

Title: Cross-Ancestry Investigation of Venous Thromboembolism Genomic Predictors

Florian Thibord PhD^{1,2*}, Derek Klarin MD^{3*}, Jennifer A. Brody BA⁴, Ming-Huei Chen PhD^{1,2}, Michael G. Levin MD^{5,6}, Daniel I. Chasman PhD^{7,8}, Ellen L. Goode PhD MPH⁹, Kristian Hveem PhD^{10,11}, Maris Teder-Laving MSc¹², Angel Martinez-Perez MSc¹³, Dylan Aïssi PhD^{14,15}, Delphine Daian-Bacq PhD^{16,17}, Kaoru Ito PhD MD¹⁸, Pradeep Natarajan MD MMSc^{19,20,21}, Pamela L. Lutsey PhD²², Girish N. Nadkarni MD MPH^{23,24,25}, Gabriel Cuellar-Partida PhD²⁶, Brooke N. Wolford PhD²⁷, Jack W. Pattee PhD^{28,29}, Charles Kooperberg PhD³⁰, Sigrid K. Braekkan PhD^{31,32}, Ruifang Li-Gao PhD³³, Noemie Saut PhD³⁴, Corriene Sept PhD³⁵, Marine Germain MS^{14,15,17}, Renae L. Judy MS³⁶, Kerri L. Wiggins MS RD⁴, Darae Ko MD^{37,2}, Christopher O'Donnell MD MPH³⁸, Kent D. Taylor PhD³⁹, Franco Giulianini PhD⁷, Mariza De Andrade PhD⁹, Therese H. Nøst PhD¹¹, Anne Boland PhD^{16,17}, Jean-Philippe Empana PhD^{40,41}, Satoshi Koyama PhD MD^{18,42,19}, Thomas Gilliland MD^{19,42,21}, Ron Do PhD^{23,24,43}, Xin Wang PhD²⁶, Wei Zhou PhD⁴⁴, Jose Manuel Soria PhD¹³, Juan Carlos Souto MD PhD^{45,13}, Nathan Pankratz PhD⁴⁶, Jeffery Haessler MS³⁰, Kristian Hindberg PhD³¹, Frits R. Rosendaal MD PhD⁴⁷, Constance Turman MS³⁵, Robert Olaso PhD^{16,17}, Rachel L. Kember MSc PhD⁴⁸, Traci M. Bartz MS⁴⁹, Julie A. Lynch PhD RN MBA^{50,51}, Susan R. Heckbert MD MPH⁵², Sebastian M. Armasu MS⁹, Ben Brumpton PhD¹¹, David M. Smadja MD PhD^{53,54}, Xavier Jouven MD PhD^{40,41}, Issei Komuro PhD MD⁵⁵, Katharine Clapham MD^{56,20,21}, Ruth J.F. Loos PhD²³, Cristen Willer PhD²⁷, Maria Sabater-Lleal PhD^{13,57}, James S. Pankow PhD²², Alexander P. Reiner MD MSc^{58,30}, Vania M. Morelli MD PhD^{31,32}, Paul M. Ridker MD MPH^{7,8}, Astrid van Hylckama Vlieg PhD⁴⁷, Jean-François Deleuze PhD^{16,59,60}, Peter Kraft PhD³⁵, Daniel J. Rader MD⁶¹, Barbara McKnight PhD⁴⁹, Global Biobank Meta-Analysis Initiative, Estonian Biobank Research Team, 23andMe Research Team, Kyung Min Lee PhD⁵⁰, Bruce M. Psaty MD PhD^{4,62,63}, Anne Heidi Skogholt PhD¹¹, Joseph Emmerich MD PhD^{64,65}, Pierre Suchon MD PhD^{34,66}, Biobank Japan⁶⁷, Stephen S. Rich PhD⁶⁸, Ha My T. Vy PhD^{23,24}, Weihong Tang MD PhD⁶⁹, Rebecca D. Jackson MD⁷⁰, John-Bjarne Hansen MD PhD^{31,32}, Pierre-Emmanuel Morange MD PhD^{34,66}, Christopher Kabrhe MD MPH^{71,72}, David-Alexandre Tréguët PhD^{14,15,17*}, Scott Damrauer MD^{6*}, Andrew D. Johnson PhD^{1,2*}, Nicholas L. Smith PhD^{49,73*}

*Denotes equal contribution

¹Population Sciences Branch, Division of Intramural Research, National Heart, Lung and Blood Institute, 73 Mt. Wayte, Suite #2, Framingham, MA, 01702, USA,

²The Framingham Heart Study, Boston University and NHLBI, 73 Mt. Wayte Ave, Suite #2, Framingham, MA, 01702, USA,

³Division of Vascular Surgery, Stanford University School of Medicine, Palo Alto, CA, 94305, USA,

⁴Cardiovascular Health Research Unit, Department of Medicine, University of Washington, 1730 Minor Ave, Suite 1360, Seattle, WA, 98101, USA,

⁵Division of Cardiovascular Medicine, Department of Medicine, University of Pennsylvania, 3400 Spruce Street, PA, 19104, USA,

⁶Medicine, Copora Michael J. Crescenzi Philadelphia VA Medical Center, 3900 Woodland Ave, PA, 19104, USA,

- ⁷Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Ave, Boston, MA, 02215, USA,
- ⁸Harvard Medical School, Boston, MA, 02115, USA,
- ⁹Department of Quantitative Health Sciences, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905, USA,
- ¹⁰HUNT Research Center, Department of Public Health and Nursing, Norwegian University of Science and Technology, Forskningsvegen 2, Levanger, 7600, Norway,
- ¹¹K.G. Jebsen Centre for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Håkon Jarls gate 11, Trondheim, 7030, Norway,
- ¹²Institute of Genomics, University of Tartu, Riia 23b, Tartu, Tartu, 51010, Estonia,
- ¹³Genomics of Complex Disease Unit, Sant Pau Biomedical Research Institute (IIB Sant Pau), Barcelona, Spain, St Quinti 77-79, Barcelona, 8041, Spain,
- ¹⁴Bordeaux Population Health Research Center, University of Bordeaux, 146 rue Léo Saignat, Bordeaux, 33076, France,
- ¹⁵UMR1219, INSERM, 146 rue Léo Saignat, Bordeaux, 33076, France,
- ¹⁶Centre National de Recherche en Génomique Humaine, CEA, Université Paris-Saclay, 2 Rue Gaston Crémieux, Evry, 91057, France,
- ¹⁷Laboratory of Excellence on Medical Genomics, France,
- ¹⁸Laboratory for Cardiovascular Genomics and Informatics, RIKEN Center for Integrative Medical Sciences,, 1-7-22 Suehirocho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan,
- ¹⁹Cardiovascular Research Center, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA, 02446, USA,
- ²⁰Program in Medical and Population Genetics and the Cardiovascular Disease Initiative, Broad Institute of Harvard & MIT, 75 Ames St, Cambridge, MA, USA,
- ²¹Department of Medicine, Harvard Medical School, Shattuck St, Boston, MA, USA,
- ²²Division of Epidemiology and Community Health, University of Minnesota, 1300 South Second Street, MN, 55454, USA,
- ²³The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Pl, New York, NY, 10029, USA,
- ²⁴Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Pl, New York, NY, 10029, USA,
- ²⁵Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Pl, New York, NY, 10029, USA,
- ²⁶23andMe, Inc., 223 N Mathilda Ave, CA, 94086, USA,
- ²⁷Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, 48109, USA,
- ²⁸Division of Biostatistics, University of Minnesota, 420 Delaware St. SE, MN, 55455, USA,
- ²⁹Center for Innovative Design & Analysis and Department of Biostatistics & Informatics, Colorado School of Public Health, 13001 East 17th Place, CO, 80045, USA,
- ³⁰Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA, 98109, USA,
- ³¹Thrombosis Research Center (TREC), UiT - The Arctic University of Norway, Universitetsvegen 57, Tromsø, 9037, Norway,

- 88 ³²Division of internal medicine, University Hospital of North Norway, Tromsø, 9038, Norway,
- 89 ³³Clinical Epidemiology, Leiden University Medical Center, PO Box 9600, 2300 RC, The
- 90 Netherlands,
- 91 ³⁴Hematology Laboratory, La Timone University Hospital of Marseille, 264 Rue Saint-Pierre,
- 92 Marseille, 13385, France,
- 93 ³⁵Department of Epidemiology, Harvard TH Chan Harvard School of Public Health, 655
- 94 Huntington Ave., Building II, Boston, MA, 02115, USA,
- 95 ³⁶Surgery, University of Pennsylvania, 3401 Walnut Street, PA, 19104, USA,
- 96 ³⁷Section of Cardiovascular Medicine, Boston University School of Medicine, 85 East Newton
- 97 Street, Boston, MA, 02118, USA,
- 98 ³⁸Cardiology, VA Boston Healthcare System, Boston, MA, 02130, USA,
- 99 ³⁹Institute for Translational Genomics and Population Sciences, The Lundquist Institute for
- 100 Biomedical Innovation, 1124 W Carson St., Torrance, CA 90502, CA, USA,
- 101 ⁴⁰Integrative Epidemiology of cardiovascular diseases, Université Paris Descartes, Sorbonne
- 102 Paris Cité, 56 rue Leblanc, Paris, 75015, France,
- 103 ⁴¹Paris Cardiovascular Research Center, Inserm U970, Université Paris Descartes, Sorbonne
- 104 Paris Cité, 20 rue Leblanc, Paris, 75015, France,
- 105 ⁴²Program in Medical and Population Genetics and the Cardiovascular Disease Initiative, Broad
- 106 Institute of Harvard & MIT, 75 Ames St, Cambridge, MA, 02142, USA,
- 107 ⁴³BioMe Phenomics Center, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Pl, New
- 108 York, NY, 10029, USA,
- 109 ⁴⁴Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General
- 110 Hospital, 55 Fruit Street, Boston, MA, 02114, USA,
- 111 ⁴⁵Unit of Thrombosis and Hemostasis, Hospital de la Santa Creu i Sant Pau, St Quinti 89,
- 112 Barcelona, 8041, Spain,
- 113 ⁴⁶Department of Laboratory Medicine and Pathology, University of Minnesota, 420 Delaware St.
- 114 SE, MN, 55455, USA,
- 115 ⁴⁷Clinical Epidemiology, Leiden University Medical Center, PO Box 9600, Leiden, The
- 116 Netherlands,
- 117 ⁴⁸Psychiatry, University of Pennsylvania, 3401 Walnut Street, PA, 19104, USA,
- 118 ⁴⁹Cardiovascular Health Research Unit, Departments of Biostatistics and Medicine, University of
- 119 Washington, 1730 Minor Ave, Suite 1360, Seattle, WA, 98101, USA,
- 120 ⁵⁰VA Informatics & Computing Infrastructure, VA Salt Lake City Healthcare System, 500 Foothills
- 121 Drive, UT, 01730, USA,
- 122 ⁵¹Epidemiology, University of Utah, 500 Foothills Drive, UT, 01730, USA,
- 123 ⁵²Department of Epidemiology, University of Washington, 1730 Minor Ave, Suite 1360, Seattle,
- 124 WA, 98101, USA,
- 125 ⁵³Hematology Department and Biosurgical Research Lab (Carpentier Foundation), European
- 126 Georges Pompidou Hospital, Assistance Publique Hôpitaux de Paris, 20 rue Leblanc, Paris,
- 127 75015, France,
- 128 ⁵⁴Innovative Therapies in Haemostasis, INSERM, Université de Paris, 4 avenue de l'Observatoire,
- 129 Paris, 75270, France,
- 130 ⁵⁵Department of Cardiovascular Medicine, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo,
- 131 Tokyo, 113-8655, Japan,

⁵⁶Division of Cardiovascular Medicine, Brigham & Women's Hospital, MA, USA,
⁵⁷Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Center for
Molecular Medicine, Stockholm, 17176, Sweden,
⁵⁸Department of Epidemiology, University of Washington, 3980 15th Ave NE, WA, 98195, USA,
⁵⁹Centre D'Etude du Polymorphisme Humain, Fondation Jean Dausset, 27 rue Juliette Dodu,
Paris, 75010, France,
⁶⁰Laboratory of Excellence on Medical Genomics,
⁶¹Cardiology, Medical Genetics, University of Pennsylvania, 3401 Walnut Street, PA, 19104,
USA,
⁶²Department of Epidemiology, University of Washington, 1730 Minor Ave, Suite #1360, Seattle,
WA, 98101, USA,
⁶³Department of Health Systems and Population Health, University of Washington, 1730 Minor
Ave, Suite #1360, Seattle, WA, USA,
⁶⁴Department of vascular medicine, Paris Saint-Joseph Hospital Group, University of Paris, 185
rue Raymond Losserand, Paris, 75674, France,
⁶⁵UMR1153, INSERM CRESS, 185 rue Raymond Losserand, Paris, 75674, France,
⁶⁶C2VN, INSERM, INRAE, Aix-Marseille University, 27, bd Jean Moulin, Marseille, 13385, France,
⁶⁷Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo,
Tokyo, 108-8639, Japan,
⁶⁸Center for Public Health Genomics, University of Virginia, 3242 West Complex, Charlottesville,
VA, 22908-0717, USA,
⁶⁹Division of Epidemiology and Community Health, University of Minnesota, 1300 South Second
Street, MN, 55454, USA,
⁷⁰College of Medicine, Ohio State University, 376 W. 10th Ave, Columbus, OH, 43210, USA,
⁷¹Emergency Medicine, Massachusetts General Hospital, Zero Emerson Place, Suite 3B, Boston,
MA, 02114, USA,
⁷²Emergency Medicine, Harvard Medical School, Zero Emerson Place, Suite 3B, Boston, 02114,
MA,
⁷³Department of Surgery, Perelman School of Medicine University of Pennsylvania,
Philadelphia, PA, 19104, USA

Contact: Nicholas L. Smith, Department of Epidemiology, University of Washington,
Seattle WA, 98195, USA. (nlsmith@u.washington.edu)

Keywords: venous thrombosis, venous thromboembolism, genetic, genome-wide
association studies, transcriptome-wide association study

Version: February 17, 2022

ABSTRACT

Venous thromboembolism (VTE) is a complex disease with environmental and genetic determinants. We present new cross-ancestry meta-analyzed genome-wide association study (GWAS) results from 30 studies, with replication of novel loci and their characterization through *in silico* genomic interrogations. In our initial genetic discovery effort that included 55,330 participants with VTE (47,822 European, 6,320 African, and 1,188 Hispanic ancestry), we identified 48 novel associations of which 34 replicated after correction for multiple testing. In our combined discovery-replication analysis (81,669 VTE participants) and ancestry-stratified meta-analyses (European, African and Hispanic), we identified another 44 novel associations, which are new candidate VTE-associated loci requiring replication. In total, across all GWAS meta-analyses, we identified 135 independent genomic loci significantly associated with VTE risk. We also identified 31 novel transcript associations in transcriptome-wide association studies and 8 novel candidate genes with protein QTL Mendelian randomization analyses. *In silico* interrogations of hemostasis and hematology traits and a large phenome-wide association analysis of the 135 novel GWAS loci provided insights to biological pathways contributing to VTE, indicating that some loci may contribute to VTE through well-characterized coagulation pathways while others provide new data on the role of hematology traits, particularly platelet function. Many of the replicated loci are outside of known or currently hypothesized pathways to thrombosis. In summary, these findings highlight new pathways to thrombosis and provide novel molecules that may be useful in the development of antithrombosis treatments with reduced risk of bleeds.

INTRODUCTION

Venous thrombosis is a vascular event resulting from an imbalance in the regulation of hemostasis, with subsequent pathologic coagulation and vascular thrombosis formation. Clinically, venous thrombosis can manifest as deep vein thrombosis (DVT), when occurring in the deep veins primarily of the legs and trunk, or as a pulmonary embolism (PE), when the thrombus embolizes and obstructs the pulmonary arteries. Collectively, these events are known as venous thromboembolism (VTE), a life-threatening condition with an incidence of 1-2 events per 1,000 person-years.¹⁻³ VTE is a complex disease with both environmental and genetic determinants. Family studies, candidate-gene approaches, and early genome-wide association studies (GWAS) primarily identified genetic risk factors in loci with well characterized effects on coagulation (*F2*, *F5*, *F11*, *FGG*, *ABO*, *SERPINC1*, *PROCR*, *PROC*, *PROS1*), supporting current therapeutic strategies that mainly target the coagulation cascade.⁴⁻⁸ In recent years, larger GWAS meta-analyses revealed unanticipated loci, such as *SLC44A2*,⁹ which was later characterized as a choline transporter involved in platelet activation,¹⁰ and in the adhesion and activation of neutrophils.^{11,12} Thus, genetic associations with VTE in larger and more diverse populations may uncover new biological pathways and molecular events contributing to the disease and potentially help identify novel targets for treatment. Most recently, 2 large efforts involving up to 30,000 VTE cases, led by the International Network Against Venous Thrombosis (INVENT) consortium¹³ and the Million Veteran Program¹⁴ (MVP), identified up to 43 genetic loci associated with VTE. To expand discovery of novel VTE risk loci, we conducted a large, cross-ancestry GWAS meta-analysis involving more than 80,000 VTE cases, along with a replication of novel loci and their characterization through downstream analyses.

METHODS

Design and Study Participants

The current cross-ancestry GWAS meta-analysis is comprised of new analyses of data from 13 studies, including the Department of Veterans Affairs Million Veteran Program (MVP, version 4),¹⁴ UK Biobank (UKB),^{15,16} FinnGen (freeze 5), Estonian Biobank (EGP),¹⁷ Biobank Japan (BBJ),¹⁸ Mass General Brigham biobank (MGB),¹⁹ BioMe, Penn Medicine BioBank (UPenn),

FARIVE,²⁰ MARTHA12,²¹ RETROVE,²² Multi-Ethnic Study of Atherosclerosis (MESA),²³ and GAIT2,^{24,25} as well as previously published data from the INVENT consortium, a 17 study analysis of prospective cohorts and case-control data (designated INVENT-2019).¹³ A detailed description of participating studies is provided in supplemental **Table S1**.

The study design (see **Figure 1**) included a discovery meta-analysis of GWAS summary data from 4 consortium/studies (INVENT-2019, MVP, FinnGen, EGP) followed by a replication of discovery loci that exceed the genome-wide significance threshold ($P < 5.00 \times 10^{-8}$). Included in the replication were the remaining 10 participating studies plus 2 external collaborations that provided association results for the queried variants: 23andMe²⁶ and Global Biobank Meta-Analysis Initiative (GBMI), limiting data to non-overlapping studies with our discovery.²⁷ The combined discovery and replication data (with the exception of 23andMe and GBMI) were then meta-analyzed, and ancestry-stratified meta-analyses were performed for African (AFR)-ancestry, European (EUR)-ancestry, and Hispanic (HIS) participants to enable further downstream ancestry-specific analyses, such as fine mapping. Participants from studies provided written informed consent for use of their genetic and health information for analysis, and the studies were individually approved by the appropriate Institutional Review Boards (see **Supplemental Methods**).

Study-specific GWAS

Genotyping arrays, imputation panels, and analyses performed by each participating study are detailed in supplemental **Table S1**. Briefly, studies performed association analyses (logistic regression analyses or generalized mixed models for case-control studies and Cox regression for cohort studies) using age and sex as covariates and further adjusting for participant relatedness, genetic principal components, and study site or other study-specific factors when applicable.

For each dataset, quality control was performed using EasyQC²⁸ to remove variants with missing information (effect and/or standard error), low imputation quality (< 0.3), and rare

variants (allele count < 5). For studies missing either imputation quality or variant frequency (BioMe, UPenn), a filter was added to remove variants with extreme effects ($|\text{Effect}| > 10$). Indels and marker names were then harmonized across all studies.

For the X chromosome, all studies performed sex-stratified GWAS, excluding variants from pseudo-autosomal regions. Results from males and females were then meta-analyzed.

Discovery, Replication, and Combined GWAS Meta-analyses

All GWAS meta-analyses were conducted with METAL,²⁹ using a fixed-effects inverse-variance weighted model. All variants were included and there was no lower minor allele frequency (MAF) limit beyond study-specific minor allele count. Genome-wide significant variants ($P < 5.00 \times 10^{-8}$) were kept if a concordant effect direction was observed in 2 or more studies and grouped into the same locus if they were within 1Mb. We used the closest gene to the lead variant to refer to each locus, except at known loci where the causal gene has been previously identified and is different from the closest gene (such as *PROCR* or *PROS1*). We defined a locus as novel if a genetic association with VTE has not been previously observed in the region according to our review of peer-reviewed published reports. We used PhenoGram³⁰ to visually represent the genomic position of the loci, their significance in EUR- or AFR-ancestry analyses, and their discovery status of novel or known.

Discovery Meta-Analysis: For the discovery cross-ancestry GWAS meta-analysis, we meta-analyzed data from 4 consortium/studies: INVENT-2019, MVP, FinnGen and EGCUT. Participants were adult men and women and included 55,330 VTE cases (either DVT and/or PE cases) and 1,081,973 controls of EUR, AFR, or HIS ancestries. At each locus with a genome-wide significant signal, the lead variant was extracted and tested in an independent replication meta-analysis.

Replication: The replication GWAS meta-analysis consisted of the remaining 10 participating studies, as well as 2 external collaborators (GBMI²⁷ and 23andMe²⁶), for a total of 91,230 VTE cases in replication. Replicating variants from the discovery were defined as those that had

concordant effect direction in the discovery and the replication, and reached a Bonferroni-corrected p-value threshold in the replication population corresponding to the number of variants tested for replication with a 1-sided hypothesis: $p\text{-value threshold} = [(0.05 \times 2) / \text{number of variants tested for replication}]$ in the replication analysis.

Combined GWAS Meta-Analysis and Stratification by Ancestry: We performed a combined, cross-ancestry GWAS meta-analysis of discovery and replication data using participating studies with genome-wide summary data. We included variants with $MAF \geq 0.01$ to maintain adequate statistical power by reducing the number of low-powered tests since replication was not available. Genome-wide data from GBMI and 23andMe data were not available and therefore excluded from combined analyses. We estimated the heterogeneity associated with each variant using Cochran's Q test and the corresponding I^2 statistic. We assessed the genomic inflation with the lambda genomic control.³¹ We report on variants exceeding the genome-wide threshold ($P < 5.00 \times 10^{-8}$) and view these as candidate novel loci associated with VTE and needing future replication.

We then stratified the analyses by ancestry and limited strata to EUR, AFR, and HIS ancestries as the remaining ancestries had too few VTE events to be informative: East Asian (EAS) in BBJ, $n=507$ VTE events; South Asian (SAS) in UKB, $n=189$ VTE events. As above, we estimated heterogeneity and assessed inflation with lambda genomic control; the LD-score intercept was computed for EUR-ancestry analysis, using the recommended Hapmap3 variants.³² We report all additional ancestry-specific variants exceeding the genome-wide threshold ($P < 5.00 \times 10^{-8}$) and view these as ancestry-specific candidate loci associated with VTE and needing future replication.

Ancestry-Stratified Analyses: Conditional Analyses and Fine-mapping

To estimate the presence of multiple independent signals, we performed conditional analyses with GCTA-COJO³³ at each locus with significant signals in the EUR- and AFR-ancestry GWAS meta-analyses. The Trans-Omics for Precision Medicine (TOPMed) trans-ancestry sequence

data (freeze 8) was used as reference panel, selecting only EUR-ancestry participants from TOPMed (N=34,890) for the EUR conditional analyses and AFR-ancestry participants (N=17,322) for the AFR analyses.³⁴ GCTA-COJO analyses were not performed for the HIS-ancestry stratum since a HIS reference panel was not available in TOPMed. Conditional analyses were performed at each locus, using a window that encompassed at least the genome-wide significant variants present in the locus with an additional buffer of ± 100 Kb. A stepwise joint regression model was used to identify secondary signals with joint p-values $< 5.00 \times 10^{-8}$ and a linkage disequilibrium (LD) $r^2 < 0.2$ with selected variants.

In addition, for each locus and for each ancestry-specific GWAS meta-analysis, we produced forest plots with the *forestplot* R library, and regional association plots with the *RACER*³⁵ and *LDlinkR*³⁶ R libraries, to visually inspect the local genetic architecture (available as supplemental **Figures 1-8**). We used the 1000 Genomes project EUR-ancestry dataset as reference panel to infer LD patterns for the EUR-ancestry participants and overall meta-analyses and the 1000 Genomes project AFR-ancestry reference panel for the AFR-ancestry meta-analysis. Furthermore, at each locus where distinct lead variants were identified in the different meta-analyses, we also extracted the lead variant from each analysis, as well as additional independent variants identified by the conditional analyses, and computed the LD between each variant (using both EUR- and AFR-ancestry reference panels) to verify the independence of the signals.

Transcriptome-wide Association Studies (TWAS)

We performed TWAS with the FUSION pipeline to accomplish 2 tasks: (1) prioritize genes for those genome-significant signals with ambiguous gene associations; and (2) identify new candidate loci by linking gene expression with VTE risk using GWAS results not reaching genome-wide significance.³⁷ This analysis was performed using the EUR-ancestry autosomal GWAS meta-analysis results, since FUSION depends on a EUR-ancestry LD reference panel (from 1000 genomes³⁸) and does not include data for chromosome X. We first performed a series of single-tissue TWAS using gene expression from eQTL datasets relevant to blood and thrombosis

disorders: whole blood, liver, lung, and spleen from GTEx v8,³⁹ whole blood from the Young Finn Study,⁴⁰ and peripheral blood from the Netherlands Twins Register.⁴¹ We then also employed an analysis using cross-tissue weights computed from GTEx v8 tissues, available as 3 canonical vectors (sCCA1-3) that capture most of the gene expression.⁴² All associations reaching a Bonferroni corrected significance threshold corresponding to the number of gene tested ($N=14,219$, $P < 3.52 \times 10^{-6}$) were deemed statistically significant. As several genes can be associated at the same locus, the TWAS results were subjected to a conditional analysis implemented in FUSION to select genes that remained conditionally independent. For each tissue, we further performed a colocalization test with COLOC⁴³ for all significant associations, to identify and select genetic signals shared by both VTE risk and gene expression with high posterior probability ($PP4 > 0.75$). Selected genes located farther than 200kb from genetic loci identified in the meta-analyses were considered novel candidate VTE genes.

Protein QTL Mendelian Randomization

Using the combined, cross-ancestry VTE GWAS meta-analysis results, we performed a proteome Mendelian randomization (MR) analysis with high-confidence genomic instruments corresponding to protein QTL (pQTL) for 1,216 circulating plasma proteins that passed consistency and pleiotropy filters, as previously described.⁴⁴ When 2 or more genetic instruments were available for the exposure-outcome pairs, we performed inverse variance-weighted MR. If a single genetic variant was present, Wald-ratio MR was used instead. In addition, when multiple SOMAmers for a protein were available, which was the case for 40 proteins, tests were conducted separately for each SOMAmer. To account for multiple testing, associations passing the Bonferroni corrected threshold corresponding to the number of SOMAmers tested ($N=1,256$, $P < 3.98 \times 10^{-5}$) were considered statistically significant.

Association of VTE Loci with Hemostasis and Hematology Traits

We conducted a series of *in silico* investigations involving hemostasis and hematology traits to better characterize the VTE-associated variants from the GWAS meta-analyses. To better understand if novel VTE-associated variants might operate through hemostasis pathways, we

extracted associations from published GWAS of 10 hemostatic traits: fibrinogen;⁴⁵ fibrin D-dimer;⁴⁶ coagulation factors VII (FVII),⁴⁷ VIII (FVIII),⁴⁸ and XI (FXI);⁴⁹ von Willebrand factor (vWF);⁴⁸ tissue plasminogen activator (tPA);⁵⁰ plasminogen-activator inhibitor 1 (PAI-1);⁵¹ activated partial thromboplastin time (aPTT); and prothrombin time (PT).⁵² Since each variant-association was investigated in 10 hemostasis traits, we set a p-value threshold of 0.005 (0.05/10 traits tested for each lead variant of a locus) to separate associations of interest from other associations.

Similarly, we extracted associations with complete blood count (CBC) measures using summary data from nearly 750,000 individuals on 15 leukocyte, erythrocyte, and platelet traits.⁵³ Given the large sample size and high statistical power of these analyses, we used a more stringent threshold of interest that was a Bonferroni correction corresponding to the number of look-ups performed ($P < 1.92 \times 10^{-5}$).

Phenome-wide Association Testing

To explore associations between VTE-associated variants and other traits agnostically, we performed a phenome-wide association study (PheWAS) using the MRC IEU infrastructure,⁵⁴ which included the datasets from the 1,500 UKB analyses performed by the Neale lab on 337,000 individuals of British ancestry (pheWAS source codes: ukb-a and ukb-d). Only variants of interest reaching genome-wide significance ($P < 5.00 \times 10^{-8}$) were extracted for each trait and presented.

RESULTS

Discovery Cross-Ancestry Meta-analysis and Replication

The primary cross-ancestry discovery analysis was based on 4 contributing consortium/studies (INVENT-2019, MVP, FinnGen, EGP) and included 55,330 participants among 3 ancestry groups with VTE (47,822 EUR, 6,320 AFR, and 1,188 HIS) and 1,081,973 participants without VTE (918,195 EUR, 118,144 AFR, and 45,634 HIS). Over the 22 autosomal and X chromosomes, 35.5 million variants were analyzed, and the observed genomic inflation factor was 1.06. We

identified 10,493 variants reaching genome-wide significance, corresponding to 85 loci, of which 48 have not been identified in previous genetic studies of VTE (see supplemental **Table S2**).

We tested these 85 variants for replication in 91,230 cases and 3,322,939 controls from the independent replication data. After meta-analyzing the results of these 85 tests in the replication population, we identified 83 variants with a concordant effect direction between the discovery and the replication, of which 68 replicated at the 1-sided Bonferroni corrected significance threshold ($p < 0.1/83 = 0.0012$) (**Table 1, Figure 2**, supplemental **Table S2**). The successfully replicated signals corresponded to 34 known and 34 novel loci. Among the 34 novel loci that replicated, heterogeneity was minimal (heterogeneity $P > 0.05$), odds ratios (ORs) ranged between 0.84-0.98 and 1.03-1.18, and MAFs were all ≥ 0.021 . The majority of variants were gene-centric (4 exonic, 16 intronic, and 3 in 3' or 5' UTR regions or immediately downstream), 3 were linked to intronic non-coding RNA, and 8 were considered intergenic. The novel replicated variants were at the following loci (in chromosomal order): *H6PD/SPSB1*, *TENT5C*, *TRIM58*, *CALCRL* (near *TFPI*), *CPS1*, *SERPINE2*, *PIK3CB/LINC01391*, *MECOM*, *LINC01968/XXYLT1*, *SEC31A*, *ARHGAP24*, *LNPEP*, *ILRUN*, *AGPAT5/XKR5*, *ZNF367/HABP4*, *MIR1265/FAM107B*, *ZMIZ1*, *PLCE1*, *ST3GAL4*, *A2ML1/PHC1*, *COPZ1*, *SH2B3*, *RCOR1*, *MAP1A*, *RORA-AS1*, *HSD3B7*, *ZFPM1*, *ALOX12-AS1*, *MAPT-AS1*, *CEP112/APOH*, *RFX2*, *GIPR*, *FUT2*, and *SYN3*. Among the 17 variants and their associated loci that failed replication, 14 were novel and remain candidate loci that merit additional replication while 3 were known loci: *PROS1*, *JAK2*, and *STAB2*.

Combined Cross-Ancestry GWAS Meta-analysis and Ancestry-Stratified Results

Combined: The combined, cross-ancestry meta-analysis of the studies with genome-wide markers included 81,669 individuals with VTE and 1,426,717 individuals without VTE. We analyzed 19.1 million common variants ($MAF \geq 0.01$) and observed a genomic factor of 1.16 which is slightly elevated but expected for large scale meta-analyses of polygenic traits.⁵⁵ We identified 16,550 variants reaching genome-wide significance located in 110 loci, of which 41

were not observed in the discovery analysis (supplemental **Table S3, Figure 2**). Of these 41 additional loci, 1 corresponded to a common variant at the known *SERPINC1* locus (rs6695940) which encodes antithrombin, 4 were previously identified in the INVENT-2019¹³ or MVP¹⁴ meta-analyses at the *PEPD*, *ABCA5*, *MPHOSPH9*, and *ARID4A* loci, and 1 was a known pathogenic missense variant located in *SERPINA1* (rs28929474, p.Glu366Lys).⁵⁶ The remaining 35 loci were novel associations and are presented in **Table 2**. Among the 35 candidate loci, all had ORs within the range of 0.93-0.97 and 1.03-1.15 and had a minimum MAF of 0.021. The majority of the variants were gene-centric (18 intronic and 3 in 3' UTR regions), 3 were intronic in non-coding RNA and 11 were considered intergenic. These candidate loci included (in chromosomal order): *ARID1A*, *AK5*, *CELSR2*, *FALEC/ADAMTSL4*, *DCST2*, *FCER1G*, *CFHR1/CFHR4*, *LINC01800*, *RFTN2/MARS2*, *LINC02029/LINC00880*, *AFAP1*, *REST*, *ARHGEF28*, *HSPA4*, *CARMIL1*, *HCG9*, *JAZF1-AS1*, *LOC101929128/LOC157273*, *ENTPD4/SLC25A37*, *BPMS*, *ANGPT1*, *MLLT10*, *RAB18/MKX*, *SOX6*, *MS4A4A/MS4A6E*, *ZBTB16*, *SNX19*, *LINC02324/MIR4708*, *LINC00924/LOC105369212*, *NDE1*, *DHODH/HP*, *CFDP1*, *C16orf74*, *TAOK1*, and *RAB4B*.

European Ancestry: The EUR-ancestry meta-analysis, which included 71,771 participants with VTE and 1,059,740 participants without VTE, had a lambda genomic factor of 1.22. As population stratification might be introduced by founder effects in Finnish participants from FinnGen, we did a sensitivity analysis by removing this cohort, and observed a similar genomic factor of 1.19. We also observed an LD-score intercept of 1.07, indicating an inflation mainly due to polygenic architecture. Out of 11.1 million variants analyzed, 16,867 were genome-wide significant and clustered into 100 regions, of which 7 did not overlap with loci identified in the discovery or overall meta-analysis: *FAM168A*, *AKIRIN2*, *LINC02103*, *LINC02652*, *SLC39A8*, *IRAG1*, and *RGS18* (**Table 2, Figure 2**, supplemental **Table S4**). For these 7 additional candidate loci, the ORs ranged from 0.94-0.97 to 1.04-1.07 and the minimum MAF was 0.058.

Conditional analyses were performed with GCTA COJO at each of the 100 significant loci and revealed a subset of 21 loci with multiple independent signals (supplemental **Table S5**). These included 3 within novel loci: rs39840 (intronic, in *ERAP1*), rs28712647 (upstream of *NEURL4*,

near *ALOX12*) and rs7412 (nonsynonymous, in *APOE*). At the remaining 18 known VTE loci, secondary signals at 5 loci had been previously observed, or were in LD with previously identified signals, at *F5*, *F11*, *ABO*, *F2* and *VWF*,¹³ while the rest were located at *EPHA3*, *PROS1*, *FGG*, the *HLA* region, *SCARA5*, *TSPAN15*, *FADS2B*, *FADS1*, *STAB2*, *F10*, *PLCG2*, *THBD*, and *A4GALT*.

African Ancestry: The AFR-ancestry meta-analysis included 7,482 participants with VTE and 129,975 participants without VTE from 7 cohorts and had a genomic inflation factor of 1.05. Here, 17.1 million variants were analyzed, of which 752 were genome-wide significant and located within 13 loci, of which 2 corresponded to novel ancestry-specific signals at *RBFOX1* (OR = 0.56; MAF = 0.04) and *COL6A2* (OR = 2.16; MAF = 0.011) (**Table 2, Figure 2**, supplemental **Table S6**).

Conditional analyses were performed with GCTA COJO at each of the 13 significant loci revealed 3 loci with additional independent signals (supplemental **Table S7**) at *EPHA3* (with LD $r^2 < 0.02$ with the primary and secondary EUR signals, using the 1000 Genomes AFR reference LD panel), *PROS1* (rs6795524, LD $r^2 = 0.94$ with rs28479320, the lead variant from the combined meta-analysis located in *ARL13B*), and *ABO* (several signals).

Hispanic Ancestry: The HIS-ancestry meta-analysis included 1,720 participants with VTE and 57,367 participants without VTE from 4 cohorts and had a genomic inflation factor of 1.02. We analyzed 11.1 million variants, of which 58 were genome-wide significant, all located at the *ABO* locus with rs2519093 as lead variant (MAF = 0.15, OR = 1.49, $P = 3.08 \times 10^{-15}$).

Comparison of Ancestry-Specific and Cross-Ancestry Meta-Analysis Results: We then investigated the lead variants from the AFR- and EUR-ancestry meta-analyses at the 11 loci (all known) identified in both analyses. At 5 loci (*PROC*, *EPHA3*, *PROS1*, *VWF* and *THBD*), none of the AFR lead variants were available in the EUR analyses, due to their low frequency in EUR (MAF < 0.0006 for all 5 lead variants in non-Finnish Europeans according to gnomAD⁵⁷). At the

remaining 6 loci, the lead variants from the AFR analysis were also genome wide significant in the EUR analysis, and shared similar effect sizes.

Across the discovery, combined, EUR, AFR and HIS meta-analyses, we identified 135 independent loci (**Figure 2**). A summary of each locus, including LD patterns between lead variants from each meta-analysis as well as independent signals and association test results across all meta-analyses, is available in supplemental **Table S8**.

Gene Prioritization with TWAS and Protein QTL MR

Transcriptome Wide Association Study: Across the 6 single-tissue and 3 cross-tissues datasets analyzed, we identified 166 significant ($P < 3.52 \times 10^{-6}$) and conditionally independent associations with a high posterior probability of colocalization (> 0.75) between gene expression and VTE risk (see supplemental **Table S9**). These associations involved 108 genes, of which 77 were mapped to 46 genome-wide significant GWAS loci, leaving an additional 31 novel candidate genes that mapped outside of genome-wide significant GWAS loci (supplemental **Table S10**). The candidate genes included (in chromosomal order): *PRDX6*, *LMOD1*, *FEZ2*, *PNKD*, *BAP1*, *PRKCD*, *CHST13*, *CHST13*, *HGFAC*, *NIPAL1*, *ARHGAP10*, *LHFPL2*, *DND1*, *THBS2*, *MEST*, *COPG2*, *SYK*, *AAMDC*, *STAC3*, *KIF5A*, *NFKBIA*, *CINP*, *COMMD4*, *DCTPP1*, *NFAT5*, *MPDU1*, *KDM6B*, *C18orf8*, *NPC1*, *RAB3D*, *ISOC2*, and *SDCBP2*. At 33 GWAS loci, an associated gene matched the gene closest to the lead variant, supporting a role as a causal gene, while associated genes at the remaining 13 GWAS loci may point to genes of interest for further investigations.

Protein QTL Mendelian Randomization: We performed agnostic MR of 1,216 plasma circulating pQTL using the combined VTE meta-analysis results and identified 23 proteins with a significant causal association ($P < 3.98 \times 10^{-5}$, **Figure 3**, supplemental **Table S11**). For 13 proteins, the gene coordinates matched a genome-wide significant GWAS locus and included 5 of the novel GWAS loci: *TFPI*, *ERAP2*, *TIMP3* (at the *SYN3* locus), *TIMP4* (at the *SYN2/PPARG* locus) and *ECM1* (at the *FALEC* locus). Among the 10 candidate genes mapping outside of GWAS loci, *HGFAC* and

PRDX6 were also candidate genes identified in the TWAS. The remaining 8 candidate genes were: *EFEMP1*, *LCT*, *CLPS*, *MSR1*, *LGALS3*, *CD97*, *LILRB5*, *APOL3*.

Association of VTE-associated Variants with Hemostasis and Hematology Traits

The association of any lead or conditionally independent variant at the 135 GWAS loci with hemostasis traits is presented in **Figure 4.A** and supplemental **Table S12**. Among the 92 novel (replicated and candidate) loci reported above, 18 (19%) had a variant associated with 1 or more of the 10 hemostasis traits: fibrinogen (*CPS1*, *SLC39A8*, *ARHGEF28*, *LNPEP*, *PLCE1*, *MS4A4A*, *SH2B3*, *MIR4708*, *MAP1A*, *HP*); vWF (*SLC39A8*, *LNPEP*, *ST3GAL4*, *SH2B3*, *MAP1A*, *HP*, *APOE*, *FUT2*); FVII (*XXYLT1*, *JAZF1-AS1*, *MS4A6E*, *RCOR1*, *MAP1A*); FVIII (*ST3GAL4*, *COPZ1*, *HP*, *APOE*); PAI-1 (*COPZ1*, *GIPR*); or PT (*APOH*).

Next, we investigated associations of the 135 GWAS loci with hematology traits, presented in **Figure 4.B** and supplemental **Table S13**. Across all 15 CBC measures and among the 92 novel loci, we observed at least 1 association at 55 (59%) novel (replicated and candidate) loci.

Phenome-wide Association Studies

We performed a pheWAS of lead and conditionally independent variants at the 135 significantly associated loci across 1,500 publicly available phenotypes involving European UKB participants (supplemental **Table S14**). For each trait, only genome-wide significant variants were retrieved, and we focused on traits sharing at least 10 loci with our VTE analyses (**Figure 5**, supplemental **Table S15**), which might indicate common biological pathways. Hematology traits, in particular platelet traits, shared the most loci with VTE (e.g. 33 for platelet count), consistent with our observations from the larger CBC GWAS ($n \sim 750,000$) sample (**Figure 4.B**). Several traits correspond to height and weight measurements, as well as enzymes mainly produced by the liver (e.g., albumin, sex-hormone binding globulin, or insulin growth factor-1), and plasma lipid-related traits (Apolipoprotein-A and B, HDL cholesterol, or triglycerides). Blood pressure (systolic and diastolic), glycated hemoglobin, calcium, cystatin C, and C-reactive protein levels were among additional traits sharing at least 10 loci with VTE. Few traits had a consistent

direction of effect with respect to VTE risk across shared loci (**Figure 5**). For example, out of 10 loci shared between bilirubin levels and VTE, 9 (90%) were associated with an increase of both bilirubin levels and VTE risk. For albumin levels, glycated hemoglobin, and systolic blood pressure, an opposite direction of effect between these traits and VTE risk was observed at more than 75% of shared loci.

DISCUSSION

We identified 135 independent genomic loci and 39 additional genes from TWAS and pQTL associated with an increased or decreased risk of VTE. This reflects a substantial increase in the number of validated and candidate loci for VTE risk beyond past genetic mapping efforts.^{13,14} Our results highlight genetic variation across the rare-to-common allele frequency spectrum in multiple ancestry groups and add new evidence of biologic predictors of VTE pathogenesis for further investigation. The *in silico* interrogations provide valuable clues regarding the putative causal gene at each locus and additional insights to biological pathways shared with VTE.

Biological Insights

Novel Replicated Loci: Our strongest evidence supports 34 loci with novel VTE associations. Except for *TFPI* and *SERPINE2*, the novel genetic loci were not in established VTE pathophysiology pathways. A subset of these loci (12 loci, 35%) was associated with plasma levels of the 10 hemostasis traits interrogated and most (26 loci, 76%) were associated with a hematology trait. This contrast should be interpreted with caution as statistical power for the hemostasis traits was much smaller than for the hematology traits.

While most of the novel associations reported had an OR in the range of [0.90-0.98; 1.03-1.10], we were able to identify and replicate 3 uncommon variants with larger effects: an intronic variant (MAF = 0.021) in the glycosyltransferase *ST3GAL4* (discovery OR = 1.21, replication OR = 1.18), which was also associated with increased vWF and FVIII levels, an intronic variant (MAF = 0.029) in the transcriptional co-activator *ZMIZ1* (discovery OR = 1.15, replication OR = 1.11), and an exonic variant (MAF = 0.027) in *MAP1A* (p.Pro2349Leu, discovery OR = 0.87, replication OR =

0.84), which was also associated with decreased levels of vWF and fibrinogen, and had a protective effect against VTE.

Variants associated with hemostasis traits provide clues that the causal gene at these loci might directly or indirectly perturb the coagulation cascade. For instance, *XXYL1* encodes a xylosyltransferase known to interact with coagulation factors⁵⁸ and had a nearby variant (OR discovery = 1.06, OR replication = 1.06) also associated with decreased FVII levels. Another example is *FUT2*, a fucosyltransferase gene involved in the synthesis of the H antigen, a building block for the production of antigens within the ABO blood group. *FUT2* had a downstream variant (OR discovery = 0.96, OR replication = 0.96) that was also associated with decreased vWF levels, mirroring results observed with vWF at the *ABO* locus. In addition, some variants were associated with several hematology traits, suggesting common genetic regulatory pathways affecting hematopoiesis, such as the replicated *RCOR1* signal on chromosome 14, and the candidate gene *REST* on chromosome 4 identified in the combined meta-analysis, 2 genes that form the transcriptional repressor CoREST, known to mediate hematopoiesis.⁵⁹

Among the 34 loci, 17 had TWAS evidence linking transcript expression with a gene in the locus and 3 were linked to protein measures. These results may help to prioritize biologically relevant genes for further investigations. Notably, at the *COPZ1* locus, the lead variant was associated with several CBC measures, including platelet count and red blood cell count, and the TWAS revealed an association with *NFE2*, known to regulate erythroid and megakaryocyte maturation.

Other Replicated and Non-Replicated Loci: Replicated variants included 2 rare variants at the known *EPHA3* (intergenic, MAF = 0.0024, OR = 2.40) and *FADS2B* (intronic, MAF = 0.0047, OR = 0.64) loci. Among variants that failed replication, only 1 rare variant displayed significant heterogeneity ($P = 0.0001$, *MYO16* locus), and 3 variants were located in known loci: *STAB2* was previously identified as associated with VTE in an independent gene-based study using exome sequencing,⁶⁰ *ARL13B* (near *PROS1*) was identified in the previous VTE GWAS from MVP,¹⁴ and

the *JAK2* V617F variant, which is known to increase the risk of myeloproliferative neoplasm, was recently identified as associated with VTE in an exome study of nearly 450,000 UKB participants.⁶¹ According to gnomAD,⁵⁷ the *ARL13B* variant identified is mostly observed in AFR-ancestry individuals (rs79324379, AFR MAF = 0.026 against MAF < 0.0003 in other ancestries) and was not in LD with the lead variant identified in the previous MVP GWAS (rs6795524, LD r^2 = 0.01 in AFR); nonetheless, we would need additional information to validate this locus as a truly independent signal—and not just a marker—from any strong, uncharacterized signal in *PROS1*. Similarly, the *STAB2* variant identified is mostly observed in Finns (rs142351376, Finns MAF = 0.020 against MAF < 0.0003 in other ancestries); the lack of Finns in the replication likely impaired our ability to replicate the association. Out of the other 12 failed replications, 5 involved rare variants (MAF < 0.01) that did not reach nominal significance ($P < 0.05$), while 6 of the remaining 7 common variants reached nominal significance, suggesting that these common variants might need a larger replication sample to be validated. One of these signals, located between *SYN2* and *PPARG*, was associated with the protein levels of *TIMP4* in a previous study.⁶² This protein, known to inhibit matrix metalloproteinases and involved in platelet aggregation and recruitment,⁶³ was confirmed by the pQTL MR analysis as a gene associated with VTE risk.

Novel Candidate Loci: Across the multiple interrogation approaches, we identified several score of candidate loci with evidence to support their association with VTE, though not yet replicated. This included 35 candidates from the combined GWAS, 7 candidates from the EUR-ancestry GWAS, and 2 candidates from the AFR-ancestry GWAS. Interestingly, the 2 variants (MAF 0.04 and 0.011) in the AFR-ancestry population were not present in EUR-ancestry participants and were associated with nearly 2-fold changes in risk of VTE. However, these 2 variants were only detected in a subset of studies, which included only 882 AFR-ancestry VTE cases out of 7,482, warranting additional investigations to confirm these 2 signals in *RBFOX1* (an RNA-binding protein) and *COL6A2* (a collagen-generating gene that contains several domains similar to *VWF* type A domains). For the remaining candidate GWAS loci, we saw similar attributes and

associations as we did with the replicated loci. With additional replication resources in the future, these candidates may become fully replicated genetic associations. In addition, the conditional analyses revealed independently associated variants mapping to distinct genes that may be of interest for further investigations, such as *BRD3* at the *ABO* locus, a chromatin reader known to associate with the hematopoietic transcription factor *GATA1*.⁶⁴ At the *EPHA3* locus, we also noted that the lead GWAS variant and the conditionally independent variant mapped upstream and downstream of *PROS2P*, a protein S pseudogene that might be of interest.

At these candidate loci, gene prioritized by the TWAS may also provide putative genes at these loci. For example *ZBTB7B*, a zinc-finger protein that represses the expression of extracellular matrix genes such as fibronectin and collagen⁶⁵ was identified by TWAS at the candidate locus *DCST2*. The 31 candidate genes identified in the TWAS as well as the additional 8 from the pQTL MR analyses, although lacking a significant genetic association at these loci, might indicate relevant genes for future investigations. For instance, SYK is a critical platelet-activation protein and tyrosine kinase inhibitors of SYK have been explored for platelet inhibition.^{66,67}

Clinical Implications

Current anticoagulation therapy to prevent or treat VTE operate through the modulation of proteins produced in the liver (coumarin-based therapies) or through direct inhibition of coagulation factors IIa (thrombin) and Xa. Although the safety profile of anticoagulation treatments has evolved, bleeding remains a life-threatening off-target outcome. New approaches to preventing thrombosis while minimizing bleeds are in development, including a focus on contact (intrinsic) pathway proteins factor XI, factor XII, prekallikrein, and high-molecular-weight kininogen.⁶⁸ Agnostic interrogations such as these may lead to discovery of novel proteins that “break the inexorable link between antithrombotic therapy and bleeding risk.”⁶⁹

Remarkably, the hematology traits investigations and the pheWAS established that CBC measures share a large number of loci with VTE, and platelet phenotypes in particular are the most frequent traits shared with VTE variants: 51 loci were associated with either platelet count, mean platelet volume, plateletcrit or platelet distribution width in the pheWAS, and 35 of these loci are novel, which represents more than a third of all novel genetic associations. Several loci associated with VTE harbor genes with known roles in hematopoiesis and megakaryocyte development, or platelet turnover: *ARID1A*, *REST* and its co-repressor *RCOR1*, *CDK6*, *MECOM*, *RBPM5*, *ANGPT1*, *RET*, *NFE2*, *ST3GAL4*, *SH2B3*, *ZFPM2* and *ZFPM1*,^{59,71–79} or platelet aggregation: *SLC44A2*, *VWF*, *FGG*, *GP6*, *RGS18*, *GRK5*, *PIK3CB*, *PLCE1*, *PLCG2*, *IRAG1*, *TIMP4*, *FCER1G*, and *ALOX12*.^{10,63,80–89} Altered platelet generation, turnover or reactivity may be a feature of VTE pathogenesis. For one, past prospective studies⁹⁰ and case-control studies^{91,92} suggest that enlarged platelets, as measured by MPV, are associated with VTE and VTE outcomes. Studies of platelet function measures with VTE have been less conclusive which may relate to the limitations of these studies in assessing comprehensive and standardized platelet reactivity mechanisms.^{93–95} Collectively, these results suggest that treatments inhibiting platelet activation such as aspirin might be beneficial in the prevention of VTE, although previous studies and trials on aspirin and combinations with anticoagulants offered mixed results.⁹⁶ Different antiplatelets, such as more targeted thrombin, PAR1 or PAR4 inhibitors, or intracellular PDE platelet signaling inhibitors like cilostazol, could be worthwhile for further study in VTE prevention.

Strengths and Limitations

The major strength of this genetic discovery effort is the large sample size of the genetic variation interrogations. As the largest genetic association study of VTE to date, we were able to increase statistical power compared with previous VTE GWAS meta-analysis efforts and increase our ability to detect new associations, many of which were replicated. We were also better powered to detect less common genetic variation. The cross-ancestry meta-analyses also increased our potential to discover novel genetic associations where the allele frequency was more common in some populations compared with others.

Several limitations deserve mention. Case ascertainment varied by study and some studies provided validated VTE events while others relied on information from electronic health records. Further, some studies only included hospitalized VTE events and did not capture events in the outpatient setting. These differences may have introduced some bias if case ascertainment and hospitalization status have genetic determinants. We included all VTE cases and did not stratify by provoked status in order to increase statistical power. Furthermore, many of the studies had not classified the VTE events as provoked and unprovoked. In addition, although the cross-ancestry approach provided benefits as described above, the numbers of VTE cases were not evenly distributed by ancestry, thus reducing our ability to detect ancestry-specific VTE variants in the under-represented ancestry groups with more modest case counts. Due to the diversity of imputation panels used by the participating studies, genetic variants had variable coverage across studies which weakened our power to detect associations. Another limitation of our approach that used summary GWAS statistics from meta-analyses is the absence of participant-specific genotype-level information. This required us to rely on LD information extracted from external datasets, which can result in variants being missed and LD patterns not accurately captured. This may have introduced some bias in analyses that relied on LD, such as the conditional analyses and the TWAS. Further, *in silico* work was performed using external data sets such as the hemostatic factors and hematology traits summary statistics, where the size (and statistical power) of the datasets varied greatly. Although different significance thresholds were employed for significance, this may have biased the detection of significant associations to those traits that had large sample sizes.

Conclusions

These cross-ancestry GWAS meta-analyses have provided a list of 34 loci that replicated discovery findings. Some of the novel loci may contribute to VTE through well-characterized coagulation pathways while others provide new data on the role of hematology traits, particularly platelet function. Many of the replicated loci are outside of known or currently hypothesized pathways to thrombosis. We also provided a list of 44 new candidate loci

712 including candidates from the combined cross-ancestry GWAS, from the EUR-ancestry GWAS,
713 from the AFR-ancestry GWAS, and also 39 candidate genes from the TWAS and pQTL MR. These
714 findings highlight new pathways to thrombosis and provide novel molecules that may be useful
715 in the development of antithrombosis treatment that reduce bleeding adverse occurrences.

Acknowledgments

The INVENT Consortium would like to acknowledge all the participants across studies that provided their health information to support these analyses.

The INVENT Consortium is supported in part by HL134894 and HL154385. The Analysis Commons was funded by R01HL131136. Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756.

Study acknowledges can be found in the **Supplemental Methods**.

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 Institute of Health, Department of Veterans Affairs, or the U.S. Department of Health and Human Services.

References

1. Silverstein, M. D. *et al.* Trends in the incidence of deep vein thrombosis and pulmonary embolism: a 25-year population-based study. *Arch Intern Med* **158**, 585–593 (1998).
2. Ghanima, W. *et al.* Incidence and prevalence of venous thromboembolism in Norway 2010–2017. *Thromb Res* **195**, 165–168 (2020).
3. Delluc, A. *et al.* Current incidence of venous thromboembolism and comparison with 1998: a community-based study in Western France. *Thromb Haemost* **116**, 967–974 (2016).
4. Smith, N. L. *et al.* Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *JAMA* **297**, 489–498 (2007).
5. Bezemer, I. D. *et al.* Gene variants associated with deep vein thrombosis. *JAMA* **299**, 1306–1314 (2008).
6. Heit, J. A. *et al.* A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *J. Thromb. Haemost.* **10**, 1521–1531 (2012).
7. Buil, A. *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* **115**, 4644–4650 (2010).
8. Tang, W. *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* **37**, 512–521 (2013).
9. Germain, M. *et al.* Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am. J. Hum. Genet.* **96**, 532–542 (2015).

- 757 10. Bennett, J. A. *et al.* The choline transporter Slc44a2 controls platelet activation and
758 thrombosis by regulating mitochondrial function. *Nat Commun* **11**, 3479 (2020).
- 759 11. Constantinescu-Bercu, A. *et al.* Activated $\alpha\text{IIb}\beta 3$ on platelets mediates flow-dependent
760 NETosis via SLC44A2. *Elife* **9**, e53353 (2020).
- 761 12. Zirka, G. *et al.* Impaired adhesion of neutrophils expressing Slc44a2/HNA-3b to VWF
762 protects against NETosis under venous shear rates. *Blood* **137**, 2256–2266 (2021).
- 763 13. Lindström, S. *et al.* Genomic and transcriptomic association studies identify 16 novel
764 susceptibility loci for venous thromboembolism. *Blood* **134**, 1645–1657 (2019).
- 765 14. Klarin, D. *et al.* Genome-wide association analysis of venous thromboembolism identifies
766 new risk loci and genetic overlap with arterial vascular disease. *Nat Genet* **51**, 1574–1579
767 (2019).
- 768 15. Klarin, D., Emdin, C. A., Natarajan, P., Conrad, M. F. & Kathiresan, S. Genetic Analysis of
769 Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates
770 Obesity as a Causal Risk Factor. *Circ Cardiovasc Genet* **10**, (2017).
- 771 16. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide
772 Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**, e1001779 (2015).
- 773 17. Mitt, M. *et al.* Improved imputation accuracy of rare and low-frequency variants using
774 population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet*
775 **25**, 869–876 (2017).
- 776 18. Nagai, A. *et al.* Overview of the BioBank Japan Project: Study design and profile. *J Epidemiol*
777 **27**, S2–S8 (2017).

- 778 19. Smoller, J. W. *et al.* An eMERGE Clinical Center at Partners Personalized Medicine. *J Pers*
779 *Med* **6**, E5 (2016).
- 780 20. Antoni, G. *et al.* A multi-stage multi-design strategy provides strong evidence that the BAI3
781 locus is associated with early-onset venous thromboembolism. *J Thromb Haemost* **8**, 2671–
782 2679 (2010).
- 783 21. Ibrahim-Kosta, M. *et al.* Minor allele of the factor V K858R variant protects from venous
784 thrombosis only in non-carriers of factor V Leiden mutation. *Sci Rep* **9**, 3750 (2019).
- 785 22. Vázquez-Santiago, M. *et al.* Short closure time values in PFA-100® are related to venous
786 thrombotic risk. Results from the RETROVE Study. *Thromb Res* **169**, 57–63 (2018).
- 787 23. Bild, D. E. *et al.* Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol*
788 **156**, 871–881 (2002).
- 789 24. Souto, J. C. *et al.* Genetic determinants of hemostasis phenotypes in Spanish families.
790 *Circulation* **101**, 1546–1551 (2000).
- 791 25. Souto, J. C. *et al.* Genetic susceptibility to thrombosis and its relationship to physiological
792 risk factors: the GAIT study. Genetic Analysis of Idiopathic Thrombophilia. *Am J Hum Genet*
793 **67**, 1452–1459 (2000).
- 794 26. Hinds, D. A. *et al.* Genome-wide association analysis of self-reported events in 6135
795 individuals and 252 827 controls identifies 8 loci associated with thrombosis. *Hum Mol*
796 *Genet* **25**, 1867–1874 (2016).
- 797 27. Zhou, W. *et al.* *Global Biobank Meta-analysis Initiative: powering genetic discovery across*
798 *human diseases*. 2021.11.19.21266436

799 <https://www.medrxiv.org/content/10.1101/2021.11.19.21266436v1> (2021)

800 doi:10.1101/2021.11.19.21266436.

801 28. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-

802 analyses. *Nat Protoc* **9**, 1192–1212 (2014).

803 29. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide

804 association scans. *Bioinformatics* **26**, 2190–2191 (2010).

805 30. Wolfe, D., Dudek, S., Ritchie, M. D. & Pendergrass, S. A. Visualizing genomic information

806 across chromosomes with PhenoGram. *BioData Min* **6**, 18 (2013).

807 31. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004

808 (1999).

809 32. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in

810 genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).

811 33. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics

812 identifies additional variants influencing complex traits. *Nat Genet* **44**, 369–375, S1-3

813 (2012).

814 34. Kowalski, M. H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed)

815 Consortium whole genome sequences improves imputation quality and detection of rare

816 variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet* **15**,

817 e1008500 (2019).

818 35. Sabik, O. L. & Farber, C. R. *RACER: A data visualization strategy for exploring multiple*

819 *genetic associations*. 495366 <https://www.biorxiv.org/content/10.1101/495366v3> (2018)

820 doi:10.1101/495366.

- 821 36. Myers, T. A., Chanock, S. J. & Machiela, M. J. LDlinkR: An R Package for Rapidly Calculating
822 Linkage Disequilibrium Statistics in Diverse Populations. *Front Genet* **11**, 157 (2020).
- 823 37. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association
824 studies. *Nat Genet* **48**, 245–252 (2016).
- 825 38. The 1000 Genomes Project Consortium. A global reference for human genetic variation.
826 *Nature* **526**, 68–74 (2015).
- 827 39. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis:
828 multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
- 829 40. Nuotio, J. *et al.* Cardiovascular risk factors in 2011 and secular trends since 2007: the
830 Cardiovascular Risk in Young Finns Study. *Scand J Public Health* **42**, 563–571 (2014).
- 831 41. Wright, F. A. *et al.* Heritability and genomics of gene expression in peripheral blood. *Nat*
832 *Genet* **46**, 430–437 (2014).
- 833 42. Feng, H. *et al.* Leveraging expression from multiple tissues using sparse canonical
834 correlation analysis and aggregate tests improves the power of transcriptome-wide
835 association studies. *PLOS Genetics* **17**, e1008973 (2021).
- 836 43. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic
837 Association Studies Using Summary Statistics. *PLOS Genetics* **10**, e1004383 (2014).
- 838 44. Zhao, H. *et al.* Proteome-wide Mendelian randomization in global biobank meta-analysis
839 reveals multi-ancestry drug targets for common diseases. 2022.01.09.21268473 (2022)
840 doi:10.1101/2022.01.09.21268473.
- 841 45. de Vries, P. S. *et al.* A meta-analysis of 120 246 individuals identifies 18 new loci for
842 fibrinogen concentration. *Hum. Mol. Genet.* **25**, 358–370 (2016).

- 843 46. Smith, N. L. *et al.* Genetic predictors of fibrin D-dimer levels in healthy adults. *Circulation*
844 **123**, 1864–1872 (2011).
- 845 47. de Vries, P. S. *et al.* A genome-wide association study identifies new loci for factor VII and
846 implicates factor VII in ischemic stroke etiology. *Blood* **133**, 967–977 (2019).
- 847 48. Sabater-Lleal, M. *et al.* Genome-Wide Association Transethnic Meta-Analyses Identifies
848 Novel Associations Regulating Coagulation Factor VIII and von Willebrand Factor Plasma
849 Levels. *Circulation* **139**, 620–635 (2019).
- 850 49. Sennblad, B. *et al.* Genome-wide association study with additional genetic and post-
851 transcriptional analyses reveals novel regulators of plasma factor XI levels. *Hum. Mol.*
852 *Genet.* **26**, 637–649 (2017).
- 853 50. Huang, J. *et al.* Genome-wide association study for circulating tissue plasminogen activator
854 levels and functional follow-up implicates endothelial STXBP5 and STX2. *Arterioscler.*
855 *Thromb. Vasc. Biol.* **34**, 1093–1101 (2014).
- 856 51. Huang, J. *et al.* Genome-wide association study for circulating levels of PAI-1 provides novel
857 insights into its regulation. *Blood* **120**, 4873–4881 (2012).
- 858 52. Tang, W. *et al.* Genetic associations for activated partial thromboplastin time and
859 prothrombin time, their gene expression profiles, and risk of coronary artery disease. *Am J*
860 *Hum Genet* **91**, 152–162 (2012).
- 861 53. Chen, M.-H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667
862 Individuals from 5 Global Populations. *Cell* **182**, 1198–1213.e14 (2020).

- 863 54. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. 2020.08.10.244293
864 <https://www.biorxiv.org/content/10.1101/2020.08.10.244293v1> (2020)
865 doi:10.1101/2020.08.10.244293.
- 866 55. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *European Journal of*
867 *Human Genetics* **19**, 807 (2011).
- 868 56. Riis, J., Nordestgaard, B. G. & Afzal, S. α 1 -Antitrypsin Z allele and risk of venous
869 thromboembolism in the general population. *J Thromb Haemost* **20**, 115–125 (2022).
- 870 57. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in
871 141,456 humans. *Nature* **581**, 434–443 (2020).
- 872 58. Minamida, S. *et al.* Detection of UDP-D-xylose: alpha-D-xyloside alpha 1--
873 >3xylosyltransferase activity in human hepatoma cell line HepG2. *J Biochem* **120**, 1002–
874 1006 (1996).
- 875 59. Saleque, S., Kim, J., Rooke, H. M. & Orkin, S. H. Epigenetic regulation of hematopoietic
876 differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1. *Mol Cell*
877 **27**, 562–572 (2007).
- 878 60. Desch, K. C. *et al.* Whole-exome sequencing identifies rare variants in STAB2 associated
879 with venous thromboembolic disease. *Blood* **136**, 533–541 (2020).
- 880 61. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants.
881 *Nature* **599**, 628–634 (2021).
- 882 62. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
- 883 63. Radomski, A. *et al.* Identification, regulation and role of tissue inhibitor of
884 metalloproteinases-4 (TIMP-4) in human platelets. *Br J Pharmacol* **137**, 1330–1338 (2002).

885 64. Lamonica, J. M. *et al.* Bromodomain protein Brd3 associates with acetylated GATA1 to
886 promote its chromatin occupancy at erythroid target genes. *Proc Natl Acad Sci U S A* **108**,
887 E159-168 (2011).

888 65. Widom, R. L., Lee, J. Y., Joseph, C., Gordon-Froome, I. & Korn, J. H. The hckrox gene family
889 regulates multiple extracellular matrix genes. *Matrix Biol* **20**, 451–462 (2001).

890 66. Perrella, G. *et al.* Role of Tyrosine Kinase Syk in Thrombus Stabilisation at High Shear. *Int J*
891 *Mol Sci* **23**, 493 (2022).

892 67. Zheng, T. J. *et al.* Assessment of the effects of Syk and BTK inhibitors on GPVI-mediated
893 platelet signaling and function. *Am J Physiol Cell Physiol* **320**, C902–C915 (2021).

894 68. Fredenburgh, J. C. & Weitz, J. I. New anticoagulants: Moving beyond the direct oral
895 anticoagulants. *J Thromb Haemost* **19**, 20–29 (2021).

896 69. Mackman, N., Bergmeier, W., Stouffer, G. A. & Weitz, J. I. Therapeutic strategies for
897 thrombosis: new targets and approaches. *Nat Rev Drug Discov* **19**, 333–352 (2020).

898 70. Hoffman, M. Coagulation in Liver Disease. *Semin Thromb Hemost* **41**, 447–454 (2015).

899 71. Han, L. *et al.* Chromatin remodeling mediated by ARID1A is indispensable for normal
900 hematopoiesis in mice. *Leukemia* **33**, 2291–2305 (2019).

901 72. Scheicher, R. *et al.* CDK6 as a key regulator of hematopoietic and leukemic stem cell
902 activation. *Blood* **125**, 90–101 (2015).

903 73. Schulze, H. & Shivdasani, R. A. Mechanisms of thrombopoiesis. *J Thromb Haemost* **3**, 1717–
904 1724 (2005).

- 905 74. Maslah, N., Cassinat, B., Verger, E., Kiladjian, J.-J. & Velazquez, L. The role of LNK/SH2B3
906 genetic alterations in myeloproliferative neoplasms and other hematological disorders.
907 *Leukemia* **31**, 1661–1670 (2017).
- 908 75. Mancini, E. *et al.* FOG-1 and GATA-1 act sequentially to specify definitive megakaryocytic
909 and erythroid progenitors. *EMBO J* **31**, 351–365 (2012).
- 910 76. Kros, J. *et al.* A mutant allele of the Swi/Snf member BAF250a determines the pool size of
911 fetal liver hemopoietic stem cell populations. *Blood* **116**, 1678–1684 (2010).
- 912 77. Ayoub, E. *et al.* EVI1 overexpression reprograms hematopoiesis via upregulation of Spi1
913 transcription. *Nat Commun* **9**, 4239 (2018).
- 914 78. Fonseca-Pereira, D. *et al.* The neurotrophic factor receptor RET drives haematopoietic stem
915 cell survival and function. *Nature* **514**, 98–101 (2014).
- 916 79. Gregory, G. D. *et al.* FOG1 requires NuRD to promote hematopoiesis and maintain lineage
917 fidelity within the megakaryocytic-erythroid compartment. *Blood* **115**, 2156–2166 (2010).
- 918 80. Keramati, A. R. *et al.* Genome sequencing unveils a regulatory landscape of platelet
919 reactivity. *Nat Commun* **12**, 3626 (2021).
- 920 81. Mitsui, T. *et al.* ALOX12 mutation in a family with dominantly inherited bleeding diathesis. *J*
921 *Hum Genet* **66**, 753–759 (2021).
- 922 82. Fukami, K. Structure, regulation, and function of phospholipase C isozymes. *J Biochem* **131**,
923 293–299 (2002).
- 924 83. Johnson, A. D. *et al.* Genome-wide meta-analyses identifies seven loci associated with
925 platelet aggregation in response to agonists. *Nat Genet* **42**, 608–613 (2010).

926 84. Moore, S. F., Smith, N. R., Blair, T. A., Durrant, T. N. & Hers, I. Critical roles for the
927 phosphatidylinositide 3-kinase isoforms p110 β and p110 γ in thrombopoietin-mediated
928 priming of platelet function. *Sci Rep* **9**, 1468 (2019).

929 85. Kuijpers, M. J. E. *et al.* Platelet CD40L Modulates Thrombus Growth Via Phosphatidylinositol
930 3-Kinase β , and Not Via CD40 and I κ B Kinase α . *Arterioscler Thromb Vasc Biol* **35**, 1374–
931 1381 (2015).

932 86. Rodriguez, B. A. T. *et al.* A Platelet Function Modulator of Thrombin Activation Is Causally
933 Linked to Cardiovascular Disease and Affects PAR4 Receptor Signaling. *Am J Hum Genet*
934 **107**, 211–221 (2020).

935 87. Antl, M. *et al.* IRAG mediates NO/cGMP-dependent inhibition of platelet aggregation and
936 thrombus formation. *Blood* **109**, 552–559 (2007).

937 88. Schinner, E., Salb, K. & Schlossmann, J. Signaling via IRAG is essential for NO/cGMP-
938 dependent inhibition of platelet activation. *Platelets* **22**, 217–227 (2011).

939 89. van Geffen, J. P. *et al.* High-throughput elucidation of thrombus formation reveals sources
940 of platelet function variability. *Haematologica* **104**, 1256–1267 (2019).

941 90. Braekkan, S. K. *et al.* Mean platelet volume is a risk factor for venous thromboembolism:
942 the Tromsø Study, Tromsø, Norway. *J Thromb Haemost* **8**, 157–162 (2010).

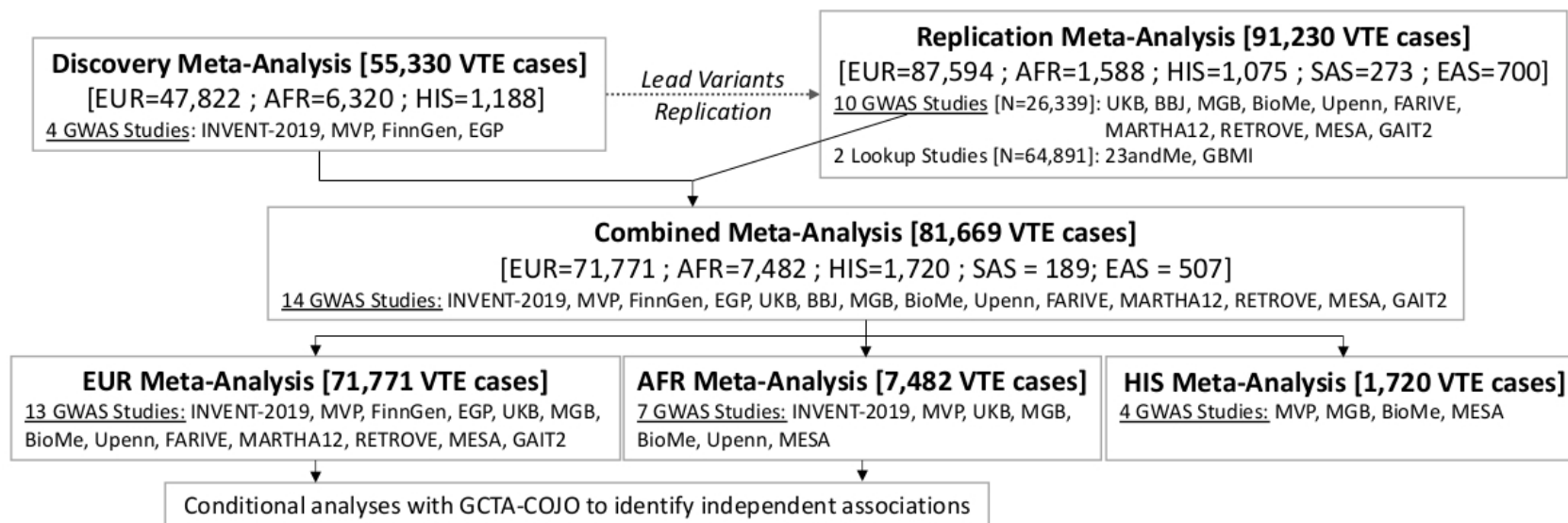
943 91. Ghaffari, S. *et al.* Prognostic value of platelet indices in patients with acute pulmonary
944 thromboembolism. *J Cardiovasc Thorac Res* **12**, 56–62 (2020).

945 92. Farah, R., Nseir, W., Kagansky, D. & Khamisy-Farah, R. The role of neutrophil-lymphocyte
946 ratio, and mean platelet volume in detecting patients with acute venous thromboembolism.
947 *J Clin Lab Anal* **34**, e23010 (2020).

93. Puurunen, M. K., Hwang, S.-J., O'Donnell, C. J., Tofler, G. & Johnson, A. D. Platelet function as a risk factor for venous thromboembolism in the Framingham Heart Study. *Thromb Res* **151**, 57–62 (2017).
94. Sokol, J., Skerenova, M., Ivankova, J., Simurda, T. & Stasko, J. Association of Genetic Variability in Selected Genes in Patients With Deep Vein Thrombosis and Platelet Hyperaggregability. *Clin Appl Thromb Hemost* **24**, 1027–1032 (2018).
95. Panova-Noeva, M. *et al.* Comprehensive platelet phenotyping supports the role of platelets in the pathogenesis of acute venous thromboembolism - results from clinical observation studies. *EBioMedicine* **60**, 102978 (2020).
96. Diep, R. & Garcia, D. Does aspirin prevent venous thromboembolism? *Hematology Am Soc Hematol Educ Program* **2020**, 634–641 (2020).

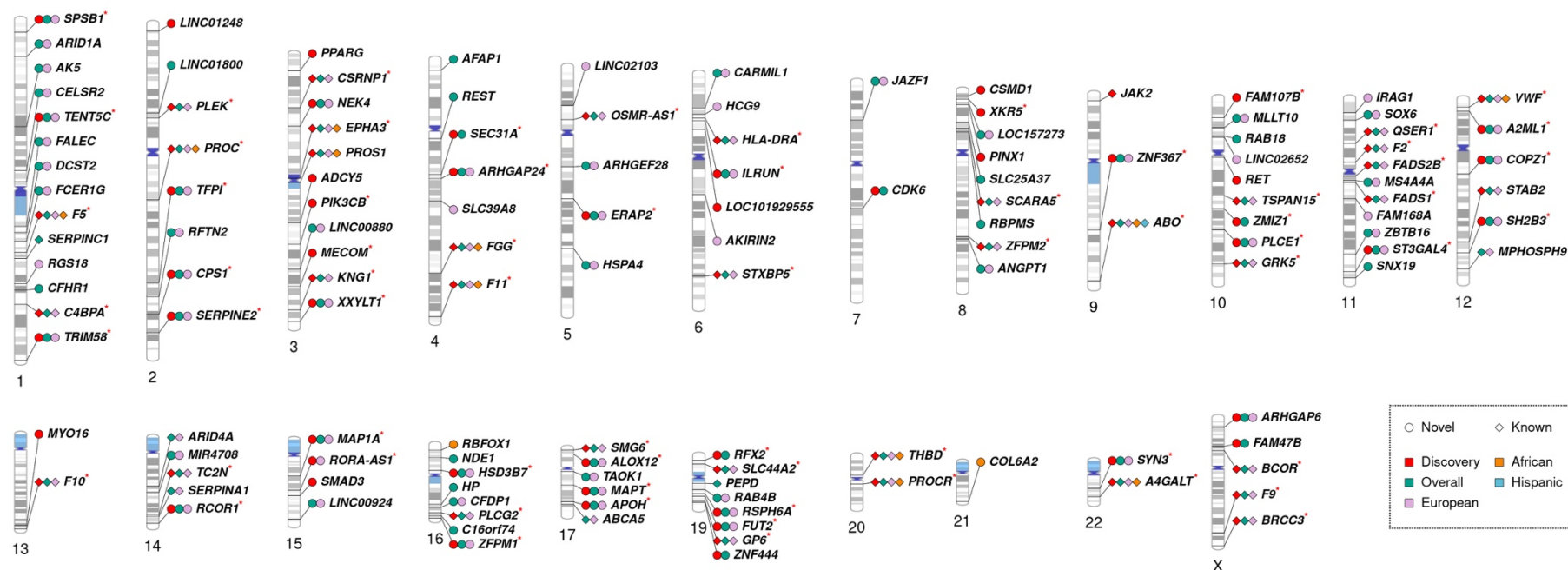
Figures and legends

Figure 1: Analyses Workflow



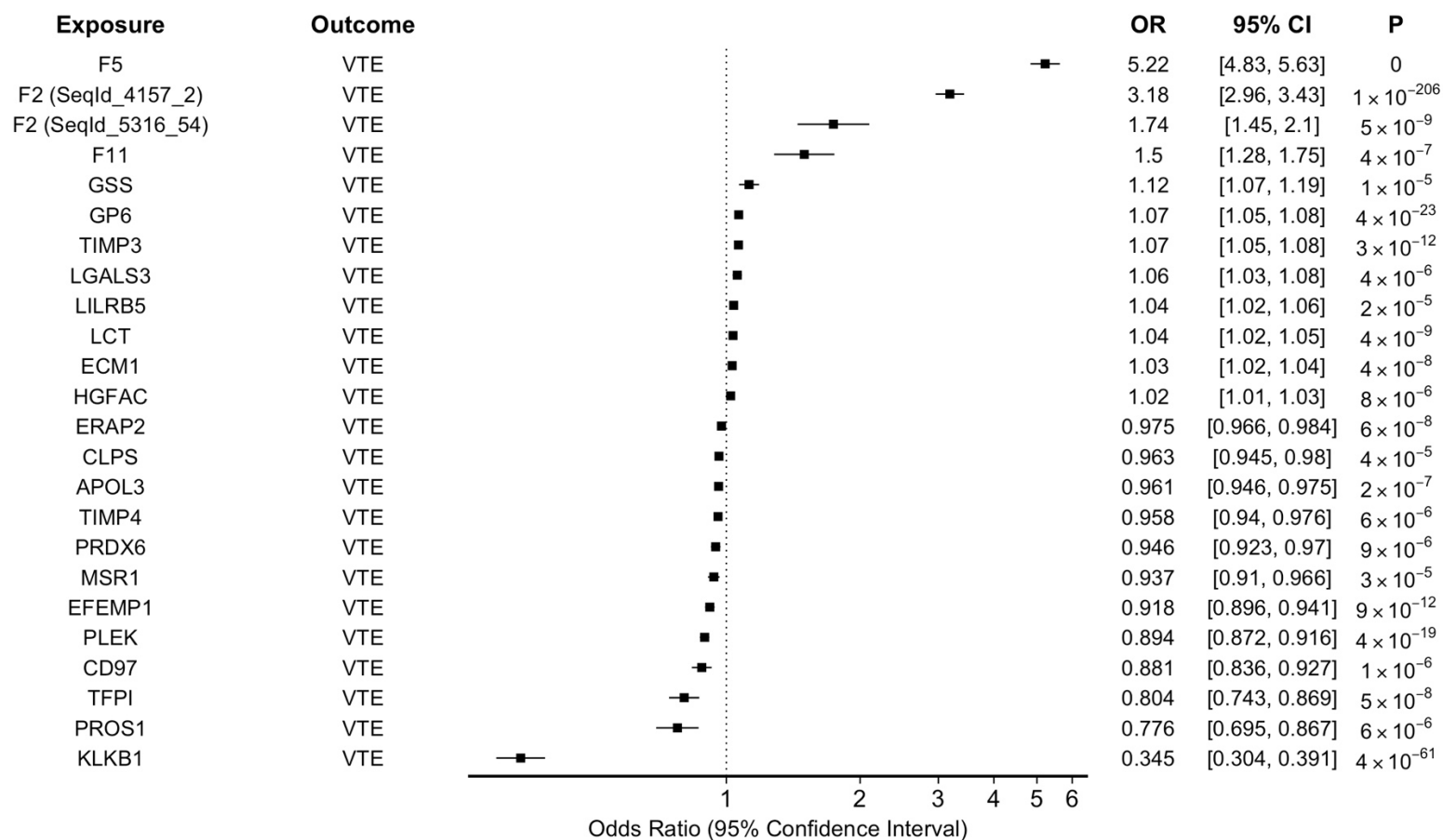
Workflow of genetic analyses conducted for this study.

987 Figure 2: Genetic loci associated with VTE



988 This figure presents the 130 loci significantly associated with VTE identified across all 4 meta-analyses: the Discovery (in blue), the
 989 overall meta-analysis (in green), the analysis restricted to individuals of European ancestry (in red) and the analysis restricted to
 990 individuals of African ancestry (in black). Novel loci are represented with circles and known loci with diamonds.
 991

997 Figure 3: Significant associations of protein QTL Mendelian Randomization

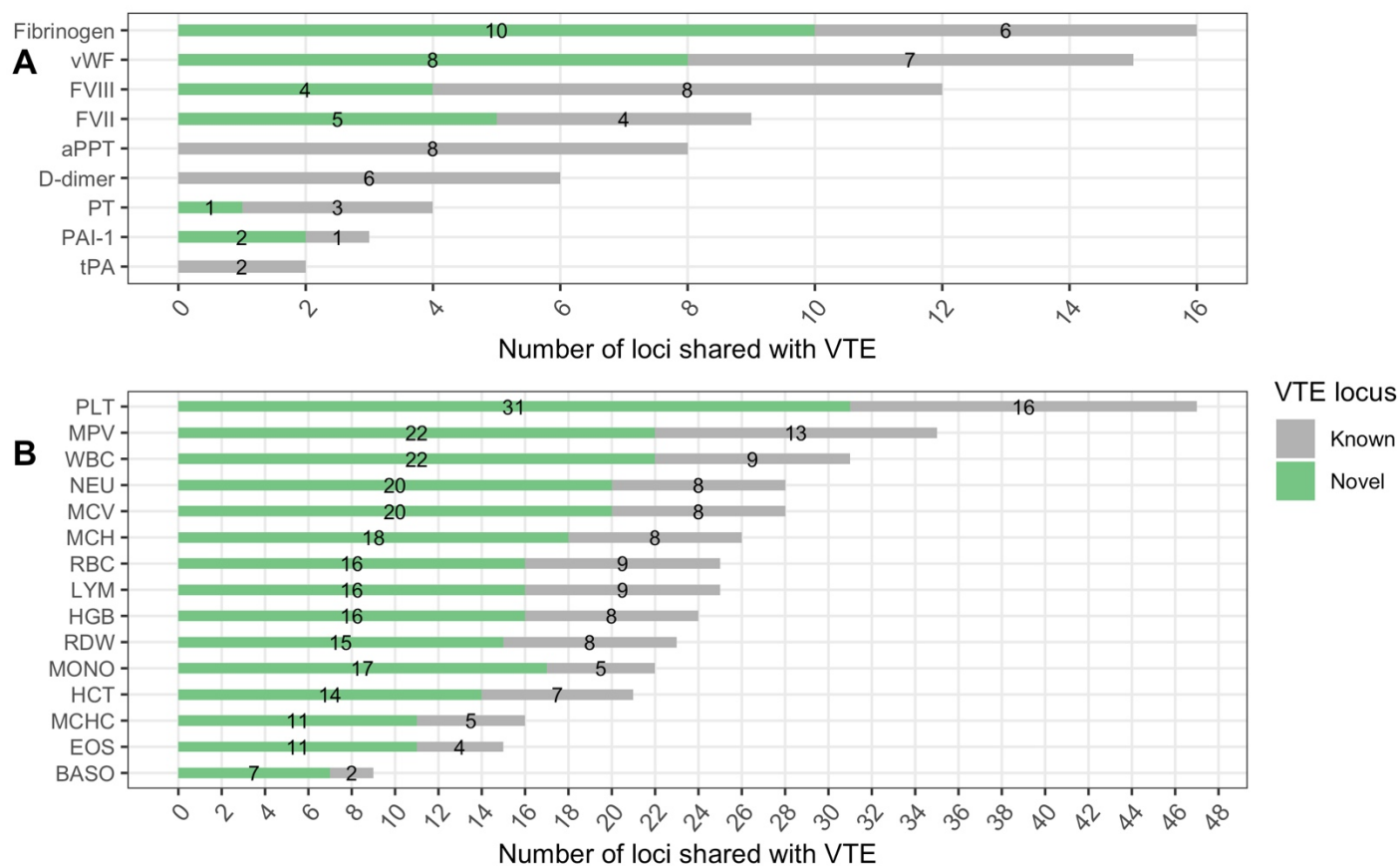


998
 999 23 genes with significant causal associations with VTE, out of 1,216 plasma protein analyzed, using the combined VTE summary
 1000 statistics.

1001

1002

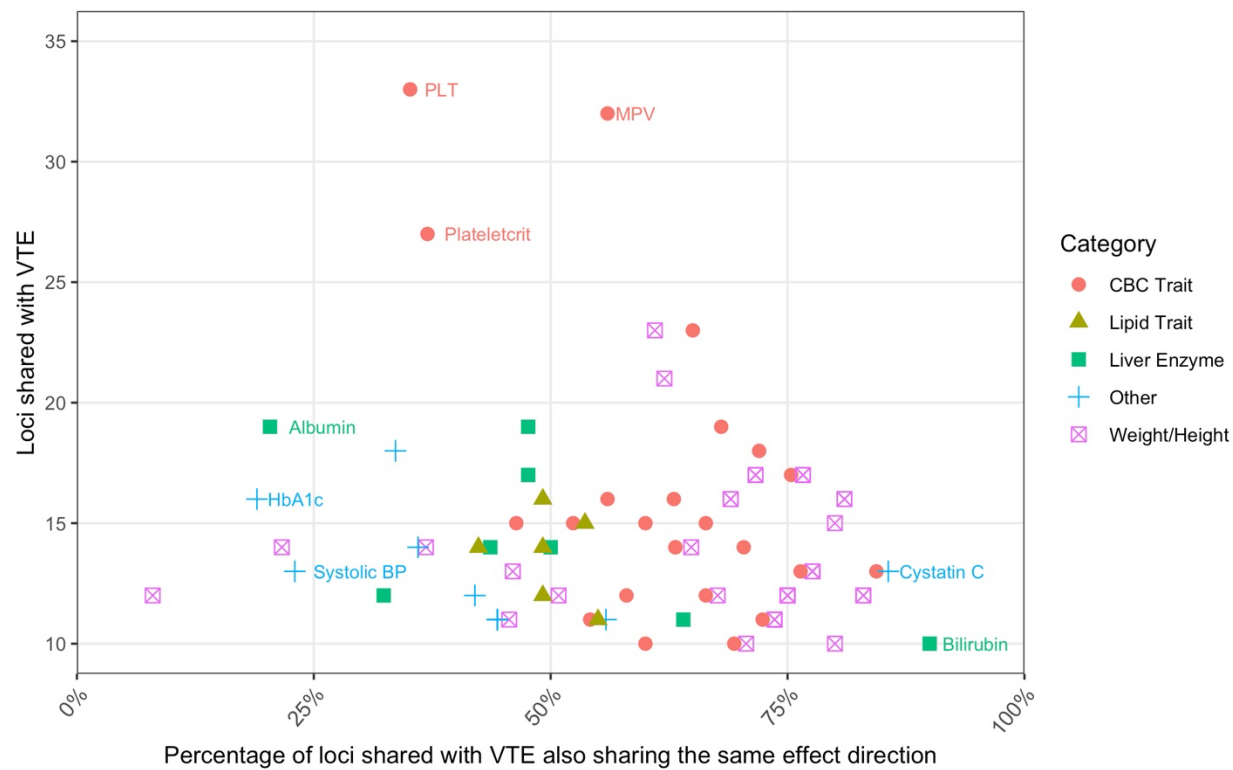
1003 Figure 4: VTE genetic loci shared with hemostatic factors and blood traits



1004

1005 (A) Number of known and novel VTE loci shared with each of the 10 hemostatic factors investigated. (B) Same analysis with
 1006 complete blood count traits: PLT (platelet count), MPV (mean platelet volume), RBC (red blood cell count), MCV (mean corpuscular
 1007 volume), HCT (hematocrit), MCH (mean corpuscular hemoglobin), MCHC (MCH concentration), HGB (hemoglobin concentration),
 1008 RDW (red cell distribution width), WBC (white blood cell count), MONO (monocyte count), NEU (neutrophil count), EOS (eosinophil
 1009 count), BASO (basophil count), LYM (lymphocyte count).

1010 Figure 5: PheWAS traits sharing at least 10 loci with VTE



1011

1012 This figure presents the pheWAS traits sharing at least 10 loci with VTE. Shape and color represent one of 5 categories: Complete
 1013 Blood Count (CBC) traits, lipid traits, liver enzyme, height and weight traits, or other (if the trait did not fit in one of the
 1014 aforementioned categories). The x-axis indicates the number of loci shared between VTE and the pheWAS trait, while the y-axis
 1015 indicates the proportion of loci where the direction of effect was identical between the pheWAS trait and VTE. As a result, traits
 1016 close to 100% have the same direction of effect than VTE at most shared loci, while traits close to 0% have an opposite direction
 1017 than VTE at most shared loci

1018 Tables

1019

1020

1021 Table 1: 68 Lead variants from the Discovery meta-analysis successfully replicated

1022

rsID	CHR.POS:EA:NEA	EA.F.Disc	OR.Disc	P.Disc	OR.Repl	P.Repl	Locus.Context	Locus.Gene
rs9442580	1:9339467:T:C	0.1551	1.06	1.83E-08	1.03	9.70E-05	intergenic	H6PD;SPSB1
rs3767812	1:118155620:A:G	0.2437	1.05	9.64E-11	1.06	1.03E-20	intronic	TENT5C
rs6025	1:169519049:T:C	0.0259	3.02	8.40E-811	3.59	9.29E-3103	exonic	F5 (p.Q534Q)
rs2842700	1:207282149:A:C	0.1092	1.11	5.95E-17	1.12	1.19E-25	intronic	C4BPA
rs3811444	1:248039451:T:C	0.3324	0.96	5.70E-09	0.95	1.53E-20	exonic	TRIM58 (p.T374M)
rs7600986	2:68636923:A:T	0.2819	1.06	3.54E-12	1.05	9.18E-19	intergenic	PLEK;FBXO48
rs182293241	2:128029746:A:G	0.0195	1.89	1.83E-27	1.55	0.0001063	intronic	ERCC3
rs6719550	2:188272460:T:C	0.6639	1.04	7.56E-09	1.05	1.93E-17	intronic	CALCRL
rs715	2:211543055:T:C	0.7022	0.95	3.51E-09	0.95	1.43E-17	UTR3	CPS1
rs13412535	2:224874874:A:G	0.2047	1.06	3.05E-10	1.08	1.10E-36	intronic	SERPINE2
rs13084580	3:39188182:T:C	0.1076	1.09	2.89E-15	1.08	9.10E-22	exonic	CSRNP1 (p.G18S)
rs562281690	3:90177913:T:G	0.0024	2.01	6.45E-15	2.40	8.68E-31	intergenic	EPHA3;NONE
rs62282204	3:138584405:T:C	0.5784	0.96	1.87E-08	0.98	6.73E-05	intergenic	PIK3CB;LINC01391
rs7613621	3:169191186:A:G	0.4467	1.04	3.21E-09	1.03	5.33E-09	intronic	MECOM
rs710446	3:186459927:T:C	0.5799	0.96	5.92E-11	0.96	1.41E-16	exonic	KNG1 (p.I581I)
rs6797948	3:194784705:T:C	0.7983	1.06	2.99E-11	1.05	7.59E-16	intergenic	LINC01968;XXYLT1
rs6826579	4:83785031:T:C	0.7914	1.05	2.38E-08	1.03	2.44E-07	intronic	SEC31A
rs17010957	4:86719165:T:C	0.8581	1.06	3.99E-09	1.05	1.00E-11	intronic	ARHGAP24
rs2066864	4:155525695:A:G	0.2585	1.23	1.98E-172	1.23	1.94E-284	UTR3	FGG
rs3756011	4:187206249:A:C	0.3903	1.23	7.48E-198	1.24	9.26E-398	intronic	F11
rs16867574	5:38708554:T:C	0.6673	0.95	2.78E-11	0.95	5.67E-16	ncRNA_intronic	OSMR-AS1
rs38032	5:96321887:T:C	0.6049	1.04	8.74E-09	1.03	1.49E-09	intronic	LNPEP
rs9268881	6:32431606:A:T	0.5727	0.96	4.17E-10	0.97	6.73E-09	intergenic	HLA-DRA;HLA-DRB5
rs145294670	6:34622561:A:AG	0.1385	1.06	6.11E-10	1.04	6.89E-06	intronic	ILRUN
rs9390460	6:147694334:T:C	0.4957	0.95	2.49E-13	0.95	1.01E-20	intronic	STXBP5
rs67694436	8:6654220:T:C	0.3486	0.96	3.94E-08	0.98	0.0001105	intergenic	AGPAT5;XKR5
rs2685417	8:27807434:C:G	0.2562	1.06	1.57E-14	1.06	2.84E-25	intronic	SCARA5
rs6993770	8:106581528:A:T	0.7142	1.08	4.48E-25	1.09	3.55E-48	intronic	ZFPM2
rs35208412	9:99194509:A:AT	0.8298	1.09	1.56E-08	1.04	5.54E-06	intergenic	ZNF367;HABP4
rs505922	9:136149229:T:C	0.6334	0.74	1.11E-425	0.69	1.55E-1043	intronic	ABO
rs1887091	10:14535113:T:C	0.4936	0.96	4.77E-08	0.98	0.001107	intergenic	MIR1265;FAM107B
rs17490626	10:71218646:C:G	0.1136	0.80	1.02E-79	0.80	3.23E-160	intronic	TSPAN15
rs16937003	10:80938499:A:G	0.0287	1.15	1.07E-08	1.11	2.11E-11	intronic	ZMIZ1
rs2274224	10:96039597:C:G	0.4414	1.04	2.55E-09	1.03	1.29E-10	exonic	PLCE1 (p.R1267P)
rs10886430	10:121010256:A:G	0.8897	0.89	7.34E-25	0.88	2.76E-64	intronic	GRK5
rs11032074	11:32993887:A:G	0.7792	1.05	5.37E-09	1.03	3.24E-06	intronic	QSER1
rs1799963	11:46761055:A:G	0.0136	2.05	2.19E-135	2.09	6.86E-420	UTR3	F2
rs141687379	11:56666822:A:G	0.9953	0.52	3.56E-31	0.64	1.06E-42	intronic	FADS2B
rs174551	11:61573684:T:C	0.6583	1.07	1.65E-19	1.07	4.90E-35	intronic	FADS1
rs35257264	11:126296816:T:C	0.0212	1.21	2.88E-14	1.18	2.28E-24	intronic	ST3GAL4
rs1558519	12:6153738:A:G	0.6175	0.93	7.73E-24	0.92	1.42E-55	intronic	VWF
rs7311483	12:9053661:T:C	0.3589	0.96	2.74E-09	0.97	2.73E-07	intergenic	A2ML1;PHC1
rs6580981	12:54723028:A:G	0.5081	0.96	3.71E-09	0.95	2.26E-23	intronic	COP21
rs3184504	12:111884608:T:C	0.4520	1.05	1.18E-11	1.04	3.30E-12	exonic	SH2B3 (p.T178T)
rs3211752	13:113787459:A:G	0.5527	0.95	1.69E-12	0.94	3.49E-25	intronic	F10
rs57035593	14:92268096:T:C	0.3202	1.07	1.08E-20	1.07	2.64E-38	intronic	TC2N
rs8013957	14:103140254:T:C	0.3699	1.04	5.33E-09	1.03	2.23E-07	intronic	RCOR1
rs55707100	15:43820717:T:C	0.0270	0.87	2.90E-08	0.84	2.49E-27	exonic	MAP1A (p.P2349L)

VTE GWAS

rs59442804	15:60899031:G:GAAAT	0.6438	0.96	4.67E-08	0.97	5.42E-10	ncRNA_intronic	RORA-AS1
rs12443808	16:30996871:C:G	0.4668	1.06	3.89E-14	1.03	1.85E-07	UTR5	HSD3B7
rs56943275	16:81898152:T:G	0.2446	1.08	4.15E-13	1.07	1.20E-26	intronic	<i>PLCG2</i>
rs28634651	16:88553198:T:C	0.6191	1.06	9.20E-13	1.04	7.62E-14	intronic	ZFPM1
rs6503222	17:1977862:A:G	0.6188	1.05	1.59E-12	1.04	5.21E-06	intronic	<i>SMG6</i>
rs7225756	17:6893691:A:G	0.4877	0.96	3.57E-08	0.98	1.20E-06	ncRNA_intronic	ALOX12-AS1
rs62054822	17:43927708:A:G	0.8028	0.95	6.39E-09	0.95	7.11E-19	ncRNA_intronic	MAPT-AS1
rs142140545	17:64191540:CTATT:C	0.1169	0.93	2.27E-08	0.95	7.83E-07	intergenic	CEP112;APOH
rs59277920	19:6077231:A:G	0.8210	0.94	1.47E-09	0.96	8.52E-06	intronic	RFX2
rs8110055	19:10739143:A:C	0.2000	0.89	5.36E-44	0.89	6.50E-70	intronic	<i>SLC44A2</i>
rs34783010	19:46180414:T:G	0.2132	0.95	3.25E-09	0.96	4.87E-10	intronic	GIPR
rs1688264	19:49209560:T:G	0.5341	0.96	2.07E-10	0.96	3.02E-15	downstream	FUT2
rs1654425	19:55538980:T:C	0.1468	0.91	2.65E-18	0.94	4.21E-14	exonic	<i>GP6 (p.S192S)</i>
rs79388863	20:23168500:A:G	0.1521	0.92	1.74E-18	0.92	4.48E-27	intergenic	<i>LINC00656;NXT1</i>
rs6060288	20:33772243:A:G	0.3417	1.12	8.19E-54	1.13	1.52E-102	intronic	<i>MMP24-AS1-EDEM2</i>
rs4820093	22:33160208:T:C	0.2693	1.05	1.04E-08	1.04	5.39E-14	intronic	SYN3
rs9611844	22:43115776:C:G	0.1286	1.10	2.09E-21	1.07	7.54E-20	intronic	<i>A4GALT</i>
rs3002416	23:39710195:T:C	0.3638	0.95	2.20E-18	0.93	2.23E-23	intergenic	<i>MIR1587;BCOR</i>
rs6048	23:138633280:A:G	0.7215	1.07	1.09E-25	1.08	1.59E-46	exonic	<i>F9 (p.T156T)</i>
rs2084408	23:154346709:T:G	0.3764	0.94	5.36E-19	0.94	6.27E-09	intronic	<i>BRCC3</i>

1023
1024 CHR: chromosome; POS: position (hg19 build); EA: effect allele; NEA: non effect allele; EAF: effect allele frequency; OR: odd
1025 ratio; P: P-value. Results from the discovery are in presented in columns suffixed with “Disc”, while results from the replication
1026 are in columns suffixed with “Repl”. Novel genetic associations are indicated as bold gene names.

1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050

1051 Table 2: Additional 44 candidate novel loci identified in the Overall, European and African meta-
1052 analyses

rsID	CHR:POS:EA:NEA	EAF	EFFECT	SE	OR	P	Locus.context	Locus.Gene
Novel loci identified in the overall meta-analysis								
rs551176418	1:27107263:T:TC	0.9248	0.0759	0.0132	1.08	9.61E-09	UTR3	ARID1A
rs6695572	1:77945635:A:G	0.1938	0.0424	0.0072	1.04	4.28E-09	intronic	AK5
rs3832016	1:109818158:CT:C	0.7627	-0.0449	0.0066	0.96	8.95E-12	UTR3	CELSR2
rs1267881263	1:150496127:CA:C	0.5468	0.0426	0.0076	1.04	2.36E-08	intergenic	FALEC;ADAMTSL4
rs905938	1:154991389:T:C	0.7448	-0.0346	0.0063	0.97	3.70E-08	intronic	DCST2
rs3557	1:161188893:T:G	0.9182	0.0654	0.0106	1.07	7.70E-10	UTR3	FCER1G
rs143410348	1:196809316:T:TAA	0.5434	0.0415	0.0074	1.04	2.44E-08	intergenic	CFHR1;CFHR4
rs78475244	2:65086804:T:C	0.0542	-0.0713	0.0128	0.93	2.52E-08	ncRNA_intronic	LINC01800
rs78872368	2:198545250:C:G	0.1919	-0.0412	0.0071	0.96	7.27E-09	intergenic	RFTN2;MARS2
rs900399	3:156798732:A:G	0.6205	0.0382	0.0060	1.04	1.46E-10	intergenic	LINC02029;LINC00880
rs9654093	4:7903763:C:G	0.1504	0.0492	0.0081	1.05	1.03E-09	intronic	AFAP1
rs781656	4:57778645:A:G	0.1963	0.0389	0.0070	1.04	2.26E-08	intronic	REST
rs7730244	5:72957088:T:C	0.5245	-0.0328	0.0057	0.97	1.04E-08	intronic	ARHGEF28
rs147133967	5:132426851:G:GTT	0.0810	-0.0659	0.0110	0.94	2.43E-09	intronic	HSPA4
rs214059	6:25536937:T:C	0.4331	0.0357	0.0055	1.04	1.01E-10	intronic	CARMIL1
rs2394251	6:29943688:G:C	0.7331	-0.0405	0.0063	0.96	1.43E-10	ncRNA_intronic	HCG9
rs1513275	7:28259233:T:C	0.7453	0.0449	0.0070	1.05	1.40E-10	ncRNA_intronic	JAZF1-AS1
rs10099512	8:9178821:C:G	0.1105	0.0608	0.0105	1.06	6.98E-09	intergenic	LOC101929128;LOC157273
rs2048528	8:23373680:A:G	0.3089	-0.0347	0.0060	0.97	5.77E-09	intergenic	ENTPD4;SLC25A37
rs2915595	8:30402817:A:G	0.2391	0.0365	0.0065	1.04	2.52E-08	intronic	RBPMS
rs4236786	8:108291878:C:G	0.2492	0.0353	0.0064	1.04	3.93E-08	intronic	ANGPT1
rs1243187	10:21907016:T:C	0.6920	-0.0341	0.0061	0.97	2.53E-08	intronic	MLLT10
rs4272700	10:27881308:A:T	0.2726	0.0395	0.0064	1.04	7.75E-10	intergenic	RAB18;MKX
rs2030291	11:16251251:A:T	0.6077	-0.0325	0.0056	0.97	8.19E-09	intronic	SOX6
rs4354705	11:60088159:C:G	0.3635	0.0315	0.0058	1.03	4.83E-08	intergenic	MS4A4A;MS4A6E
rs2846027	11:114003415:T:C	0.3112	-0.0344	0.0061	0.97	1.42E-08	intronic	ZBTB16
rs7107568	11:130779668:T:C	0.5610	-0.0303	0.0056	0.97	4.71E-08	intronic	SNX19
rs2127869	14:65794352:T:C	0.3350	-0.0340	0.0062	0.97	4.68E-08	intergenic	LINC02324;MIR4708
rs7183672	15:96101018:A:G	0.6432	-0.0358	0.0062	0.96	7.34E-09	intergenic	LINC00924;LOC105369212
rs71376077	16:15738114:C:G	0.9728	0.1408	0.0249	1.15	1.57E-08	intronic	NDE1
rs7197453	16:72079127:C:G	0.3572	0.0315	0.0057	1.03	3.19E-08	intergenic	DHODH;HP
rs77246010	16:75429853:T:C	0.4489	0.0408	0.0069	1.04	4.12E-09	intronic	CFDP1
rs8049403	16:85778651:A:G	0.0214	0.1365	0.0248	1.15	3.91E-08	intronic	C16orf74
rs71138827	17:27833678:A:AGATT	0.4288	0.0336	0.0058	1.03	5.89E-09	intronic	TAOK1
rs2545774	19:41287674:T:C	0.2528	-0.0378	0.0065	0.96	6.80E-09	intronic	RAB4B
Additional novel loci identified in the European meta-analysis								
rs4540639	1:192104320:C:G	0.4675	0.0346	0.0060	1.04	6.88E-09	intergenic	LINC01680;RGS18
rs35225200	4:103146888:A:C	0.9190	-0.0645	0.0115	0.94	1.89E-08	intergenic	BANK1;SLC39A8
rs112367053	5:28379046:T:G	0.6662	0.0586	0.0107	1.06	4.07E-08	intergenic	LINC02103;LSP1P3
rs2754251	6:88385949:A:G	0.0584	0.0715	0.0129	1.07	2.65E-08	intronic	AKIRIN2
rs10763665	10:28771491:C:G	0.5783	-0.0342	0.0062	0.97	3.13E-08	ncRNA_intronic	LINC02652
rs7122100	11:10732560:A:C	0.2411	0.0410	0.0075	1.04	4.93E-08	intergenic	IRAG1;CTR9
rs1145656	11:73305859:A:C	0.8171	-0.0442	0.0079	1.05	2.00E-08	upstream	FAM168A
Additional novel loci identified in the African meta-analysis								
rs76668186	16:6686083:A:T	0.9597	-0.5776	0.1056	0.56	4.52E-08	intronic	RBFOX1
rs114102448	21:47523605:A:G	0.0114	0.9527	0.1725	2.60	4.11E-08	intronic	COL6A2

1054
1055 CHR: chromosome; POS: position (hg19 build); EA: effect allele; NEA: non effect allele; EAF: effect allele frequency; SE: Standard
1056 Error of Effect; OR: odd ratio; P: P-value.