Gene Variants Associated With Venous Thrombosis: A Replication Study in a Brazilian Multicentre Study

Anna Virgínia Calazans Romano, MSc[1] , Aline Barnabé, PhD[1],
Telma Barbosa Gadelha, PhD, MD[2], João Carlos de Campos Guerra, PhD, MD[3],
Rodrigo Secolin, PhD[4], Fernanda Loureiro de Andrade Orsi, PhD, MD[1],
Gizele de Castro Sousa Campanate, PhD, MD[2], Nelson Wolosker, PhD, MD[5],
and Joyce Maria Annichino-Bizzacchi, PhD, MD[1]

Abstract

Single nucleotide polymorphisms (SNP) associated with Venous Thromboembolism (VTE) risk have been identified in European and American populations. Replicate SNPs associated with VTE in a Brazilian multicenter case-control study of the Southeast region. Patients with previous VTE assisted at the Outpatient Clinics of 3 centers of the Southeast Brazilian region were compared to normal controls of the same geographic region. We evaluated 29 SNPs associated with VTE risk in other populations, and 90 SNPs for stratification analysis of the population. Due to high admixture of Brazilian population and lack of previous studies, the calculation of the sample power was performed after genotyping. Sample size, allelic frequency and Hardy-Weinberg equilibrium were estimated. The association and odds ratio analyses were estimated by logistic regression and the results were adjusted for multiple tests using Bonferroni correction. The evaluation of the genetic structure similarity in the cases and controls was performed by AMOVA. 436 cases and 430 controls were included. It was demonstrated that this sample has a statistical power to detect a genetic association of 79.4%. AMOVA showed that the genetic variability between groups was 0.0% and 100% within each group. None of the SNPs showed association with VTE in our population. A Brazilian multicenter case-control study with adequate sample power, high genetic variability though no stratification between groups, showed no replication of SNPs associated with VTE. The high admixture of Brazilian population may be responsible for these results, emphasizing the influence of the population genetic structure in association studies.

Keywords

GWAS, genetic association, TEV, SNP, single nucleotide polymorphism, venous thromboembolism, pulmonary embolism, thrombosis

Introduction

Venous Thromboembolism (VTE) is a multicausal disease with an annual incidence of approximately 1 in 1,000[1] and is the result of a complex interaction between gene variants and environmental exposures.[2] Despite rare and common genetic risk factors having been identified during the last years, these cannot explain the majority of spontaneous VTE. Indeed, the prevalence of these genetic factors can vary according to ancestry.[3-6] Furthermore, the genetic or environmental factor responsible for increased thrombosis susceptibility associated to elevation of factor VIII, factor IX and factor FXI has not yet been established.

The HapMap project and the introduction of Genome Wide Association Study (GWAS) strategy enabled the identification of genetic alterations associated to complex diseases.[3] This encouraged the search for additional genetic factors that could

[1] Hematology and Hemotherapy Center, University of Campinas (UNICAMP), Campinas, Sao Paulo, Brazil
[2] Hematology, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
[3] Clinical Pathology, Israelite Hospital Albert Einstein, São Paulo, Brazil
[4] Hematology and Hemotherapy Center, Faculty of Medical Sciences, University of Campinas-Unicamp, Campinas, Brazil
[5] Vascular Surgery, Israelite Hospital Albert Einstein, São Paulo, Brazil

Corresponding Author:
Anna Virgínia Calazans Romano, Hematology and Hemotherapy Center, University of Campinas (UNICAMP), Campinas, Sao Paulo, Brazil.
Email: vivicalazans@gmail.com

contribute to the knowledge of the widespread clinical heterogeneity of VTE.

Identification of Single Nucleotide Polymorphisms (SNPs) associated with VTE were based on prospective population studies, case-control, and cohorts of patients and families. Many loci have been robustly demonstrated to harbor common and low frequency susceptibility alleles, including ABO, F2, F5, F11, FGG, PROCR and others.[7]

However, these GWAS studies included only European-ancestry adults from French and Dutch case-control studies, and cohort and case-control studies from the United States, and mostly of Caucasian ethnicity.[4-25] The Brazilian population is a highly admixed population and no study has been previously performed in South America.

Therefore, the aim of the present study was to replicate 29 SNPs previously associated with VTE risk, in a multicenter study covering 3 Brazilian centers of the Southeast region. We also evaluated the correlation between SNP genotyping and activity of the corresponding coagulation parameter in samples of participants of one of the centers.

Objective

We sought to replicate 29 SNPs previously associated with VTE risk, in a multicenter study covering 3 Brazilian centers of the Southeast region.

Methods

Selection of Study Participants

All participants provided written informed consent and the study protocol was approved by the institutional Research Ethics Committee (CAAE: 07970212.2.0000.5404).

This is a multicenter case-control study including outpatients and controls from the Southeast region of Brazil. The cases comprised Caucasian and African descendent unrelated patients with a previous diagnosis of deep venous thrombosis (DVT) of the lower limb or pulmonary embolism (PE). They were consecutively recruited from the Outpatient Clinic of Hemocentro Unicamp (Hemocentro / UNICAMP) in Campinas-SP, Thrombophilia Clinic of the Federal University of Rio de Janeiro (UFRJ) in Rio de Janeiro-RJ, and Thrombophilia Clinic of the Hospital Israelita Albert Einstein (HIAE), in São Paulo-SP, from 2009 to 2013. The controls were represented by Brazilian individuals without a history of VTE, from the same geographical area of the cases, matched by age and sex. Control subjects were not-inbred with the participating patients.

Two investigators from each center interviewed all participants and collected information using a non-validated structured questionnaire. For patients, information regarding gender, age at VTE event, risk factors for VTE (prolonged immobilization, local and general trauma, hospitalization, air travel lasting more than 6 hours or road travel without mobilization por more than 10 hours, 30 days before the VTE episode, surgery, cancer, use of oral contraceptive or hormonal

replacement therapy, pregnancy, puerperium, central venous catheter, nephrotic syndrome, personal or familial history of thrombosis, antiphospholipid syndrome) and thrombosis site were collected.

The inclusion criteria included age between 18 and 60 years and confirmed diagnosis of DVT by Doppler ultrasonography and PE by angiotomography, angioresonance or ventilation/perfusion lung scan.

The exclusion criteria were VTE related to cancer, antiphospholipid syndrome, renal, liver or autoimmune disease, and hereditary thrombophilia (deficiency of protein C—PC, protein S—PS, antithrombin—AT, homozygosity or double heterozygosity for FV Leiden and F2 20210A).

For controls, information regarding gender, age at data collection, clinical comorbidities, use of oral contraceptive or hormonal replacement therapy, pregnancy and surgical history, medicine intake including acetylsalicylic acid were collected.

Exclusion criteria for controls included personal or familial history of cardiovascular or VTE episode, non-thrombotic antiphospholipid syndrome, renal, liver or autoimmune disease, hereditary thrombophilia and hormonal therapy or acetylsalicylic acid use.

SNP Selection

Twenty-nine SNPs were selected due to previous association with VTE irrespective of ethnicity or country of origin[5-25] (Table 1). In order to represent variability across candidate genes, we performed a systematic SNPtag selection using SNPBROWSERTM v.4.0 software (Applied Biosystems, Foster City, California, USA), based on the International Hap-Map Project (http://hapmap.ncbi.nlm.nih.gov/), a minimum allele frequency (MAF) > 0.05, and a pairwise linkage disequilibrium (LD) of r2 >0.8 (Table 1). SNPs were previously tested by the manufacturer and demonstrated good quality performance.

Furthermore, we used an additional 90 SNPs, randomly selected, based on our previous study and not ancestry informative[26] for population stratification analysis. A previous study performed in European individuals showed that any number of SNPs greater than 65 was adequate to evaluate population stratification.[27] Based in our previous study[26] we developed a bioinformatics pipeline to select randomly SNPs for the population stratification analysis. Briefly, using the International HapMap Project, a SNP frequency vector was built based on the following parameters: MAF > 0.05 and minimum distance between 2 SNPs = 300 kb, to avoid possible LD. In the next step, the pipeline used the sample function in R software v.3.2.3 (http://www.r-project.org) to randomly select the 90 SNPs from the SNP frequency vector. We concentrated SNP selection on chromosomes 2, 7, 10, 15 and 22.[28]

Table 2 contains information of previous studies showing the odds ratios and their respective confidence intervals as well as the sample size of each study.

Table 1. Selected SNPs.

Locus	SNP	Alleles*	Global MAF
NME7	rs16861990	A / C	0.062
F5	rs2420371	A / G	0.031
SERPINC1	rs2227589	A / G	0.129
C4BPB	rs3813948	A / G	0.123
STAB2	rs4981021	A / G	0.322
SCARA5	rs9644133	C / T	0.241
BAI3	rs9363864	A / G	0.334
STAB2	rs4981022	C / T	0.249
STXBP5	rs1039084	A / G	0.457
SCARA5	rs2726953	C / T	0.263
STX2	rs7978987	A / G	0.286
vWF	rs1063857	C / T	0.336
TC2N	rs1884841	C / T	0.479
CLEC4M	rs868875	A / G	0.208
SERPINF2	rs8074026	C / T	0.279
SLC44A2	rs2288904	C / T	0.183
GP6	rs1613662	A / G	0.157
GCKR	rs1260326	C / T	0.293
EDEM2	rs6120849	C / T	0.157
PROCR	rs6088735	C / T	0.221
PROCR	rs867186	A / G	0.096
DNAJC6	rs1413885	C / T	0.218
BAZ1B	rs17145713	C / T	0.172
KNG1	rs710446	A / G	0.415
CYP4V2	rs13146272	A / C	0.441
FII	rs2036914	C / T	0.394
FII	rs2289252	C / T	0.319
HIVEP1	rs169713	C / T	0.337
ABO	rs687621	C / T	0.381

*: reference/risk allele; MAF: Minor Allele Frequency.

Genotyping

The SNPs study was carried out using a high-performance genotyping Platform (TaqMan® OpenArray Genotype System—Applied Biosystems), based on real-time PCR, and the plate used was120 SNPs for 24 samples.

Genomic DNA was extracted and purified using standard procedure.[29] The concentration of the extracted DNA samples was determined in a spectrophotometer (Thermo Scientific NanoDrop™ 1000 Spectrophotometer). The DNA samples had A260/A230 and A260/A280 ratios from 1.8 to 2.0; the concentrations of the samples were adjusted to 50 ng/μl.

DNA analysis for both, FV Leiden (G1691A) and F2 G20210A mutation was performed using a polymerase chain reaction (PCR) method using a TaqMan® SNP Genotyping Assay (Applied Biosystems, Foster City, CA, USA).

The OpenArray TaqMan assays for genotyping were performed using the QuantStudio™ 12 K Flex Real-Time PCR System (Applied Biosystems, CA, USA) according to the manufacturer's recommended operating guideline. DNase/RNase-free distilled water was used in each assay as a non-template control (NTC).

The genotype data were analyzed using the OpenArray® SNP Genotyping Analysis software (version 1.0.5) (Applied Biosystems, CA, USA).

Statistical Analysis

We evaluated the allele frequency of genotyped SNPs in terms of Hardy-Weinberg equilibrium (p-value > 0.05), MAF > 0.05 and pairwise LD (r2 > 0.8) by HAPLOVIEW v.4.2 software.[30]

In order to evaluate whether population substructure is similar, we performed the Analysis of MOlecular VAriance (AMOVA) by ARLEQUIN v3.5.1.2 software,[31] using the 90 additional SNPs as analysis factors. AMOVA partitions the source of genetic variance into 2 components: within-groups, between-groups, and measures the population differentiation due to genetic structure by the fixation index (Fst). We considered Fst <0.05 as the absence of population stratification,[26] as well as the between-groups variance component < 0.7%, as this was the smallest value found when comparing populations within regions worldwide.[32]

The analysis of genetic associations was performed by logistic regression test. The power to detect genetic associations was determined from the post hoc statistical power of the sample in the software GPOWER[33] with the following parameters: logistic regression test; 2-tailed; OR = 1.4; α = 0.002; n = 875.

The correction of the multiple testing was performed using the Bonferroni correction.

Results

Study Population

In this case-control study, patients and controls of both genders, aged between 18 and 60 years were included.

In the period from 2009 to 2013, 1258 samples of participants were collected, of which 149 were from HIAE, 638 were from UFRJ and were 471 from Unicamp. Of these, 392 samples were discarded for not meeting purity and/or DNA quantification standards.

Clinical and Demographic Characteristics

Of the 436 cases and 430 controls, 206 cases and 230 controls came from Hemocentro /Campinas, 134 cases and 148 controls came from UFRJ/Rio de Janeiro, and 96 cases and 52 controls came from HIAE/São Paulo. Of those patients 4,8% were heterozygous for FV Leiden and 5,1% heterozygous for F2 20210A mutation. This sample has a statistical power to detect a genetic association of 79.4%. The age and gender of patients and controls separated by each study center are shown in Table 3.

Table 4 shows the site of thrombosis (pulmonary or lower limbs) and the risk factors for provoked VTE.

Population Stratification Analysis

AMOVA results showed Fst<0.05 overall loci, and a between-group variance component of 0.0% and within population of 100%, indicating no differences in genetic structure among populations, and 100% within populations. These results indicated no population stratification (Table 5).

Table 2. SNPs in Other Studies.

Locus	SNP	OR [95% CI] and sample size
NME7	rs16861990	1.79 [1.47–2.18] 1407 cases and 1418 controls
F5	rs2420371	1.39 [1.17–1.64] 1150 cases and 801 controls 1.44 [1.07–1.93] 607 cases and 607 controls
SERPINC1	rs2227589	1.29 [1.11–1.49] 1294 cases and 2836 controls 1.24 [1.05–1.47] 1398 cases and 1757 controls
C4BPB	rs3813948	1.17 [0.91–1.51] 1110 cases and 789 controls 1.20 [0.88–1.64] 596 cases and 590 controls
STAB2	rs4981021	1.10 [not provided] 419 cases and 1228 controls
SCARA5	rs9644133	1.12 [not provided] 419 cases and 1228 controls
BAI3	rs9363864	0.58 [0.42–0.80] 1166 cases and 1408 controls
STAB2	rs4981022	0.97 [0.74–1.27] 656 cases and 710 controls 0,92 [not provided] 1150 cases and 801 controls
STXBP5	rs1039084	0.78 [0.65–0.93] 656 cases and 710 controls
SCARA5	rs2726953	0.78 [0.65–0.93] 656 cases and 710 controls
STX2	rs7978987	1.01 [0.83–1.22] 656 cases and 710 controls
vWF	rs1063857	-
TC2N	rs1884841	1.17 [0.92–1.50] 1150 cases and 801 controls 1.27 [0.94–1.72] 2163 cases and 2617 controls
CLEC4M	rs868875	1.07 [0.67–1.72] 656 cases and 710 controls
SERPINF2	rs8074026	-
SLC44A2	rs2288904	1.19 [1.12–1.26] 7507 cases and 52632 controls 1.28 [1.16–1.40] 3009 cases and 2586 controls
GP6	rs1613662	1.08 [0.90–1.30] 1150 cases and 801 controls 1.15 [not provided] 1314 cases and 2877 controls
GCKR	rs1260326	-
EDEM2	rs6120849	-
PROCR	rs6088735	1.35 [1.19–1.54] 1542 cases and 1110 controls
PROCR	rs867186	1.33 [1.11–1.60] 1542 cases and 1110 controls
DNAJC6	rs1413885	-
BAZ1B	rs17145713	-
KNG1	rs710446	1.20 [1.07–1.34] 1542 cases and 1110 controls
CYP4V2	rs13146272	1.22 [1.10–1.49] 443 cases and 453 controls 1.19 [1.07–1.32] 1398 cases and 1757 controls
F11	rs2036914	1.33 [1.11–1.59] 3155 cases and 5087 controls 1.38 [1.17–1.64] 1841 cases and 2210 controls
F11	rs2289252	1.49 [1.25–1.76] 3155 cases and 5087 controls 1.31 [1.20–1.43] 1841 cases and 2210 controls
HIVEP1	rs169713	1.20 [1.13–1.27] 5862 cases and 7112 controls 1.11 [0,97–1.26] 1542 cases 1110 controls
ABO	rs687621	1.52 [1.28–1.80] 656 cases and 710 controls

OR: Odds Ratio from different studies; Not provided: confidence interval [95% C.I.] information was not provided by the author.

Table 3. Characteristic of Patients and Controls, Separated by Each Study Center and Total.

	HIAE		Unicamp		UFRJ		Total	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
	(n = 96)	(n = 52)	(n = 206)	(n = 230)	(n = 134)	(n = 148)	(n = 436)	(n = 430)
Men, n (%)	46 (47.92%)	27 (51.92%)	61 (29.61%)	65 (28.26%)	48 (35.82%)	55 (37.16%)	155 (35.55%)	147 (34.19%)
Age mean, y	41*	42	42*	40	38*	40	40*	41

*For patients, age at VTE n = number of participants; HIAE = Hospital Israelita Albert Einstein; UNICAMP = University of Campinas; UFRJ = Federal University of Rio de Janeiro.

Association Analysis of SNPs With VTE Risk

The SNP average genotypes call rate was 99.7% (standard deviation = 4.7%) in the candidate genes and 99.4% (standard deviation = 12.2%) for SNPs used in the stratification analysis.

Among the 29 SNPs genotyped in the candidate genes, 4 SNPs (rs2227589, rs169713, rs687621, rs2288904) in the cases and 4 SNPs in the controls (rs2726953, rs687621, rs1884841, rs868875) did not follow the Hardy-Weinberg equilibrium (HWE) and were excluded. None of the SNPs failed frequency test (MAF < 0.05) (Table 6). With these results we excluded SNPs without Hardy-Weinberg equilibrium in the controls.

Our replication results showed that none of the 25 candidate SNPs was associated with VTE risk in this Brazilian population. Results of the association analyses are reported in Table 7.

Discussion

A central goal of human genetics is to identify genetic risk factors for common, complex diseases. GWAS, though powerful to identify SNPs associated to VTE risk, present limitations.[34]

Table 4. Clinical Characteristics of Patients.

	HIAE	Unicamp	UFRJ	Total
Lower limbs	58 (60,4%)	174 (84,5%)	115 (85,8%)	347 (79,6%)
Pulmonary Embolism—PE	16 (16,7%)	21 (10,2%)	18 (13,4%)	55 (12,6%)
Lower limbs associated with PE	22 (22,9%)	11 (5,3%)	1 (0,7%)	34 (7,8%)
Spontaneous	43 (44,8%)	52 (25,2%)	89 (66,4%)	184 (42,2%)
Provoked VTE	53 (55,2%)	154 (74,8%)	45 (33,6%)	252 (57,8%)
Hormonal anticonceptive hormone replacement therapy	10 (18,9%)	62 (40,3%)	2 (4,4%)	74 (29,4%)
Surgery	4 (7,5%)	34 (22,1%)	13 (28,9%)	51 (20,2%)
Pregnancy/Puerperium	4 (7,5%)	36 (23,4%)	6 (13,3%)	46 (18,3%)
Trauma	12 (22,6%)	7 (4,5%)	9 (20%)	28 (11,1%)
Immobilization	14 (26,4%)	8 (5,2%)	12 (26,7%)	34 (13,5%)
Travel	9 (17%)	7 (4,5%)	3 (6,7%)	19 (7,5%)

Table 5. Analysis of Molecular Variance (AMOVA).

Source of variation	Sum of squares	Variance components	Percentage variation
Among Populations	14.191	−0.00089	0.0000
Within populations	24.020.005	1.491.077	100.00
Total	24.034.196	1.490.988	

Recently GWAS of VTE linked SNPs located in genes that encode proteins associated with coagulation and haemostasis, such as receptor glycoprotein VI (GP6), AT (SERPINC1), PC receptor (PROCR), F11 and F5. There are SNPs in genes associated to protein involved in pathways associated to DVT, such as HIVEP1 and GCR. Otherwise, some SNPs are functional variants and others may only be markers in linkage

Table 6. Presented HWE in Case and Controls and MAF for Each SNPs.[a]

SNP	Gene	Chr.	HWE, Cases	HWE, Controls	HWE, Total	MAF%
rs1413885	DNAJC6	1	0.3598	0.2784	0.8875	0.4568
rs16861990	NME7	1	0.5585	0.2212	0.1969	0.3218
rs2420371	F5	1	0.7807	1	1	0.1005
rs2227589*	SERPINC1	1	0.01876	0.4754	0.2597	0.2208
rs3813948	C4BPB	1	0.1947	0.4376	0.7832	0.2583
rs1260326	GCKR	2	0.3104	0.1946	0.102	0.4439
rs710446	KNG1	3	0.7599	0.7697	0.6221	0.4833
rs13146272	CYP4V2	4	0.2066	0.5986	0.6413	0.3446
rs2036914	F11	4	1	0.2244	0.3802	0.404
rs2289252	F11	4	0.7199	0.9096	0.869	0.3073
rs169713*	HIVEP1	6	0.01995	0.2999	0.4549	0.2081
rs9363864	BAI3	6	0.6818	0.6087	0.8853	0.4173
rs1039084	STXBP5	6	0.08157	0.2322	0.8301	0.4369
rs17145713	BAZ1B	7	0.3928	0.4771	0.2389	0.3911
rs2726953**	SCARA5	8	0.2058	0.01769	0.0142	0.08067
rs9644133	SCARA5	8	0.5652	0.1492	0.1455	0.237
rs687621* **	ABO	9	0.01406	0.0313	0.0008949	0.4727
rs1063857	VWF	12	0.622	1	0.6404	0.1846
rs4981022	STAB2	12	0.4788	0.6255	0.9439	0.4821
rs4981021	STAB2	12	0.8199	0.7467	1	0.3389
rs7978987	STX2	12	0.7266	0.5844	0.9365	0.3294
rs1884841**	TC2N	14	0.8147	0.004088	0.0266	0.3113
rs8074026	SERPINF2	17	0.8206	0.2692	0.4766	0.3303
rs868875**	CLEC4M	19	0.5892	0.04017	0.06269	0.3846
rs2288904*	SLC44A2	19	0.04182	0.1167	0.8097	0.1783
rs1613662	GP6	19	0.6439	0.3756	0.8106	0.3286
rs6120849	EDEM2	20	0.6115	1	0.6722	0.4623
rs6088735	PROCR	20	0.4411	0.1251	0.6509	0.2639
rs867186	PROCR	20	0.4672	0.72	0.395	0.2901

[a]Seven SNPs Were Not in HWE.
SNPs not following Hardy-Weinberg equilibrium in cases (*) and in controls (**).
Chr, chromosome; HWE Hardy-Weinberg equilibrium; MAF, minor allele frequency.

Table 7. Results of Logistic Regression.

SNP	Gene	Chr	p Value	OR (95%)	Beta
rs1413885	DNAJC6	1	0.8326	0.97 (0.81-1.19)	−0.2114
rs16861990	NME7	1	0.7783	0.97 (0.79-1.19)	−0.2815
rs2420371	F5	1	0.8701	0.97 (0.70-1.35)	−0.1635
rs2227589	SERPINC1	1	0.7179	1.04 (0.83-1.31)	0.3612
rs3813948	C4BPB	1	0.6074	1.06 (0.85-1.33)	0.5138
rs1260326	GCKR	2	0.1021	1.18 (0.97-1.45)	1.635
rs710446	KNG1	3	0.2365	0.88 (0.73-1.08)	−1.184
rs13146272	CYP4V2	4	0.1545	0.86 (0.70-1.06)	−1.424
rs2036914	F11	4	0.8942	1.01 (0.83-1.23)	0.133
rs2289252	F11	4	0.7541	0.96 (0.78-1.20)	−0.3133
rs169713	HIVEP1	6	0.8959	0.98 (0.77-1.26)	−0.1309
rs9363864	BAI3	6	0.6209	1.25 (1.03-1.53)	2.244
rs1039084	STXBP5	6	0.7562	1.03 (0.85-1.26)	0.3104
rs17145713	BAZ1B	7	0.5033	0.93 (0.76-1.14)	−0.6693
rs9644133	SCARA5	8	0.3685	0.90 (0.72-1.13)	−0.8993
rs1063857	VWF	12	0.5455	1.08 (0.84-1.39)	0.6046
rs4981022	STAB2	12	0.6194	0.95 (0.78-1.16)	−0.4967
rs4981021	STAB2	12	0.7009	0.96 (0.78-1.18)	−0.3841
rs7978987	STX2	12	0.5746	0.94 (0.77-1.16)	−0.5613
rs8074026	SERPINF2	17	0.9083	1.01 (0.86-1.23)	0.1152
rs2288904	SLC44A2	19	0.4278	1.11 (0.86-1.44)	0.793
rs1613662	GP6	19	0.9988	1.00 (0.81-1.23)	0.001454
rs6120849	EDEM2	20	0.5842	1.05 (0.87-1.29)	0.5473
rs6088735	PROCR	20	0.5641	1.06 (0.86-1.33)	0.5768
rs867186	PROCR	20	0.5256	1.07 (0.86-1.34)	0.6348

Chr: chromosome.

disequilibrium (LD) with functional variants that remain to be identified, like MYH7B, EDEM2, BAZ1B.[14,18]

It is important to note that although SNPs are the most frequent genetic variation in the human genome, the majority has a minimal biological impact and confer a modest increase of disease risk,[35] and this is not different for VTE.

GWAS that identified SNPs linked to VTE included predominantly Caucasian populations, and only one study included Afro descendants,[36] with a poor representation of other areas and admixed populations. We performed the first case-control study in South America, in a Brazilian admixed population, with the aim to replicate some of these SNPs previously demonstrated in other populations. However, our replication results showed that none of the 25 candidates SNPs were associated with VTE risk in this Brazilian population.

Factors that can explain the fact by which GWAS findings in European populations may not be transferable to admixed American cohorts are differences in population substructure, including a specific distribution of ancestry tracts in the genome,[37] a positive selection of alleles among populations and even within a population, and genetic variants that could have functional effects.[38]

Brazil is a genetically trihybrid population, with the genomic inheritance of European, sub-Saharan African, and Native-American populations.[39-42] It has been described different proportion of global ancestries among these 3 populations and other admixed American populations,[13,22,23] and in Brazilian individuals from diverse geographical regions.[19,20]

We recently performed an analysis of an admixed population of the same geographical area of the individuals included in this study and demonstrated that some non-inverted haplotypes were more similar to Native-American haplotypes than to European haplotypes, in contrast to what was found in other American admixed populations. We showed that individuals who display the same global ancestry could exhibit remarkable differences in the distribution of local ancestry blocks.[43]

A SNP identified in a study with a sufficient sample size and replicated in an independent dataset from the same population, confirmed the effect in the GWAS target population.[44] However, reviews of literature based on associations between SNPs and VTE have not always been robustly confirmed in studies aimed for replication, even in the same population.[15,17,19,36] Failure to replicate in a different population does not necessarily invalidate the original study and suggests that the SNP has an ethnic-specific effect.

Austin et al., (2011) evaluated 10 SNPs identified in 3 European case-control studies as risk factors for VTE in Caucasian and Afro descendants Americans. Some SNPs were replicated in Caucasians; however, the association with F11 SNPs after adjustment for multiple comparisons among Afro descendants was not confirmed. In this setting, we correctly performed Bonferroni correction in our analysis.[43]

An important point is the relation of one SNP with another established genetic factor. In a case-control study,[45] the authors described an association of VTE with 4 SNPs (ATP1B1, NME7, BLZF1, and SLC19A2). However, after a replication study and control for FV Leiden (F5 rs6025), no additional information over that provided by patients characteristics and F5 rs6025T alone could be added. They found that FV Leiden is inherited as a haplotype block that spans 7 genes on chromosome 1q24.2, including those that have been associated with VTE risk, such as ATP1B1, NME7, BLZF1, C1orf114, SLC19A2, F5 and SELP.

Phenotype comparable to those reported in the initial study should be used in all replication studies. Our patients presented phenotype very similar to those described in the original studies, including VTE of the lower limbs and PE, which is a strength of our study.

The possibility of differences in selection criteria when compared to other European and American studies would have influenced the confirmation of the association of some of the SNPs with VTE risk. MARTHA cases are patients referred to thrombophilia centers generally younger than FARIVE cases, recruited from the general population.[11,13] In addition, MARTHA population was enriched with Factor V Leiden or F2 G20210A mutations, 50% in cases and 41% in controls, compared to 18% and 8% in FARIVE, respectively. All 3 centers that selected patients for our study were thrombophilia reference centers. We excluded patients with known hereditary thrombophilia associated with a high risk of VTE that could explain per se the symptomatic thrombosis, like anticoagulant deficiency, and homozygosity or double heterozygosity FV Leiden and F2 20210A.

We genotyped FV Leiden and F2 20210A mutation in 1104 VTE patients assisted at Hemocentro Unicamp from 2004 to 2019 and found a low prevalence of these mutations than that described in Europe (4,4% and 4,1%, respectively). These results emphasize the ethnic differences of a Brazilian admixed population when compared to European.

We noted a great variability in sex distribution between the 3 participating centers, with a higher number of women patients from Unicamp. The healthcare of Unicamp include a secondary/tertiary general hospital, a hematology clinic at Hemocentro / Unicamp, and a woman hospital (CAISM). Patients that presented DVT were assisted at outpatient clinic of Hemocentro / Unicamp, after hospital discharge. This could explain the higher number of women in this sample. It is important to point that we performed statistical analysis according to sex, but no difference was found (data not shown).

About 58% of our patients presented provoked VTE, but 55,5% were associated to mild risk factors. This percentage was even higher in those patients assisted at Hemocentro / Unicamp (68,2%), which provided the most patients and controls included in this study. The inclusion of only spontaneous DVT would be desirable to avoid underestimation of genetic influence. Otherwise, VTE is a complex disease and the association between acquired and genetic risk factors can play a role to trigger the thrombotic episode.

Lack of reproducibility of genetic associations can be attributed to lack of comparability between cases and controls, increasing the risk of biases as there can be heterogeneity in exposure to environmental challenges and population stratification, particularly in admixed populations.[26,39,42] In our study, as we were aware of these critical points, we included patients and controls of the same geographical area. Indeed, we genotyped 90 SNPs to exclude stratification of the population analyzed and our results confirmed that the 2 groups had similar genetic structures, which allowed us to perform an unbiased genetic association analysis.

Small sample size is a frequent problem and can result in insufficient power to detect minor contributions of one or more alleles. We calculated our sample size according to previous studies in populations of European ancestry, which presented a large variation in OR.[17,24] Our sample size presented 79.4% of statistical power to detect genetic association, with allele frequencies greater than 5%. However, all of our OR were lower than 1.1, and our study was underpowered to determine whether any association between the SNPs and DVT existed in our population. Despite the attention given to many of the important points in a replication study, we consider this a limiting factor of our study.

Despite having included a large number of SNPs, as those associated with biochemical traits related to VTE and fibrinolytic system, the allele frequency of the SNPs was lower than expected. A recent meta-analysis that included 7,507 VTE and 52,632 controls, only confirmed 9 loci previously associated with VTE.[46]

The genetic heterogeneity of the Brazilian population probably in part explains this lower allele frequencies, highlighting the influence of population's genetic structure in association studies.

Further studies designed to investigate new genes, common, rare and regulatory variants will possibly contribute to the knowledge and the use of genetic analysis to determine individual VTE risk in the clinical practice.

Conclusions

In a Brazilian multicenter case-control study, with high genetic variability though no stratification between groups, there was no replication of SNPs associated with VTE in GWAS, previously described for European and United States adults, mostly of Caucasian ethnicity. The high admixture of the Brazilian population may be a critical point for these results, emphasizing the influence of the population genetic structure in association studies.

To our knowledge this was the first study that evaluated SNPs associated with VTE in the South America, in Brazilian population.

Declaration of Conflicting Interests

Funding

ORCID iD

Anna Virgínia Calazans Romano 🔟 https://orcid.org/0000-0002-5729-5271

References

1. Raskob GE, Angchaisuksiri P, Blanco AN, et al. Thrombosis: a major contributor to global disease burden. *Arterioscler Thromb Vasc Biol*. 2014;34(11):2363-2371.
2. Rosendaal FR. Venous thrombosis: a multicausal disease. *Lancet*. 1999;353(9159):1167-1173.
3. International HapMap Consortium. The International HapMap Project. *Nature*. 2003;426(6968):789-796.
4. Van der Meer FJ, Koster T, Vandenbroucke JP, Briet E, Rosendaal FR. The Leiden Thrombophilia Study (LETS). *Thromb Haemost*. 1997;78(1):631-635.
5. Blom JW, Doggen CJ, Osanto S, Rosendaal FR. Malignancies, prothrombotic mutations, and the risk of venous thrombosis. *JAMA*. 2005;293(6):715-722.
6. Buil A, Soria JM, Souto JC, et al. Protein C levels are regulated by a quantitative trait locus on chromosome 16: results from the Genetic Analysis of Idiopathic Thrombophilia (GAIT) Project. *Arterioscler Thromb Vasc Biol*. 2004;24(7):1321-1325.
7. Rosendaal FR. Venous thrombosis: the role of genes, environment, and behavior. *Hematology Am Soc Hematol Educ Program*. 2005;2005(1):1-2.
8. Bezemer ID, Bare LA, Doggen CJ, et al. Gene variants associated with deep vein thrombosis. *JAMA*. 2008;299(11):1306-1314.

9. Li Y, Bezemer ID, Rowland CM, et al. Genetic variants associated with deep vein thrombosis: the F11 locus. *J Thromb Haemost*. 2009;7(11):1802-1808.

10. Souto JC, Almasy L, Borrell M, et al. Genetic determinants of hemostasis phenotypes in Spanish families. *Circulation*. 2000; 101(13):1546-1551.

11. Buil A, Trégouët DA, Souto JC, et al. C4BPB/C4BPA is a new susceptibility locus for venous thrombosis with unknown protein S–independent mechanism: results from genome-wide association and gene expression analyses followed by case-control studies. *Blood*. 2010;115(23):4644-4650.

12. Athanasiadis G, Buil A, Souto JC, et al. A genome-wide association study of the protein C anticoagulant pathway. *PLoS One*. 2011;6(12):e29168.

13. Morange PE, Oudot-Mellakh T, Cohen W, et al. KNG1 Ile581Thr and susceptibility to venous thrombosis. *Blood*. 2011;117(13): 3692-3694.

14. Oudot-Mellakh T, Cohen W, Germain M, et al. Genome wide association study for plasma levels of natural anticoagulant inhibitors and protein C anticoagulant pathway: the MARTHA project. *Br J Haematol*. 2012;157(2):230-239.

15. Houlihan LM, Davies G, Tenesa A, et al. Common variants of large effect in F12, KNG1, and HRG are associated with activated partial thromboplastin time. *Am J Hum Genet*. 2010;86(4): 626-631.

16. Antoniak M, Pugliatti M, Hubbard R, et al. Vascular factors and risk of dementia: design of the three-city study and baseline characteristics of the study population. *Neuroepidemiology*. 2003; 22(6):316-325.

17. Germain M, Saut N, Greliche N, et al. Genetics of venous thrombosis: insights from a new genome wide association study. *PLoS One*. 2011;6(9):e25581.

18. Tang W, Basu S, Kong X, et al. Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study. *Blood*. 2010;116(23):5032-5036.

19. Tregouet DA, Heath S, Saut N, et al. Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. *Blood*. 2009; 113(21):5298-5303.

20. Morange PE, Bezemer I, Saut N, et al. A follow-up study of a genome-wide association scan identifies a susceptibility locus for venous thrombosis on chromosome 6p24. 1. *Am J Hum Genet*. 2010;86(4):592-595.

21. Antoni G, MORANGE PE, Luo Y, et al. A multi-stage multi-design strategy provides strong evidence that the BAI3 locus is associated with early-onset venous thromboembolism. *J Thromb Haemost*. 2010;8(12):2671-2679.

22. Breteler MM, Van Den Ouweland FA, Grobbee DE, Hofman A. A community-based study of dementia: the Rotterdam Elderly Study. *Neuroepidemiology*. 1992;11(Suppl 1):23-28.

23. Psaty BM, O'donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic epidemiology (CHARGE) consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet*. 2009;2(1):73-80.

24. Smith NL, Chen MH, Dehghan A, et al. Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: the CHARGE consortium. *Circulation*. 2010;121(12):1382.

25. Antoni G, Oudot-Mellakh T, Dimitromanolakis A, et al. Combined analysis of three genome-wide association studies on vWF and FVIII plasma levels. *BMC Med Genet*. 2011;12(1):102.

26. Silva-Alves MS, Secolin R, Carvalho BS, et al. A prediction algorithm for drug response in patients with mesial temporal lobe epilepsy based on clinical and genetic information. *PLoS One*. 2017;12(1):e0169214.

27. Turakulov R, Easteal S. Number of SNPS loci needed to detect population structure. *Hum Hered*. 2003;55(1):37-45. PMID: 12890924.

28. Baye TM. Inter-chromosomal variation in the pattern of human population genetic structure. *Hum Genomics*. 2011;5(4):220-240. doi:10.1186/1479-7364-5-4-220. PMID:21712187.

29. Maniatis T, Fritsch EF, Sambrook J. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press; 1989.

30. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005; 21(2):263-265. doi:10.1093/bioinformatics/bth457. PMID: 15297300.

31. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10(3):564-567. doi:10. 1111/j.1755-0998.2010.02847.x. PMID:21565059.

32. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science*. 2002;298(5602):2381-2385. doi:10. 1126/science.1078311. PMID:12493913.

33. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G* power 3.1: tests for correlation and regression analyses. *Behav Res Methods*. 2009;41(4):1149-1160.

34. Morange PE, Tregouet DA. Lessons from genome-wide association studies in venous thrombosis. *J Thromb Haemost*. 2011;9:258-264.

35. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467(7319):1061.

36. Austin H, De Staercke C, Lally C, Bezemer ID, Rosendaal FR, Hooper WC. New gene variants associated with venous thrombosis: a replication study in White and Black Americans. *J Thromb Haemost*. 2011;9(3):489-495.

37. Martin AR, Gignoux CR, Walters RK, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet*. 2017;100(4):635-649.

38. Goetz LH, Uribe-Bruce L, Quarless D, Libiger O, Schork NJ. Admixture and clinical phenotypic variation. *Hum Hered*. 2014; 77(1-4):73-86.

39. Rodrigues de Moura R, Coelho AV, de Queiroz Balbino V, Crovella S, Brandão LA. Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. *Am J Hum Biol*. 2015;27(5):674-680.

40. Kehdy FS, Gouveia MH, Machado M, et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc Natl Acad Sci U S A*. 2015;112(28): 8696-8701.

41. Mychaleckyj JC, Havt A, Nayak U, et al. Genome-wide analysis in Brazilians reveals highly differentiated native American genome regions. *Mol Biol Evol*. 2017;34(3):559-574.

42. Passos-Bueno MR, Bertola D, Horovitz DD, de Faria Ferraz VE, Brito LA. Genetics and genomics in Brazil: a promising future. *Mol Genet Genomic Med*. 2014;2(4):280-291.

43. Secolin R, Mas-Sandoval A, Arauna LR, et al. Distribution of local ancestry and evidence of adaptation in admixed populations. *Sci Rep*. 2019;9(1):1-12.

44. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype–phenotype associations. *Nature*. 2007;447(7145):655-660.

45. Heit JA, Armasu SM, Asmann YW, et al. A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *J Thromb Haemost*. 2012;10(8):1521-1531.

46. Trégouët DA, Delluc A, Roche A, et al. Is there still room for additional common susceptibility alleles for venous thromboembolism? *J Thromb Haemost*. 2016;14(9):1798-1802.