Antithrombin, protein C and protein S: Genome and transcriptome wide association

studies identify 7 novel loci regulating plasma levels

Yuekai Ji,MSc1*, Gerard Temprano-Sagrera MSc2*, Lori A Holle PhD3, Allison Bebo MSc4, Jennifer Brody5, Ngoc-Quynh Le MSc5, Michael R Brown4, Angel Martinez-Perez6, Colleen M Sittani PhD6, Pierre Suchon MD PhD7, Marcus E Kleber PhD8, David B Emmert BA9, Ayse Bilge Ozel PhD2, Dre’Von A Dobson BS3, WeiHong Tang PhD13, Dolors Llobet PhD14, Russell P Tracy PhD15, Jean-François Deleuze PhD16,17,18, Graciela E Delgado MSc19, Martin Gögele MSc20, Kerri L Wiggins MS, RD21, Juan Carlos Souto MD, PhD22, James S Pankow PhD23, Kent D Taylor PhD24, David-Alexandre Trégouët PhD25,26, Angela P Moissl MSc10 21,22, Christian Fuchsberger PhD21,22, Frits R Rosendaal23, Alanna C Morrison PhD4, Jose Manuel Soria PhD2, Mary Cushman MD24, Pierre-Emmanuel Morange MD PhD7,8, Winfried März MD10 25, Andrew A Hicks PhD11, Karl C Desch MD26, Andrew D Johnson PhD27, Paul S de Vries PhD2, CHARGE Consortium Hemostasis Working Group, INVENT Consortium, Alisa S Wolberg PhD2, Nicholas L Smith PhD38, 39,30**, Maria Sabater-Lleal PhD2,31 **

Affiliations:

1Cardiovascular Division, Department of Medicine, University of Minnesota, MN, USA, 2Unit of genomics of Complex Disease, Institut d'Investigació Biomedica Sant Pau (IIB SANT PAU), Barcelona, Spain, 3Department of Pathology and Laboratory Medicine and UNC Blood Research Center, University of North Carolina at Chapel Hill, NC, USA, 4Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, TX, USA, 5Department of Medicine, University of Washington, WA, USA, 6Cardiovascular Health Research Unit, Department of Medicine, University of Washington, WA, USA, 7C2VN, INSERM, INRAE, Aix Marseille Univ, France, 8Laboratory of Haematology, La Timone Hospital, France 8SYNLAB MVZ für Humangenetik Mannheim, Germany, 9Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Germany, 10Institute for Biomedicine (affiliated to the University of Lübeck), Eurac Research, Italy, 11Department of Human Genetics, University of Michigan, C.S. Mott Children's Hospital, MI, USA, 12Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, MN, USA, 13Unit of Thrombosis and Hemostasis, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain, 14Department of Pathology and Laboratory Medicine, University of Vermont College of Medicine, VT, USA, 15Centre National de Recherche en Génomique Humaine, CEA, France, 16Centre d'Etude du Polymorphisme Humain, Fondation Jean Daussset, France, 17Laboratory of Excellence on Medical Genomics (GenMed), France, 18The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, CA, USA, 19INSERM UMR 1219, Bordeaux Population Health Research Center, France, 20Institute of Nutritional Sciences, Friedrich Schiller University Jena, Germany, 21Competence Cluster for Nutrition and Cardiovascular Health(nutriCARD) Halle-Jena-Leipzig, Germany, 22Department of Clinical Epidemiology, Leiden University Medical Center, the Netherlands, 23Larner College of Medicine, University of Vermont, VT, USA, 24Synlab Academy, Synlab Holding Deutschland GmbH, Germany, 25Department of Pediatrics, University of Michigan, C.S. Mott Children's Hospital, MI, USA, 26National Heart Lung and Blood Institute, Division of Intramural Research, Population Sciences Branch, The Framingham Heart Study, MA, USA, 27Department of Epidemiology, University of Washington, WA, USA, 28Kaiser Permanente Washington Health Research Institute, Kaiser Permanente, WA, USA, 29Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research and Development, WA, USA, 30Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Center for Molecular Medicine, Stockholm, Sweden

*These authors contributed equally as first authors.
**These authors contributed equally as last authors.

Running title: Genome-wide associations with natural anticoagulants

Corresponding authors:

Maria Sabater-Lleal, PhD
Genomics of Complex Disease Unit, Sant Pau Biomedical Research Institute (IIB Sant Pau)
St Quintí 77-79, 08041, Barcelona
Phone +34932919000; Email: msabater@santpau.cat

Nicholas L. Smith, PhD
Cardiovascular Health Research Unit, University of Washington
1730 Minor Ave, Suite 1360, Seattle WA 98101
Phone: 206-287-2777; Email: nsmith@u.washington.edu

Descriptives: abstract: 245 words; manuscript length: 5,785; 1 table; 4 figures; plus supplemental methods, figures, and tables.
ABSTRACT

Objective: Antithrombin, protein C (PC) and protein S (PS) are circulating natural-anticoagulant proteins that regulate hemostasis and of which partial deficiencies are causes of venous thromboembolism. Previous genetic association studies involving antithrombin, PC, and PS were limited by modest sample sizes or by being restricted to candidate genes. In the setting of the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium, we meta-analyzed across ancestries the results from 10 genome-wide association studies (GWAS) of plasma levels of antithrombin, PC, PS free and PS total.

Approach and Results: Study participants were of European and African ancestries and genotype data were imputed to TOPMed, a dense multi-ancestry reference panel. Each of 10 studies conducted a GWAS for each phenotype and summary results were meta-analyzed, stratified by ancestry. We also conducted transcriptome-wide association analyses and multi-phenotype analysis to discover additional associations. Novel GWAS findings were validated by in vitro functional experiments. Mendelian randomization was performed to assess the causal relationship between these proteins and cardiovascular outcomes.

GWAS meta-analyses identified 4 newly associated loci: 3 with antithrombin levels (GCKR, BAZ1B, and HP-TXNL4B) and 1 with PS levels (ORM1-ORM2). TWAS identified 3 newly associated genes: 1 with antithrombin level (FCGRT), 1 with PC (GOLM2), and 1 with PS (MYL7). In addition, we replicated 7 independent loci reported in previous studies. Functional experiments provided evidence for the involvement of GCKR, SNX17, and HP genes in antithrombin regulation.
**Conclusion:** The use of larger sample sizes, diverse populations, and a denser imputation reference panel allowed the detection of 7 novel genomic loci associated with plasma antithrombin, PC, and PS levels.

**ABBREVIATIONS**

TOPMed: Trans-Omic for Precision Medicine

PC: protein C

PS: protein S

VTE: venous thromboembolism

CAD: coronary artery disease

PAD: peripheral artery disease

IS: ischemic stroke

GWAS: genome-wide association study

TWAS: transcription-wide association study

EA: European ancestry

AA: African ancestry

eQTL: expression quantitative trait locus
INTRODUCTION

Antithrombin, protein C (PC), and protein S (PS) are circulating anticoagulant proteins, and low levels or low activity of these proteins are associated with the risk of venous thromboembolism (VTE)\textsuperscript{1-5}. Variation in the protein-coding genes for antithrombin, PC, and PS (\textit{SERPINC1}, \textit{PROC}, and \textit{PROS1}, respectively)\textsuperscript{6-8} has been studied for decades, and rare mutations have been associated both with low protein levels and with risk of VTE\textsuperscript{6,9-12}. There have been at least 6 agnostic genome-wide association studies (GWAS) for antithrombin, PC, and PS, with sample sizes ranging from 351 (GAIT, antithrombin) to 13,968 (ARIC, PC). For antithrombin, no additional genome-wide significant loci beyond \textit{SERPINC1} were identified\textsuperscript{13,14}. For PC, significant loci at the \textit{GCKR} and \textit{BAZ1B} genes had been identified in European ancestry (EA) populations\textsuperscript{15,16}, and the \textit{CELSR2-PSRC1-SORT1}, \textit{PROC} and \textit{PROCR} loci were identified in both EA and African ancestry (AA) populations\textsuperscript{14,16-18}. For PS, no genome-wide significant associations have been found. In this report, using larger sample sizes, diverse populations, and a denser imputation reference panel, we sought to identify novel genomic loci associated with plasma antithrombin, PC, and PS levels.

METHODS

Overview

We used densely imputed genotypes to perform cross-ancestry (antithrombin and PC) and EA-only (PS) GWAS meta-analyses and attempted replication of the lead variants using available summary data from a proteomics-based study\textsuperscript{19}. This was followed by a multi-phenotype analysis and transcriptome-wide association analyses (TWAS) in EA individuals. For characterization and prioritization of genes, we used colocalization and fine-mapping analyses, and novel GWAS findings were functionally interrogated. Last, we conducted Mendelian randomization (MR) analyses to assess causal relationships with cardiovascular clinical events. Figure 1 is a schematic summarizing our approach.
Study Design and Participating Studies

The setting for the meta-analysis is the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Hemostasis Working Group\textsuperscript{20}. We included data from 10 studies from the US and Europe that measured 1 or more of the 3 natural anticoagulants in plasma, by antigen or activity methods. Study details including genotype and phenotype measurement, study design, population, and baseline time are found in Supplemental Tables S1, S2, and Supplementary Materials\textsuperscript{14,16,21-31}. In total, 27,606 EA and 2,688 AA participants were included. All studies were approved by appropriate research ethics committees and all participants provided informed consent.

Discovery Analysis

Study-Specific Genome-Wide Association Analyses

Each study imputed measured genotypes to the Trans-Omic for Precision Medicine (TOPMed) reference panel before association analyses\textsuperscript{32}. Study-specific quality control was implemented before the analysis. Details about genotyping platforms and specific quality control parameters can be found in Supplementary Table S2. Each study followed a common analysis plan that required performing linear regression within each ancestry group, adjusting for sex, age, principal components, and study-specific variables, which included a kinship matrix when necessary to account for family structure. Residuals from regression were inverse-normal transformed and were re-scaled by the standard deviation (SD) of the pre-transformed values for antithrombin and PS. Because different studies had different unit measures for PC, we did not re-scale by SD, and used the inverse-normal transformed levels for the PC analyses. Details of the measures of the 3 natural anticoagulants can be found in Supplemental Table S2. Associations with imputed genotypes were then tested using an additive genetic model between each imputed dosage and the residuals for each re-scaled (antithrombin, PS) or inverse-normal transformed (PC) phenotype using linear regression and adjusting for all the covariates used in
the phenotype regression. The X chromosome was additionally stratified by sex where women and men were coded as 0, 1, 2 and 0, 2, respectively.

**Population-Specific and Cross-Ancestry Meta-Analysis**

Quality control across studies was conducted using EasyQC\textsuperscript{33}. Details of meta-analysis quality control can be found in Supplementary Materials. We meta-analyzed study-level summary results, first by phenotype measure (antigen or activity), then by ancestry. Only variants appearing in at least 2 cohorts were retained in the final meta-analyses. Cross-ancestry meta-analyses were conducted on those phenotypes that included EA and AA participants (AT and PC). Meta-analyses were performed by 2 analysts in parallel.

The significance threshold\textsuperscript{34} was set at $5 \times 10^{-9}$. A locus was defined as 1 Mb upstream and downstream of the variant with the lowest p-value. Genome-wide significant variants with MAF < 1%, present in 2 cohorts or less, or with inconsistent beta directions between cohorts were not considered.

**Conditional Analysis**

We performed approximate conditional and joint analyses for all variants with MAF > 1% using summary statistics from ancestry-specific meta-analyses using COJO (Conditional & Joint; gcta—cojo--slct)\textsuperscript{35}, implemented in the Genome-Wide Complex Trait Analysis (GCTA) software\textsuperscript{36}, to identify additional independent signals at the associated loci.

**Replication**

We sought for replication of associations for the identified lead variants in an external dataset, using available summary data from DeCODE Genetics (available at https://www.decode.com/summarydata)\textsuperscript{19}. DeCODE Genetics used the SOMAscan multiplexed

\textsuperscript{19}DeCODE Genetics used the SOMAscan multiplexed
proteomics assay to obtain proteomic measurements on 35,559 individuals of Icelandic origin, for which antithrombin, PC and PS data is available. The significance threshold p-value of the replication cohort was set at $4.2 \times 10^{-3}$, after correcting for the number of identified lead variants ($n = 12; 0.05/12 = 4.2 \times 10^{-3}$).

**Transcriptome-Wide Association Analyses**

We used GWAS results and S-PrediXcan and S-MultiXcan\textsuperscript{37,38} to perform transcriptome-wide analyses for each phenotype within the EA populations in order to infer significant associations between the \textit{cis} component of gene expression and the phenotypes. See detailed methods in \textbf{Supplementary methods}. Only tissues with a potential role in the synthesis or regulation of anticoagulants proteins (artery aorta, artery coronary, artery tibial, liver and whole blood) were considered to reduce false positives from more distally related tissues. The significance threshold was established as a Bonferroni correction to the number of genes interrogated: up to 66,745 genes/0.05 = 7.5 x 10$^{-7}$.

**Multi-Phenotype Meta-Analysis**

We jointly analyzed the 4 meta-analyses results (cross-ancestry meta-analyses for antithrombin and PC, and the 2 EA PS meta-analyses) using a multi-phenotype method implemented in the metaUSAT R package 1.17\textsuperscript{39}. Significant multi-phenotype associations were defined as any genome-wide significant lead variants in the multivariate analysis ($p$-values$_{\text{multivariate}}$ for the lead variant < $5 \times 10^{-9}$), that were also nominally significant in a least 2 of the phenotypes individually ($p$-value$_{\text{univariate}}$ < 0.005)\textsuperscript{40}. Additionally, we considered novel variants to be those that were not genome-wide significant for any of the 4 phenotypes individually, or that had not been associated with antithrombin, PC, PS free or total in a previous GWAS for antithrombin, PC, PS free or total. Lead variants for each phenotype found in the discovery (\textbf{Table 1}) were queried using the HaploR R package v4.0.6 to extract functional annotations and biological information.
(Table 1 and Supplementary Table S3). Further details are reported in Supplementary Methods.

Characterization and Prioritization of Candidate Loci

Fine-Mapping and Colocalization

To prioritize causal genes among those residing at associated locus, we performed fine-mapping and colocalization. Detailed methods for fine-mapping and colocalization can be found in the Supplementary Methods.

In vitro Functional Validation

Functional validation of prioritized candidates was performed by in vitro silencing of candidate genes in a liver-derived hepatoblastoma cell (HepG2) expression system. Briefly, HepG2 cells were reverse transfected with small interfering RNA (siRNA) against candidate genes. Cells were counted, and target proteins and genes were characterized by immunoblot of cell supernatants and RT-qPCR, respectively. Normalized, log transformed data were compared using one-way ANOVA with Dunnett’s multiple comparisons test. Details on cell culture, transfection, RNA extraction, RT-qPCR, and immunoblotting methods can be found in Supplementary Methods.

Mendelian Randomization

Two-sample summary statistics-based MR was used to assess the association of genetically determined levels of antithrombin and PC with the risk of thrombotic outcomes, VTE, peripheral artery disease (PAD) (31,307 cases and 211,753 controls), coronary artery disease (CAD) (60,801 cases and 123,504 controls), and ischemic stroke (IS) (60,341 cases and 454,450 controls). Given the small proportion of variance explained by the identified PS
variants we did not investigate PS (PS\textsubscript{free} and PS\textsubscript{total}) in MR analyses because of insufficient number of genetic instruments.

All analyses were performed using the ‘TwoSampleMR’ v0.4.26 and ‘MRPRESSO’ v1.0 R packages. Further details are reported in the Supplementary Methods.

RESULTS

Antithrombin activity (% or IU/mL\texttimes{}100, n = 26,999) or antigen (IU/mL\texttimes{}100; n = 932) was measured in 9 cohorts, PC activity (% or IU/mL\texttimes{}100; n = 6,734) or antigen (μg/mL; n = 12,551) was measured in 8 cohorts, PS Total (PS\textsubscript{total}) activity (% or IU/mL\texttimes{}100 or IU/mL; n = 5,045 ) or antigen (IU/mL or μg/dL; n = 1,363) was measured in 7 cohorts, and PS Free (PS\textsubscript{free}) activity (μg/dL or %; n = 1,998) or antigen (IU/mL\texttimes{}100; n = 2,115) was measured in 6 cohorts. See Supplementary Table S4.

\textit{Antithrombin}

\textit{GWAS}: The antithrombin meta-analysis include 25,243 EA and 2,688 AA participants. After quality control and filtering, 80,168,840 variants remained in the meta-analysis. All λ\textsubscript{GC} for individual GWAS were 1.04 or below for all chromosomes. Additional details about quality control are provided in Supplementary Table S5 and Supplementary Material. Manhattan plots for the overall cross-ancestry meta-analyses are shown in Figure 2. A quantile-to-quantile plot (QQ plot) of p-values for these variants is presented in Supplementary Figure S1 and Manhattan plots for the EA and AA population specific analyses are available at Supplementary Figure S2.
In total, 402 variants in 4 loci, exceeded the established genome-wide significance level in the cross-ancestry analysis, 394 (2 loci) in the EA-specific analysis and 57 (1 locus) in the AA-specific analysis. Forest plots for significant variants can be found in Supplementary Figure S3. Loci at \textit{SNX17-GCKR-NRBP1} (2p23.3), \textit{MLXIPL-BAZ1B-BCL7B} (7q11.23) and \textit{HP-TXNL4B} (16q22.2) were new associations. The association at \textit{HP-TXNL4B} (16q22.2) was only found in the AA population. Lead variants in the cross-ancestry meta-analysis in each region are listed in Table 1 along with the meta-analysis p-value, ancestry specific p-value, effect allele frequency (EAF), beta estimates, and closest gene.

Conditional analyses using the population specific meta-analyses (Supplementary Table S6), identified no additional independent variants on \textit{SNX17-GCKR-NRBP1} and \textit{HP-TXNL4B} surrounding regions. On chromosome 1 locus (\textit{SERPINC1}), we found 1 variant (rs182221508, MAF = 0.0017) intronic to \textit{RABGAP1L} gene (600 kb upstream the lead variant), that was independent from the lead missense variant rs2227624 on \textit{SERPINC1} gene.

Supplementary Table S7 shows the lead variants with the strongest associations in the EA and AA meta-analyses. There was 1 significant locus in the AA population specific analysis at chromosomal position 16q22.2 (\textit{HP-TXNL4B}), which also appeared in cross-ancestry analysis. In the EA-specific population analysis, the results reflected cross-ancestry findings at 1q25.1 (\textit{SERPINC1}) and 2p23.3 (\textit{SNX17-GCKR-NRBP1}), with a different lead variant on chromosome 2: rs4665972, located in an intronic region of \textit{SNX17}, was the lead variant in the cross-ancestry analysis, while rs11127048,150 kb upstream rs4665972 and located in an intergenic region between \textit{SNX17} and \textit{GCKR} genes was the lead variant in the EA-specific analysis. We did not find significant signals at 7q11.23 in the EA-specific analysis. The proportion of variance explained by the independent lead variants was 1.4% in EA and 4.3% in AA, of the total antithrombin variance.
All lead variants from GWAS were replicated in the deCODE summary results derived from SOMAscan measures of these anticoagulants, except for the lead variant of the chromosome 16 locus, that was specific for the AA population and was not present in the DeCODE data (Table 1 and Supplementary Table S8).

**TWAS:** TWAS analyses identified associated genes in 4 different loci (Figure 2A). Associations on chromosomes 1 (SERPINC1), 2 (GKCR) and 7 (MLXIPL), identified by the strongest associated gene in the TWAS, matched associated loci found in the GWAS. Additionally, the FCGRT gene represented a new association on chromosome 19. The smallest GWAS p-value for this region approached significance and was for a rare intronic variant (rs111981233) in FCGRT gene (Figure 2 and Supplementary Table S9) that was replicated in the DeCODE cohort (Table 1 and Supplementary Table S8).

**Fine Mapping:** EA-specific fine-mapping results prioritized the SERPINC1 gene on chromosome 1 and the NRBP1 gene on chromosome 2. Given that FOCUS only prioritizes GWAS hits at TWAS risk loci, loci on chromosomes 16 (only GWAS) or 19 (only TWAS) could not be further explored for gene prioritization. In addition, after correcting for LD and pleiotropic effects, none of genes in chromosome 7 locus was included in the credible set, suggesting a regulation mechanism that does not involve gene expression (Supplementary Table S10).

**Colocalization:** We obtained 2 significant colocalizations in lead variants located in the new antithrombin loci (CPC > 0.8) and gene expression of nearby genes. On chromosome 2, GTF3C2-AS2 (at SNX17-GCKR-NRBP1 locus) gene expression in artery tibial tissue colocalized with antithrombin plasma levels and on chromosome 16 locus, HP gene expression
in liver and whole blood also colocalized with antithrombin plasma regulation (Supplementary Table S11).

**Functional Validation:** We selected 1-3 genes per locus for functional analysis (5 genes total): SNX17, GCKR, NRBP1 (Chr 2), and HP (Chr 16). LCMT2 (Chr 15) was also included for its association in the multi-phenotype analysis at MAP1A locus. We transfected HepG2 cells with siRNA against each candidate gene and confirmed that target genes were knocked down more than 60% using RT-qPCR (data not shown). We then characterized effects of the gene knockdowns on cell count. Finally, we quantified antithrombin expression by immunoblot of cell supernatants and SERPINC1 expression by RT-qPCR. As expected, control experiments showed that treatment of HepG2 cells with lipofectamine (alone) or siRNA against PROC did not significantly alter antithrombin (protein) or SERPINC1 (gene) expression, whereas silencing SERPINC1 significantly suppressed antithrombin and SERPINC1 expression (Figure 4A-B). Quantification of immunoblots revealed that silencing GCKR enhanced, whereas silencing SNX17 and HP suppressed, antithrombin protein production (Figure 4A). The GCKR-dependent increase in antithrombin was associated with a significant increase in SERPINC1 expression, suggesting GCKR negatively regulates antithrombin gene expression (Figure 4B). The SNX17-dependent loss of antithrombin was associated with a significant decrease in SERPINC1 expression, suggesting SNX17 positively regulates antithrombin gene expression (Figure 4B). Interestingly, HP-dependent loss of antithrombin was not accompanied by a significant decrease in SERPINC1 expression (Figure 4B) suggesting that HP modifies antithrombin production in a post-transcriptional manner.

**MR analysis:** We used 4 genetic instruments (Supplementary Table S12) to investigate the association between antithrombin levels and VTE and PAD, and 3 to investigate its association with CAD and IS. We detected a significant deleterious effect of genetically determined low
antithrombin levels and risk of VTE (IVW OR 0.84 [0.72-0.97], P-value: 0.015; Figure 3A).

Sensitivity analyses showed consistent effect in size and direction with MR Egger, MR weighted median, and MR weighted mode (Supplementary Table S13 and Supplementary Figure S4). Leave-one-out sensitivity analyses showed homogeneity of effects among the instruments. No significant results were found for the association of genetically determined antithrombin levels with IS, CAD or PAD (Figure 3A and Supplementary Figure S4).

Protein C

GWAS: The PC meta-analysis included 16,597 EA and 2,688 AA participants. After quality control, 72,929,079 variants were included. All $\lambda_{GC}$ for individual GWAS were 1.04 or below for autosomal chromosomes (1.18 for X chromosome). Additional details about quality control are provided in Supplementary Table S5 and Supplementary Material.

Manhattan and QQ-plots showing the cross-ancestry meta-analysis results are presented in Figure 2 and Supplementary Figure S1, respectively. Briefly, 2,198 variants exceeded the genome-wide significance level in the main analysis, identifying 5 regions associated with PC levels. All loci, located near CELSR2-PRSC1 (1p13.3), PROC (2q14.3), SNX17-GCKR-NRBP1 (2p23.3), MLXIPL-TBL2 (7q11.23), and PROCR (20q11.22) genes, have been previously reported to be associated with PC. Coefficients, p-values, ancestry stratified EAF and p-values, and closest genes are listed in Table 1. Forest plots of significant signals found in the GWAS analysis can be found at Supplementary Figure S5.

In the conditional analysis at 1p13.3 (CELSR2-PRSC1), 2p23.3 (SNX17-GCKR-NRBP1), and 7q11.23 (MLXIPL-TBL2) loci in the EA population, no additional independent variants were identified (Supplementary Table S6). Within the PROC locus on chromosome 2, an additional independent variant (rs74392719, MAF = 0.01, 300 bases upstream of the lead variant) was
identified in the EA population, located within the **PROC** gene. Finally, an additional independent variant (rs6060300, MAF = 0.2, 13 kb upstream of the lead variant) was found in the EA population, intronic to **PROCR**.

No significant heterogeneity was found in the direction or magnitude of beta coefficients for any of the lead variants associated with PC, within or between ancestries. AA and EA population-specific results are shown in **Supplementary Table S7** and **Supplementary Figure S2**. The AA population analysis had findings at 2q14.3 (**PROC**) and 20q11.22 (**PROCR**); the EA population analysis recapitulated all the candidate loci found in cross-ancestry analysis. The proportion of variance explained by the identified independent variants was 12.7% in EA and 7.4% in AA. The lead variants at the **PROCR** locus (rs11907011 and rs867186) alone explain 9.5% and 9% of the total variance in the EA and AA meta-analyses, respectively. All lead variants from GWAS were replicated in the DeCODE data (**Table 1** and **Supplementary Table S8**).

**TWAS**: For PC levels, TWAS (**Figure 2B**) identified associated genes at 6 loci, matching all loci found in the cross-ancestry and EA GWAS, of which, the most significant based on TWAS z-score values were **PSRC1** (chromosome 1, **CELSR2-PRSC1** locus), **GCKR** (chromosome 2, **SNX17-GCKR-NRBP1** locus), **PROC** (chromosome 2), **MLXIPL** (chromosome 7, **MLXIPL-TBL2** locus) and **PROCR** (chromosome 20). Additionally, 3 new associations with PC were found in 1 locus on chromosome 15 for **GOLM2**, **LCMT2** and **CATSPER2** genes (**Supplementary Table S9**) (**Table 1** and **Supplementary Table S9**).

**Fine Mapping**: Fine-mapping results for PC prioritized the **PSRC1** gene on chromosome 1, **NRBP1** and **PROC** on chromosome 2 (**SNX17-GCKR-NRBP1** and **CELSR2-PRSC1** locus,
respectively), MLXIPL and TBL2 on chromosome 7 (MLXIPL-TBL2 locus), and PROCR on chromosome 20. (Supplementary Table S10).

**MR analysis**: For PC, 4 variants were initially selected as genetic instruments. After examination of pleiotropic effects, the variant at the PROCR gene (rs1799809) was excluded to avoid violations of MR assumptions. Moreover, additional evidence indicates that this variant is strongly associated with several hemostasis and thrombosis phenotypes and has opposite effect directions for venous and arterial thrombosis reflecting distinct pleiotropic biological mechanisms. Details of selected genetic instruments can be found in Supplementary Table S12. There was a significant deleterious effect of genetically determined lower PC levels on VTE and CAD risk (VTE IVW OR:0.83 (0.76-0.92), P-value: < 0.001; CAD IVW OR: 0.92 (0.84-0.99), P-value: 0.031; Figure 3B). Sensitivity analyses showed consistent significant associations (Supplementary Table S13 and Supplementary Figure S4). No significant associations were found between genetically determined PC with PAD or IS (Figure 3B and Supplementary Figure S4).

**Protein S**

**GWAS**: The PS meta-analysis included 4,113 EA individuals in PS\textsubscript{free} analyses and 6,408 EA individuals in PS\textsubscript{total} analyses. A total of 19,791,246 variants were investigated in the analysis of PS\textsubscript{free} and 25,365,467 in the analysis of PS\textsubscript{total}. All $\lambda_{GC}$ for individual GWAS were 1.04 or below for autosomal chromosomes (1.19 for X chromosome). Additional details about quality control are provided in Supplementary Table S5 and Supplementary Material. Manhattan and QQ-plots describing the main results are shown in Figures 2C/D and Supplementary Figure S1 for PS\textsubscript{free} and PS\textsubscript{total}, and main associated variants are listed in Table 1. Forest plots of significant signals for PS\textsubscript{free} and PS\textsubscript{total} can be found in Supplementary Figure S8.
We identified 1 novel genome-wide significant locus associated with PS\textsubscript{free} and PS\textsubscript{total} near ORM1 and ORM2 genes (9q32) and a known association located near PROS1 gene (3q11.1) for PS\textsubscript{free}. The lead variant at PROS1 locus (rs121918472, EA P-value = 2.04 x 10\textsuperscript{-16}, PS\textsubscript{free} EAF (G) = 0.0108) was a missense variant located in the protein S coding gene PROS1. In our analysis, this variant was associated with PS\textsubscript{free} level, but genome-wide significance was not observed in PS\textsubscript{total} (PS\textsubscript{total} P-value = 2 x 10\textsuperscript{-4}) although there was a consistent direction of effect.

Nominally significant heterogeneity p-values were detected in the ORM1/ORM2 locus lead variant (PS\textsubscript{total} Heterogeneity P-value = 0.03), indicating minor differences between the 2 measurement methods. No additional independent variants were found with conditional analyses (Supplementary Table S6). The variance explained by the identified variants in PS\textsubscript{free} is 6% of the total variance of PS\textsubscript{free} while the variance explained by the unique identified variant in PS\textsubscript{total} is 1% of the phenotypic variance for PS\textsubscript{total}.

Variants at both loci replicated in the DeCODE data (Table 1 and Supplementary Table S8).

**TWAS:** PS\textsubscript{free} TWAS results recapitulated the 2 significant GWAS associations at chromosomes 3 (PROS1) and 9 (ORM2) and additionally revealed a new association at MYL7 gene on chromosome 7 (Figure 2 and Supplementary Table S9).

**Fine Mapping:** Fine-mapping results did not prioritize any genes for PS\textsubscript{free} or PS\textsubscript{total}.

**Colocalization:** There was a significant colocalization for both PS phenotypes and ORM2 gene expression in liver (Supplementary Table S11).

**MR analysis:** Given the small proportion of variance explained by the limited number of genetic instruments (< 3), we did not investigate PS (PS\textsubscript{free} and PS\textsubscript{total}) in MR analyses.
Antithrombin, Protein C and S Multi-phenotype Analysis

Multi-phenotype analyses between antithrombin, PC, PS\textsubscript{free} and PS\textsubscript{total} revealed 1 additional novel GWAS association close to the \textit{MAP1A} gene\textsuperscript{48}, on chromosome 15 (Table 1), found in the PC TWAS (\textit{GOLM2-LCMT2-CATSPER2} locus in PC). The lead variant is a missense variant on the \textit{MAP1A} gene (rs55707100, P-value = 1.64 x 10\textsuperscript{-13}, EAF EA [T] = 0.03, EAF AA [T] = 0.0042) that was nominally associated in the GWAS for antithrombin and PC individually (antithrombin P-value = 1.04 x 10\textsuperscript{-6}, PC P-value = 4.76 x 10\textsuperscript{-8}) and was not significantly associated to either of the PS phenotypes (PS\textsubscript{total} P-value = 0.2717, PS\textsubscript{free} P-value = 0.9937). The colocalization results were significant (CPC > 0.8) between antithrombin and PC, suggesting the existence of a common variant as regulator of both phenotypes.

DISCUSSION

In this study, we performed GWAS for 4 natural anticoagulant hemostasis phenotypes (antithrombin, PC, PS\textsubscript{total}, PS\textsubscript{free}) using larger sample sizes and better imputation panels than previously reported and detected 4 novel associations: 3 loci for antithrombin (\textit{SNX17-GCKR-NRBP1}, \textit{MLXIPL-BAZ1B-BCL7B}, and \textit{HP-TXNL4B}) and 1 locus for PS (\textit{ORM1-ORM2}). For 3 genes within the newly associated loci with antithrombin (\textit{SNX17}, \textit{GCKR}, and \textit{HP}), \textit{in vitro} gene silencing in liver cell experiments provided functional evidence. Using TWAS methods, we detected 3 more novel associations that did not reach significance in individual GWAS: \textit{FCGRT} for antithrombin; \textit{GOLM2} for PC; and \textit{MYL7} for PS. Using MR, we also identified a causal relationship of antithrombin and PC levels with VTE, and of PC levels with CAD. This investigation elucidated genetic regulation of the anticoagulant pathway and provides new information that could identify therapeutic targets in VTE prevention or treatment.
Additionally, we replicated 7 known loci. These loci are SERPINC1 for antithrombin\textsuperscript{6}; CELSR2-PRSC\textsubscript{16}, PROC\textsubscript{14,16,17}, SNX17-GCKR-NRBP\textsubscript{14,16}, MLXIPL-TBL\textsubscript{214,16}, and PROCR\textsubscript{14,16-18} for PC; and PROS1 for PS\textsuperscript{8}. Two of the PC loci, SNX17-GCKR-NRBP1 and MLXIPL-BAZ1B-BCL7B-TBL2, also had novel associations with antithrombin, demonstrating some genetic overlap between different anticoagulant proteins. This was also reflected in the multi-phenotype analysis results where MAP1A was identified.

**Characterization of Novel Loci**

**Antithrombin-associated Loci**

More than 45 rare variants within the SERPINC1 gene have already been described using non-GWAS approaches\textsuperscript{49}. Our lead variant, rs2227624, is a known missense variant causing a Val to Glu amino-acid substitution that leads to antithrombin deficiency\textsuperscript{50-52} and increases risk of VTE\textsuperscript{53}. We identified a second independent variant at this locus at RAPBGAP1L and believe this rare variant is likely in LD with a SERPINC1 variant but we were unable to demonstrate this due to the limitation of the CGTA conditional analyses.

On chromosome 2, lead variants in locus SNX17-GCKR-NRBP1 differed by ancestry. In the cross-ancestry analysis, rs4665972 was in an intronic region of SNX17 whereas, in the EA-specific analysis, the lead variant (rs11127048) was located in an intergenic region between the SNX17 and GCKR genes. Neither rs4665972 nor rs11127048 were significant in AA population suggesting that these variants are tagging an association within a large LD block in EA population. Consistent with this observation, conditional results indicate that the lead variant (rs4665972) is the only independent variant on this locus. Given limited power in the AA-specific analysis, we could not refine the region with AA data (Supplementary Figure S10).
Functional validation in liver-derived cells suggest that *SNX17* positively, and *GCKR* negatively, alters plasma antithrombin levels via effects on *SERPINC1* expression. In contrast, *HP* appears to suppress antithrombin levels through an as-yet unidentified post-transcriptional mechanism.

*SNX17* is a regulator of low density lipoprotein (LDL) receptors and has not been previously associated to antithrombin levels but has been associated with CAD. *GCKR* is a highly pleiotropic gene, that has been found significantly associated to PC, Factor VII (FVII), Factor XI (FXI) and C-reactive protein (CRP) in previous GWAS meta-analyses. In previous candidate gene studies, variant rs1260326 in *GCKR* was found to be related to multiple cardiometabolic traits, including total and LDL cholesterol, fasting plasma glucose, liver fat content and metabolic syndrome, suggesting that *GCKR* might act as a broad regulator of hepatocyte function.

On chromosome 7, the lead variant (rs13244268) was located in an intronic region of *BAZ1B* gene and was only significant in the EA population. This gene has been previously associated with PC and in our PC meta-analysis, but not with antithrombin. rs13244268 was also found significant in bivariate and univariate GWAS of CRP and high-density lipoprotein. TWAS results confirmed an association between *BCL7B* and *MLXIPL* genes and antithrombin levels. Given the differences in LD blocks observed for this region in different populations, we sought to confirm the most plausible candidate genes in this locus with *in vitro* silencing studies in liver cells. Within the 3 closest genes in the region (*BAZ1B, MLXIPL* and *BCL7B*), *BAZ1B* and *BCL7B* are involved in chromatin remodeling and DNA repair and were not prioritized for functional validation. *MLXIPL* is a highly expressed transcription factor in liver that activates triglyceride synthesis in response to carbohydrates. However, this gene is substantially downregulated in HepG2 cells, and we therefore excluded it from further validation. As such, we did not have a good candidate for the functional work.
The lead variant on \textit{HP-TXNL4B} locus (rs5471) was in an intronic region of the \textit{TXNL4B} gene and 5' UTR of the \textit{HP} gene and was only significant in the AA population. Colocalization results performed using cross-ancestry data suggested the existence of a common regulatory variant between \textit{HP} gene expression and antithrombin levels in liver and whole blood, and suggested that higher expression of \textit{HP} in liver and blood were associated with higher levels of antithrombin in plasma. In the same direction, functional validation in HepG2 cells suggested a significant reduction of antithrombin levels upon \textit{HP} silencing. \textit{HP} codes for haptoglobin (Hp), which serves as a binding protein of hemoglobin, and affects the release of hemoglobin from red blood cells\textsuperscript{65}. Its phenotype Hp2-2 was identified as a potential regulator of inflammation and reverse cholesterol transportation and has been suggested to have higher prevalence in VTE patients\textsuperscript{66-68}. Binding of Hp to hemoglobin could prevent the generation of oxidized LDL\textsuperscript{69} from the activation pathway of free hemoglobin. Oxidized LDL could increase prothrombinase activity in vivo\textsuperscript{70}, and affect the level of factors VII, IX and XI\textsuperscript{71}. Overall, previous evidence suggests a potential role of Hp in the inflammation-induced thrombosis, and our results suggest \textit{HP} is a potentially direct regulator of antithrombin production.

Finally, TWAS results suggested a novel locus associated to antithrombin levels on the \textit{FCGRT} gene. Colocalization results suggested the existence of a common regulatory variant between antithrombin levels and the expression of \textit{FCGRT}, \textit{RPS11} and \textit{RCN3} in the aorta, tibial artery, and whole blood. In GWAS analysis, rs111981233 (intronic to \textit{FCGRT}) nearly reached genome-wide significance levels. \textit{FCGRT} encodes a receptor that binds immunoglobulin G and transfers immunoglobulin G antibodies from mother to fetus across the placenta\textsuperscript{72} and previous studies demonstrate that \textit{FCGRT} is also expressed in the liver\textsuperscript{73,74}. Additional work is needed to further elucidate the role of this gene in antithrombin regulation.
Protein C-associated Loci

We found 5 loci associated with PC in the present GWAS meta-analysis, all of which had been previously described. In addition, 3 genes (GOLM2, LCMT2, and CATSPER2) were associated in a novel locus on chromosome 15 in the TWAS analysis. GOLM2 encodes for a transmembrane protein predicted to colocalize in the Golgi apparatus with no known function; LCMT2 acts as a checkpoint regulator in the cell cycle; and CATSPER2 codes for a protein crucial for correct function of sperm cells. Interestingly, a variant near this locus was significant when in a multi-phenotype GWAS analysis, and colocalization results suggested the existence of a common variant between antithrombin and PC. The lead variant in the multi-phenotype analysis is a missense variant located on MAP1A gene that has been associated with lipid metabolism and platelet count. MAP1A had low expression in hepatocytes and could not be included for further functional analyses.

Protein S-associated Loci

For PS, genome-wide associations found at the ORM1-ORM2 locus represented novel findings for both PS phenotypes, and colocalization analysis suggests the existence of a common regulatory variant between PS_free and PS_total levels and ORM2 expression in liver. ORM1 is responsible for encoding acute phase plasma protein orosomucoid (ORM, also known as α1-acid-glycoprotein, AGP), which is increased with acute inflammation. Previous genetic results suggested that ORM1 was associated with thrombin generation potential and the discovery was further confirmed with in vitro experiments. ORM1 has also been associated with cell-free DNA levels in plasma, a surrogate marker of neutrophil extracellular traps that contribute to immunothrombosis. Moreover, AGPs encoded by the ORM1 and ORM2 genes strongly bind to the vitamin K antagonist warfarin that reaches circulation, suggesting that these genes could be relevant in regulating the response to oral anticoagulation. Supporting this hypothesis, ORM1, ORM2 and PROC were nominally associated with warfarin dose requirement in a study.
of candidate gene analysis with 201 patients. This is interesting, since it is widely known that one of the challenges in oral anticoagulation is the wide variation in response among patients. Confirming novel genomic regulators of anticoagulant response could help explain the mechanisms of action of these drugs and move towards a personalized treatment based on genomic background. Our results suggest an involvement of ORM genes in PS regulation.

MYL7, associated with PS levels in TWAS analyses, is the coding gene for myosin light chain 7, and is related to calcium ion binding activity. Variants in this gene have been associated with fasting glucose levels and type II diabetes, probably for their proximity to the glucokinase (GCK) gene, which lies 1.9 kb upstream of MYL7, and is essential for producing glucose-6-phosphate. Variants at GCK have been associated with multiple types of diabetes and hemoglobin A1c measurement.

Implication with for disease outcomes:

The present MR results confirm a causal relationship between genetically determined plasma levels of antithrombin and PC with VTE events, and for PC with CAD outcomes. Specifically, we observed a 19% VTE risk increase per 1 SD decrease in antithrombin plasma levels, a 20% VTE risk increase per 1 SD decrease of PC plasma levels, and a 9% CAD risk increase per 1 SD decrease in PC plasma levels. Our findings of a causal relationship of antithrombin and PC with VTE agree with previous epidemiological studies that report an increased VTE risk in individuals with deficiencies of these anticoagulants. The causal relationship between PC and CAD was also reported in previous epidemiological and MR studies. Overall, these results support previous data suggesting that AT and PC are relevant proteins that regulate the risk of VTE, confirmed the causal association between PC levels and CAD, and corroborated that intervention in the anticoagulant system could be considered for VTE or CAD prevention.
**Strengths and limitations:**

A major strength of this study is in the modestly large sample size, including around 30,000 individuals, compared with more limited studies in the previous discovery efforts. Additionally, the TOPMed imputation panel, provides better imputation quality for low-frequency variants compared with previous panels, which increases our power to detect rare variation. However, the present study was not designed to provide a detailed evaluation of rare variation within coding genes, and some rare variants within these genes were excluded from the analyses if they were presented in less than 2 cohorts. Larger studies combined with whole genome sequencing data will help identify novel rare (familial) associations for these phenotypes and may provide better instruments that will improve the power for MR studies.

Inclusion of AA ancestry individuals has allowed the identification of novel associated loci for antithrombin in this population. There is a recent debate on transferability of results from GWAS studies to non-European populations, given the overwhelming majority of GWAS results in EA populations for most phenotypes. Although our sample was predominantly of EA, we were able to observe differences in LD blocks between EA and AA ancestry groups, which allowed us to detect novel associations in variants with lower frequency in the EA population, and to refine loci where the linkage blocks differed between ancestries. However, some of the follow-up methods (TWAS, approximate conditional analyses) depend on population reference panels and were limited to the EA population.

Finally, to reduce the risk of false positives, we used a stringent significance threshold ($5 \times 10^{-9}$), sought replication of the main findings in an external proteomics cohort, and provided additional post-GWAS evidence for our novel findings. We included functional validation using *in vitro* silencing to provide evidence for causality of candidate genes and help understand the biological mechanism. We believe this strengthens the credibility of our results. However, liver
cell-derived expression system is only able to assess effects of candidate genes on synthetic mechanisms (e.g., transcription, translation), and is not able to assess potential effects on protein stability and/or clearance. Thus, genes that did not demonstrate an effect, as well as genes that were not selected for testing in this system, could regulate circulating anticoagulant protein expression via synthesis-independent mechanisms.

Summary:
Using cross-ancestry GWAS and TWAS methods, we report 7 novel associations for antithrombin, PC, and PS plasma levels: 4 novel loci regulating antithrombin plasma levels, 2 novel loci regulating PS plasma levels, and 1 novel locus regulating PC plasma levels. Post-GWAS analyses and functional work suggest both SNX17 and GCKR are regulators of antithrombin on the chromosome 2 locus and validate an AA-specific HP gene locus. MR analyses provided evidence implicating low antithrombin levels in VTE risk and low PC levels in VTE and CAD risk. Overall, our findings identified novel pathways regulating the main anticoagulant proteins in hemostasis and strengthen their implication on disease outcomes.
ACKNOWLEDGMENTS

We would also like to thank MEGASTROKE, DeCODE genetics, INVENT Consortium, CARDIoGRAMplusC4D and Million Veterans Program for making their data publicly available. The MEGASTROKE project received funding from sources specified at http://www.megastroke.org/acknowledgments.html. Appendix 1 contains a list with all authors that contributed to MEGASTROKE project (https://www.megastroke.org/authors.html). Data on coronary artery disease has been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG. Data from DeCODE genetics was accessed through DeCode genetics web page at decode.com. VTE and PAD data from de Million Veteran Program were accessed through dbGAPs (phs001672 v7.p1).

SOURCES OF FUNDING

This study is supported in part by the National Heart, Lung, and Blood Institute grant HL134894 and HL139553; infrastructure for the CHARGE Consortium is supported in part by HL105756. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the US Department of Health and Human Services. G. Temprano-Sagrera is supported by the Pla Estratègic de Recerca i Innovació en Salut (PERIS) grant from the Catalan Department of Health for junior research personnel (SLT017/20/000100). P de Vries is supported by American Heart Association grant 17POST33350042. M. Sabater-Lleal is supported by a Miguel Servet contract from the ISCIII Spanish Health Institute (CP17/00142) and co-financed by the European Social Fund, and acknowledges funding from the CERCA Programme/Generalitat de Catalunya. Sources of funding for the specific cohorts can be found in the online-only Data Supplement.
DISCLOSURES

None
REFERENCES:


93. Wessler S, Gaston LW. Anticoagulant therapy in coronary artery disease. Circulation. 1966;34:856-864. doi: 10.1161/01.cir.34.5.856


### Table 1. GWAS and post-GWAS evidence of candidate genes

<table>
<thead>
<tr>
<th>Chr:Pos:A1:A2</th>
<th>rsID</th>
<th>EAF</th>
<th>Beta (Std Err)</th>
<th>N</th>
<th>P-value</th>
<th>Consequence</th>
<th>Status</th>
<th>Post-GWAS Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antithrombin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:173914872:A:T</td>
<td>rs2227624</td>
<td>0.9940</td>
<td>24414 EA 2688 AA</td>
<td>5.31 x 10^{-10}</td>
<td>3.33 x 10^{-10} EA</td>
<td>MV to SERPNC1</td>
<td>Known association</td>
<td>SERPNC1; CATSPER2; PSMA5; PPM1G; SNX17</td>
</tr>
<tr>
<td>2:27375230:T:C</td>
<td>rs4665972</td>
<td>0.9432</td>
<td>25242 EA 2688 AA</td>
<td>6.74 x 10^{-10}</td>
<td>7.87 x 10^{-10} EA</td>
<td>IV to SNX17</td>
<td>Novel GWAS Association with replication (3.18 x 10^{-9})</td>
<td>GOX; NRP1; GTF3C2-AS2; SNX17; GOX</td>
</tr>
<tr>
<td>7:73497513:T:C</td>
<td>rs1244268</td>
<td>0.9111</td>
<td>25095 EA 2688 AA</td>
<td>3.31 x 10^{-10}</td>
<td>5.88 x 10^{-10} EA</td>
<td>IV to BAZ1B</td>
<td>Novel GWAS Association with replication (6.001 x 10^{-11})</td>
<td>MLXPL; BCL7B</td>
</tr>
<tr>
<td>16:72054562:A:C</td>
<td>rs5471</td>
<td>0.9744</td>
<td>12774 EA 2688 AA</td>
<td>0.0311 EA 1.37 x 10^{-10} AA</td>
<td>MV to SERPINC1</td>
<td>Known association</td>
<td>SERPINC1; TNN</td>
<td></td>
</tr>
<tr>
<td>19:49513222:T:G</td>
<td>rs111981233</td>
<td>0.9197</td>
<td>25095 EA 2688 AA</td>
<td>3.28 x 10^{-10}</td>
<td>0.7230 AA</td>
<td>IV to FCGRT</td>
<td>Novel TWAS association</td>
<td>FCGRT; FCGRT; RCP3; RPS11</td>
</tr>
</tbody>
</table>

**Protein C**

<table>
<thead>
<tr>
<th>Chr:Pos:A1:A2</th>
<th>rsID</th>
<th>EAF</th>
<th>Beta (Std Err)</th>
<th>N</th>
<th>P-value</th>
<th>Consequence</th>
<th>Status</th>
<th>Post-GWAS Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:100274968:T:G</td>
<td>rs12740374</td>
<td>0.2556</td>
<td>15565 EA 2688 AA</td>
<td>1.52 x 10^{-8}</td>
<td>1.09 x 10^{-8} EA</td>
<td>3' UTR to CELSR2</td>
<td>Known association</td>
<td>PSRC1; PSMAS; PSRC1</td>
</tr>
<tr>
<td>2:127418299:A:G</td>
<td>rs1799809</td>
<td>0.5429</td>
<td>16597 EA 2688 AA</td>
<td>5.71 x 10^{-8}</td>
<td>5.10 x 10^{-8} EA</td>
<td>0.1 KB 5' to PROC</td>
<td>Known association</td>
<td>PROC; PROC</td>
</tr>
<tr>
<td>2:27375230:T:C</td>
<td>rs4665972</td>
<td>0.4555</td>
<td>16597 EA 2688 AA</td>
<td>7.10 x 10^{-10}</td>
<td>1.78 x 10^{-10} EA</td>
<td>IV to SNX17</td>
<td>Known association</td>
<td>GOX; PRMT6; NRP1; KRTAP43; 2LC4; ATP2G16</td>
</tr>
<tr>
<td>7:73425076:C:G</td>
<td>rs35493968</td>
<td>0.8154</td>
<td>16597 EA 2688 AA</td>
<td>5.87 x 10^{-8}</td>
<td>6.64 x 10^{-8} EA</td>
<td>2KB 5' to MLXPL</td>
<td>Known association</td>
<td>MLXPL; MLXPL</td>
</tr>
<tr>
<td>15:42960963:T:C</td>
<td>rs529330569</td>
<td>0.0059</td>
<td>15341 EA 2688 AA</td>
<td>1.2 x 10^{-10}</td>
<td>1.2 x 10^{-10} EA</td>
<td>IV to UBR1</td>
<td>Novel TWAS association</td>
<td>GDCM3; LCM3; CATSPER2; LCM3;</td>
</tr>
<tr>
<td>20:35179667:T:C</td>
<td>rs11907011</td>
<td>0.0512</td>
<td>16597 EA 2688 AA</td>
<td>4.72 x 10^{-10}</td>
<td>2.43 x 10^{-10} AA</td>
<td>IV to PROC</td>
<td>Known association</td>
<td>PROC; PROC</td>
</tr>
</tbody>
</table>

**Protein S**

<table>
<thead>
<tr>
<th>Chr:Pos:A1:A2</th>
<th>rsID</th>
<th>EAF</th>
<th>Beta (Std Err)</th>
<th>N</th>
<th>P-value</th>
<th>Consequence</th>
<th>Status</th>
<th>Post-GWAS Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:50086695:T:C</td>
<td>rs24128538</td>
<td>0.0108</td>
<td>4006</td>
<td>2.04 x 10^{-10}</td>
<td></td>
<td>1KB 5' to PROST</td>
<td>Known association</td>
<td>PROST</td>
</tr>
<tr>
<td>7:4442456171:T:C</td>
<td>rs141292986</td>
<td>0.0051</td>
<td>3718</td>
<td>9.24 x 10^{-10}</td>
<td></td>
<td>IV to VDDX5</td>
<td>Novel TWAS association</td>
<td>NRP1; MLXPL;</td>
</tr>
<tr>
<td>9:114321523:A:C</td>
<td>rs150611042</td>
<td>0.0680</td>
<td>4006</td>
<td>9.24 x 10^{-13}</td>
<td></td>
<td>2KB 5' to ORM1</td>
<td>Novel GWAS Association with replication (7.53 x 10^{-10})</td>
<td>ORM1; ORM2;</td>
</tr>
<tr>
<td>7:4123101A:T</td>
<td>rs35595692</td>
<td>0.0015</td>
<td>1975</td>
<td>5.41 x 10^{-10}</td>
<td></td>
<td>3KB 85' to MLX7</td>
<td>Novel TWAS association</td>
<td>MLX7;</td>
</tr>
<tr>
<td>9:114321523:A:C</td>
<td>rs150611042</td>
<td>0.0680</td>
<td>6257</td>
<td>7.65 x 10^{-10}</td>
<td></td>
<td>2KB 5' to ORM1</td>
<td>Novel GWAS Association with replication (7.53 x 10^{-10})</td>
<td>ORM1;</td>
</tr>
</tbody>
</table>

### Multi-phenotype

<table>
<thead>
<tr>
<th>Chr:Pos:A1:A2</th>
<th>rsID</th>
<th>EAF</th>
<th>Beta (Std Err)</th>
<th>N</th>
<th>P-value</th>
<th>Consequence</th>
<th>Status</th>
<th>Post-GWAS Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>15:435286519:T:C</td>
<td>rs557070100</td>
<td>0.0039</td>
<td>51955 EA 5376 AA</td>
<td>1.64 x 10^{-10}</td>
<td>MV to MAP1A</td>
<td>Novel Multi-phenotype association</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

A1: Effect Allele; A2: Other Allele; AA: African ancestry; EA: European ancestry; IV: Intronic Variant; MV: missense variant; 3'UTR: 3 Prime Untranslated Region; 5'UTR: 5 Prime Untranslated Region; PST: protein S total; PSF: protein S free.
Figure 1. Schematic view of the analysis’s workflow
Figure 2. Manhattan plots for discovery meta-analyses of GWAS (up) and TWAS (down) results (A) Antithrombin (B) Protein C (C) Protein S Free (D) Protein S Total. Dots represent all allelic variants (GWAS) or genes (TWAS) sorted by chromosome and position throughout the X-axis. Y-axis report inverse log transformed p-value for the associations.
Figure 3. Forest plot showing inverse variance weighted mendelian randomization results for multiple outcomes using antithrombin (A) and protein C (B) as exposure. Squares indicate OR (95% CI).
Figure 4. Knockdown of GCKR, SNX17, and HP alter antithrombin production in HepG2 cells. A) Antithrombin secreted into the culture supernatant was detected by immunoblot (upper) and quantified by densitometry (lower). B) SERPINC1 expression was measured by RT-qPCR. Bars and error bars indicate mean and standard error of the mean; Numbers indicate biological replicates; *P-value < 0.05; ***P-value < 0.0005; ****P-value < 0.0001.