Received: 17 August 2018 | Revised: 26 November 2018 | Accepted: 11 December 2019

DOI: 10.1002/gepi.22187

RESEARCH ARTICLE

WILEY Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

A large-scale exome array analysis of venous thromboembolism

Sara Lindström[1,2] | Jennifer A. Brody[3] | Constance Turman[4] | Marine Germain[5] | Traci M. Bartz[3,6] | Erin N. Smith[7,8] | Ming-Huei Chen[9] | Marja Puurunen[10] | Daniel Chasman[11] | Jeffrey Hassler[2] | Nathan Pankratz[12] | Saonli Basu[13] | Weihua Guan[13] | Beata Gyorgy[14,15] | Manal Ibrahim[16,17,18] | Jean-Philippe Empana[19,20] | Robert Olaso[21] | Rebecca Jackson[22] | Sigrid K. Brækkan[23] | Barbara McKnight[6] | Jean-Francois Deleuze[21] | Cristopher J. O'Donnell[24] | Xavier Jouven[19,25] | Kelly A. Frazer[7,8,26] | Bruce M. Psaty[1,3,27,28] | Kerri L. Wiggins[3] | Kent Taylor[29] | Alexander P. Reiner[1] | Susan R. Heckbert[1] | Charles Kooperberg[2] | Paul Ridker[11] | John-Bjarne Hansen[8,23] | Weihong Tang[30] | Andrew D. Johnson[9] | Pierre-Emmanuel Morange[16,17,18] | David A. Trégouët[5] | Peter Kraft[4,31] | Nicholas L. Smith[1,28,32] | Christopher Kabrhel[33,34,35] |

on behalf of the INVENT Consortium

[1]Department of Epidemiology, University of Washington, Seattle, Washington

[2]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington

[3]Department of Medicine, University of Washington, Seattle, Washington

[4]Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, Massachusetts

[5]University of Bordeaux, Inserm 1219, Bordeaux Population Health Research Center, Bordeaux, France

[6]Department of Biostatistics, University of Washington, Seattle, Washington

[7]Department of Pediatrics, Rady Children's Hospital, University of California, San Diego, California

[8]Department of Clinical Medicine, UiT—The Arctic University of Norway, K.G. Jebsen Thrombosis Research and Expertise Center (TREC), Tromsø, Norway

[9]Population Sciences Branch, National Heart, Lung and Blood Institute's The Framingham Heart Study, Framingham, Massachusetts

[10]School of Medicine, Boston University, Boston, Massachusetts

[11]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts

[12]Department of Laboratory Medicine and Pathology, School of Medicine, University of Minnesota, Minneapolis, Minnesota

[13]Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota

[14]Team Genomics & Pathophysiology of Cardiovascular Diseases, Sorbonne Universités, UPMC Univ. Paris 06, INSERM, UMR_S 1166, Paris, France

[15]ICAN Institute for Cardiometabolism and Nutrition, Paris, France

[16]Laboratory of Haematology, La Timone Hospital, Marseille, France

[17]Aix-Marseille University, INSERM, INSERM, INRA, C2VN, Marseille, France

[18]CRB Assistance Publique Hopitaux de Marseille HemoVasc, Marseille, France

[19]Department of Epidemiology, Université Paris Descartes, Sorbonne Paris Cité, INSERM UMR_S 970, Paris, France

[20]Faculté de Médecine, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

[21]Centre National de Recherche en Génomique Humaine (CNRGH), Direction de la Recherche Fondamentale, CEA, Institut de Biologie François Jacob, Evry, France

 © 2019 Wiley Periodicals, Inc.

[22]Ohio State University, Columbus, Ohio
[23]Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway
[24]Department of Cardiology, Boston Veteran's Administration Healthcare, Boston, Massachusetts
[25]Department of Cardiology, Georges Pompidou European Hospital, APHP, Paris, France
[26]Institute for Genomic Medicine, University of California, San Diego, California
[27]Department of Health Services, University of Washington, Seattle, Washington
[28]Kaiser Permanente Washington Research Institute, Seattle, Washington
[29]LA BioMed, Torrance, California
[30]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota
[31]Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, Massachusetts
[32]Department of Veteran Affairs Office of Research and Development, Seattle Epidemiologic Research and Information Center, Seattle, Washington
[33]Department of Emergency Medicine, Center for Vascular Emergencies, Massachusetts General Hospital, Boston, Massachusetts
[34]Channing Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts
[35]Harvard Medical School, Boston, Massachusetts

Correspondence
Christopher Kabrhel, Center for Vascular Emergencies, Department of Emergency Medicine, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114-2696.
Email: ckabrhel@partners.org

Funding information
ICAN Institute for Cardiometabolism and Nutrition, Grant/Award Number: ANR-10-IAHU; GENMED Laboratory of Excellence on Medical Genomics, Grant/Award Number: ANR-10-LABX-0013; National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center, Grant/Award Number: DK063491; National Heart, Lung, and Blood Institute, Grant/Award Numbers: R01HL116854, HL74745, R01HL35464, HL060739, HHSN268200800007C, HHSN268201700003I, N01HC55222, N01HC85083, N01-HC-25195, N01HC85079, HL080467, HHSN268201600018C, R01CA67262, HL073410, HL043851, N01HC85082, U01HL080295, HHSN268201600001C, R01HL120393, HL085251, R01HL59367, N01HC85086, HL053375, HHSN268201200036C, HL068986, HL095080, R01HL034594, R01HL088521, HHSN268201800001C, R01HL068986, HHSN268201700005I, HHSN268201700001I, N01HC85081, HHSN268201600002C, N01HC85080, HHSN268201600003C, HL040628, HL68639, R01HL103612, HL043201, R01HL087652, R01HL130114, P01CA055075, HL099355, R01HL105756, HHSN268201600004C, R01CA50385, HHSN268201700002I; US Department of Health and Human Services, National Institutes of Health, National Institute of Aging, Grant/Award Number: R01AG023629; NIH American Recovery and

Abstract

Although recent Genome-Wide Association Studies have identified novel associations for common variants, there has been no comprehensive exome-wide search for low-frequency variants that affect the risk of venous thromboembolism (VTE). We conducted a meta-analysis of 11 studies comprising 8,332 cases and 16,087 controls of European ancestry and 382 cases and 1,476 controls of African American ancestry genotyped with the Illumina HumanExome BeadChip. We used the seqMeta package in R to conduct single variant and gene-based rare variant tests. In the single variant analysis, we limited our analysis to the 64,794 variants with at least 40 minor alleles across studies (minor allele frequency [MAF] ~0.08%). We confirmed associations with previously identified VTE loci, including *ABO*, *F5*, *F11*, and *FGA*. After adjusting for multiple testing, we observed no novel significant findings in single variant or gene-based analysis. Given our sample size, we had greater than 80% power to detect minimum odds ratios greater than 1.5 and 1.8 for a single variant with MAF of 0.01 and 0.005, respectively. Larger studies and sequence data may be needed to identify novel low-frequency and rare variants associated with VTE risk.

KEYWORDS
exome, genetic association, venous thromboembolism

Reinvestment Act, Grant/Award Number: 5RC2HL102419; National Cancer Institute, Grant/Award Numbers: P01CA87969, CA047988, UM1CA182913, R01CA49449; National Center for Advancing Translational Sciences, Grant/Award Number: UL1TR000124

1 | INTRODUCTION

Candidate gene studies and Genome-Wide Association Studies (GWAS) have identified multiple genetic variants that are associated with venous thromboembolism (VTE), a condition spanning both pulmonary embolism (PE) and deep vein thrombosis (DVT). The majority of genetic variants associated with VTE have been located in genes known to be involved in hemostasis, such as *ABO*, *F2*, *F5*, *F11*, *FGG*, *and PROCR* (Crous-Bou, Harrington, & Kabrhel, 2016). Associations have also been observed for variants located in genes outside known hemostasis pathways, such as *TSPAN15* and *SLC44A2* (Germain et al., 2015), and the exact mechanisms by which these genes influence VTE risk have yet to be determined.

Despite these successes, much of the heritability of VTE remains unexplained. A recent study based on 3,290 VTE cases and 116,868 controls from the UK Biobank estimated the heritability due to genotyped and imputed single-nucleotide polymorphisms (SNPs) to be ~30% (Klarin, Emdin, Natarajan, Conrad, & Kathiresan, 2017), and twin studies have estimated VTE heritability to be as high as ~50% (Heit et al., 2004). However, the UK Biobank study also noted that known variants only explain 5% of VTE heritability. Thus, additional genetic variants that contribute to VTE risk remain to be discovered.

Recently developed exome arrays (Huyghe et al., 2013) allow for cost-efficient genotyping of 240,000 coding variants identified through the NHLBI Exome Sequencing Project (Tennessen et al., 2012). Based on exome and whole-genome sequencing data from 9,000 subjects of European ancestry, 2,000 subjects of African ancestry and 500 subjects each of Hispanic and Asian ancestry, 240,000 SNPs were selected for inclusion on the exome array. To be selected, nonsynonymous variants had to be seen at least three times in at least two data sets whereas splice and stop variants had to be seen at least two times and in at least two data sets. The exome array has proven to be an efficient tool for identifying low-frequency coding variants associated with blood and cardiovascular traits including: hypertension (Liu et al., 2016; Surendran et al., 2016), hematological traits (Chen et al., 2017; Eicher et al., 2016; Chami et al., 2016; Mousas et al., 2017), lipid levels (Peloso et al., 2014), coronary artery disease (Myocardial Infarction Genetics et al., 2016), and atrial fibrillation (Christophersen et al., 2017). However, no study has published a comprehensive investigation of the association between low-frequency exonic variants and VTE.

We hypothesized that exonic, low-frequency genetic variation would be associated with VTE. We meta-analyzed exome array genotype data from 11 European and US studies, totaling 8,723 VTE cases and 17,563 controls. We conducted both single-variant and gene-based tests to identify novel genetic variants associated with VTE risk.

2 | MATERIALS AND METHODS

2.1 | Study participants

All study participants were either of European or African American ancestry and came from eight US-based cohorts (ARIC, CHS, FHS, HPFS, NHS, NHSII, WGHS, and WHI), one US-based case–control study (HVH), one Norwegian case-control study (Tromsø), and one French case-control study (MARTHA; ARIC Investigators, 1989, Fried et al., 1991; Germain et al., 2011; Giovannucci et al., 1995; Hankinson et al., 1995; Heckbert, Li, Cummings, Smith, & Psaty, 2008, Kannel, Dawber, Kagan, Revotskie, & Stokes, 1961; Psaty et al., 2009; Ridker et al., 2008; Solomon et al., 2016; Tell et al., 1993; Tworoger, Sluss, & Hankinson, 2006; Women's Health Initiative Study Group, 1998; Table 1). Details of each study have been previously published (ARIC Investigators, 1989, Fried et al., 1991; Germain et al., 2011; Giovannucci et al., 1995; Hankinson et al., 1995; Heckbert et al., 2008; Kannel et al., 1961; Ridker et al., 2008; Solomon et al., 2016; Tell et al., 1993; Tworoger et al., 2006; Women's Health Initiative Study Group, 1998). Physician-diagnosed VTE was identified either through hospital records or validated self-reports, supplemented by review of medical records. A detailed description of the study-specific design and characteristics is presented in Supporting Information Table 1. All participating studies were approved by their respective institutional review board and informed consent for genetic analyses was obtained from each study participant.

TABLE 1 Studies included in the VTE exome array analysis

Ethnicity	Study	Country	Cases	Controls
African American	ARIC	US	202	807
African American	CHS	US	30	120
African American	HVH	US	58	181
African American	WHI	US	92	368
European	ARIC	US	433	1,734
European	CHS	US	112	448
European	FHS	US	212	848
European	HPFS/NHS/NHSII	US	2,321	2,301
European	HVH	US	841	1,788
European	MARTHA	France	2,628	3,414
European	Tromsø	Norway	528	526
European	WGHS	US	610	2,404
European	WHI	US	656	2,624
Total			8,341	16,087

2.2 | Genotyping and quality control

Genotyping was conducted using either the Illumina HumanExome BeadChip v1.0 (Illumina, San Diego, CA), Illumina HumanExome BeadChip v1.1 (Illumina, San Diego, CA), or the Illumina HumanCore Exome BeadChip v1.1 (Illumina, San Diego, CA), depending on study. Genotypes from 765 samples from the Tromsø study were obtained from exome sequencing rather than genotyping. Genotypes were called using either GenomeStudio or Zcall. Each study conducted data cleaning and quality assurance checks following a common protocol. Details of study-specific genotype calling and quality control can be found in Supporting Information Table 1.

We conducted gene set enrichment analysis using the GSEAPreranked algorithm as implemented in the GenePattern software and the KEGG, Gene Ontology, and Hallmarks pathway sets (Mootha et al., 2003; Subramanian et al., 2005). We applied GSEAPreranked to four sets of results (a) burden test of variants using a broad definition, (b) burden test of variants using a strict definition, (c) Sequence Kernel Association Test (SKAT) test of variants using a broad definition, and (d) SKAT test of variants using a strict definition.

One of the included studies (MARTHA) genotyped cases and controls on separate platforms (cases were genotyped on Illumina HumanExome 12v.1-2_A and controls were genotyped with the Illumina Human Exome 12v.1_A array). We identified a moderate inflation in test statistics ($\lambda_{1000} = 1.09$) in MARTHA for the single variant analysis with minor allele frequency (MAF) > 0.005. Therefore, we re-meta-analyzed the data while excluding MARTHA, leaving a total of 6,095 cases and 14,149 controls in a sensitivity analysis.

2.3 | Statistical methods

Each study conducted individual analysis following a common protocol. To avoid Type 1 error inflation (Ma, Blackwell, Boehnke, Scott, & Go, 2013), studies with more than four controls per case randomly selected a maximum of four controls for each case (i.e., 1:4 case: control ratio). Those studies that performed control selection (ARIC, CHS, FHS, WGHS, and WHI) reviewed the distribution of age and sex following control selection to ensure roughly equal distributions among cases and controls. Each study conducted both single variant and gene-based analysis. Association analysis were based on logistic regression adjusting for age, sex, principal components, and other study-specific variables (as needed). Analyses were conducted using the seqMeta (Voorman, Chen, Lumley, & David, 2017) package in R which produces study-specific results. Each study sent their study-specific results to the co-ordinating center at Harvard TH Chan School of Public Health where the meta-analyses took place.

For single variant analysis, we conducted a meta-analysis of the study-specific score statistics based on an additive coding. We limited our analysis to the 64,794 variants that had at least 40 minor alleles across studies (corresponding to a MAF of ~0.08%). Bonferroni correction for the number of variants tested was used to set the significance threshold for the analysis corresponding to $p < 7.7 \times 10^{-7}$ (0.05/64,794 variants).

For the gene-based rare variant analyses, we conducted two tests: weighted-sum burden (WSB; Madsen & Browning, 2009) test as implemented in seqMeta (Voorman et al., 2017) and SKAT (Wu et al., 2011). We applied two different sets of criteria to select variants, based on coding variant annotation from five prediction algorithms (PolyPhen2, HumDiv and HumVar, LRT, MutationTaster and SIFT; Purcell et al., 2014). The "broad" definition included variants with a MAF < 0.01 that were nonsense, stop-loss, splice site, as well as missense variants that are annotated as damaging by at least one prediction algorithm. The "strict" definition included only variants with a MAF < 0.01 that were nonsense, stop-loss, splice-site, as well as missense variants annotated as damaging by all five algorithms. For the SKAT analysis, variants were weighted according to the β density function as previously described (Wu et al., 2011). We excluded

all genes that had fewer than two variants included in the analysis. In total, we tested 15,041 genes using the broad definition and 5,749 genes using the strict definition. Bonferroni correction for the number of genes and tests performed was used to set the significance threshold for the gene-based analysis corresponding to $p < 1.2 \times 10^{-6}$ (0.05/41,580 tests [(5,749 + 15,041) genes × 2 tests]).

One of the included studies (MARTHA) genotyped cases and controls on separate platforms (cases were genotyped on Illumina HumanExome 12v.1–2_A and controls were genotyped with the Illumina HumanExome 12v.1_A array. We identified a moderate inflation in test statistics ($\lambda_{1000} = 1.09$) in MARTHA for the single variant analysis with MAF > 0.005. Therefore, we reran the analysis excluding MARTHA, leaving a total of 6,095 cases and 14,149 controls in a sensitivity analysis.

3 | RESULTS

3.1 | Single variant analysis

After excluding all variants with MAC < 40 (corresponding to a MAF of ~0.08%) in the combined study population, single variant meta-analysis showed no sign of genomic inflation ($\lambda_{1000} = 1.03$; Supporting Information Figure 1). The strongest association was observed for rs635634 at the *ABO* locus (odds ratio [OR]: 1.60; 95% confidence interval [CI]: 1.52–1.68; $p = 1.51 \times 10^{-73}$). In addition, we observed significant associations for previously known genes including the *F5*, *FGG*, and *F11* (Supporting Information Table 2). The most strongly associated rare (MAF < 0.01) variant we observed was rs121918472 (OR: 1.93; 95% CI: 1.46–2.56; $P = 3.55 \times 10^{-6}$), a nonsynonymous variant located in the Protein S (*PROS1*) gene, also known as the p.Ser501Pro or PS Herleen mutation, MAF = 0.005. This variant is also known to be associated with VTE (Suchon et al., 2017). After excluding known loci, only one single variant remained significant after adjusting for multiple testing but this signal was driven exclusively by the MARTHA study ($p = 1.96 \times 10^{-15}$ when including MARTHA and $p = 0.37$ after excluding MARTHA). As the signal at this locus is most likely due to technical issues, we removed it from further analysis. No other variant reached genome-wide significance ($p < 7.7 \times 10^{-7}$). The strongest subthreshold association was observed for rs755109, a common (MAF = 0.37) variant previously associated with thyroid stimulating hormones (OR: 1.10; 95% CI: 1.06–1.16; $p = 3.31 \times 10^{-6}$).

3.2 | Weighted-sum burden (WSB) rare variant analysis

No gene reached the predetermined significance threshold of $p < 1.2 \times 10^{-6}$ (Figure 1). The top three associated genes

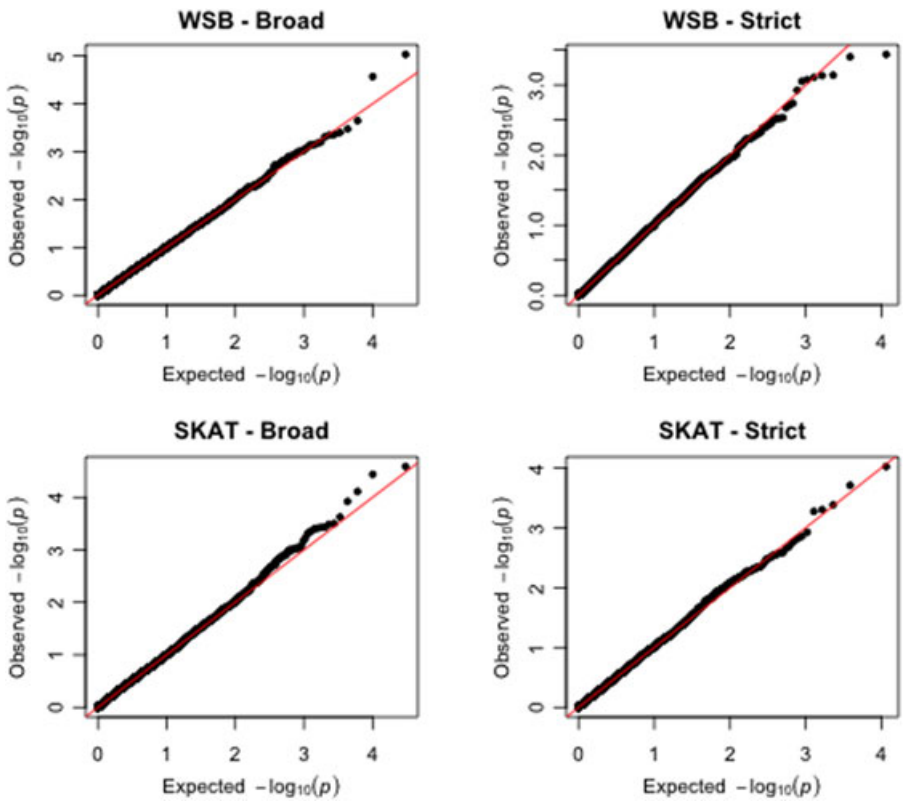

FIGURE 1 QQ plots for gene burden tests including nonsynonymous variants with MAF ≤ 0.01. The WSB test using a broad definition of variant inclusion (upper left panel), the WSB test using a strict definition of variant inclusion (upper right panel), the SKAT test using a broad definition of variant inclusion (lower left panel), the SKAT test using a strict definition of variant inclusion (lower right panel). MAF: minor allele frequency; SKAT: Sequence Kernel Association Test; WSB: weighted-sum burden

TABLE 2 Association results for the three strongest associations from the rare variant weighted-sum burden test

Variant inclusion	Genes	β	*SE*	*p* Value	CMAF[a]	No. of variants[b]
Broad	*FAM71C*	3.86	0.871	9.30E−06	0.0002	2
	FOXB2	1.56	0.372	2.71E−05	0.0009	4
	SERPINA10	0.28	0.077	0.0002	0.017	8
Strict	*DGAT2*	0.64	0.179	0.0004	0.003	4
	NUDT12	3.11	0.877	0.0004	0.0002	2
	SERPINA10	0.28	0.082	0.0007	0.015	6

[a]Cumulative minor allele frequency (CMAF) for single-nucleotide polymorphisms included in the analysis.
[b]Number of variants included in the analysis.

using the "broad" and "strict" definitions are shown in Table 3 and all associations with $p < 0.01$ are shown in Supporting Information Tables 5 and 6. The *SERPINA10* gene on chromosome 14 was the third strongest-associated gene using the broad ($p = 0.0002$) and the strict ($p = 0.0007$) definition. *SERPINA10* is expressed primarily in the liver and mutations in this gene have previously been linked to VTE (Corral et al., 2006; Van de Water et al., 2004).

3.3 | SKAT rare variant analysis

No gene reached the predetermined significance threshold of $p < 1.2 \times 10^{-6}$ (Figure 1). The top three associated genes using the using the "broad" and "strict" definitions are shown in Table 3 and all associations with $p < 0.01$ are shown in Supporting Information Tables 5 and 6.

3.4 | Gene set enrichment analysis

Gene set enrichment analysis based on the results obtained from WSB and SKAT did not yield any significant pathway after adjusting for number of pathways tested (all FDR $q > 0.08$, data not shown).

4 | DISCUSSION

To assess the contribution of rare coding variation to VTE risk, we combined data from 11 studies spanning four countries, resulting in exome array data on 8,723 VTE cases and 17,563 controls. By comparison, the number of cases included in this study is larger than the largest GWAS of VTE published to date (Germain et al., 2015). Beyond known associations, we did not observe evidence that additional low-frequency and rare coding variants with moderate-to-large effects contribute to VTE risk.

Although our study is the largest genomic study of VTE to date, our ability to identify rare variants associated with VTE was limited by low statistical power. Given our sample size, we had greater than 80% power to detect minimum odds ratios of 1.56 and 1.81 for a single variant with MAF = 0.01 and 0.005, respectively. It is estimated that the exome array includes 97–98% of nonsynonymous variants and 94–95% of stop variants that would have been detected in an average genome through exome sequencing (https://genome.sph.umich.edu/wiki/Exome_Chip_Design). Thus, it is possible that we missed coding variants associated with VTE, especially if such variants are particularly rare in individuals who are not affected with VTE. Another limitation with this study is the limited contribution of non-European populations to our analyses, with 93% of our study population being of European ancestry.

Identifying risk factors for VTE, including genetic risk factors, is of great public health importance. VTE affects 1–2/1,000 Americans yearly. The incidence has been increasing and mortality from PE remains high (Arshad, Isaksen, Hansen, & Braekkan, 2017; Heit, 2005; Heit et al., 2017; Silverstein et al., 1998). The mortality of VTE is greatest in the first 24 hr, and for one-fourth of PE

TABLE 3 Association results for the three strongest associations from the rare variant Sequence Kernel Association Test

Variant inclusion	Genes	Q_{meta}[a]	*p* Value	CMAF[b]	No. of variants[c]
Broad	*CREB3L1*	596343.13	2.59E−05	0.016	7
	FAM71C	9844.60	3.65E−05	0.0002	2
	PHC3	514613.94	7.76E−05	0.017	10
Strict	*SRR*	37032.58	9.52E−05	0.0007	4
	ABCF3	15590.02	0.0002	0.0007	2
	DSC1	50186.59	0.0004	0.0014	5

[a]The SKAT Q statistic, defined as $\sum_j w_j U_j^2$, where w_j is the weight given to SNP j, and U_j^2 is the associated score statistic.
[b]Cumulative minor allele frequency (CMAF) for single-nucleotide polymorphisms included in the analysis.
[c]Number of variants included in the analysis.

patients, the initial clinical presentation is sudden death (Cohen et al., 2007; Courtney & Kline, 2001; Goldhaber, Visani, & De Rosa, 1999). Therefore, our ability to improve mortality hinges on primary prevention, identifying patients at risk for VTE, and understanding the underlying pathophysiology of the disease. Despite the accumulated evidence that genetic factors play a major role in the pathophysiology of VTE, only 35% of VTE patients undergoing testing for thrombophilia carry a polymorphism known to increase VTE risk (Cushman, 2005). Additional efforts to identify genetic variants associated with VTE risk are still needed.

The INVENT Consortium is a well-established collaboration of genetic studies of VTE and, to our knowledge, our meta-analysis includes data from the vast majority of exome array studies of VTE. Additional large studies, potentially including comprehensive sequencing data, may be needed to identify novel low-frequency rare coding variants associated with VTE. Further research into the genetic basis of VTE is needed to aid in the primary prevention of this potentially fatal disease.

ACKNOWLEDGEMENTS

Atherosclerosis Risk in Communities (ARIC): The ARIC study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, NIH, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN 268201700005I). The authors thank the staff and participants of the ARIC study for their important contributions. Funding for LITE was supported by R01HL59367. Funding support for "Building on GWAS for NHLBI-diseases: the US CHARGE consortium" was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA; 5RC2HL102419). Cardiovascular Health Study (CHS): This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL 087652, R01HL105756, R01HL103612, R01HL120393, R01HL130114, and R01HL068986 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC; Grant DK063491) to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Framingham Heart Study (FHS): The FHS is conducted and supported by the NHLBI in collaboration with Boston University (Contract No. N01-HC-25195). Genotyping, quality control, and calling of the Illumina HumanExome BeadChip in the FHS were supported by funding from the National Heart, Lung and Blood Institute Division of Intramural Research (Daniel Levy and Christopher J. O'Donnell, Principal Investigators). Support for the centralized genotype calling was provided by Building on GWAS for NHLBI-diseases: the US CHARGE consortium through the National Institutes of Health (NIH) American Recovery and Reinvestment Act of 2009 (5RC2HL102419). M.H.C. and A.D.J. were supported by National Heart, Lung and Blood Institute Division of Intramural Research funds. The views expressed in this manuscript 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 US Department of Health and Human Services. HVH: The research of the Heart and Vascular Health Studies has been funded in part by National Institute of Health grants HL040628, HL043201, HL053375, HL060739, HL68639, HL068986, HL073410, HL74745, HL085251, HL095080. MARTHA genetics project was supported by the GENMED Laboratory of Excellence on Medical Genomics (ANR-10-LABX-0013), the French Clinical Research Infrastructure Network on Venous Thrombo-Embolism (F-CRIN INNOVTE) and the ICAN Institute for Cardiometabolism and Nutrition (ANR-10-IAHU), three research programs managed by the National Research Agency (ANR) part of the French Investment for the Future initiative. NHS/NHSII/HPFS: These studies received grant supports P01CA87969, R01CA49449, R01HL034594, R01HL088521, R01CA50385, R01CA67262, P01CA055075, R01HL35464, R01HL116854. We thank all of the researchers and staff whose hard work made each of the contributing studies possible. Tromsø: This study was supported by an independent grant from Stiftelsen Kristian Gerhard Jebsen in Norway. WGHS: The Women's Genome Health Study is funded by the National Heart, Lung, and Blood Institute (HL043851, HL080467, and HL099355) and the National Cancer Institute (CA047988 and UM1CA182913), with support for genotyping provided by Amgen. WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN 268201600003C, and HHSN268201600004C.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

ORCID

Saonli Basu http://orcid.org/0000-0003-1200-4546
Christopher Kabrhel http://orcid.org/0000-0002-8699-7176

REFERENCES

Aric Investigators (1989). The Atherosclerosis Risk in Communities (ARIC) Study: Design and objectives. *American Journal of Epidemiology*, *129*, 687–702.

Arshad, N., Isaksen, T., Hansen, J. B., & Braekkan, S. K. (2017). Time trends in incidence rates of venous thromboembolism in a large cohort recruited from the general population. *European Journal of Epidemiology*, *32*, 299–305.

Chami, N., Chen, M. H., Slater, A. J., Eicher, J. D., Evangelou, E., Tajuddin, S. M., ... Zonderman, A. B. (2016). Exome genotyping identifies pleiotropic variants associated with red blood cell traits, *American Journal of Human Genetics 99*, 8–21.

Chen, M. H., Yanek, L. R., Backman, J. D., Eicher, J. D., Huffman, J. E., Ben-Shlomo, Y., ... Faraday, N. (2017). Exome-chip meta-analysis identifies association between variation in ANKRD26 and platelet aggregation. *Platelets*, 1–10. Advance online publication. https://doi.org/10.1080/09537104.2017.1384538

Christophersen, I. E., Rienstra, M., Roselli, C., Yin, X., Geelhoed, B., Barnard, J., ... Ellinor, P. T. (2017). Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nature Genetics*, *49*, 946–952.

Cohen, A., Agnelli, G., Anderson, F., Arcelus, J., Bergqvist, D., Brecht, J., ... Spannagl, M. (2007). Venous thromboembolism (VTE) in Europe. The number of VTE events and associated morbidity and mortality. *Thrombosis and Haemostasis*, *98*, 756–764.

Corral, J., Gonzalez-Conejero, R., Soria, J. M., Gonzalez-Porras, J. R., Perez-Ceballos, E., Lecumberri, R., ... Vicente, V. (2006). A nonsense polymorphism in the protein Z-dependent protease inhibitor increases the risk for venous thrombosis. *Blood*, *108*, 177–183.

Courtney, D. M., & Kline, J. A. (2001). Identification of prearrest clinical factors associated with outpatient fatal pulmonary embolism. *Academic Emergency Medicine*, *8*, 1136–1142.

Crous-Bou, M., Harrington, L., & Kabrhel, C. (2016). Environmental and genetic risk factors associated with venous thromboembolism. *Seminars in Thrombosis and Hemostasis*, *42*, 808–820.

Cushman, M. (2005). Inherited risk factors for venous thrombosis. *Hematology*, *2005*, 452–457.

Eicher, J. D., Chami, N., Kacprowski, T., Nomura, A., Chen, M. H., Yanek, L. R., ... Deary, I. J. (2016). Platelet-related variants identified by exomechip meta-analysis in 157,293 individuals, *American Journal of Human Genetics 99*, 40–55.

Fried, L. P., Borhani, N. O., Enright, P., Furberg, C. D., Gardin, J. M., Kronmal, R. A., ... Weiler, P. G. (1991). The Cardiovascular Health Study: Design and rationale. *Annals of Epidemiology*, *1*, 263–276.

Germain, M., Chasman, D. I., De Haan, H., Tang, W., Lindstrom, S., Weng, L. C., ... Morange, P. E. (2015). Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism, *American Journal of Human Genetics 96*, 532–542.

Germain, M., Saut, N., Greliche, N., Dina, C., Lambert, J. C., Perret, C., ... Morange, P. E. (2011). Genetics of venous thrombosis: Insights from a new genome wide association study. *PLOS One*, *6*, e25581.

Giovannucci, E., Ascherio, A., Rimm, E. B., Stampfer, M. J., Colditz, G. A., & Willett, W. C. (1995). Intake of carotenoids and retinol in relation to risk of prostate cancer. *Journal of the National Cancer Institute*, *87*, 1767–1776.

Goldhaber, S. Z., Visani, L., & De rosa, M. (1999). Acute pulmonary embolism: Clinical outcomes in the International Cooperative Pulmonary Embolism Registry (ICOPER). *Lancet*, *353*, 1386–1389.

Hankinson, S. E., Colditz, G. A., Hunter, D. J., Manson, J. E., Willett, W. C., Stampfer, M. J., ... Speizer, F. E. (1995). Reproductive factors and family history of breast cancer in relation to plasma estrogen and prolactin levels in postmenopausal women in the Nurses' Health Study (United States). *Cancer Causes & Control*, *6*, 217–224.

Heckbert, S. R., Li, G., Cummings, S. R., Smith, N. L., & Psaty, B. M. (2008). Use of alendronate and risk of incident atrial fibrillation in women. *Archives of Internal Medicine*, *168*, 826–831.

Heit, J. A. (2005). Venous thromboembolism: Disease burden, outcomes and risk factors. *Journal of Thrombosis and Haemostasis*, *3*, 1611–1617.

Heit, J. A., Ashrani, A. A., Crusan, D. J., McBane, R. D., Petterson, T. M., & Bailey, K. R. (2017). Reasons for the persistent incidence of venous thromboembolism. *Thrombosis and Haemostasis*, *117*, 390–400.

Heit, J. A., Phelps, M. A., Ward, S. A., Slusser, J. P., Petterson, T. M., & De Andrade, M. (2004). Familial segregation of venous thromboembolism. *Journal of Thrombosis and Haemostasis*, *2*, 731–736.

Huyghe, J. R., Jackson, A. U., Fogarty, M. P., Buchkovich, M. L., Stančáková, A., Stringham, H. M., ... Mohlke, K. L. (2013). Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nature Genetics*, *45*, 197–201.

Kannel, W. B., Dawber, T. R., Kagan, A., Revotskie, N., & Stokes, J., 3RD (1961). Factors of risk in the development of coronary heart disease—six year follow-up experience. The Framingham Study. *Annals of Internal Medicine*, *55*, 33–50.

Klarin, D., Emdin, C. A., Natarajan, P., Conrad, M. F., & Kathiresan, S. (2017). Genetic analysis of venous thromboembolism in UK Biobank identifies the ZFPM2 locus and implicates obesity as a causal risk factor. *Circulation: Cardiovascular Genetics*, *10*, 10.

Liu, C., Kraja, A. T., Smith, J. A., Brody, J. A., Franceschini, N., Bis, J. C., ... Chasman, D. I. (2016). Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nature Genetics*, *48*, 1162–1170.

Ma, C., Blackwell, T., Boehnke, M., Scott, L. J., & Go, T. D. I. (2013). Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. *Genetic Epidemiology*, *37*, 539–550.

Madsen, B. E., & Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLOS Genetics*, *5*, e1000384.

Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., ... Groop, L. C. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, *34*, 267–273.

Mousas, A., Ntritsos, G., Chen, M. H., Song, C., Huffman, J. E., Tzoulaki, I., ... Reiner, A. P. (2017). Rare coding variants pinpoint genes that control human hematological traits. *PLOS Genetics*, *13*, e1006925.

Myocardial Infarction Genetics, C. A. E. C. Investigators, Stitziel, N. O., Stirrups, K. E., Masca, N. G., Erdmann, J., ... Ardissino, D. (2016). Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease, *New England Journal of Medicine 374*, 1134–1144.

Peloso, G. M., Auer, P. L., Bis, J. C., Voorman, A., Morrison, A. C., Stitziel, N. O., ... Cupples, L. A. (2014). Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks, *American Journal of Human Genetics 94*, 223–232.

Psaty, B. M., O'Donnell, C. J., Gudnason, V., Lunetta, K. L., Folsom, A. R., Rotter, J. I., ... Boerwinkle, E. (2009). Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circulation: Cardiovascular Genetics*, *2*, 73–80.

Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., ... Sklar, P. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, *506*, 185–190.

Ridker, P. M., Chasman, D. I., Zee, R. Y. L., Parker, A., Rose, L., Cook, N. R., & Buring, J. E., Women's Genome Health Study Working (2008). Rationale, design, and methodology of the Women's Genome Health Study: A genome-wide association study of more than 25,000 initially healthy american women. *Clinical Chemistry*, *54*, 249–255.

Silverstein, M. D., Heit, J. A., Mohr, D. N., Petterson, T. M., O'Fallon, W. M., & Melton, L. J., 3RD (1998). Trends in the incidence of deep vein thrombosis and pulmonary embolism: A 25-year population-based study. *Archives of Internal Medicine*, *158*, 585–593.

Solomon, T., Smith, E. N., Matsui, H., Braekkan, S. K., Wilsgaard, T., Njølstad, I., ... Frazer, K. A. (2016). Associations between common and rare exonic genetic variants and serum levels of 20 cardiovascular-related proteins: The Tromso study. *Circulation: Cardiovascular Genetics*, *9*, 375–383.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 15545–15550.

Suchon, P., Germain, M., Delluc, A., Smadja, D., Jouven, X., Gyorgy, B., ... Trégouët, D. A. (2017). Protein S Heerlen mutation heterozygosity is associated with venous thrombosis risk. *Scientific Reports*, *7*, 45507.

Surendran, P., Drenos, F., Young, R., Warren, H., Cook, J. P., Manning, A. K., ... Munroe, P. B. (2016). Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nature Genetics*, *48*, 1151–1161.

Tell, G. S., Fried, L. P., Hermanson, B., Manolio, T. A., Newman, A. B., & Borhani, N. O. (1993). Recruitment of adults 65 years and older as participants in the Cardiovascular Health Study. *Annals of Epidemiology*, *3*, 358–366.

Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., ... Akey, J. M. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*, *337*, 64–69.

Tworoger, S. S., Sluss, P., & Hankinson, S. E. (2006). Association between plasma prolactin concentrations and risk of breast cancer among predominately premenopausal women. *Cancer Research*, *66*, 2476–2482.

Van De Water, N., Tan, T., Ashton, F., O'grady, A., Day, T., Browett, P., ... Harper, P. (2004). Mutations within the protein Z-dependent protease inhibitor gene are associated with venous thromboembolic disease: A new form of thrombophilia, *British Journal of Haematology 127*, 190–194.

Voorman, A. B. J., Chen, H., Lumley, T., David, B. (2017). seqMeta: An R package for meta-analyzing region-based tests of rare DNA variants. R package version 1.6.7.

Women's Health Initiative Study Group (1998). Design of the women's health initiative clinical trial and observational study. *Controlled Clinical Trials*, *19*, 61–109.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics*, *89*, 82–93.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Lindström S, Brody JA, Turman C, et al. A large-scale exome array analysis of venous thromboembolism. *Genet. Epidemiol.* 2019;43:449–457. https://doi.org/10.1002/gepi.22187