20	0.19	2	2.10×10^{-7}	-5.27	1.34×10^{-7}	0.995

Level of significance: $P = 3.13 \times 10^{-6}$.

Chr, chromosome; SNP, single nucleotide polymorphism.

*Number of SNPs included in the final gene prediction model.

†The lowest SNP-specific *P* value in the region.

‡PP of colocalization between GWAS and eQTL signal.

Clinical significance

Although genetic predictors of VTE have been known since the 1960s, consensus on the clinical and economic utility of assessing genetic risks has not been reached.⁸⁹ In this article, we expand the number of independent variants associated with VTE from 23 to 37, 32 of which have strong replication evidence. We found limited statistical evidence that the 11 novel VTE variants were associated with plasma levels of the 10 hemostasis traits, suggesting that the majority of these variants affect VTE risk through other mechanisms or pathways yet to be characterized. A GRS constructed from these variants identified a risk gradient of risk alleles that could be tested for utility in the setting of shortening (those in the lowest fifth percentile of the GRS distribution were at a 50% decreased risk for VTE relative to those with mid-range GRSs) or prolonging (those in the highest fifth percentile of the GRS distribution were at threefold increased risk for VTE relative to those with mid-range GRSs) prophylaxis for incident VTE events. A better understanding of the biologic pathways through which individual or aggregated (GRS) variants contribute to risk is needed, but evidence from this study suggest that little of any additional risk or protection from the novel VTE variants contributing to the GRS are mediated by the 10 hemostasis traits interrogated. Given the roles of anticoagulation therapy and, to some extent, antiplatelet therapy in reducing the risk for incident and recurrent VTE, findings from this GWAS meta-analysis, especially those involved in platelet biology, may eventually be used to identify those who are at highest risk for an event, incident or otherwise, and provide guidance for the most effective treatment strategy.^{90,91}

Strengths and limitations

The present study constitutes the largest GWAS of VTE to date and the first transancestry GWAS meta-analysis of VTE. The meta-analysis approach performs equally well in terms of efficiency compared with a combined (mega) analysis approach.⁹² Although we had good statistical power for common variants with modest effect sizes, uncommon and familial rare variants (eg, those in *PROS1* and *PROC*), even those with large effect sizes, are not likely to be detected by GWAS, especially if the variants are poorly imputed. However, we were able to detect a variant with a MAF of 0.01 and effect size $\geq 30\%$ with 82% power. We assessed the association of previously reported

VTE variants in our GWAS (supplemental Table 10) and found that some previously reported variants were not associated with VTE ($P > .1$) in our data. We had modest numbers of AA participants, which limited our power to identify associations unique to them and to test differences across ancestry. Other investigators have attempted to identify variants unique to the AA population.⁹³ Further, *P* values were generally higher in EA-only meta-analyses than in combined EA and AA analyses, suggesting that differential LD with any causative variant may have diluted some genetic signals in the transancestry analyses. Although all 11 newly associated loci exhibit genome-wide significant signals, replication in an independent study was only successful for 6; the remaining 5 were not replicated because of the low statistical power given MAF and effect size (range of power from 48% to 72% for α 0.0045) or as a result of a false-positive GWAS signal. Adjusting for winner's curse of the original point estimate would further diminish power. For conditional analyses, we relied on a genotype-imputed data set of >9000 individuals from 4 cohorts, and it is possible that we might have missed additional independent signals as a result of mismatching LD patterns or incomplete coverage of the genetic variation in the reference population. However, variants in strong LD with one another can each be causally related to the outcome. To our knowledge, this is the first attempt to apply TWASs to VTE. Using transcription information to inform discovery is particularly powerful when many variants in a region contribute to disease risk, but no single variant has an association strong enough to reach genome-wide significance by itself. The TWAS approach aims to detect situations in which gene expression and phenotype are associated, which may not reflect the complete underlying biology. For example, in the case in which the variant leads to a change in the protein structure, such as *F5* rs6025, the variant is not expected to be linked to expression. For 2 of the TWAS signals (*ERAP1* and *SH2B3*), we did not observe strong evidence for a colocalization between the eQTL and GWAS signal. Both of these regions only showed moderate single-variant associations with VTE (lowest single-variant *P* values were $P = 7.8 \times 10^{-6}$ for *ERAP1* and $P = 2.9 \times 10^{-6}$ for *SH2B3*). Thus, these results likely reflect a situation in which there are strong eQTL signals with multiple variants/loci associated with gene expression, each with a small effect on VTE.

Conclusions

We identified 34 independent genetic signals for VTE risk from GWAS meta-analysis, of which 14 are newly reported associations. This included 11 newly associated genetic loci and 3 new independent signals in known genes. Further, TWAS identified 5 additional genetic loci. In some of the loci, we found that the association signal for novel and previously known regions is likely mediated by gene expression of nearby genes. We further show that a GRS based on 37 VTE-susceptibility variants can identify a subset in the population that is at particularly high risk for developing VTE. The MR analyses suggested that blood traits may contribute to the underlying biology of VTE risk.

Acknowledgments

The INVENT Consortium thanks all of the participants across the 18 studies who provided their health information to support these analyses.

This work was supported in part by National Institutes of Health, National Heart, Lung, and Blood Institute grants HL134894, HL139553, and HL141791 (N.L.S.). The K.G. Jepsen Thrombosis Research and Expertise Center is supported by an independent grant from Stiftelsen Kristian Gerhard Jepsen (J.H.) and HL116854 (C. Kabrhel). Sources of funding for the specific cohorts can be found in the supplemental Materials.

The views expressed in this article are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health, or the U.S. Department of Health and Human Services.

Authorship

Contribution: S.L., E.N.S., A.v.H.V., M.d.A., D.I.C., P.S., N.P., M.K.P., S.R.H., S.M.D., P.N., D.K., P.S.d.V., M.S.-L., J.E.H., K.A.F., J.S.P., A.P.R., M.E.G., C.J.O., J.-B.H., F.R.R., J.A.H., B.M.P., W.T., C. Kabrhel, K.H., P.M.R., P.-E.M., A.D.J., C. Kooperberg, D.-A.T., and N.L.S. designed the study; S.L., L.W., W.G., J.A.B., J.W.P., J.H., B.M.B., M.-H.C., C.T., M.G., K.L.W., J.M., S.K.B., S.M.A., S.M.D., P.N., D.K., P.S.d.V., M.S.-L., J.E.H., T.K.B., B.M.M., K.T., J.-F.D., B.M., P.K., D.-A.T., and N.L.S. designed and/or performed analyses; S.L., P.-E.M., A.D.J., C. Kabrhel, D.-A.T., and N.L.S. wrote 1 or more parts of the manuscript; and all authors reviewed and provided comments on the work.

Conflict-of-interest disclosures: B.M.P. serves on the Steering Committee of the Yale Open Data Access project funded by Johnson & Johnson.

REFERENCES

- Silverstein MD, Heit JA, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ III. Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. *Arch Intern Med*. 1998;158(6):585-593.
- Heit JA, Silverstein MD, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ III. Predictors of survival after deep vein thrombosis and pulmonary embolism: a population-based, cohort study. *Arch Intern Med*. 1999;159(5):445-453.
- Egeberg O. Thrombophilia caused by inheritable deficiency of blood anti-thrombin. *Scand J Clin Lab Invest*. 1965; 17(1):92.
- Tréguët DA, Heath S, Saut N, et al. 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(21):5298-5303.
- 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(8):1521-1531.
- Tang W, Teichert M, Chasman DI, et al. A genome-wide association study for venous thromboembolism: the extended cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Genet Epidemiol*. 2013;37(5):512-521.
- Germain M, Chasman DI, de Haan H, et al; Cardiogenics Consortium. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am J Hum Genet*. 2015; 96(4):532-542.
- Klarin D, Emdin CA, Natarajan P, Conrad MF, Kathiresan S; INVENT Consortium. Genetic analysis of venous thromboembolism in UK Biobank identifies the ZFPM2 locus and implicates obesity as a causal risk factor. *Circ Cardiovasc Genet*. 2017;10(2):e001643.
- Hinds DA, Buil A, Ziemek D, et al; META-STROKE Consortium, INVENT Consortium. Genome-wide association analysis of self-reported events in 6135 individuals and 252,827 controls identifies 8 loci associated with thrombosis. *Hum Mol Genet*. 2016;25(9):1867-1874.
- Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016; 48(3):245-252.
- Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383.
- Winkler TW, Day FR, Croteau-Chonka DC, et al; Genetic Investigation of Anthropometric Traits (GIANT) Consortium. Quality control

P.N. has received grant support from Amgen, Apple, and Boston Scientific and is a scientific advisor to Apple. S.M.D. receives institutional funding from CytoVas LLC and RenalytixAI. The remaining authors declare no competing financial interests.

D.K., Emma Busenkell, Renae Judy, Julie Lynch, Michael Levin, J.H., Krishna Aragam, Mark Chaffin, Mary Haas, S.L., Themistocles L. Assimes, Jie Huang, Kyung Min Lee, Qing Shao, J.E.H., C. Kabrhel, Yunfeng Huang, Yan V. Sun, Marijana Vujkovic, Danish Saleheen, Donald R. Miller, Peter Reaven, Scott DuVall, William Boden, Saiju Pyarajan, A.P.R., D.-A.T., Peter Henke, C. Kooperberg, J. Michael Gaziano, John Concato, Daniel J. Rader, Kelly Cho, Kyong-Mi Chang, Peter W. F. Wilson, N.L.S., C.J.O., Philip S. Tsao, Sekar Kathiresan, Andrea Obi, S.M.D., and P.N. represent the Million Veteran Program. The INVENT Consortium consists of all the authors of the present paper.

ORCID profiles: M.d.A., 0000-0003-2329-2686; J.A.B., 0000-0001-8509-148X; B.M.B., 0000-0002-3058-1059; D.I.C., 0000-0003-3357-0862; C.T., 0000-0002-7790-8311; K.L.W., 0000-0003-2749-1279; J.M., 0000-0002-7328-7626; J.B.N., 0000-0002-6654-2852; P.N., 0000-0001-8402-7435; M.S.-L., 0000-0002-0128-379X; J.E.H., 0000-0002-9672-2491; J.S.P., 0000-0001-7076-483X; J.K., 0000-0001-7096-820X; F.R.R., 0000-0003-2558-7496; J.A.H., 0000-0002-5509-1953; P.M.R., 0000-0003-1249-4522; C. Kabrhel, 0000-0002-8699-7176.

Correspondence: Nicholas L. Smith, Cardiovascular Health Research Unit, University of Washington, 1730 Minor Ave, Suite 1360, Seattle WA 98101; e-mail: nlsmith@u.washington.edu.

Footnotes

Submitted 3 March 2019; accepted 17 July 2019. Prepublished online as *Blood* First Edition paper, 16 August 2019; DOI 10.1182/blood.2019000435.

Please contact Nicholas L. Smith (nlsmith@u.washington.edu) or David-Alexandre Tréguët (david-alexandre.tregouet@u-bordeaux.fr) for information on INVENT Consortium data sharing.

The online version of this article contains a data supplement.

There is a *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

- and conduct of genome-wide association meta-analyses. *Nat Protoc.* 2014;9(5):1192-1212.
13. Verma SS, de Andrade M, Tromp G, et al. Imputation and quality control steps for combining multiple genome-wide datasets. *Front Genet.* 2014;5:370.
 14. Michailidou K, Lindström S, Dennis J, et al; ConFab/AOCS Investigators. Association analysis identifies 65 new breast cancer risk loci. *Nature.* 2017;551(7678):92-94.
 15. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.
 16. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet.* 2008;17(R2):R122-R128.
 17. Magi R, Lindgren CM, Morris AP. Meta-analysis of sex-specific genome-wide association studies. *Genet Epidemiol.* 2010;34(8):846-853.
 18. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol.* 2008;32(3):227-234.
 19. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol.* 2008;32(4):381-385.
 20. Yang J, Lee SH, Goddard ME, Visscher PM. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol.* 2013;1019:215-236.
 21. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet.* 2012;44(4):369-375, S1-3.
 22. Lindström S, Loomis S, Turman C, et al. A comprehensive survey of genetic variation in 20,691 subjects from four large cohorts. *PLoS One.* 2017;12(3):e0173997.
 23. de Vries PS, Chasman DI, Sabater-Lleal M, et al. A meta-analysis of 120 246 individuals identifies 18 new loci for fibrinogen concentration. *Hum Mol Genet.* 2016;25(2):358-370.
 24. Smith NL, Huffman JE, Strachan DP, et al. Genetic predictors of fibrin D-dimer levels in healthy adults. *Circulation.* 2011;123(17):1864-1872.
 25. de Vries PS, Sabater-Lleal M, Huffman JE, et al; MEGASTROKE Consortium of the International Stroke Genetics Consortium. A genome-wide association study identifies new loci for factor VII and implicates factor VII in ischemic stroke etiology. *Blood.* 2019;133(9):967-977.
 26. Sabater-Lleal M, Huffman JE, de Vries PS, et al. Genome-wide association trans-ethnic meta-analyses identifies novel associations regulating coagulation factor VIII and von Willebrand factor plasma levels. *Circulation.* 2019;139(5):620-635.
 27. Sennblad B, Basu S, Mazur J, et al. Genome-wide association study with additional genetic and post-transcriptional analyses reveals novel regulators of plasma factor XI levels. *Hum Mol Genet.* 2017;26(3):637-649.
 28. Huang J, Huffman JE, Yamakuchi M, et al; CHARGE Consortium Hemostatic Factor Working Group. Genome-wide association study for circulating tissue plasminogen activator levels and functional follow-up implicates endothelial STXBP5 and STX2 [published correction appears in *Arterioscler Thromb Vasc Biol.* 2014;34(8):E19]. *Arterioscler Thromb Vasc Biol.* 2014;34(5):1093-1101.
 29. Huang J, Sabater-Lleal M, Asselbergs FW, et al; CARDIOGENICS Consortium. Genome-wide association study for circulating levels of PAI-1 provides novel insights into its regulation. *Blood.* 2012;120(24):4873-4881.
 30. Tang W, Schwienbacher C, Lopez LM, et al. Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease. *Am J Hum Genet.* 2012;91(1):152-162.
 31. Klarin D, Busenkell E, Judy R, et al. Genetic analysis of venous thromboembolism highlights the genetic overlap of thrombosis with lipids and arterial vascular disease. *Nat Genet.* In press.
 32. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010;42(7):565-569.
 33. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76-82.
 34. Zaitlen N, Kraft P. Heritability in the genome-wide association era. *Hum Genet.* 2012;131(10):1655-1664.
 35. Dudbridge F. Power and predictive accuracy of polygenic risk scores [published correction appears in *PLoS Genet.* 2013;9(4)]. *PLoS Genet.* 2013;9(3):e1003348.
 36. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012;40(Database issue):D930-D934.
 37. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585.
 38. Wright FA, Sullivan PF, Brooks AI, et al. Heritability and genomics of gene expression in peripheral blood. *Nat Genet.* 2014;46(5):430-437.
 39. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* 2013;37(7):658-665.
 40. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG; EPIC-InterAct Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol.* 2015;30(7):543-552.
 41. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol.* 2015;44(2):512-525.
 42. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol.* 2016;40(4):304-314.
 43. Bulik-Sullivan BK, Loh PR, Finucane HK, et al; Schizophrenia Working Group of the Psychiatric Genomics Consortium. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-295.
 44. Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell.* 2016;167(5):1415-1429.e19.
 45. Coppinger JA, Cagney G, Toomey S, et al. Characterization of the proteins released from activated platelets leads to localization of novel platelet proteins in human atherosclerotic lesions. *Blood.* 2004;103(6):2096-2104.
 46. Fröbel J, Cadeddu RP, Hartwig S, et al. Platelet proteome analysis reveals integrin-dependent aggregation defects in patients with myelodysplastic syndromes. *Mol Cell Proteomics.* 2013;12(5):1272-1280.
 47. Arita K, South AP, Hans-Filho G, et al. Oncostatin M receptor-beta mutations underlie familial primary localized cutaneous amyloidosis. *Am J Hum Genet.* 2008;82(1):73-80.
 48. Smith NL, Chen MH, Dehghan A, et al; Wellcome Trust Case Control Consortium. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation.* 2010;121(12):1382-1392.
 49. Smith NL, Rice KM, Bovill EG, et al. Genetic variation associated with plasma von Willebrand factor levels and the risk of incident venous thrombosis. *Blood.* 2011;117(22):6007-6011.
 50. Danjou F, Zoledziwska M, Sidore C, et al. Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels. *Nat Genet.* 2015;47(11):1264-1271.
 51. van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res.* 2018;122(3):433-443.
 52. Yen TL, Lu WJ, Lien LM, et al. Amarogentin, a secoiridoid glycoside, abrogates platelet activation through PLC γ 2-PKC and MAPK pathways. *BioMed Res Int.* 2014;2014:728019.
 53. Eicher JD, Chami N, Kacprowski T, et al; Myocardial Infarction Genetics Consortium. Platelet-related variants identified by exomechip meta-analysis in 157,293 individuals. *Am J Hum Genet.* 2016;99(1):40-55.
 54. Bertina RM, Koeleman BP, Koster T, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature.* 1994;369(6475):64-67.
 55. Smith NL, Hindorf LA, Heckbert SR, et al. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *JAMA.* 2007;297(5):489-498.

56. Buil A, Tréguët DA, Souto JC, et al. C4BPB/C4BPA is a new susceptibility locus for venous thrombosis with unknown protein S-independent mechanism: results from genome-wide association and gene expression analyses followed by case-control studies. *Blood*. 2010;115(23):4644-4650.
57. 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(13):4176-4183.
58. Bezemer ID, Bare LA, Doggen CJ, et al. Gene variants associated with deep vein thrombosis. *JAMA*. 2008;299(11):1306-1314.
59. Li Y, Bezemer ID, Rowland CM, et al. Genetic variants associated with deep vein thrombosis: the F11 locus. *J Thromb Haemost*. 2009;7(11):1802-1808.
60. Jick H, Slone D, Westerholm B, et al. Venous thromboembolic disease and ABO blood type. A cooperative study. *Lancet*. 1969;1(7594):539-542.
61. Dennis J, Johnson CY, Adediran AS, et al. 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(10):2392-2400.
62. Germain M, Saut N, Greliche N, et al. Genetics of venous thrombosis: insights from a new genome wide association study. *PLoS One*. 2011;6(9):e25581.
63. Schneppenheim R, Hellermann N, Brehm MA, et al. The von Willebrand factor Tyr2561 allele is a gain-of-function variant and a risk factor for early myocardial infarction. *Blood*. 2019;133(4):356-365.
64. Hanson E, Nilsson S, Jood K, et al. Genetic variants of coagulation factor XI show association with ischemic stroke up to 70 years of age. *PLoS One*. 2013;8(9):e75286.
65. Austin H, De Staercke C, Lally C, Bezemer ID, Rosendaal FR, Hooper WC. New gene variants associated with venous thrombosis: a replication study in White and Black Americans. *J Thromb Haemost*. 2011;9(3):489-495.
66. Morange PE, Oudot-Mellakh T, Cohen W, et al. KNG1 Ile581Thr and susceptibility to venous thrombosis. *Blood*. 2011;117(13):3692-3694.
67. Wells PS, Anderson JL, Scarvelis DK, Doucette SP, Gagnon F. Factor XIII Val34Leu variant is protective against venous thromboembolism: a HuGE review and meta-analysis. *Am J Epidemiol*. 2006;164(2):101-109.
68. Morange PE, Bezemer I, Saut N, et al. A follow-up study of a genome-wide association scan identifies a susceptibility locus for venous thrombosis on chromosome 6p24.1. *Am J Hum Genet*. 2010;86(4):592-595.
69. Tsantes AE, Nikolopoulos GK, Bagos PG, et al. Association between the plasminogen activator inhibitor-1 4G/5G polymorphism and venous thrombosis. A meta-analysis. *Thromb Haemost*. 2007;97(6):907-913.
70. Gohil R, Peck G, Sharma P. The genetics of venous thromboembolism. A meta-analysis involving approximately 120,000 cases and 180,000 controls. *Thromb Haemost*. 2009;102(2):360-370.
71. Morange PE, Saut N, Antoni G, Emmerich J, Tréguët DA. Impact on venous thrombosis risk of newly discovered gene variants associated with FVIII and VWF plasma levels. *J Thromb Haemost*. 2011;9(1):229-231.
72. Bezemer ID, Arellano AR, Tong CH, et al. F9 Malmö, factor IX and deep vein thrombosis. *Haematologica*. 2009;94(5):693-699.
73. Hernandez W, Gamazon ER, Smithberger E, et al. Novel genetic predictors of venous thromboembolism risk in African Americans. *Blood*. 2016;127(15):1923-1929.
74. Kirino Y, Bertsias G, Ishigatsubo Y, et al. Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B*51 and ERAP1. *Nat Genet*. 2013;45(2):202-207.
75. Tsoi LC, Spain SL, Knight J, et al; Wellcome Trust Case Control Consortium 2. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet*. 2012;44(12):1341-1348.
76. International Genetics of Ankylosing Spondylitis Consortium (IGAS); Cortes A, Hadler J, Pointon JP, et al. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*. 2013;45(7):730-738.
77. Ungprasert P, Sanguankeo A, Upala S, Suksaranjit P. Psoriasis and risk of venous thromboembolism: a systematic review and meta-analysis. *QJM*. 2014;107(10):793-797.
78. Yazici H, Fresko I, Yurdakul S. Behçet's syndrome: disease manifestations, management, and advances in treatment. *Nat Clin Pract Rheumatol*. 2007;3(3):148-155.
79. Eriksson JK, Jacobsson L, Bengtsson K, Askling J. Is ankylosing spondylitis a risk factor for cardiovascular disease, and how do these risks compare with those in rheumatoid arthritis? *Ann Rheum Dis*. 2017;76(2):364-370.
80. Lee EJ, Dykas DJ, Leavitt AD, et al. Whole-exome sequencing in evaluation of patients with venous thromboembolism. *Blood Adv*. 2017;1(16):1224-1237.
81. Malik R, Chauhan G, Traylor M, et al; MEGASTROKE Consortium. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes [published correction appears in *Nat Genet*. 2019;51(7):1192-1193] *Nat Genet*. 2018;50(4):524-537.
82. Ellinghaus D, Jostins L, Spain SL, et al; Psoriasis Association Genetics Extension (PAGE). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet*. 2016;48(5):510-518.
83. Bucciarelli P, Maino A, Felicetta I, et al. Association between red cell distribution width and risk of venous thromboembolism. *Thromb Res*. 2015;136(3):590-594.
84. Ellingsen TS, Lapppegård J, Skjelbakken T, Brækkan SK, Hansen JB. Red cell distribution width is associated with incident venous thromboembolism (VTE) and case-fatality after VTE in a general population. *Thromb Haemost*. 2015;113(1):193-200.
85. Rezende SM, Lijfering WM, Rosendaal FR, Cannegieter SC. Hematologic variables and venous thrombosis: red cell distribution width and blood monocyte count are associated with an increased risk. *Haematologica*. 2014;99(1):194-200.
86. Gieger C, Radhakrishnan A, Cvejic A, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature*. 2011;480(7376):201-208.
87. Koupenova M, Clancy L, Corkrey HA, Freedman JE. Circulating platelets as mediators of immunity, inflammation, and thrombosis. *Circ Res*. 2018;122(2):337-351.
88. Byrnes JR, Wolberg AS. Red blood cells in thrombosis. *Blood*. 2017;130(16):1795-1799.
89. De Stefano V, Rossi E. Testing for inherited thrombophilia and consequences for antithrombotic prophylaxis in patients with venous thromboembolism and their relatives. A review of the Guidelines from Scientific Societies and Working Groups. *Thromb Haemost*. 2013;110(4):697-705.
90. Hall S. Venous thromboembolism prophylaxis after hip or knee arthroplasty. *N Engl J Med*. 2018;378(19):1848.
91. Marik PE, Cavallazzi R. Extended anticoagulant and aspirin treatment for the secondary prevention of thromboembolic disease: a systematic review and meta-analysis. *PLoS One*. 2015;10(11):e0143252.
92. Lin DY, Zeng D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol*. 2010;34(1):60-66.
93. Heit JA, Armasu SM, McCauley BM, et al. Identification of unique venous thromboembolism-susceptibility variants in African-Americans. *Thromb Haemost*. 2017;117(4):758-768.