



HAL
open science

A genetic association study of circulating coagulation factor VIII and von Willebrand factor levels

Paul S de Vries, Paula Reventun, Michael R Brown, Adam S Heath, Jennifer E Huffman, Ngoc-Quynh Le, Allison Bebo, Jennifer A Brody, Gerard Temprano-Sagrera, Laura M Raffield, et al.

► **To cite this version:**

Paul S de Vries, Paula Reventun, Michael R Brown, Adam S Heath, Jennifer E Huffman, et al.. A genetic association study of circulating coagulation factor VIII and von Willebrand factor levels. *Blood*, 2024, 143 (18), pp.1845-1855. 10.1182/blood.2023021452 . hal-04701077

HAL Id: hal-04701077

<https://hal.science/hal-04701077v1>

Submitted on 18 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



de Vries, P. S., Reventun, P., Brown, M. R., Heath, A. S., Beswick, A. D., Brumpton, B. M., Ben-Shlomo, Y., Morrison, A. C., & al, E. (2024). A genetic association study of circulating coagulation Factor VIII and von Willebrand Factor levels. *Blood*, 143(18), 1845-1855. Advance online publication. <https://doi.org/10.1182/blood.2023021452>

Peer reviewed version

License (if available):
CC BY

Link to published version (if available):
[10.1182/blood.2023021452](https://doi.org/10.1182/blood.2023021452)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via American Society of Hematology at <https://doi.org/10.1182/blood.2023021452>. Please refer to any applicable terms of use of the publisher.

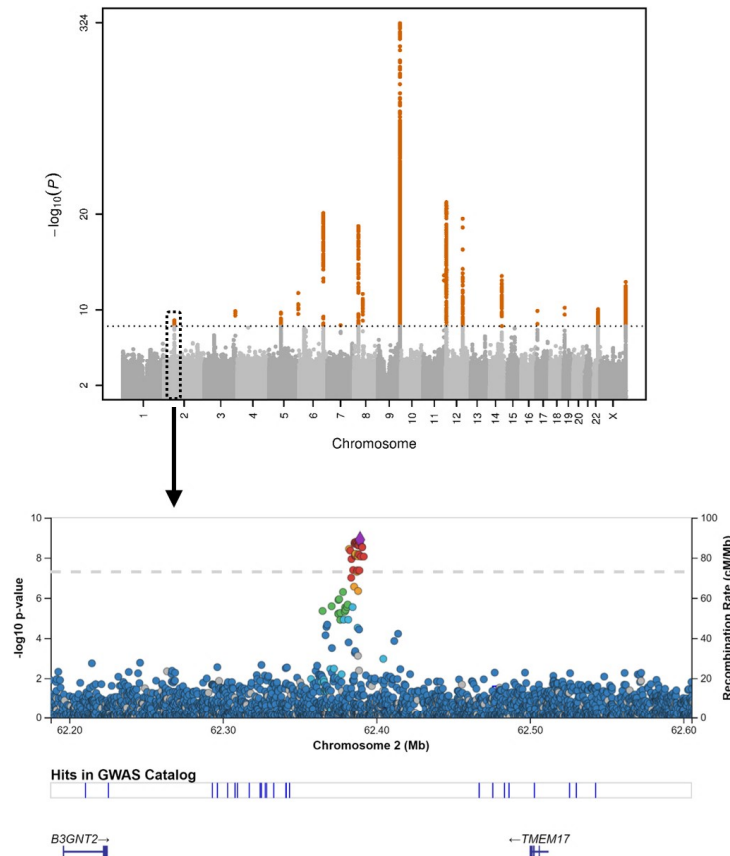
University of Bristol - Explore Bristol Research

General rights

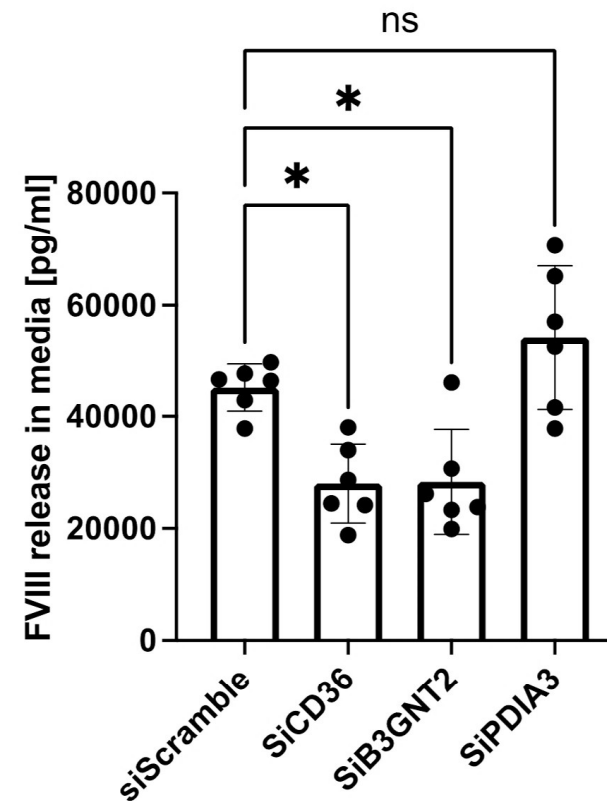
This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

A genetic association study of circulation coagulation Factor VIII and von Willebrand Factor levels

- Genome-wide association study of Factor VIII and von Willebrand Factors



- Gene silencing of potential causal genes *CD36*, *B3GNT2*, and *PDIA3*



- We identified seven new genetic regions for Factor VIII levels, one for von Willebrand factor levels, and three in a combined analysis.
- de Vries et al, Blood, DOI 10.1182/blood.XXXXXXX*

A genetic association study of circulating coagulation Factor VIII and von Willebrand Factor levels

Paul S. de Vries^{1,*}, Paula Reventun^{2,*}, Michael R. Brown¹, Adam S. Heath¹, Jennifer E. Huffman³, Ngoc-Quynh Le⁴, Allison Bebo¹, Jennifer A. Brody⁵, Gerard Temprano-Sagrera⁴, Laura M. Raffield⁶, Ayse Bilge Ozel⁷, Florian Thibord⁸, Deepti Jain⁹, Joshua P. Lewis¹⁰, Benjmain A.T. Rodriguez⁸, Nathan Pankratz¹¹, Kent D. Taylor¹², Ozren Polasek¹³, Ming-Huei Chen⁸, Lisa R. Yanek², German D. Carrasquilla^{14,15}, Riccardo E. Marioni¹⁶, Marcus E. Kleber^{17,18}, David-Alexandre Trégouët¹⁹, Jie Yao¹², Ruifang Li-Gao²⁰, Peter K. Joshi²¹, Stella Trompet²², Angel Martinez-Perez⁴, Mohsen Ghanbari²³, Tom E. Howard²⁴, Alex P. Reiner²⁵, Marios Arvanitis², Kathleen A. Ryan¹⁰, Traci M. Bartz²⁶, Igor Rudan²¹, Nauder Faraday²⁷, Allan Linneberg^{28,29}, Lynette Ekunwe³⁰, Gail Davies³¹, Graciela E. Delgado¹⁸, Pierre Suchon^{32,33}, Xiuqing Guo¹², Frits R. Rosendaal²⁰, Lucija Klaric³⁴, Raymond Noordam²², Frank van Rooij²³, Joanne E. Curran²⁴, Marsha M. Wheeler³⁵, William O. Osburn², Jeffrey R. O'Connell¹⁰, Eric Boerwinkle^{1,36}, Andrew Beswick³⁷, Bruce M. Psaty^{38,5}, Ivana Kolcic¹³, Juan Carlos Souto^{39,4}, Lewis C. Becker², Torben Hansen¹⁴, Margaret F. Doyle⁴⁰, Sarah E. Harris³¹, Angela P. Moissl^{41,42}, Jean-François Deleuze^{43,44}, Stephen S. Rich⁴⁵, Astrid van Hylckama Vlieg²⁰, Harry Campbell²¹, David J. Stott⁴⁶, Jose Manuel Soria⁴, Moniek P.M. de Maat⁴⁷, Laura Almasy^{48,49}, Lawrence C. Brody⁵⁰, Paul L. Auer⁵¹, the Trans-Omics for Precision Medicine (TOPMed) program, INVENT consortium, Braxton D. Mitchell^{10,52}, Yoav Ben-Shlomo⁵³, Myriam Fornage^{54,1}, Caroline Hayward³⁴, Rasika A. Mathias², Tuomas O. Kilpeläinen¹⁴, Leslie A. Lange⁵⁵, Simon R. Cox³¹, Winfried März^{56,18}, Pierre-Emmanuel Morange^{32,33}, Jerome I. Rotter¹², Dennis O. Mook-Kanamori^{20,57}, James F. Wilson²¹, Pim van der Harst⁵⁸, J. Wouter Jukema^{59,60}, M. Arfan Ikram²³, John Blangero²⁴, Charles Kooperberg²⁵, Karl C. Desch⁶¹, Andrew D. Johnson⁸, Maria Sabater-Lleal^{4,62,*}, Charles J. Lowenstein^{2,*}, Nicholas L. Smith^{63,64,65,*}, and Alanna C. Morrison^{1,*}

¹Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA.

²Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

³Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA, USA.

⁴Unit of genomics of Complex Disease, Institut de Recerca Sant Pau (IR SANT PAU), Barcelona, Spain.

⁵Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA.

⁶Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

⁷Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA.

⁸National Heart Lung and Blood Institute, Division of Intramural Research, Population Sciences Branch, The Framingham Heart Study, Framingham, MA, USA.

⁹Genetic Analysis Center, Department of Biostatistics, School of Public Health, University of Washington

¹⁰Department of Medicine, University of Maryland, Baltimore, MD, USA.

¹¹Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA.

¹²The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA.

¹³Faculty of Medicine, University of Split, Split, Croatia.

¹⁴Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark.

- ¹⁵Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.
- ¹⁶Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, Scotland.
- ¹⁷SYNLAB MVZ Humangenetik Mannheim, Mannheim, Germany.
- ¹⁸Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany.
- ¹⁹INSERM UMR 1219, Bordeaux Population Health Research Center, Bordeaux, France.
- ²⁰Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands.
- ²¹Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, Scotland.
- ²²Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden, the Netherlands.
- ²³Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, the Netherlands.
- ²⁴Department of Human Genetics and South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA.
- ²⁵Public Health Sciences, Fred Hutchinson Cancer Center, Seattle, WA, USA.
- ²⁶Cardiovascular Health Research Unit, Departments of Biostatistics and Medicine, University of Washington, Seattle, WA, USA.
- ²⁷Department of Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.
- ²⁸Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark.
- ²⁹Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
- ³⁰Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA.
- ³¹Lothian Birth Cohorts, Department of Psychology, University of Edinburgh, Edinburgh, Scotland.
- ³²C2VN, INSERM, INRAE, Aix Marseille Univ, Marseille, France.
- ³³Laboratory of Haematology, La Timone Hospital, Marseille, France.
- ³⁴MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, Scotland.
- ³⁵Department of Genome Sciences, University of Washington, WA, USA.
- ³⁶Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA.
- ³⁷Translational Health Sciences, University of Bristol, Bristol, UK.
- ³⁸Departments of Epidemiology and Health Systems and Population Health, Seattle, WA, USA.
- ³⁹Unit of Thrombosis and Hemostasis, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain.
- ⁴⁰Department of Pathology and Laboratory Medicine, Larner College of Medicine, University of Vermont, Colchester, VT, USA.
- ⁴¹Institute of Nutritional Sciences, Friedrich-Schiller-University Jena, Jena, Germany.
- ⁴²Competence Cluster for Nutrition and Cardiovascular Health (nutriCARD) Halle-Jena-Leipzig, Jena, Germany.
- ⁴³Centre National de Recherche en Génomique Humaine, Université Paris-Saclay, CEA., Evry, France.
- ⁴⁴Centre d'Etude du Polymorphisme Humain, Fondation Jean Dausset, Paris, France.
- ⁴⁵Department of Public Health Sciences, Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA.
- ⁴⁶Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, Scotland.

- ⁴⁷Department of Hematology, Erasmus MC University Medical Center, Rotterdam, the Netherlands.
- ⁴⁸Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
- ⁴⁹Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA.
- ⁵⁰National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.
- ⁵¹Department of Biostatistics, Medical College of Wisconsin, WI, USA.
- ⁵²Geriatric Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA.
- ⁵³Population Health Sciences, University of Bristol, Bristol, UK.
- ⁵⁴Brown Foundation Institute of Molecular Medicine, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA.
- ⁵⁵Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA.
- ⁵⁶Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany.
- ⁵⁷Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands.
- ⁵⁸Department of Cardiology, Division of Heart and Lungs, Utrecht University, University Medical Center Utrecht, Utrecht, the Netherlands.
- ⁵⁹Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands.
- ⁶⁰Netherlands Heart Institute, Utrecht, the Netherlands.
- ⁶¹Department of Pediatrics, University of Michigan, C.S. Mott Children's Hospital, Ann Arbor, MI, USA.
- ⁶²Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Center for Molecular Medicine, Stockholm, Sweden.
- ⁶³Department of Epidemiology, University of Washington, Seattle, WA, USA.
- ⁶⁴Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle, WA, USA.
- ⁶⁵Seattle Epidemiologic and Information Center, Department of Veterans Affairs Office of Research and Development, Seattle, WA, USA.

*Contributed equally as junior or senior authors.

Correspondence

Paul S. de Vries

Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences,
School of Public Health, The University of Texas Health Science Center at Houston

1200 Pressler St., Suite E-429, Houston, TX 77030

Tel: +1 (713) 500-9809

Email: paul.s.devries@uth.tmc.edu

Alanna C. Morrison

Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences,
School of Public Health, The University of Texas Health Science Center at Houston

1200 Pressler St., Suite E-429, Houston, TX 77030

Tel: +1 (713) 500-9913

Email: alanna.c.morrison@uth.tmc.edu

Data availability statement

Whole genome sequencing data from the Trans-Omics for Precision Medicine (TOPMed) program are available on dbGap: the study-specific accession numbers are listed in **Supplementary Table 14**.

Summary statistics from genome-wide association analyses will be made available on dbGaP, accession number phs001974.

Counts

Abstract word count: 246

Text word count: 4,399

Table count: 2

Figure count: 2

Reference count: 70

Key Points

- We identified seven new genetic regions for Factor VIII levels, one for von Willebrand factor levels, and three in a combined analysis.
- Silencing *B3GNT2* and *CD36* reduced Factor VIII release *in vitro*. Silencing *B3GNT2*, *CD36*, and *PDIA3* reduced von Willebrand Factor release.

Abstract

Coagulation Factor VIII (FVIII) and its carrier protein von Willebrand factor (VWF) are critical to coagulation and platelet aggregation. We leveraged whole genome sequence data from the Trans-Omics for Precision Medicine (TOPMed) program along with TOPMed-based imputation of genotypes in additional samples to identify genetic associations with circulating FVIII and VWF levels in a single variant meta-analysis including up to 45,289 participants. Gene-based aggregate tests were implemented in TOPMed. We identified three candidate causal genes and tested their functional effect on FVIII release from human liver endothelial cells (HLECs) and VWF release from human umbilical vein endothelial cells (HUVECs). Mendelian randomization was also performed to provide evidence for causal associations of FVIII and VWF with thrombotic outcomes. We identified associations ($P < 5 \times 10^{-9}$) at seven new loci for FVIII (*ST3GAL4*, *CLEC4M*, *B3GNT2*, *ASGR1*, *F12*, *KNG1*, and *TREM1/NCR2*) and one for VWF (*B3GNT2*). *VWF*, *ABO*, and *STAB2* were associated with FVIII and VWF in gene-based analyses. Multi-phenotype analysis of FVIII and VWF identified another three new loci, including *PDIA3*. Silencing of *B3GNT2* and the previously reported *CD36* gene decreased release of FVIII by HLECs, while silencing of *B3GNT2*, *CD36*, and *PDIA3* decreased release of VWF by HVECs. Mendelian randomization supports causal association of higher FVIII and VWF with increased risk of thrombotic outcomes. Seven new loci were identified for FVIII and one for VWF, with evidence supporting causal associations of FVIII and VWF with thrombotic outcomes. *B3GNT2*, *CD36*, and *PDIA3* modulate the release of FVIII and/or VWF *in vitro*.

Introduction

Coagulation factor VIII (FVIII) and its carrier protein von Willebrand factor (VWF) are essential hemostasis proteins that play important roles in bleeding and thrombosis.¹ FVIII markedly enhances Factor IX-catalyzed Factor X activation, and ultimately fibrin clot formation, via amplification of the intrinsic coagulation pathway.² VWF promotes thrombosis by mediating platelet adhesion and aggregation.¹ VWF also binds to and stabilizes FVIII, transports FVIII in the blood, limits proteolysis of FVIII, and delivers FVIII to local sites of vascular injury,³ resulting in a high phenotypic and genetic correlation between plasma levels of these two factors.^{4,5}

FVIII and VWF levels are quantitative traits with a substantial genetic component as estimates of their heritabilities range from 60-75%.^{6,7} Prior genome-wide association studies (GWAS) identified 22 loci associated with FVIII or VWF levels,⁸⁻¹¹ of which nine were identified for both FVIII and VWF levels. An additional two loci were identified through bivariate GWAS of FVIII and VWF levels.⁸ Mendelian randomization studies using genetic variants at these loci as instrumental variables provide evidence suggestive of a causal association of FVIII and/or VWF levels on venous thromboembolism (VTE), coronary artery disease (CAD), ischemic stroke, and peripheral artery disease (PAD).^{8,12,13}

Whole genome sequencing enables the simultaneous interrogation of rare and population-specific genetic variants that were poorly covered or absent in previous GWAS. Furthermore, using TOPMed as a reference panel for genotype imputation considerably improves imputation quality for rare and population-specific variants, expanding coverage of many of these variants to studies that lack whole genome sequencing data.^{14,15} To more comprehensively evaluate the genetic architecture of FVIII and VWF, we performed a genomic analysis of FVIII and VWF levels based on whole genome sequencing data from the TOPMed program in combination with TOPMed-imputed genotype data from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium.¹⁶

Methods

Study Population

The study population included 42,125 participants with measured FVIII levels and 45,289 participants with measured VWF levels. These participants represent ten TOPMed studies (**Supplementary Table 1**) and 20 CHARGE studies (**Supplementary Table 2**).^{14,16} Several new studies were added compared to our previous GWAS of FVIII and VWF,⁸ including CaPS, FHS OMNI, Inter99, LBC1921, LBC1936, NEO, PREVEND, retrove, SAFS, and WHI. At the same time, several studies that contributed to the previous GWAS did not contribute to the present study, including AGES, B58C, MEGA, and ORCADES. In TOPMed studies, participants were included if they had freeze 6 whole genome sequencing data and measures of FVIII or VWF levels. CHARGE studies included participants with imputed genotype data and measures of FVIII or VWF levels. Because some studies were analyzed both as a part of TOPMed and a part of CHARGE, overlapping participants were removed from CHARGE analyses. Summaries of the study design of each of the included studies can be found in the **Supplementary Methods**. The total study population across TOPMed and CHARGE was diverse, with reported membership of race/ethnicity population groups as follows: Asian ($N_{\text{FVIII}} = 609$; $N_{\text{VWF}} = 210$), Black ($N_{\text{FVIII}} = 7,970$; $N_{\text{VWF}} = 4,445$), Hispanic/Latino ($N_{\text{FVIII}} = 1,311$; $N_{\text{VWF}} = 656$), White ($N_{\text{FVIII}} = 32,092$; $N_{\text{VWF}} = 39,617$), and other/unknown ($N_{\text{FVIII}} = 143$; $N_{\text{VWF}} = 151$). All studies were approved by their respective institutional review committees, and all included participants gave written informed consent.

Measurement of FVIII and VWF

Circulating FVIII activity or antigen levels and VWF antigen levels were measured in plasma of TOPMed and CHARGE participants and reported in IU/dL. All studies except the Jackson Heart Study measured FVIII activity, while the Jackson Heart Study measured FVIII antigen. Further study-specific details on these measurements are provided in the **Supplementary Methods**. Within each TOPMed study and each population group, FVIII and VWF levels were adjusted for age, sex, and study-specific covariates, and the residuals were rank normalized and rescaled by multiplying by the original standard deviation, so that the transformed phenotype data have the same variances as on the original scale. Each CHARGE study

prepared harmonized phenotypes using a similar strategy as described above, with two key differences: 1) CHARGE studies stratified their phenotype harmonization by population group, and 2) only CHARGE studies with family structure adjusted for the corresponding genetic relatedness matrix.

Whole genome sequencing and genotyping

Whole genome sequencing in TOPMed studies was conducted at a mean depth of >30X using Illumina HiSeq X Ten instruments at six sequencing centers. Variant and sample-level QC are described in the **Supplementary Methods**.

Genetic data in CHARGE studies was based on genotyping arrays and imputation. The studies imputed genotypes using TOPMed Freeze 6 as the reference panel,¹⁴ except for the Rotterdam Study and Inter99, which were imputed to the Haplotype Reference Consortium (HRC) panel.¹⁷ Since TOPMed Freeze 6 used genome build 38 whereas the HRC panel used genome build 37, genomic positions from studies with HRC imputation were converted to build 38 using liftOver. EasyQC was used to harmonize data across studies and perform standard quality control.¹⁸ We calculated the effective minor allele count of each variant by multiplying the minor allele count by the imputation quality. Variants with imputation quality < 0.3 or effective minor allele count ≤ 5 were excluded.

Genetic Analyses

We used two approaches to analyze associations with genetic variants: single variant analyses and gene-based aggregation analyses. Single variant analyses were carried out across TOPMed and CHARGE, while gene-based aggregation analyses were restricted to the whole genome sequencing data from TOPMed to ensure accurate coverage of rare variants.

Single variant analysis: The transformed phenotype data from TOPMed were pooled together into a single analysis, using a heteroscedastic linear mixed model allowing for different residual variances in each study/population group,¹⁹ adjusting for age, sex, study, population group, sequence center, and top ten ancestry-informative principal components as fixed-effects covariates,²⁰ and including a genetic

relatedness matrix calculated by mixed model analysis for pedigrees and populations (MMAP) to model the random effects for relatedness. Association analyses in TOPMed were performed on the Analysis Commons using SMMAT for all variants with MAC > 5.^{21,22} In CHARGE, variants with MAC > 5 were analyzed separately in each study, stratified by population group. METAL was used to perform inverse variance weighted (IVW) fixed-effects meta-analysis across summary statistics from the TOPMed analyses and each of the CHARGE studies.²³ Genomic control corrections were applied to each set of summary statistics prior to meta-analysis. Variants with total minor allele count < 20 in the meta-analysis were excluded. We used a genome-wide significance threshold of 5×10^{-9} , as suggested for whole genome sequencing data.²⁴ Variants within 1 Mb of each index variants were considered to belong to the same locus.

To assess how significant loci affect FVIII and VWF levels differently, for each index variant we calculated the ratio between the beta for FVIII and the beta for VWF. We used the Genotype-Tissue Expression (GTEx) project version 8 to explore whether index variants at loci significantly associated with FVIII and/or VWF levels were associated with expression levels of genes across a wide range of tissues.²⁵ Additional post-GWAS analyses based on single variant summary statistics are described in the **Supplementary Methods**, including multi-phenotype analysis, conditional analysis, transcriptome-wide association analysis. The transcriptome-wide association analyses were limited to White participants because these analyses require the use of a population-specific reference panel.

Gene-based aggregate analysis in TOPMed: We used gene-based aggregate analyses to increase the power to detect associations with low-frequency and rare variants. These analyses were also implemented using SMMAT on the Analysis Commons.^{21,22} We employed three complementary strategies for variant selection: 1) loss of function variants of all frequencies, 2) loss of function and deleterious missense variants with minor allele frequency < 5%, and 3) coding, enhancer, and promoter variants with minor allele frequency < 5%. Further methods are provided in the **Supplementary Methods**.

Evaluation of associations with VWF propeptide to VWF ratio: We examined whether variants associated with FVIII and/or VWF levels, including those from the multi-phenotype analysis, were also associated with VWF propeptide to VWF ratio. This phenotype reflects clearance of VWF from the circulation,²⁶ and was available in 3,238 participants from 2 studies: Genes and Blood Clotting Study (GABC) and Trinity Student Study (TSS). Results from these studies were meta-analyzed using METAL as described above.²³ The significance threshold was adjusted using the Bonferroni correction, dividing 0.05 by the number of tested genetic variants.

Mendelian randomization: We performed two-sample summary statistics-based Mendelian randomization to assess the association of genetically determined levels of FVIII and VWF with the risk of several thrombotic outcomes, including VTE, CAD, myocardial infarction (MI), PAD, ischemic stroke, and ischemic stroke subtypes (large artery stroke, cardioembolic stroke, small vessel stroke). Summary statistics for these outcomes were obtained from large-scale published GWAS.²⁷⁻²⁹ Mendelian randomization analyses are described in the **Supplementary Methods**.

Functional validation of identified loci

We previously used human umbilical vein endothelial cells (HUVECs) as gene silencing models for FVIII and VWF.⁸ Although these cells released abundant VWF into the media, not enough FVIII was released to be detected *in vitro*.⁸ Thus, we identified and used human liver endothelial cells (HLECs, also known as liver sinusoidal endothelial cells) as an alternative cell model to study release of FVIII. Methods for the gene silencing experiments in primary HUVECs and HLECs are provided in the **Supplementary Methods**.

Results

Baseline characteristics

Baseline characteristics of the study participants are shown in **Supplementary Table 1** for TOPMed studies and **Supplementary Table 2** for CHARGE studies. The mean age ranged from 21.2 to 75.4 years

across the included studies. In total, 54.7% of participants were female for FVIII and 46.1% were female for VWF. Mean FVIII levels ranged from 68.2 to 231.4 IU/dL across the studies, while mean VWF antigen levels ranged from 57.4 to 200.7 IU/dL.

Single variant analysis

A total of 34,234,667 genetic variants were included in the cross-population single variant analyses for FVIII, and 35,749,360 were included for VWF. QQ plots are shown in **Supplementary Figure 1**. The genomic inflation factors that were adjusted for were < 1.1 for all contributing studies (**Supplementary Table 3**). The genomic inflation factors of the cross-population meta-analyses were 1.07 for FVIII and 1.02 for VWF. A total of 2,543 variants at 17 loci were significantly associated with FVIII (**Supplementary Table 4**), while 2,824 variants at 15 loci were significantly associated with VWF (**Supplementary Table 5**). Of these loci, 11 were genome-wide significant ($P < 5 \times 10^{-9}$) for both FVIII and VWF. Manhattan plots are shown in **Supplementary Figure 2**, and regional plots are shown in **Supplementary Figures 3 and 4**. Among the 17 significant FVIII loci, 11 were reported by previous GWAS,^{8,11} while six were newly identified associated loci, including: *ST3GAL4*, *CLEC4M*, *B3GNT2*, *ASGR1*, *F12*, and *KNG1* (**Table 1**). *KNG1* has previously been reported in a multi-phenotype GWAS of FVIII and VTE,⁵ but not in a GWAS of FVIII alone. All 15 loci associated with VWF have been reported by previous GWAS. Analyses restricted to White participants identified an additional association for FVIII at the *TREMI1/NCR2* locus that had not been identified in previous studies and an additional association for VWF at *B3GNT2* (**Table 1**), while analyses restricted to Black participants did not identify additional new loci (**Supplementary Table 6**). Conditional analysis, based on summary statistics using GCTA, indicates that there are multiple independently associated variants at *ABO*, *VWF*, *STAB2*, *SCARA5* and *F8* (**Supplementary Table 7**). The independently associated variants identified by GCTA across all significant loci explained 19.6% of the variance in FVIII in White participants and 24.8% in Black participants, as well as 19.8% of the variance in VWF in White participants and 28.9% in Black participants.

A previous study found that each 1% increase in VWF levels results in a 0.54% increase in FVIII activity.³⁰ For each of the index variants at the FVIII-associated loci we calculated the ratio between the beta for FVIII levels and the beta for VWF levels (**Supplementary Table 4**). Except for the *F8* locus (beta ratio = -1.5), all values for this ratio were positive and ranged from 0.62 to 5.87, indicating that the index variants at most significant loci share the same effect direction for FVIII as VWF, and that many loci have direct effects on both FVIII and VWF as opposed to only affecting FVIII indirectly through VWF. Index variants at *ABO*, *SCARA5*, and *VWF* were also associated with VWF propeptide to VWF ratio ($P < 8.5 \times 10^{-4}$, adjusting for 59 tested variants), suggesting that the mechanism at these loci may involve clearance of VWF (**Supplementary Table 8**). Association of the index variants with gene expression across tissues and cell types are shown in **Supplementary Table 9**.

Multi-phenotype analysis

Multi-phenotype analysis of FVIII and VWF levels using metaUSAT identified an additional two previously reported loci and three novel loci: *RASIP1/FUT1/FUT2*, *CKMT1A/PDIA3*, and *PLCL1/SATB2* (**Supplementary Table 10**). All three of these new loci had similar effect sizes for FVIII and VWF, and were suggestively associated with both phenotypes (**Table 1**).

Gene-based rare variant analysis

Across the three aggregation strategies and two phenotypes, 178,012 gene-based aggregation units were analyzed using two statistical tests, , resulting in a P-value threshold of $0.05 / (2 * 178,012) = 1.4 \times 10^{-7}$. As shown in **Table 2**, gene-based rare variant analyses identified associations of FVIII and VWF levels with genes at three loci (*ABO*, *STAB2*, and *VWF*). Results for aggregate tests involving genes annotated to single variant associations are shown in **Supplementary Table 11**. **Supplementary Figures 5 and 6** respectively show leave-one-out analyses for gene-based aggregation units associated with FVIII and VWF, in order to identify the individual rare variants driving the gene-based associations. Some aggregation units appear to be driven by a subset of one or more driver variants: for example, the aggregate test that included loss-of-function and deleterious missense *VWF* variants with FVIII and VWF

levels is markedly attenuated when excluding rs1800386 (12:6018667:T:C). This variant was also independently associated with FVIII and VWF in the conditional analysis with GCTA (**Supplementary Table 7**). In *STAB2*, rs149382223 (12:103695776:T:C) appears to be an important driving variant for both FVIII and VWF.

Transcriptome-wide association study

72,086 transcript-tissue pairs across four tissues were included in transcriptome-wide association analyses for FVIII and VWF levels, which aggregates results across tissues. After Bonferroni correction (P-value threshold = $0.05/72,086 = 6.94 \times 10^{-7}$), genetically determined expression of 83 transcript-tissue pairs was significantly associated with FVIII, involving 27 unique genes at 10 loci (**Supplementary Table 13**). Among these 10 loci, 6 were not reported in the single variant analysis, including *ADSSL1*, *GALC*, *ZNF337*, *SHANK3*, *PDHB*, and *HLA-C*. Similarly, 109 transcript-tissue pairs were associated with VWF, involving 36 unique genes in 13 loci, all of which were reported in the single variant analysis (**Supplementary Table 13**). Genes that were prioritized using FOCUS for fine-mapping are shown in **Supplementary Table 14**.

Mendelian randomization

Up to 22 genetic instruments were used for FVIII levels and up to 19 were used for VWF levels. Forest plots that list the included genetic instruments and show each of their respective causal association estimates are provided in **Supplementary Figures 7 and 8**. When causal estimates were meta-analyzed across these genetic instruments, genetically determined circulating FVIII and VWF levels were associated with increased risk of VTE, PAD, CAD, MI, ischemic stroke, cardioembolic stroke, large artery stroke, but no significant association was found with small vessel stroke (**Figure 1**). These associations were generally consistent across sensitivity analyses (**Supplementary Figures 9**), including the exclusion of genetic instruments pertaining to the ABO locus. When the ABO variant was excluded from the analyses, confidence intervals widened for all outcomes.

Functional validation of newly identified loci

FVIII mRNA expression was not detected in HUVECs, but it was detected in the HLECs (**Supplementary Figure 10a**). We tested *CD36*, *B3GNT2*, and *PDIA3*, three genes that 1) were expressed in HLECs (**Supplementary Figure 10b-d**), 2) had not been reported in association with FVIII or VWF levels at the time of these experiments, and 3) did not clearly relate to a known biological mechanism. An association of *CD36* with FVIII was reported by Pankratz et al. after the experiments had been conducted.¹¹ The three genes were successfully silenced in HLECs (**Supplementary Figure 11**). Silencing of *CD36* and *B3GNT2* decreased FVIII release into the media, while silencing of *PDIA3* did not (**Figure 2**).

We then tested the effect of *CD36*, *B3GNT2*, and *PDIA3* silencing in HUVECs on VWF release into the media, with and without stimulation with histamine. Silencing of the three genes in HUVECs was successful (**Supplementary Figure 12**). Silencing of *CD36* only decreased VWF levels in media under basal conditions, but not after histamine stimulation (**Supplementary Figure 13a**). Silencing of *B3GNT2* decreased VWF release with and without histamine stimulation (**Supplementary Figure 13b**). The direction of the effects of *CD36* and *B3GNT2* silencing in HUVECs on VWF release were thus consistent with the direction of the effects in HLECs on FVIII release. In contrast, while *PDIA3* silencing in HLECs did not significantly affect FVIII release, *PDIA3* silencing in HUVECs decreased VWF release (**Supplementary Figure 13c**).

Discussion

We used whole genome sequencing data from the TOPMed program, along with TOPMed-imputed genotype data from the CHARGE consortium, to perform genetic association studies of FVIII and VWF levels. We identified seven novel and 11 known loci for FVIII levels, one novel and 15 known loci for VWF levels, and three novel and one two known loci in a multi-phenotype analysis of FVIII and VWF levels. Of the loci discovered for FVIII and VWF, 11 overlapped at genome-wide significance in the

cross-population meta-analysis. All VWF loci were associated with FVIII levels at nominal significance ($P < 0.05$), but three FVIII loci were not associated with VWF ($P > 0.05$), including *F12*, *RP1*, and *KNG1*.

B3GNT2 is a newly identified locus for FVIII and VWF levels. It encodes Beta-1,3-N-acetylglucosaminyltransferase 2, which is involved in the synthesis of N-acetylglucosamine chains. Specifically, *B3GNT2* catalyzes the addition of N-acetylglucosamine to N-acetylglucosamine.³¹ Variants at this locus were associated with the levels of both FVIII and VWF. FVIII and VWF are both heavily glycosylated, and there is evidence that VWF carries N-acetylglucosamine chains.³² However, it is unclear from these associations whether only VWF or both FVIII and VWF are specific substrates of *B3GNT2*. Our gene silencing experiments show that *B3GNT2* silencing in HLECs decreases the release of FVIII, and *B3GNT2* silencing in HUVECs decreases the release of VWF. These results suggest that both FVIII and VWF are glycosylated by *B3GNT2*, and that this glycosylation promotes the production and/or secretion of these proteins, and/or decreases the clearance of these proteins from the medium.

PDIA3 was identified as a novel locus in the multi-phenotype analysis, with the same effect direction for FVIII and VWF. We identified *PDIA3* as a candidate causal gene at this locus because it regulates the folding of glycoproteins by catalyzing the formation of disulfide bonds in the endoplasmic reticulum,³³ and both FVIII and VWF rely on multiple disulfide bonds established in the endoplasmic reticulum to dimerize.^{34,35} To test whether *PDIA3* regulates FVIII, VWF, or both, we silenced *PDIA3* in HLECs and HUVECs. We did not find evidence an effect of *PDIA3* silencing on FVIII release from HLECs, but we did identify a decrease in VWF release from HUVECs. This is consistent with a previous report that indicated that VWF dimerization is likely to be performed by either PDI or *PDIA3*.³⁶

We identified several loci for FVIII levels that have not been reported in previous GWAS and relate to mechanisms that are known to interact with or regulate FVIII. *ASGR1* encodes the Ashwell receptor, which is involved in catabolism of circulating glycoproteins, including VWF and FVIII.^{37,38} *F12* and *KNG1* were associated only with FVIII levels and respectively encode Factor XII and high molecular weight kininogen, two contact pathway proteins that activate coagulation via the intrinsic pathway. They

may involve differences in the rate of contact factor activation and hence on the time for initial fibrin clot formation via the intrinsic and common coagulation pathways. Indeed, variants in *F12* and *KNG1* respectively explain 12% and 6% of the variance in activated partial thromboplastin time (aPTT),³⁹ a global test of coagulation potential of the intrinsic and common pathways. Variants in *F12* and *KNG1* are the most significant genetic contributors to this commonly used hemostatic measure. Importantly, FVIII activity is measured using an aPTT-based clotting time assay. It is therefore unclear whether the identified variants in *F12* and *KNG1* affect circulating FVIII levels *in vivo*, or whether they only influence the laboratory measurement of FVIII levels.

In the multi-phenotype analysis of FVIII and VWF we identified an association at the locus harboring the *FUT1* and *FUT2* genes. *FUT1* encodes the H antigen that is expressed in individuals with ABO blood group O and upon which the A and/or B antigens are attached in those with blood groups A, B, or AB. ABO blood group is the most important genetic determinant of FVIII and VWF levels, explaining as much as 13-16% of their variance.⁸ *FUT2* encodes a protein that is required for the secretion of A, B, and H antigens into plasma. Secretor status based on *FUT2* alleles has been associated with VWF levels,⁴⁰ and rare variants in *FUT1* and *FUT2* have been linked to markedly lower VWF levels.⁴¹

Our study also provides new insights into the *CD36* locus, for which we recently reported an association with FVIII levels in an exome-wide association study.¹¹ While loss-of-function variant rs3211938 in *CD36* reached genome-wide significance for FVIII levels, it was also suggestively associated with VWF levels (P=0.0037). The associated loss-of-function variant has also been reported to be under selective pressure and associated with multiple hematological and cardiometabolic traits.⁴²⁻⁴⁷ This variant is largely restricted to populations with African ancestry, and our study population included considerably more Black participants for FVIII (N=7,448) than for VWF (N=4,445). It is thus possible that *CD36* is associated with both FVIII and VWF, but that there was less power to detect this association for VWF. The CD36 receptor is a glycoprotein on the surface of endothelial cells, platelets, and monocytes. While there are no prior reports that CD36 regulates FVIII or VWF levels, our gene

knockdown experiments show that silencing CD36 directly decreased FVIII levels in HLECs and VWF levels in HUVECs. This is consistent with evidence from our GWAS in which the loss-of-function allele for rs3211938 was associated with lower FVIII levels. It remains unclear how FVIII and VWF are regulated by CD36, and whether this involves an effect of CD36 on the synthesis, secretion, or clearance of these proteins.

Several of the new loci identified for FVIII levels have been previously described in association with VWF levels, such as *CLEC4M*, *ST3GAL4*, and *PDHB*.^{8,9} *CLEC4M* appears to affect FVIII indirectly through clearance of VWF,⁴⁸ but also directly by interacting with mannose-exposed glycans on FVIII to promote its clearance, both in the presence of absence of VWF.⁴⁹ While *ST3GAL4* has not been reported in association with FVIII levels by any previous GWAS, it was associated with FVIII in a candidate gene study.⁵⁰ In contrast to *CLEC4M*, *ST3GAL4* appears to prevent clearance of VWF.⁵¹ Whether or not *ST3GAL4* also targets FVIII directly has not been studied. The mechanism linking *PDHB* to VWF and FVIII is unclear, but we previously showed that *PDHB* silencing in HUVECs increased VWF release into media.⁸

A potential benefit of using whole genome sequencing data is the ability to study rare variants. In our study, we identified several gene-based associations that included rare variants, but all of these mapped to loci that were also found by previous studies as well as in our single variant analysis, including *ABO*, *VWF*, and *STAB2*.^{10,11} In our conditional analysis we identified an association of VWF with rs1800386 (Tyr1584Cys), a rare variant of uncertain clinical significance that may contribute to von Willebrand disease.⁵² All in all, our results highlight the ability of whole genome sequencing to provide insights into the role of rare variants, but at the same time suggest that these insights may largely be limited to loci already known to harbor associations with common variants.

Mendelian randomization analyses using improved genetic instruments as a result of the genetic discovery in this study confirm previous reports that genetically determined FVIII and VWF levels are associated with the risk of a range of thrombotic outcomes.^{8,13} However, because of the extensive genetic

overlap between FVIII and VWF, it is unclear whether the effector underlying these associations is FVIII or VWF, or both. It may be possible to determine the causal relationships by relying on FVIII-specific and/or VWF-specific loci. The identification of several FVIII-specific loci in this study is therefore important, but additional FVIII-specific and/or VWF-specific loci will need to be identified in order to distinguish between the effect of FVIII and VWF in future Mendelian randomization studies.

Our study had several limitations. First, while our study represents the most diverse genomic analysis of FVIII and VWF levels to date, many population groups remain underrepresented. Second, because our aggregated rare variant analyses were restricted to only participants with whole genome sequencing data, the sample size and power for these analyses was lower than for the single variant analysis. Third, although sample overlap between each of the exposure and outcome GWAS used in the Mendelian randomization analyses was low, we cannot rule out that sample overlap has influenced the results of the Mendelian randomization analysis.⁵³ Fourth, because of the modest effect size of the genetic variants used as instruments in the Mendelian randomization analyses, it is not clear whether the effect sizes of FVIII and VWF on thrombotic outcomes can be extrapolated to the full range of FVIII and VWF values, including extremely low or high values.

In conclusion, we identified seven novel loci associated with FVIII levels, one with VWF levels, and three with both FVIII and VWF in multi-phenotype analysis. For the newly identified *B3GNT2* locus, as well as the recently reported *CD36* locus, we show that siRNA-mediated gene silencing in HLECs led to reduced levels of FVIII. Many of the newly identified variants are located in or near plausible candidate causal genes, some of which relate to known mechanisms, while others, such as *B3GNT2*, may represent new mechanisms that should be explored further by *in vitro* and *in vivo studies*. Our findings have several translational implications. The novel loci we identified can contribute to more accurate genetic risk scores that may be used to predict risk of bleeding or thrombosis. Additionally, these newly identified loci may represent biological pathways that improve our understanding of new potential therapeutic targets. Finally, since some of the variants are associated with FVIII levels, but not VWF levels, our study emphasizes that VWF stability is not the only determinant of FVIII levels.

Acknowledgements

We thank the CARDIoGRAMplusC4D and MEGASTROKE consortium investigators for making their summary statistics publicly available, as well as the MVP study investigators for making their summary statistics available through dbGaP.

This work and the CHARGE-TOPMed Hemostasis Working Group was funded in part by National Heart, Lung and Blood Institute (NHLBI) grant numbers R01 HL134894 and R01 HL139553. Infrastructure for the CHARGE consortium was supported by R01HL105756. The Analysis Commons was funded by R01 HL131136. Paula Reventun is supported by the American Heart Association and the DC Women's Board 23POST1021300. Charles Lowenstein was supported by R33 HL141791 and the Michel Mirowski M.D. Professorship in Cardiology. Caroline Hayward and James Wilson were supported by the UK Medical Research Council grant U. MC_UU_00007/10. Gerard Temprano-Sagrera is supported by the "Pla Estratègic de Recerca i Innovació en Salut (PERIS)" grant from the Catalan Department of Health (SLT017/20/000100). Maria Sabater-Lleal is supported by a Miguel Servet contract from the ISCIII Spanish Health Institute (CPII22/00007) and co-financed by the European Social Fund. Laura Raffield was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant KL2TR002490.

Molecular data for the TOPMed program was supported by the NHLBI. Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). See the TOPMed Omics Support Table (**Supplementary Table 15**) for study specific omics support information. Further study-specific funding and acknowledgements are detailed in the **Supplementary Methods**. We gratefully acknowledge the participants and staff of the contributing studies.

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

Author contributions

ACM, APR, AL, AnB, ADJ, AvHV, BDM, BMP, CH, DJS, DAT, DOMK, FRR, HC, IR, JWJ, JFW, JFD, JIR, JEC, JB, JMS, JPL, JCS, KCD, KDT, LMR, LA, LCB, LAL, LCB, LE, MFD, MAI, MHC, MF, NF, NLS, OP, PEM, RAM, SRC, SSR, THE, WM, and YBS contributed to the design and/or funding of the included studies from the TOPMed program and/or CHARGE consortium. ACM, APR, AL, AnB, ADJ, AMP, AvHV, ABO, BATR, BDM, BMP, CH, DJS, DAT, DJ, DOMK, FvR, FRR, GD, GDC, IR, IK, JWJ, JFW, JFD, JROC, JEH, JEC, JB, JPL, JCS, KCD, KAR, KDT, LA, LCB, LRY, MEK, MSL, MRB, MHC, MPMdM, MF, NP, NF, NQL, OP, PSdV, PS, PEM, PvdH, RAM, RLG, SHE, SRC, ST, THE, TH, TOK, and YBS contributed to the acquisition of genotype and/or phenotype data. ASH, ACM, AIB, ADJ, AMP, APM, ABO, BATR, BDM, CH, DAT, DJ, FT, GD, GTS, GDC, GED, JFD, JROC, JEH, JAB, JY, JPL, KCD, KAR, LMR, LRY, LK, MEK, MSL, MRB, MHC, MG, NQL, NLS, PSdV, PKJ, RAM, RN, REM, RLG, ST, TMB, and XG contributed to the genetic epidemiology analyses and interpretation. CJL, MA, PR, and WOB contributed to the design of the gene silencing experiments. PR performed the gene silencing experiments. PR and CJL contributed to data analysis and interpretation of the gene silencing experiments. ASH, ACM, CJL, MRB, MSL, NLS, PR, PSdV drafted the initial manuscript. All coauthors participated in critical review of the manuscript.

Conflict of interest disclosures

Charles Lowenstein received funding in 2020 from Novartis for a research grant. Bruce Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Laura Raffield is a consultant for the TOPMed Administrative Coordinating Center through WeStat. Ruifang Li-Gao is a part-time consultant for Metabolon, Inc.

References

1. Swami A, Kaur V. von Willebrand Disease: A Concise Review and Update for the Practicing Physician. *Clin Appl Thromb Hemost*. 2017;23(8):900-910.
2. Chavin SI. Factor VIII: structure and function in blood clotting. *Am J Hematol*. 1984;16(3):297-306.
3. Terraube V, O'Donnell JS, Jenkins PV. Factor VIII and von Willebrand factor interaction: biological, clinical and therapeutic importance. *Haemophilia*. 2010;16(1):3-13.
4. Lippi G, Franchini M, Salvagno GL, Montagnana M, Poli G, Guidi GC. Correlation between von Willebrand factor antigen, von Willebrand factor ristocetin cofactor activity and factor VIII activity in plasma. *J Thromb Thrombolysis*. 2008;26(2):150-153.
5. Temprano-Sagrera G, Sitlani CM, Bone WP, et al. Multi-phenotype analyses of hemostatic traits with cardiovascular events reveal novel genetic associations. *J Thromb Haemost*. 2022;20(6):1331-1349.
6. de Lange M, Snieder H, Ariens RA, Spector TD, Grant PJ. The genetics of haemostasis: a twin study. *Lancet*. 2001;357(9250):101-105.
7. Swystun LL, Lillicrap D. Genetic regulation of plasma von Willebrand factor levels in health and disease. *J Thromb Haemost*. 2018;16(12):2375-2390.
8. Sabater-Lleal M, Huffman JE, de Vries PS, et al. Genome-Wide Association Transethnic Meta-Analyses Identifies Novel Associations Regulating Coagulation Factor VIII and von Willebrand Factor Plasma Levels. *Circulation*. 2019;139(5):620-635.
9. Smith NL, Chen MH, Dehghan A, et al. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. *Circulation*. 2010;121(12):1382-1392.
10. Huffman JE, de Vries PS, Morrison AC, et al. Rare and low-frequency variants and their association with plasma levels of fibrinogen, FVII, FVIII, and vWF. *Blood*. 2015;126(11):e19-29.
11. Pankratz N, Wei P, Brody JA, et al. Whole-exome sequencing of 14 389 individuals from the ESP and CHARGE consortia identifies novel rare variation associated with hemostatic factors. *Hum Mol Genet*. 2022;31(18):3120-3132.
12. Guo Y, Rist PM, Sabater-Lleal M, et al. Association Between Hemostatic Profile and Migraine: A Mendelian Randomization Analysis. *Neurology*. 2021;96(20):e2481-e2487.
13. Small AM, Huffman JE, Klarin D, et al. Mendelian Randomization Analysis of Hemostatic Factors and Their Contribution to Peripheral Artery Disease-Brief Report. *Arterioscler Thromb Vasc Biol*. 2021;41(1):380-386.
14. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590(7845):290-299.
15. Kowalski MH, Qian H, Hou Z, et al. Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet*. 2019;15(12):e1008500.
16. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet*. 2009;2(1):73-80.
17. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283.
18. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9(5):1192-1212.
19. Conomos MP, Laurie CA, Stilp AM, et al. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet*. 2016;98(1):165-184.

20. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol.* 2015;39(4):276-293.
21. Chen H, Huffman JE, Brody JA, et al. Efficient Variant Set Mixed Model Association Tests for Continuous and Binary Traits in Large-Scale Whole-Genome Sequencing Studies. *Am J Hum Genet.* 2019;104(2):260-274.
22. Brody JA, Morrison AC, Bis JC, et al. Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. *Nat Genet.* 2017;49(11):1560-1563.
23. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.
24. Lin DY. A simple and accurate method to determine genomewide significance for association tests in sequencing studies. *Genet Epidemiol.* 2019;43(4):365-372.
25. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369(6509):1318-1330.
26. Ozel AB, McGee B, Siemieniak D, et al. Genome-wide studies of von Willebrand factor propeptide identify loci contributing to variation in propeptide levels and von Willebrand factor clearance. *J Thromb Haemost.* 2016;14(9):1888-1898.
27. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet.* 2018;50(4):524-537.
28. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet.* 2015;47(10):1121-1130.
29. Klarin D, Lynch J, Aragam K, et al. Genome-wide association study of peripheral artery disease in the Million Veteran Program. *Nat Med.* 2019;25(8):1274-1279.
30. Song J, Chen F, Campos M, et al. Quantitative Influence of ABO Blood Groups on Factor VIII and Its Ratio to von Willebrand Factor, Novel Observations from an ARIC Study of 11,673 Subjects. *PLoS One.* 2015;10(8):e0132626.
31. Hao Y, Crequer-Grandhomme A, Javier N, et al. Structures and mechanism of human glycosyltransferase beta1,3-N-acetylglucosaminyltransferase 2 (B3GNT2), an important player in immune homeostasis. *J Biol Chem.* 2021;296:100042.
32. Canis K, McKinnon TA, Nowak A, et al. Mapping the N-glycome of human von Willebrand factor. *Biochem J.* 2012;447(2):217-228.
33. Bourdi M, Demady D, Martin JL, et al. cDNA cloning and baculovirus expression of the human liver endoplasmic reticulum P58: characterization as a protein disulfide isomerase isoform, but not as a protease or a carnitine acyltransferase. *Arch Biochem Biophys.* 1995;323(2):397-403.
34. McMullen BA, Fujikawa K, Davie EW, Hedner U, Ezban M. Locations of disulfide bonds and free cysteines in the heavy and light chains of recombinant human factor VIII (antihemophilic factor A). *Protein Sci.* 1995;4(4):740-746.
35. Denis CV. Molecular and cellular biology of von Willebrand factor. *Int J Hematol.* 2002;75(1):3-8.
36. Lippok S, Kolsek K, Lof A, et al. von Willebrand factor is dimerized by protein disulfide isomerase. *Blood.* 2016;127(9):1183-1191.
37. Grewal PK, Uchiyama S, Ditto D, et al. The Ashwell receptor mitigates the lethal coagulopathy of sepsis. *Nat Med.* 2008;14(6):648-655.
38. Bovenschen N, Rijken DC, Havekes LM, van Vlijmen BJ, Mertens K. The B domain of coagulation factor VIII interacts with the asialoglycoprotein receptor. *J Thromb Haemost.* 2005;3(6):1257-1265.
39. Weng LC, Cushman M, Pankow JS, et al. A genetic association study of activated partial thromboplastin time in European Americans and African Americans: the ARIC Study. *Hum Mol Genet.* 2015;24(8):2401-2408.
40. O'Donnell J, Boulton FE, Manning RA, Laffan MA. Genotype at the secretor blood group locus is a determinant of plasma von Willebrand factor level. *Br J Haematol.* 2002;116(2):350-356.

41. O'Donnell JS, McKinnon TA, Crawley JT, Lane DA, Laffan MA. Bombay phenotype is associated with reduced plasma-VWF levels and an increased susceptibility to ADAMTS13 proteolysis. *Blood*. 2005;106(6):1988-1991.
42. Fry AE, Ghansa A, Small KS, et al. Positive selection of a CD36 nonsense variant in sub-Saharan Africa, but no association with severe malaria phenotypes. *Hum Mol Genet*. 2009;18(14):2683-2692.
43. Sun Q, Graff M, Rowland B, et al. Analyses of biomarker traits in diverse UK biobank participants identify associations missed by European-centric analysis strategies. *J Hum Genet*. 2022;67(2):87-93.
44. Hu Y, Stilp AM, McHugh CP, et al. Whole-genome sequencing association analysis of quantitative red blood cell phenotypes: The NHLBI TOPMed program. *Am J Hum Genet*. 2021;108(5):874-893.
45. Wojcik GL, Graff M, Nishimura KK, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019;570(7762):514-518.
46. Ellis J, Lange EM, Li J, et al. Large multiethnic Candidate Gene Study for C-reactive protein levels: identification of a novel association at CD36 in African Americans. *Hum Genet*. 2014;133(8):985-995.
47. Auer PL, Johnsen JM, Johnson AD, et al. Imputation of exome sequence variants into population-based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO Exome Sequencing Project. *Am J Hum Genet*. 2012;91(5):794-808.
48. Rydz N, Swystun LL, Notley C, et al. The C-type lectin receptor CLEC4M binds, internalizes, and clears von Willebrand factor and contributes to the variation in plasma von Willebrand factor levels. *Blood*. 2013;121(26):5228-5237.
49. Swystun LL, Notley C, Georgescu I, et al. The endothelial lectin clearance receptor CLEC4M binds and internalizes factor VIII in a VWF-dependent and independent manner. *J Thromb Haemost*. 2019;17(4):681-694.
50. Song J, Xue C, Preisser JS, et al. Association of Single Nucleotide Polymorphisms in the ST3GAL4 Gene with VWF Antigen and Factor VIII Activity. *PLoS One*. 2016;11(9):e0160757.
51. Ellies LG, Ditto D, Levy GG, et al. Sialyltransferase ST3Gal-IV operates as a dominant modifier of hemostasis by concealing asialoglycoprotein receptor ligands. *Proc Natl Acad Sci U S A*. 2002;99(15):10042-10047.
52. Bowen DJ, Collins PW, Lester W, et al. The prevalence of the cysteine1584 variant of von Willebrand factor is increased in type 1 von Willebrand disease: co-segregation with increased susceptibility to ADAMTS13 proteolysis but not clinical phenotype. *Br J Haematol*. 2005;128(6):830-836.
53. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol*. 2016;40(7):597-608.

Tables

Table 1: Index variants at novel loci discovered in the cross-population, population-specific, and multi-phenotype analyses.

rsID	Chr:pos	Gene	Alleles	Factor VIII levels				Von Willebrand Factor levels				Multi-phenotype
				Freq	Beta	SE	P-value	Freq	Beta	SE	P-value	P-value
<i>Factor VIII levels – Cross-population</i>												
rs75077631	5:177413083	<i>GRK6 / F12</i>	G/GC	37.2%	-2.56	0.36	1.8E-12	31.2%	-0.44	0.40	2.8E-01	6.0E-13
rs62061426	17:7170806	<i>ASGR1</i>	A/G	18.8%	2.42	0.38	1.3E-10	19.0%	1.37	0.41	7.9E-04	1.4E-09
rs710446	3:186742138	<i>KNG1</i>	T/C	57.0%	-1.85	0.29	1.4E-10	57.7%	-0.55	0.32	8.1E-02	9.1E-10
rs6727115	2:62389406	<i>B3GNT2</i>	T/C	76.4%	-2.13	0.35	1.3E-09	78.0%	-2.00	0.39	2.4E-07	6.9E-11
rs35257264	11:126426921	<i>ST3GAL4 / KIRREL3</i>	T/C	2.2%	7.45	0.98	2.5E-14	2.4%	6.67	1.07	3.7E-10	8.1E-16
rs868875	19:7766280	<i>CLEC4M</i>	A/G	72.0%	2.10	0.32	6.0E-11	70.9%	2.36	0.35	1.7E-11	2.3E-15
<i>Factor VIII levels – White participants</i>												
rs35206772	6:41325017	<i>TREM1 / NCR2</i>	A/G	67.6%	-2.02	0.33	9.3E-10	67.8%	-0.72	0.43	9.0E-02	2.3E-07
<i>von Willebrand Factor levels – White participants</i>												
rs61558368	2:62414160	<i>B3GNT2</i>	T/C	34.9%	1.24	0.33	1.9E-04	35.6%	2.09	0.35	2.0E-09	1.2E-08
<i>Multi-phenotype analysis of Factor VIII levels and von Willebrand Factor levels – Cross-population</i>												
rs12979891	19:48723999	<i>RASIP1 / FUT1 / FUT2</i>	T/C	44.2%	1.37	0.30	3.9E-06	49.0%	1.83	0.32	1.8E-08	2.6E-09
rs145633869	15:43719006	<i>CKMT1A / PDIA3</i>	CAGAG/C	98.3%	6.55	1.22	8.0E-08	98.1%	6.15	1.26	1.1E-06	4.1E-09
rs184336448	2:198846960	<i>PLCLI / SATB2</i>	A/T	0.43%	11.01	2.53	1.3E-05	0.53%	14.61	2.59	1.6E-08	4.4E-09

Gene shows the closest genes at each locus; Alleles shows the coded/noncoded alleles; Frequency (freq) and beta apply to the coded alleles. SE refers to the standard error of the beta.

Table 2: Significantly associated gene-based aggregation units for circulating Factor VIII and von Willebrand Factor levels based on whole genome sequencing data from the Trans-Omics for Precision Medicine (TOPMed) program.

Aggregation strategy	Gene	nVar	MAC	P_{burden}	P_{SMMAT}
<i>Factor VIII levels</i>					
Coding, promoter, and enhancer variants	<i>STAB2</i>	61	315	2.71E-06	2.85E-10
LOF and deleterious missense variants	<i>STAB2</i>	158	611	1.97E-16	4.93E-19
LOF and deleterious missense variants	<i>VWF</i>	86	353	8.62E-18	1.23E-18
LOF variants	<i>STAB2</i>	53	297	2.44E-06	1.92E-11
LOF variants	<i>ABO</i>	17	16763	3.36E-11	7.29E-11
<i>Von Willebrand Factor levels</i>					
Coding, promoter, and enhancer variants	<i>STAB2</i>	36	123	3.64E-08	1.08E-08
LOF and deleterious missense variants	<i>VWF</i>	76	222	3.34E-15	8.83E-16
LOF and deleterious missense variants	<i>STAB2</i>	98	307	9.50E-15	8.76E-15
LOF variants	<i>VWF</i>	22	44	5.30E-07	5.30E-10
LOF variants	<i>STAB2</i>	29	114	7.64E-09	6.14E-10
LOF variants	<i>ABO</i>	15	12438	2.67E-08	9.39E-08

nVar shows the number of genetic variants included in the aggregation test. MAC shows the cumulative minor allele count of the included variants. LOF refers to loss-of-function.

Figure Legends

Figure 1: Results of the inverse-variance weighted Mendelian randomization analyses of a) Factor VIII levels and b) von Willebrand factor levels on a range of thrombotic outcomes. Odds ratios are shown per 10 IU/dL change in Factor VIII and von Willebrand Factor levels.

a) Factor VIII levels

b) Von Willebrand Factor levels

Figure 2: Candidate genes identified by GWAS regulate Factor VIII release from human liver endothelial cells.

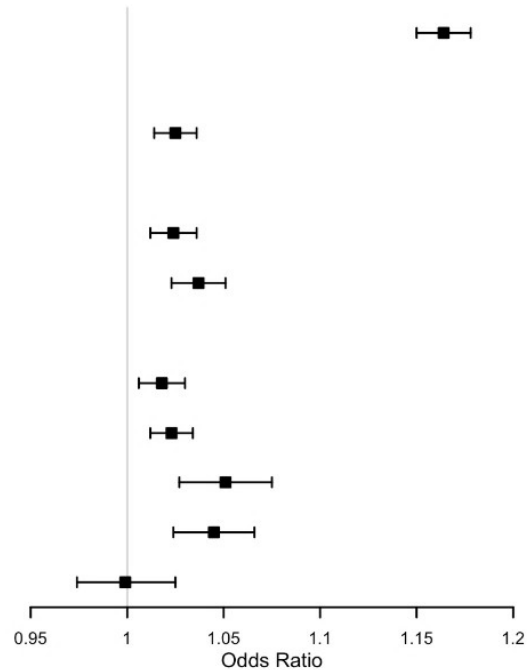
Human liver endothelial cells were transfected with small interfering RNA (siRNA), including control siRNA (Scramble Ct), siRNA directed against CD36 (siCD36), siRNA directed against B3GNT2 (siB3GNT2), and siRNA directed against PDIA3 (siPDIA3). The cells were washed and fed with fresh media, cultured for 24 h, and the media was collected. FVIII concentrations in the cell culture media were measured by ELISA ($n = 6 \pm$ S.D. * for $p < 0.05$ compared to Scramble Ct).

Fig 1

a)

Outcome

Venous thromboembolism
 Peripheral arterial disease
 Coronary artery disease
 Myocardial infarction
 Any stroke
 Any ischemic stroke
 Cardioembolic stroke
 Large artery stroke
 Small vessel stroke



OR

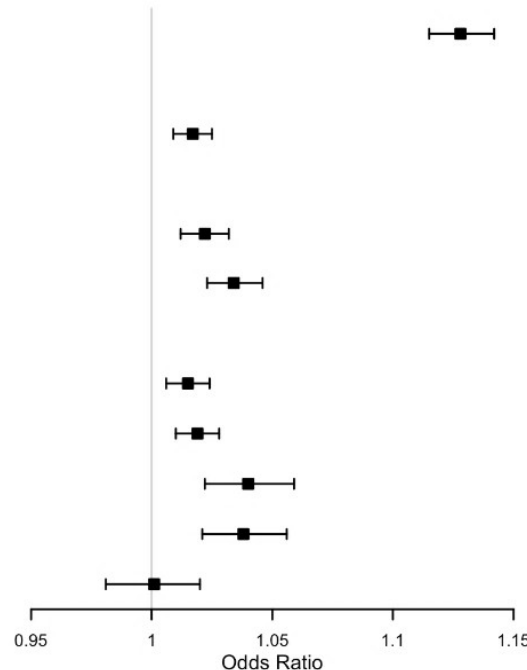
P-Value

95% CI

b)

Outcome

Venous thromboembolism
 Peripheral arterial disease
 Coronary artery disease
 Myocardial infarction
 Any stroke
 Any ischemic stroke
 Cardioembolic stroke
 Large artery stroke
 Small vessel stroke



OR

P-Value

95% CI

Fig 2

