Genetic determinants of venous thrombosis

Haan, H.G. de

CHAPTER 3

Genome-wide association study identifies a novel genetic risk factor for recurrent venous thrombosis

de Haan HG, van Hylckama Vlieg A, Germain M, Baglin TP, Deleuze JF, Trégouët DA, Rosendaal FR.

Circ Genom Precis Med. 2018;11:e001827.

ABSTRACT

Background

Genetic risk factors for a first venous thrombosis (VT) seem to have little effect on recurrence risk. Therefore, we aimed specifically to identify novel genetic determinants of recurrent VT. So far, genome-wide association studies are lacking.

Methods and Results

We performed a genome-wide association scan in 1279 patients from the MEGA follow-up study; 832 patients with a first VT only and 447 recurrent VT patients. We analysed genotype probabilities of about 8.6 million variants, imputed to the Genome of the Netherlands project reference panel, with a minor allele frequency ≥ 1% for an association with recurrent VT. One region exceeded genome-wide significance (P-value ≤ $5x10^{-8}$), mapping to the well-known FV Leiden locus. Conditional association analyses on FV Leiden did not yield any secondary association signals. We also identified 52 suggestive association signals (P-value < $1x10^{-5}$) at 17 additional loci. None of these loci were previously implicated in VT risk. Replication analyses for 17 lead variants were performed in 350 recurrent VT patients and 1866 patients with a single VT event from the MEGA follow-up study, THE-VTE study, and LETS study. We observed an association with recurrence for an intergenic variant at 18q22.1 with an odds ratio of 1.7 (95% CI 1.2-2.6) per copy of the minor allele.

Conclusions

We confirmed the association of FV Leiden and identified a novel risk locus at 18q22.1 in the first large genetic study on recurrent VT.

INTRODUCTION

Approximately 20 to 30% of patients with a first venous thrombosis (VT) develop a recurrence within five years of the first event,[1,2] and therefore predicting and preventing recurrence is of crucial importance. However, risk factors for a first event do no predict recurrence well and hence risk profiling is difficult.[3-6] Recurrence risk is the highest amongst patients whose thrombotic event was not provoked by transient risk factors such as surgery and immobilization.[1,2,7-9] In particular, previous studies have shown that patients with a first unprovoked event have a two to three-fold increased risk of recurrence compared with patients with a first provoked event.[7-9] This suggests that patients with recurrent VT are enriched for genetic risk factors. The minor effects of determinants of first events on recurrence on the relative risk scale can be explained by the difference in absolute risks of first and recurrent VT, and index event bias.[10,11]

In addition, different genetic variants may play a role in recurrence than in first thrombosis, for example factors that affect clot lysis or the recanalization of the vein after a thrombotic event. So far, few studies have focused on recurrence-specific genetic risk factors. Zee and colleagues studied a panel of 86 variants in 56 candidate genes and observed suggestive associations with recurrent VT for four variants.[12] In addition, homozygosity of Ser128Arg in the E-selectin gene and length of a GT-dinucleotide repeat in the promoter of the gene encoding heme oxygenase 1 have been linked to recurrent VT in an Austrian study.[13,14] However, none of these findings have sofar been confirmed in large independent studies.

In order to identify novel genetic determinants of recurrent VT, we performed the first genome-wide association study (GWAS) on recurrence in 447 patients with recurrent VT and a sample of 832 patients who remained recurrence-free in the Multiple Environmental and Genetic Assessment of risk factors for venous thrombosis (MEGA) follow-up study.[15] To validate our findings, we additionally performed a replication study of the newly identified risk variants among 350 recurrent VT patients and a sample of 1866 patients with a single event only from three cohort studies.

MATERIAL AND METHODS

GWAS analysis

Study population

We included patients from the MEGA follow-up study, a large population-based cohort study on risk factors for recurrent VT. Details of this study have been described elsewhere.[15] In short, 4956 patients with a first deep vein thrombosis (DVT) of the leg or pulmonary embolism (PE), who were enrolled in the MEGA case-control study between 1999 and 2004[16], were invited to participate. Follow-up started at the date of the first event. Between 2008 and 2009, questionnaires related to recurrent VT were sent to the patients. Occurrence of recurrent VT was determined by information from patients, anticoagulation clinics, and treating physicians according to a decision rule.[15] Follow-up ended when a recurrent VT occurred, the patient died or migrated, or when the questionnaire was returned, whichever occurred first. For the patients who died, information on the cause of death was retrieved from the national registry of death certificates. If no questionnaire was returned, patients were considered lost to follow-up.

For the GWAS analysis, 1499 patients were selected according to the following process (Flow diagram is shown in Supplemental Figure 1). First, patients who had not provided a high-quality blood sample or buccal swap for DNA analysis were excluded (667 out of 4956 eligible patients). In addition, we excluded all patients who had been diagnosed with cancer (N=457). We then selected all patients for whom a recurrent VT event was reported at time of sample selection for the current analysis (N=542). Of these, 16 recurrences were classified as uncertain recurrences according to the decision rule,[15] and these patients were subsequently analyzed as recurrence-free patients. In addition, we randomly sampled 957 patients, totaling 973 patients who remained without a recurrent event during a median period of 7.1 years (interquartile range [IQR] 5.5-8.4). Follow-up was incomplete for 19.5% of these patients, as some died without recurrence (N=11), whereas others were last seen at the anticoagulation clinic (N=77) or at time of blood sampling for the MEGA case-control study (N=102). Patients with incomplete follow-up were followed for a median period of 312 days (range 60 days to 9.7 years). As these patients did not or no longer visit the anticoagulation clinic, which monitor anticoagulant treatment, it is unlikely that these patients suffered from a recurrent

VT and, therefore, these patients were considered as recurrence-free patients in the GWAS analysis. We performed a sensitivity analysis for the top GWAS findings in which we excluded patients with incomplete follow-up and patients who had an uncertain recurrent event.

This study was approved by the Medial Ethics Committee of the Leiden University Medical Center, and all participants gave written informed consent.

GWAS quality-control and imputation

Genome-wide genotyping was performed with the Illumina Human660-Quad v.1 BeadChip (Illumina Inc., San Diego, USA) at Centre National de Génotypage (Institut de Génomique, Evry, France). Genotyping was successfully completed for 1461 patients, of whom 1426 had a call rate of at least 98%. Additional exclusions at the individual level included discrepancy between self-reported and genotypic sex, abnormal level of autosomal heterozygosity (false discovery rate <1%), and ethnic outliers based on multidimensional scaling analysis of the identity-by-state matrix. Furthermore, 32 patients withdrew their consent for the MEGA follow-up study, leaving a total of 1279 patients for imputation and association analyses (447 patients with a recurrence during follow-up and 832 recurrence-free patients). The following exclusions were applied to identify a final set of 497,563 high-quality variants: minor allele frequency (MAF) below 1%, genotyping call rate below 98%, significant deviation from Hardy-Weinberg equilibrium (P-value $<1x10^{-6}$) in patients with a first event only. All quality-control procedures were performed with the R-package GenABEL.[17]

Following the conversion of the genomic positions from hg18 to hg19 using the UCSC Genome Browser LiftOver tool, imputation of 19.6 million autosomal variants was performed using IMPUTE2 software[18] according to the Genome of the Netherlands reference panel (GoNL release 4).[19] Prior to the association analyses, we excluded variants with a MAF below 1% or an imputation quality score I below 0.5.

Statistical analysis

Imputed genotypes of 8.6 million variants were tested for an association with recurrent VT using SNPTEST version 2[20] by means of logistic regression with the missing data likelihood score test, which takes the uncertainty of the imputed genotypes into account. All analyses were adjusted for age and sex. We assumed an additive mode of

inheritance. The level of genome-wide significance was set at P-value $<5x10^{-8}$, whereas the threshold for highly suggestive association signals was set at P-value $<1x10^{-5}$. In order to identify independent secondary association signals at a locus, we performed conditional analyses on the lead variant or the previously reported VT risk variant. In addition, we grouped associated variants in clumps based on linkage disequilibrium (LD) and genomic distance according to standard settings in PLINK.[21] Regional association plots were created with LocusZoom[22] and functional annotation of the variants was performed with AnnoVar.[23]

The quantile-quantile plot of the genome-wide test statistics against the expected null distribution showed no appreciable evidence of inflation due to population stratification or genotyping artefacts (Supplemental Figure 2). Likewise, the genomic inflation factor (lambda[24]) before and after imputation was 1.033 and 1.001, respectively. None of the first four principal components were associated with recurrent VT, and these were therefore not included as covariates in the association analyses.

Look-up of previously reported risk variants

In order to validate previously reported genetic associations with (recurrent) venous thrombosis that may not have attained genome-wide significance in our study, we specifically explored the association results for 17 variants. Selected variants were either previously shown to be associated with recurrence only[12-14,25] or reached genome-wide significance in one of the two recent GWAS studies on first VT.[26,27] Effects were calculated per copy of the risk allele based on the reporting in the original studies. Additional information on the selected variants is provided in Supplemental Table 1. Two variants (rs3025058 and rs3074372) could not be studied due to the absence of (tagging) variants in the GWAS, one variant (rs114209171) could not be studied as it was located on the X chromosome.

Replication analysis

Study population

The replication analysis was conducted in 350 patients with recurrent VT and a sample of 1866 patients with a single event only. These individuals were included from three European studies into VT risk, that is the MEGA follow-up study, the Leiden Thrombophilia Study (LETS) study[4], and the Thrombophilia, Hypercoagulability and

Environmental Risks in Venous Thromboembolism (THE-VTE) study[28]. From the MEGA follow-up study, we included 155 recurrent VT patients who had not been included in the original GWAS or who were excluded during the quality-control procedures of the GWAS. In addition, we randomly sampled 929 patients with a single VT event only, of whom 72.9% had complete follow-up.

LETS and THE-VTE study are both population-based case-control studies into risk factors for VT with subsequent follow-up of the VT patients. The study designs are similar to that of the MEGA study and have been described in detail previously.[4,28] In LETS, 474 consecutive patients with a first DVT in the leg or arm were recruited at three anticoagulation clinics in or near Leiden. Patients were subsequently followed for recurrence until 2000 using repeated questionnaires. Follow-up started 90 days after the date of the first event and ended at the date of recurrence, date of death, date of emigration, or the end of the study, whichever occurred first.[4] A total of 471 patients had a DNA sample available for genotyping. Of these, 90 patients developed a recurrence during a median follow-up of 8.0 years (IQR 6.8-9.0). Follow-up was complete for 88.2% of the recurrence-free patients. THE-VTE is a two-center case-control study, in which 796 consecutive patients with a first VT were enrolled in Leiden and Cambridge (UK).[28] Patients were subsequently followed for recurrence starting at the date of the first event. In Leiden, follow-up ended when a recurrent event occurred, when a patient died or migrated, or when patients were untraceable, whichever occurred first. For patients included in Cambridge, recurrence status was checked on 1 July 2013 using hospital records. In the absence of recurrence or death, this date was registered as the end of follow-up. For the current analysis, we excluded patients who did not have a DNA sample available (N=135). During a median follow-up of 5.4 years (IQR 4.2-6.6), 105 of the 661 patients experienced a recurrent VT event. Follow-up was complete for 88.5% of the patients with a single VT event only. In both LETS and THE-VTE, individuals with a recent cancer diagnosis were not enrolled.

All participants gave written informed consent. The THE-VTE and LETS study were both approved by the Medical Ethics Committee of the Leiden University Medical Center. In addition, THE-VTE was also approved by the NHS Research Ethics Committee in Cambridge, UK.

Genotyping

For each novel locus that showed a highly significant association with recurrent VT in the discovery GWAS, we selected the lead variant or the variant with the largest functional impact. These variants were genotyped with predesigned or custom-made TaqMan assays (Life Technologies, Thermo Fisher Scientific, USA) according to manufacturer's specifications. Primer design failed for three variants (rs9834479 in *ROBO1*, rs61504683 in *LPPR3*, and rs111750150 in *TSPEAR*), which were subsequently replaced by variants in high LD (r^2>0.8) in our GWAS study population or based on the CEU 1000 Genomes population using SNAP software[29].

Statistical analysis

Association with recurrent VT was assessed using logistic regression analyses adjusting for age, sex, study, and study center in case of THE-VTE. Patients who were lost to follow-up were analyzed as recurrence-free. These patients remained without a recurrent event during a median follow-up period of 1.2 years (IQR 0.7-3.4). To account for multiple hypothesis testing, the threshold for statistical significance was set at 0.05 divided by the number of variants tested in the replication analyses. We also calculated the false discover rate (FDR). In addition, we performed a sub-analysis including only the patients from LETS and THE-VTE in a Cox regression model to calculate hazard ratios with 95% confidence intervals (95% CI). In this analysis, patients who were lost to follow-up were censored at the last date known to be recurrence-free. To ensure comparability of follow-up time between the LETS and THE-VTE study in the Cox regression analysis, we recalculated the follow-up time in THE-VTE to start 90 days after the date of the first event.

For the variant that replicated, we performed a meta-analysis of the results obtained in the replication cohorts and in the original GWAS in order to obtain the most robust estimate of its effect size. For this, we used a fixed-effects model based on inverse-variance weighting as implemented in the METAL software.[30] Heterogeneity was assessed by the Cochran's Q statistic and the I^2 index.

Discriminative value

To explore the potential clinical value of the two identified and validated genetic risk loci, we assessed the discriminative accuracy of two prediction models: a clinical model

and a combined model to which we added dosages of two genetic variants (rs6025 and rs9946608). The clinical model included sex, age, event type (DVT only versus PE with or without a DVT), and provoking status (recent surgery, trauma, immobilization, hormone use, pregnancy, and travel). We fitted both models in the GWAS population, which had complete clinical information for 1260 individuals (443 recurrence patients and 817 patients with a first VT only). Areas under the receiver-operating characteristic curves (AUC) were constructed using the predicted risks derived from logistic regression models. We calculated and compared the AUCs of the two prediction models using DeLong's test for correlated ROC curves as implemented in R package "pROC".[31]

RESULTS

GWAS analysis

Population characteristics

After quality-control assessments, 447 patients with a recurrent VT and 832 patients with a single VT event were included in the genome-wide association analyses. Overall, these patients had been followed for a median period of 6.1 years (IQR 2.2-7.9). Seventeen percent of the recurrence-free patients did not complete follow-up, as some died without recurrence (N=9) or had an uncertain recurrent event (N=10), whereas others were last seen at the anticoagulation clinic (N=46) or at time of blood sampling for the MEGA case-control study (N=75). The mean age at time of the first event was 48.1 years (standard deviation [sd] 12.9) and 49% of the patients was a man. Sixty-one percent of the patients had a first DVT of the leg, whereas 29% had a PE and 10% of the patients were diagnosed with both. Compared with patients with a single VT event, patients who experienced a recurrence were more often men and had more often a first unprovoked event (Table 1).

Table 1. Characteristics of GWAS study population

	Patients with a first VT only N=832	**Patients with a recurrent VT** N=447
Age at first event, mean years (SD)	47.0 (12.8)	50.2 (12.7)
Male sex, N (%)	339 (40.7)	287 (64.2)
Body mass index, kg/m^2	26.8 (4.7)	27.1 (4.5)
Smoking, N (%)	297 (35.7)	144 (32.9)
First event was unprovoked*, N (%)	248 (29.8)	220 (49.2)
Duration of anticoagulant therapy, median days (IQR)	183 (110-213)	185 (111-212)
Type of first event:		
DVT, N (%)	497 (59.7)	283 (63.3)
PE, N (%)	265 (31.9)	102 (22.8)
DVT and PE, N (%)	70 (8.4)	61 (13.8)

VT venous thrombosis, DVT deep vein thrombosis, PE pulmonary embolism, SD standard deviation, IQR interquartile range

*Provoking factors: recent surgery, immobilization (plaster cast, bedridden at home, hospitalization), hormone use, pregnancy or post-partum, and travel.

Association analyses

We assessed the association between 8.6 million variants and recurrent VT. The Manhattan plot of the GWAS results is shown in Supplemental Figure 3. Nineteen variants, all mapping to the *F5* region, were associated with recurrent VT at genome-wide significance (Supplemental Table 2). The lead variant mapped to a non-coding sequence in *F5* (rs2213868, MAF 14%, P-value 2.67x10^{-9}). The *F5* locus also included the established VT-associated variant FV Leiden (rs6025, MAF 9.6%, P-value 1.28x10^{-8}), of which the T-allele was associated with a 2.4-fold increased risk of recurrent VT (95% CI 1.75-3.15). Conditional analyses on rs6025 did not reveal any secondary association signals at the locus (Supplemental Figure 4). Of the genome-wide significant variants, the lowest remaining P-value was 0.02 for rs2213868 (Supplemental Table 2).

We additionally identified 52 variants that showed suggestive evidence of an association (P-value <1.0x10^{-5}) with recurrent VT (Supplemental Table 3). Of these, nine variants

were part of the *F5* locus and were no longer associated with recurrent VT when conditioning on FV Leiden. The other 43 variants mapped to 17 loci, mainly at non-coding sequence. None of the variants or gene regions have previously been implicated in the risk of recurrent or a first VT. We did not identify independent association signals at any of these loci when conditioning on the lead variant of each locus (data not shown). The effect estimates of the lead variants did not materially change in a sensitivity analysis excluding patients who were lost to follow-up, although confidence intervals became wider due to the smaller sample size (Supplemental Table 4). Likewise, all lead variants remained associated with recurrence risk, with similar effect sizes, in a sensitivity analysis adjusting for provoking status (Supplemental Table 5).

Furthermore, we aimed to replicate previous genetic associations with recurrent VT and to explore associations for variants recently reported in GWAS analyses on first VT. Results are reported in Table 2. We assessed the association of eight variants that reached genome-wide significance in two recent GWAS studies. Besides the association with FV Leiden, we observed a nominal association with recurrent VT for *FGG* rs2066865 (OR 1.30, 95% CI 1.09-1.56) and *F5* rs4524 (OR 1.25, 95% CI 1.02-1.54). The recently identified risk variants in *SCL44A2* and *TSPAN15* showed no evidence of an association with the risk of recurrence (rs2288904, OR 1.14, 95% CI 0.90-1.44 and rs78707713, OR 1.14, 95% CI 0.85-1.54, respectively). In addition, five variants that have previously been linked to recurrent VT risk were not associated with recurrence in the present GWAS analysis (Table 2).

Table 2. GWAS look-up of genetic variants previously associated with first or recurrent venous thrombosis

Literature	rs ID	Chr.	Position	Gene	A1/A2	EAF	Info	OR (95% CI)	P-value
First VT	rs4524	1	169511755	*F5*	C/T	0.800	1.00	1.25 (1.02-1.54)	0.032
	rs6025	1	169519049	*F5*	C/T	0.096	0.94	2.35 (1.75-3.15)	$1.28x10^{-8}$
	rs2066865	4	155525276	*FGG*	G/A	0.339	1.00	1.30 (1.09-1.56)	0.003
	rs4253417	4	187199005	*F11*	T/C	0.483	0.95	1.17 (0.99-1.39)	0.068
	rs529565	9	136149500	*ABO*	T/C	0.447	1.00	1.19 (1.00-1.42)	0.055
	rs78707713	10	71245276	*TSPAN15*	C/T	0.915	0.98	1.14 (0.85-1.54)	0.381
	rs1799963	11	46761055	*F2*	G/A	0.021	0.77	1.25 (0.64-2.42)	0.516
	rs2288904	19	10742170	*SLC44A2*	A/G	0.839	1.00	1.14 (0.90-1.44)	0.265
	rs6087685	20	33777612	*PROCR*	G/C	0.381	0.98	1.05 (0.89-1.25)	0.543
Recurrence	rs5361	1	169701060	*SELE*	T/G	0.121	1.00	1.14 (0.89-1.47)	0.296
	rs1799864	3	46399208	*CCR5*	G/A	0.067	0.83	0.91 (0.63-1.32)	0.622
	rs805297	6	31622606	*APOM*	C/A	0.313	0.98	1.07 (0.90-1.28)	0.447
	rs662	7	94937446	*PON1*	T/C	0.298	1.00	0.93 (0.77-1.11)	0.405
	rs1800775	16	56995236	*CETP*	C/A	0.493	1.00	1.03 (0.87-1.22)	0.718

Chr. chromosome, A1 reference allele, A2 effect allele, EAF effect allele frequency, Info imputation quality info score, OR odds ratio, CI confidence interval, VT venous thrombosis

Effects were calculated per copy of the risk allele, as reported in the original study, and adjusted for age and sex assuming an additive mode of inheritance.

Table 3. Main findings of GWAS and replication study for lead variants at highly suggestive loci of GWAS

					GWAS				Replication		
rs ID	**Chr**	**Position**	**(Nearest) Gene**	**A1/A2**	**info**	**MAF**	**OR (95% CI)**	**P-value**	**MAF**	**OR (95% CI)**	**P-value**
rs112349920	1	159933483	*LINC01133*	C/T	0.98	0.011	6.90 (3.06-15.6)	$3.36x10^{-6}$	0.006	0.21 (0.03-1.55)	0.125
rs144482539	2	164295170	*FIGN*	A/G	0.60	0.015	8.11 (3.33-19.8)	$4.14x10^{-6}$	0.011	1.85 (0.92-3.71)	0.084
rs34029315	3	10571102	*ATP2B2*	A/G	0.95	0.081	2.16 (1.57-2.97)	$2.48x10^{-6}$	0.101	1.01 (0.76-1.34)	0.931
rs41499647	3	50525154	*CACNA2D2*	C/T	0.91	0.205	1.61 (1.31-1.99)	$8.90x10^{-6}$	0.204	0.81 (0.65-1.01)	0.063
rs6548639*	3	79687975	*ROBO1*	C/T	0.70	0.380	1.59 (1.29-1.96)	$1.39x10^{-5}$	0.545	1.03 (0.87-1.23)	0.699
rs114497105	5	13759735	*DNAH5*	C/T	0.70	0.014	6.91 (2.97-16.1)	$7.55x10^{-6}$	0.012	1.26 (0.60-2.62)	0.547
rs142454359	5	135637432	*TRPC7*	G/A	0.58	0.036	4.06 (2.28-7.25)	$2.07x10^{-6}$	0.000	∞ (0.00-∞)	1.000
rs79438589	5	158397065	*EBF1*	G/T	0.80	0.022	4.32 (2.30-8.11)	$5.21x10^{-6}$	0.018	0.93 (0.49-1.77)	0.820
rs78069640	6	8859837	*RP11-314C16.1*	C/T	0.63	0.014	7.69 (3.18-18.6)	$6.01x10^{-6}$	0.020	0.86 (0.45-1.64)	0.639
rs2334321†	6	110567409	*METTL24*	G/A	0.97	0.084	0.50 (0.37-0.68)	$7.86x10^{-6}$	0.077	0.82 (0.59-1.15)	0.256
rs142720518	9	39156170	*CNTNAP3*	T/C	0.69	0.074	0.41 (0.28-0.60)	$6.11x10^{-6}$	0.065	0.88 (0.61-1.28)	0.515
rs4766986	12	113076475	*PTPN11*	C/T	0.85	0.057	2.42 (1.64-3.58)	$8.75x10^{-6}$	0.054	0.74 (0.49-1.13)	0.159
rs9946608	18	65817281	*RP11-638L3.1*	T/C	0.94	0.037	2.91 (1.83-4.61)	$5.76x10^{-6}$	0.033	1.73 (1.16-2.59)	0.008
rs351995*	19	809732	*PTBP1*	C/A	0.60	0.449	0.62 (0.50-0.78)	$2.05x10^{-5}$	0.472	0.97 (0.82-1.14)	0.676
rs203551	20	1192766	*C20orf202*	T/G	0.94	0.237	1.61 (1.31-1.97)	$5.26x10^{-6}$	0.246	1.00 (0.83-1.21)	0.987
rs78571420	21	36377390	*RUNX1*	T/A	0.99	0.042	2.58 (1.70-3.93)	$9.88x10^{-6}$	0.041	0.79 (0.50-1.25)	0.311
rs117161628*	21	46138322	*TSPEAR*	C/T	NA	NA	NA	NA	0.000	∞ (0.00-∞)	1.000

Chr. chromosome, A1 major allele, A2 minor allele, info imputation quality info score, MAF minor allele frequency, OR odds ratio, CI confidence interval, NA not applicable

* Primer design failed for rs9834479 in *ROBO1*, rs61504683 in *LPPR3*, and rs111750150 in *TSPEAR* and each of these were replaced by tagging variants.

For rs111750150, a tagging variant was selected based on 1000G CEU population using SNAP software[28], as there was no tagging variant available in the GWAS.

† This variant was not the lead variant at this locus, but it was selected on its functional impact (LD with lead variant r^2 0.93)

GWAS: effects calculated per copy of the minor allele, adjusted for age and sex

Replication: effects calculated per copy of the minor allele using a logistic regression model, adjusted for age, sex, study and, country of study

Replication analyses

To eliminate false-positive findings, we next performed a replication study in 350 patients with recurrent VT and 1866 patients with a single event only from three population-based cohorts. Overall, patients were followed for recurrence for a median period of 6.1 years (IQR 3.8-7.8), albeit follow-up started at different moments in time (see Material and Methods). Follow-up was complete for 83.7 percent of the patients.

For each of the 17 loci, we genotyped either the lead variant or the variant with substantial functional impact, and tested these for an association with recurrent VT in the replication cohorts. Results of the replication analyses are presented in Table 3. For two variants, rs142454359 and rs117161628, we observed only one carrier and, therefore, these variants could not be studied in detail. We observed an association with recurrent VT for one variant, whereas the remaining variants showed no evidence of an association with recurrent VT. Variant rs9946608 is located in an intergenic region at 18q22.1 and was associated with a 1.7-fold (95% CI 1.16-2.59, P-value 0.008, FDR 0.136) increased recurrence risk per copy of the minor allele. Similarly, we observed a hazard ratio of 1.69 (95% CI 1.18-2.42) per copy of the minor allele of rs9946608 for recurrence risk in a sub-analysis of patients from the LETS and THE-VTE cohorts. When we meta-analyzed the results obtained in the replication cohorts and the discovery GWAS, the minor allele of rs9946608 was associated with a 2.2-fold increased recurrence risk (Table 4, 95% CI 1.62-2.98, P-value 4.83×10^{-7}). There was no evidence for heterogeneity across the three replication cohorts (Q-statistic 1.12, I^2 0.00, P-value 0.57), nor across the replication cohorts and the discovery GWAS (Q-statistic 3.66, I^2 18.1, P-value 0.30).

We subsequently interrogated several publicly available databases for potential mechanistic information on rs9946608. No significant expression quantitative trait loci have been reported in GTEx[32] for rs9946608 or any of the linked variants (r^2>0.8). We used RegulomeDB[33], which integrates information from the ENCODE[34] and Roadmap Epigenomic[35] projects, to assess whether rs9946608 or linked variants may have a regulatory function. There is minimal evidence that several variants at this locus, including rs9946608, may affect transcription factor binding affinity. In some cell lines, DNase peaks in the chromatin structure have been identified using DNase-sequencing. Genes located nearby, which could be potential target genes, are two long intergenic non-coding RNA (lincRNA) genes *(RPH11-526H11-1* and *RP11-638L3.1)* and protein-coding gene *TMX3*. The latter encodes thioredoxin-related transmembrane protein

3 (TMX3), which has been detected in human megakaryocytes, platelets, and at the platelet surface of both resting and stimulated platelets.[36]

Table 4. Association results of rs9946608 in three replication cohorts

	TT	TC	CC	MAF	OR (95% CI)
MEGA					
recurrent VT patients	132	10	0	0.035	1.43 (0.71-2.87)
first VT patients	850	43	1	0.025	reference
LETS					
recurrent VT patients	66	10	1	0.078	2.40 (1.17-4.90)
first VT patients	330	25	1	0.038	reference
THE-VTE					
recurrent VT patients	95	9	1	0.052	1.61 (0.78-3.29)
first VT patients	519	36	0	0.032	reference
Meta-analysis					1.76 (1.17-2.65)
Combined with GWAS					2.20 (1.62-2.98)

MAF minor allele frequency, OR odds ratio, CI confidence interval, VT venous thrombosis, GWAS genome-wide association study

Results were meta-analyzed 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. Across the three replication cohorts the heterogeneity measures were as follows: Q 1.12, I^2 0.00, P-value 0.57. For the three replication studies and the discovery GWAS, we observed a Q of 3.66, I^2 18.1, P-value 0.30. In the GWAS, the MAF of rs9946608 was 0.583 in recurrence patients and 0.256 in patients with a first VT only.

Discriminative value

In a preliminary analysis, we explored the added discriminative value of FV Leiden and rs9946608 to a prediction model with clinical risk factors alone. The AUC of the clinical prediction model, which included sex, age, event type, and provoking status, was 0.65 (95% CI 0.61-0.68). Predictive accuracy of recurrence risk significantly improved when adding the two genetic risk variants to the model (AUC 0.68, 95% CI 0.65-0.71).

DISCUSSION

This GWAS is the first large-scale genetic discovery effort for recurrent VT. Previous studies were either small or focussed on candidate gene variants, such as FV Leiden and prothrombin G20210A. The high recurrence rate of VT, especially in patients with a first unprovoked event, and the subsequent lifelong treatment with anticoagulants make it important to uncover the genetic and biological architecture of recurrent VT. Here, we confirm the association of FV Leiden with recurrence and identify a novel potential risk locus at chromosome 18q22.1.

Genome-wide significance was attained by several variants at the *F5* locus, which included the well-known risk variant FV Leiden. We observed a 2.4-fold increased risk of recurrence per copy of the T-allele of FV Leiden, which is slightly higher than previously reported,[3,4] albeit still lower than the risk estimates observed for a first VT.[26,27] There were no secondary association signals observed at the *F5* locus. Known VT risk variant rs4524, which has been shown to affect the risk of a first thrombotic event independent of FV Leiden,[26,37] was only nominally associated with recurrent VT. This may suggest that FV Leiden is the key determinant at the *F5* locus of recurrence risk.

We additionally identified 43 variants at 17 novel loci associated with recurrent VT at suggestive significance (P-value $<1.0\times10^{-5}$). We sought to replicate these findings in independent samples from three studies. Our results suggest that carriers of rs9946608-C have a 1.7-fold increased recurrence risk compared with non-carriers. We observed little evidence for statistical heterogeneity between the replication studies which could explain our findings. Formal replication is needed to confirm the association between rs9946608 and recurrent VT, as the meta-analysis of the GWAS and the replication studies did not reach genome-wide significance. From a clinical perspective, it would also be interesting to evaluate whether this variant has a differential effect on recurrent DVT or PE, which was now impossible to study due to low number of patients.

Variant rs9946608 and proxies map to noncoding sequence at chromosome 18q22.1 and have not been implicated in disease risk before. If the association with recurrence risk is true, this intergenic locus has most likely a regulatory function. We observed some evidence of transcription factor binding affinity and DNase peaks in the chromatin structure of some cell lines. Additional work, including fine-mapping of the GWAS signal to identify the functional variant, is needed to unravel the potential underlying

mechanism. Candidate genes could be nearby lincRNA genes *RPH11-526H11-1* and *RP11-638L3.*1. Increasing evidence suggests that lincRNAs may play an important role in epigenetic and post-transcriptional regulation in health and (cardiovascular) disease.[38,39] However, the characteristics and function of the majority of these RNAs are currently not known. Interrogation of several publicly available databases, such as GTEx[32] and several long noncoding RNA databases, did not yield additional information. The nearest protein-coding gene, *TMX3*, lies over 500Kb away, but could also be a target given its biological function. As TMX3 has been detected at the platelet surface,[36] it may play a role in platelet functioning, in line with other members of the protein disulphide isomerase family. Functional follow-up experiments could help to identify and characterize the potential role of these genes in recurrent VT. In addition, long-range chromatin interaction analyses using chromosome conformation capture technologies, such as 4C and Hi-C, might aid to identify other potential target genes.

Another notable finding is that almost all variants, which have previously been linked to a first VT at genome-wide significance[26,27] including the novel risk variants at *TSPAN15* and *SLC44A2*, were not or only nominally associated with the risk of recurrent VT. This is in line with previous reports on the risk variants which have been studied for recurrence risk.[3-6] Several explanations for this discrepancy have been proposed. To some extent, this can be explained by the difference in absolute risks for first and recurrent VT, resulting in the incomparability of effects on a relative risk scale between first and recurrent VT.[11] In addition, research into risk factors for recurrence risk may be hindered by index event bias, although this could lead to both under- and overestimation of the risk estimate.[10] Of note, as all candidate risk variants had effects in the expected direction and three out of nine variants were associated with recurrence risk at a significance level of 0.05, which is more than expected by chance, our results provide some evidence that these variants may also impact recurrence risk. In particular, *FGG* rs2066865 might be promising, as earlier studies have also reported some evidence of an association.[5,6]

The main limitation of this study is the small sample size with 447 and 345 recurrent VT patients in the discovery GWAS and the combined replication studies, respectively. As a result, we may have missed associations between recurrent VT and variants with a small effect or a low MAF. The small sample size may also explain why we failed to replicate most suggestively associated variants identified in our GWAS. We therefore emphasise

the need of a large international collaborative effort to substantially increase the sample size for recurrent VT analyses. Of note, mainly patients of Northwest-European origin were included in our analyses and, therefore, caution is needed in generalizing our results to other populations. In addition, the X chromosome was not interrogated in the discovery GWAS.

In both the GWAS and the replication analyses, patients who were lost to follow-up or who experienced an uncertain recurrent VT were considered to be recurrence-free. This could have affected our results, as we cannot rule out that these patients experienced a recurrent thrombotic event. However, this is unlikely, since these patients did not visit the anticoagulation clinics, which monitor anticoagulant treatment. In addition, the results of the sensitivity GWAS, in which these patients were excluded, did not materially differ from the discovery GWAS. Likewise, we obtained a similar effect estimate for rs9946608 in the logistic regression model and the time-to-event analysis, in which patients who were lost to follow-up were censored. Together, this suggests that the impact of misclassification in our study was probably low.

Our findings could lead to a better understanding of the biological mechanism underlying recurrent VT. In addition, we have previously shown the potential clinical value of genetic risk factors in the risk stratification of first and recurrent VT.[5,40] In a preliminary analysis, we showed that adding FV Leiden and rs9946608 to a clinical prediction model slightly improved the risk discrimination of recurrence. Identification of novel risk variants may further improve risk prediction of recurrent VT. Although additional replication and functional analyses are required, we identified a potential risk locus at chromosome 18q22.1 and confirmed the role of FV Leiden in recurrent VT pathophysiology.

REFERENCES

1. Prandoni P, Lensing AW, Cogo A, Cuppini S, Villalta S, Carta M, et al. The long-term clinical course of acute deep venous thrombosis. *Ann Intern Med*. 1996;125(1):1-7.
2. Hansson PO, Sörbo J, Eriksson H. Recurrent venous thromboembolism after deep vein thrombosis: incidence and risk factors. *Arch Intern Med*. 2000;160(6):769-74.
3. Ho WK, Hankey GJ, Quinlan DJ, Eikelboom JW. Risk of recurrent venous thromboembolism in patients with common thrombophilia: a systematic review. *Arch Intern Med*. 2006;166(7):729-36.
4. Marchiori A, Mosena L, Prins MH, Prandoni P. The risk of recurrent venous thromboembolism among heterozygous carriers of factor V Leiden or prothrombin G20210A mutation. A systematic review of prospective studies. *Haematologica*. 2007;92(8):1107-14.
5. van Hylckama Vlieg A, Flinterman LE, Bare LA, Cannegieter SC, Reitsma PH, Arellano AR, et al. Genetic variations associated with recurrent venous thrombosis. *Circ Cardiovasc Genet*. 2014;7(6):806-13.
6. Bruzelius M, Ljungqvist M, Bottai M, Bergendal A, Strawbridge RJ, Holmström M, et al. F11 is associated with recurrent VTE in women. A prospective cohort study. *Thromb Haemost*. 2016;115(2):406-14.
7. Baglin T, Luddington R, Brown K, Baglin C. Incidence of recurrent venous thromboembolism in relation to clinical and thrombophilic risk factors: prospective cohort study. *Lancet*. 2003;362(9383):523-6.
8. Christiansen SC, Cannegieter SC, Koster T, Vandenbroucke JP, Rosendaal FR. Thrombophilia, clinical factors, and recurrent venous thrombotic events. *JAMA*. 2005;293(19):2352-61.
9. Prandoni P, Noventa F, Ghirarduzzi A, Pengo V, Bernardi E, Pesavento R, et al. The risk of recurrent venous thromboembolism after discontinuing anticoagulation in patients with acute proximal deep vein thrombosis or pulmonary embolism. A prospective cohort study in 1,626 patients. *Haematologica*. 2007;92(2):199-205.
10. Dahabreh IJ, Kent DM. Index event bias as an explanation for the paradoxes of recurrence risk research. *JAMA*. 2011;305(8):822-3.
11. Cannegieter SC, van Hylckama Vlieg A. Venous thrombosis: understanding the paradoxes of recurrence. *J Thromb Haemost*. 2013;11 Suppl 1:161-9.
12. Zee RY, Bubes V, Shrivastava S, Ridker PM, Glynn RJ. Genetic risk factors in recurrent venous thromboembolism: A multilocus, population-based, prospective approach. *Clin Chim Acta*. 2009;402(1-2):189-92.
13. Jilma B, Kovar FM, Hron G, Endler G, Marsik CL, Eichinger S, et al. Homozygosity in the single nucleotide polymorphism Ser128Arg in the E-selectin gene associated with recurrent venous thromboembolism. *Arch Intern Med*. 2006;166(15):1655-9.

14. Mustafa S, Weltermann A, Fritsche R, Marsik C, Wagner O, Kyrle PA, et al. Genetic variation in heme oxygenase 1 (HMOX1) and the risk of recurrent venous thromboembolism. *J Vasc Surg*. 2008;47(3):566-70.
15. Timp JF, Lijfering WM, Flinterman LE, van Hylckama Vlieg A, le Cessie S, Rosendaal FR, et al. Predictive value of factor VIII levels for recurrent venous thrombosis: results from the MEGA follow-up study. *J Thromb Haemost*. 2015;13(10):1823-32.
16. Blom JW, Doggen CJ, Osanto S, Rosendaal FR. Malignancies, prothrombotic mutations, and the risk of venous thrombosis. *JAMA*. 2005;293(6):715-22.
17. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: a R library for genome-wide association analysis. *Bioinformatics*. 2007 23(10):1294-6.
18. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529.
19. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet*. 2014;46(8):818-25.
20. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39(7):906-13.
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-75.
22. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336-7.
23. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
24. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55(4):997-1004.
25. Ahmad A, Sundquist K, Zöller B, Dahlbäck B, Svensson PJ, Sundquist J, et al. Identification of polymorphisms in Apolipoprotein M gene and their relationship with risk of recurrent venous thromboembolism. *Thromb Haemost*. 2016;116(3):432-41.
26. Germain M, Chasman DI, de Haan H, Tang W, Lindström S, Weng LC, et al. Meta-analysis of 65,734 individuals identifies TSPAN15 and SLC44A2 as two susceptibility loci for venous thromboembolism. *Am J Hum Genet*. 2015 Apr 2;96(4):532-42.
27. Hinds DA, Buil A, Ziemek D, Martinez-Perez A, Malik R, Folkersen L, et al. 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(9):1867-74.
28. 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(9):1642-52.

29. Johnson AD, Handsaker RE, Pulit S, Nizzari MM, O'Donnell CJ, de Bakker P I. SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008;24(24):2938-9.
30. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190-1.
31. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
32. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648-60.
33. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-7.
34. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.
35. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010;28(10):1045-8.
36. Holbrook LM, Watkins NA, Simmonds AD, Jones CI, Ouwehand WH, Gibbins JM. Platelets release novel thiol isomerase enzymes which are recruited to the cell surface following activation. *Br J Haematol*. 2010;148(4):627-37.
37. Smith NL, Hindorff LA, Heckbert SR, Lemaitre RN, Marciante KD, Rice K, et al. Association of genetic variations with nonfatal venous thrombosis in postmenopausal women. *JAMA*. 2007;297: 489–98.
38. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet*. 2011;12(12):861-74.
39. Uchida S, Dimmeler S. Long noncoding RNAs in cardiovascular diseases. *Circ Res*. 2015;116(4):737-50.
40. de Haan HG, Bezemer ID, Doggen CJ, Le Cessie S, Reitsma PH, Arellano AR, et al. Multiple SNP testing improves risk prediction of first venous thrombosis. *Blood*. 2012;120(3):656-63.

SUPPLEMENTAL FIGURES AND TABLES

Supplemental Table 1. Previously reported associations with first or recurrent VT

Supplemental Table 2. GWAS associations with recurrent VT at genome-wide significance

Supplemental Table 3. GWAS associations with recurrent VT at significance threshold of $P<1x10^{-5}$

Supplemental Table 4. Association results for lead variants in sensitivity analysis excluding patients who were lost to follow-up

Supplemental Table 5. Association results for lead variants in sensitivity analysis adjusting on provoking status

Supplemental Figure 1. Flow diagram of patients included and excluded from GWAS analyses

Supplemental Figure 2. Quantile-quantile plot of the genome-wide test statistics

Supplemental Figure 3. Manhattan plot of the GWAS association results

Supplemental Figure 4. Regional association plots at the *F5* locus before and after conditioning on FV Leiden

Supplemental Table 1. Previously reported associations with first or recurrent VT

A. Previously reported genome wide significant associations with first VT

rs ID	Chr.	Position	Gene	Effect	Alleles	A2	Analysis	OR (95% CI)	P-value	Study
rs4524	1	169511755	*F5*	missense	T/C	T	additive	1.20 (1.14-1.26)	$2.65x10^{-11}$	Germain *et al.*
rs6025	1	169519049	*F5*	missense	C/T	T	additive	3.25 (2.91-3.64)	$1.10x10^{-96}$	Germain *et al.*
rs2066865	4	155525276	*FGG*	3'UTR	G/A	A	additive	1.24 (1.18-1.31)	$1.03x10^{-16}$	Germain *et al.*
rs4253417	4	187199005	*F11*	intronic	T/C	C	additive	1.27 (1.22-1.34)	$1.21x10^{-23}$	Germain *et al.*
rs529565	9	136149500	*ABO*	intronic	T/C	C	additive	1.55 (1.48-1.63)	$4.23x10^{-75}$	Germain *et al.*
rs78707713	10	71245276	*TSPAN15*	intronic	T/C	T	additive	1.28 (1.19-1.39)	$5.74x10^{-11}$	Germain *et al.*
rs1799963	11	46761055	*F2*	intronic	G/A	A	additive	2.29 (1.75-2.99)	$1.73x10^{-9}$	Germain *et al.*
rs2288904	19	10742170	*SLC44A2*	missense	G/A	G	additive	1.19 (1.12-1.26)	$1.07x10^{-9}$	Germain *et al.*
rs6087685	20	33777612	*PROCR*	intronic	G/C	C	additive	1.15 (1.10-1.21)	$1.65x10^{-8}$	Germain *et al.*
rs114209171	X	154278797	*FUNCD2**	intronic	T/C	T	additive	1.15 (1.11-1.20)	$7.0x10^{-13}$	Hinds *et al.*

Chr. Chromosome, A2 effect allele, OR odds ratio, CI confidence interval, UTR untranslated region *near *F8*

B. Previously reported associations with recurrent VT specifically

rs ID	Chr.	Position	Gene	Effect	Alleles	A2	Analysis	HR (95% CI)	P-value	Study
rs5361	1	169701060	*SELE*	missense	T/G	G	homozygous	4.1 (1.5-11.4)	0.01	Jilma *et al.*
rs1799864	3	46399208	*CCR2*	missense	G/A	A	additive	2.00 (1.15-3.48)	0.014	Zee *et al.*
rs805297	6	31622606	*APOM*	intronic	C/A	A	additive, men	1.72 (1.03-2.88)	0.038	Ahmad *et al.*
rs662	7	94937446	*PON1*	missense	T/C	C	additive	1.79 (1.08-2.95)	0.023	Zee *et al.*
rs3025058	11	102715948-102715949	*MMP3*	upstream	5A/6A	6A	additive	1.66 (1.10-2.49)	0.015	Zee *et al.*
rs1800775	16	56995236	*CETP*	upstream	C/A	A	additive	0.63 (0.40-0.98)	0.041	Zee *et al.*
rs3074372	22	35776887-35776888	*HMOX1*	5'UTR	GT-Repeat	Long	heterozygous	2.2 (1.4-3.4)	0.001	Mustafa *et al.*

Chr. Chromosome, A2 effect allele, HR hazard ratio, CI confidence interval, UTR untranslated region

Supplemental Table 2. GWAS associations with recurrent VT at genome-wide significance

rs ID	Chr	Position	Effect	(Nearest) Gene	A1/ A2	MAF overall	Info	OR (95%CI)	P-value	OR_{cond} (95%CI)	$P\text{-value}_{cond}$
rs1894692	1	169467654	intergenic	*SLC19A2*	A/G	0.094	0.93	2.38 (1.76-3.20)	$1.18x10^{-8}$	2.39 (0.39-14.7)	0.349
rs4264045	1	169470748	intergenic	*F5*	G/T	0.137	0.98	2.07 (1.62-2.65)	$6.38x10^{-9}$	1.57 (1.01-2.44)	0.045
rs6670848	1	169472899	intergenic	*F5*	G/A	0.137	0.98	2.07 (1.62-2.64)	$6.99x10^{-9}$	1.56 (1.00-2.42)	0.048
rs10737547	1	169476052	intergenic	*F5*	G/A	0.137	0.98	2.06 (1.61-2.63)	$7.56x10^{-9}$	1.55 (1.00-2.41)	0.050
rs6687813	1	169477574	intergenic	*F5*	C/A	0.137	0.98	2.06 (1.61-2.63)	$7.80x10^{-9}$	1.55 (1.00-2.40)	0.051
rs970740	1	169479974	intergenic	*F5*	T/C	0.137	0.98	2.06 (1.61-2.63)	$7.91x10^{-9}$	1.55 (1.00-2.40)	0.052
rs6427194	1	169481121	downstream	*F5*	A/T	0.137	0.98	2.06 (1.61-2.63)	$7.92x10^{-9}$	1.55 (1.00-2.40)	0.052
rs6427195	1	169481176	downstream	*F5*	T/A	0.137	0.98	2.06 (1.61-2.63)	$7.92x10^{-9}$	1.55 (1.00-2.40)	0.052
rs6427196	1	169481223	3' UTR	*F5*	G/C	0.137	0.98	2.06 (1.61-2.63)	$7.92x10^{-9}$	1.55 (1.00-2.40)	0.052
rs9332666	1	169486641	intronic	*F5*	C/G	0.137	0.98	2.05 (1.61-2.62)	$8.24x10^{-9}$	1.54 (0.99-2.39)	0.054
rs2420370	1	169490392	intronic	*F5*	C/G	0.137	0.99	2.05 (1.60-2.62)	$9.05x10^{-9}$	1.53 (0.99-2.37)	0.058
rs6682179	1	169490401	intronic	*F5*	C/T	0.137	0.99	2.05 (1.60-2.62)	$9.05x10^{-9}$	1.53 (0.99-2.37)	0.058
rs2420371	1	169491555	intronic	*F5*	A/G	0.138	0.99	2.03 (1.59-2.59)	$1.16x10^{-8}$	1.50 (0.97-2.32)	0.066
rs2420372	1	169498056	intronic	*F5*	G/A	0.138	0.99	2.02 (1.58-2.57)	$1.37x10^{-8}$	1.48 (0.96-2.29)	0.075
rs6009	1	169498834	intronic	*F5*	C/T	0.138	0.99	2.02 (1.58-2.57)	$1.37x10^{-8}$	1.48 (0.96-2.29)	0.075
rs6427197	1	169500590	intronic	*F5*	A/C	0.138	0.99	2.02 (1.58-2.57)	$1.43x10^{-8}$	1.48 (0.96-2.29)	0.077
rs1018827	1	169514006	intronic	*F5*	G/A	0.140	0.98	2.07 (1.62-2.64)	$4.34x10^{-9}$	1.63 (1.07-2.50)	0.024
rs6025	1	169519049	missense	*F5*	C/T	0.096	0.94	2.35 (1.75-3.15)	$1.28x10^{-8}$	NA	NA
rs2213868	1	169521553	intronic	*F5*	A/G	0.138	0.93	2.14 (1.67-2.75)	$2.67x10^{-9}$	1.71 (1.10-2.68)	0.018

Chr Chromosome, A1 major allele, A2 minor allele, MAF minor allele frequency, OR odds ratio, CI confidence interval, UTR untranslated region, NA not applicable, cond conditional, info imputation quality info measure

GWAS analyses were adjusted for age and sex, assuming an additive mode of inheritance. Subsequently, we performed a conditional analysis on FV Leiden (rs6025).

Supplemental Table 3. GWAS associations with recurrent VT at significance threshold of $P<1x10^{-5}$

rs ID	Chr.	Position	Effect	(Nearest) Gene	A1/A2	MAF overall	info	OR (95%CI)	P-value
rs111438240	1	159902335	synonymous	*IGSF9*	G/A	0.011	0.82	7.59 (3.18-18.1)	$5.07x10^{-6}$
rs112349920	1	159933483	intronic	*LINC01133*	C/T	0.011	0.98	6.90 (3.06-15.6)	$3.36x10^{-6}$
rs111272082	1	159940634	intronic	*LINC01133*	A/G	0.011	0.98	6.92 (3.06-15.6)	$3.37x10^{-6}$
rs145163454	1	169090748	intronic	*ATP1B1*	T/C	0.091	0.95	2.19 (1.63-2.95)	$2.38x10^{-7}$
rs77979353	1	169115022	intronic	*NME7*	T/C	0.119	0.96	1.81 (1.39-2.35)	$8.97x10^{-6}$
rs144737447	1	169160458	intronic	*NME7*	C/T	0.093	0.97	2.15 (1.60-2.87)	$2.72x10^{-7}$
rs2227246	1	169208179	intronic	*NME7*	T/C	0.091	0.96	2.16 (1.61-2.91)	$3.03x10^{-7}$
rs2040445	1	169216412	intronic	*NME7*	G/C	0.093	0.97	2.15 (1.61-2.87)	$2.58x10^{-7}$
rs6692824	1	169226218	intronic	*NME7*	G/C	0.094	0.96	2.16 (1.62-2.90)	$2.07x10^{-7}$
rs1209731	1	169324793	intronic	*NME7*	C/T	0.094	0.97	2.18 (1.63-2.92)	$1.42x10^{-7}$
rs2678166	1	169435027	3' UTR	*SLC19A2*	T/C	0.091	0.96	2.21 (1.64-2.96)	$1.43x10^{-7}$
rs6696217	1	169460726	intergenic	*SLC19A2*	G/A	0.147	0.97	1.80 (1.42-2.28)	$9.80x10^{-7}$
rs144482539	2	164295170	intergenic	*FIGN*	A/G	0.015	0.60	8.11 (3.33-19.8)	$4.14x10^{-6}$
rs34029315	3	10571102	intergenic	*ATP2B2*	A/G	0.081	0.95	2.16 (1.57-2.97)	$2.48x10^{-6}$
rs41499647	3	50525154	intronic	*CACNA2D2*	C/T	0.205	0.91	1.61 (1.31-1.99)	$8.90x10^{-6}$
rs9834479	3	79687062	intronic	*ROBO1*	T/A	0.411	0.73	1.61 (1.32-1.96)	$3.10x10^{-6}$
rs114497105	5	13759735	intronic	*DNAH5*	C/T	0.014	0.70	6.91 (2.97-16.1)	$7.55x10^{-6}$
rs142454359	5	135637432	intronic	*TRPC7*	G/A	0.036	0.58	4.06 (2.28-7.25)	$2.07x10^{-6}$
rs77962281	5	135710076	intergenic	*TRPC7*	T/C	0.036	0.59	3.87 (2.19-6.86)	$3.37x10^{-6}$
rs79438589	5	158397065	intronic	*EBF1*	G/T	0.022	0.80	4.32 (2.30-8.11)	$5.21x10^{-6}$

Supplemental Table 3. Continued

rs ID	Chr.	Position	Effect	(Nearest) Gene	A1/A2	MAF overall	info	OR (95%CI)	P-value
rs78069640	6	8859837	intergenic	*RP11-314C16.1*	C/T	0.014	0.63	7.69 (3.18-18.6)	$6.01x10^{-6}$
rs2334321	6	110567409	missense	*METTL24*	G/A	0.084	0.97	0.50 (0.37-0.68)	$7.86x10^{-6}$
rs72935918	6	110579775	intronic	*METTL24*	A/G	0.085	0.98	0.50 (0.37-0.68)	$7.07x10^{-6}$
rs72935927	6	110616710	intronic	*METTL24*	C/T	0.085	0.95	0.50 (0.37-0.68)	$9.86x10^{-6}$
rs10974147	9	39113963	intronic	*CNTNAP3*	C/A	0.073	0.67	0.41 (0.28-0.61)	$9.60x10^{-6}$
rs150155153	9	39130703	intronic	*CNTNAP3*	A/G	0.069	0.72	0.41 (0.28-0.61)	$9.74x10^{-6}$
rs1692979	9	39145910	intronic	*CNTNAP3*	T/C	0.071	0.70	0.41 (0.28-0.61)	$7.96x10^{-6}$
rs115460012	9	39146124	intronic	*CNTNAP3*	G/C	0.095	0.60	0.43 (0.30-0.63)	$9.22x10^{-6}$
rs142720518	9	39156170	intronic	*CNTNAP3*	T/C	0.074	0.69	0.41 (0.28-0.60)	$6.11x10^{-6}$
rs115361037	9	39158211	intronic	*CNTNAP3*	A/C	0.071	0.71	0.42 (0.28-0.61)	$9.29x10^{-6}$
rs233713	12	113049776	intergenic	*PTPN11*	G/A	0.057	0.85	2.43 (1.64-3.59)	$8.85x10^{-6}$
rs4766986	12	113076475	intergenic	*PTPN11*	C/T	0.057	0.97	2.42 (1.64-3.58)	$8.75x10^{-6}$
rs58823953	18	65794561	intergenic	*RP11-638L3.1*	A/G	0.036	0.97	2.87 (1.81-4.54)	$6.82x10^{-6}$
rs78684150	18	65797020	intergenic	*RP11-638L3.1*	A/G	0.036	0.97	2.87 (1.81-4.54)	$6.84x10^{-6}$
rs116330040	18	65799997	intergenic	*RP11-638L3.1*	T/C	0.036	0.97	2.87 (1.81-4.54)	$6.82x10^{-6}$
rs58636937	18	65801415	intergenic	*RP11-638L3.1*	C/T	0.036	0.97	2.87 (1.81-4.54)	$6.85x10^{-6}$
rs115198055	18	65804403	intergenic	*RP11-638L3.1*	A/C	0.036	0.97	2.87 (1.81-4.54)	$6.91x10^{-6}$
rs150366483	18	65804614	intergenic	*RP11-638L3.1*	G/T	0.036	0.97	2.87 (1.81-4.54)	$6.91x10^{-6}$
rs114858887	18	65809262	intergenic	*RP11-638L3.1*	A/G	0.036	0.97	2.87 (1.81-4.55)	$6.82x10^{-6}$
rs77684223	18	65810005	intergenic	*RP11-638L3.1*	T/A	0.036	0.97	2.87 (1.81-4.54)	$6.83x10^{-6}$

Supplemental Table 3. Continued

rs ID	Chr.	Position	Effect	(Nearest) Gene	A1/A2	MAF overall	info	OR (95%CI)	P-value
rs74443278	18	65810476	intergenic	*RP11-638L3.1*	T/C	0.036	0.97	2.87 (1.81-4.55)	$6.82x10^{-6}$
rs77238268	18	65813987	intergenic	*RP11-638L3.1*	C/G	0.036	0.96	2.88 (1.82-4.56)	$6.57x10^{-6}$
rs7228982	18	65814097	intergenic	*RP11-638L3.1*	A/G	0.036	0.96	2.88 (1.82-4.56)	$6.55x10^{-6}$
rs9946608	18	65817281	intergenic	*RP11-638L3.1*	T/C	0.037	0.94	2.91 (1.83-4.61)	$5.76x10^{-6}$
rs145772467	18	65825166	intergenic	*RP11-638L3.1*	T/C	0.037	0.88	2.99 (1.85-4.83)	$7.21x10^{-6}$
rs118036929	18	65825184	intergenic	*RP11-638L3.1*	T/C	0.037	0.88	2.99 (1.85-4.83)	$7.21x10^{-6}$
rs61504683	19	816919	intronic	*LPPR3*	C/T	0.468	0.58	0.60 (0.48-0.74)	$3.80x10^{-6}$
rs7269259	20	1179082	intergenic	*C20orf202*	C/T	0.237	1.00	1.57 (1.29-1.92)	$7.45x10^{-6}$
rs126622	20	1191140	intergenic	*C20orf202*	T/C	0.238	0.95	1.60 (1.30-1.96)	$5.86x10^{-6}$
rs203551	20	1192766	intergenic	*C20orf202*	T/G	0.237	0.94	1.61 (1.31-1.97)	$5.26x10^{-6}$
rs78571420	21	36377390	intronic	*RUNX1*	T/A	0.042	0.99	2.58 (1.70-3.93)	$9.88x10^{-6}$
rs111750150	21	45960880	intronic	*TSPEAR*	C/T	0.020	0.52	8.25 (3.56-19.1)	$8.46x10^{-7}$

Chr. Chromosome, A1 major allele, A2 minor allele, MAF minor allele frequency, OR odds ratio, CI confidence interval, UTR untranslated region
GWAS analyses were adjusted for age and sex, assuming an additive mode of inheritance. Variants in or near *ATP1B1*, *NME7*, or *SLC19A2* mapped to the F5 locus.

Supplemental Table 4. Association results for lead variants in sensitivity analysis excluding patients who were lost to follow-up

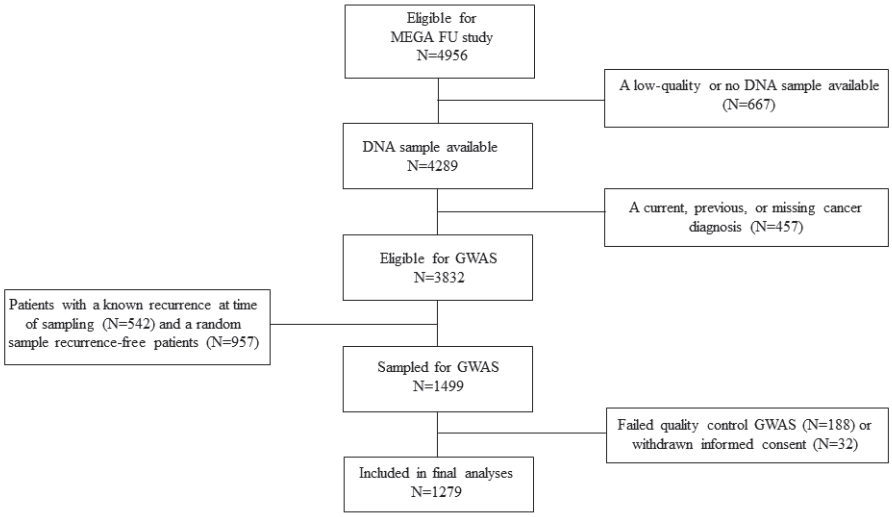
rs ID	Chr.	Position	Effect	(Nearest) Gene	A1/A2	MAF overall	OR (95% CI)	P-value
rs112349920	1	159933483	intronic	*LINC01133*	C/T	0.011	6.99 (3.05-16.0)	$4.47x10^{-6}$
rs2213868	1	169521553	intronic	*F5*	A/G	0.143	2.05 (1.59-2.66)	$4.07x10^{-8}$
rs144482539	2	164295170	intergenic	*FIGN*	A/G	0.016	6.74 (2.79-16.2)	$2.17x10^{-5}$
rs34029315	3	10571102	intergenic	*ATP2B2*	A/G	0.084	2.10 (1.52-2.92)	$7.66x10^{-6}$
rs41499647	3	50525154	intronic	*CACNA2D2*	C/T	0.209	1.59 (1.28-1.97)	$3.05x10^{-5}$
rs9834479	3	79687062	intronic	*ROBO1*	T/A	0.416	1.59 (1.29-1.95)	$9.82x10^{-6}$
rs114497105	5	13759735	intronic	*DNAH5*	C/T	0.015	5.39 (2.35-12.4)	$7.04x10^{-5}$
rs142454359	5	135637432	intronic	*TRPC7*	G/A	0.037	3.71 (2.07-6.62)	$9.65x10^{-6}$
rs79438589	5	158397065	intronic	*EBF1*	G/T	0.022	4.94 (2.57-9.49)	$1.70x10^{-6}$
rs78069640	6	8859837	intergenic	*RP11-314C16.1*	C/T	0.015	6.65 (2.75-16.1)	$2.70x10^{-5}$
rs72935918	6	110579775	intronic	*METTL24*	A/G	0.079	0.53 (0.38-0.73)	$1.04x10^{-4}$
rs142720518	9	39156170	intronic	*CNTNAP3*	T/C	0.073	0.37 (0.25-0.56)	$1.84x10^{-6}$
rs4766986	12	113076475	intergenic	*PTPN11*	C/T	0.060	2.20 (1.49-3.27)	$8.51x10^{-5}$
rs9946608	18	65817281	intergenic	*RP11-638L3.1*	T/C	0.038	2.78 (1.74-4.44)	$1.97x10^{-5}$
rs61504683	19	816919	intronic	*LPPR3*	C/T	0.464	0.60 (0.48-0.75)	$1.08x10^{-5}$
rs203551	20	1192766	intergenic	*C20orf202*	T/G	0.239	1.62 (1.31-2.00)	$8.27x10^{-6}$
rs78571420	21	36377390	intronic	*RUNX1*	T/A	0.043	2.62 (1.70-4.04)	$1.21x10^{-5}$
rs111750150	21	45960880	intronic	*TSPEAR*	C/T	0.020	9.09 (3.84-21.5)	$5.00x10^{-7}$

Chr. Chromosome, A1 major allele, A2 minor allele, MAF minor allele frequency, OR odds ratio, CI confidence interval
GWAS analyses were adjusted for age and sex, assuming an additive mode of inheritance.

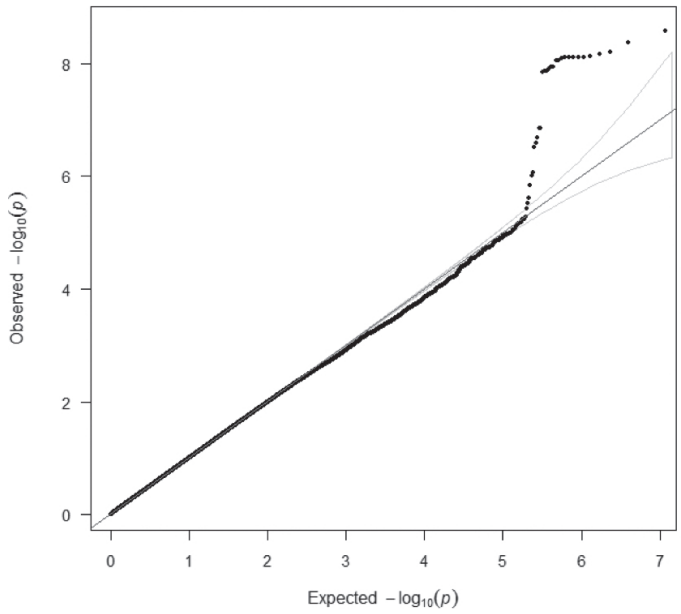
Supplemental Table 5. Association results for lead variants in sensitivity analysis adjusting on provoking status

rs ID	Chr.	Position	Effect	(Nearest) Gene	A1/A2	MAF overall	OR (95% CI)	P-value
rs112349920	1	159933483	intronic	*LINC01133*	C/T	0.011	6.59 (2.87-15.1)	$8.85x10^{-6}$
rs2213868	1	169521553	intronic	*F5*	A/G	0.139	2.14 (1.66-2.76)	$3.67x10^{-9}$
rs144482539	2	164295170	intergenic	*FIGN*	A/G	0.015	8.72 (3.56-21.4)	$2.20x10^{-6}$
rs34029315	3	10571102	intergenic	*ATP2B2*	A/G	0.081	2.24 (1.62-3.10)	$1.05x10^{-6}$
rs41499647	3	50525154	intronic	*CACNA2D2*	C/T	0.206	1.58 (1.28-1.96)	$2.24x10^{-5}$
rs9834479	3	79687062	intronic	*ROBO1*	T/A	0.413	1.58 (1.29-1.93)	$9.49x10^{-6}$
rs114497105	5	13759735	intronic	*DNAH5*	C/T	0.014	7.62 (3.22-18.1)	$3.85x10^{-6}$
rs142454359	5	135637432	intronic	*TRPC7*	G/A	0.036	4.05 (2.27-7.24)	$2.25x10^{-6}$
rs79438589	5	158397065	intronic	*EBF1*	G/T	0.022	4.48 (2.37-8.46)	$3.93x10^{-6}$
rs78069640	6	8859837	intergenic	*RP11-314C16.1*	C/T	0.014	7.17 (2.94-17.5)	$1.54x10^{-5}$
rs72935918	6	110579775	intronic	*METTL24*	A/G	0.084	0.49 (0.36-0.67)	$6.13x10^{-6}$
rs142720518	9	39156170	intronic	*CNTNAP3*	T/C	0.074	0.40 (0.27-0.58)	$3.25x10^{-6}$
rs4766986	12	113076475	intergenic	*PTPN11*	C/T	0.058	2.44 (1.65-3.61)	$8.85x10^{-6}$
rs9946608	18	65817281	intergenic	*RP11-638L3.1*	T/C	0.037	2.93 (1.84-4.65)	$5.32x10^{-6}$
rs61504683	19	816919	intronic	*LPPR3*	C/T	0.469	0.59 (0.47-0.77)	$2.92x10^{-6}$
rs203551	20	1192766	intergenic	*C20orf202*	T/G	0.236	1.58 (1.28-1.94)	$1.43x10^{-5}$
rs78571420	21	36377390	intronic	*RUNX1*	T/A	0.043	2.47 (1.62-3.78)	$2.80x10^{-5}$
rs111750150	21	45960880	intronic	*TSPEAR*	C/T	0.019	8.70 (3.64-20.8)	$1.14x10^{-6}$

Chr. Chromosome, A1 major allele, A2 minor allele, MAF minor allele frequency, OR odds ratio, CI confidence interval
GWAS analyses were adjusted for age, sex, and provoking status, assuming an additive mode of inheritance.
Provoking status at time of the first event was defined by the following factors: recent surgery, immobilization (plaster cast, bedridden at home, hospitalization), hormone use, pregnancy or post-partum, and travel.

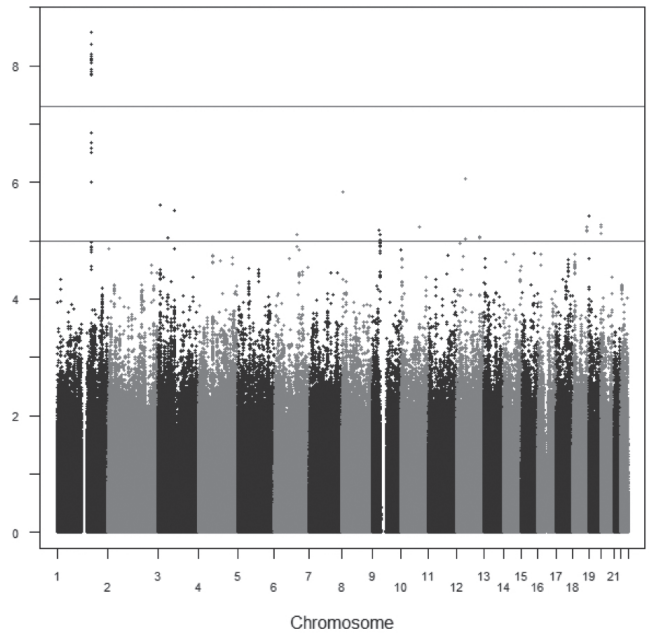


Supplemental Figure 1. Flow diagram of patients included and excluded from GWAS analyses



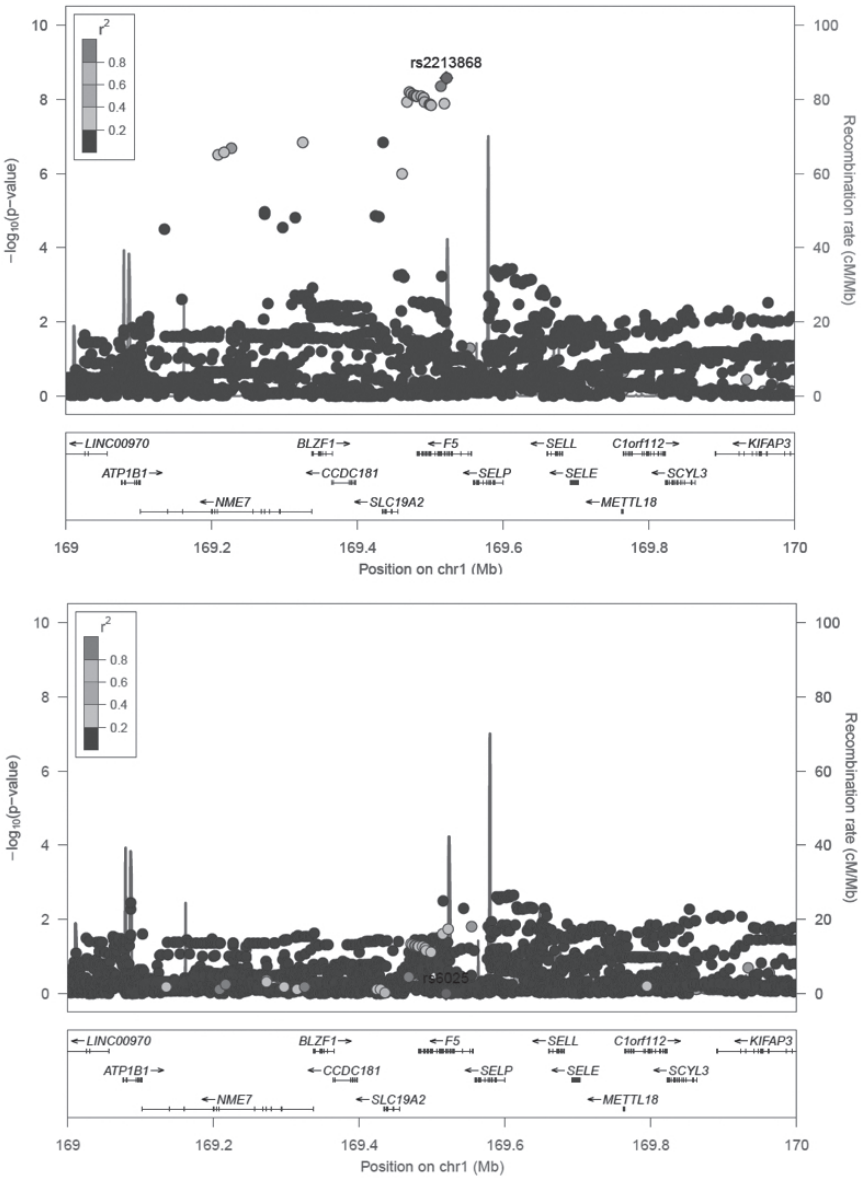
Supplemental Figure 2. Quantile-quantile plot of the genome-wide test statistics

The test statistics of the GWAS are plotted against the expected null distribution. Results are shown as $-\log_{10}$(P-values).



Supplemental Figure 3. Manhattan plot of the GWAS association results

Manhattan plot of $-\log_{10}$(P-values) for the associations between genotyped and imputed variants with recurrent venous thrombosis. We used logistic regression models to calculate the effects per copy of the minor allele, adjusted for age and sex. A total of 8.6 million autosomal variants were tested for an association with recurrent VT. The upper horizontal line at $5x10^{-8}$ represents the genome-wide significance threshold, whereas the lower line at $1x10^{-5}$ indicates the highly suggestive threshold.



Supplemental Figure 4. Regional association plots at the *F5* locus before and after conditioning on FV Leiden.

Results are shown as $-\log_{10}$(P-values) for both genotyped and imputed variants. The most associated variant in the discovery GWAS is shown as a triangle (rs2213868, upper panel). The colors of the other variants reflect the extent of linkage disequilibrium with the lead variant. The lower panel shows the association plot for recurrent VT after conditioning on the well-known FV Leiden variant (rs6025). The plots were generated using LocusZoom software.