SUPPLEMENTARY MATERIALS

ABO O blood group as a risk factor for platelet reactivity in heparin-induced thrombocytopenia

Jason H Karnes[1,2*], Jerome Rollin[3,4], Jason B Giles[1], Kiana L Martinez[1], Heidi E Steiner[1], Christian M Shaffer[5], Yukihide Momozawa[6], Chihiro Inai[6], Andrei Bombin[5], Mingjian Shi[2], Jonathan D Mosley[2,5], Ian Stanaway[7], Kathleen Selleng[8], Thomas Thiele[8], Taisei Mushiroda[6], Claire Pouplard[3,4], Nancy M Heddle[9], Michiaki Kubo[6], Elizabeth J Phillips[5,10], Theodore E Warkentin[9], Yves Gruel[3,4*], Andreas Greinacher[8*], Dan M Roden[2,5,10*]

*Authors jointly supervised this work

1) Department of Pharmacy Practice and Science, University of Arizona College of Pharmacy, Tucson, AZ, USA
2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
3) Regional University Hospital Centre Tours, Department of Hemostasis, Tours, France
4) University of Tours, EA7501 GICC, Tours, France
5) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
6) RIKEN Center for Integrative Medical Sciences (IMS), Yokohama, Kanagawa, Japan
7) Department of Medicine, Kidney Research Institute, University of Washington, Seattle, WA, USA
8) Institute of Immunology and Transfusion Medicine, University of Greifswald, Greifswald, Germany
9) Department of Medicine, McMaster University, Hamilton, ON, Canada
10) Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA

Correspondence to:
Jason H Karnes, PharmD, PhD
Associate Professor
University of Arizona College of Pharmacy
1295 N Martin AVE
Tucson, AZ 85721
520-626-1447
karnes@pharmacy.arizona.edu

SUPPLEMENTARY METHODS

Patient Cohorts

Discovery Cohort (Greifswald University Medicine [Greifswald, Germany])

The observational discovery cohort was collected at the platelet laboratory at Greifswald University Medicine, which is utilized as a central testing facility for heparin-induced thrombocytopenia (HIT) diagnostic assays across multiple institutions. Patients underwent laboratory testing for platelet factor 4 (PF4)/heparin antibodies and heparin-induced platelet activation (HIPA) assay. The HIPA test has been validated extensively in comparison with the platelet aggregation test, by analysis of prospective clinical study data, in comparison with the serotonin release assay (SRA), and in international workshops data[1-4]. The cut off of a lag time of platelet activation of 30 minutes gave the best discrimination between specific heparin-dependent platelet activation and non-specific platelet aggregates which may occur after steering of platelets for more than 30 minutes[5]. For the HIPA test, at least two of four donors must become activated within 30 minutes. We tested a panel of four random platelet donors, accounting for the interindividual variability insensitivity towards HIT sera.

Samples for this genome-wide association study (GWAS) were tested between 2002 and 2015. DNA was extracted from buffy coats by the Vanderbilt Technologies for Advanced Genomics (VANTAGE) core using AUTOPURE LS® Automated Nucleic Acid Purification Instrument (Qiagen®, Hilden, Germany). Basic demographic data, including age and sex was available, but detailed clinical data, such as 4Ts score, thromboembolic complications, unfractionated heparin (UFH) dosing, and platelet counts, were not available in the discovery cohort. PF4/heparin antibodies were measured via polyclonal enzyme-linked immunosorbent assay (ELISA), including quantitation of IgG, IgA, and IgM. PF4/heparin antibody ELISAs used an optical density threshold of 0.5 for a positive test result. Sera were tested for anti-PF4/heparin IgG, IgA, and IgM antibodies by ELISA,[6] including a high heparin inhibition step (100 IU/mL ≥ 40%) as described.[7] The discovery cohort was ultimately divided into three groups, including HIPA

positive cases, PF4/heparin antibody positive controls with negative HIPA test result, and controls with both negative PF4/heparin antibody and negative HIPA test results. Samples for this GWAS were tested between 2002 and 2015 and were completely de-identified prior to GWAS genotyping. This research received ethical approved from the institutional Ethics committee at Greifswald University (Reference Number BB 097/15).

Replication Cohort (University of Tours [Tours, France])

The replication cohort was made up of two previously described cohorts recruited prospectively at the University of Tours. The first cohort contained definite serotonin release assay (SRA)-confirmed HIT cases and PF4/heparin antibody positive controls that tested negative for SRA. Details of this cohort have been previously published[8-11]. Serotonin release assays were performed as previously described[12]. All HIT cases developed a platelet count fall bearing a temporal relationship with preceding heparin exposure consistent with an immune response, a significant decrease in platelet count after a heparin treatment, or both. Both PF4-specific ELISA (HAT; GTI, Brookfield, WI) and SRA were positive in every HIT case. The SRA result was negative in every patient grouped as PF4/heparin antibody positive. All PF4/heparin antibody positive (but SRA negative) patients had undergone cardio-pulmonary bypass surgery (CPB) and had been treated with heparin postoperatively. All developed significant levels of PF4-specific antibodies using an ELISA optical density threshold of 0.5 but without significant abnormal evolution in platelet count in the postoperative period. Genomic DNA was extracted from citrated whole blood using the Flexigen DNA kit (Qiagen, Courtaboeuf, France) according to the manufacturer's instructions. Blood samples were collected after obtaining informed consent according to the Helsinki declaration principles. Both the University of Tours institutional ethics committee and the Ministry of Research had previously approved collection of DNA from HIT patients and healthy controls for genetic studies (DC2008-308).

The second cohort named "Facteurs de RIsque Génétiques de Thrombopénie Induite par l'Héparine" (FRIGTIH) included patients cases recruited between 2011 and 2015[13]. Clinical and

biological data were recorded using a standardized case report form with 22 questions and 64 response items. Plasma and DNA samples were collected during the acute phase, stored frozen at –80°C, and then transferred to the University Hospital in Tours. After HIT was clinically suspected, each center detected PF4-specific antibodies using a PF4-specific immunoassay. In addition, a PF4/polyvinylsulfonate IgG-specific ELISA was performed on all patients for whom sufficient plasma was available. Positive PF4/heparin antibodies were determined using an optical density threshold of 0.5. SRA was also performed for all patients with a positive PF4/heparin antibody test and the diagnosis of HIT only confirmed when SRA was positive. All HIT cases developed a decrease of ≥40% compared with the initial platelet count during heparin therapy. Alternative non-heparin antithrombotic therapy was initiated in most HIT cases after heparin discontinuation. Thrombocytopenia was considered resolved when platelet count increased by at least 100% from nadir value and remained above 100 G/L without any relapse. The Institutional Ethics Committee approved this study (Number # 2011-N7).

Genome-wide Genotyping, Genomic Imputation, and Quality Control

GWAS genotyping was performed at the RIKEN Laboratory for Genotyping Development, Center for Integrative Medical Sciences (IMS). Genotyping was performed as a result of funded proposals for the Pharmacogenomics Research Network (PGRN)-RIKEN Global Alliance in Round 13 (GWAS genotyping) and Round 16 (*ABO* sequencing). All samples in the discovery and replication cohorts were genotyped using the Illumina® Infinium HumanOmniExpressExome BeadChip, which contains 958,497 markers including 273,000 functional exonic markers. Quality control was accomplished using sample and marker call rate ≥98%, sex mismatch, duplicate concordance, and removal of related samples (first cousin or closer kinship) through identity by descent, including removal of one sample from pairs with any genetic relationship greater than first cousin. For the sensitivity analysis using mixed models in GCTA, related individuals were retained[14,15]. Genomic imputation was performed for samples using Minimac4 on the MI imputation server and the phase3 v5 reference panel data from the 1000 genomes project[16].

Imputed single nucleotide polymorphisms (SNPs) were retained if their R^2 value was 0.3 or greater. The Illumina® Infinium HumanOmniExpressExome BeadChip genome-wide platform does not contain copy number variation and only single nucleotide and insertion/deletion polymorphisms were used in this analysis. Genomic imputation was also performed using the Trans-Omics for Precision Medicine (TOPMed) Imputation Server[17]. The quality-controlled autosomes and X chromosomes from the discovery cohort were imputed with Eagle v2.4 and imputed with the TOPMed r2 reference panel. The r2 filter option was set to 0.3 thus excluding >70% of poorly imputed SNPs. During statistical analyses, SNPs were removed for Hardy-Weinberg equilibrium p value<0.001 and minor allele frequency (MAF) less than 0.01.

Analysis and Adjustment of Ancestry

Principal components analysis (PCA) was performed using the SmartPCA R package. PCs were generated using all directly genotyped markers alongside HapMap reference populations and visually inspected for adequate separation of race/ethnic groups using ggplot2 v3.3.5 in R[18] (Supplementary Figures S1-S2). To minimize confounding by population stratification, GWAS analyses were adjusted for the first three principal components (PCs). Due to the overwhelming majority of patients in both cohorts being of European ancestry, GWAS analyses were restricted to individuals of European ancestry. Estimated proportions of global ancestry were determined using ancestry informative markers (AIMs) input into STRUCTURE with HapMap reference populations[18]. Individuals were considered to be of European ancestry if genetically-defined global ancestry was at least 80 percent European. Similarly, individuals were considered to have African or Asian ancestry if relevant ancestral proportions were at least 80 percent. Otherwise, individuals were considered admixed since patient-reported race/ethnicity data was not available. In order to incorporate samples with non-European ancestry and to assess consistency of associations across ancestry groups, we performed a mixed model analysis using GCTA[14,15]. This mixed model analysis included all samples regardless of ancestry and all related

individuals. The GCTA analysis was performed in the combined discovery and replication cohorts with adjustment for age, sex, cohort (discovery versus replication), and PCs 1-3.

Statistical Analysis

Q-Q plots were generated to determine potential for false positive results and were assessed in conjunction with genomic inflation values (λ) for potential of false positive results. Pairwise linkage disequilibrium (LD) as measured by r^2 was estimated in discovery and replication cohorts using Plink v1.90 using the –r2 flag. Manhattan plots were created with heatmaps for indication of effect sizes of variants. Meta-analysis association results for SNPs in the *ABO* genomic region were plotted for visual inspection against local recombination rates, pairwise linkage disequilibrium between SNPs, and posterior probabilities from PAINTOR (described below; Figure 2 in main text). *ABO* blood types (A, B, AB, and O) were derived based on haplotypes of rs8176719 and rs8176746, as previously described, and tested for association with functional assay positive status in both cohorts combined[19,20]. Logistic regressions for association of blood groups compared the relevant blood group (i.e. O, A, B, or AB) to all other blood groups combined with adjustment for age, sex, PCs 1-3, and cohort (discovery versus replication). Sensitivity analyses were also performed for blood group associations, including functional assay positive cases compared to PF4/heparin antibody positive controls alone (excluding antibody negative controls), functional assay positive cases compared to PF4/heparin antibody negative controls alone (excluding antibody positive controls), and including adjustment for PF4/heparin antibody IgG levels. Forest plots were generated for blood group analyses. All statistical analyses were performed using Rv4.04, PLINK v1.90[21] and SAS v9.4 (SAS, Cary, NC).

Polygenic Risk Score Analysis

For calculating polygenic risk scores (PRS), we used 297 variants that were shown to be correlated with venous thromboembolism in a prior study[22]. In order to exclude variants located near the *ABO* gene, we excluded all SNPs located on chromosome nine. The remaining SNPs that matched variants that had been present in imputed populations, including 242 SNPs in the

Discovery population and 241 SNPs in the Replication population, were used for computing PRS scores with PLINK v1.90b6.17[23]. In order to validate the thrombosis PRS published by Klarin et al, PRS were also generated for patients genotyped on the Multi-Ethnic Global Array (MEGA) in BioVU, the Vanderbilt University DNA biorepository coupled to de-identified electronic medical records (EMRs). A phenome-wide association study (PheWAS) was performed using phecodes derived using previously described approaches with at least two phecode instances necessary for diagnosis[24,25]. The PheWAS included covariates for the first five principal components, median age, and sex. PheWAS results indicated that the strongest association with thrombosis PRS was with the phecodes “Other venous embolism and thrombosis” (n=3,146 cases, n=40,444 controls, odds ratio 1.291, $p=2.367\times10^{-53}$), “Primary hypercoagulable state” (n=426 cases, n=43,779 controls, odds ratio 1.630, $p=7.367\times10^{-39}$) and “Deep vein thrombosis” (n=1,744 cases, n=40,424 controls, odds ratio 1.314, $p=1.350\times10^{-36}$) (Supplementary Figure S3). Following confirmation of association between PRS and thromboembolism in BioVU, PRS was used to adjust the association of positive functional assay status in random effects meta-analysis of functional assay positive status in the discovery and replication cohorts. Meta-analysis was performed in PLINK v1.90 with adjustment for age, sex, PCs 1-3, and PRS score.

ABO Sequencing

Sequencing was performed to acquire full coverage of the *ABO* gene +/- 5000 base pairs (kbps). Multiplex PCR-based target sequencing was performed at RIKEN IMS, as previously described.[26] Briefly, primers were designed using the Primer 3 software (ver. 2.3.4)[27]. PCR product sizes were designed to be 180-300 bp to cover the amplicon by the sequencing reads. We added CGCTCTTCCGATCTCTG to the 5′ end of the forward primers and CGCTCTTCCGATCTGAC to the 5′ end of the reverse primers to perform second PCR[28]. Multiplex PCR reactions were carried out in 10 µl containing 10 ng genomic DNA, 5 µl 2× Platinum Multiplex PCR Master Mix (Life Technologies), and 0.1 pmol each primer. Cycling

conditions were 95 °C for 2 min, 25 cycles of 95 °C for 30 s, 60 °C for 90 s and 72 °C for 3 min, followed by an additional extension step at 72 °C for 10 min using the GeneAmp PCR System 9700 (Life Technologies).

We modified a previously reported barcode technique to use dual barcodes[28,29]. The PCR reactions were carried out in 10-μl reactions containing 1 μl of the first PCR product, 0.1 U KAPA HiFi HotStart DNA Polymerase (KAPA), 2 μl 5× KAPA HiFi Fidelity Buffer, 0.3 μl 10 mM dNTP, and 0.2 pmol each primer. Cycling conditions were 98 °C for 45 s, 4 cycles of 98 °C for 15 s, 65 °C for 30 s, and 72 °C for 1 min, followed by an additional extension step at 72 °C for 1 min. All second PCR products were pooled for one sequencing run. After each library was purified using Agencourt AMPure XP (Beckman Coulter) to eliminate primer dimers, the library was applied to a bioanalyzer (Agilent Technologies) to check the size distribution and then quantified using the KAPA library quantification kit (KAPA) on an ABI Prism 7700 sequence detection system (Life Technologies). We obtained 2 × 150-bp paired-end reads with dual 8-bp barcode sequences on a HiSeq 2500 instrument. We applied a 'dark cycle' to discard the first 3 bp of both reads that originated from the end of the adapter sequences used for the second PCR. Sequence reads were aligned to GRCh37 using the Burrows-Wheeler Aligner (ver. 0.7.9a) and variants were called using Samtools (ver. 0.1.19) and VarScan software[30,31]. For quality control, individuals were excluded from further analysis if the proportion of covered bases of ≥20 reads in the target region was < 95% and variants were excluded that did not follow Hardy-Weinberg equilibrium ($p<1\times10^{-6}$).

Fine-Mapping of GWAS Association

PAINTOR

The Probabilistic Annotation INtegraTOR (PAINTOR V3.0)[32,33] software was used for the prioritization of variants from *ABO* sequence data. PAINTOR is a Bayesian probabilistic framework algorithm that integrates association strength from summary statistics with functional genomic annotation data to improve accuracy in selecting plausible SNPs. A logistic regression

was performed for *ABO* sequence data using PLINK v1.90 in combined replication and discovery cohorts with adjustment for sex, age, and PCs 1-3. For quality control, a minor allele frequency cutoff of 0.01, Hardy-Weinberg equilibrium p value cutoff of 0.001, and genotyping rate cutoff of 98% was used, consistent with GWAS analyses. PAINTOR requires 3 input files, 1) a Z-score file, 2) a linkage disequilibrium file and 3) an annotations file containing functional information on each variant in the locus. The Z-score file containing the Z-statistic for each variant along with the chromosome number, location, and SNP identifier was derived from PLINK association results. The linkage disequilibrium file, including a pairwise matrix of linkage for each pair of variants, was also generated in PLINK using the --r2 flag. Functional annotation datasets were download from the PAINTOR Github webpage (https://github.com/gkichaev/PAINTOR_V3.0). The functional annotations datasets included in this analysis utilized data from FANTOM5[34,35], GENCODE[36] and Transcription Factor Binding Site (TFBS) datasets. Using the Python script provided by PAINTOR (AnnotateLocus.py), the *ABO* variants were cross-referenced with these functional annotation datasets to generate a matrix of functional annotations. Default parameters for the program were used, including the -enumerate flag set to 3 causal SNPs, -variance flag on the causal effect sizes scaled by sample size set at 30, and -MI flag for Maximum iterations for algorithm to run set at 10. In addition to default specification analysis, we tested the robustness of the results by varying the number of pre-defined causal SNPs. Furthermore, we tested the consistency of causal SNP determination by removing subsets of annotated datasets. The PAINTOR program output posterior probabilities for which variants were likely causal in the dataset (Supplementary Table S3). The output of the program was robust to changes in annotated datasets (FANTOM5, GENCODE, and TFBS) and to changing the number of casual variants for selection using the -enumerate flag.

FUMA

FUnctional Mapping and Annotation of genome-wide association studies (FUMA) is a web-based application to annotate, prioritize and interpret GWAS results[37]. FUMA incorporates

18 biological data repositories and tools to process GWAS summary statistics. FUMA generates data for independently significant variants and their surrounding genomic loci depending on the linkage disequilibrium structure of the surrounding region. Meta-analysis association results from combined discovery and replication cohort GWAS analysis generated in PLINK were input into FUMA's web-based software using the following default settings: 5×10^{-8} minimal p-value for lead SNPs, maximum p=0.05 to be included in the annotation, 250 kb maximum distance between LD blocks, sample size n=5,335, reference population=EUR (1000 Genomes/Phase3), and r^2=0.6 threshold used to define independent significant SNPs. The SNP2GENE function of FUMA was utilized to determine causal variants from our GWAS results. Lead SNPs are identified from GWAS summary statistic data and variants within a defined LD block were then mapped for functional consequences. FUMA integrates ANNOVAR[38] for gene function, Combined Annotation–Dependent Depletion (CADD)[39] for deleteriousness score, and RegulomeDB[40] for potential regulatory functions. Using the pre-selected Bonferroni cutoff, only 1 locus reached the threshold to be analyzed by FUMA. The lead SNP in the region was a C>TC insertion (rs71281258), located on chromosome 9 (9:136143136). Gene function annotations (ANNOVAR) identified the variant located at 9:136132908 (rs8176719) as the most likely casual variant based on LD (r^2) and functional consequences (Supplementary Table S4). No other exonic SNPs were identified from the analysis as being potentially the casual SNP. CADD identified the same SNP (rs8176719) as the top SNP based on deleterious score, with an overall CADD score of 15.84, twice as large as the next closest variant. The RegulomeDB (RDB) score of this top variant (rs8176719) was not available (NA), indicating no evidence exists that this variant affects binding and/or is linked to expression of a gene target, since rs8176719 is an exonic SNP. The highest score for any SNP using the RDB scoring system (range 1-6) was 3a, which was assigned to six variants total. This score indicates a given variant is "less likely to affect binding" but does show transcription factor binding according to the RDB scoring convention[40]. The aggregate results of FUMA indicated the rs8176719 variant as the casual SNP.

SUPPLEMENTARY TABLES

Supplementary Table S1: Quantity of genetic markers prior to imputation and included in GWAS analysis after imputation.

Cohort	Genotyped markers (prior to imputation)[a]	Sample Call Rate (genotyped markers)[a]	Markers included in GWAS (after imputation)[a]	Sample Call Rate (imputed markers)[a]
Discovery (n=4,166)	n=667,956	0.99969	n=9,065,510	0.99998
Replication (n=786)	n=671,575	0.99974	n=9,048,838	0.99998

GWAS indicates genome-wide association study.

[a]Number of markers and sample call rate calculated after all quality control procedures, including Hardy-Weinberg equilibrium p value<0.001 and MAF less than 0.01.

Supplementary Table S2. SNPs previously associated with thromboembolism and their association with positive assay functional status in meta-analysis of discovery and replication cohorts

Gene	SNP	Chr	Position[a]	Risk Allele	EUR Freq[b]	Study Freq.[c]	Study Odds Ratio[d]	Study P value[d]
ABO[41]	rs9411377	9	136145404	A	0.340	0.365	0.07918	0.003703
ABO[42-44]	rs8176719	9	136132908	C	0.395	0.402	0.7581	2.89×10^{-9}
F5[41,45,46]	rs6025	1	169519049	T	0.012	0.027	1.0708	0.6131
VWF[47]	rs1063856	12	6153534	C	0.367	0.362	0.9807	0.6738
FGG[41]	rs2066865	4	155525276	A	0.221	0.252	1.0803	0.1316
STXBP5[47]	rs1039084	6	147635413	G	0.523	0.549	1.0286	0.5309
F2[41,48,49]	rs1799963	11	46761055	A	0.008	0.005	-	-
F11[50]	rs4253417	4	187199005	C	0.395	0.413	1.0111	0.8058
F11[51-53]	rs2036914	4	187192481	C	0.529	0.531	0.9897	0.815
F11[52,53]	rs2289252	4	187207381	T	0.389	0.409	1.0038	0.9334
KNG1[54,55]	rs710446	3	186459927	C	0.417	0.419	0.9806	0.6593
ZFPM2[41]	rs4734879	8	106583124	A	0.708	0.736	0.9042	0.05047
F13A1[56]	rs5985	6	6318795	A	0.242	0.260	1.0097	0.8778
TSPAN15[41]	rs78707713	10	71245276	T	0.874	0.886	0.9352	0.3389
EDEM2[41]	rs10747514	20	33775369	A	0.277	0.323	0.995	0.9168
GP6[57]	rs1613662	19	55536595	A	0.855	0.845	0.9198	0.1811

Chr indicates chromosome; EUR, 1000Genomes European ancestry superpopulation; freq, frequency; SNP, single nucleotide polymorphism.

[a]Position of SNP in GRCh37 genome build.

[b]Frequency of risk allele in 1000Genomes European ancestry superpopulation.[16]

[c]Frequency of risk allele in combined discovery and replication cohort.

[d]Odds ratios and p values generated using random effect meta-analysis of discovery and replication cohorts in logistic regressions adjusted for age, sex, and principal components 1-3 in an additive model.

Supplementary Table S3: Posterior probabilities for top variants identified by PAINTOR

CHR	Position[a]	SNP	Z Score[b]	Posterior Probability[c]
9	136132908	rs8176719	-5.776	0.994
9	136139907	rs60937319	-6.11	0.962
9	136139833	rs141515001	-0.04792	0.381
9	136138841	rs1378007910	3.6	7.80×10^{-5}
9	136147160	rs554833	-5.945	4.11×10^{-5}
9	136142217	rs644234	-6.164	2.17×10^{-5}
9	136142355	rs643434	-6.137	1.97×10^{-5}
9	136147012	rs7046674	3.305	1.49×10^{-5}
9	136147295	rs8176649	3.586	1.44×10^{-5}
9	136143212	rs544873	-5.833	1.09×10^{-5}

CHR indicates chromosome number; SNP, single nucleotide polymorphism.

[a]Position of SNP in GRCh37 genome build.

[b]Z Score generated from PLINK analysis in the combined discovery and replication cohorts with adjustment for age, sex, and PCs 1-3.

[c]Posterior probabilities generated using PAINTOR using PLINK association results using three annotated datasets (FANTOM5[34,35], GENCODE[36], and TFBS)

Supplementary Table S4: Aggregated results from FUMA functional analysis.

CHR	Position[a]	SNP	CADD[b]	RegulomeDB[c]	ANNOVAR[d]
9	136132908	rs8176719	15.84	NA	1
9	136139907	rs60937319	8.308	NA	5
9	136149098	rs8176645	7.176	5	46
9	136137106	rs687289	5.544	NA	3
9	136149229	rs505922	4.81	5	47
9	136142203	rs514659	4.295	6	8
9	136147160	rs554833	4.196	3a	44
9	136149095	rs8176646	3.882	NA	45
9	136137065	rs687621	3.717	NA	2
9	136144427	rs8176663	3.389	NA	22

CADD indicates Combined Annotation–Dependent Depletion score; CHR, chromosome number; FUMA, FUnctional Mapping and Annotation of genome-wide association studies; NA, not available; SNP, single nucleotide polymorphism.

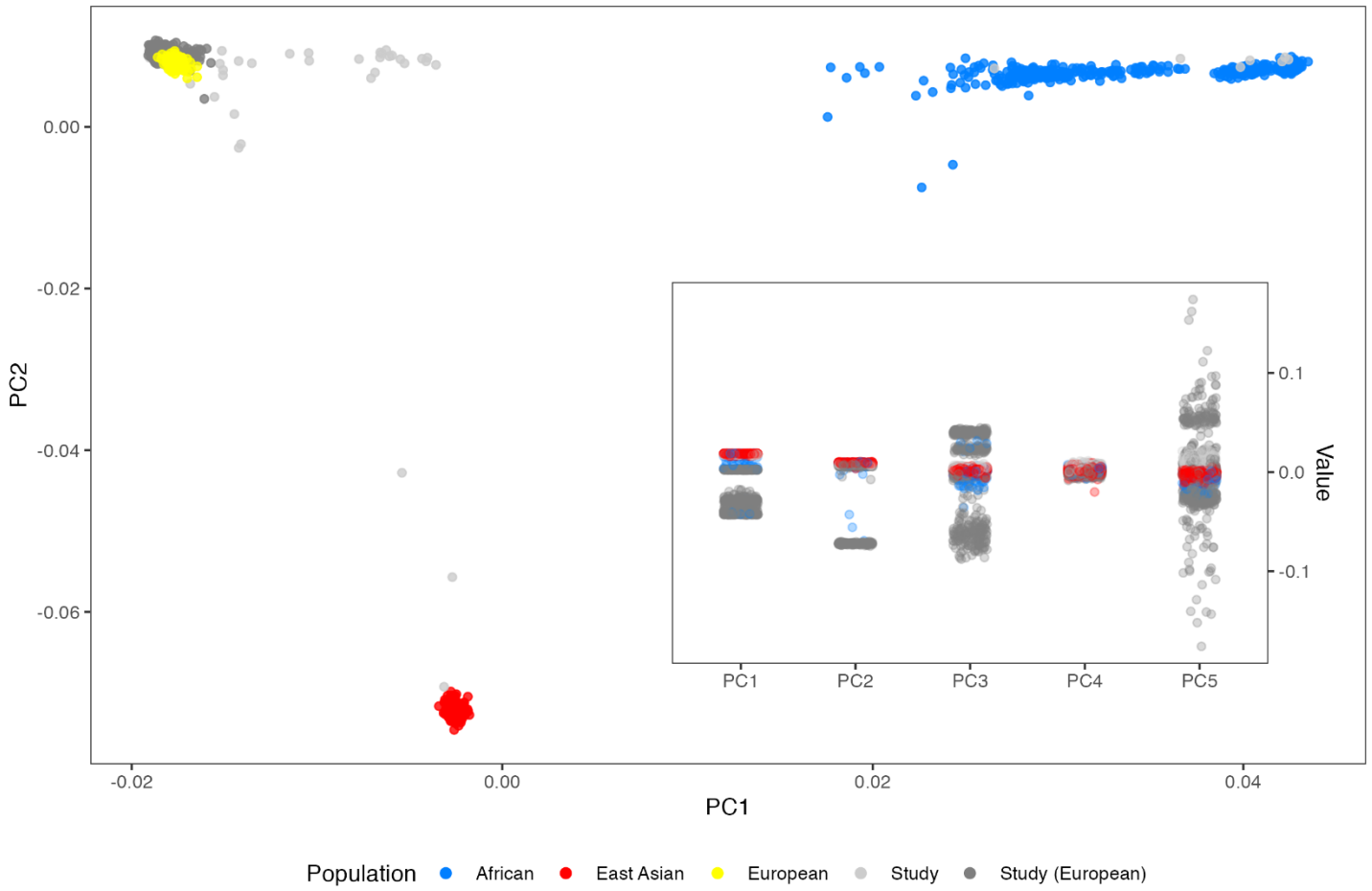
[a]Position of SNP in GRCh37 genome build.

[b]Combined Annotation–Dependent Depletion scores (CADD)[39] for deleteriousness produced from FUMA using meta-analysis result from combined discovery and replication cohort GWAS analysis.

[c]RegulomeDB[40] for potential regulatory functions produced from FUMA using meta-analysis result from combined discovery and replication cohort GWAS analysis (1-6 scale).

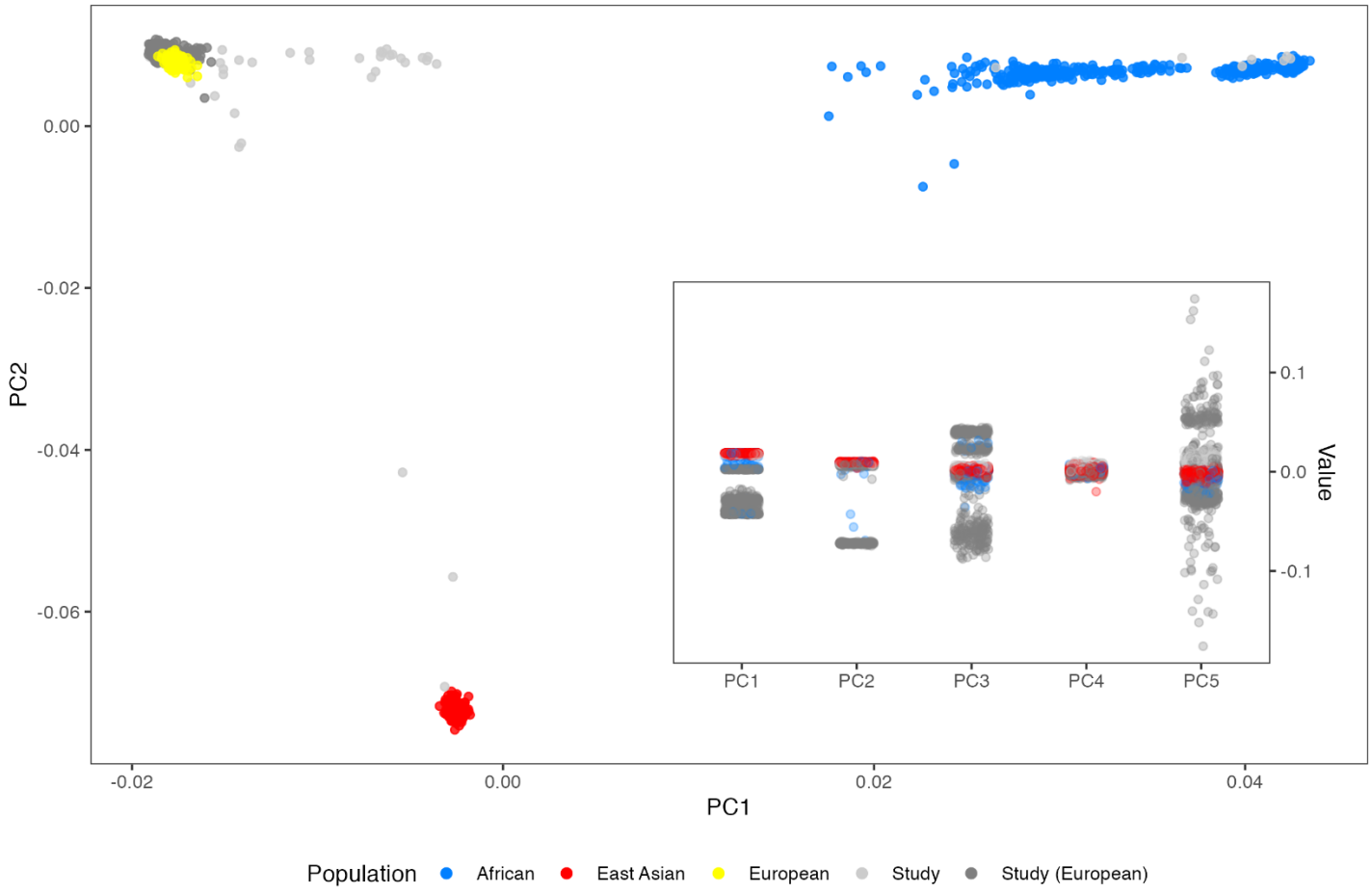
[d]ANNOVAR[38] rankings based on functional annotation via ANNOVAR implemented in FUMA using meta-analysis result from combined discovery and replication cohort GWAS analysis.

SUPPLEMENTARY FIGURES:



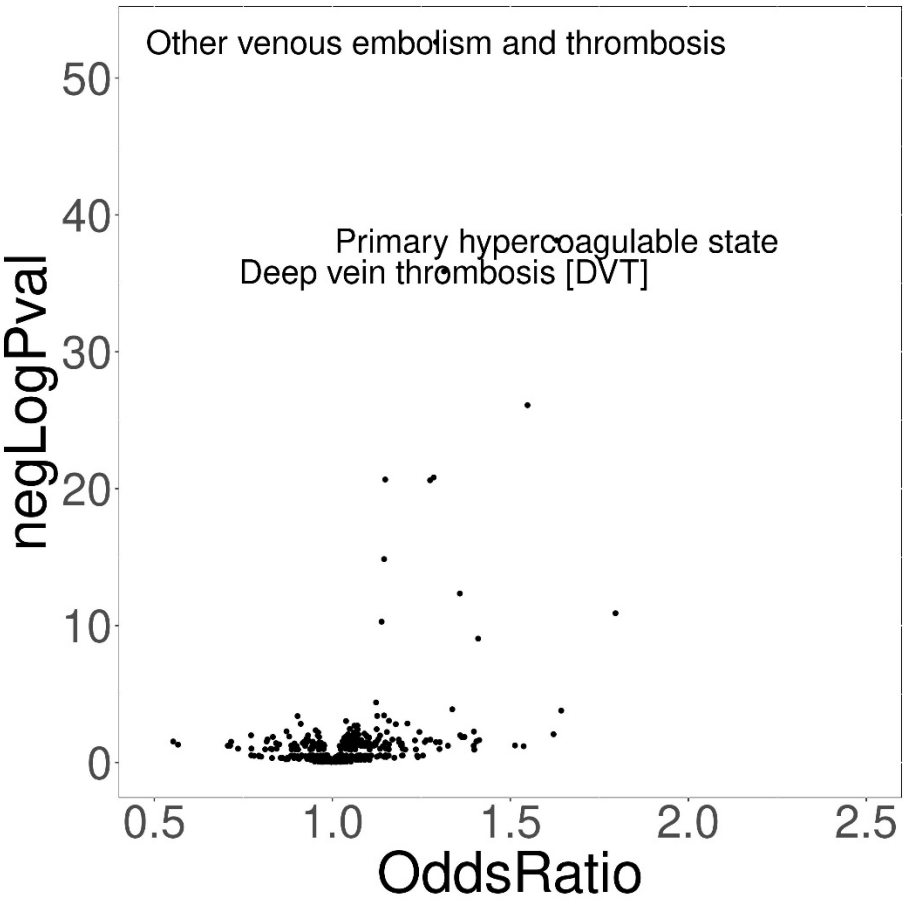
Supplementary Figure S1: Principal components 1 and 2 for discovery cohort plotted alongside HapMap reference populations. Individuals were considered to be of European ancestry if genetically-defined global ancestry, determined using ancestry

informative markers (AIMs) input into STRUCTURE with HapMap reference populations, was at least 80 percent European. PC indicates principal component. The African ancestry population includes the Yoruba HapMap population (YRI), the East Asian ancestry population includes the combined Han Chinese and Japanese HapMap populations (CHB and JPT), and the European ancestry population includes the HapMap Utah residents population with Northern and Western European ancestry (CEU). The inset plot indicates the linear spread of the first five principal components.

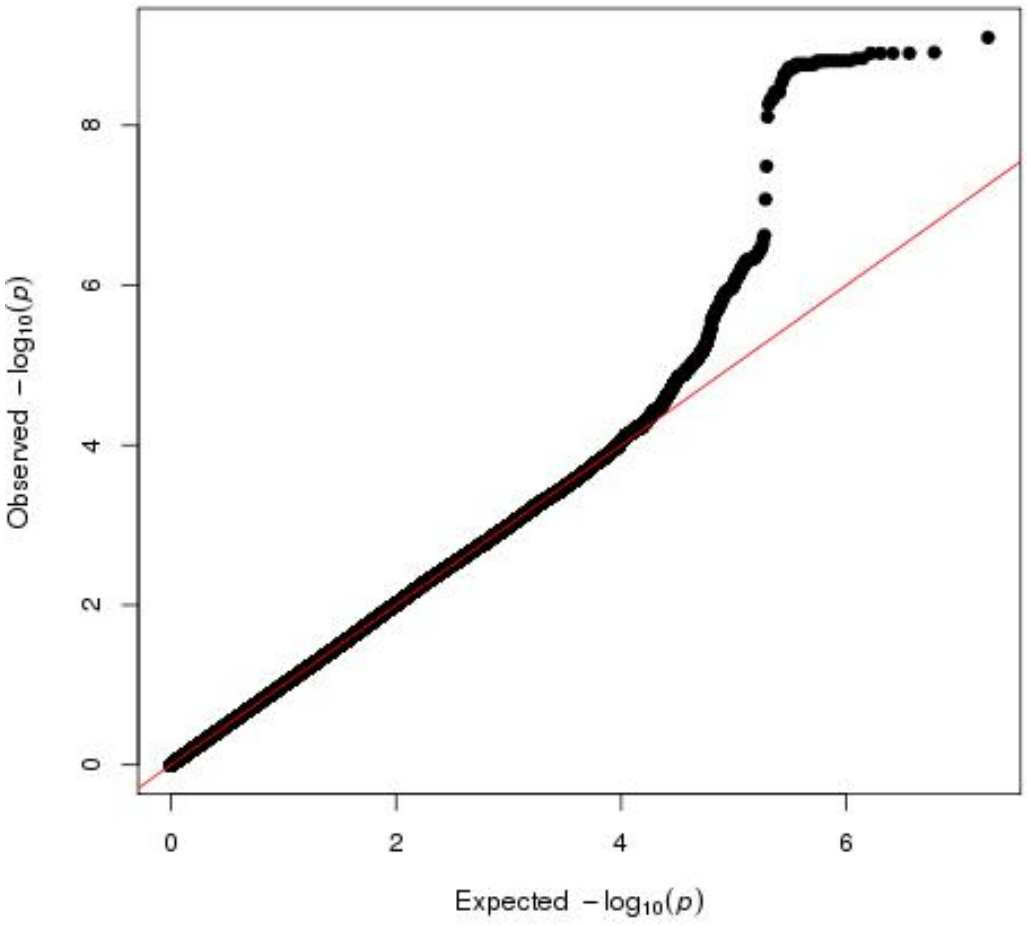


<u>Supplementary Figure S2</u>: Principal components 1 and 2 for replication cohort plotted alongside HapMap reference populations. Individuals were considered to be of European ancestry if genetically-defined global ancestry, determined using ancestry informative markers (AIMs) input into STRUCTURE with HapMap reference populations, was at least 80 percent European. PC

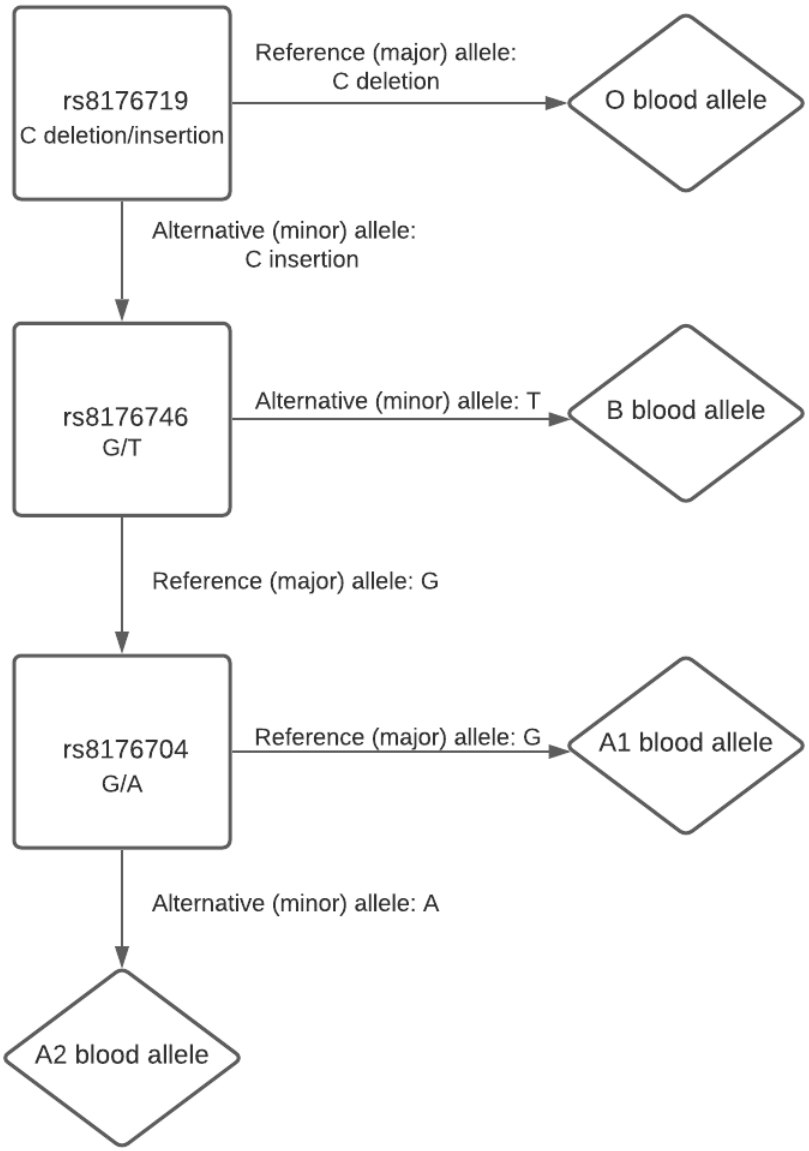
indicates principal component. The African ancestry population includes the Yoruba HapMap population (YRI), the East Asian ancestry population includes the combined Han Chinese and Japanese HapMap populations (CHB and JPT), and the European ancestry population includes the HapMap Utah residents population with Northern and Western European ancestry (CEU). The inset plot indicates the linear spread of the first five principal components.



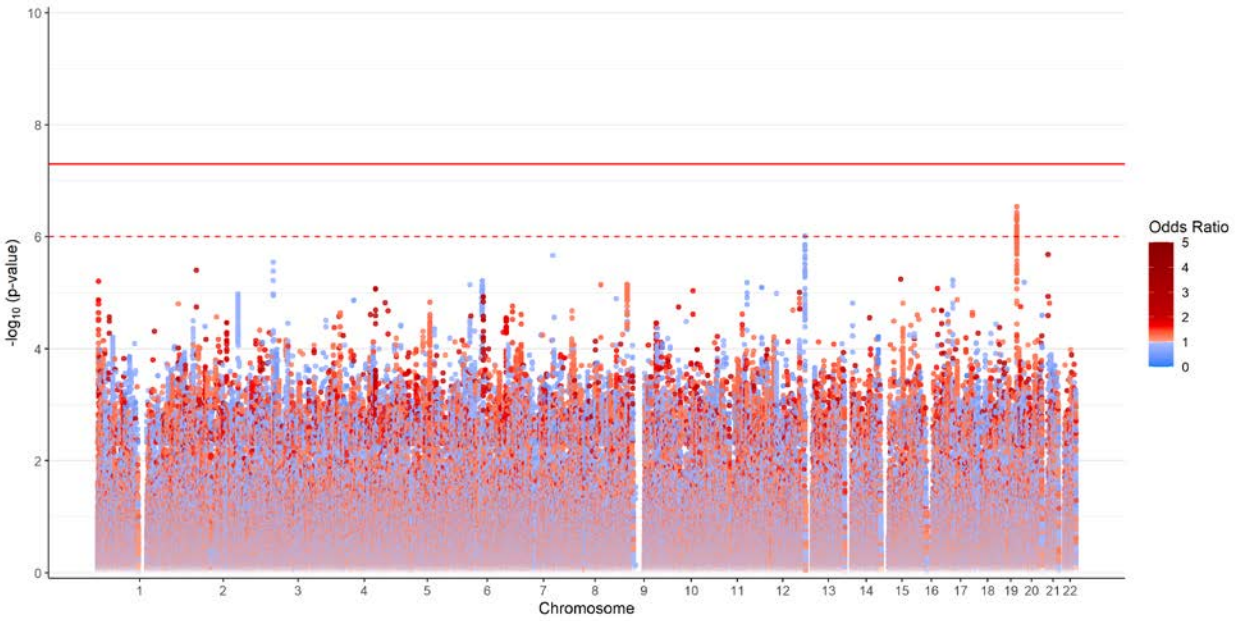
Supplementary Figure S3. Phenotypes associated with the calculated polygenic risk score for venous thromboembolism derived based on Klarin et al.[58] The phenome-wide association study (PheWAS) was performed using phecodes derived using previously described approaches with at least two phecode instances necessary for diagnosis[24,25]. The PheWAS included covariates for the first five principal components, median age, and sex. PheWAS was performed in the BioVU population genotyped on the Multi-Ethnic Global Array.



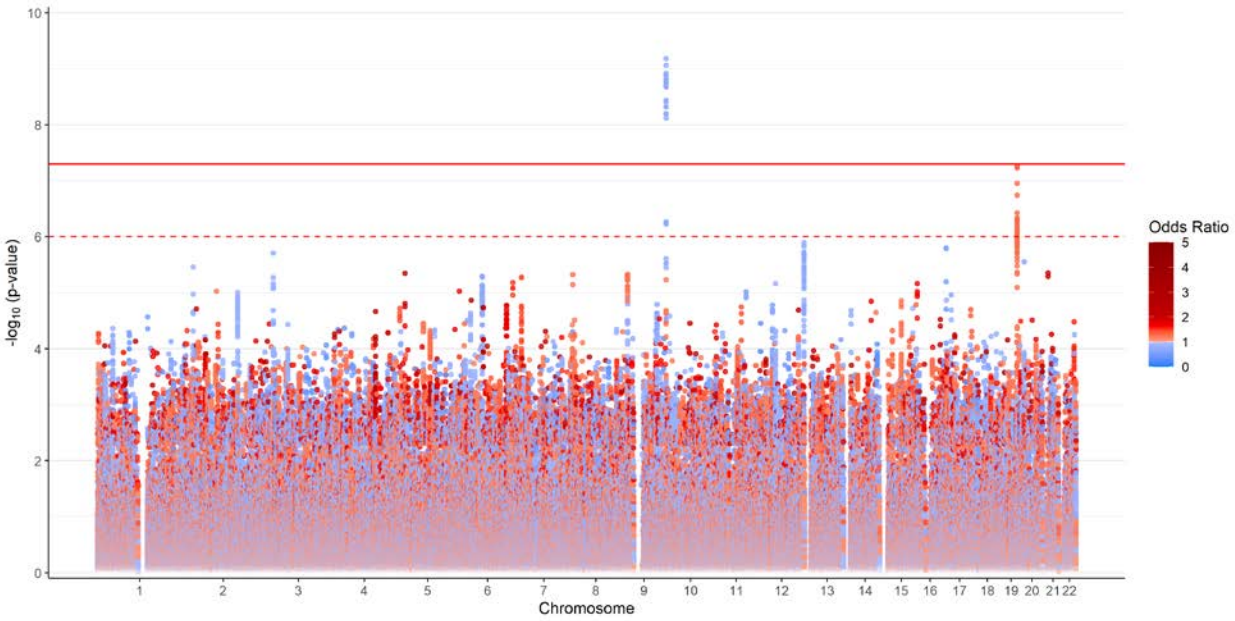
<u>Supplementary Figure S4</u>. QQ plot for genome-wide association with functional assay positive status in the discovery cohort. P values were generated using logistic regression adjusted for age, sex, and principal components 1-3 in an additive model. Observed P values on the –log10 scale are plotted on the left vertical axis and expected P values on the –log10 scale are plotted along the horizontal axis. Genomic inflation (λ) equals 0.998.



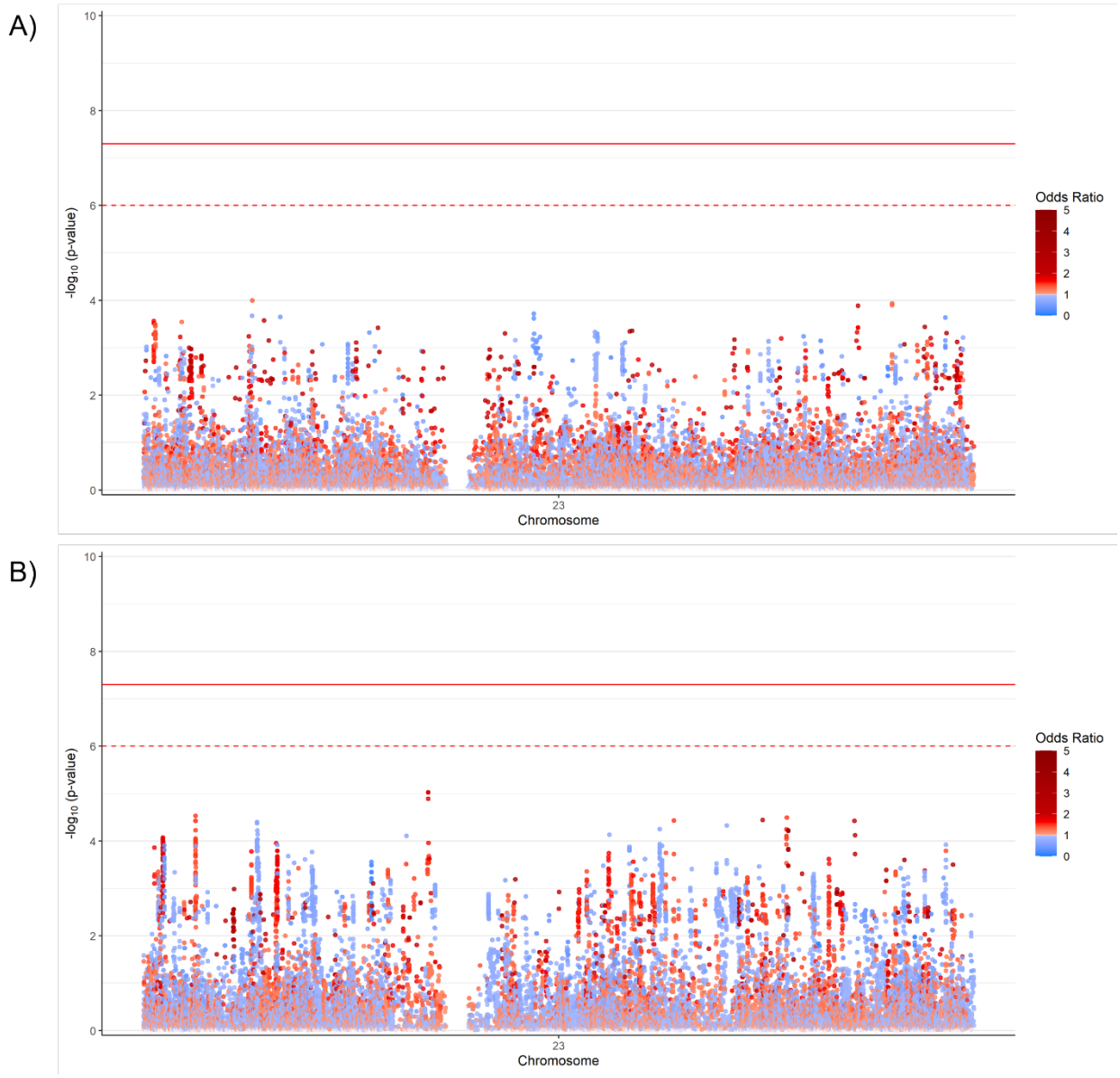
Supplementary Figure S5. Reference and alternative alleles for polymorphisms that encode major *ABO* blood group alleles. In the genome-wide association analysis, all odds ratios characterize risk for the less frequent (minor) allele, which is referred to as the alternative allele. Deletion of the *ABO* rs8176719 C allele produces O blood group individuals, whereas insertion of the rs8176719 C allele produces A, B, or AB blood group individuals



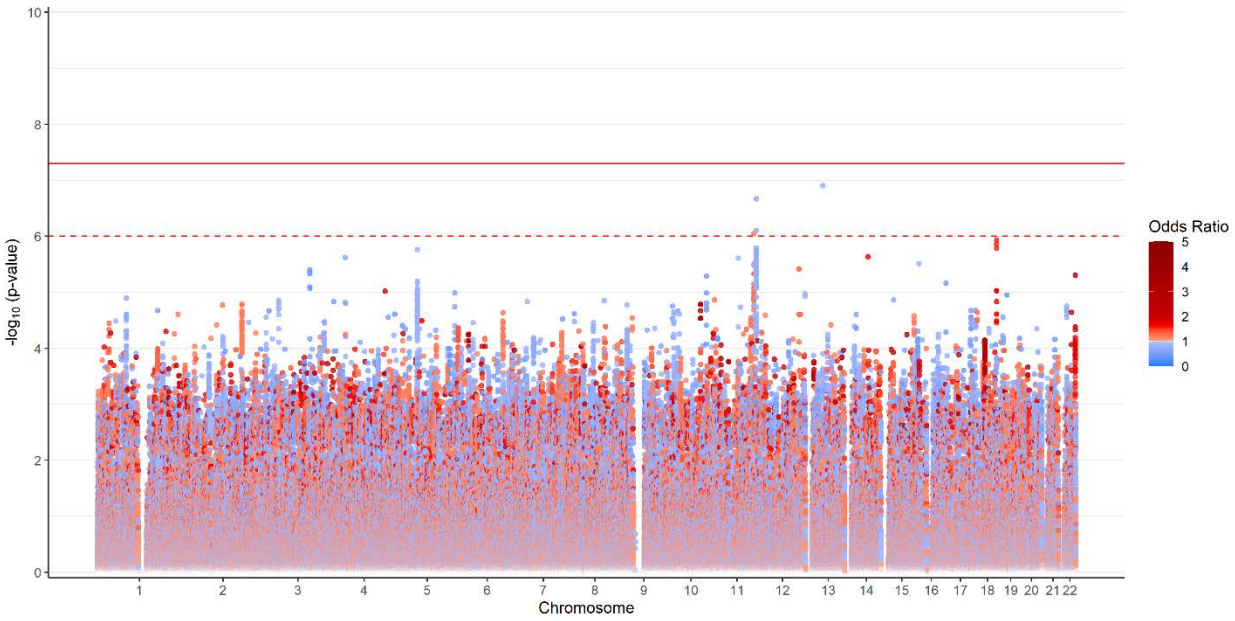
Supplementary Figure S6. Genome-wide association with functional assay positive case status after adjustment for rs8176719 in the discovery cohort. Logistic regressions were adjusted for age, sex, principal components 1-3, and rs8176719 genotypes. P values on the –log10 scale are plotted on the left vertical axis and the chromosomal position is plotted along the horizontal axis. The significance threshold of 5×10^{-8} is indicated by the red horizontal line. Odds ratios are indicated by dot color as described in the legend in the lower right corner.



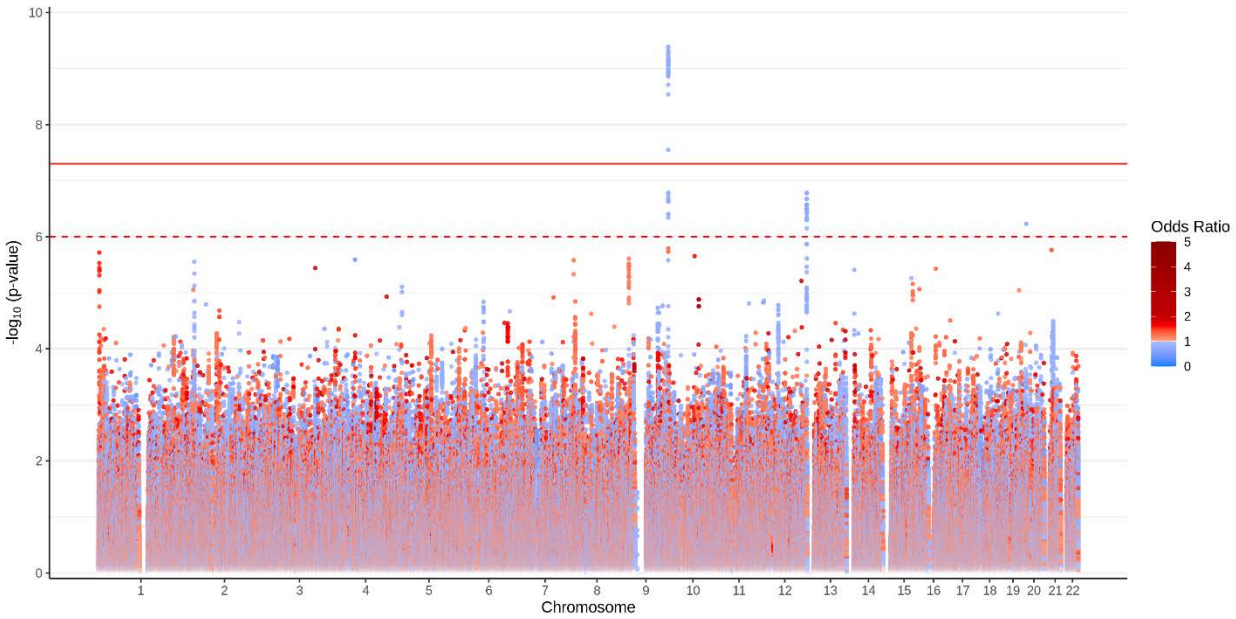
Supplementary Figure S7. Genome-wide association with functional assay positive case status using genomic imputation results from the TOPMed Imputation Server[17]. Logistic regressions were adjusted for age, sex, and principal components 1-3 in an additive model. P values on the –log10 scale are plotted on the left vertical axis and the chromosomal position is plotted along the horizontal axis. The significance threshold of 5×10^{-8} is indicated by the red horizontal line. Odds ratios are indicated by dot color as described in the legend in the lower right corner.



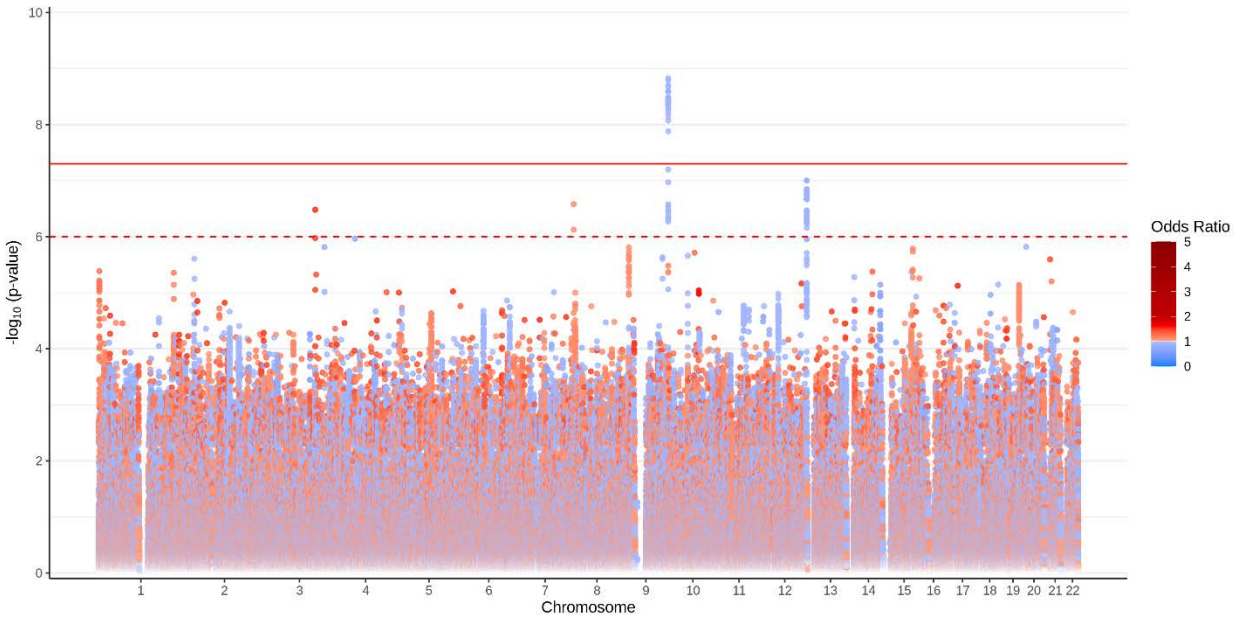
Supplementary Figure S8. X chromosome variant association with functional assay positive case status stratifies by sex in A) females and B) males. Logistic regressions were adjusted for age, sex, and principal components 1-3 in an additive model. P values on the –log10 scale are plotted on the left vertical axis and the chromosomal position is plotted along the horizontal axis. The significance threshold of 5×10^{-8} is indicated by the red horizontal line. Odds ratios are indicated by dot color as described in the legend in the lower right corner.



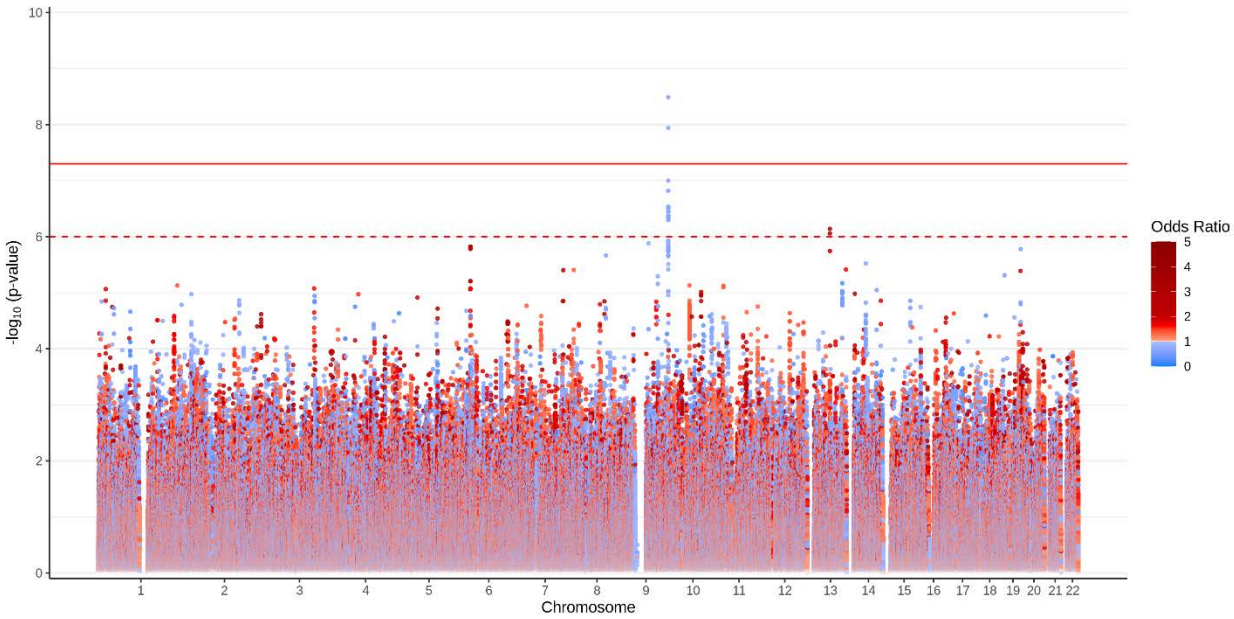
Supplementary Figure S9. Secondary genome-wide association analysis with positive PF4/heparin antibody status as the outcome in the discovery cohort. Logistic regressions were adjusted for age, sex, and principal components 1-3 in an additive model. P values on the –log10 scale are plotted on the left vertical axis and the chromosomal position is plotted along the horizontal axis. The significance threshold of 5×10^{-8} is indicated by the red horizontal line. Odds ratios are indicated by dot color as described in the legend in the lower right corner.



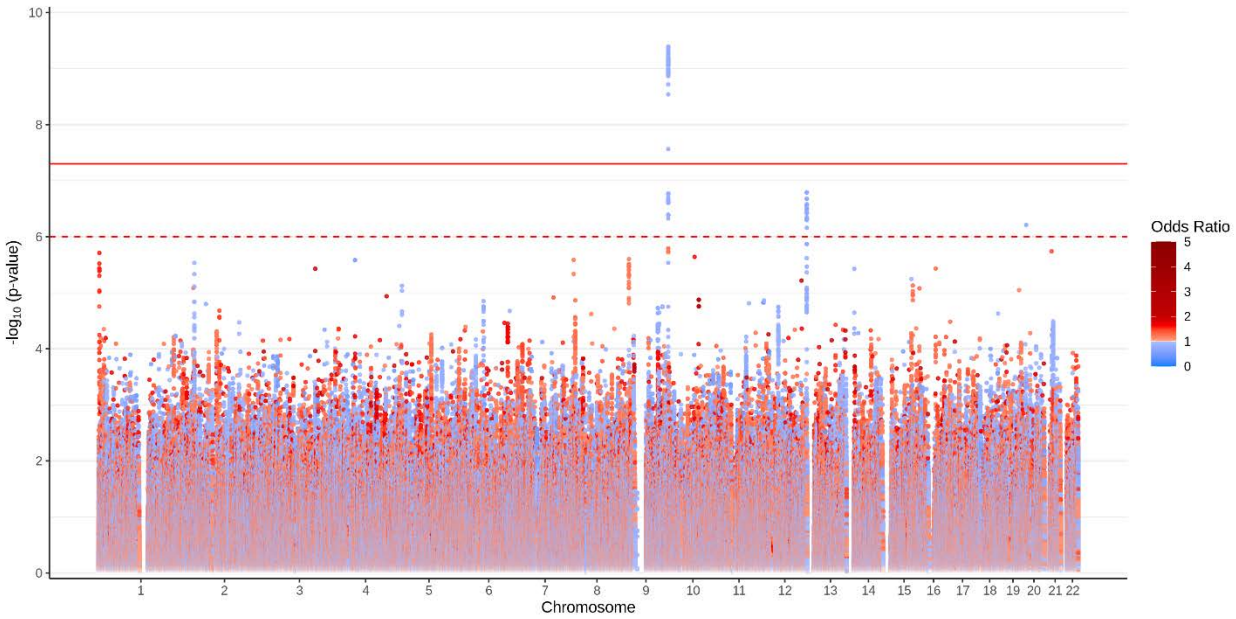
Supplementary Figure S10. Genome-wide association with functional assay positive case status in combined discovery and replication cohorts using random effects meta-analysis of logistic regressions adjusted for age, sex, and principal components 1-3 in an additive model. P values on the –log10 scale are plotted on the left vertical axis and the chromosomal position is plotted along the horizontal axis. The significance threshold of 5×10^{-8} is indicated by the red horizontal line. Odds ratios are indicated by dot color as described in the legend in the lower right corner.



Supplementary Figure S11. Genome-wide association with functional assay positive case status in combined discovery and replication cohorts using multi-ethnic mixed model association implemented in GCTA. All patients with genome-wide data are included regardless of ancestry and relatedness. P values were generated using logistic regression adjusted for age, sex, cohort, principal components 1-3, and cohort (discovery versus replication) in an additive model. P values on the –log10 scale are plotted on the left vertical axis and the chromosomal position is plotted along the horizontal axis. The significance threshold of 5×10^{-8} is indicated by the red horizontal line. Odds ratios are indicated by dot color as described in the legend in the lower right corner.



Supplementary Figure S12. Genome-wide association with functional assay-positive status in discovery and replication cohorts with adjustment for PF4/heparin IgG optical density levels. P values were generated using random effect meta-analysis of logistic regressions adjusted for age, sex, PF4/heparin IgG levels, and principal components 1-3 in an additive model. P values on the –log10 scale are plotted on the left vertical axis and the chromosomal position is plotted along the horizontal axis. The significance threshold of 5×10^{-8} is indicated by the red horizontal line. Odds ratios are indicated by dot color as described in the legend in the lower right corner.



Supplementary Figure S13. Genome-wide association with functional assay-positive status in discovery and replication cohorts with adjustment for thromboembolism polygenic risk scores (PRS). P values were generated using random effect meta-analysis of logistic regressions adjusted for age, sex, PRS, and principal components 1-3 in an additive model. P values on the –log10 scale are plotted on the left vertical axis and the chromosomal position is plotted along the horizontal axis. The significance threshold of 5×10^{-8} is indicated by the red horizontal line. Odds ratios are indicated by dot color as described in the legend in the lower right corner.

REFERENCES

1. Greinacher A, Amiral J, Dummel V, Vissac A, Kiefel V, Mueller-Eckhardt C. Laboratory diagnosis of heparin-associated thrombocytopenia and comparison of platelet aggregation test, heparin-induced platelet activation test, and platelet factor 4/heparin enzyme-linked immunosorbent assay. *Transfusion.* 1994;34(5):381-385.
2. Lo GK, Juhl D, Warkentin TE, Sigouin CS, Eichler P, Greinacher A. Evaluation of pretest clinical score (4 T's) for the diagnosis of heparin-induced thrombocytopenia in two clinical settings. *J Thromb Haemost.* 2006;4(4):759-765.
3. Eekels JJM, Althaus K, Bakchoul T, et al. An international external quality assessment for laboratory diagnosis of heparin-induced thrombocytopenia. *J Thromb Haemost.* 2019;17(3):525-531.
4. Greinacher A, Ittermann T, Bagemuhl J, et al. Heparin-induced thrombocytopenia: towards standardization of platelet factor 4/heparin antigen tests. *J Thromb Haemost.* 2010;8(9):2025-2031.
5. Warkentin TE, Greinacher A, Gruel Y, et al. Laboratory testing for heparin-induced thrombocytopenia: a conceptual framework and implications for diagnosis. *J Thromb Haemost.* 2011;9(12):2498-2500.
6. Juhl D, Eichler P, Lubenow N, Strobel U, Wessel A, Greinacher A. Incidence and clinical significance of anti-PF4/heparin antibodies of the IgG, IgM, and IgA class in 755 consecutive patient samples referred for diagnostic testing for heparin-induced thrombocytopenia. *Eur J Haematol.* 2006;76(5):420-426.
7. Althaus K, Strobel U, Warkentin TE, Greinacher A. Combined use of the high heparin step and optical density to optimize diagnostic sensitivity and specificity of an anti-PF4/heparin enzyme-immunoassay. *Thromb Res.* 2011;128(3):256-260.
8. Gruel Y, Pouplard C, Lasne D, Magdelaine-Beuzelin C, Charroing C, Watier H. The homozygous FcgammaRIIIa-158V genotype is a risk factor for heparin-induced thrombocytopenia in patients with antibodies to heparin-platelet factor 4 complexes. *Blood.* 2004;104(9):2791-2793.
9. Rollin J, Pouplard C, Gratacap MP, et al. Polymorphisms of protein tyrosine phosphatase CD148 influence FcgammaRIIA-dependent platelet activation and the risk of heparin-induced thrombocytopenia. *Blood.* 2012;120(6):1309-1316.
10. Karnes JH, Cronin RM, Rollin J, et al. A genome-wide association study of heparin-induced thrombocytopenia using an electronic medical record. *Thromb Haemost.* 2015;113(4):772-781.
11. Rollin J, Pouplard C, Gruel Y. Risk factors for heparin-induced thrombocytopenia: Focus on Fcgamma receptors. *Thromb Haemost.* 2016;116(5):799-805.
12. Pouplard C, Amiral J, Borg JY, Vissac AM, Delahousse B, Gruel Y. Differences in specificity of heparin-dependent antibodies developed in heparin-induced thrombocytopenia and consequences on cross-reactivity with danaparoid sodium. *Br J Haematol.* 1997;99(2):273-280.
13. Gruel Y, Vayne C, Rollin J, et al. Comparative Analysis of a French Prospective Series of 144 Patients with Heparin-Induced Thrombocytopenia (FRIGTIH) and the Literature. *Thromb Haemost.* 2020;120(7):1096-1107.
14. Yang J, Lee SH, Goddard ME, Visscher PM. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol.* 2013;1019:215-236.
15. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76-82.
16. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467(7319):1061-1073.

17. 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.
18. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945-959.
19. Groot HE, Villegas Sierra LE, Said MA, Lipsic E, Karper JC, van der Harst P. Genetically Determined ABO Blood Group and its Associations With Health and Disease. *Arterioscler Thromb Vasc Biol.* 2020;40(3):830-838.
20. Kolin DA, Kulm S, Christos PJ, Elemento O. Clinical, regional, and genetic characteristics of Covid-19 patients from UK Biobank. *PLoS One.* 2020;15(11):e0241264.
21. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.
22. Klarin D, Busenkell E, Judy R, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. 2019;51(11):1574-1579.
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. 2007;81(3):559-575.
24. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics.* 2010;26(9):1205-1210.
25. Karnes JH, Bastarache L, Shaffer CM, et al. Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. *Sci Transl Med.* 2017;9(389).
26. Momozawa Y, Akiyama M, Kamatani Y, et al. Low-frequency coding variants in CETP and CFB are associated with susceptibility of exudative age-related macular degeneration in the Japanese population. *Hum Mol Genet.* 2016;25(22):5027-5034.
27. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol.* 2000;132:365-386.
28. Harismendy O, Schwab RB, Bao L, et al. Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. *Genome Biol.* 2011;12(12):R124.
29. Forshew T, Murtaza M, Parkinson C, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med.* 2012;4(136):136ra168.
30. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
31. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-2079.
32. Kichaev G, Yang W-Y, Lindstrom S, et al. Integrating Functional Data to Prioritize Causal Variants in Statistical Fine-Mapping Studies. *PLoS Genetics.* 2014;10(10):e1004722.
33. Kichaev G, Roytman M, Johnson R, et al. Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics.* 2017;33(2):248-255.
34. Lizio M, Harshbarger J, Shimoji H, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biology.* 2015;16(1):22.
35. Abugessaisa I, Ramilowski JA, Lizio M, et al. FANTOM enters 20th year: expansion of transcriptomic atlases and functional annotation of non-coding RNAs. *Nucleic Acids Research.* 2021;49(D1):D892-D898.
36. Frankish A, Diekhans M, Ferreira A-M, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Research.* 2019;47(D1):D766-D773.
37. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications.* 2017;8(1).

38. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research.* 2010;38(16):e164-e164.
39. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics.* 2014;46(3):310-315.
40. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research.* 2012;22(9):1790-1797.
41. Klarin D, Busenkell E, Judy R, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nature Genetics.* 2019;51(11):1574-1579.
42. Heit JA, Cunningham JM, Petterson TM, Armasu SM, Rider DN, De Andrade M. Genetic variation within the anticoagulant, procoagulant, fibrinolytic and innate immunity pathways as risk factors for venous thromboembolism. *Journal of Thrombosis and Haemostasis.* 2011;9(6):1133-1142.
43. Wu O, Bayoumi N, Vickers MA, Clark P. ABO(H) blood groups and vascular disease: a systematic review and meta-analysis. *Journal of Thrombosis and Haemostasis.* 2007;6(1):62-69.
44. Bruzelius M, Bottai M, Sabater-Lleal M, et al. Predicting venous thrombosis in women using a combination of genetic markers and clinical risk factors. *Journal of Thrombosis and Haemostasis.* 2015;13(2):219-227.
45. Bertina RM, Koeleman BPC, Koster T, et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature.* 1994;369(6475):64-67.
46. Rinde LB, Morelli VM, Småbrekke B, et al. Effect of prothrombotic genotypes on the risk of venous thromboembolism in patients with and without ischemic stroke. The Tromsø Study. *Journal of Thrombosis and Haemostasis.* 2019;17(5):749-758.
47. Smith NL, Rice KM, Bovill EG, et al. Genetic variation associated with plasma von Willebrand factor levels and the risk of incident venous thrombosis. *Blood.* 2011;117(22):6007-6011.
48. Bergendal A, Bremme K, Hedenmalm K, et al. Risk factors for venous thromboembolism in pre-and postmenopausal women. *Thrombosis Research.* 2012;130(4):596-601.
49. Poort, Sr., Rosendaal F, Reitsma P, Bertina R. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood.* 1996;88(10):3698-3703.
50. Sennblad B, Basu S, Mazur J, et al. Genome-wide association study with additional genetic and post-transcriptional analyses reveals novel regulators of plasma factor XI levels. *Human Molecular Genetics.* 2017:ddw401.
51. Smith NL, Hindorff LA, Heckbert SR, et al. Association of Genetic Variations With Nonfatal Venous Thrombosis in Postmenopausal Women. *JAMA.* 2007;297(5):489.
52. Li Y, Bezemer ID, Rowland CM, et al. Genetic variants associated with deep vein thrombosis: theF11locus. *Journal of Thrombosis and Haemostasis.* 2009;7(11):1802-1808.
53. Austin H, De Staercke C, Lally C, Bezemer ID, Rosendaal FR, Hooper WC. New gene variants associated with venous thrombosis: a replication study in White and Black Americans. *Journal of Thrombosis and Haemostasis.* 2011;9(3):489-495.
54. Morange P-E, Oudot-Mellakh T, Cohen W, et al. KNG1 Ile581Thr and susceptibility to venous thrombosis. *Blood.* 2011;117(13):3692-3694.
55. Wang Q, Cheng G, Wang X, et al. Genetic effects of BDKRB2 and KNG1 on deep venous thrombosis after orthopedic surgery and the potential mediator. *Scientific Reports.* 2018;8(1).

56. Wells PS, Anderson JL, Scarvelis DK, Doucette SP, Gagnon F. Factor XIII Val34Leu Variant Is Protective against Venous Thromboembolism: A HuGE Review and Meta-Analysis. *American Journal of Epidemiology.* 2006;164(2):101-109.
57. Trégouët D-A, Heath S, Saut N, et al. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. *Blood.* 2009;113(21):5298-5303.
58. Klarin D, Busenkell E, Judy R, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat Genet.* 2019;51(11):1574-1579.