



FACULDADE DE MEDICINA DA UNIVERSIDADE DE COIMBRA

MESTRADO INTEGRADO EM MEDICINA – TRABALHO FINAL

CATARINA MARÇAL DA SILVA

Single Nucleotide Polymorphisms associated with venous thromboembolism - Factor XI, ABO blood group and Fibrinogen loci – Evaluation in a Portuguese population group

ARTIGO CIENTÍFICO

ÁREA CIENTÍFICA DE HEMATOLOGIA

Trabalho realizado sob a orientação de:

PROFESSORA DOUTORA ANA BELA SARMENTO ANTUNES CRUZ RIBEIRO

LICÍNIO MANUEL MENDES MANCO

PATRÍCIA JOÃO RODRIGUES FERNANDES MARTINHO SIMÕES

TERESA DE JESUS SEMEDO FIDALGO

FEVEREIRO/2017

Single Nucleotide Polymorphisms associated with venous thromboembolism - Factor XI, ABO blood group and Fibrinogen loci – Evaluation in a Portuguese population group

Catarina Marçal da Silva¹, Licínio Manuel Mendes Manco^{2,3}, Teresa de Jesus Semedo Fidalgo³, Patrícia João Rodrigues Fernandes Martinho Simões⁴, Ana Bela Sarmiento Antunes Cruz Ribeiro^{1,5,6,7,8}

¹Faculdade de Medicina da Universidade de Coimbra (FMUC), 3000-548 Coimbra, Portugal

²Centro de Investigação em Antropologia e Saúde (CIAS), 3000-456 Coimbra

³Universidade de Coimbra, Departamento de Ciências da Vida, 3000-456 Coimbra

⁴Serviço de Hematologia Clínica, Unidade de Trombose e Hemostase – Centro Hospitalar e Universitário de Coimbra - CHUC, EPE, Hospital Pediátrico de Coimbra, 3000-602 Coimbra

⁵Clínica Universitária de Hematologia e Unidade de Biologia Molecular Aplicada - Laboratório de Oncobiologia e Hematologia, 3000-548 Coimbra, Portugal

⁶Centro de Investigação em Meio Ambiente, Genética e Oncobiologia (CIMAGO), FMUC, 3000-548 Coimbra, Portugal

⁷Centro de Neurociências e Biologia Celular & Instituto de Imagem Biomédica e Ciências da Vida (CNC.IBILI), 3004-504 Coimbra, Portugal

⁸Serviço de Hematologia Clínica, Centro Hospitalar Universitário de Coimbra (CHUC, EPE), 3000-548 Coimbra, Portugal

TABLE OF CONTENTS

TABLE INDEX.....	iii
FIGURE INDEX.....	iii
ABBREVIATIONS.....	iv
RESUMO	1
ABSTRACT.....	3
BACKGROUND.....	5
MATERIALS AND METHODS	8
Sample Selection	8
Ethical Statement.....	9
Genotyping	9
Statistical Analyses.....	11
RESULTS.....	12
Population characteristics	12
Evaluation of <i>FGG</i> , <i>ABO</i> and <i>F11</i> SNP allelic frequencies.....	13
Analysis of the SNP in <i>FGG</i> , <i>ABO</i> and <i>F11</i> genes with the risk of VTE.....	14
Analysis of the association between the total number of risk alleles in <i>FGG</i> and <i>ABO</i> genes with the risk of VTE.....	16
DISCUSSION AND CONCLUSIONS.....	17
CONFLICT OF INTEREST STATEMENT.....	21
ACKNOWLEDGMENTS	21
REFERENCES.....	22

TABLE INDEX

TABLE 1: DEMOGRAPHIC, CLINICAL AND MOLECULAR CHARACTERISTICS OF VTE PATIENTS AND CONTROLS.	13
TABLE 2: ASSOCIATION OF SNP ALLELIC FREQUENCIES IN <i>ABO</i> , <i>FGG</i> AND <i>FII</i> GENES WITH THE RISK OF VTE	15
TABLE 3: ASSOCIATION BETWEEN THE TOTAL NUMBER OF RISK ALLELES AND THE RISK FOR VTE.....	16

FIGURE INDEX

FIGURE 1: Sample selection according to the inclusion and exclusion criteria.	12
FIGURE 2: RFLP gel electrophoresis using DdeI restriction enzyme for <i>ABO</i> rs2519093 SNP.	14

ABBREVIATIONS

CI Confidence interval

DVT Deep vein thrombosis

FGG Fibrinogen

FXI Factor XI

HWE Hardy- Weinberg equilibrium

MAF Minor allele frequency

OR Odds ratio

PCR Polymerase chain reaction

PE Pulmonary Embolism

RFLP Restriction fragment length polymorphism

SISA Simple Interactive Statistical Analysis

SNP Single nucleotide polymorphism

VTE Venous thromboembolism

RESUMO

Introdução: O Tromboembolismo venoso (TEV) é uma doença que resulta da interação de fatores de risco genéticos e adquiridos. Com o desenvolvimento dos *Genome wide association studies* (GWAS), foi possível demonstrar a associação do TEV com *loci* de suscetibilidade comuns ou de baixa frequência que podem ajudar a explicar a heritabilidade do TEV. De entre as variantes de risco destacam-se os alelos de risco de três *single nucleotide polymorphisms* (SNPs), alelo T do *ABO* rs2519093, alelo C do *F11* rs2036914 e alelo T do *FGG* rs2066865, que nunca foram testadas na população portuguesa. Como tal, este estudo pretende avaliar a frequência dos três alelos de risco nos *loci* *ABO*, *F11* e *FGG* e averiguar a associação com o risco de TEV numa amostra populacional Portuguesa.

Metodologia: Realizou-se um estudo caso-controlo retrospectivo com 95 casos de TEV sem fatores genéticos ou adquiridos de elevado ou moderado risco (2012-2015) e 132 controlos saudáveis de origem Portuguesa. Os SNP's *F11* rs2036914 e *FGG* rs2066865 foram genotipados utilizando PCR em tempo real com sondas TaqMan. O SNP *ABO* rs2519093 foi genotipado por *restriction fragment length polymorphism* (RFLP) recorrendo à enzima de restrição DdeI. Foram estimadas as frequências alélicas, avaliada a conformidade com o equilíbrio Hardy-Weinberg e testada a associação entre os *loci* em estudo e o risco de TEV por regressão logística através do cálculo do OR (*odds ratio*), intervalos de confiança (IC) de 95% e valores de *p*, recorrendo ao software PLINK. A associação entre o número cumulativo de alelos de risco e o risco para TEV foi calculado através do teste de χ^2 de Pearson utilizando o programa *Simple Interactive Statistical Analysis* (SISA).

Resultados: Na população estudada, as frequências alélicas estimadas do alelo de risco foram: 0,194 para *FGG* rs2066865 (alelo T), 0,63 para o *F11* rs2036914 (alelo C) e 0,302 para *ABO* rs2519093 (alelo T). O teste de regressão logística no modelo aditivo mostrou que o

polimorfismo *FGG* rs2066865 se encontra associado ao risco de TEV (nominal $p=0,02$; OR=1,71, IC 95% 1,08-2,69) e o polimorfismo *ABO* rs2519093 aproximou-se de níveis marginais de significância estatística (nominal $p=0,087$ ajustado à idade e sexo, OR=1,42, IC 95% 0,93-2,16) para TEV. O loci *F11* rs2036914 não mostrou associação estatisticamente significativa ao risco de TEV ($p=0,64$). Observou-se igualmente um aumento do risco para TEV associado a um número cumulativo de alelos de risco (0 vs. 2 ou mais alelos de risco: $\chi^2=5,50$, $p=0,019$, OR=2,31).

Conclusões: Os resultados obtidos sugerem que o alelo de risco T do SNP *FGG* rs2066865 e do *ABO* rs2519093 contribuem para o risco de TEV nesta amostra da população Portuguesa, individual e cumulativamente.

Palavras-chave: Tromboembolismo Venoso, Polimorfismos de nucleótido único, Fibrinogénio, Grupo de Sangue do sistema ABO, Fator XI, Alelos de risco.

ABSTRACT

Background: Venous Thromboembolism (VTE) is a disease resulting from the interactions of inherited and non-inherited risk factors. Genome wide association studies (GWAS) enable the association between VTE and common or low-frequency susceptibility loci, which can explain the heritability of VTE. From the risk variants, there are three single nucleotide polymorphisms (SNP) *ABO* rs2519093 (allele T), *F11* rs2036914 (allele C) and *FGG* rs2066865 (allele T) that have not yet been assessed in the Portuguese population. Therefore, we intend to evaluate the frequency of these three risk alleles and to assess the association between risk alleles in the *ABO*, *F11*, *FGG* loci and the risk of VTE.

Materials and Methods: A retrospective case-control study was conducted with 95 cases of VTE without strong or moderate inherited or non-inherited predisposing factors (2012-2015) and 132 healthy controls of Portuguese origin. *F11* rs2036914 and *FGG* rs2066865 SNP's were genotyped by real-time PCR with TaqMan probes. *ABO* rs2519093 SNP was genotyped by restriction fragment length polymorphism (RFLP) with DdeI restriction enzyme. PLINK software was used to determine the allelic frequencies, concordance with Hardy Weinberg equilibrium and association between risk alleles and VTE through logistic regression estimating OR with 95% confidence intervals (95% CI) and *p*-values. The association between the cumulative number of risk alleles and the risk of VTE was assessed through Pearson χ^2 using the *Simple Interactive Statistical Analysis software* (SISA).

Results: In the analyzed population, the estimated risk allele frequencies were: 0.194 for *FGG* rs2066865 (allele T), 0.63 for *F11* rs2036914 (allele C) and 0.302 for *ABO* rs2519093 (allele T). The logistic regression under an additive model showed that *FGG* rs2066865 was associated with VTE (nominal *p*=0.02; OR=1.71, CI 95% 1.08-2.69) and *ABO* rs2519093 attained near marginal significance in the association with VTE (nominal *p*=0.087 adjusted

for age and sex; OR=1.42, IC 95% 0.93-2.16). *FII* rs2036914 never reach statistical significant association with VTE ($p=0.64$). In addition, there was an increased risk of VTE associated with the increment in the total number of risk alleles (0 vs. 2 or more risk alleles: $\chi^2=5.50$, $p=0.019$, OR=2.31).

Conclusion: The obtained results suggest that the risk allele T of *FGG* rs2066865 and *ABO* rs2519093 SNPs contribute to the VTE risk (individually and additively) in this Portuguese population sample.

Key-words: Venous Thromboembolism, Single nucleotide polymorphisms, Fibrinogen, ABO Blood-Group system, Factor XI, Risk alleles

BACKGROUND

Venous thromboembolism (VTE) results from a disequilibrium between procoagulant and anticoagulant factors which are related with the presence of endothelial dysfunction, hypercoagulability status and hemodynamic changes (Virchow's triad).¹⁻³ VTE can affect any branch of the venous circulation⁴ and usually manifest as deep vein thrombosis (DVT) or pulmonary embolism (PE).⁵

VTE is the third most frequent cardiovascular disease⁶ with an estimated annual incidence rate of 148 per 100,000 and 95 per 100,000 for DVT and PE, respectively, in the European Union.⁷ After the first VTE event, the recurrence rate is approximately 30% within 10 years.⁴ It is a very significant cause of death in Europe and in the USA.⁸ Among the survivors, 1 to 4% will have chronic thromboembolic pulmonary hypertension and 25 to 50% will develop long-term debilitating health problems like post-thrombotic syndrome (after DVT).⁹⁻¹¹

VTE is a multifactorial disease resulting from the interaction between non-inherited and inherited risk factors.^{1, 3, 4, 12, 13} Amongst the non-inherited predisposing factors for VTE, some are considered strong/moderate (odds ratio >2) provoking factors, such as, major surgery, active cancer with or without chemotherapy, spinal cord injury, major trauma, immobilization, auto-immune diseases, pregnancy or puerperium, hormone replacement therapy, oral contraception or long haul flights (>4-6 hours).^{1, 4, 6, 14, 15} Nevertheless, 50% of VTE occur in the absence of such risk factors and the event is categorized as "unprovoked".^{1,}

¹⁶ Regarding the inherited risk factors, VTE has a strong heritability (estimated around 50%).^{1, 5} Amongst the inherited factors, there are classical factors, such as, rare deficiencies in natural coagulation inhibitors (antithrombin, protein C and protein S) and coagulation factors prothrombin G20120A mutation and Factor V Leiden, but also weak genetic risk factors like

common and low-frequency susceptibility alleles for VTE^{1, 17}, such as, allele T on rs2519093 SNP (*locus ABO*); allele C on rs2036914 (*locus F11*) and allele T on rs2066865 (*locus FGG*).^{1, 18, 19} Nevertheless, the genetic factors identified so far explain only 5% of VTE heritability.¹ Furthermore, association studies shown that VTE results from multigenic interaction. Thus, even each SNP associated with VTE has little predictive value, their combination may improve the predictive ability and could be used to prevent VTE.^{3, 20}

In search for novel susceptibility loci associated with VTE, the genome wide association studies (GWAS) approach, and its meta-analysis were used especially since 2008.^{1, 3} However, results mostly confirm what it was already known, in particular for genetic variants within the coagulation system including *F5*, *FXI*, *FGG*, *GP6* and *PROCR* genes.²¹ The GWA studies also confirmed the effect of *ABO* locus previously associated with VTE susceptibility.^{22, 23} As a result, in 2015, 12 genes have been robustly demonstrated to harbor common and low-frequency risk variants for VTE: *ABO*, *F2*, *F5*, *F11*, *FGG*, *GP6*, *KNGL1*, *PROCR*, *SLC44A2*, *STXBP5*, *TSPAN15* and *VWF*.¹

Among these loci, three single nucleotide polymorphism (SNP) with an odds ratio (OR) for VTE in the order of 1.5 in individuals of Caucasian origin, were not yet previously assessed in Portuguese patients: rs2519093 on *locus ABO* (OR=1.68)¹⁹; rs2036914 on *locus F11* (OR=1.35)¹⁸; rs2066865 on *locus FGG* (OR=1.47).¹

The rs2519093 SNP is located in intron 1 of the *ABO* gene (chromosome 9: 136.141.870) and the risk allele T (0.24 frequency in reference population) remained significantly associated with VTE after controlling for non-O blood type¹⁹. Moreover, this intronic region doesn't affect RNA splicing or harbor any suppressor RNA elements, it isn't close to any other *ABO* variants with known function, and it isn't in linkage disequilibrium with *ABO* non-O blood type (*ABO* rs8176719) located in exon 6.²⁴ The rs2036914 SNP is located in the intronic region of the *F11* gene (chromosome 4: 187.429.475).^{18, 25} The risk allele C (0.52 frequency

in reference population) was associated with increased levels of factor XI (FXI), which could explain, at least in part, the association with VTE.¹⁸ The rs2066865 SNP is located in 3' UTR region of *FGG* gene (chromosome 4: 155.744.726)²⁶. The risk allele T (0.25 frequency in reference population)¹ was associated with reduced γ' fibrinogen levels and γ' fibrinogen/total fibrinogen ratio below the 10th percentile, which increases VTE risk.^{3, 5, 20, 27}

The aims of this study were i) to evaluate the frequency of susceptibility alleles of the *ABO*, *F11* and *FGG* loci in a Portuguese population sample with personal history of VTE; and ii) to assess the association between risk alleles and the risk of VTE (individually and additively).

MATERIALS AND METHODS

Sample Selection

We conducted a population-based case-control study that included 95 cases and 132 controls. All enrolled patients with VTE were recruited from “Centro Hospitalar e Universitário de Coimbra, EPE” (CHUC) in Portugal between 2012 and 2015 and underwent screening for thrombophilia – with no known strong genetic risk factor for VTE including deficiencies of antithrombin, protein C or protein S, homozygous for Factor V Leiden or *F2 G20210A* mutations or composed Factor V Leiden and *F2 G20210A* heterozygosity. The cases included 25 males and 70 females, between 18 and 64 years old at the date of the first DVT or PE event. In order to maximize our ability to detect an association, we excluded VTE patients with transient or permanent environmental risk factors such as: history of cancer (5 years before and 6 months latter), major trauma or surgery (with general anesthesia lasting ≥ 30 minutes in the previous 3 months), prolonged immobilization, antiphospholipid syndrome, auto-immune diseases, heart failure, patients who were taking combined hormonal contraceptives when the VTE occurred, pregnancy and puerperium – factors that imputed a moderate/strong risk for VTE. These data were collected retrospectively from the patient’s clinical processes, being selected patients with confirmed PE or DVT.

Controls were recruited from healthy workers of CHUC and students at the University of Coimbra, without VTE prior diagnosis, being 34 males and 98 females aged between 18-64 years.

Ethical Statement

Written Informed Consent for genetic studies were obtained prior to the participation in this study from all of the individuals, in accordance with the Declaration of Helsinki. The Ethics Committee of Faculty of Medicine of University of Coimbra (Coimbra, Portugal) approved all research methodology.

Genotyping

Within the common and low-frequency susceptibility alleles robustly associated with VTE in the literature up to 2015, we selected three SNP's in the *ABO*, *F11*, *FGG* loci (rs2519093, rs2036914 and rs2066865) with a relatively high OR (~1.5) based on GWA studies conducted in individuals of Caucasian origin.¹

Peripheral blood samples were collected from all participants with the exception of the 34 samples obtained from students of University of Coimbra from which a buccal swab was obtained. Genomic DNA was extracted from EDTA whole blood by automatic isolation on iPrep™ instrument using gDNA Blood Kit (Invitrogen, Carlsbad, CA, USA). Regarding the buccal swabs, genomic DNA was extracted from buccal cells using the the Favor-Prep™ Genomic DNA Mini Kit (Favorgen Biotech Corp, Taiwan), according to instructions of the manufacturer.

Samples were genotyped for rs2036914 and rs2066865 SNPs by allelic discrimination assays using pre-designed TaqMan probes (C_12066124_10 for rs2036914 and C__11503414_10 for rs2066865) (Applied Biosystems, Foster City, CA, USA). All Polymerase Chain Reactions (PCR) were done in a final volume of 20 µL containing in 10 µL of 2x SsoAdvanced™ Universal Probes Supermix (Bio-Rad, Hercules, CA, USA), 0.5 µL of 20x

working stock of TaqMan® SNP Genotyping Assay (Applied Biosystems), 7.5 µL of DNase-free water and 2 µL (about 50 ng) of high quality genomic DNA (gDNA). Thermal cycling conditions were 10 min at 95°C, and 40 cycles each of 92°C for 15 seconds and 60°C for 1 min. The fluorescence was observed through a MiniOpticon real time PCR system (Bio-Rad, Hercules, CA, USA).

Regarding rs2519093, due to a lack of specific TaqMan probe to conduct a real time PCR, we genotyped this SNP by restriction fragment length polymorphism (RFLP) using the restriction enzyme DdeI. All samples were previously amplified by using the PCR method in a total volume of 25 µL containing: 2.5 µL of 10x Taq Buffer, 0.5 µL of dNTP Mix (10 mM each), 0.5 µL of forward (5'-CAATATCTCAGGGACCCCATAA-3') and reverse (5'-AAGGCCCAAACCATAAAAATG-3') primers (10 mM each), 3 µL of MgCl₂ (25 mM), 0.5 U of Taq DNA Polymerase (Thermo Fischer Scientific), 1 µL (25-50 ng) of gDNA and DNase-free water (up to 25 µL). The PCR cycling conditions were 5 min at 94°C (initial denaturation), and 35 cycles of 94°C for 30 seconds (denaturation), 58°C for 45 seconds (annealing) and 72°C for 45 seconds (extension), with a final extension at 72°C for 5 min. PCR amplifications were performed using a TProfessional Thermocycler (Biometra, Göttingen, Germany). Afterwards, the PCR products (208 bp) were digested using the restriction enzyme DdeI, which discriminated allele C (three restriction sites enabling fragments with 7, 62, 21 and 118 bp) from allele T (two restriction sites, enabling fragments with 7, 62 and 139 bp). These fragments were visualized by electrophoresis on 10% polyacrylamide gels.

To assess genotyping reproducibility, 10% of random samples were selected and re-genotyped for rs2036914 and rs2066865 SNPs using TaqMan SNP probes with 100% concordance. Regarding rs2036914, since the frequency of the risk allele was estimated to be 0,52 in the reference population, and to confirm the TaqMan-assigned genotypes, three

samples for each genotype identified by TaqMan assay were direct-sequenced by the Sanger's dideoxy chain termination reaction using Big-Dye Terminator v1.1 Cycle Sequencing kit (Applied Biosystems, Foster City, USA) and the ABI 3130 sequencer (Applied Biosystems, Foster City, USA) after PCR amplification using oligonucleotides 5'-AGCCATCAGATGCTGTCAGA-3' (forward) and 5'-GGCTTCAATGGAATTGTGCT-3' (reverse).

Statistical Analyses

Genotypic and allelic frequencies of all polymorphisms were estimated by direct counting and Hardy–Weinberg equilibrium probability values were achieved using an exact test. A logistic regression was conducted to evaluate the association of each individual SNP with VTE risk in an additive genetic model, in which allele dosage (0 to 2 copies of the minor allele) of imputed SNP was analyzed, estimating *p*-values, odds ratio (OR), and 95% confidence intervals (CI), crude and adjusted for sex and age. A significant *p*-value was considered below 0.017 (0.05/3) by applying a Bonferroni correction for multiple testing, and a *p*-value between 0.017 and 0.05 was considered nominally significant. All these statistical analyses were done using the set-based tests implemented on PLINK software v.1.07 (<http://pngu.mgh.harvard.edu/purcell/plink/>)²⁸.

The effect of the cumulative number of risk alleles on VTE outcome was assessed by Pearson's 2x2 contingency table chi-square test, using the Simple Interactive Statistical Analysis software (SISA, <http://www.quantitativeskills.com/sisa/>).

An exact test of sample differentiation between populations by using allele frequencies was achieved with the Arlequin software ver 3.5.2 (<http://cmpg.unibe.ch/software/arlequin35/>).

RESULTS

Population characteristics

Among 203 cases previously studied for VTE risk factors from 2012 to 2015, 95 satisfied the inclusion criteria and were admitted for analyses (Figure 1).

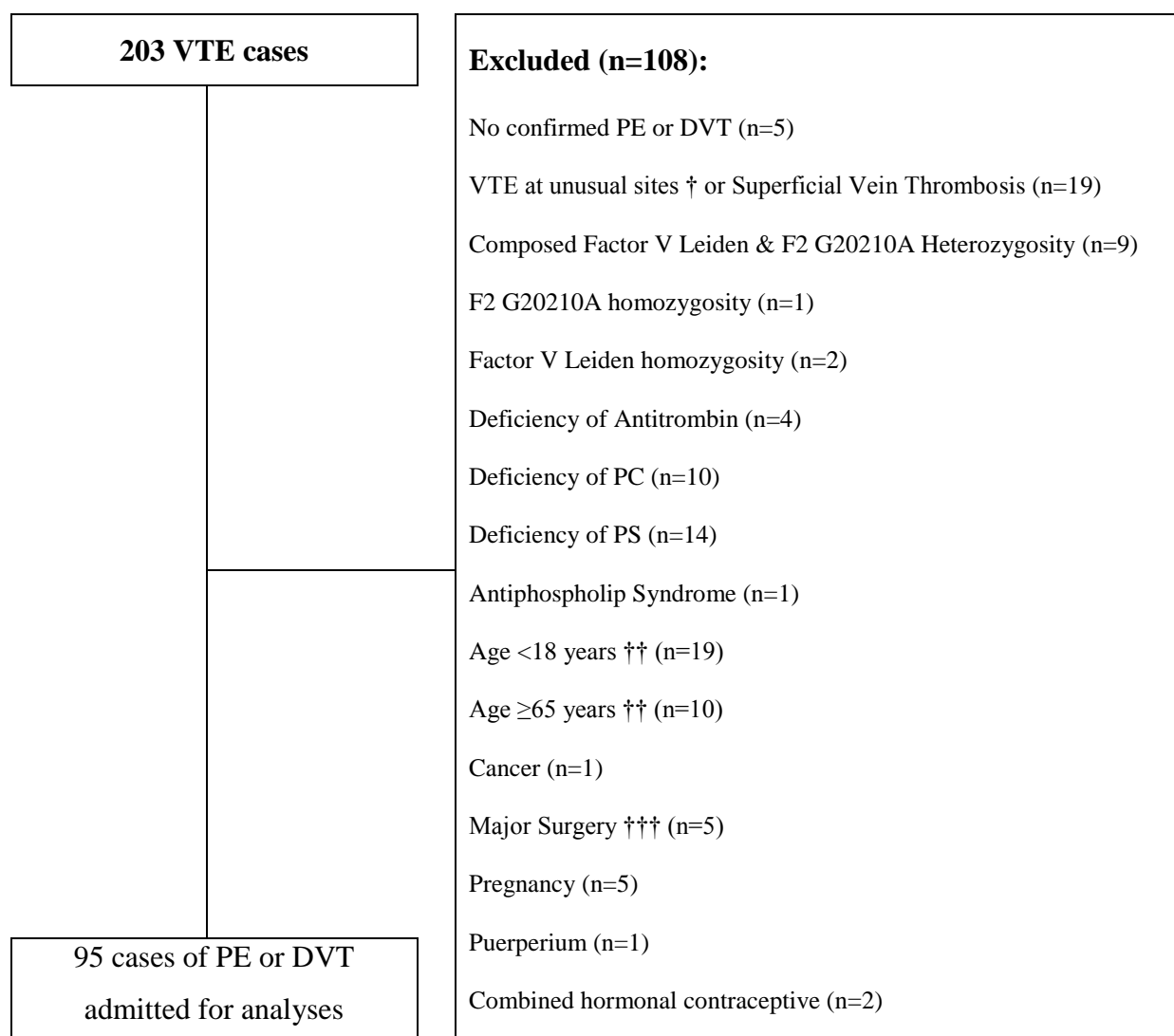


Figure 1. Sample selection according to the inclusion and exclusion criteria.

VTE – Venous Thromboembolism; DVT – Deep vein thrombosis; PE – Pulmonary embolism; †VTE at unusual sites: cerebral veins, splanchnic veins, retinal veins, jugular veins; †† - Age at the date of the first VTE event; ††† - Major Surgery with general anesthesia lasting ≥30 minutes in the previous 3 months.

The main demographic and clinical data of the 95 cases of VTE and 132 healthy controls were summarized in Table 1.

Table 1: Demographic, clinical and molecular characteristics of VTE patients and controls.

Characteristics	Cases (n=95)	Controls (n=132)
Demographic data		
<i>Gender</i>		
Female N (%)	70 (74%)	98 (74%)
Male N (%)	25 (26%)	34 (26%)
Median age – years (range)	39 (19 - 64)	34 (20 - 62)
Clinical data		
<i>Diagnosis</i>		
DVT	66	-
PE	29	-
<i>Genetic risk factors</i>		
Heterozygous Factor V Leiden	14	-
Heterozygous F2 G20210A	7	-

DVT – Deep vein thrombosis; PE – Pulmonary embolism; N – number of samples.

Evaluation of *FGG*, *ABO* and *F11* SNP allelic frequencies

The genotyping success rate was 100% for the three studied SNPs rs2519093 on locus *ABO*, rs2036914 on locus *F11* and rs2066865 on locus *FGG*. Regarding the rs2519093 since the DdeI restriction enzyme created an additional restriction site on major allele C we were able to discriminate the three genotypes (Figure 2).

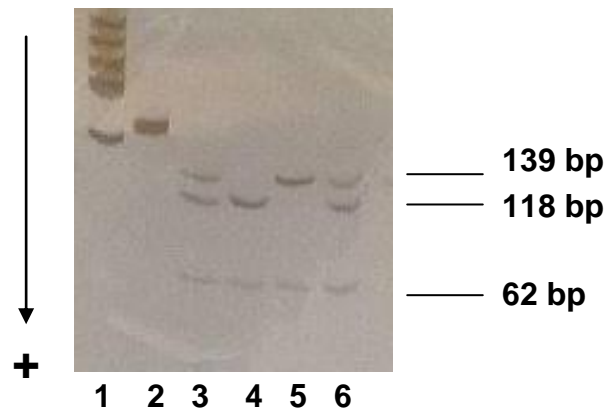


Figure 2. RFLP gel electrophoresis using DdeI restriction enzyme for *ABO* rs2519093

SNP. The photo shows the restriction bands of 62 and 118 bp for allele C and 62 and 139 bp band for allele T (the 6 and 21 bp bands for allele C and 7 bp band for allele T were not shown because of the small size). Lines: 1, 2 kb ladder marker (nzytech); 2, PCR product (208 pb); 3 and 6, heterozygous subjects (genotype CT); 4, homozygous CC; 5, homozygous TT.

Allelic frequencies observed for the three studied SNPs in the total population, cases and controls are displayed in Table 2. Risk allele frequencies in the total population were 30.2% for rs2519093 on locus *ABO*, 63% for rs2036914 on locus *F11* and 19.4% for rs2066865 on locus *FGG*. The genotype distributions were in agreement with the Hardy-Weinberg equilibrium ($p > 0.05$) for all SNPs in the total population (Table 2), cases (*ABO* rs2519093 $p = 1.0$; *F11* rs2036914 $p = 0.38$; *FGG* rs2066865 $p = 0.58$) and controls; (*ABO* rs2519093 $p = 0.28$; *F11* rs2036914 $p = 0.85$; *FGG* rs2066865 $p = 0.31$).

Analysis of the SNP in *FGG*, *ABO* and *F11* genes with the risk of VTE

To test the association between individual SNPs and VTE we used a logistic regression, under an additive model, as shown in Table 2. We can observe a nominal significant association between rs2066865 on the *FGG* locus and VTE, unadjusted (OR=1.71; $p = 0.021$) as also

using age and sex as covariates (OR=1.70; $p=0.024$), since age and sex have been suggested to be a risk factor for VTE^{4, 29}. Furthermore, the *FGG* rs2066865 remained significant in this population assuming a dominant model (OR=1.94; $p=0.02$) and after adjusting for age and sex (OR=1.93; $p=0.025$). The *ABO* rs2519093 polymorphism attained near marginal levels of nominal significance unadjusted (OR=1.42, $p=0.10$) and adjusted to age and sex (OR=1.46, $p=0.087$). The *F11* rs2036914 polymorphism did not reach statistical significant association with VTE, crude (OR=0.91; $p=0.64$) or after adjusting for sex and age (OR=0.82; $p=0.33$). After Bonferroni correction for multiple testing ($p=0.017$ ($0.05/3$)), only a marginally significant p -value ($p=0.02$) was found for the *FGG* minor T-allele.

Table 2: Association of SNP allelic frequencies in *ABO*, *FGG* and *F11* genes with the risk of

		VTE						
Polymorphisms	Alleles	Total (N=227)	p -HWE	Cases (N=95)	Controls (N=132)	OR (CI 95%)	p (a)	p (b)
<i>ABO</i>	<u>T</u>	0.302	0.436	0.342	0.273	1.42 (0.93-2.16)	0.10	0.087
rs2519093	C	0.698		0.658	0.727			
<i>F11</i>	T	0.370	0.476	0.358	0.379	0.91 (0.61-1.36)	0.64	0.33
rs2036914	<u>C</u>	0.630		0.642	0.621			
<i>FGG</i>	<u>T</u>	0.194	0.139	0.247	0.155	1.71 (1.08-2.69)	<u>0.021</u>	<u>0.024</u>
rs2066865	C	0.806		0.753	0.845			

The study was performed considering individuals with VTE vs controls. †HWE – Hardy-Weinberg Equilibrium; OR – odds ratio; CI – confidence interval; N – number of samples. The OR and p value were obtained by logistic regression under an additive model. The OR was estimated for the minor allele and p values were unadjusted (a) and adjusted for age and sex (b). Nominal significant results ($p<0.05$) are in bold and underlined. The risk alleles for each SNP are in bold and underlined.

Analysis of the association between the total number of risk alleles in *FGG* and *ABO* genes with the risk of VTE

Considering the total number of risk alleles in each individual for both polymorphism in *FGG* and *ABO* genes, that showed significant and near marginal significant association, respectively, Pearson χ^2 *P* values indicated that a higher number of risk alleles increased the risk for VTE. It was not found a statistically significant difference whether the individuals had one risk allele or none (0 risk alleles vs. 1 risk allele: $\chi^2=1.98$, $p=0.159$), however when testing 0 risk alleles vs. 2-4 risk alleles a statistically significant value was found: $\chi^2= 5.50$ ($p=0.019$). According to the contingency table 2x2, the presence of 2 or more risk alleles increased 2.31 times the risk for VTE (OR=2.31; 95% CI 1.14 – 4.68), compared with zero risk alleles (Table 3).

Table 3: Association between the total number of risk alleles and the risk for VTE

	Number of risk alleles			Total
	0	1	≥ 2	
Group 1:Control	52	53	27	132
Group 2:VTE	25	40	30	95
Model 1	Chi sq= 1.98 ($p=0.159$)			
(0 vs. 1 risk allele)	OR=1.57 (95% CI 0.837 – 2.95)			
Model 2	Chi sq=5.50 ($p=0.019$)			
(0 vs. 2-4 risk alleles)	OR=2.31 (95% CI 1.14 – 4.68)			

OR – odds ratio; CI – confidence interval; Chi sq – chi-square

DISCUSSION AND CONCLUSIONS

In this work we report a study of three polymorphisms located at the *locus ABO*, *F11* and *FGG*, previously found associated with VTE in individuals of Caucasian origin.¹ The obtained allele frequencies in the control population sample for *F11* rs2036914 and *FGG* rs2066865 SNPs, were found significantly different from previous studies using European-descendent individuals: the risk allele for *F11* rs2036914 (C) showed a higher frequency in the Portuguese population (0.62 vs. 0.52)¹⁸ (Exact *p*-value = 0.02), and the risk allele for *FGG* rs2066865 (T) showed a lower frequency in the Portuguese population (0.155 vs. 0.25)¹ (Exact *p*-value = 0.019). The frequency of the *ABO* rs2519093 was in conformity with former studies in Caucasian populations (minor allele frequency (MAF) 0.27 vs. 0.24)¹⁸ (Exact *p*-value = 0.389).

Among the three studied loci, the *FGG* rs2066865 minor T-allele was found to be significantly associated (nominal *p*=0.02) with VTE in the Portuguese population (marginally significant after Bonferroni correction for multiple testing). This result is in accordance with previous studies with populations of Caucasian origin reporting similar significant associations with VTE for *FGG*.¹ The association between *ABO* rs2519093 SNP and VTE attained near marginal levels of nominal significance (*p*=0.087 adjusted for age and sex). This suggests that the *ABO* SNP may be associated with VTE in the Portuguese population, resembling results from previous studies in European populations but further testing with a larger sample size is needed to confirm this association.¹ Regarding the *F11* rs2036914, this SNP never reach statistical significance in the Portuguese study sample (*p*=0.64). Despite having gathered a subject sampling sufficiently random, as suggested by the HWE *p*-values, the sample size probably did not provide sufficient power to detect significant association, or

this common SNP have a more modest effect than *FGG* or *ABO* SNPs. The higher frequency of the rs2036914 risk allele observed in the Portuguese population in relation to other populations of European origin may also explain the absence of significant association as the differences in population genetic backgrounds are strong candidates to explain heterogeneity in disease-association studies between populations.³⁰

In addition, we observed a statistical association between the number of risk alleles and VTE risk when considering the *FGG* and *ABO* risk alleles, with an OR=2.31 for 2 or more risk alleles. In this test an additive effect was assumed for the different genotypes, however we cannot exclude a gene-gene and gene-environment interactions.²⁰ Since VTE is a multifactorial disease, each individual SNP have little predictive value as a result of their modest effect on risk, but the combination of gene variants may improve the predictive ability for VTE risk.^{3, 20} For this matter, several studies have constructed genetic risk scores to assess the overall genetic predisposition to VTE risk.³¹ De Haan et al. calculated a genetic risk score based on 31 SNPs consistently associated with VTE to predict the first VTE event. This study with 2712 cases and 4634 controls concluded that the addition of the 5 most strongly associated SNPs risk score to a risk scoring system based on nongenetic risk factors significantly improved the VTE risk prediction (with a 100-fold difference in risk between the lowest and highest number of risk alleles).²⁰ The rs2066865 (*FGG*) and rs2036914 (*F11*) SNPs were present among the five strongly associated (which also included rs6025 (Factor V Leiden, rs1799963 (F2 G20210A) and *ABO* rs8176719).²⁰ Bruzelius et al. used a combination of 18 SNPs (including *FGG* rs2066865 and *F11* rs2036914) with clinical risk factors in a case-control study with 1443 female cases and 1402 controls to predict the first occurrence of VTE.³² The results attained in both studies suggest that the genetic risk score (that included *FGG* rs2066865 and *F11* rs2036914 SNPs) predicts the first VTE event more reliably than the family history alone.^{20, 32} However, in Cohen et al study, that used a cohort of family

members, the major predictor of the VTE risk was the family history. Nevertheless, Cohen et al suggested that the inclusion of other variables, such as risk allele T of *FGG* rs2066865, allowed a more accurate definition of the individual VTE risk within these families.³³ In addition, Van Hylckama Vlieg et al.³⁴ showed that multiple genetic SNP analysis is also useful in the prediction of recurrent thromboembolism. Using a 5-SNP genetic risk score (same as De Haan et al study), including *FGG* rs2066865 and *F11* rs2036914, the score was able to stratify patients into high or low risk of recurrence, with an over 2-fold difference. Since long-term anticoagulation is associated with a risk of major hemorrhage of 2% per year, this assessment might be useful to determine those whom this therapy might be safely discontinued.³⁴

Even though the VTE incidence is too low in the general population to justify genotyping for SNPs consistently associated with VTE, in the coming era of personalized medicine, the number of risk alleles may be useful in combination with nongenetic risk factors for a more precise clinical decision in high-risk populations.^{20, 31}

Some limitations of the present study need to be taken into account when interpreting the results obtained. This was a small population-based study, with a retrospective data collection. Thus there maybe some confounding factors that it was not possible to control within the study. Even though the Hardy-Weinberg analysis did not reveal any evidence of population disequilibrium and included 95 unprovoked cases of VTE, the diminished total size sample does not allow the generalisability of the study results. To overcome these limitations, it would be interesting to conduct a prospective study with a larger sample size.

Since individual risk to VTE is affected by multiple acquired and inherited factors³³ and their interactions³¹, in the future the physician might need to evaluate the presence of common and low-susceptibility SNPs in combination with other known factors to assess the risk of VTE. Furthermore, according to global ENDORSE study, in Portugal there is a high prevalence of

hospitalized patients at risk for VTE (52.7%) and less than two-thirds of them receive appropriate prophylaxis according to the American College of Chest Physicians (64.2% of the surgical patients and 64.5% of medical patients).⁸ Thereby, a national, multicentric and prospective study would be interesting to validate a risk assessment model (with acquired and inherited factors, including common SNPs) in a Portuguese population, similarly to the ARTE study³⁵.

In conclusion, our results suggest that risk allele T on both *FGG* rs2066865 and *ABO* rs2519093 SNPs may contribute to the risk of VTE in this Portuguese population sample, individually and additively. In spite of the limitations of the study and the modest effect of each studied SNP, the data collected emphasizes the need to incorporate common and low-frequency susceptibility SNPs in validated risk assessment models for VTE.

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest in this paper.

ACKNOWLEDGMENTS

I thank all the help provided by my tutors and Centro de Investigação em Antropologia e Saúde (CIAS) – Universidade de Coimbra who funded this study, without which this work would not be possible.

REFERENCES

1. Morange PE, Suchon P, Tregouet DA. Genetics of Venous Thrombosis: update in 2015. *Thromb Haemost* 2015;114(5):910-9.
2. Walton BL, Byrnes JR, Wolberg AS. Fibrinogen, red blood cells, and factor XIII in venous thrombosis. *J Thromb Haemost* 2015;13 Suppl 1:S208-15.
3. Morange PE, Tregouet DA. Current knowledge on the genetics of incident venous thrombosis. *J Thromb Haemost* 2013;11 Suppl 1:111-21.
4. Heit JA. Epidemiology of venous thromboembolism. *Nat Rev Cardiol* 2015;12(8):464-74.
5. Tang W, Teichert M, Chasman DI, Heit JA, Morange PE, Li G, et al. A genome-wide association study for venous thromboembolism: the extended cohorts for heart and aging research in genomic epidemiology (CHARGE) consortium. *Genet Epidemiol* 2013;37(5):512-21.
6. Konstantinides SV, Torbicki A, Agnelli G, Danchin N, Fitzmaurice D, Galie N, et al. 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism. *Eur Heart J* 2014;35(43):3033-69, 3069a-3069k.
7. Cohen AT, Agnelli G, Anderson FA, Arcelus JI, Bergqvist D, Brecht JG, et al. Venous thromboembolism (VTE) in Europe. The number of VTE events and associated morbidity and mortality. In: *Thromb Haemost*. Germany; 2007. p. 756-64.
8. Franca A, Reis A, Paulino A, Lohman C, Cartucho D, Campello G, et al. Venous thromboembolism risk factors and practices of prophylaxis: ENDORSE study results in Portugal. *Acta Med Port* 2011;24(6):951-60.

9. Reis A. [Prevention and treatment of venous thromboembolism: the place of new oral anticoagulants]. *Rev Port Cardiol* 2012;31 Suppl 1:45-50.
10. Germain M, Chasman DI, de Haan H, Tang W, Lindstrom 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;96(4):532-42.
11. Wells P, Anderson D. The diagnosis and treatment of venous thromboembolism. *Hematology Am Soc Hematol Educ Program* 2013;2013:457-63.
12. Martinelli I, De Stefano V, Mannucci PM. Inherited risk factors for venous thromboembolism. *Nat Rev Cardiol* 2014;11(3):140-56.
13. Stevens SM, Woller SC, Bauer KA, Kasthuri R, Cushman M, Streiff M, et al. Guidance for the evaluation and treatment of hereditary and acquired thrombophilia. *J Thromb Thrombolysis* 2016;41(1):154-64.
14. Goldhaber SZ, Bounameaux H. Pulmonary embolism and deep vein thrombosis. *Lancet* 2012;379(9828):1835-46.
15. Kreidy R. Influence of acquired and genetic risk factors on the prevention, management, and treatment of thromboembolic disease. *Int J Vasc Med* 2014;2014:859726.
16. Kearon C, Ageno W, Cannegieter SC, Cosmi B, Geersing GJ, Kyrle PA. Categorization of patients as having provoked or unprovoked venous thromboembolism: guidance from the SSC of ISTH. *J Thromb Haemost* 2016;14(7):1480-3.
17. Rosendaal FR, Reitsma PH. Genetics of venous thrombosis. *J Thromb Haemost* 2009;7 Suppl 1:301-4.

18. Li Y, Bezemer ID, Rowland CM, Tong CH, Arellano AR, Catanese JJ, et al. Genetic variants associated with deep vein thrombosis: the F11 locus. *J Thromb Haemost* 2009;7(11):1802-8.
19. Heit JA, Armasu SM, Asmann YW, Cunningham JM, Matsumoto ME, Petterson TM, 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):1521-31.
20. 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.
21. Reitsma PH. Genetics in thrombophilia. An update. *Hamostaseologie* 2015;35(1):47-51.
22. Jick H, Slone D, Westerholm B, Inman WH, Vessey MP, Shapiro S, et al. Venous thromboembolic disease and ABO blood type. A cooperative study ABO blood group genotypes and the risk of venous thrombosis: effect of factor V Leiden. In: *Lancet*. England; 1969. p. 539-42.
23. Morelli VM, De Visser MC, Vos HL, Bertina RM, Rosendaal FR. ABO blood group genotypes and the risk of venous thrombosis: effect of factor V Leiden. In: *J Thromb Haemost*. England; 2005. p. 183-5.
24. Heit JA, Cunningham JM, Petterson TM, Armasu SM, Rider DN, M DEA. Genetic variation within the anticoagulant, procoagulant, fibrinolytic and innate immunity pathways as risk factors for venous thromboembolism. *J Thromb Haemost* 2011;9(6):1133-42.
25. El-Galaly TC, Severinsen MT, Overvad K, Steffensen R, Vistisen AK, Tjønneland A, et al. Single nucleotide polymorphisms and the risk of venous thrombosis: results from a Danish case-cohort study. *Br J Haematol* 2013;160(6):838-41.

26. Germain M, Saut N, Greliche N, Dina C, Lambert JC, Perret C, et al. Genetics of venous thrombosis: insights from a new genome wide association study. *PLoS One* 2011;6(9):e25581.
27. Uitte de Willige S, de Visser MC, Houwing-Duistermaat JJ, Rosendaal FR, Vos HL, Bertina RM. Genetic variation in the fibrinogen gamma gene increases the risk for deep venous thrombosis by reducing plasma fibrinogen gamma' levels. *Blood* 2005;106(13):4176-83.
28. 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.
29. Roach RE, Lijfering WM, Tait RC, Baglin T, Kyrle PA, Cannegieter SC, et al. Sex difference in the risk of recurrent venous thrombosis: a detailed analysis in four European cohorts. *J Thromb Haemost* 2015;13(10):1815-22.
30. Myles S, Davison D, Barrett J, Stoneking M, Timpson N. Worldwide population differentiation at disease-associated SNPs. *BMC Med Genomics* 2008;1:22.
31. Crous-Bou M, Harrington LB, Kabrhel C. Environmental and Genetic Risk Factors Associated with Venous Thromboembolism. *Semin Thromb Hemost* 2016.
32. Bruzelius M, Bottai M, Sabater-Lleal M, Strawbridge RJ, Bergendal A, Silveira A, et al. Predicting venous thrombosis in women using a combination of genetic markers and clinical risk factors. *J Thromb Haemost* 2015;13(2):219-27.
33. Cohen W, Castelli C, Suchon P, Bouvet S, Aillaud MF, Brunet D, et al. Risk assessment of venous thrombosis in families with known hereditary thrombophilia: the MARseilles-NImes prediction model. *J Thromb Haemost* 2014;12(2):138-146.
34. van Hylckama Vlieg A, Flinterman LE, Bare LA, Cannegieter SC, Reitsma PH, Arellano AR, et al. Genetic variations associated with recurrent venous thrombosis. In:

Circ Cardiovasc Genet. United States: 2014 American Heart Association, Inc.; 2014.
p. 806-13.

35. Franca A, De Sousa JA, Felicissimo P, Ferreira D. [Venous thromboembolism's risk assessment: rationale, objectives, and methodology--the ARTE study]. Acta Med Port 2011;24 Suppl 2:575-82.