Genetic variation and susceptibility to venous thrombosis : Etiology and risk assessment

Bezemer, I.D.

Gene Variants Associated With Deep Vein Thrombosis

Chapter 5

ID Bezemer
LA Bare
CJM Doggen
AR Arellano
C Tong
CM Rowland
J Catanese
BA Young
PH Reitsma
JJ Devlin
FR Rosendaal

JAMA. 2008;299(11):1306-1314

ABSTRACT

Context The genetic causes of deep vein thrombosis (DVT) are not fully understood.

Objective To identify single-nucleotide polymorphisms (SNPs) associated with DVT.

Design, Setting, and Patients We used 3 case-control studies of first DVT. A total of 19 682 gene-centric SNPs were genotyped in 443 cases and 453 controls from the Leiden Thrombophilia Study (LETS, 1988-1992). Twelve hundred six SNPs associated with DVT were reinvestigated in the Multiple Environmental and Genetic Assessment of Risk Factors for Venous Thrombosis study (MEGA-1, 1999-2004) in a subset of 1398 cases and 1757 controls. Nine SNPs associated with DVT in both LETS and MEGA-1 were investigated a third time in 1314 cases and 2877 controls from MEGA-2, a second subset of MEGA. Additional SNPs close to one SNP in *CYP4V2* were genotyped in LETS and MEGA-1.

Main Outcome Measure Odds ratios (ORs) for DVT were estimated by logistic regression. False discovery rates served to investigate the effect of multiple hypothesis testing.

Results Of 9 SNPs genotyped in MEGA-2, 3 were strongly associated with DVT ($P < .05$; false discovery rate ≤.10): rs13146272 in *CYP4V2* (risk allele frequency, 0.64), rs2227589 in *SERPINC1* (risk allele frequency, 0.10), and rs1613662 in *GP6* (risk allele frequency, 0.84). The OR for DVT per risk allele was 1.24 (95% confidence interval [95%CI], 1.11-1.37) for rs13146272, 1.29 (95% CI, 1.10-1.49) for rs2227589, and 1.15 (95% CI, 1.01-1.30) for rs1613662. In the region of *CYP4V2*, we identified 4 additional SNPs (in *CYP4V2*, *KLKB1*, and *F11*) that were also associated with both DVT (highest OR per risk allele, 1.39; 95% CI, 1.11-1.74) and coagulation factor XI level (highest increase per risk allele, 8%; 95% CI, 5%-11%).

Conclusions We identified SNPs in several genes that were associated with DVT. We also found SNPs in the region around the SNP in *CYP4V2* (rs13146272) that were associated with both DVT and factor XI levels. These results show that common genetic variation plays an important role in determining thrombotic risk.

INTRODUCTION

The incidence of deep vein thrombosis (DVT) is 1 per 1000 person-years [1]. The 10-year recurrence risk is 30% [158]. Deep vein thrombosis can lead to life-threatening pulmonary embolism [159]. Deep vein thrombosis is caused by acquired and genetic risk factors. Acquired risk factors include age, hospitalization, cancer, pregnancy, hormone therapy, and surgery [158]. Family and twin studies indicate that genetics accounts for about 60% of the risk for DVT [160,161]. Deficiencies of natural anticoagulants antithrombin, protein C, and protein S are strong risk factors for DVT; however, the variants causing these deficiencies are rare and explain only about 1% of all DVTs [6]. Two more common genetic variants, Factor V Leiden (FVL) and prothrombin G20210A, have been consistently found to be associated with DVT [51,162] but still only explain a fraction of the DVT events [6]. It has been suggested that 2 or more risk factors are needed for thrombosis [6,36,163].

The identification of additional common gene variants associated with DVT will improve the ability to predict risk for DVT and increase understanding of this disease. Therefore, we investigated whether any of 19 682 primarily missense single-nucleotide polymorphisms (SNPs) were associated with DVT in 3 large case-control studies.

METHODS

Study Populations and Data Collection

The 3 studies (LETS, MEGA-1 and MEGA-2) in the present analysis are derived from 2 large population-based case-control studies: the Leiden Thrombophilia Study (LETS) [17] and the Multiple Environmental and Genetic Assessment of Risk Factors for Venous Thrombosis (MEGA study) [19]. These studies were approved by the Medical Ethics Committee of the Leiden University Medical Center, Leiden, the Netherlands. All participants gave oral informed consent for LETS and written for MEGA to participate.

LETS Population

Collection and ascertainment of DVT events in LETS has been described previously.[11] Briefly, 474 consecutive patients, 70 years or younger, without a known malignancy were recruited between January 1, 1988, and December 30, 1992, from 3 anticoagulation clinics in the Netherlands. For each patient, an age- and sex-matched control participant without a history of DVT was enrolled. Participants completed a questionnaire on risk factors for DVT and provided a blood sample. No ethnicity information was collected. After exclusion of 52 participants due to inadequate sample, 443 cases and 453 controls remained in the analyses.

MEGA-1 and MEGA-2 Studies

Collection and ascertainment of DVT events in MEGA has been described previously [19,20]. MEGA enrolled consecutive patients aged 18 to 70 years who presented with their first diagnosis of DVT or pulmonary embolism (PE) at any of 6 anticoagulation clinics in the Netherlands between March 1, 1999, and May 31, 2004. Control subjects included partners of patients and random population control subjects frequency-matched on age and sex to the patient group. Participants completed a questionnaire on risk factors for DVT and provided a blood or buccal swab sample. The questionnaire included an item on parent birth country as a proxy for ethnicity.

For the present analyses, we split the MEGA study to form 2 case-control studies, based on recruitment date and sample availability (blood or buccal swab). We excluded those with isolated pulmonary embolism or a history of malignant disorders to obtain a study population similar to that of the LETS population. The first subset, MEGA-1, included 1398 cases and 1757 controls who all donated a blood sample. The remaining 1314 cases and 2877 controls who donated either a blood sample or a buccal swab sample were included in MEGA-2.

SNP Association Study

The 19 682 SNPs tested in this study are located in 10 887 genes and were

selected because of their potential to affect gene function or expression [164]. Most SNPs (69%) are missense. Another 24% of the SNPs are located in transcription factor binding sites or in untranslated regions of mRNA, which could affect messenger RNA expression or stability. Ninety-one percent of the SNPs studied have minor allele frequencies of at least 5% in whites. Information on all SNPs tested and primer sequences are available on request.

The design of the SNP association study is presented in the Figure. First, all 19 682 SNPs were tested in pooled DNA samples of LETS (http://www.ncbi.nlm.nih.gov/projects/SNP). Single-nucleotide polymorphisms that were associated with DVT ($P \leq .05$) were tested in pooled DNA samples of MEGA-1. Single-nucleotide polymorphisms that were associated in both LETS and MEGA-1 pools ($P \leq .05$) were confirmed by genotyping individual samples of LETS and MEGA-1. Single-nucleotide polymorphism genotypes consistently associated with DVT in LETS and MEGA-1 ($P \leq .05$) were genotyped in MEGA-2.

Allele Frequency and Genotype Determination

DNA concentrations were standardized to 10 ng/µL using PicoGreen (Molecular Probes, Invitrogen Corp, Carlsbad, California) fluorescent dye. DNA pools, typically of 30 to 100 samples, were assembled based on case-control status, sex, age, and factor V Leiden status. DNA pools were made by mixing equal volumes of standardized DNA solution from each individual sample. Each allele was amplified separately by polymerase chain reaction (PCR) using 3 ng of pooled DNA. In the pooled stage, we used 6 case pools and 4 control pools for LETS, and 13 case pools and 18 control pools for MEGA-1. Allele frequencies in pooled DNA were determined by kinetic polymerase chain reaction (kPCR) [165]. Duplicate kPCR assays were run for each allele and the amplification curves from these assays were used to calculate the allele frequencies of the SNP [165]. Genotyping of individual DNA samples was similarly performed using 0.3 ng of DNA in kPCR assays [165] or using multiplexed oligo ligation assays [166]. Genotyping accuracy of the multiplex method and kPCR has been assessed in

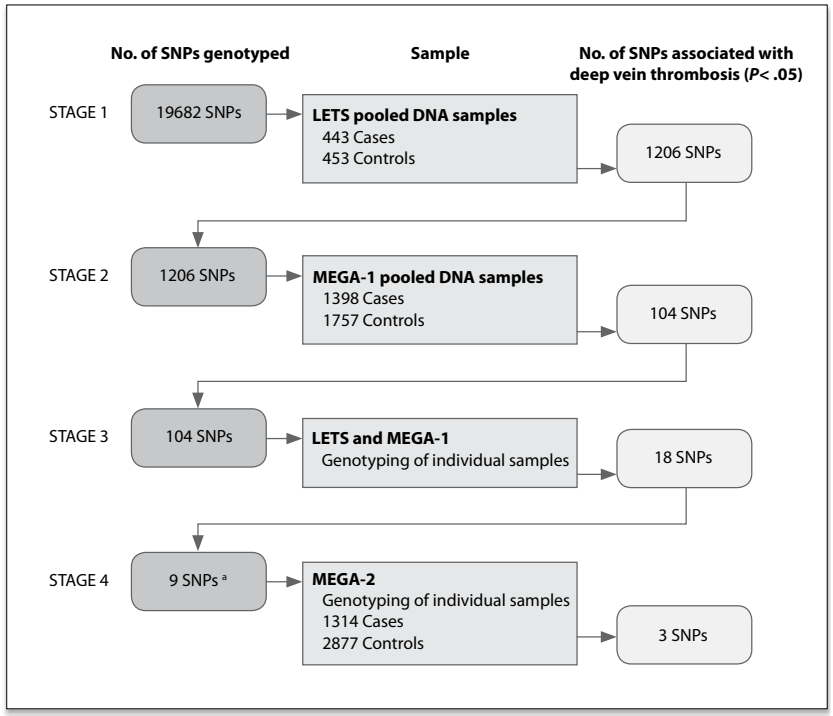

Figure. Flowchart of the Approach Used to Identify SNPs Associated With Deep Vein Thrombosis

[a] Only 9 SNPs were subsequently tested in MEGA-2 because assays for the other 9 were not available.

3 previous studies, and the overall concordance of the genotype calls from these 2 methods was greater than 99% [164,167,168]. The SNPs associated with DVT in MEGA-2 were successfully genotyped in more than 95% of the participants in LETS, MEGA-1, and MEGA-2.

Gene Variants and DVT Risk in the *CYP4V2* Region

The rs13146272 SNP in the gene *CYP4V2* was most strongly associated with DVT in the SNP association study. To investigate whether other SNPs in this region are associated with DVT, we used results from the HapMap Project [169] to identify a region surrounding rs13146272 (chromosome 4:187,297,249-

187,467,731). This region contained 149 SNPs with allele frequencies of more than 2% (HapMap NCBI build 36). Allele frequencies and linkage disequilibrium were calculated from the SNP genotypes in the HapMap Centre d'Etude du Polymorphisme Humain (CEPH) population, which includes Utah residents with ancestry from northern and western Europe. We selected 48 of these 149 SNPs for genotyping, as surrogates for 142 of the 149 SNPs in this region that were either directly genotyped or in strong linkage disequilibrium (r^2>0.8) with at least 1 of the 48 genotyped SNPs (the remaining 7 of the 149 SNPs were in low-linkage disequilibrium with rs13146272 (r^2<0.2) and therefore not likely to be the cause of the observed association). The 48 SNPs were chosen using pairwise tagging in Tagger (implemented in Haploview [170]). The 48 SNPs were initially investigated in LETS, and SNPs that were equally or more strongly associated with DVT than rs13146272 were investigated in MEGA-1.

Factor XI Assays

Factor XI antigen measurements in LETS were described previously [171]. In MEGA, factor XI levels were measured on a STA-R coagulation analyzer (Diagnostica Stago, Asnières, France). STA calcium chloride solution was used as an activator, STA Unicalibrator was used as a reference standard, and Preciclot plus I (normal factor XI range) was used as control plasma. The intraassay coefficient of variation was 5.8% (10 assays). The interassay coefficient of variation was 8.7% (48 assays).

Statistical Analysis

Deviations from Hardy-Weinberg expectations were assessed using an exact test in controls [172]. For pooled DNA analysis, a Fisher exact test was used to evaluate allele frequency differences between cases and controls. For the final set of SNPs, logistic regression models were used to calculate the odds ratio (OR), 95% confidence interval (95% CI), and 2-sided *P* value for the association of each SNP with DVT and to adjust for age and sex. For each SNP, we calculated the OR per genotype relative to noncarriers of the risk allele, and the risk allele OR from an additive model. This risk allele OR

can be interpreted as the risk increase per copy of the risk allele, and the corresponding *P* value was used to decide whether the SNP was associated with DVT ($P \leq .05$). For SNPs on the X chromosome the analysis was conducted separately in men and women.

The OR (95% CI) for SNPs in the *CYP4V2* region was estimated by logistic regression with adjustment for factor XI levels and other SNPs in the region. Differences in factor XI level between groups were tested with *t* tests, and changes in factor XI level per allele were estimated by linear regression.
Analyses were done using SAS version 9 (SAS Institute Inc, Cary, North Carolina) and SPSS for Windows, 14.0.2 (SPSS Inc, Chicago, Illinois).

False Discovery Rate

Studies of thousands of SNPs can lead to false-positive associations. Therefore, we performed 2 replications after the initial discovery stage in LETS and calculated the false discovery rate for the SNPs genotyped in MEGA-2. The false discovery rate estimates the expected fraction of false positives among a group of SNPs; and is a function of the *P* values and the number of tests [173]. False discovery rates were estimated using the 2-sided, unadjusted *P* value from the additive model. We used a false discovery rate of 0.10 as a criterion for further analysis (for a false discovery rate of 0.10, one would expect 10% of the SNPs in the group considered associated to be false positives).

RESULTS

Baseline characteristics of the participants are presented in Table 1.

SNPs Associated With DVT in LETS and MEGA-1

In LETS, we investigated 19 682 SNPs by comparing the allele frequencies of patients and controls using pooled DNA samples [165]. We found that 1206 of these 19 682 SNPs were associated ($P \leq .05$) with DVT. These 1206 SNPs were then investigated in patients and controls from MEGA-1 using pooled

Table 1. Characteristics of Cases and Controls in LETS, MEGA-1, and MEGA-2

	LETS		MEGA-1		MEGA-2	
	Cases (n=443)	Controls (n=453)	Cases (n=1398)	Controls (n=1757)	Cases (n=1314)	Controls (n=2877)
Men, No. (%)	190 (43)	192 (42)	652 (47)	843 (48)	633 (48)	1348 (47)
Age, Mean (SD)	45 (14)	45 (14)	47 (13)	48 (12)	48 (13)	47 (12)
Both parents born in North-West Europe, N(%)[a]	-	-	1247 (91)	1609 (92)	1149 (90)	2527 (89)

[a] No information on birth country was collected in LETS.

DNA samples. The SNPs that were associated with DVT in both LETS and MEGA-1 were confirmed by genotyping in both studies, and we found that 18 SNPs were consistently (with the same risk allele) associated with DVT ($P \leq .05$) in both LETS and MEGA-1 (Table 2).

SNPs Associated with DVT in MEGA-2

Nine of these 18 SNPs were subsequently tested in MEGA-2 for association with DVT (Table 3); assays for the other 9 SNPs were not available at the time. The genotypes of these 9 SNPs did not deviate from the Hardy-Weinberg equilibrium ($P \leq .01$) in the LETS and MEGA controls.

To account for the many tests, we estimated the false discovery rate for the SNPs tested in MEGA-2. In Table 2, factor V Leiden and the prothrombin G20210A mutation are presented for reference. Because these variants were not included in the SNP association study, we did not calculate their false discovery rate. For the SNP in *F9* (rs6048), we only included men because in women, no association with DVT was observed in LETS and MEGA-1. We found that 3 SNPs were again associated with DVT in MEGA-2 ($P \leq .05$), with false discovery rates ≤ .10. These 3 SNPs were in the genes *CYP4V2, SERPINC1*, and *GP6*. The 4 SNPs with the next lowest *P* values (ranging from .06-.15) also had low false discovery rates (≤ .20). These SNPs were in the genes *RGS7, NR1I2, NAT8B*, and *F9*. The risk allele frequencies for

Table 2. Association of 18 SNPs From the SNP Association Study and Factor V Leiden and Prothrombin G20210A With Deep Vein Thrombosis in the LETS and MEGA-1 Studiesa

Chr	Gene	SNP ID	SNP Type[b]	Study	Risk Allele	No. (%) of Alleles Cases	Controls	OR (95% CI)[c]	P-Value
3	NR1I2	rs1523127	5′UTR	LETS	C	373 (42)	300 (33)	1.44 (1.19-1.73)	<.001
				MEGA-1		1185 (42)	1373 (39)	1.15 (1.04-1.27)	.008
19	GP6	rs1613662	Ser219Pro	LETS	A	749 (85)	725 (80)	1.36 (1.07-1.74)	.01
				MEGA-1		2318 (84)	2823 (81)	1.21 (1.06-1.38)	.004
17	APOH	rs1801690	Ser335Trp	LETS	C	850 (97)	852 (95)	1.65 (1.02-2.68)	.04
				MEGA-1		2676 (96)	3312 (94)	1.42 (1.12-1.79)	.004
2	NAT8B	rs2001490	Ala112Gly	LETS	C	382 (43)	348 (38)	1.23 (1.01-1.49)	.04
				MEGA-1		1118 (40)	1301 (37)	1.14 (1.03-1.26)	.01
1	SERPINC1	rs2227589	Intronic	LETS	T	105 (12)	78 (9)	1.42 (1.04-1.94)	.03
				MEGA-1		303 (11)	313 (9)	1.24 (1.05-1.47)	.01
7	MET	rs2237712	Intronic	LETS	G	45 (5)	27 (3)	1.68 (1.05-2.70)	.03
				MEGA-1		119 (4)	110 (3)	1.38 (1.06-1.80)	.02
11	EPS8L2	rs3087546	Leu101Leu	LETS	T	522 (60)	487 (54)	1.26 (1.04-1.52)	.02
				MEGA-1		1637 (59)	1964 (56)	1.12 (1.01-1.24)	.03
6	CASP8AP2	rs369328	Lys93Lys	LETS	A	461 (52)	406 (45)	1.35 (1.11-1.63)	.002
				MEGA-1		1420 (51)	1680 (48)	1.13(1.02-1.24)	.02
1	SELP	rs6131	Asn331Ser	LETS	T	196 (22)	161 (18)	1.29 (1.03-1.62)	.03
				MEGA-1		589 (21)	636 (18)	1.21 (1.06-1.36)	.003
19	ZNF544	rs6510130	Asp203His	LETS	G	33 (4)	13 (1)	2.54 (1.34-4.83)	.004
				MEGA-1		78 (3)	64 (2)	1.56 (1.11-2.18)	.01
1	RGS7	rs670659	Intronic	LETS	C	617 (70)	584 (64)	1.27 (1.04-1.54)	.02
				MEGA-1		1864 (67)	2249 (64)	1.13 (1.01-1.25)	.03
2	TACR1	rs881	3′UTR	LETS	C	745 (85)	713 (80)	1.38 (1.07-1.77)	.01
				MEGA-1		2356 (85)	2894 (83)	1.15 (1.01-1.32)	.04
4	CYP4V2	rs13146272	Lys259Gln	LETS	A	611 (69)	588 (65)	1.22 (1.00-1.49)	.05
				MEGA-1		1896 (68)	2245 (64)	1.19 (1.07-1.32)	.001
1	F5	rs4524	Arg858Lys	LETS	T	708 (80)	671 (74)	1.36 (1.09-1.69)	.006
				MEGA-1		2184 (79)	2608 (74)	1.26 (1.12-1.42)	<.001
1	SMOYKEEBO/F5	rs6016	Ile736Ile	LETS	G	704 (80)	668 (74)	1.35 (1.09-1.69)	.006
				MEGA-1		2188 (79)	2615 (75)	1.27 (1.13-1.43)	< .001
1	C1orf114	rs3820059	Ser172Phe	LETS	A	320 (36)	269 (30)	1.34 (1.10-1.64)	.004
				MEGA-1		1065 (38)	1169 (33)	1.22 (1.10-1.35)	<.001
X	F9	rs6048	Ala194Thr	LETS Men	A	146 (77)	128 (67)	1.74 (1.10-2.74)	.02
				LETS Women		225 (70)	238 (68)	1.09 (0.84-1.42)	.50
				MEGA-1 Men		464 (73)	566 (68)	1.26 (1.00-1.58)	.05
				MEGA-1 Women		674 (72)	818 (71)	1.04 (0.90-1.21)	.61
X	ODZ1	rs2266911	Intronic	LETS Men	C	161 (85)	147 (77)	1.66 (0.99-2.79)	.06
				LETS Women		422 (83)	418 (80)	1.26 (0.91-1.73)	.17
				MEGA-1 Men		556 (85)	671 (80)	1.47 (1.12-1.94)	.006
				MEGA-1 Women		1234 (82)	1430 (78)	1.25 (1.05-1.48)	.01
1	F5 (Leiden)	rs6025	Arg534Gln	LETS	A	95 (11)	14 (2)	7.19 (4.05-(12.77)	<.001
				MEGA-1		291 (10)	96 (3)	4.10 (3.23-5.21)	<.001
11	F2 (G20210A)	rs1799963	3′UTR	LETS	A	28 (3)	10 (1)	2.98 (1.43-6.20)	<.001
				MEGA-1		81 (3)	37 (1)	2.89 (1.94-4.29)	<.001

[a] All gene symbols, rs numbers, SNP types and chromosome numbers are from NCBI build 36.

[b] The first amino acid corresponds to the non risk allele.

[c] ORs were estimated by logistic regression using an additive model. Sex was included as a covariate in logistic regression models containing markers residing on the X chromosome and the number of risk alleles for these SNPs were coded as 0 or 1 for males and 0, 1 or 2 for females.

Table 3. Associations of SNPs From the SNP Association Study With Deep Vein Thrombosis in MEGA-2[a]

				No. (%) of genotypes					
Chromo-some	Gene	SNP	Risk Allele[b]	Genotype[c]	Cased	Controld	OR (95% CI)	P value	FDR[e]
4	CYP4V2	rs13146272	A	CC	121 (10)	352 (13)	1 [Reference]		
				CA	478 (41)	1178 (45)	1.18 (0.94-1.49)		
				AA	561 (48)	1094 (42)	1.49 (1.19-1.88)		
				Additive	(69)	(64)	1.24 (1.11-1.37)	<.001	<.001
1	SERPINC1	rs2227589	T	CC	1001 (77)	2325 (82)	1 [Reference]		
				CT	278 (21)	483 (17)	1.34 (1.13-1.58)		
				TT	15 (1)	28 (1)	1.24 (0.66-2.34)		
				Additive	(12)	(11)	1.29 (1.10-1.49)	<.001	0.004
19	GP6	rs1613662	A	GG	29 (2)	89 (3)	1 [Reference]		
				GA	355 (27)	835 (29)	1.31 (0.84-2.02)		
				AA	915 (70)	1924 (68)	1.46 (0.95-2.24)		
				Additive	(84)	(82)	1.15 (1.01-1.30)	0.03	0.10
1	RGS7	rs670659	C	TT	129 (10)	355 (13)	1 [Reference]		
				TC	615 (48)	1326 (47)	1.28 (1.02-1.60)		
				CC	548 (42)	1153 (41)	1.31 (1.04-1.64)		
				Additive	(66)	(64)	1.10 (1.00-1.22)	0.06	0.13
3	NR1I2	rs1523127	C	AA	480 (37)	1097 (39)	1 [Reference]		
				AC	598 (46)	1340 (47)	1.02(0.88-1.18)		
				CC	220 (17)	409 (14)	1.23 (1.01-1.50)		
				Additive	(40)	(38)	1.09 (0.99-1.20)	0.07	0.13
2	NAT8B	rs2001490	C	GG	490 (38)	1122 (39)	1 [Reference]		
				GC	603 (46)	1334 (47)	1.04 (0.90-1.19)		
				CC	205 (16)	394 (14)	1.19 (0.98-1.45)		
				Additive	(39)	(37)	1.08 (0.98-1.19)	0.12	0.18
X	F9 (men)	rs6048	A	Additive	(73)	(70)	1.17 (0.94-1.45)	0.15	0.20
X	F9 (women)	rs6048	A	GG	56 (8)	148 (10)	1 [Reference]		
				GA	275 (41)	615 (41)	1.18 (0.84-1.66)		
				AA	343 (51)	752 (50)	1.21 (0.86-1.68)		
				Additive	(71)	(70)	1.07 (0.93-1.23)	0.37	NA
19	ZNF544	rs6510130	G	CC	1192 (95)	2626 (95)	1 [Reference]		
				CG	60 (5)	137 (5)	0.97 (0.71-1.32)		
				GG	0 (0)	4 (0)	-		
				Additive	(2)	(3)	0.91 (0.67-1.24)	0.56	0.63
11	MET	rs2237712	G	AA	1183 (93)	2528 (93)	1 [Reference]		
				AG	86 (7)	184 (7)	1.00 (0.77-1.30)		
				GG	3 (0)	3 (0)	2.14 (0.43-10.6)		
				Additive	(4)	(4)	1.03 (0.80-1.33)	0.79	0.79
1	F2	rs1799963	A	GG	1219 (94)	2794 (98)	1 [Reference]		
				GA	76 (6)	55 (2)	3.17 (2.22-4.51)		
				AA	0 (0)	0 (0)	-		
				Additive	(3)	(1)	3.17 (2.22-4.51)	<.001	NA
1	F5	rs6025	A	GG	1029 (81)	2646 (95)	1 [Reference]		
				GA	235 (18)	140 (5)	4.32 (3.46-5.39)		
				AA	8 (0)	2 (0)	10.30 (2.18-48.52)		
				Additive	(10)	(3)	4.24 (3.42-5.26)	<.001	NA

Abbreviations: NA, not applicable, not in FDR analysis.

[a] All gene symbols, rs numbers, SNP types and chromosome numbers are from NCBI build 36.

[b] Risk increasing allele identified in LETS and MEGA-1.

[c] In the additive model, the increase in risk per copy of the risk allele is calculated

[d] For the additive model, only the allele frequency is presented, not the count.

[e] P value from the additive model was used for FDR estimation. Factor V Leiden and the prothrombin 20210A mutation are presented for reference. Because these variants were not included in the SNP association study, we did not calculate their FDR. F9 FDR was calculated for men only

these 7 SNPs ranged from 11% to 82% among the controls. The OR for homozygous carriers, compared with homozygotes of the other allele, ranged from 1.19 to 1.49. The 2 SNPs most strongly associated with DVT were in *CYP4V2* (rs13146272, $P < .001$, false discovery rate 0.0006) and *SERPINC1* (rs2227589, $P < .001$, false discovery rate, 0.004).

For the 2 SNPs on chromosome 1 (rs2227589 and rs670659), we investigated linkage disequilibrium with FVL. The SNP (rs2227589) in *SERPINC1*, which encodes antithrombin, is 4.37 megabases away from the FVL variant. The SNP in *RGS7* (rs670659) is 71.48 megabases from FVL. Each was in weak linkage disequilibrium with FVL (r^2<.01). Restricting analyses to noncarriers of FVL did not appreciably change the risk estimate of either SNP (data not shown).

SNPs in *CYP4V2* Region and DVT Risk

The SNP with the strongest association with DVT was rs13146272, located in the gene encoding a member of the cytochrome P450 family 4 (*CYP4V2*). We genotyped 48 SNPs in this region in the LETS population (Table 4) and estimated the OR for DVT per copy of the risk-increasing allele. For many of the 48 SNPs, including rs13146272, the common allele was the risk allele. In LETS, rs13146272 had an OR for DVT of 1.22 (95% CI, 1.00-1.49). Higher ORs were observed for 9 of the other SNPs tested in this region. These SNPs were located in the *CYP4V2*, *KLKB1* (coding for prekallikrein), and *F11* (coding for coagulation factor XI) genes.

We then selected the 9 of the 48 SNPs that had an OR of more than 1.22 (the OR of rs13146272) and investigated them in MEGA-1. We found that, in addition to rs13146272, four of these SNPs were associated with DVT in both LETS and MEGA-1: rs3087505, rs3756008, rs2036914, and rs4253418 (Table 5). The rs3087505 SNP in *KLKB1* had the highest risk estimate: OR 3.61 (95% CI, 1.48-8.82) for the major allele homozygotes vs minor allele homozygotes. Mutual adjustment among these 5 SNPs did not indicate that any of these 5 associations were explained by the other 4 SNPs (data not shown).

Table 4. 48 SNPs in *CYP4V2* Region Genotyped in the Leiden Thrombophilia Study[a]

rs number	Gene[b]	SNP Type[c]	Risk allele (%) Case	Control	OR (95% CI)[d]	*P*
rs7686244	-	intergenic	39	37	1.11 (0.91 - 1.34)	*0.30*
rs4862650	*DKFZP564J102*	Gly41Lys	11	10	1.10 (0.82 - 1.49)	*0.52*
rs4862653	*DKFZP564J102*	Gly146Lys	11	10	1.09 (0.80 - 1.48)	*0.58*
rs2276922	*DKFZP564J102*	Pro241Leu	27	25	1.10 (0.89 - 1.36)	*0.38*
rs2276921	*DKFZP564J102*	intronic	49	48	1.04 (0.87 - 1.26)	*0.65*
rs2276920	*DKFZP564J102*	intronic	21	19	1.09 (0.86 - 1.38)	*0.47*
rs1877321	*DKFZP564J102*	intronic	79	76	1.19 (0.95 - 1.48)	*0.13*
rs2276919	*DKFZP564J102*	intronic	74	73	1.04 (0.85 - 1.28)	*0.70*
rs13141433	*DKFZP564J102*	intronic	87	85	1.12 (0.87 - 1.45)	*0.37*
rs11733307	*DKFZP564J102*	intronic	46	43	1.13 (0.94 - 1.35)	*0.21*
rs2241818	*DKFZP564J102*	intronic	21	19	1.16 (0.92 - 1.46)	*0.22*
rs6552959	*DKFZP564J102*	intronic	35	33	1.08 (0.89 - 1.30)	*0.46*
rs10017419	-	intergenic	42	41	1.04 (0.86 - 1.24)	*0.70*
rs7676755	*CYP4V2*	intronic	19	18	1.08 (0.84 - 1.38)	*0.55*
rs13146272	*CYP4V2*	Gln259Lys	69	65	1.22(1.00 - 1.49)	*0.05*
rs7687961	*CYP4V2*	intronic	83	81	1.19 (0.93 - 1.51)	*0.16*
rs3817184	*CYP4V2*	splice site	47	44	1.15 (0.96 - 1.39)	*0.13*
rs3736456	*CYP4V2*	Cys282Cys	96	94	1.49 (0.96 - 2.32)	*0.08*
rs2276917	*CYP4V2*	intronic	64	63	1.07 (0.89 - 1.30)	*0.46*
rs3733402	*KLKB1*	Ser143Asn	56	55	1.04 (0.87 - 1.25)	*0.67*
rs4253259	*KLKB1*	intronic	95	94	1.27 (0.84 - 1.92)	*0.25*
rs4253260	*KLKB1*	intronic	85	84	1.14 (0.88 - 1.47)	*0.31*
rs4253301	*KLKB1*	Ala381Ser	89	88	1.15 (0.86 - 1.55)	*0.35*
rs2292423	*KLKB1*	intronic	47	43	1.16 (0.97 - 1.40)	*0.11*
rs3775302	*KLKB1*	intronic	89	86	1.24 (0.95 - 1.64)	*0.12*
rs4253325	*KLKB1*	Gln560Arg	92	89	1.37 (1.01 - 1.86)	*0.04*
rs925453	*KLKB1*	Asn587Asn	71	71	1.00 (0.81 - 1.23)	*0.99*
rs3087505	*KLKB1*	3'UTR	92	90	1.26 (0.91 - 1.76)	*0.17*
rs3822055	*KLKB1*	3'near gene	20	18	1.14 (0.89 - 1.46)	*0.30*
rs6844764	-	intergenic	60	56	1.16 (0.96 - 1.40)	*0.13*
rs13135645	-	intergenic	86	83	1.22 (0.94 - 1.58)	*0.14*
rs3756008	*F11*	5'near gene	47	42	1.22 (1.02 - 1.46)	*0.03*

rs number	Gene[b]	SNP Type[c]	Risk allele (%) Case	Control	OR (95% CI)[d]	*P*
rs3822056	*F11*	5'near gene	90	89	1.07 (0.79 - 1.46)	*0.65*
rs3733403	*F11*	5'near gene	90	89	1.10 (0.81 - 1.50)	*0.53*
rs2036914	*F11*	intronic	60	54	1.25 (1.04 - 1.51)	*0.02*
rs4253408	*F11*	intronic	10	7	1.43 (1.01 - 2.01)	*0.04*
rs1593	*F11*	intronic	90	88	1.16 (0.85 - 1.57)	*0.36*
rs4253414	*F11*	intronic	3	3	1.07 (0.61 - 1.87)	*0.82*
rs4253418	*F11*	intronic	96	95	1.42 (0.88 - 2.27)	*0.15*
rs5974	*F11*	Thr267Thr	87	86	1.01 (0.77 - 1.32)	*0.93*
rs4253423	*F11*	intronic	85	83	1.13 (0.88 - 1.46)	*0.33*
rs5971	*F11*	Arg604Arg	96	96	1.01 (0.64 - 1.59)	*0.97*
rs4253430	*F11*	3'near gene	67	65	1.13 (0.92 - 1.37)	*0.24*
rs11938564	-	intergenic	81	79	1.14 (0.91 - 1.43)	*0.27*
rs13136269	-	intergenic	76	73	1.18 (0.95 - 1.46)	*0.13*
rs10025152	-	intergenic	85	85	1.01 (0.78 - 1.32)	*0.93*
rs12500826	-	intergenic	67	64	1.12 (0.93 - 1.36)	*0.24*
rs13133050	-	intergenic	71	68	1.13 (0.93 - 1.38)	*0.22*

[a] Gene symbols, rs numbers, SNP types, and chromosome numbers are from National Center for Biotechnology Information build 36.

[b] Some SNPs were located between genes, indicated in the "SNP type" column as "intergenic."

[c] The first amino acid corresponds to the nonrisk allele.

[d] Odds ratios were estimated by logistic regression using an additive model.

SNPs in *CYP4V2* Region and Factor XI Levels

Because the *F11* gene is located close to rs13146272 and because factor XI levels have been previously reported to be associated with DVT in the LETS population [171], we investigated whether an association between SNPs and factor XI levels explained the association between the SNPs and DVT. In LETS, factor XI levels above the 90th percentile had been shown to be associated with a 2-fold increased risk of DVT [171]. We found that high factor XI levels (>90th percentile) were also associated with DVT in MEGA (OR, 1.9; 95% CI, 1.6-2.3).

The 5 SNPs from the *CYP4V2* region that were associated with DVT were all associated with factor XI levels in LETS and MEGA-1, with higher factor XI levels for those who carried the risk-increasing alleles (Table 5). We investigated whether factor XI levels mediate the association between these 5 SNPs and DVT by adjusting for factor XI levels in the combined LETS and MEGA-1 studies. For all 5 SNPs, adjustment for factor XI levels weakened the association with DVT but none of the associations disappeared. Interestingly, the 5 SNPs that were not associated with DVT in the combined analysis of LETS and MEGA-1 (rs3736456, rs4253259, rs4253408, rs4253325, and rs3775302) were also not associated with factor XI levels in LETS.

All analyses were performed with and without adjustment for age and sex, and analyses in MEGA-1 and MEGA-2 were performed with and without restriction to the group with both parents born in northwestern Europe. Because neither influenced the results, we presented the unadjusted OR.

COMMENT

We identified 7 SNPs that were associated with DVT in 3 large, well-characterized populations including 3155 cases and 5087 controls. The evidence was strongest for the 3 SNPs in the *CYP4V2, SERPINC1*, and *GP6* genes. It is interesting to note that these SNPs are in or near genes that have a clear role in blood coagulation. This may indicate that the coagulation system is well characterized.

Testing 19 682 SNPs will result in false-positive associations. Therefore, we investigated the SNPs in 3 large studies and estimated the false discovery rate for the SNPs tested in the third study. The 3 SNPs in genes *CYP4V2*, *SERPINC1*, and *GP6* were associated with DVT with a false discovery rate of less than 10%, which means that less than 10% of these 3 SNPs would be expected to be false positive. Relaxing the false discovery rate to less than 20% would add 4 SNPs, in *RGS7, NR1I2, NAT8B,* and *F9* as associated with DVT.

Table 5. Association of 10 SNPs in *CYP4V2* Region With Deep Vein Thrombosis and Factor XI Levels in the Combined LETS and MEGA-1 Studies.

SNP	Gene	Risk Allele	Genotype	Risk allele, No. (%) Case	Control	Factor XI[a] % Difference (95% CI)	Deep Vein Thrombosis OR (95% CI)	OR[b] (95% CI)
rs13146272	*CYP4V2*	A	CC	181 (10)	296 (13)	[Reference]	1 [Reference]	1 [Reference]
			CA	808 (44)	995 (45)	3 (1 to 6)	1.32 (1.07 to 1.62)	1.26 (1.03 to 1.56)
			AA	850 (46)	919 (42)	7 (4 to 9)	1.50 (1.22 to 1.84)	1.36 (1.10 to 1.68)
			Additive	(68)	(64)	3 (2 to 4)	1.20 (1.09 to 1.31)	1.14(1.04 to 1.25)
rs3736456	*CYP4V2*	T	CC	7 (0)	0 (0)			
			CT	163 (9)	222 (10)	[Reference]	1 [Reference]	1 [Reference]
			TT	1663 (91)	1973 (90)	1 (-1 to 4)	1.16 (0.93 to 1.43)	1.15 (0.93 to 1.42)
			Additive	(95)	(95)	1 (-2 to 4)	1.06 (0.86 to 1.30)	1.05 (0.85 to 1.28)
rs3087505	*KLKB1*	C	TT	6 (0)	25 (1)	[Reference]	1 [Reference]	1 [Reference]
			TC	317 (17)	438 (20)	11 (6 to 16)	3.02 (1.22 to 7.44)	2.59 (1.05 to 6.40)
			CC	1509 (82)	1743 (79)	19 (14 to 24)	3.61 (1.48 to 8.82)	2.81 (1.15 to 6.89)
			Additive	(91)	(89)	8 (6 to 10)	1.27 (1.09 to 1.47)	1.15 (0.99 to 1.34)
rs4253259	*KLKB1*	C	AA	5 (0)	6 (0)	[Reference]	1 [Reference]	1 [Reference]
			AC	168 (9)	219 (10)	0 (-18 to 18)	0.92 (0.28 to 3.07)	0.95 (0.28 to 3.20)
			CC	1652 (91)	1978 (90)	0 (-19 to 18)	1.00 (0.31 to 3.29)	1.03 (0.31 to 3.43)
			Additive	(95)	(95)	0 (-3 to 2)	1.08 (0.88 to 1.32)	1.08 (0.88 to 1.32)
rs4253408	*F11*	A	GG	1526 (83)	1869 (85)	[Reference]	1 [Reference]	1 [Reference]
			GA	293 (16)	317 (14)	4 (2 to 6)	1.13 (0.95 to 1.35)	1.06 (0.89 to 1.27)
			AA	15 (2)	17 (1)	13 (-1 to 27)	1.08 (0.54 to 2.17)	0.97 (0.48 to 1.98)
			Additive	(9)	(8)	5 (3 to 7)	1.11 (0.95 to 1.30)	1.05 (0.89 to 1.23)
rs4253325	*F11*	G	AA	21 (1)	23 (1)	[Reference]	1 [Reference]	1 [Reference]
			AG	308 (17)	392 (18)	5 (0 to 16)	0.86 (0.47 to 1.58)	0.86 (0.46 to 1.59)
			GG	1507 (82)	1785 (81)	8 (-3 to 13)	0.93 (0.51 to 1.68)	0.88 (0.48 to 1.62)
			Additive	(90)	(90)	3 (1 to 5)	1.05 (0.91 to 1.21)	1.01 (0.87 to 1.17)
rs3775302	*KLKB1*	A	GG	1418 (77)	1686 (77)	[Reference]	1 [Reference]	1 [Reference]
			GA	380 (21)	481 (22)	-1 (-3 to 1)	0.94 (0.81 to 1.09)	0.97 (0.83 to 1.13)
			AA	38 (2)	34 (2)	-4 (-11 to 3)	1.33 (0.83 to 2.12)	1.34 (0.83 to 2.14)
			Additive	(12)	(12)	-1 (-3 to 0)	1.00 (0.87 to 1.14)	1.02 (0.89 to 1.16)
rs3756008	*F11*	T	AA	526 (29)	788 (36)	[Reference]	1 [Reference]	1 [Reference]
			AT	903 (49)	1032 (47)	7 (6 to 9)	1.31 (1.14 to 1.51)	1.21 (1.04 to 1.39)
			TT	408 (22)	384 (17)	15 (12-17)	1.59 (1.33 to 1.90)	1.32 (1.09 to 1.59)
			Additive	(47)	(41)	7 (6 to 8)	1.27 (1.16 to 1.38)	1.16 (1.05 to 1.27)

SNP	Gene	Risk Allele	Genotype	Risk allele, No. (%) Case	Control	Factor XI[a] % Difference (95% CI)	Deep Vein Thrombosis OR (95% CI)	OR[b] (95% CI)
rs2036914	*F11*	C	TT	302 (17)	505 (23)	[Reference]	1 [Reference]	1 [Reference]
			TC	895 (49)	1081 (49)	7 (5 to 9)	1.38 (1.17 to 1.64)	1.27 (1.07 to 1.51)
			CC	633 (35)	620 (28)	14 (12 to 16)	1.71 (1.43 to 2.05)	1.43 (1.19 to 1.73)
			Additive	(59)	(53)	7 (6 to 8)	1.30 (1.19 to 1.42)	1.19 (1.08 to 1.30)
rs4253418	*F11*	G	AA	3 (0)	4 (0)	[Reference]	1 [Reference]	1 [Reference]
			AG	120 (7)	199 (9)	14 (10 to 19)	0.80 (0.18 to 3.65)	0.69 (0.15 to 3.14)
			GG	1710 (93)	2000 (91)	22 (18 to 26)	1.14 (0.26 to 5.10)	0.88 (0.20 to 3.94)
			Additive	(97)	(95)	8 (5 to 11)	1.39 (1.11 to 1.74)	1.24 (0.99 to 1.56)

The 3 SNPs with the strongest evidence for association with DVT were in the genes *CYP4V2*, *SERPINC1*, and *GP6*. The *CYP4V2* gene encodes a member of the CYP450 family 4 that is not known to be related to thrombosis [174,175]. The *CYP4V2* gene is located on chromosome 4 in a region containing genes encoding coagulation proteins prekallikrein (*KLKB1*) and factor XI (*F11*). We also found 4 other SNPs in the *CYPV42/KLKB1/F11* locus that were associated with DVT. No previous reports exist of genetic variants in *CYP4V2* and *KLKB1* and their association with DVT. There exists no evidence for an association between prekallikrein levels and DVT [176], while there is evidence for elevated factor XI levels [160,171]. It remains unclear whether only one of these SNPs, or all of them affect DVT risk.

The *SERPINC1* gene encodes antithrombin, a serine protease inhibitor located on chromosome 1 that plays a central role in natural anticoagulation. Deficiencies of antithrombin are rare but result in a strong thrombotic tendency [177]. The SNP in *SERPINC1* (rs2227589) had a minor allele frequency of about 10% in the controls and was associated with a modest thrombotic tendency. The *GP6* gene encodes glycoprotein VI, a 58-kDa platelet membrane glycoprotein that plays a crucial role in the collagen-induced activation and aggregation of platelets [178] and may play a role in DVT [179].

The SNPs in the genes *F9*, *NR1I2*, *RGS7*, and *NAT8B* are of interest for further validation. The *F9* gene encodes factor IX, a vitamin K–dependent coagulation factor, of which high levels have been shown to increase the risk of DVT [180]. The SNP rs6048, also known as *F9* Malmö, is a common polymorphism at the third amino acid residue of the activation peptide of factor IX [181].

The SNP in *CYP4V2* (rs13146272) is located close to the gene encoding coagulation factor XI. Factor XI levels have been reported to be associated with DVT in LETS [171] and in a large analysis of pedigrees [160]. We confirmed the association between DVT and factor XI levels in MEGA. Interestingly, the 5 SNPs in the *CYP4V2* region that were associated with DVT in both LETS and MEGA-1 were also associated with factor XI levels. However, the association between these 5 SNPs and DVT does not seem to be completely explained by variation in factor XI levels because adjusting for factor XI level did not remove the excess DVT risk of these 5 SNPs. Thus, if only part of the risk associated with these genetic variants is mediated through levels of factor XI, some of the risk might also be due to effects on protein function.

Several variants in the *F11* gene (rs5974, rs5970, rs5971, rs5966, rs5976, and rs5973) were previously tested for association with factor XI levels in patients with DVT and atherosclerosis, but no relationship was observed [182]. In the present study, rs5974 (r^2 = 1.0 with 5970) and rs5971 (r^2 = 1.0 with 5966 and rs5976) were not associated with DVT in LETS. Rs5973 was not genotyped because its minor allele frequency was lower than 2% (HapMap CEPH population). In a study of West African volunteers [129], rs3822056 and rs3733403 were associated with transcription factor binding affinity and slightly increased factor XI levels, but neither SNP was associated with DVT in LETS. In a study among white postmenopausal women [183], rs3822057 and rs2289252 were associated with DVT. Both of these associations were indirectly confirmed in the present study because 2 of the 5 SNPs in the *CYPV42* region that were consistently associated with DVT and factor XI levels are in linkage disequilibrium with rs3822057 (r^2 = 0.9 with rs2036914) and rs2289252 (r^2 = 0.8 with rs3756008).

The association between genetic variants and DVT may depend on clinical variables or other risk factors for DVT, such as surgery or the use of oral contraceptives. Because we aimed to identify variants that are associated with DVT in general and from a large set of SNPs, we did not study subgroups. Clinical utility, however, may well depend on interaction with these clinical variables and should form a focus of subsequent studies.

The associations between SNPs and DVT were modest, for instance homozygous carriership of the AA genotype of rs13146272 in *CYP4V2* increased risk 1.49-fold. However, because the variants are common, they might be useful risk indicators especially when combined with other risk factors. Moreover, the associations found might represent a diluted effect of an unmeasured SNP in linkage disequilibrium or indicate a region with several variants involved in DVT susceptibility. The results from the *CYP4V2* region illustrate the need for further study, because some ORs found in that region were higher than initially found for rs13146272.

Only 9 of 18 stage 3 SNPs were indeed tested in MEGA-2 DNA in stage 4. The reason for this was that in order to save MEGA-2 DNA, stage 4 SNPs were genotyped using multiplexed oligoligation assays, and assays for only 9 stage 3 SNPs were available at the time of this study. Therefore, a future extension of this study may yield additional SNPs associated with DVT.

The replication criteria that we used to identify SNPs associated with DVT may have caused us to miss some truly associated positive variants. Although the statistical power to detect associations between DVT and uncommon genetic variation was high, a rare variant with a modest association with DVT may have been missed.

Our analysis was limited to a northwestern European population. Confounding in a genetic study may arise from population stratification, ie, the presence of ethnic groups with different allele and disease frequencies within a study. In LETS, no information on ethnicity was collected. However,

we do not think that population stratification biased our results because MEGA participants were recruited from the same population as LETS but 10 years later and 90% of MEGA had both parents born in northwestern Europe. Furthermore, restricting the analyses to this 90% of MEGA did not modify our results.

CONCLUSIONS

We tested thousands of SNPs for association with DVT in unrelated individuals, and found 7 genetic variants consistently associated with risk. In the *CYP4V2* region we identified several SNPs that were associated with both DVT and factor XI levels. Although most variants had a modest effect on risk, they were common and could therefore be responsible for as many thrombotic events in the population as stronger but rarer variants. Clinical utility may stem from the determinants being frequent and affecting many people, as well as from interactions with environmental risk factors (high-risk situations) and interactions with other genes. Subsequent studies will be needed to further our knowledge on these issues.

ACKNOWLEDGEMENT

Funding/Support: The Leiden Thrombophilia Study was supported by grant 89.063 from the Netherlands Heart Foundation. The Multiple Environmental and Genetic Assessment of Risk Factors for Venous Thrombosis was supported by grant NHS 98.113 from the Netherlands Heart Foundation, grant RUL 99/1992 from the Dutch Cancer Foundation, and grant 912-03-033| 2003 from the Netherlands Organisation for Scientific Research. Celera reimbursed the Leiden University Medical Center for collecting, providing, and shipping the samples. Ms Bezemer received support for training in genetic epidemiology from a Transatlantic Network for Excellence in Cardiovascular Research grant from Fondation Leducq, Paris, France (LINAT project). The funding organizations played no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

Additional Contributions: We acknowledge the contributions of the Celera High Throughput Laboratory and Computational Biology group and thank John J. Sninsky, PhD, and Thomas J. White, PhD, both Celera employees for helpful comments on this manuscript. We thank Rob van Eck, BSc, Jeroen van der Meijden, and Petra J. Noordijk all from Leiden University Medical Center for performing factor XI measurements. We also thank Hans L. Vos, PhD, of Leiden University Medical Center for supervision of laboratory analysis and for his expertise in genetic data interpretation. All were compensated as part of their regular duties.