Genetic determinants of venous thrombosis

Haan, H.G. de

CHAPTER 2

Targeted sequencing to identify novel genetic risk factors for deep vein thrombosis: a study of 734 genes

de Haan HG, van Hylckama Vlieg A, Lotta LA, Gorski MM, Bucciarelli P, Martinelli I, Baglin TP, Peyvandi F, Rosendaal FR, INVENT consortium.

J Thromb Haemost. 2018;16(12):2432-2441.

ABSTRACT

Background

Although several genetic risk factors for deep vein thrombosis (DVT) are known, almost all related to hemostasis, a large genetic component remains unexplained.

Objectives

We aimed to identify novel genetic determinants using targeted DNA sequencing.

Patients/Methods

We included 899 DVT patients and 599 controls from three case-control studies (DVT-Milan, MEGA, and THE-VTE) for sequencing of the coding regions of 734 genes involved in hemostasis or related pathways. We performed single-variant association tests for common variants (minor allele frequency [MAF]≥1%) and gene-based tests for rare variants (MAF≤1%), accounting for multiple testing by the false discovery rate (FDR).

Results

Sixty-two out of 3,617 common variants were associated with DVT risk (FDR<0.10). Most of these mapped to *F5, ABO, FGA-FGG,* and *CYP4V2-KLKB1-F11*. Lead variant at *F5* was rs6672595 (odds ratio [OR] 1.58, 95% confidence interval [CI] 1.29-1.92), in moderate linkage with known variant rs4524. Reciprocal conditional analyses suggested that intronic variation might drive this association. We also observed a secondary association at the *F11* region: missense *KLKB1* variant rs3733402 remained associated conditional on known variants rs2039614 and rs2289252 (OR 1.36, 95% CI 1.10-1.69). Two novel variant associations were observed, in *CBS* and *MASP1*, but these did not replicate in the meta-analysis data from the INVENT consortium. There was no support for a burden of rare variants contributing to DVT risk (FDR>0.2).

Conclusions

We confirmed associations between DVT and common variants in *F5, ABO, FGA-FGG,* and *CYP4V2-KLKB1-F11* and observed secondary signals in *F5* and *CYP4V2-KLKB1-F11* that warrant replication and fine-mapping in larger studies.

INTRODUCTION

The hemostatic system ensures the delicate balance between clotting and bleeding. Disturbance of this balance towards clotting may lead to venous thrombosis (VT), mainly manifested as pulmonary embolism (PE) or deep vein thrombosis (DVT).[1,2] Abnormal levels of both fibrinolytic and coagulation factors have been associated with VT risk.[3-6] The role of platelets as risk factor is less well studied, with conflicting results being reported for associations between VT and several platelet markers.[7,8] In addition, genetic variants predominantly in genes encoding proteins of the hemostatic system have been linked to VT risk.[9] Deficiencies of the natural anticoagulants, antithrombin, protein C and protein S, were among the first identified genetic causes of VT, and by now hundreds of (mainly rare) mutations have been reported.[10] Two recent meta-analyses of genome-wide association studies (GWAS), each including over 6000 patients and a multifold of controls, confirmed the association of six loci and identified three novel loci.[11,12] The established loci all map to genes related to hemostasis, specifically: *F5, FGG, F11, ABO, F2,* and *PROCR*.[9-12] Two of the novel loci (*TSPAN15* and *SLC44A2*), and potentially a third locus at *HIVEP1* identified in an earlier GWAS[13] but not confirmed in the latest meta-analyses,[11,12] are the only replicated loci not directly connected to the hemostatic system. This suggests that genes regulating (components of) the hemostatic system are the main genetic contributors to VT risk.

While VT has a strong genetic basis, with heritability estimates of 50-60%,[14-16] the established genetic risk factors only explain a small proportion of the phenotypic variance.[17] In addition, the genetic component remains unknown in 30% of families with multiple family members affected by VT.[18] GWAS efforts have had limited success in identifying novel genetic risk factors, which were mainly common variants in hemostatic-related genes conferring small effects on VT risk. Therefore, a focus on rare and low-frequency variants in coding regions of the genome, may help to discover novel determinants of VT. As such, we have previously shown that a burden of rare coding *ADAMTS13* variants is associated with a 4.8-fold increased DVT risk.[19]

To extend the GWAS efforts, we performed targeted DNA sequencing of the coding regions of 734 genes that were or could be related to the hemostatic system in 899 DVT patients and 599 controls. We subsequently sought replication for associated variants using meta-analysis data from the International Network against Thrombosis (INVENT) collaboration.[11]

PATIENTS AND METHODS

Study population

We set up the Milan Leiden Sequencing study (MILES), in which we included patients with a first VT and controls without a history of VT from three population-based case-control studies: DVT-Milan, Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis (MEGA), and the Thrombophilia, Hypercoagulability and Environmental Risks in Venous Thromboembolism (THE-VTE) study. All studies have been previously described in detail.[19-21] Briefly, DVT-Milan recruited 2,139 consecutive patients with a first DVT at the Angelo Bianchi Bonomi Hemophilia and Thrombosis Center in Milan (Italy) between 1995 and 2010.[19] As controls served non-consanguineous relatives, partners or friends who accompanied patients to center visits. In MEGA, 4,956 consecutive patients with a first DVT or PE were recruited at six anticoagulation clinics in the Netherlands between 1999 and 2004.[20] Partners of patients were invited to participate as a control subject. Additional controls were recruited from the general population using random digit dialling. Patients and controls were invited to provide a blood sample until 2002, after which we switched, for logistical reasons, to buccal swabs. THE-VTE is a two-center case-control study, with a similar design as MEGA, in which 796 consecutive patients with a first DVT or PE and 531 controls were enrolled in Leiden (the Netherlands) and Cambridge (United Kingdom) between 2003 and 2008.[21] Again, partners of eligible patients were invited to participate as control subject.

From each study we included patients and controls based on the following criteria: high-quality DNA sample available from blood, European ancestry as defined by self-reported country of birth of the parents, no major surgery or cancer diagnosis related to the index date, and no deficiency of the natural anticoagulant proteins defined as having normal levels of protein C, protein S, and antithrombin. To eliminate two major genetic causes of VT, we included patients and controls who did not carry factor V (FV) Leiden (rs6025) or prothrombin (PT) G20210A (rs1799963). In addition, we oversampled patients who had a recurrence during the follow-up studies of MEGA and THE-VTE (N=241), as these are more likely to carry genetic risk factors for VT. To ensure a sufficient sample size, we allowed recurrent VT patients to carry FV Leiden or PT G20210A (N=94). In total, 899 DVT patients and 599 controls were selected for sequencing. An overview of the participants per study is presented in Supplemental Table 1.

All participants provided written informed consent. DVT-Milan was approved by the Institutional Review Board of the Fondazione IRCCS Ca' Granda–Ospedale Maggiore Policlinico, whereas MEGA and THE-VTE were approved by the Medical Ethics Committee of the Leiden University Medical Center. THE-VTE was also approved by the NHS Research Ethics Committee in Cambridge.

Targeted DNA sequencing

We selected pathways involved in thrombosis and hemostasis, including the coagulation system, fibrinolysis, platelet function, inflammation, and the complement system. Using literature and gene ontology databases, we extracted genes belonging to these pathways. From the ThromboGenomics database,[22] we included additional genes that have been linked to inherited clotting, platelets or bleeding disorders. In total, we included 734 genes, of which we sequenced the coding regions plus 10 base pairs flanking the exons to cover the splice junctions. For a subset of 48 genes, we additionally sequenced the 3' and 5' untranslated regions (UTR). In addition, we performed whole gene sequencing including 10 kilo base pairs promoter area of three genes, that is *F5, VWF*, and *F8*, which are of particular interest for VT. *F5* harbours the strongest genetic risk factor for VT, that is FV Leiden, in the general population. Von Willebrand factor and factor VIII, encoded by *VWF* and *F8*, are tightly interconnected proteins of which levels are strongly associated with first and recurrent VT risk.[5,23] We also targeted 179 single nucleotide variants, consisting of 28 variants previously associated with VT and 151 ancestry-informative markers. To facilitate the capture, we allowed some 200 base pairs of target region surrounding each variant. A list of the targeted genes and variants can be found in Supplemental Table 2.

The target area was designed with the Reference Sequence (RefSeq) Database using tools in the UCSC Genome Browser[24] and sent to NimbleGen (Roche NimbleGen, Madison, WI, USA) for probe design. Next-generation DNA sequencing was subsequently performed at the Human Genome Sequencing Center (HGSC), Baylor College of Medicine (Houston, USA). A complete sequencing protocol can be accessed on the HGSC website (https://www.hgsc.bcm.edu/content/protocols-sequencing-library-construction). Briefly, DNA samples were constructed into Illumina paired-end pre-capture libraries according to the manufacturer's protocol (Illumina Multiplexing_SamplePrep_Guide_1005361_D) with some minor modifications. We multiplexed 24 samples per capture and included

two capture pools per HiSeq lane. Enriched samples were sequenced using the HiSeq 2000 platform (Illumina, San Diego, CA, USA).

Sequence analysis was performed using the Mercury analysis pipeline.[25] In short, sequence reads and base-call confidence values were generated for de-multiplexed pools using the vendor's primary analysis software (CASAVA). Next, reads and qualities were mapped to reference genome hg19 using the Burrows-Wheeler aligner,[26] resulting in BAM files per sample.[27] Realignment around insertions and deletions (indels), and recalibration of quality scores was performed with the Genome Analysis Toolkit.[28] Variant calling was conducted using the Atlas2suite,[29] followed by variant annotation as implemented in the Cassandra annotation suite. Individual variant files were subsequently merged into a project-level file to generate a genotype matrix of all identified variants.

Initial exclusion criteria for variant calls were as follows: variant posterior probability <0.95, number of variant reads <3, variant read ratio <0.1, variant reads in a single strand direction, total coverage <6 or >1024 reads. Called variants that passed quality control in at least one individual were included in the project-level variant file. In total, 31,540 variants were identified in 1495 individuals with sequencing data available (897 DVT patients and 598 controls). We subsequently performed additional filtering using VCFtools[30] to identify high-quality variants, requiring a sequencing depth ≥10 reads, call rate ≥80%, Phred score ≥30, and Hardy-Weinberg equilibrium $P>1.0x10^{-4}$ in the controls separately per study. A total of 20,054 variants passed quality control.

Statistical analysis

We conducted single-variant association analyses for 3,617 low-frequency and common variants, defined as a minor allele frequency (MAF) ≥1%, using logistic regression as implemented in PLINK.[31] We calculated effect estimates as odds ratios (OR) with corresponding 95% confidence intervals (95% CI) per risk allele copy and adjusted for sex, age, (study) origin, carriership of FV Leiden per allele copy, and carriership of PT G20210A. We assumed that X-chromosomal loci undergo complete inactivation. Linkage disequilibrium (LD) between variants was assessed in Europeans from the 1000 Genomes Project.[32] To identify secondary associations, we performed conditional analyses by adjusting for the lead variant at a locus (defined as region within 1 Mb of the lead variant). The Bonferroni threshold for significance was set at $1.38 x10^{-5}$ (0.05

divided by 3,617 variants) to account for multiple testing. We additionally calculated false discovery rates (FDR) and variants with a FDR <0.10 were carried forward for replication.

Rare variants (MAF ≤1%) were collapsed per gene and analysed with the T1 burden test and the Sequence Kernel Association Test (SKAT),[33] the latter allowing differential effect directions. In total, we analysed 16,188 variants in 647 genes with a cumulative minor allele count (cMAC) ≥5. Analyses were adjusted for sex, age, (study) origin, carriership of FV Leiden, and PT G20210A. In the burden test, we used adaptive permutations to calculate empirical P-values, which were stratified by Northwest versus South European origin. We calculated FDRs to take multiple testing into account. To identify which rare variant contributed to an association signal, we excluded one variant at a time and repeated the analyses. The gene-based association tests were performed with the PLINK/SEQ suite.

Replication

Novel associations between common and low-frequency variants and DVT (FDR <0.10) were examined in meta-analysis data from INVENT. Details on the meta-analysis and the included studies are provided elsewhere.[11] In short, GWAS data from 12 studies, totalling 7,507 VT patients and 52,632 controls, were meta-analysed using an inverse-variance weighting fixed-effects model. Of note, there was a small amount of overlap in VT patients (N=384) between the discovery and the replication analyses, as some patients were also included in the meta-analysis of INVENT.

RESULTS

Targeted DNA sequencing was successfully performed in 897 DVT patients and 598 controls. The study population characteristics are presented in Table 1. In total, 20,054 high-quality variants were identified, of which 11,268 were singletons (median of 7 singletons per person, interquartile range 4-10). An overview of the functional classes and the MAF distribution is shown in Supplemental Figure 1. The majority of the variants was rare and mapped to protein-coding sequence (N=10,131), including several stop-loss and -gain variants. We also observed 168 indels and 530 splice variants. In addition, we identified a total of 5,210 variants which had not been reported in any database.

Table 1. Study population characteristics

	DVT patients	Controls
N	897	598
Age in years, mean (SD)	48.1 (13.7)	47.1 (13.3)
male sex, N (%)	449 (50.1)	277 (46.3)
North-west European origin, N (%)	599 (67.8)	300 (50.2)
DVT only, N (%)	755 (84.2)	NA
*Carriers PT, N (%)	15 (1.67)	NA
*Carriers FVL, N (%)	75 (8.36)	NA

DVT deep vein thrombosis; SD standard deviation; FVL factor V Leiden; PT prothrombin G20210A; NA not applicable

* These were part of a subgroup of 241 DVT patients who had a recurrence during follow-up in MEGA and THE-VTE (prevalence of FVL and PT in that subgroup of 31.1% and 6.2%, respectively).

Single variant association analyses

We tested 3,617 low-frequency and common variants for an association with DVT risk. The quantile-quantile plot of the observed P-values versus the expected distribution is shown in Supplemental Figure 2. Statistically significant associations at the Bonferroni threshold were observed for 12 variants in four loci: *ABO, FGA-FGG, CYP4V2-KLKB1-F11*, and *F5* (Table 2). All four loci harbour established genetic risk factors for VT. Interestingly, only three of the 12 variants mapped to coding sequence. Exclusion of recurrent VT patients in a sensitivity analysis resulted in similar associations with DVT risk (Supplemental Table 3). Lead variant in *ABO* was the well-known risk variant rs8176719 (frameshift variant, risk allele frequency (RAF) 45%), encoding non-O blood groups. C-carriers had a 1.9-fold (95% CI 1.61-2.24) increased DVT risk per allele copy. The intronic *ABO* variant rs4962040 also reached statistical significance (RAF 59%, OR 1.53, 95% CI 1.28-1.83), though this association was diminished upon conditioning on rs8176719 ($OR_{adjusted}$ 1.12, 95% CI 0.88-1.41). Likewise, none of the other 22 *ABO* variants were associated with DVT risk conditional on rs8176719 (Supplemental Table 4). In *CYP4V2-KLKB1-F11*, lead variant was intronic *F11* variant rs2036914 (RAF 60%, OR 1.65, 95% CI 1.38-1.97), which has been linked to increased FXI levels and VT.[34,35] Three additional variants were associated with DVT risk at the Bonferroni threshold, of which one remained associated upon conditioning on rs2036914 (rs3733402 in *KLKB1*, $OR_{adjusted}$ 1.33, 95% CI 1.08-1.64). Conditioning on a second known *F11* risk variant (rs2289252), did not materially change this association ($OR_{adjusted}$ 1.36, 95% CI 1.10-1.69).

The *KLKB1* missense variant (p.Ser143Asn) leads to reduced binding of prekallikrein to its cofactor high-molecular weight kininogen,[36] affecting the initiation of the intrinsic coagulation cascade. In the *FGA-FGG* locus, the association with DVT was driven by missense *FGA* variant rs6050 (RAF 39%, OR 1.66, 95% CI 1.37-2.02) and downstream *FGG* variant rs2066865 (RAF 35%, OR 1.60, 95% CI 1.33-1.92), which have both been linked to increased γ' fibrinogen levels and VT risk.[37,38] rs6050 and rs2066865 were in high LD (r^2 0.90) and reciprocal conditional analysis showed that they represented the same association signal (Supplemental Table 5). We did not identify additional associations after conditioning on the lead variants (Supplemental Table 4). Four intronic *F5* variants were associated with DVT risk at the Bonferroni threshold, which were in almost complete LD (lowest r^2 between any pair was 0.90) and represented the same association signal. Carriers of the lead variant (rs6672595, RAF 76%) had a 1.6-fold increased DVT risk (95% CI 1.29-1.92) per risk allele. The variants were also in high LD (r^2 0.77) with *F5* missense variant rs4524, for which an association with VT independent of FV Leiden has been reported.[39] In our study, carriers of rs4524 (RAF 73%) had a 1.3-fold higher DVT risk (95% CI 1.11-1.60) per allele copy, which attenuated with adjustment for lead variant rs6672595 ($OR_{adjusted}$ 1.10, 95% CI 0.74-1.63). On the other hand, the association between rs6672595 (and its proxies) and DVT risk remained, albeit with wider confidence intervals, with adjustment for rs4524 (Supplemental Table 6). No secondary association signals were observed in the *F5* region (Supplemental Figure 3).

Table 2. Associations between common variants and first deep vein thrombosis ($P < 1.38 \times 10^{-5}$)

							Discovery analysis		*Conditional analysis	
rsID	**Chr.**	**Position**	**Class**	**Gene**	**A_1/A_2**	**RAF**	**OR (95%CI)**	**P**	***OR (95%CI)**	**P**
rs3766110	1	169515183	intronic	*F5*	C/A	0.774	1.54 (1.27-1.86)	1.07×10^{-5}	NA	NA
rs3766111	1	169515204	intronic	*F5*	C/T	0.773	1.57 (1.29-1.91)	6.82×10^{-6}	NA	NA
rs3766113	1	169515307	intronic	*F5*	G/A	0.770	1.55 (1.28-1.88)	8.59×10^{-6}	NA	NA
rs6672595	1	169515536	intronic	*F5*	T/C	0.757	1.58 (1.29-1.92)	6.11×10^{-6}	NA	NA
rs6050	4	155507590	missense	*FGA*	T/C	0.393	1.66 (1.37-2.02)	2.33×10^{-7}	NA	NA
rs2066865	4	155525276	downstream	*FGG*	G/A	0.352	1.60 (1.33-1.92)	4.86×10^{-7}	1.36 (0.81-2.31)	0.245
rs3733402	4	187158034	missense	*KLKB1*	G/A	0.573	1.55 (1.30-1.86)	1.27×10^{-6}	1.33 (1.08-1.64)	0.006
rs4253399	4	187188094	intronic	*F11*	T/G	0.458	1.50 (1.27-1.76)	8.34×10^{-7}	1.16 (0.90-1.49)	0.246
rs3822057	4	187188152	intronic	*F11*	A/C	0.545	1.44 (1.23-1.70)	6.74×10^{-6}	0.91 (0.62-1.35)	0.642
rs2036914	4	187192481	intronic	*F11*	T/C	0.602	1.65 (1.38-1.97)	2.47×10^{-8}	NA	NA
rs8176719	9	136132908	frameshift	*ABO*	T/TC	0.451	1.90 (1.61-2.24)	1.39×10^{-14}	NA	NA
rs4962040	9	136133531	intronic	*ABO*	G/A	0.594	1.53 (1.28-1.83)	3.67×10^{-6}	1.12 (0.88-1.41)	0.355

Chr. chromosome; A_1 reference allele; A_2 risk allele; RAF risk allele frequency; OR odds ratio; CI confidence interval; P P value; NA not applicable

Single variants association analyses for 3,617 low-frequency and common variants (MAF > 1%) were conducted using logistic regression assuming an additive mode of inheritance. Analyses were adjusted for sex, age, (study) origin, carriership of FV Leiden per copy of the risk allele, and carriership of PT G20210A.

*We conducted conditional logistic regression analyses in which we adjusted for the lead variant per locus (highlighted in bold, i.e. *F5* rs6672595, *FGA* rs6050, *F11* rs2036914, and *ABO* rs8176719).

In addition, we observed 50 variants that did not exceed the Bonferroni threshold for statistical significance, but did have low FDR (<0.10). Almost all of these mapped to the four main loci and did not represent new association signals (Supplemental Table 7). We additionally identified two novel, suggestive variant associations with DVT risk (Table 3). In *MASP1*, we observed an association with DVT for 3' UTR variant rs72549167 (RAF 1.6%, FDR 9%). Carriers of the risk allele had a 3.5-fold increased DVT risk (95% CI 1.62-7.67) per allele copy. The *MASP1* gene encodes mannan-binding lecture serine peptidase 1, which is involved in the lectin pathway of complement activation and has crosslinks with the clotting cascade.[40,41] In particular, activated by thrombin and activated platelets,[42] MASP1 can cleave several coagulation factors, including prothrombin, thrombin-activatable fibrinolysis inhibitor, and factor XIII.[41] Of the other 16 *MASP1* variants, one was also associated with DVT risk (Supplemental Table 8), which was in complete LD with rs72549167. The other novel variant association mapped to a synonymous variant in *CBS*, encoding cystathionine beta-synthase, associated with DVT risk with an allelic OR of 1.31 (95% CI 1.11-1.55, FDR 9%). Cystathionine beta-synthase catalyses the conversion of homocysteine to cystathionine and specific genetic defects in *CBS* lead to homocystinuria, a disorder which has been linked to increased VT risk.[43] We observed two additional common variants in *CBS*, all not associated with rs1801181, and none of these were associated with DVT risk (Supplemental Table 9). We next aimed to replicate the two novel variant associations using the meta-analysis data from INVENT, which included 7,507 VT patients and 52,632 controls (Table 3). There was no clear evidence for an association of DVT with rs72549167 in *MASP1* (OR 1.21, 95% CI 0.96-1.52), nor with rs1801181 in *CBS* (OR 1.00, 95% CI 0.96-1.05).

Table 3. Novel variant associations with deep vein thrombosis (FDR < 0.10) and replication effort

						Discovery analysis				*Replication			
rs ID	**Chr.**	**Position**	**Class**	**Gene**	**A_1/A_2**	**RAF**	**OR (95% CI)**	**P**	**FDR**	**RAF**	**OR (95% CI)**	**P**	**Info (SD)**
rs1801181	21	44480616	synonymous	*CBS*	G/A	0.370	1.31 (1.11-1.55)	0.002	0.09	0.364	1.00 (0.95-1.05)	0.926	0.96 (0.02)
rs72549167	3	186952375	3' UTR	*MASP1*	C/G	0.016	3.52 (1.62-7.67)	0.002	0.09	0.010	1.21 (0.96-1.52)	0.102	0.86 (0.12)

Chr. chromosome; A_1 reference allele; A_2 risk allele; RAF risk allele frequency; OR odds ratio; CI confidence interval; P P value; info mean imputation quality score; SD standard deviation; UTR untranslated region

Discovery analysis was performed using logistic regression assuming an additive mode of inheritance. Analyses were adjusted for sex, age, (study) origin, carriership of FV Leiden per copy of the risk allele, and carriership of PT G20210A.

*Replication was performed in data from the INVENT consortium. GWAS results from 12 studies were meta-analysed using a fixed-effect meta-analysis model based on inverse-variance weighting. Heterogeneity was assessed by the Cochran's Q statistic and the I^2 index. For rs1801181 we observed a Q 8.69, I^2 0.00, P-value 0.65. For rs72549167, we observed a Q of 9.06, I^2 0.00, P-value 0.62.

Gene-based association analyses

The impact of 16,188 rare variants mapping to 647 genes (cMAC ≥5) on DVT risk was assessed with aggregation tests. The results from the SKAT-based joint analyses of all rare variants per gene did not provide support for an association between rare variants and DVT risk. The most suggestive association signal was observed for *F2RL2* (P 0.0013, FDR 60%), encoding proteinase-activated receptor-3 (PAR-3). The burden tests identified one gene suggestive of an association with DVT risk. DVT patients had a burden of rare variants in *KLK5* (P 0.0003, FDR 21%), which encodes a serine protease named kallikrein related peptidase 5 and is involved in inflammatory responses through the PAR-2 system.[44] Of the 10 rare variants identified in *KLK5*, including five singletons, 26 variant alleles were observed in DVT patients compared with three alleles in controls. All 10 variants mapped to protein-coding sequence. None of the variants was solely driving the association signal (data not shown).

DISCUSSION

To identify novel genetic risk factors for DVT which have been missed by GWAS, we sequenced the coding regions of 734 genes related to hemostasis in 899 DVT patients and 599 controls. Our targeted sequencing approach confirmed several established risk loci. Specifically, lead variants at *ABO*, *FGA-FGG*, and *CYP4V2-KLKB1-F11* have all previously been implicated in VT risk, both directly or via proxy variants.[11-13,19,34-36] The effect sizes observed in our study were slightly higher than in earlier reports, which may in part be explained by our selection of individuals without a cancer diagnosis or recent surgery. Differences in genetic effects on PE versus DVT could also have played a role, in line with the so-called 'FV Leiden paradox'.[45] Although we did not discover novel risk loci, the secondary risk loci identified at *F5* and *CYP4V2-KLKB1-F11* may provide leads for a better understanding of the biological mechanism underlying these loci.

Interestingly, almost all associated variants mapped to non-coding sequence, while our sequencing design mainly targeted coding variation. In *F5* and *CYP4V2-KLKB1-F11*, there was little evidence that the (lead) associations could be explained by linkage to common, coding variants. This may point to non-coding variation as causal risk factor, potentially influencing DVT risk by affecting gene regulation. Four co-inherited intronic variants in *F5* were associated with DVT risk at the Bonferroni threshold, which have

not been implicated in VT risk. Missense *F5* variant rs4524, an established risk variant independent of FV Leiden[38] and in moderate LD with the associated *F5* variants, did not attain a high level of statistical significance in our study. Furthermore, its effect on DVT risk was strongly diminished when adjusting for our lead *F5* variant (rs6672595). Both variants are part of a large, strongly-linked cluster of variants, which spans across several introns and exons of *F5*. Additional fine-mapping in a large study is necessary to uncover the most likely causal variant. Another notable finding was the suggestive, secondary association signal at *CYP4V2-KLKB1-F11*, missense *KLKB1* variant rs3733402, which remained associated with DVT risk with an allelic odds ratio of 1.4 upon adjusting for rs2036914 and rs2289252. We are not the first to report an association signal at *CYP4V2-KLKB1-F11* secondary to rs2289252 and rs2036914,[11,34] although the previously reported variants are not in LD with rs3733402, suggesting that this locus may indeed harbour multiple causal variants. In addition, we were unable to disentangle the effects of *FGA*-rs6050 and *FGG*-rs2068865 on DVT risk due to their strong, though imperfect, linkage. However, a previously reported haplotype analysis did not show an independent association with VT for the haplotype carrying *FGA*-rs6050.[36]

In addition to the associations at the known loci, we identified two variants, which have not been linked to VT risk, with low FDR but association tests that did not pass the Bonferroni threshold. These were a synonymous variant in *CBS* and a 3' UTR variant in *MASP1*. Both variants did, however, not replicate in the meta-analysis data from INVENT. Imputation quality was sufficient and there was no evidence of statistical heterogeneity. We cannot rule out that differences in the discovery and the replication study populations, for example due to the inclusion of DVT patients versus patients with any VT event, could have explained the lack of replication. Alternatively, the associations in the discovery analysis might have been chance findings, taking into account the FDR of 9% for both variants.

The gene-based analyses did not support the hypothesis of a burden of rare, mainly coding variants in hemostasis-related genes contributing to DVT risk. We observed a potential association for a burden of rare variants in *KLK5* with 26 alleles observed in DVT patients compared with 3 alleles in controls, though the FDR was relatively high (21%). The lack of significant gene associations may be explained by our limited sample size. Gene-based analyses for complex diseases generally require large study sizes given the likely modest effect sizes and the expected proportion of causal variants.[46]

Therefore, we might have missed associations between genes with rare variants and DVT risk. We also did not distinguish between rare variants with or without a predicted deleterious consequence, as advocated by some,[46,47] since this would have further increased the multiple testing burden and lowered cMAC counts. As the effects of VT on fitness are limited, we also did not expect strong purifying selection on deleterious variants. In addition, our group has previously reported an association between DVT and a burden of rare coding variants in *ADAMTS13* (17 alleles in DVT patients compared with 4 alleles in controls, N=192 individuals).[19] In the present study, we observed a nominal association for a burden of rare variants in *ADAMTS13* with DVT risk (P 0.048, 84 alleles in DVT patients compared with 42 alleles in controls). Although the majority of studied rare *ADAMTS13* variants mapped to coding sequence (75%), the inclusion of noncoding variants may explain the difference in the results of the burden analyses. However, when only focusing on rare coding variation in *ADAMTS13*, we observed a similar association with DVT risk (P 0.066, 55 alleles in DVT patients compared with 27 alleles in controls). Larger studies are needed to elucidate the role of rare coding and noncoding variants in *ADAMTS13* on DVT risk.

The major limitation of our study is its limited sample size, which did not allow us to detect associations across the entire allele frequency spectrum. Given the multicausal nature of DVT, genetic effect estimates on DVT risk are expected to be modest, requiring an even larger sample size. We attempted to maximize our statistical power by studying genetic variation in biologically plausible genes in a well-characterized study population. Specifically, we selected genetically enriched DVT patients, without some of the major clinical risk factors. In addition, we oversampled VT patients who had developed a recurrence and are therefore more likely to carry genetic risk variants. Except for a small number of patients with recurrent VT, we selected individuals not carrying FV Leiden and PT G20210A, and, therefore, we could not study these variants or those in strong LD. Another limitation is the lack of generalizability of our findings to non-European populations. In addition, by design, our targeted sequencing approach did not allow us to study variation in regulatory regions outside our target area nor variation in genes not previously linked to the hemostatic system. Therefore, we were unable to identify variants in untargeted regions of the candidate genes, novel DVT-associating genes outside the hemostatic system, and to assess variation in the recently identified risk loci *SLC44A2* and *TSPAN15*.[11]

In conclusion, our targeted sequencing approach confirmed the association of several of the established VT risk loci. The secondary loci identified at *F5* and *CYP4V2-KLKB1-F11* suggest that the underlying biological mechanism might be more complex than initially thought. In addition, we did not find evidence of a burden of rare variants in hemostasis-related genes affecting DVT risk.

REFERENCES

1. Reitsma PH, Versteeg HH, Middeldorp S. Mechanistic view of risk factors for venous thromboembolism. Arterioscler Thromb Vasc Biol. 2012;32:563-8.
2. Reitter EM, Kaider A, Ay C, Quehenberger P, Marosi C, Zielinski C, Pabinger I. Longitudinal analysis of hemostasis biomarkers in cancer patients during antitumor treatment. J Thromb Haemost. 2016;14:294-305.
3. van Tilburg NH, Rosendaal FR, Bertina RM. Thrombin activatable fibrinolysis inhibitor and the risk for deep vein thrombosis. Blood. 2000;95:2855-9.
4. Meltzer ME, Lisman T, de Groot PG, Meijers JC, le Cessie S, Doggen CJ, Rosendaal FR. Venous thrombosis risk associated with plasma hypofibrinolysis is explained by elevated plasma levels of TAFI and PAI-1. Blood. 2010;116:113-21.
5. Tsai AW, Cushman M, Rosamond WD, Heckbert SR, Tracy RP, Aleksic N, Folsom AR. Coagulation factors, inflammation markers, and venous thromboembolism: the longitudinal investigation of thromboembolism etiology (LITE). Am J Med. 2002;113:636-42.
6. Cushman M, O'Meara ES, Folsom AR, Heckbert SR. Coagulation factors IX through XIII and the risk of future venous thrombosis: the Longitudinal Investigation of Thromboembolism Etiology. Blood. 2009;114:2878-83.
7. Braekkan SK, Mathiesen EB, Njølstad I, Wilsgaard T, Størmer J, Hansen JB. Mean platelet volume is a risk factor for venous thromboembolism: the Tromsø Study, Tromsø, Norway. J Thromb Haemost. 2010;8:157-62.
8. Puurunen MK, Hwang SJ, O'Donnell CJ, Tofler G, Johnson AD. Platelet function as a risk factor for venous thromboembolism in the Framingham Heart Study. Thromb Res. 2017;151:57-62.
9. Rosendaal FR, Reitsma PH. Genetics of venous thrombosis. J Thromb Haemost. 2009; 7 Suppl 1:301–304.
10. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, Cooper DN. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet;136:665-677.
11. Germain M, Chasman DI, de Haan H, Tang W, Lindström S, Weng LC, de Andrade M, de Visser MC, Wiggins KL, Suchon P, Saut N, Smadja DM, Le Gal G, van Hylckama Vlieg A, Di Narzo A, Hao K, Nelson CP, Rocanin-Arjo A, Folkersen L, Monajemi R, et al. Meta-analysis of 65,734 Individuals Identifies TSPAN15 and SLC44A2 as Two Susceptibility Loci for Venous Thromboembolism. Am J Hum Genet. 2015;96:532-42.

12. Hinds DA, Buil A, Ziemek D, Martinez-Perez A, Malik R, Folkersen L, Germain M, Mälarstig A, Brown A, Soria JM, Dichgans M, Bing N, Franco-Cereceda A, Souto JC, Dermitzakis ET, Hamsten A, Worrall BB, Tung JY; METASTROKE Consortium, INVENT Consortium., Sabater-Lleal M. Genome-wide association analysis of self-reported events in 6135 individuals and 252 827 controls identifies 8 loci associated with thrombosis. Hum Mol Genet. 2016;25:1867-74.
13. Morange PE, Bezemer I, Saut N, Bare L, Burgos G, Brocheton J, Durand H, Biron-Andreani C, Schved JF, Pernod G, Galan P, Drouet L, Zelenika D, Germain M, Nicaud V, Heath S, Ninio E, Delluc A, Münzel T, Zeller T, et al. A follow-up study of a genome-wide association scan identifies a susceptibility locus for venous thrombosis on chromosome 6p24.1. Am J Hum Genet. 2010;86:592-5.
14. Souto J, Almasy L, Borrell M, Blanco-Vaca F, Mateo J, Soria J, Coll I, Felices R, Stone W, Fontcuberta J, Blangero J. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Genetic Analysis of Idiopathic Thrombophilia. Am J Hum Genet. 2000;67:1452–9.
15. Larsen T, Sorensen H, Skytthe A, Johnsen S, Vaupel J, Christensen K. Major genetic susceptibility for venous thromboembolism in men: a study of Danish twins. Epidemiology. 2003;14(3):328–32.
16. Heit J, Phelps M, Ward S, Slusser J, Petterson T, De Andrade M. Familial segregation of venous thromboembolism. J Thromb Haemost. 2004;2:731–6.
17. Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, Cohen W, Oudot-Mellakh T, Antoni G, Alessi MC, Zelenika D, Cambien F, Tiret L, Bertrand M, Dupuy AM, Letenneur L, Lathrop M, Emmerich J, Amouyel P, Trégouët DA, Morange PE. Genetics of venous thrombosis: insights from a new genome wide association study. PLoS One. 2011;6:e25581.
18. Bezemer ID, van der Meer FJ, Eikenboom JC, Rosendaal FR, Doggen CJ. The value of family history as a risk indicator for venous thrombosis. Arch Intern Med. 2009;169(6):610-5.
19. Lotta LA, Tuana G, Yu J, Martinelli I, Wang M, Yu F, Passamonti SM, Pappalardo E, Valsecchi C, Scherer SE, Hale W 4th, Muzny DM, Randi G, Rosendaal FR, Gibbs RA, Peyvandi F. Next-generation sequencing study finds an excess of rare, coding single-nucleotide variants of ADAMTS13 in patients with deep vein thrombosis. J Thromb Haemost. 2013;11:1228-39.
20. Blom JW, Doggen CJ, Osanto S, Rosendaal FR. Malignancies, prothrombotic mutations, and the risk of venous thrombosis. JAMA. 2005;293:715–722.
21. van Hylckama Vlieg A, Baglin CA, Luddington R, MacDonald S, Rosendaal FR, Baglin TP. The risk of a first and a recurrent venous thrombosis associated with an elevated D-dimer level and an elevated thrombin potential: results of the THE-VTE study. J Thromb Haemost. 2015;13:1642-52.

22. Simeoni I, Stephens JC, Hu F, Deevi SV, Megy K, Bariana TK, Lentaigne C, Schulman S, Sivapalaratnam S, Vries MJ, Westbury SK, Greene D, Papadia S, Alessi MC, Attwood AP, Ballmaier M, Baynam G, Bermejo E, Bertoli M, Bray PF, et al. A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. Blood. 2016;127:2791-803.
23. Timp JF, Lijfering WM, Flinterman LE, van Hylckama Vlieg A, le Cessie S, Rosendaal FR, Cannegieter SC. Predictive value of factor VIII levels for recurrent venous thrombosis: results from the MEGA follow-up study. J Thromb Haemost. 2015;13:1823-32.
24. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002;12:996-1006.
25. Reid JG, Carroll A, Veeraraghavan N, Dahdouli M, Sundqvist A, English A, Bainbridge M, White S, Salerno W, Buhay C, Yu F, Muzny D, Daly R, Duyk G, Gibbs RA, Boerwinkle E. Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. BMC Bioinformatics. 2014;15,30.
26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
28. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43,491–8.
29. Challis D, Yu J, Evani US, Jackson AR, Paithankar S, Coarfa C, Milosavljevic A, Gibbs RA, Yu F. An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics. 2012;13:8.
30. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics. 2011;27:2156-8.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-75.
32. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. Nature. 2015;526:68-74.
33. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X: Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89:82-93.

34. Li Y, Bezemer ID, Rowland CM, Tong CH, Arellano AR, Catanese JJ, Devlin JJ, Reitsma PH, Bare LA, Rosendaal FR. Genetic variants associated with deep vein thrombosis: the F11 locus. J Thromb Haemost. 2009;7:1802-8.
35. Bezemer ID, Bare LA, Doggen CJ, Arellano AR, Tong C, Rowland CM, Catanese J, Young BA, Reitsma PH, Devlin JJ, Rosendaal FR. Gene variants associated with deep vein thrombosis. JAMA. 2008;299:1306-14.
36. Katsuda I, Maruyama F, Ezaki K, Sawamura T, Ichihara Y. A new type of plasma prekallikrein deficiency associated with homozygosity for Gly104Arg and Asn124Ser in apple domain 2 of the heavy-chain region. Eur J Haematol. 2007;79:59–68.
37. Uitte de Willige S, de Visser MC, Houwing-Duistermaat JJ, Rosendaal FR, Vos HL, Bertina RM. Genetic variation in the fibrinogen gamma gene increases the risk for deep venous thrombosis by reducing plasma fibrinogen gamma' levels. Blood. 2005;106:4176-83.
38. Lovely RS, Yang Q, Massaro JM, Wang J, D'Agostino RB Sr, O'Donnell CJ, Shannon J, Farrell DH. Assessment of genetic determinants of the association of γ' fibrinogen in relation to cardiovascular disease. Arterioscler Thromb Vasc Biol. 2011;31:2345-52.
39. Smith NL, Hindorff LA, Heckbert SR, Lemaitre RN, Marciante KD, Rice K, Lumley T, Bis JC, Wiggins KL, Rosendaal FR, Psaty BM. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. JAMA. 2007;297:489–98.
40. Dobó J, Schroeder V, Jenny L, Cervenak L, Závodszky P, Gál P. Multiple roles of complement MASP-1 at the interface of innate immune response and coagulation. Mol Immunol. 2014;61:69-78.
41. Hess K, Ajjan R, Phoenix F, Dobó J, Gál P, Schroeder V. Effects of MASP-1 of the complement system on activation of coagulation factors and plasma clot formation. PLoS One. 2012;7:e35690.
42. Kozarcanin H, Lood C, Munthe-Fog L, Sandholm K, Hamad OA, Bengtsson AA, Skjoedt MO, Huber-Lang M, Garred P, Ekdahl KN, Nilsson B. The lectin complement pathway serine proteases (MASPs) represent a possible crossroad between the coagulation and complement systems in thromboinflammation. J Thromb Haemost. 2016;14:531-45.
43. Mudd SH, Skovby F, Levy HL, Pettigrew KD, Wilcken B, Pyeritz RE, Andria G, Boers GH, Bromberg IL, Cerone R, Fowler B, Gröbe H, Schmidt H, Schweitzeret L. The natural history of homocystinuria due to cystathionine β-synthase deficiency. Am J Hum Genet. 1997;37:1-31.
44. Oikonomopoulou K, Hansen KK, Saifeddine M, Vergnolle N, Tea I, Blaber M, Blaber SI, Scarisbrick I, Diamandis EP, Hollenberg MD. Kallikrein-mediated cell signalling: targeting proteinase-activated receptors (PARs). Biol Chem. 2006;387:817-24.
45. Desmarais S, de Moerloose P, Reber G, Minazio P, Perrier A, Bounameaux H. Resistance to activated protein C in an unselected population of patients with pulmonary embolism. Lancet. 1996;347:1374-75.

46. Moutsianas L, Agarwala V, Fuchsberger C, Flannick J, Rivas MA, Gaulton KJ, Albers PK, GoT2D Consortium, McVean G, Boehnke M, Altshuler D, McCarthy MI. The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. PLoS Genet. 2015;11:e1005165.
47. Richardson TG, Campbell C, Timpson NJ, Gaunt TR. Incorporating Non-Coding Annotations into Rare Variant Analysis. PLoS One. 2016;11:e0154181.

SUPPLEMENTAL INFORMATION

Supplemental Table 1. Included participants per study

Supplemental Table 2. Targeted genes and variants

Supplemental Table 3. Sensitivity analysis excluding recurrent VT patients

Supplemental Table 4. Discovery and conditional association analyses of variants in *ABO*, *CYP4V2-KLKB1-F11*, and *FGA-FGG* ($P > 1.38 \times 10^{-5}$)

Supplemental Table 5. Reciprocal conditional association analyses rs6050 and rs2066865

Supplemental Table 6. Association analysis at the *F5* association locus conditional on rs4524

Supplemental Table 7. Suggestive single variant associations (FDR <0.10)

Supplemental Table 8. Single variant association analyses of common *MASP1* variants

Supplemental Table 9. Single variant association analyses of common *CBS* variants

Supplemental Figure 1. Minor allele frequency distribution of identified variants (left) and overview of functional classes (right)

Supplemental Figure 2. Quantile-quantile plot of single variant association analyses

Supplemental Figure 3. Regional association plots for single variant associations in the *F5* region

Supplemental Table 1. Included participants per study

	MEGA		THE-VTE		DVT-Milan	
	DVT patients	Control subjects	DVT patients	Control subjects	DVT patients	Control subjects
N	459	230	140	70	298	298
Men, N (%)	246 (53.6)	130 (56.5)	86 (61.4)	27 (38.6)	117 (39.3)	120 (40.3)
Age in years, mean (SD)	49.0 (12.4)	49.6 (11.5)	54.4 (12.9)	55.7 (10.9)	43.6 (14.7)	43.2 (13.8)
FV Leiden carriers, N	69	0	10	0	0	0
PT G20210A carriers, N	11	0	2	0	2	0

DVT, deep vein thrombosis; SD, standard deviation; PT, prothrombin

Supplemental Table 2. Targeted genes and variants

Target area	Gene names
Exons and intron-exon boundaries (N=670)	*ABCB6, ABCG2, ABO, ACTB, ACTG1, ACVRL1, ADAM17, ADIPOQ, ADRA1A, ADRA1B, ADRA1D, ADRA2A, ADRA2B, ADRA2C, ADRB1, ADRB2, ADRB3, ADRBK1, ADRBK2, AGPAT1, ALOX5, ALOX5AP, ANKRD26, ANO6, ANXA1, ANXA10, ANXA11, ANXA13, ANXA2, ANXA3, ANXA4, ANXA5, ANXA6, ANXA7, ANXA8, ANXA8L1, ANXA8L2, ANXA9, AP3B1, APCS, APOA1, APOA2, APOA4, APOA5, APOB, APOC1, APOC2, APOC3, APOC4, APOD, APOE, APOF, APOL1, APOL2, APOL3, APOL4, APOL5, APOL6, ARHGDIB, ARHGEF1, ARHGEF3, ARNT, ARNT2, ARNT3, ASIC2, ATP2A3, AXL, B2M, BAI3, BAZ1B, BIRC5, BLOC1S3, C1GALT1, C1orf114, C1QA, C1QB, C1QC, C1R, C1S, C2, C21orf7, C2orf88, C3, C4A, C4B, C5, C6, C6orf25, C7, C8A, C8B, C8G, C9, CA2, CADM1, CADM2, CADM3, CADM4, CALM1, CALM2, CALM3, CALR, CASK, CASP8AP2, CBS, CCL5, CD34, CD36, CD4, CD40, CD40LG, CD46, CD55, CD59, CDH1, CDKN1A, CDKN2D, CFB, CFD, CFH, CFI, CFP, CHST12, CHST14, CLEC4M, CLU, CMTM5, COL10A1, COL11A1, COL11A2, COL12A1, COL13A1, COL14A1, COL15A1, COL16A1, COL17A1, COL18A1, COL19A1, COL1A1, COL1A2, COL20A1, COL21A1, COL22A1, COL23A1, COL24A1, COL25A1, COL27A1, COL28A1, COL2A1, COL3A1, COL4A1, COL4A2, COL4A3, COL4A3BP, COL4A4, COL4A5, COL4A6, COL5A1, COL5A2, COL5A3, COL6A1, COL6A2, COL6A3, COL6A5, COL7A1, COL8A1, COL8A2, COL9A1, COL9A2, COL9A3, COMP,CR1, CR2, CRP, CRTAM, CTSA, CTSG, CYCS, CYP1A2, CYP2A6, CYP2C9, CYP3A5, CYP4V2, DTNBP1, EBI3, EDEM2, EDIL3, ELANE, EMID2, ENG, ENTPD1, EPAS1, EPB41L1, EPB41L2, EPB41L3, EPR1, EPS8L2, ERBB2, ERBB3, ESR1, FCGR2A, FCGR2B, FERMT3, FLI1, FLNA, FOXA1, FOXA2, FOXA3, FTH1, FTO, GAS6, GATA1, GBGT1, GCKR, GFI1B, GGCX, GLT6D1, GNA12, GNA13, GNAQ, GNAS, GNB2L1, GNG11, GP1BA, GP1BB, GP5, GP6, GP9, GPR30, GPX1, GPX2, GPX3, GPX4, GPX5, GPX6, GPX7, GPX8, GRAP2, HIF1A, HIF3A, HIST1H2AC, HIST1H2BH, HIST1H2BJ, HIST1, H2BK, HIST1H3H, HIVEP1, HLA-A, HLA-B, HLA-C, HOXA11, HPS1, HPS3, HPS4, HPS5, HPS6, HRG, HRH1, HRH2, HRH3, HRH4, HS6ST2, HTR1A, HTR1B, HTR1D, HTR1E, HTR1F, HTR2A, HTR2B, HTR2C, HTR3A, HTR3B, HTR3C, HTR3D, HTR3E, HTR4, HTR5A, HTR6, HTR7, ICAM1, ICAM2, ICAM3, ICAM4, ICAM5, IL10, IL10RA, IL10RB, IL11, IL11RA, IL12A, IL12B, IL12RB1, IL12RB2, IL13, IL13R IL15, IL15RA, IL16, IL17A, IL17B, IL17C, IL17D, IL17F, IL17RA, IL17RB, IL18, IL18R1, IL19, IL1A, IL1B, IL1R1, IL1R2, IL1RN, IL2, IL20, IL20RA, IL20RB, IL21, IL21R, IL22, IL22RA1, IL22RA2, IL23A, IL23R, IL24, IL25, IL26, IL27, IL27RA, IL28A, IL28B, IL28RA, IL29, IL2RA, IL2RB, IL2RG, IL3, IL31, IL31RA, IL32, IL33, IL3RA, IL3RB, IL4, IL4R, IL5, IL5RA, IL6, IL6RA, IL6RB, IL7, IL7RA, IL8, IL8RA, IL8RB, IL9, IL9R, IRAK1, IRAK2, IRAK3, IRAK4, ITFG2, ITGA1, ITGA10, ITGA11, ITGA2, ITGA2B, ITGA3, ITGA4, ITGA5, ITGA6, ITGA7, ITGA8, ITGA9, ITGAD, ITGAE, ITGAL, ITGAM, ITGAV, ITGAX, ITGB1, ITGB2, ITGB3, ITGB4, ITGB5, ITGB6, ITGB7, ITGB8, ITM2B, JAK1, JAK2, JAK3, KLF2, KLK1, KLK10, KLK11, KLK12, KLK13, KLK14, KLK15, KLK2, KLK3, KLK4, KLK5, KLK6, KLK7, KLK8, KLK9, LDLR, LFNG, LOC401913, LPA, LPAR4, LPAR6, LY6E, LY6G6F, LYST, MARCKS, MASP1, MASP2, MASTL, MBL2, MBTPS1, MERTK, MET, MFGE8, MMP2, MMP24, MMP9, MMRN1, MMRN2, MPL, MPP1, MPP3, MPP6, MTHFR, MYBPC3, MYH9, MYL6, MYL9, NAT8B, NBEA, NBEAL2, NCOA1, NFKB1, NFKB2, NNMT, NOS1, NOS2, NOS3, NQO1, NR1I2, NRGN, OAZ1, ODZ1, OS9, OST4, P2RX1, P2RX2, P2RX3, P2RX4, P2RX5, P2RX6, P2RX7, P2RY1, P2RY10,P2RY11, P2RY12, P2RY13, P2RY14, P2RY2, P2RY4, P2RY6, P2RY8, PCSK9, PDGFA, PDIA2, PDLIM1, PDZK1IP1, PECAM1, PF4, PGRMC1, PIGM, PKM2, PKN1, PKN2, PKN3, PLA1A, PLA2G10, PLA2G12A, PLA2G12B, PLA2G1B, PLA2G2A, PLA2G2C, PLA2G2D, PLA2G2E, PLA2G2F, PLA2G3, PLA2G4A, PLA2G4B, PLA2G4C, PLA2G4D,*

Supplemental Table 2. Continued

Target area	Gene names
Exons and intron-exon boundaries (N=670)	*PLA2G4E, PLA2G4F, PLA2G5, PLA2G6, PLA2G7, PLB1, PLCB1, PLCB2, PLCB3, PLCB4, PLCD1, PLCD3, PLCD4, PLCE1, PLCG1, PLCG2, PLCH1, PLCH2, PLCL1, PLCL2, PLCZ1, PLD1, PLD2, PPBP, PPDPF, PPP1CA, PPP1CB, PPP2CA, PPP2CB, PPP3CA, PPP3CB, PRKAR2B, PRKCA, PRKCB, PRKCD, PRKCE, PRKCG, PRKCH, PRKCI, PRKCQ, PRKCZ, PRKD1, PRKD2, PRKD3, PRTN3, PSEN1, PTAFR, PTGDR, PTGDS, PTGER1, PTGER2, PTGER3, PTGER4, PTGES, PTGES2, PTGES3, PTGFR, PTGIR, PTGIS, PTGS1, PTGS2, PTK2B, PTX3, PVRL3, RAC1, RAP1B, RASGRP2, RBM8A, RGS10, RGS18, RGS7, RND2, RUNX1, SAA1, SAA2, SAA4, SCARA5, SCARB1, SCIN, SDCBP, SDPR, SELE, SELL, SELP, SERF2, SERPINA1, SERPINA11, SERPINA12, SERPINA13, SERPINA2, SERPINA3, SERPINA4, SERPINA6, SERPINA7, SERPINA8, SERPINB1, SERPINB10, SERPINB11, SERPINB12, SERPINB13, SERPINB2, SERPINB3, SERPINB4, SERPINB5, SERPINB6, SERPINB7, SERPINB8, SERPINB9, SERPINE2, SERPINF1, SERPING1, SERPINH1, SERPINI1, SERPINI2, SH3BGRL2, SH3BGRL3, SHBG, SMAD4, SMOX, SNCA, SPARC, SPINT1, SPINT2, SPINT3, SPINT4, SRGN, STAB2, STX2, STXBP3, STXBP5, SULT1A1, SULT1E1, TACR1, TAGLN2, TAOK1, TBX2, TBXA2R, TBXAS1, TC2N, TFPI2, TGFB1, THBS1, THBS2, THBS3, THBS4, THPO, TLN1, TLR1, TLR10, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9, TNF, TNFRSF1A, TPH1, TREML1, TRIM58, TUBA4A, TUBB1, TXNLA,TYK2, TYRO3, UGT1A1, UGT1A3, UGT1A9, UGT2B7, USF1, VASP, VCAM1, VHL, VHLL, VKORC1, VTN, WAS, WASF2, WDR66, WIPF1, ZNF185, ZNF544*
Exons, intron-exon boundaries, and UTRS (N=48)	*A2M, ADAMTS13, APOH, C4BPA, C4BPB, CALU, CPB2, F10, F11, F12, F13A1, F13B, F2, F2R, F2RL1, F2RL2, F2RL3, F3, F7, F9, FGA, FGB, FGG, HABP2, HNF1A, HNF4A, KLKB1, KNG1, LMAN1, LRP1, MCFD2, PLAT, PLAU, PLAUR, PLG, PROC, PROCR, PROS1, PROZ, SERPINA10, SERPINA5, SERPINA9, SERPINC1, SERPIND1, SERPINE1, SERPINF2, TFPI, THBD*
Entire gene including promoter (N=3)	*F5, F8, VWF*
Variants (N=178)	rs988436, rs942793, rs1368136, rs2060983, rs1404402, rs1016120, rs1414411, rs2014303, rs1030626, rs1517661, rs764138, rs2218497, rs725379, rs1377724, rs1406121, rs869538, rs1905471, rs764681, rs1280100, rs723211, rs1368666, rs1896974, rs1366578, rs262981, rs262980, rs262982, rs2876712, rs2366267, rs234693, rs32524, rs3845765, rs746600, rs2227139, rs2132055, rs869537, rs2389530, rs1964428, rs1377725, rs710616, rs2312211, rs1343747, rs719985, rs768131, rs1428555, rs1073197, rs1398064, rs31096, rs718255, rs725105, rs1937147, rs2351002, rs1383279, rs1393358, rs1376264, rs1854673, rs1585217, rs727864, rs461081, rs1573275, rs1343809, rs1986710, rs959284, rs956377, rs448041, rs222948, rs952993, rs959100, rs743337, rs1074342, rs1395479, rs1476873, rs950668, rs1846466, rs1531124, rs1401535, rs953177, rs722626, rs952784, rs1327806, rs2225251, rs1931621, rs1412917, rs950228, rs718092, rs1825015, rs728337, rs2383188, rs754798, rs721544, rs361263, rs958967,

Supplemental Table 2. Continued

Target area	Gene names
Variants (N=178)	rs1416464, rs726364, rs1526952, rs1390049, rs1485254, rs1368666, rs829669, rs1362489, rs703698, rs10108270, rs10236187, rs1040045, rs1040404, rs10496971, rs10510228, rs10512572, rs10513300, rs10839880, rs11227699, rs11652805, rs12130799, rs12439433, rs12544346, rs12629908, rs12657828, rs1296819, rs1325502, rs13400937, rs1369093, rs1407434, rs1471939, rs1513181, rs1760921, rs1837606, rs1871428, rs1950993, rs200354, rs2030763, rs2073821, rs2125345, rs214678, rs2306040, rs2330442, rs2357442, rs2397060, rs2416791, rs2504853, rs260690, rs2627037, rs2702414, rs2946788, rs2986742, rs3118378, rs316598, rs316873, rs32314, rs3737576, rs3745099, rs3784230, rs1799963, rs8176719, rs2066865, rs2036914, rs2069951, rs2289252, rs4149755, rs2069952, rs2227589, rs169713, rs3136520, rs1799809, rs867186, rs1613662, rs3136516, rs1039084, rs2001490, rs6003, rs670659, rs6048, rs5985, rs8176592, rs3822057, rs1523127, rs3742264, rs710446, rs3813948. rs12941510

Supplemental Table 3. Sensitivity analysis excluding recurrent VT patients

							Discovery analysis		Sensitivity analysis	
rsID	**Chr.**	**Position**	**A_1/A_2**	**RAF**	**Class**	**Gene**	**OR (95%CI)**	**P**	***OR (95%CI)**	**P**
rs3766110	1	169515183	C/A	0.774	intronic	*F5*	1.54 (1.27-1.86)	$1.07x10^{-5}$	1.49 (1.22-1.81)	$9.87x10^{-5}$
rs3766111	1	169515204	C/T	0.773	intronic	*F5*	1.57 (1.29-1.91)	$6.82x10^{-6}$	1.52 (1.24-1.86)	$6.27x10^{-5}$
rs3766113	1	169515307	G/A	0.77	intronic	*F5*	1.55 (1.28-1.88)	$8.59x10^{-6}$	1.51 (1.24-1.85)	$5.28x10^{-5}$
rs6672595	1	169515536	T/C	0.757	intronic	*F5*	1.58 (1.29-1.92)	$6.11x10^{-6}$	1.54 (1.25-1.89)	$3.67x10^{-5}$
rs6050	4	155507590	T/C	0.393	missense	*FGA*	1.66 (1.37-2.02)	$2.33x10^{-7}$	1.55 (1.27-1.90)	$1.47x10^{-5}$
rs2066865	4	155525276	G/A	0.352	downstream	*FGG*	1.60 (1.33-1.92)	$4.86x10^{-7}$	1.56 (1.29-1.89)	$5.43x10^{-6}$
rs3733402	4	187158034	G/A	0.573	missense	*KLKB1*	1.55 (1.30-1.86)	$1.27x10^{-6}$	1.49 (1.24-1.79)	$2.78x10^{-5}$
rs4253399	4	187188094	T/G	0.458	intronic	*F11*	1.50 (1.27-1.76)	$8.34x10^{-7}$	1.43 (1.21-1.69)	$3.48x10^{-5}$
rs3822057	4	187188152	A/C	0.545	intronic	*F11*	1.44 (1.23-1.70)	$6.74x10^{-6}$	1.35 (1.15-1.60)	$3.80x10^{-4}$
rs2036914	4	187192481	T/C	0.602	intronic	*F11*	1.65 (1.38-1.97)	$2.47x10^{-8}$	1.55 (1.29-1.86)	$3.02x10^{-6}$
rs8176719	9	136132908	T/TC	0.451	Frameshift	*ABO*	1.90 (1.61-2.24)	$1.39x10^{-14}$	1.88 (1.59-2.23)	$2.28x10^{-13}$
rs4962040	9	136133531	G/A	0.594	intronic	*ABO*	1.53 (1.28-1.83)	$3.67x10^{-6}$	1.51 (1.26-1.82)	$1.21x10^{-5}$

Chr chromosome; A1 reference allele; A2 risk allele; RAF risk allele frequency; OR odds ratio; CI confidence interval
After exclusion of recurrent VT patients, single-variant association analyses were performed in 656 patients and 598 controls, adjusting for age, sex, and (study) origin.

Supplemental Table 4. Discovery and conditional association analyses of variants in *ABO*, *CYP4V2-KLKB1-F11*, and *FGA-FGG* (P >1.38x10^{-5})

							Discovery analyses		$^{\$}$Conditional analyses	
rsID	**Chr**	**Position**	**A1/A2**	**RAF**	**Functional class**	**Gene**	**OR (95%CI)**	**P**	**OR (95%CI)**	**P**
rs2070022	4	155504948	A/G	0.828	3' UTR	*FGA*	1.31 (1.06-1.62)	0.013	1.06 (0.81-1.39)	0.68
rs2070011	4	155511897	C/T	0.406	5' UTR	*FGA*	1.37 (1.16-1.62)	2.18x10^{-4}	0.90 (0.67-1.21)	0.478
rs1049636	4	155525970	G/A	0.761	3' UTR	*FGG*	1.30 (1.08-1.56)	0.005	1.03 (0.82-1.31)	0.778
rs13146272	4	187120211	C/A	0.692	missense	*CYP4V2*	1.35 (1.12-1.61)	0.001	1.20 (0.99-1.46)	0.07
rs3817184	4	187122304	C/T	0.456	intronic	*CYP4V2*	1.43 (1.21-1.69)	2.24x10^{-5}	1.21 (1.00-1.46)	0.054
rs3736455	4	187122319	T/G	0.706	synonymous	*CYP4V2*	1.31 (1.09-1.56)	0.003	1.18 (0.97-1.44)	0.104
rs34745240	4	187122332	G/A	0.046	missense	*CYP4V2*	1.42 (0.95-2.11)	0.083	1.18 (0.79-1.78)	0.42
rs4253301	4	187173012	G/T	0.868	missense	*KLKB1*	1.31 (1.03-1.66)	0.029	1.00 (0.75-1.33)	0.978
rs925453	4	187179210	T/C	0.715	synonymous	*KLKB1*	1.25 (1.05-1.49)	0.012	1.14 (0.94-1.39)	0.187
rs3087505	4	187179486	A/G	0.912	3' UTR	*KLKB1*	1.20 (0.92-1.56)	0.188	0.97 (0.71-1.33)	0.854
rs3733403	4	187187135	G/C	0.898	5' UTR	*F11*	1.05 (0.79-1.38)	0.752	1.28 (0.96-1.72)	0.093
rs4253398	4	187188061	C/T	0.705	intronic	*F11*	1.16 (0.97-1.39)	0.115	0.87 (0.69-1.09)	0.222
rs35709976	4	187188141	G/GAT	0.898	intronic	*F11*	1.33 (1.03-1.72)	0.027	1.12 (0.83-1.51)	0.469
rs2289252*	4	187207381	C/T	0.528	intronic	*F11*	1.39 (1.17-1.65)	1.94x10^{-4}	1.15 (0.92-1.44)	0.211
rs5976	4	187209729	A/G	0.949	synonymous	*F11*	1.21 (0.84-1.76)	0.307	0.89 (0.58-1.35)	0.579
rs4253429*	4	187210033	G/A	0.803	3' UTR	*F11*	1.28 (1.04-1.57)	0.021	1.01 (0.78-1.31)	0.958
rs4253430*	4	187210064	C/G	0.590	3' UTR	*F11*	1.38 (1.16-1.63)	2.50x10^{-4}	1.05 (0.83-1.31)	0.694
rs4253865*	4	187210090	A/G	0.950	3' UTR	*F11*	1.20 (0.83-1.73)	0.334	0.87 (0.57-1.32)	0.499
rs1062547*	4	187210247	T/A	0.605	3' UTR	*F11*	1.33 (1.13-1.57)	7.97x10^{-4}	1.01 (0.80-1.26)	0.954

Supplemental Table 4. Continued

							Discovery analyses		[$]Conditional analyses	
rsID	**Chr**	**Position**	**A1/A2**	**RAF**	**Functional class**	**Gene**	**OR (95%CI)**	**P**	**OR (95%CI)**	**P**
rs186377697	4	187210319	T/A	0.987	3' UTR	*F11*	1.39 (0.69-2.83)	0.359	1.32 (0.60-2.92)	0.495
rs8176749	9	136131188	C/T	0.087	ncRNA	*ABO*	1.53 (1.15-2.04)	0.004	1.09 (0.80-1.49)	0.569
rs8176748	9	136131289	T/C	0.774	ncRNA	*ABO*	1.37 (1.13-1.66)	0.001	1.01 (0.82-1.26)	0.898
rs8176747	9	136131315	C/G	0.092	ncRNA	*ABO*	1.60 (1.19-2.14)	0.002	1.13 (0.82-1.55)	0.452
rs41302905	9	136131316	T/C	0.982	ncRNA	*ABO*	1.38 (0.77-2.48)	0.276	2.09 (1.14-3.81)	0.017
rs8176746	9	136131322	G/T	0.096	ncRNA	*ABO*	1.56 (1.18-2.08)	0.002	1.10 (0.81-1.50)	0.527
rs8176745	9	136131347	A/G	0.766	ncRNA	*ABO*	1.37 (1.13-1.66)	0.001	1.02 (0.82-1.26)	0.882
rs8176744	9	136131350	T/G	0.967	ncRNA	*ABO*	1.72 (1.10-2.67)	0.017	1.37 (0.86-2.19)	0.188
rs8176743	9	136131415	C/T	0.099	ncRNA	*ABO*	1.52 (1.14-2.02)	0.004	1.07 (0.79-1.46)	0.654
rs8176742	9	136131437	T/C	0.760	ncRNA	*ABO*	1.39 (1.15-1.68)	$8.42x10^{-4}$	1.01 (0.81-1.25)	0.959
rs8176741	9	136131461	G/A	0.094	ncRNA	*ABO*	1.57 (1.18-2.08)	0.002	1.14 (0.84-1.54)	0.399
rs8176740	9	136131472	T/A	0.759	ncRNA	*ABO*	1.34 (1.10-1.63)	0.003	0.98 (0.79-1.23)	0.882
rs8176739	9	136131523	A/G	0.983	ncRNA	*ABO*	1.63 (0.86-3.08)	0.136	1.25 (0.65-2.41)	0.506
rs7853989	9	136131592	G/C	0.093	ncRNA	*ABO*	1.37 (1.03-1.81)	0.029	0.96 (0.71-1.31)	0.808
rs1053878	9	136131651	G/A	0.046	ncRNA	*ABO*	1.24 (0.85-1.82)	0.266	0.89 (0.60-1.34)	0.584
rs8176720	9	136132873	C/T	0.618	ncRNA	*ABO*	1.08 (0.91-1.29)	0.374	1.00 (0.84-1.20)	0.979
rs75179845	9	136132954	T/C	0.098	intronic	*ABO*	1.49 (1.11-2.00)	0.008	1.07 (0.78-1.46)	0.684
rs8176718	9	136132957	T/C	0.766	intronic	*ABO*	1.39 (1.15-1.68)	$7.46x10^{-4}$	1.03 (0.83-1.27)	0.821
rs512770	9	136133506	A/G	0.819	ncRNA	*ABO*	1.36 (1.11-1.67)	0.003	1.11 (0.89-1.38)	0.348

Supplemental Table 4. Continued

							Discovery analyses		$Conditional analyses	
rsID	**Chr**	**Position**	**A1/A2**	**RAF**	**Functional class**	**Gene**	**OR (95%CI)**	**P**	**OR (95%CI)**	**P**
rs549443	9	136135237	A/G	0.789	ncRNA	*ABO*	1.32 (1.08-1.61)	0.006	0.97 (0.78-1.21)	0.792
rs549446	9	136135238	T/C	0.785	ncRNA	*ABO*	1.38 (1.14-1.67)	0.001	1.00 (0.81-1.25)	0.968
rs688976	9	136136770	A/C	0.803	ncRNA	*ABO*	1.31 (1.06-1.61)	0.012	0.97 (0.77-1.22)	0.807
rs8176696	9	136136773	T/C	0.980	ncRNA	*ABO*	1.65 (0.91-2.97)	0.098	1.15 (0.62-2.13)	0.656

Chr chromosome; A1 reference allele; A2 risk allele; RAF risk allele frequency; UTR untranslated region; ncRNA non-coding RNA; OR odds ratio; CI confidence interval; NA not applicable

** F11* variants rs2289252, rs4253429, rs4253430, rs4253865, and rs1062547 also map to a non-coding RNA transcript of *LOC285441 (F11-AS1).*

$ We conducted conditional logistic regression analyses in which we adjusted for the lead variant per locus (i.e. *F5* rs6672595, *FGA* rs6050, *F11* rs2036914, and *ABO* rs8176719).

Supplemental Table 5. Reciprocal conditional association analyses rs6050 and rs2066865

rsID	**Chr**	**Position**	**A1/A2**	**RAF**	**Class**	**Gene**	***LD R^2 (D')**	**OR (95% CI)**	**P**	**OR_{cond} (95% CI)**	**P_{cond}**
rs6050	4	155507590	T/C	0.393	missense	*FGA*	0.90 (0.99)	1.66 (1.37-2.02)	$2.33x10^{-7}$	1.24 (0.73-2.12)	0.430
rs2066865	4	155525276	G/A	0.352	downstream	*FGG*	0.90 (0.99)	1.60 (1.33-1.92)	$4.86x10^{-7}$	1.37 (0.81-2.31)	0.245

Chr chromosome; A1 reference allele; A2 risk allele; RAF risk allele frequency; LD linkage disequilibrium; OR odds ratio; CI confidence interval; cond conditional

* Linkage disequilibrium based on data of European population of 1000 Genomes Project.

Supplemental Table 6. Association analysis at the *F5* association locus conditional on rs4524

								Discovery analyses		Conditional analyses[$]	
rsID	Chr	Position	A1/A2	RAF	Functional class	[*]LD R^2 (D')	Gene	OR (95% CI)	P	OR (95% CI)	P
rs2009814	1	169471917	T/C	0.721	intergenic	0.96 (0.99)	.	1.42 (1.18-1.71)	$2.29x10^{-4}$	1.91 (0.73-5.00)	0.186
rs974793	1	169478654	T/C	0.712	intergenic	0.96 (0.99)	.	1.36 (1.13-1.64)	$1.44x10^{-3}$	1.32 (0.47-3.69)	0.596
rs2187952	1	169481950	A/G	0.697	3' UTR	0.96 (0.99)	*F5*	1.39 (1.14-1.68)	$9.44x10^{-4}$	1.21 (0.30-4.98)	0.789
rs4656685	1	169483844	T/C	0.746	intronic	0.96 (0.99)	*F5*	1.40 (1.17-1.68)	$3.02x10^{-4}$	0.74 (0.20-2.67)	0.645
rs2227244	1	169489358	C/T	0.728	intronic	0.96 (0.99)	*F5*	1.36 (1.13-1.63)	$1.12x10^{-3}$	0.77 (0.13-4.72)	0.778
rs2213867	1	169489585	C/T	0.727	intronic	0.96 (0.99)	*F5*	1.38 (1.15-1.66)	$6.34x10^{-4}$	0.59 (0.10-3.29)	0.544
rs9332655	1	169490592	G/A	0.730	intronic	0.97 (1.00)	*F5*	1.36 (1.13-1.64)	$1.03x10^{-3}$	0.29 (0.03-2.68)	0.277
rs9332652	1	169491021	AT/A	0.706	intronic	0.97 (1.00)	*F5*	1.35 (1.14-1.60)	$6.30x10^{-4}$	1.06 (0.61-1.83)	0.847
rs9332627	1	169497820	A/G	0.720	intronic	0.97 (1.00)	*F5*	1.41 (1.17-1.71)	$2.80x10^{-4}$	0.39 (0.04-3.78)	0.414
rs2420373	1	169498181	T/C	0.709	intronic	0.97 (1.00)	*F5*	1.44 (1.19-1.75)	$1.70x10^{-4}$	NA	NA
rs2187953	1	169499381	C/A	0.732	intronic	0.97 (1.00)	*F5*	1.40 (1.17-1.68)	$3.25x10^{-4}$	0.38 (0.04-3.73)	0.407
rs9332620	1	169499951	T/C	0.717	intronic	0.98 (1.00)	*F5*	1.39 (1.15-1.68)	$5.88x10^{-4}$	NA	NA
rs9332619	1	169500348	A/G	0.696	intronic	0.98 (1.00)	*F5*	1.37 (1.13-1.66)	$1.43x10^{-3}$	0 (0-∞)	0.968
rs6670393	1	169502533	C/A	0.723	intronic	0.98 (1.00)	*F5*	1.35 (1.12-1.62)	$1.50x10^{-3}$	NA	NA
rs10800453	1	169507076	A/T	0.772	intronic	1.00 (1.00)	*F5*	1.41 (1.16-1.71)	$5.24x10^{-4}$	NA	NA
rs9287090	1	169510380	A/G	0.714	synonymous	1.00 (1.00)	*F5*	1.36 (1.12-1.64)	$1.65x10^{-3}$	1.48 (0.09-24)	0.782
rs6032	1	169511555	C/T	0.725	missense	1.00 (1.00)	*F5*	1.37 (1.14-1.65)	$7.90x10^{-4}$	NA	NA
rs6021	1	169512027	C/T	0.729	synonymous	1.00 (1.00)	*F5*	1.36 (1.13-1.63)	$1.35x10^{-3}$	NA	NA

Supplemental Table 6. Continued

								Discovery analyses		Conditional analyses[$]	
rsID	**Chr**	**Position**	**A1/A2**	**RAF**	**Functional class**	**[*]LD R^2 (D')**	**Gene**	**OR (95% CI)**	**P**	**OR (95% CI)**	**P**
rs6016	1	169512120	A/G	0.710	synonymous	1.00 (1.00)	*F5*	1.40 (1.16-1.70)	$4.71x10^{-4}$	NA	NA
rs2239851	1	169512497	A/C	0.721	intronic	1.00 (1.00)	*F5*	1.38 (1.14-1.66)	$7.59x10^{-4}$	NA	NA
rs6675244	1	169512562	C/T	0.731	intronic	0.99 (1.00)	*F5*	1.37 (1.14-1.64)	$8.40x10^{-4}$	NA	NA
rs6662593	1	169512594	A/G	0.724	intronic	1.00 (1.00)	*F5*	1.36 (1.13-1.63)	$1.21x10^{-3}$	NA	NA
rs6662696	1	169512651	A/G	0.725	intronic	1.00 (1.00)	*F5*	1.35 (1.12-1.62)	$1.49x10^{-3}$	NA	NA
rs9332600	1	169512913	T/C	0.727	intronic	1.00 (1.00)	*F5*	1.43 (1.19-1.72)	$1.50x10^{-4}$	1.94 (0.17-22)	0.592
rs9287092	1	169513436	A/C	0.747	intronic	0.86 (0.94)	*F5*	1.41 (1.17-1.71)	$2.95x10^{-4}$	1.44 (0.85-2.43)	0.171
rs9332595	1	169514355	C/G	0.758	intronic	0.77 (0.93)	*F5*	1.51 (1.24-1.83)	$4.10x10^{-5}$	1.81 (1.21-2.72)	$4.17x10^{-3}$
rs929130	1	169514779	A/G	0.760	intronic	0.77 (0.93)	*F5*	1.55 (1.27-1.89)	$1.42x10^{-5}$	1.59 (1.07-2.37)	0.022
rs3766110	1	169515183	C/A	0.774	intronic	0.77 (0.93)	*F5*	1.54 (1.27-1.86)	$1.07x10^{-5}$	1.60 (1.09-2.36)	0.017
rs3766111	1	169515204	C/T	0.773	intronic	0.77 (0.93)	*F5*	1.57 (1.29-1.91)	$6.82x10^{-6}$	1.52 (1.02-2.27)	0.041
rs3766112	1	169515296	G/C	0.748	intronic	0.77 (0.93)	*F5*	1.55 (1.27-1.90)	$1.53x10^{-5}$	1.77 (1.17-2.69)	0.007
rs3766113	1	169515307	G/A	0.770	intronic	0.77 (0.93)	*F5*	1.55 (1.28-1.88)	$8.59x10^{-6}$	1.57 (1.06-2.32)	0.024
rs6672595	1	169515536	T/C	0.757	intronic	0.77 (0.93)	*F5*	1.58 (1.29-1.92)	$6.11x10^{-6}$	1.65 (1.10-2.48)	0.016
rs13306345	1	169515874	A/T	0.775	intronic	0.75 (0.90)	*F5*	1.46 (1.20-1.77)	$1.33x10^{-4}$	1.63 (1.09-2.43)	0.016
rs1894695	1	169517833	G/C	0.763	intronic	0.75 (0.90)	*F5*	1.47 (1.21-1.78)	$9.68x10^{-5}$	1.72 (1.15-2.56)	0.008
rs1894696	1	169517975	T/C	0.784	intronic	0.75 (0.90)	*F5*	1.44 (1.18-1.76)	$2.71x10^{-4}$	1.47 (1.00-2.15)	0.051
rs72248387	1	169518819	T/TCA	0.766	intronic	0.75 (0.90)	*F5*	1.52 (1.25-1.84)	$2.56x10^{-5}$	1.47 (1.00-2.17)	0.052
rs10158595	1	169520364	T/C	0.765	intronic	0.74 (0.89)	*F5*	1.49 (1.23-1.81)	$5.22x10^{-5}$	1.50 (1.02-2.20)	0.038

Supplemental Table 6. Continued

rsID	Chr	Position	A1/A2	RAF	Functional class	*LD R^2 (D')	Gene	Discovery analyses OR (95% CI)	Discovery analyses P	Conditional analyses$ OR (95% CI)	Conditional analyses$ P
rs2420375	1	169520459	C/G	0.796	intronic	0.74 (0.89)	*F5*	1.41 (1.16-1.72)	$5.98x10^{-4}$	1.43 (0.98-2.08)	0.065
rs2420376	1	169520549	A/G	0.788	intronic	0.74 (0.89)	*F5*	1.45 (1.20-1.76)	$1.29x10^{-4}$	1.52 (1.04-2.21)	0.03
rs2420377	1	169520592	A/G	0.788	intronic	0.74 (0.89)	*F5*	1.42 (1.17-1.72)	$4.12x10^{-4}$	1.46 (0.99-2.15)	0.055

Chr chromosome; A1 reference allele; A2 risk allele; RAF risk allele frequency; LD linkage disequilibrium; OR odds ratio; CI confidence interval; cond conditional; UTR untranslated region; NA not applicable

* Linkage disequilibrium with rs4524, based on data of European population of 1000 Genomes Project.

$ Conditional analyses on rs4525

Supplemental Table 7. Suggestive single variant associations (FDR <0.10)

rsID	Chr.	Position	A1/A2	RAF	Functional class	Gene	Discovery analyses OR (95% CI)	Discovery analyses P	FDR	Conditional analyses$ OR (95% CI)	Conditional analyses$ P
rs2009814	1	169471917	T/C	0.721	intergenic	.	1.42 (1.18-1.71)	$2.29x10^{-4}$	0.032	1.12 (0.78-1.61)	0.544
rs974793	1	169478654	T/C	0.712	intergenic	.	1.36 (1.13-1.64)	0.001	0.090	1.04 (0.71-1.51)	0.846
rs2187952	1	169481950	A/G	0.697	3' UTR	*F5*	1.39 (1.14-1.68)	$9.44x10^{-4}$	0.072	0.97 (0.66-1.43)	0.88
rs4656685	1	169483844	T/C	0.746	intronic	*F5*	1.40 (1.17-1.68)	$3.02x1^{-4}$	0.035	0.95 (0.65-1.40)	0.80
rs2227244	1	169489358	C/T	0.728	intronic	*F5*	1.36 (1.13-1.63)	0.001	0.08	0.94 (0.63-1.39)	0.746
rs2213867	1	169489585	C/T	0.727	intronic	*F5*	1.38 (1.15-1.66)	$6.34x10^{-4}$	0.058	0.93 (0.63-1.37)	0.711
rs9332655	1	169490592	G/A	0.730	intronic	*F5*	1.36 (1.13-1.64)	0.001	0.075	0.91 (0.62-1.34)	0.629
rs9332652	1	169491021	AT/A	0.706	intronic	*F5*	1.35 (1.14-1.60)	$6.30x10^{-4}$	0.058	1.01 (0.73-1.38)	0.964

Supplemental Table 7. Continued

							Discovery analyses			Conditional analyses[$]	
rsID	Chr.	Position	A1/A2	RAF	Functional class	Gene	OR (95% CI)	P	FDR	OR (95% CI)	P
rs9332627	1	169497820	A/G	0.720	intronic	*F5*	1.41 (1.17-1.71)	$2.80x10^{-4}$	0.035	0.94 (0.64-1.39)	0.77
rs2420373	1	169498181	T/C	0.709	intronic	*F5*	1.44 (1.19-1.75)	$1.70x10^{-4}$	0.027	0.97 (0.65-1.44)	0.888
rs2187953	1	169499381	C/A	0.732	intronic	*F5*	1.40 (1.17-1.68)	$3.25x10^{-4}$	0.037	0.94 (0.64-1.39)	0.765
rs9332620	1	169499951	T/C	0.717	intronic	*F5*	1.39 (1.15-1.68)	$5.88x10^{-4}$	0.058	0.90 (0.60-1.33)	0.59
rs9332619	1	169500348	A/G	0.696	intronic	*F5*	1.37 (1.13-1.66)	0.001	0.09	0.96 (0.64-1.43)	0.841
rs6670393	1	169502533	C/A	0.723	intronic	*F5*	1.35 (1.12-1.62)	0.002	0.09	0.89 (0.60-1.31)	0.554
rs10800453	1	169507076	A/T	0.772	intronic	*F5*	1.41 (1.16-1.71)	$5.24x10^{-4}$	0.054	1.00 (0.65-1.55)	0.992
rs9287090	1	169510380	A/G	0.714	synonymous	*F5*	1.36 (1.12-1.64)	0.002	0.096	0.97 (0.66-1.43)	0.872
rs6032	1	169511555	C/T	0.725	missense	*F5*	1.37 (1.14-1.65)	$7.90x10^{-4}$	0.065	0.99 (0.66-1.46)	0.939
rs4525	1	169511734	C/T	0.736	missense	*F5*	1.38 (1.15-1.66)	$6.40x10^{-4}$	0.058	1.02 (0.69-1.51)	0.93
rs6021	1	169512027	C/T	0.729	synonymous	*F5*	1.36 (1.13-1.63)	0.001	0.088	0.95 (0.64-1.40)	0.787
rs6016	1	169512120	A/G	0.710	synonymous	*F5*	1.40 (1.16-1.70)	$4.71x10^{-4}$	0.05	0.97 (0.65-1.44)	0.875
rs2239851	1	169512497	A/C	0.721	intronic	*F5*	1.38 (1.14-1.66)	$7.59x10^{-4}$	0.065	0.97 (0.66-1.45)	0.90
rs6675244	1	169512562	C/T	0.731	intronic	*F5*	1.37 (1.14-1.64)	$8.40x10^{-4}$	0.066	0.95 (0.65-1.41)	0.815
rs6662593	1	169512594	A/G	0.724	intronic	*F5*	1.36 (1.13-1.63)	0.001	0.082	0.96 (0.64-1.42)	0.822
rs6662696	1	169512651	A/G	0.725	intronic	*F5*	1.35 (1.12-1.62)	0.001	0.09	0.95 (0.64-1.40)	0.788
rs9332600	1	169512913	T/C	0.727	intronic	*F5*	1.43 (1.19-1.72)	$1.50x10^{-4}$	0.025	0.99 (0.67-1.47)	0.96
rs9287092	1	169513436	A/C	0.747	intronic	*F5*	1.41 (1.17-1.71)	$2.95x10^{-4}$	0.035	0.69 (0.40-1.18)	0.175
rs9332595	1	169514355	C/G	0.758	intronic	*F5*	1.51 (1.24-1.83)	$4.10x10^{-5}$	0.009	NA	NA

Supplemental Table 7. Continued

rsID	Chr.	Position	A1/A2	RAF	Functional class	Gene	Discovery analyses			Conditional analyses[$]	
							OR (95% CI)	P	FDR	OR (95% CI)	P
rs929130	1	169514779	A/G	0.760	intronic	*F5*	1.55 (1.27-1.89)	$1.42x10^{-5}$	0.004	NA	NA
rs3766112	1	169515296	G/C	0.748	intronic	*F5*	1.55 (1.27-1.90)	$1.53x10^{-5}$	0.004	NA	NA
rs13306345	1	169515874	A/T	0.775	intronic	*F5*	1.46 (1.20-1.77)	$1.33x10^{-4}$	0.023	1.05 (0.09-11.7)	0.967
rs1894695	1	169517833	G/C	0.763	intronic	*F5*	1.47 (1.21-1.78)	$9.68x10^{-5}$	0.018	0.59 (0.06-5.78)	0.647
rs1894696	1	169517975	T/C	0.784	intronic	*F5*	1.44 (1.18-1.76)	$2.71x10^{-4}$	0.035	0.56 (0.06-5.49)	0.616
rs72248387	1	169518819	T/TCA	0.766	intronic	*F5*	1.52 (1.25-1.84)	$2.56x10^{-5}$	0.006	0.61 (0.06-6.01)	0.673
rs10158595	1	169520364	T/C	0.765	intronic	*F5*	1.49 (1.23-1.81)	$5.22x10^{-5}$	0.01	0.62 (0.12-3.23)	0.567
rs2420375	1	169520459	C/G	0.796	intronic	*F5*	1.41 (1.16-1.72)	$5.98x10^{-4}$	0.058	0.59 (0.11-3.07)	0.528
rs2420376	1	169520549	A/G	0.788	intronic	*F5*	1.45 (1.20-1.76)	$1.29x10^{-4}$	0.023	0.60 (0.11-3.13)	0.543
rs2420377	1	169520592	A/G	0.788	intronic	*F5*	1.42 (1.17-1.72)	$4.12x10^{-4}$	0.045	0.31 (0.04-2.65)	0.282
rs2070011	4	155511897	C/T	0.406	5' UTR	*FGA*	1.37 (1.16-1.62)	$2.19x10^{-4}$	0.031	0.90 (0.67-1.21)	0.478
rs13146272	4	187120211	C/A	0.692	missense	*CYP4V2*	1.35 (1.12-1.61)	0.001	0.083	1.20 (0.99-1.46)	0.07
rs3817184	4	187122304	C/T	0.456	intronic	*CYP4V2*	1.43 (1.21-1.69)	$2.24x10^{-5}$	0.005	1.21 (1.00-1.46)	0.054
rs2289252*	4	187207381	C/T	0.528	intronic	*F11*	1.39 (1.17-1.65)	$1.94x10^{-4}$	0.029	1.15 (0.92-1.44)	0.211
rs4253430*	4	187210064	C/G	0.590	3' UTR	*F11*	1.38 (1.16-1.63)	$2.50x10^{-4}$	0.033	1.05 (0.83-1.31)	0.694
rs1062547*	4	187210247	T/A	0.605	3' UTR	*F11*	1.33 (1.13-1.57)	$7.97x10^{-4}$	0.065	1.01 (0.80-1.26)	0.954
rs8176748	9	136131289	T/C	0.774	ncRNA	*ABO*	1.37 (1.13-1.66)	0.001	0.081	1.01 (0.82-1.26)	0.898
rs8176745	9	136131347	A/G	0.766	ncRNA	*ABO*	1.37 (1.13-1.66)	0.001	0.081	1.02 (0.82-1.26)	0.882
rs8176742	9	136131437	T/C	0.760	ncRNA	*ABO*	1.39 (1.15-1.68)	$8.42x10^{-4}$	0.066	1.01 (0.81-1.25)	0.959

Supplemental Table 7. Continued

rsID	Chr.	Position	A1/A2	RAF	Functional class	Gene	Discovery analyses OR (95% CI)	P	FDR	Conditional analyses[$] OR (95% CI)	P
rs8176718	9	136132957	T/C	0.766	intronic	*ABO*	1.39 (1.15-1.68)	$7.46x10^{-4}$	0.065	1.03 (0.83-1.27)	0.821
rs549446	9	136135238	T/C	0.785	unknown	*ABO*	1.38 (1.14-1.67)	0.001	0.075	1.00 (0.81-1.25)	0.968

Chr. chromosome; A1 reference allele; A2 risk allele; RAF risk allele frequency; UTR untranslated region; ncRNA; non-coding RNA; OR odds ratio; CI confidence interval; FDR false discovery rate; NA not applicable

** F11* variants rs2289252, rs4253430, and rs1062547 also map to a non-coding RNA transcript of *LOC285441 (F11-AS1).*

[$] We conducted conditional logistic regression analyses in which we adjusted for the lead variant per locus (i.e. *F5* rs6672595, *FGA* rs6050, *F11* rs2036914, and *ABO* rs8176719).

Supplemental Table 8. Single-variant association analyses of common *MASP1* variants

rsID	Chr.	Position	A1/A2	RAF	Functional class	Gene	[*]LD R^2 (D')	OR (95%CI)	P
rs62294422	3	186952037	G/A	0.067	3' UTR	*MASP1*		1.06 (0.78-1.45)	0.707
rs12489890	3	186952415	A/G	0.821	3' UTR	*MASP1*	0.00 (1.00)	1.04 (0.84-1.28)	0.730
rs72549168	3	186952569	C/T	0.016	3' UTR	*MASP1*	1.00 (1.00)	3.11 (1.42-6.81)	0.005
rs850314	3	186952587	T/C	0.622	3' UTR	*MASP1*	0.01 (1.00)	1.03 (0.87-1.21)	0.766
rs1109452	3	186952588	A/G	0.699	3' UTR	*MASP1*	0.01 (1.00)	1.15 (0.97-1.38)	0.117
rs72549262	3	186952914	C/G	0.905	3' UTR	*MASP1*	0.00 (1.00)	1.12 (0.85-1.49)	0.424
rs874603	3	186953037	A/G	0.059	3' UTR	*MASP1*	0.00 (1.00)	1.10 (0.79-1.53)	0.582
rs850313	3	186953226	C/G	0.746	3' UTR	*MASP1*	0.01 (1.00)	1.17 (0.98-1.40)	0.083
rs78393224	3	186953244	T/A	0.976	3' UTR	*MASP1*	0.00 (1.00)	1.53 (0.90-2.60)	0.118

Supplemental Table 8. Continued

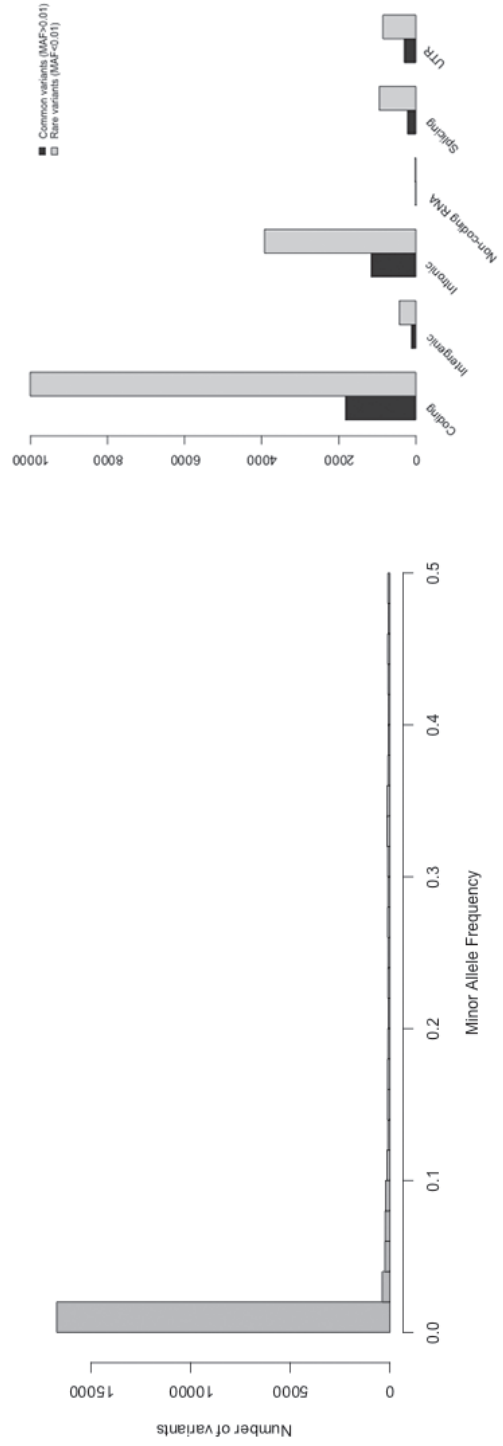
rsID	Chr.	Position	A1/A2	RAF	Functional class	Gene	*LD R^2 (D')	OR (95%CI)	P
rs67143992	3	186953321	T/C	0.816	3' UTR	*MASP1*	0.00 (1.00)	1.14 (0.93-1.40)	0.196
rs850312	3	186953808	C/T	0.364	synonymous	*MASP1*	0.00 (0.27)	1.05 (0.89-1.24)	0.548
rs72549154	3	186953932	C/A	0.028	missense	*MASP1*	0.00 (1.00)	1.04 (0.66-1.65)	0.861
rs3774268	3	186954324	G/A	0.133	synonymous	*MASP1*	0.01 (0.38)	1.04 (0.82-1.32)	0.775
rs698090	3	186964300	C/T	0.681	3' UTR	*MASP1*	0.01 (0.85)	1.01 (0.85-1.21)	0.877
rs16848736	3	186964312	T/C	0.023	3' UTR	*MASP1*	0.00 (1.00)	1.23 (0.71-2.12)	0.455
rs72549254	3	187009412	A/G	0.812	Splicing	*MASP1*	0.00 (0.62)	1.17 (0.93-1.46)	0.186

Chr. chromosome; A1 reference allele; A2 risk allele; RAF risk allele frequency; UTR untranslated region; LD linkage disequilibrium; OR odds ratio; CI confidence interval
* Linkage disequilibrium with rs72549167, based on data of European population of 1000 Genomes Project.

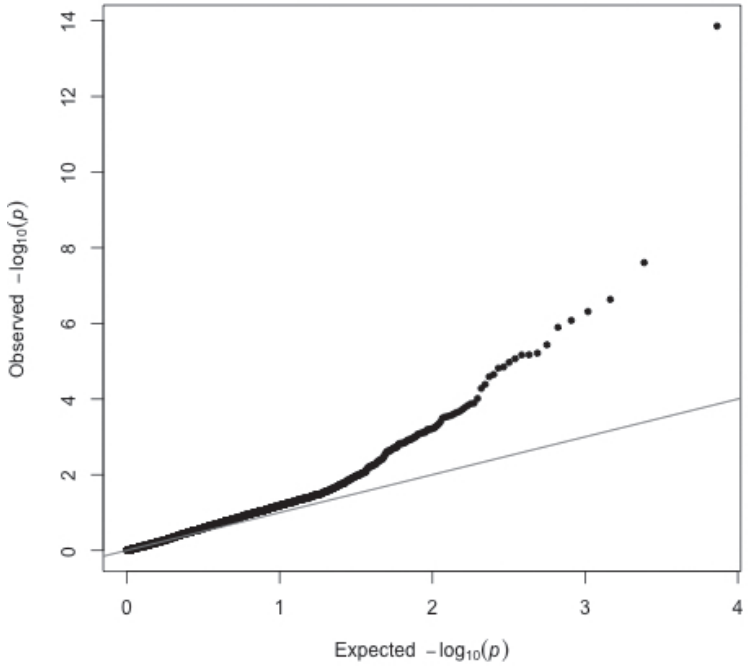
Supplemental Table 9. Single-variant association analyses of common CBS variants

rsID	Chr.	Position	A1/A2	RAF	Functional class	Gene	*LD R^2 (D')	OR (95%CI)	P
rs9978104	21	44473980	T/G	0.894	3' UTR	*CBS*	0.06 (1.00)	1.16 (0.88-1.52)	0.295
rs234706	21	44485350	A/G	0.640	synonymous	*CBS*	0.17 (0.79)	1.15 (0.97-1.36)	0.110

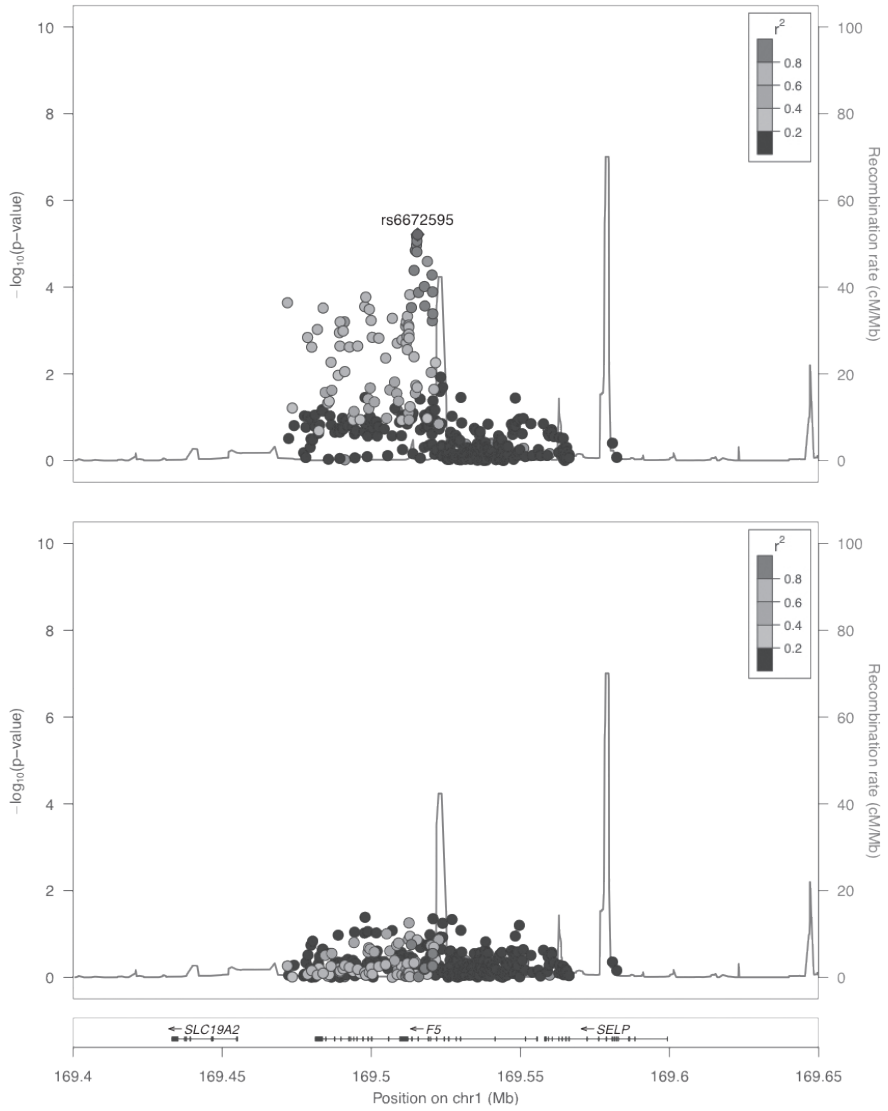
Chr. chromosome; A1 reference allele; A2 risk allele; RAF risk allele frequency; UTR untranslated region; LD linkage disequilibrium; OR odds ratio; CI confidence interval
* Linkage disequilibrium with rs1801181, based on data of European population of 1000 Genomes Population.



Supplemental Figure 1. Minor allele frequency distribution of identified variants (left) and overview of functional classes (right)



Supplemental Figure 2. Quantile-Quantile plot of the single variant association analyses



Supplemental Figure 3. Regional association plots for single variants in the *F5* region

Regional association plots showing single variant association results between common variants in the *F5* region and DVT risk before (upper panel) and after conditioning (lower panel) on lead variant rs6672595.