

TITLE

Familial thrombotic risk based on the genetic background of Protein C Deficiency in a Portuguese Study

RUNNING TITLE

Genetic *background* of PC deficiency as a risk factor for VT

Teresa Fidalgo, Patrícia Martinho, Ramon Salvado, Licínio Manco, Ana C. Oliveira, Catarina S. Pinto, Elsa Gonçalves, Dalila Marques, Teresa Sevivas, Natália Martins and Maria Letícia Ribeiro

Teresa Fidalgo	Corresponding author: teresafidalgo@netcabo.pt Departamento de Hematologia - Unidade de Hemostase. Hospital Pediátrico de Coimbra, Av Afonso Romão , 3000-602 Coimbra, Portugal Tel: +351 239 480 370, Fax: +351 239 717 216,	Department of Haematology Centro Hospitalar e Universitário de Coimbra (CHUC)
Patrícia Martinho	pmartinhosimoes@gmail.com	
Ramon Salvado	ramon.salvado@hotmail.com	
Licínio Manco	Imanco@antrop.uc.pt	Dep of Life Sciences University of Coimbra
Ana C. Oliveira	acristina_oliveira@sapo.pt	
Catarina S. Pinto	c.silvapinto@sapo.pt	
Elsa Gonçalves	egoncalves.bioq@gmail.com	
Dalila Marques	dalila.f.marques@gmail.com	Department of Haematology
Teresa Sevivas	teresa.sevivas@hotmail.com	Centro Hospitalar e Universitário de Coimbra (CHUC)
Natália Martins	nataliamartins@chc.min-saude.pt	
Maria Letícia Ribeiro	leticia.ribeiro@chc.min-saude.pt	

Abstract:

Introduction: Inherited Protein C (PC) deficiency is a well-known risk factor for venous thrombosis (VT). Plasma PC levels are reliable in moderate to severe deficiencies; however, in mildly deficient individuals, the levels may overlap with those considered normal. Genetic studies of *PROC*, which encodes PC, could help identify carriers; genome-wide association studies (GWAS) have shown that approximately 50% of phenotypic variation in PC deficiency is caused by the cumulative effects of mutations in several other *loci*, namely in the *PROCR*.

Patients and Methods: With the main objective of determining the genotype/phenotype correlation in 59 Portuguese individuals from 26 unrelated families with history of thrombosis and repeatedly low/borderline PC plasma levels, we conducted a molecular study by direct sequencing of *PROC*; *PROC* promoter haplotypes and *PROCR* c.4600A>G polymorphism (rs867186), which are known to influence plasma PC concentrations, were also screened.

Results: Twelve different *PROC* mutations were identified, one of them not previously reported, p.Cys105Arg. The mutation types and locations as well as haplotype combinations correlated with the phenotypic severity. The most frequent mutation, p.Arg199X, correlated with the CGTC haplotype and was identified in 9 families containing patients with higher numbers of VT episodes. This mutation in homozygous individuals for the CGTG haplotype is a significant risk factor for VT in Portuguese.

Conclusion: These genetic family studies allowed the identification of the unknown carriers and individuals at a higher thrombotic risk within each family, thus permitting the evaluation of the need for prophylactic measures, particularly in at-risk situations.

Key words: Hereditary protein C deficiency, genotype, phenotype, risk factor, venous thrombosis

Introduction

Protein C (PC) is a serine-protease that belongs to the vitamin K-dependent plasma glycoprotein and zymogen group, which plays a major regulatory role in blood coagulation and thrombosis. PC is activated by the thrombin/thrombomodulin complex on the surface of endothelial cells, where it binds with an endothelial PC receptor (EPCR) and proteolytically inactivates factors Va (FVa) and FVIIIa in the presence of protein S, phospholipids and calcium (1).

Numerous mutations in the gene coding for PC (*PROC*) on chromosome 2q14, lead to several types of PC deficiency (OMIM #176860) (2): in type I deficiency levels of both antigen and activity are reduced, whereas in type II deficiency antigen levels are normal, but one or more functional defects lead to a decreased activity (3). Hereditary PC deficiency is mainly an autosomal dominant inherited disorder and is an established risk factor for venous thrombosis (VT) (4), arterial ischemic stroke in children (5) and has been associated to arterial thrombosis events at a young age (6). The relative risk increases as the PC concentration falls, with an odds ratio (OR) of ≈ 4 in individuals with PC levels of <65% (7). PC deficiency appears to be quite rare, with a prevalence of 0.14–0.50% in the general population and 2.5 to 6% in patients with venous thrombosis. Heterozygous individuals have an approximately 7-fold increased risk of VT compared with normal individuals (8). The rarer form of PC deficiency, severe homozygous (or compound heterozygous), was first described as a cause of neonatal *purpura fulminans* resulting in rapid death caused by massive venous thrombosis (9,10). However, a number of adult patients with very low or undetectable plasma PC levels have been found; they exhibit milder symptoms and/or a later onset of the disease (11).

The diagnosis of PC deficiency based on plasma measurements is often difficult in patients with borderline plasma PC levels, due to the fact that not all heterozygotes for a given mutation have a low plasma PC level and because of the large overlap between heterozygotes and non-carriers (3,7,12,13). It should also be noted that the decrease in plasma PC concentrations is influenced by age, where the PC concentration level may remain below the adult reference range until adolescence (14,15).

Genetic studies offer a powerful complementary tool, although the alterations found in *PROC* are due to either rare mutations or common single nucleotide polymorphisms (SNPs). However, in about 10% to 30% of families with PC deficiency the mutations are not detectable (16).

The SNPs within the 5' region of *PROC*, that causally contribute to down transcriptional regulation, show a correlation with low plasma levels of PC and are associated with adverse outcomes in a variety of clinical states. These SNPs, that extends from the promoter region to the end of intron 2, -1657C/T (rs1799808); -1644A/G (rs1799809); -1479A/T (rs1799810) and -141T/C (rs1158867), show a high degree of *linkage disequilibrium* (LD) (7,12,17). The CGT haplotype of -1657C/T, -1644 A/G and -1479A/T is generally related to lower plasma PC levels than the complementary TAA haplotype and to a higher risk of thrombotic events. The transcriptional efficiency of *PROC* in these haplotypes has been described as driven by the SNPs at positions -1657 and -1644. The other 2 SNPs, -1479A/T and -141T/C, are in complete LD ($r^2=1$) and may be involved in regulating splicing events with moderate thrombotic risk (17).

The search for additional polymorphisms, through the genome-wide association studies (GWAS), has been conducted for loci affecting plasma PC levels and it has been found that $\approx 50\%$ of the phenotypic variation is caused by the additive effect of SNPs in other genes (18,19). With this in mind, many studies have been done but only SNPs in the EPCR gene

(*PROCR*) located on chromosome 20q11.2 have been robustly found to be associated with the risk of VT in diverse GWAS (18,20,21). The *PROCR* polymorphism that shows the highest association with PC levels is the c.4600A>G (rs867186), causing a Serine to Glycine substitution at amino acid 219 (p.S219G), located in the transmembrane domain of the EPCR protein. This functional variant showed increased soluble EPCR plasma levels, which explains the increase of PC levels in plasma and contributes to about 10% of their variability (22,23).

PC deficiency can also be found in association with other genetic defects predisposing for VT including the Factor V Leiden (F5L) and prothrombin G20210A gene polymorphisms. It has been shown that individuals with both PC deficiency and F5L have a higher risk of developing thrombosis at a younger age as compared with carriers of PC deficiency alone (24,25). In individuals with PC deficiency, environmental factors (e.g. pregnancy, *puerperium*, oral contraceptives, surgery, immobilization, and trauma) are cumulative high-risk factors for thrombosis.

In order to determine the genotype/phenotype correlation of 59 Portuguese individuals, from 26 unrelated families with a history of personal and/or familial thrombosis, who had repeated low or borderline PC plasma levels, we carried out the molecular study and the haplotype analysis for those polymorphisms which are known to influence the plasma PC concentration.

Patients and Methods

This study includes a group of 59 patients of Portuguese origin belonging to 26 apparently unrelated families, diagnosed with PC deficiency between 2002 and 2012, attending the Department of Haematology at the Centro Hospitalar de Coimbra, Portugal.

Clinical description of the probands and their families

The probands were attended in our Center for thrombophilia screening after the thrombotic event or because of their family history of thrombosis. The clinical characteristics of patients were recorded, through a validated questionnaire, focusing on personal and family history of thrombosis disease and acquired risk. Among the 26 probands, 22 had a personal history of thrombosis: 2 had had arterial thrombosis; 2 obstetrical complications; 13 venous thromboembolism (11 lower limbs deep vein thrombosis, 1 lower limbs deep vein thrombosis with pulmonary embolism and 1 portal vein thrombosis) and 5 superficial vein thrombosis. The remaining 4 probands were asymptomatic and studied because first- and second-degree relatives had had venous thrombosis (The detailed list of clinical data can be found in Table 1).

The mean age of the probands at diagnosis was 39.27 ± 17.25 years with a sex distribution 17F:9M; the mean age at the time of the first thrombotic episode was 24.67 ± 16.94 years; 11 (42%) had experienced more than one thrombotic episode. Nineteen probands had acquired risk factors: contraceptive drugs, varicose veins, trauma, pregnancy, dyslipidemia, hypertension and immobilization. Three probands aged 24.33 ± 7.57 of whom had more than one thrombotic episode, had no identified risk factors and identified risk factors and were considered as having had an idiopathic deep venous thrombosis (DVT) (Table 1). Of the 26 probands, fourteen (53.8%) had a positive family history of VT; an examination of their relatives revealed 33 with PC deficiency but only 7 (21%) had had thrombotic episodes.

Controls

Fifty two healthy volunteers acted as a control group for measuring of PC activity/antigen levels of adult's normal range, and for determining the *PROC* promoter haplotypes and p.S219G genotypes in *PROCR*. The control group include unrelated individuals from laboratory and medical staff (38F:14M), without an individual or family history of thrombosis, with a mean age of 33.5 years \pm 11.12. Absence of thrombotic events or a family history of thrombosis was verified by means of a validated questionnaire. The thrombophilic study was also carried out in the control group subjects.

In accordance with the Helsinki Declaration, informed consent was obtained from patients and controls.

Samples and sample processing

All samples of thrombophilia patients were taken with a minimum of three months after the last episode of thrombosis and off anticoagulant therapy for at least 30 days.

Functional and immunological studies were performed on blood collected into vacuum tubes containing 3.2% sodium citrate, and centrifuged within 15 min at room temperature for 20 min at 2500 *g*, the obtained platelet poor plasma is then separated into aliquots and kept frozen at -80°C until use. The separated plasma was later subjected to the functional thrombophilia tests.

Genomic DNA was extracted from EDTA whole blood by automatic isolation on "iPrepTM" instrument using gDNA Blood Kit (Invitrogen, Carlsbad, USA).

Thrombophilia screening

The following tests were performed in all probands and their relatives:

(i) *Functional thrombophilia tests* - Antithrombin activity, PC amidolytic activity, Protein S free and detection of Lupus Anticoagulant run on ACL TOP (Instrumentation Laboratory,

Milan, Italy). PC antigen, Protein S total were determined by enzyme-linked immunosorbent assay (DG-EIA– Grifols, Corgenix™).

(ii) Molecular thrombophilia tests - The screening of F5L and the prothrombin 20210A allele was carried out using modified Multiplex Allele Specific – PCR (polymerase chain reaction).

Diagnosis of PC deficiency

The PC amidolytic activity was determined using a chromogenic end-point assay (HemosIL™ Protein C, Instrumentation Laboratory, Milano, Italy), because it is less prone to interferences (26); although when a low ratio of activity to antigen and in order to detect deficiencies type II, the PC anticoagulant activity was also determined by a coagulation end-point assay using ProClot assay (Instrumentation Laboratory), derived from Protac® performed in accordance with the manufacturer's instructions.

The adult's normal ranges were established in 52 healthy controls (mean \pm 2SD): PC activity 73% to 143% and PC antigen 71% to 135%. The sample values of children and adolescents were compared with the published pediatrics reference ranges with standardized age groups (14,15).

The diagnosis of potential probands with deficiency of PC was established on the basis of the following criteria: plasma levels of PC below the lower normal range cut-off values in at least 2 different samples.

PROC and PROCR genotype analysis

All exons, intronic boundaries, 5' and 3' UTRs of the *PROC* gene were PCR amplified using primers as described previously (27) with the exception of the following primers used for amplification of: (i) *PROC* promoter SNP's rs1799808, rs1799809 and rs1799810: Prm_5'-CCTCCCCTGCCCGCAGA-3'; Prm_5'-CGTGATTCCTGGGCGATGTATT-3'; (ii) SNP

rs1158867: Int1_5'-ACTGCATTCTGGAGCTGCTT; Int1_GAAAGGGCTGTCCCTGTGT; (iii) *PROCR* SNP rs867186: SG_5'- GCTCCTA CACTTCGCTGGTC; SG_5'-GAGCTGAAACTTTCC CTTGC. After amplification, the PCR products for all samples were sequenced on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) using Big Dye Terminator v1.1 Cycle Sequencing Chemistry (Applied Biosystems), according to protocols recommended by the manufacturer.

DNA samples from patients lacking identified mutations using PCR/direct sequencing were screened using multiplexed ligation-dependent probe amplification (MLPA) method with the Kit P265-B1 *PROC* (MRC, Amsterdam, Netherlands).

New sequence variants and amino acid changes were designated according to current nomenclature and the recommendations of the Human Genome Variation Society (HGVS) - accession number JX030036.

In silico analysis of amino acid changes

The possible impact of coding sequence changes (amino acid substitutions) on the structure and function of PC was assessed using three bioinformatics tools, Sorting Intolerant From Tolerant (SIFT, <http://sift.jcvi.org>), Polymorphism Phenotyping-2 (PolyPhen-2, <http://genetics.bwh.harvard.edu/pph2>) and MutationTaster (<http://mutationtaster.org/>).

Statistical Analysis

For continuous variables, differences between groups were analyzed by a Student's t-test or Mann-Whitney *U*-test, depending on the normality of the data. A distribution of haplotypes between the studied groups was analyzed using Fisher's exact test. Tests were assumed significant whenever at 2-tailed $p < 0.05$. These statistical analyses were performed using GraphPad Prism 6.02 for Windows (GraphPad Software, La Jolla California USA, www.graphpad.com).

Allele frequencies estimated by simple counting, Hardy-Weinberg equilibrium probability values, as well as linkage phase statistical inference from diploid data with the (Bayesian) ELB algorithm and haplotype frequencies, were obtained using the software Arlequin, version 3.01 (Excoffier et al., 2005; <http://cmpg.unibe.ch/software/arlequin3/>).

Results

Laboratory haemostasis findings

Forty-three patients were determined to have plasma PC activity levels below 60%; the remaining 16 patients received oral anticoagulation (n = 4) or had slightly decreased values (60%–73%) along with family histories that suggested PC deficiency (n = 12). Therefore, 57 individuals were classified with type I PC deficiency and 2 with type IIa deficiency (PC22; PC24) with normal antigen concentrations and reductions in both amidolytic and anticoagulant activities.

These 59 patients with suspected PC deficiency were subjected to *PROC* gene analysis. Twelve different mutations were identified, of which only 1 was not previously reported in the Human Gene Mutation Database (HGMD). The majority of these were missense (9/12; Fig. 1), and included p.Cys105Arg (not previously described), p.Pro210Leu, p.Gly215Glu, p.Arg220Gln, p.Gly239Arg, p.Val339Met, p.Pro369Leu, p.Arg394Trp and p.Trp444Cys, whereas the remaining 3 were a nucleotide change (-13A>G) at the promoter region, a nonsense mutation (p.Arg199X) and a small deletion (p.Leu212Hisfs*2). No mutations were found in 1 family with low PC levels. Patients among the 60%–73% of those with PC activity levels overlapping the minimal values of the control group mostly belonged to the group with type 1 missense mutations (Fig. 2).

Screening for other inherited thrombophilias revealed the co-inheritance of heterozygous F5L in 5 individuals from 3 families (PC13, PC18, PC23) and heterozygosity for a prothrombin G20210A variant in 3 individuals from 2 families (PC13, PC14). One patient was homozygous for F5L (PC23) and another was heterozygous for F5L and prothrombin G20210A (PC13.I).

Influence of PROC and PROCR polymorphisms on plasma PC levels

The distribution of *PROC* promoter haplotypes based on the SNPs -1657C/T, -1644A/G, -1479A/T and -141T/C as well as the c.4600A>G, p.S219G (*PROCR*) genotypes among the 52 controls and 26 probands is shown in Table 2. Genotype distributions among the control group were in agreement with Hardy–Weinberg equilibrium for all 5 polymorphisms ($p > 0.05$). Although the CGTC, TAAT and CAAT *PROC* promoter haplotypes were observed in both controls and patients, the CGTC frequency was significantly higher among patients than among controls ($p = 0.005$). However, the *PROCR* c.4600A>G polymorphism allele frequencies were similar in both groups (Minor Allele Frequency: controls 0.106 vs. patients 0.096; $p=0.713$).

We further identified the haplotypes associated with each *PROC* mutation using the 4 evaluated *PROC* promoter SNPs (Table 3). The 3 different haplotypes, CAAT, TAAT and CGTC, were observed both in normal and mutation carrier chromosomes. Mutations that occur more than once in unrelated probands were always found in the context of a single haplotype. From the twelve identified mutations, the most common nonsense mutation (c.595C>T; chr=9) and 4 other missense mutations were associated with the CGTC haplotype; 4 other different missense mutations were associated with the CAAT haplotype; the remaining three different types of mutation (one missense, one promoter, and a small deletion) were associated with the TAAT haplotype (Table 3).

The respective influences of the 3 haplotypes (CGTC, TAAT and CAAT) and the c.4600A>G genotypes on plasma PC activity were then analysed in the controls and patients (Fig. 3A). In the control group, the mean PC activity values in the CG haplotype carrier groups exhibited a downward trend from CGTC/CAAT = 108.7% (95% CI, 93.62–123.8) to CGTC/TAAT = 106% (95%CI, 95.44–116.6) and finally, CGTC/CGTC = 100%

(95% CI, 88.5–111.5). In contrast, the PC activity values varied while exhibiting an upward trend from CAAT/CAAT = 103.2% (95% CI, 74.93–131.15) to TAAT/CAAT = 112.9% (95% CI, 97.74–128) and finally, TAAT/TAAT = 115.6% (95% CI, 104.7–126.5; Fig. 3A). This downward to upward tendency was supported by the significantly different PC activity values between CGTC/CGTC and TAAT/TAAT carriers ($p = 0.0362$). The mean PC activity values observed in controls regarding the *PROCR* c.4600A>G genotypes differed significantly ($p = 0.0087$), with rates of 119.6% (95% CI, 109.7–129.6) with the c.4600AG genotype versus 104.4% (95% CI, 99.08–109.7) with the c.4600AA genotype (Fig. 3B). In the patient group, the 3 *PROC* haplotypes and *PROCR* c.4600A>G genotypes did not induce significant variability, given the already low levels of PC (Fig. 3A, 3B).

Genetic Background of Protein C Deficiency

The influence of polymorphisms on plasma PC levels was analyzed in detail on each of the 26 families, as well as potentiates the phenotype of 12 different mutations. The genotype/phenotype correlation was assessed for each of the mutations, their positions, PC activities and antigen concentrations (Table 4). Only 1 patient exhibited mutational homozygosity (p.Arg220Gln); all others were heterozygous. The four individuals aged ≤ 17 years (one neonate, one aged 7 years, and two prepubertal patients aged 13 and 16 years) who carried different mutations exhibited significantly lower PC activity levels as compared with older relatives carrying the same mutations (Mann–Whitney *U*-test, $p = 0.010$). The individuals aged ≤ 17 years and the homozygous subject were analyzed and reported separately because of these differences in activity levels (Table 4).

Promoter mutation. The -13A>G mutation (La Jolla IV) was shown to abolish 1 of 2 putative HNF-3 binding sites affecting PC promoter transactivation (28). This TAAT haplotype-associated mutation correlated with type I deficiency in 8 individuals from 4 families (Table

4); members of 2 of these families were also heterozygous for *PROCR* c.4600 A>G polymorphism. The highest PC activity (70%–65%) and antigen (69%–61%) levels belonged to 2 relatives who carried the CAAT haplotype in *trans* and heterozygous for the c.4600 A>G polymorphism, 1 of whom was also a F5L carrier. Two probands (PC.1, PC.18) experienced VT episodes associated with acquired risk factors (immobilization due to a fractured meniscus and oral contraceptive use), with ages of onset of 35 and 17 years, respectively. The latter also carried the CGTC haplotype in *trans* and the c.4600 AG genotype.

Nonsense mutation. The c.595C>T (p.Arg199X) mutation in exon 7 is on a hypermutable CpG site and introduces a premature termination codon in the connecting dipeptide (Lys198–Arg199) of the light and heavy chains, thus causing disturbances in thrombin-mediated PC activation (29). This mutation, which is associated with type I PC deficiency, was identified in a heterozygous state in 9 probands and 9 relatives and was thus the most prevalent mutation among the patients (36%). All 18 patients carried the *PROC* promoter CGTC haplotype: 9 homozygotes and 9 heterozygotes. Among these, 11/18 patients (61%) exhibited early VT complications (age range, 19–44 years) and in 1 case, pulmonary embolism was also diagnosed (Table 1). Eight of the 9 individuals homozygous for CGTC had thrombosis. One proband (PC.13) and his 2 relatives also had F5L, with 1 relative additionally carrying the prothrombin G20210A variant although this individual was asymptomatic. Of these 3 individuals, only the proband was homozygous for the CGTC haplotype. It is also noteworthy that the only 2 patients from this group (2 probands and 1 relative) with idiopathic DVT were homozygous for the CGTC haplotype; 1 also carried the c.4600 AG genotype. The fact that the 18 p.Arg199X mutation-carrier chromosomes are in complete linkage disequilibrium with a unique CGTC haplotype suggests that at least within

the Portuguese population, the c.595C>T nucleotide change occurred once on this progenitor haplotype.

Small deletion. The c.633_634 del GC (Leu212Hisfs*2) mutation is located in the activation cleavage site at Arg211→Leu212, where thrombin cleavage releases a dodecapeptide from the heavy chain; this mutation thus interferes with an indispensable role of the PC structure (1). This mutation was identified in 2 families (3 individuals) with type I PC deficiency and existed with the TAAT haplotype (Tables 3, 4). Only 1 individual who was also heterozygous for CGTC haplotype (in *trans*) developed idiopathic DVT; the other 2 were asymptomatic (1 adult and 1 child) and presented TAAT haplotype heterozygosity and homozygosity, respectively, thus suggesting a protective effect of this haplotype. None of these individuals carried the *PROCR* c.4600AG genotype. Besides this study, this deletion has been reported once by another Portuguese group, suggesting that it may have a common ancient origin (30).

Missense mutations. Seven were associated with type I PC deficiency: 1 in the EGF1 domain, p.Cys105Arg, 1 in the activation peptide p.Pro210Leu and 5 in the serine–protease domain, specifically p.Arg220Gln, p.Gly239Arg, p.Val339Met, p.Pro369Leu and p.Trp444Cys. The 2 mutations associated with type II deficiency both occurred in the serine–protease domain: p.Gly215Glu and p.Arg394Trp.

The bioinformatics analyses of these mutations revealed that 8 were classified as ‘Probably damaging’ and 1 as ‘Benign’ by PolyPhen, whereas 7 were classified as ‘Damaging’ and 2 as ‘Tolerated’ according to SIFT; additionally, the Mutation Taster prediction was ‘Disease causing’ for 7 and ‘polymorphism’ for 2 (Table 4). All missense mutations occurred in amino acids that were conserved across species except for p.Arg220Gln and p.Pro369Leu.

The 3 deleterious mutations in the catalytic region, p.Gly239Arg, p.Val339Met and p.Trp444Cys, occurred in 4 families with type I deficiency and were predicted by PolyPhen, SIFT and Mutation Taster as 'damaging', 'very high risk' and 'Disease mutation'. These 3 PC amino acids are highly conserved across orthologs and are thus structurally important. Specifically, in Trp444, the replacement of large tryptophan aromatic ring with a small cysteine hydrophilic side-chain leads to tertiary structure destabilization (31). On the other hand, the p.Pro369Leu substitution affects a non-conserved amino acid.

The new mutation p.Cys105Arg was associated with a type I deficiency in 1 family (3 members). The 3 individuals carried the same promoter *PROC* haplotype combination of CAAT/TAAT and *PROCR* genotype c.4600AA; however, only the proband developed thrombosis associated with an acquired risk factor (dyslipidemia). This variant was absent in all 50 healthy individuals, suggesting that it is not a common polymorphism.

The p.Pro210Leu mutation that includes the *PROC* haplotype combination (CGTC/CAAT) and *PROCR* genotype c.4600AA was associated with type I deficiency in 1 family (4 members). The proband with a previous risk factor (oral contraceptive use) developed a DVT during a long car trip.

Another mutation found in the proximity of the thrombin activation site was p.Gly215Glu, which was detected with a type II variant as mentioned in the database HGMD (Protein C Pec's 3) (29). This mutation was found in a renal transplant patient who had been asymptomatic until age 47 years, the time at which the patient had developed a DVT as a result of an arteriovenous prosthetic device (PC.12). The patient's haplotype combination was CAAT/TAAT and *PROCR* genotype c.4600AA.

The missense mutation identified in family PC.14, p.Arg220Gln, was reported as the most recurrent in the database HGMD (32). This locus exhibits 3 different amino acid substitutions (Gln, Gly and Pro), with the Gln substitution predicted by both PolyPhen and

SIFT as the most 'benign' and 'tolerated', respectively. Four members of family PC.14 were heterozygous and 1 member was homozygous for this substitution, which caused a PC type I deficiency phenotype. Only 2 heterozygous subjects had suffered thrombotic events and the homozygous subject, despite having much lower PC levels, remained asymptomatic at 22 years of age. However, this family member is quite young and the possibility that she developed thrombosis cannot be ruled out upon exposure to risk factors such as trauma, infection and pregnancy. The proband, who was heterozygous for the prothrombin G20210A variant, suffered a DVT after surgery for varicose veins. The promoter *PROC* haplotypes were CAAT/CAAT (n = 3), CAAT/TAAT (n = 1) and CAAT/CGTC (n = 1), and all 5 individuals carried *PROCR* c.4600AA genotype.

The mutation p.Gly239Arg in the serine–protease domain (33) was reported in a 60-year-old man who remained asymptomatic until age 56 years (PC.11), the time at which DVT occurred in the presence of the acquired risks of varicose veins, smoking and dyslipidemia. The promoter *PROC* haplotype was homozygous CAAT/CAAT with the *PROCR* genotype c.4600AA. The effect of this deleterious mutation may have been attenuated by the polymorphism genotypes, and DVT may have been triggered by the acquired risks.

The mutation p.Val339Met was present in 2 families (4 individuals). Among all missense mutations, this was associated with the lowest PC values and the 2 probands had developed DVT at earlier ages (24 and 26 years), thus confirming the deleterious nature of this mutation as described by other authors (34). Besides the contributions of acquired risk factors (oral contraceptive use and delivery) to DVT development, the promoter *PROC* haplotype CGTC was present in homozygosity (n = 1) and in combination with TAAT (n = 3) and the *PROCR* genotype c.4600AA.

The p.Pro369Leu substitution has been associated with a mild phenotype (2). However, this substitution was found in a woman with early-onset VT (20 years of age) associated with

pregnancy and varicose veins; this patient's promoter *PROC* haplotype combination was CGTC/TAAT with the *PROCR* genotype c.4600AG. In this case, CGTC in *cis* and the c.4600AG genotype may have contributed to the cumulative thrombotic effect of this mutation.

The p.Arg394Trp substitution occurs in the immediate vicinity of the active site S402 (2) and has been associated with a type II deficiency state that was detected in proband PC.14. The carrier of this mutation was a young woman (age, 24 years) who had experienced recurrent first-quarter miscarriages; the promoter *PROC* haplotype was homozygous for CGTC and *PROCR* genotype c.4600AA. Co-inheritance of the homozygous CGTC haplotype may have had a cumulative effect and contributed to the increased phenotypic severity.

The proband of the family carrying the p.Trp444Cys mutation (31) was an infant who suffered an episode of severe neonatal thrombosis. An arterial cerebral thrombosis during the infant's first day of life followed by subsequent venipuncture for phenobarbital administration led to ischemia requiring amputation to the proximal interphalangeal joints of the second, third and fourth fingers of the right hand. This family (3 individuals) exhibited F5L co-inheritance; the infant and her mother were found to be homozygous and heterozygous, respectively. The mother developed thrombophlebitis during *puerperium* and the infant's sister remained asymptomatic until 20 years of age. It is known that the inheritance of F5L and *PROC* gene lesions is associated with a cumulative thrombotic risk (10). In this case, the inheritance of a single allele with *PROC* mutation along with the 2 F5L alleles was a potentially critical precipitating event in the early-onset thrombotic episodes in the affected infant. Curiously, the promoter *PROC* haplotype was present in both the protective homozygous TAAT/TAAT (proband and her sister) and heterozygous TAAT/CGTC (mother) forms; all 3 individuals carried the *PROCR* genotype c.4600AA. We

may therefore speculate that in the infant's sister, the effect of the Cys444-mutated allele on the PC levels may have been attenuated by the presence of the TAAT haplotype and were thus higher than expected. Regarding the proband child, this haplotype combination (TAAT/TAAT) may have had a positive modulatory effect on the Cys444 mutation and co-inherited homozygous F5L.

Remaining unidentified molecular lesions. Although the PC activity levels of proband PC.12 were low and phenotypically consistent with type I PC deficiency (activity, 14%; antigen, 17%), no molecular alterations in *PROC* were identified in this family (7 individuals) via promoter, exon and splice-flanking region sequencing. Three of 7 individuals experienced VT episodes. The promoter *PROC* haplotype CGTC was present in both the heterozygous states CGTC/CAAT (n = 3) and CGTC/TAAT (CGTA, n = 3) and the homozygous state (n = 1) along with the *PROCR* genotype c.4600AG. The MLPA analysis screening, which is crucial for identifying regions with large molecular defects, did not reveal any alterations. In this case, mutations may have occurred in intronic regions or locus-controlling regions within *PROC* or in regions outside of the gene that are important for its expression. mRNA studies may help to elucidate the underlying mechanism.

Genotype/Phenotype Analysis

Assuming that the PC activity correlated inversely with the risk of thrombosis and that genetic risk factors have a cumulative effect, the proportion of thrombotic episodes in the patients was analysed according to the 12 different mutations and respective haplotypes. The patients were separated according to the occurrence or not of thrombotic episodes in order to determine the background genetic differences (Fig. 4). It should be noted that patients with PC activity levels ranging from 60%–73% were mostly included in the group without thrombosis. Conversely, in the group of patients with thrombosis, nearly all had PC levels below 60% and the majority was homozygous for the CGTC haplotype. When the

thrombotic events were compared relative to haplotypes (Fig. 5A), patients homozygous for the CGTC haplotype were found to have experienced a significantly higher number of thrombotic events than were other patients ($p = 0.0048$), with a relative risk of 2.3 (95% CI = 1.5–3.6). As shown in Fig. 5B, the greater proportion of patients homozygous for CGTG (82%) carried the common nonsense mutation c.595C>T (p.Arg199X), and this association conferred a relative thrombosis development risk of 2.1 (95% CI = 1.3–3.2).

Discussion

Thrombosis is a complex disease in which each individual's susceptibility to the disease is determined by the effects of both genetic and acquired risk factors (23,26). PC deficiency is an established risk factor for thrombosis; however, diagnosis based on laboratory plasma level cut-offs is often difficult, and therefore establishing patients' disease severity and risk of thrombosis becomes challenging. However, family studies are crucial because as proven by several studies, genetic factors make an approximate 60% contribution to VT (35).

This study comprised a genotype/phenotype correlation in 26 unrelated families with suspected inherited PC deficiency. Twelve different *PROC* gene mutations were identified in 25 of the 26 families (96%). The genotype/phenotype correlation was analysed while considering some known modulating factors such as age, acquired factors and polymorphisms in the *PROC* and *PROCR* genes.

Variability in the plasma PC levels according to age was marked in these families with PC deficiency, which supports the use of age-related normal reference intervals (15). The lowest PC levels observed in individuals <17 years (two children and two prepubertal individuals) could be explained by a genetic mechanism that underlies the age-related regulation of *PROC* gene expression. This mechanism acts as a puberty-onset gene switch with an age-related stability element (ASE) that stabilizes gene expression at the pre-puberty, which is then followed by an age-related increase in gene expression (36)

A correlation between *PROC* promoter haplotypes and the *PROCR* c.4600AG polymorphism with respect to plasma PC levels was noted in the 52 controls. The mean PC activity level in individuals homozygous for the *PROC* CGTC haplotype was approximately 15.6% less than that of individuals homozygous for the TAAT haplotype. Regarding *PROCR* c.4600AG, the mean PC activity level in individuals carrying the c.4600G allele was 15% higher. These results from the control group confirm that the PC levels in healthy individuals are determined by genetic variations. As mentioned in other studies (7,12,16,17,24), the higher CGTC haplotype frequency in the patients in the present study ($p = 0.005$) demonstrated the influence on the thrombotic risk. This association was proven by the higher number of thrombotic episodes experienced by patients homozygous for the CGTC haplotype; conversely, patients homozygous for the TAAT haplotype had the lowest incidence of such events, supporting the notion of a protective effect as suggested previously by other authors (7).

Each of the 12 background genetic mutations justifies the worsening or improvement of a different phenotype (thrombotic risk). These differences are associated with the combinations of mutation types as well as the co-inherited *PROC* promoter haplotype. Therefore, it was possible to observe a positive modulatory effect on the -13A>G and p.Leu212Hisfs*2 mutations conferred by the TAAT haplotype (Fig. 4, Table 4). Similarly, the homozygous patient for Arg220Gln mutation showed relevant PC levels (19–20%), what could be explained by a cumulative effect of the benign characteristic of this missense mutation (PolyPhen-2 prediction score = 0.357) associated with the haplotype CAAT/CAAT. In contrast, individuals with mutations that co-inherited the CGTC haplotype such as p.Arg199X and 4 missense mutations (p.Pro210Leu, p.Val339Met, p.Pro369Leu and p.Arg394Trp) suffered from a negative modulatory effect. The most frequent p.Arg199X mutation was identified in 9 families that contained patients with higher numbers of VT

episodes. Regarding missense mutations, we can speculate that the influence of the CGTC haplotype is reflected in the associated earlier mean age of onset relative to that in patients with other mutations (Table 4).

In this small study, the authors did not set out to make new discoveries related to etiological and PC deficiency mechanisms. Large studies with that intent have already been performed and have provided new insights into thrombosis development (37). However, it is necessary to apply these insights in the context of familial studies to determine their usefulness for individual risk prediction (38). We have integrated some of this recent knowledge about the influence of SNPs on plasma PC levels in families with thrombotic events. We performed a genotype/phenotype correlation and noted the thrombotic event severity (type and number) with respect to the background genetic factors and acquired risk factors present in each family. More specifically, the correlation of the mutation type and location as well as the *PROC* CGTC haplotype and *PROCR* c.4600AG genotypes with the severity of the phenotype could be verified. For each family, it was possible to establish the genetic background and assess the thrombotic risk associated with each element. The common mutation p.Arg199X correlates with the *PROC* CGTC haplotype and an increased risk of VT in the Portuguese population. The SNP c.4600A>G in *PROCR* may explain differences in the PC plasma levels in individuals with the same mutation. These facts support the hypothesis that next generation sequencing (NGS) studies will allow the identification of other risk factors; however, this does not exclude the need for correlation genotype/phenotype analyses as these are required to establish family overviews and to elucidate phenotypic discrepancies and possibly cases of PC deficiency without identified mutations (23,37,39). This study, although geographically limited, illustrates the advantages of identifying the most prevalent mutations in a region.

In conclusion, we demonstrated in this study the importance of establishing the familial thrombotic risk based on the genetic background. Familial mutations, in association with *PROC* haplotypes, correlated strongly with PC levels as well as the thrombotic episode number and age of onset. Therefore, familial genetic studies of 52 individuals with low or borderline PC plasma levels allowed the identification of suspected carriers and individuals with a higher thrombotic risk that will permit evaluations of the need for prophylactic measures, especially in risky situations.

Addendum

T. Fidalgo was responsible for study design, coordination and wrote the manuscript; P. Martinho carried out the molecular analysis with contribution of A. Oliveira that also performed data collection of controls; L. Manco contributed to the analysis of data and revision the manuscript; C. Silva Pinto, D. Marques, E. Gonçalves, gave technical support in the lab; R. Salvado and T. Sevivas provided clinical support. N. Martins (senior responsible of Haemostasis Unit) and ML Ribeiro (Director of Department) were responsible for revisions of the manuscript.

Acknowledgments

We like also thank to all the medical and patients who contributed to this study.

Disclosure of Conflict of Interests

The authors state that they have no conflict of interest.

References

1. Esmon CT. The Protein C Pathway. *Chest*. 2003;124:26–32.
2. Reitsma PH. Protein C deficiency: a database of mutations, 1995 update. *Thromb Haemost*. 1995;73(5):876–89.
3. Cooper PC, Hill M, Maclean RM. The phenotypic and genetic assessment of protein C deficiency. *Int J Lab Hematol*. 2012 Feb 9;1–11.
4. Allaart CF, Poort SR, Rosendaal FR, Reitsma PH, Bertina RM, Briët E. Increased risk of venous thrombosis in carriers of hereditary protein C deficiency defect. *Lancet*. 1993 Jan 16;341(8838):134–8.
5. Kenet G, Lütkehoff LK, Albisetti M, Bernard T, Bonduel M, Brandao L, et al. Impact of thrombophilia on risk of arterial ischemic stroke or cerebral sinovenous thrombosis in neonates and children: a systematic review and meta-analysis of observational studies. *Circulation*. 2010 Apr 27;121(16):1838–47.
6. Mahmoodi BK, Brouwer J-LP, Veeger NJGM, van der Meer J. Hereditary deficiency of protein C or protein S confers increased risk of arterial thromboembolic events at a young age: results from a large family cohort study. *Circulation*. 2008 Oct 14;118(16):1659–67.
7. Aiach M, Nicaud V, Alhenc-Gelas M, Gandrille S, Arnaud E, Amiral J, et al. Complex Association of Protein C Gene Promoter Polymorphism With Circulating Protein C Levels and Thrombotic Risk. *Arterioscler Thromb Vasc Biol*. 1999 Jun 1;19(6):1573–6.
8. Koster T, Rosendaal FR, Briët E, van der Meer FJ, Colly LP, Trienekens PH, et al. Protein C deficiency in a controlled series of unselected outpatients: an infrequent but clear risk factor for venous thrombosis (Leiden Thrombophilia Study). *Blood*. 1995 May 15;85(10):2756–61.
9. Marlar RA, Montgomery RR BA. Diagnosis and treatment of homozygous protein C deficiency. Report of the working party on homozygous protein C deficiency of the subcommittee on protein C and protein S. International Committee of Thrombosis and Haemostasis. *J J Paediatr*. 1989;114:528–34.
10. Millar DS, Johansen B, Berntorp E, Minford A, Bolton-Maggs P, Wensley R, et al. Molecular genetic analysis of severe protein C deficiency. *Hum Genet*. 2000 Jun;106(6):646–53.
11. Tuddenham EG, Takase T, Thomas a E, Awidi a S, Madanat FF, Abu Hajir MM, et al. Homozygous protein C deficiency with delayed onset of symptoms at 7 to 10 months. *Thromb Res*. 1989 Mar 1;53(5):475–84.
12. Spek C a, Koster T, Rosendaal FR, Bertina RM, Reitsma PH. Genotypic variation in the promoter region of the protein C gene is associated with plasma protein C levels and thrombotic risk. *Arterioscler Thromb Vasc Biol*. 1995 Feb;15(2):214–8.
13. Bucciarelli P, Passamonti SM, Biguzzi E, Gianniello F, Franchi F, Mannucci PM, et al. Low borderline plasma levels of antithrombin, protein C and protein S are risk factors for venous thromboembolism. *J Thromb Haemost*. 2012 Sep;10(9):1783–91.

14. Williams MD, Chalmers E a, Gibson BES. The investigation and management of neonatal haemostasis and thrombosis. *Br J Haematol.* 2002 Nov;119(2):295–309.
15. Flanders MM, Crist RA RW and RG. Pediatric reference intervals for uncommon bleeding and thrombotic disorders. *J Pediatr.* 2006;149:275–7.
16. Buil A, Soria JM, Souto JC, Almasy L, Lathrop M, Blangero J, 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 Jul;24(7):1321–5.
17. Thain KR, Nakada T-A, Boyd JH, Russell J a, Walley KR. A common polymorphism in the 5' region of the human protein c gene binds USF1. *Thromb Res.* Elsevier Ltd; 2012 Mar 16;1–7.
18. Tang W, Basu S, Kong X, Pankow JS, Aleksic N, Tan A, et al. Genome-wide association study identifies novel loci for plasma levels of protein C: the ARIC study. *Blood.* 2010 Dec 2;116(23):5032–6.
19. Athanasiadis G, Buil A, Souto JC, Borrell M, López S, Martinez-Perez A, et al. A genome-wide association study of the Protein C anticoagulant pathway. *PLoS One.* 2011 Jan;6(12):e29168.
20. Pintao MC, Roshani S, de Visser MCH, Tieken C, Tanck MWT, Wichers IM, et al. High levels of protein C are determined by PROCR haplotype 3. *J Thromb Haemost.* 2011 May;9(5):969–76.
21. Dennis J, Johnson CY, Adediran AS, de Andrade M, Heit J a, Morange P-E, et al. The endothelial protein C receptor (PROCR) Ser219Gly variant and risk of common thrombotic disorders: a HuGE review and meta-analysis of evidence from observational studies. *Blood.* 2012 Mar 8;119(10):2392–400.
22. Medina P, Navarro S, Estellés A, España F. Polymorphisms in the endothelial protein C receptor gene and thrombophilia. *Thromb Haemost.* 2007;98:564–9.
23. Morange PE, Tregouet D a. Lessons from genome-wide association studies in venous thrombosis. *J Thromb Haemost.* 2011 Jul;9 Suppl 1:258–64.
24. Pomp ER, Doggen CJM, Vos HL, Reitsma PH, Rosendaal FR. Polymorphisms in the protein C gene as risk factor for venous thrombosis. *Thromb Haemost.* 2009 Dec 6;101:62–7.
25. Navarro S, Medina P, Mira Y, Estellés A, Villa P, Ferrando F, et al. Haplotypes of the EPCR gene, prothrombin levels, and the risk of venous thrombosis in carriers of the prothrombin G20210A mutation. *Haematologica.* 2008 Jun;93(6):885–91.
26. Baglin T, Gray E, Greaves M, Hunt BJ, Keeling D, Machin S, et al. Clinical guidelines for testing for heritable thrombophilia. *Br J Haematol.* 2010 Apr;149(2):209–20.
27. Reitsma PH, Poort SR, Allaart CF, Briët E, Bertina RM. The spectrum of genetic defects in a panel of 40 Dutch families with symptomatic protein C deficiency type I: heterogeneity and founder effects. *Blood.* 1991 Aug 15;78(4):890–4.
28. Spek C a, Greengard JS, Griffin JH, Bertina RM, Reitsma PH. Two mutations in the promoter region of the human protein C gene both cause type I protein C deficiency by disruption of two HNF-3 binding sites. *J Biol Chem.* 1995 Oct 13;270(41):24216–21.

29. Dávid M, Losonczy H, Sas G, Nagy A, Kutscher G, Meyer M. Identification of mutations in 15 Hungarian families with hereditary protein C deficiency. *Br J Haematol.* 2000 Oct;111(1):129–35.
30. David D, Ferreira C, Ventura C, Freire I, Moreira I, Gago T. Genetic defects in Portuguese families with inherited protein C deficiency. *Thromb Res.* 2011 Sep;128(3):299–302.
31. Romeo G, Hassan HJ, Staempfli S, Roncuzzi L, Cianetti L, Leonardi A, et al. Hereditary thrombophilia: identification of nonsense and missense mutations in the protein C gene. *Proc Natl Acad Sci U S A.* 1987 May;84(9):2829–32.
32. Soria JM, Morell M, Estivill X, Sala N. Recurrence of the PROC gene mutation R178Q: independent origins in Spanish protein C deficiency patients. *Hum Mutat.* 1996 Jan;8(1):71–3.
33. Ireland, H; Thompson, E; Lane, DA; Chan, LC; Conard, J; De Caterina, M; Rocco, V; De Stefano, V; Leone, G; Finazzi, G; Halil, O; Laffan, M; Machin, S; Woodcock B. Gene mutations in 21 unrelated cases of phenotypic heterozygous Protein C deficiency and thrombosis. *Thromb Haemost.* 1996;76(6):867–73.
34. Gandrille S, Aiach M. Identification of Mutations in 90 of 121 Consecutive Symptomatic French Patients With a Type I Protein C Deficiency. *Blood.* 1995;7(7):2598–605.
35. Germain M, Saut N, Greliche N, Dina C, Lambert J-C, Perret C, et al. Genetics of venous thrombosis: insights from a new genome wide association study. *PLoS One.* 2011 Jan;6(9):e25581.
36. Zhang K, Kurachi S, Kurachi K. Genetic mechanisms of age regulation of protein C and blood coagulation. *J Biol Chem.* 2002 Feb 8;277(6):4532–40.
37. Lowe GD. Epidemiology of venous thromboembolism : the need for large (including prospective) studies and meta-analyses. *J Thromb Haemost.* 2012;10:2186–8.
38. Rosendaal FR. Etiology of venous thrombosis : the need for small original studies. *J Thromb Haemost.* 2012;10:2189–90.
39. Oudot-Mellakh T, Cohen W, Germain M, Saut N, Kallel C, Zelenika D, et al. Genome wide association study for plasma levels of natural anticoagulant inhibitors and protein C anticoagulant pathway: the MARTHA project. *PLoS One.* 2012 Apr;6(9):e255581.

Table 1 Clinical data of 26 probands with Protein C (PC) deficiency

<i>ID</i>	<i>Age (years) / gender</i>	<i>Thrombotic history (Age of onset)</i>	<i>Risk factors</i>	<i>Thrombotic episode (age of relapse)</i>	<i>Number of kindreds with VT</i>
PC.1	45/F	SVT (35)	Meniscus fracture, immobilization	RT (39,44)	0
PC.2	31/F	DVT(19)	No	RT (20,21)	1
PC.3	64/F	SVT (44)	Hypertension and diabetes	PE (64)	0
PC.4	33/F	2 stillbirths 34/30 weeks by placental thrombosis			0
PC.5	33/F	Asymptomatic*	-		1
PC.6	24/F	DVT (24)	Oral contraceptive		0
PC.7	21/M	SVT (21)	No	RT (22,23); DVT (29)	1
PC.8	23/F	DVT; PE (19)	Oral contraceptive, 2 miscarriages (18,23)		0
PC.9	28/F	DVT (26)	Oral contraceptive, long trip		0
PC.10	35/M	DVT (33)	No	CTPV (33)	0
PC.11	56/M	DVT (56)	Varicose veins, smoker and dyslipidemia	SVT (57)	1
PC.12	26/F	DVT (20)	Oral contraceptive, sedentary lifestyle	DVT (27)	2
PC.13	68/F	DVT (34)	Delivery		1
PC.14	48/M	DVT (46)	Surgery varicose veins, obesity, sedentary lifestyle	DVT (47)	0
PC.15	33/M	SVT (33)	Varicose veins		1
PC.16	67/M	Asymptomatic*	-		2
PC.17	13/M	Asymptomatic*	-		1
PC.18	19/F	DVT(17)	Oral contraceptive		0
PC.19	42/F	Stroke (42)	Dyslipidemia		0
PC.20	45/F	DVT (20)	Pregnancy, varicose veins	DVT, T (34)	2
PC.21	16/M	Asymptomatic*	-		1
PC.22	47/M	DVT (47)	Renal transplant patient		0
PC.23	Neonate/F	arterial cerebral thrombosis, ischemia and amputation of the fingers of the right hand			1
PC.24	23/F	recurrent miscarriages			0
PC.25	33/F	SVT (19)	Hypertension, varicose veins, obesity	RT (22,26,33)	1
PC.26	29/F	DVT (26)	Delivery		1

VT, venous thrombosis; DVT, deep vein thrombosis; SVT, superficial vein thrombosis; PE, pulmonary embolism; RT, recurrent superficial vein thrombosis; Cavernous transformation of the portal vein (CTPV); *referred to thrombophilia screening by their family history of VT.

Table 2 Distribution of the promoter haplotypes (*PROC*) and c.4600A>G, p.S219G (*PROCR*) allele frequencies in controls and 26 probands.

PROC	Controls n (%)	Probands n (%)
Haplotype combinations *		
CGTC/ CGTC	9 (17.3)	8 (30.8)
CGTC/ TAAT	15 (28.8)	8 (30.8)
CGTC/ CAAT	7(13.5)	2 (7.7)
TAAT/ TAAT	8 (15.4)	3 (11.5)
TAAT /CAAT	8 (15.4)	3 (11.5)
CAAT/ CAAT	5 (9.6)	2 (7.7)
	52 (100)	26 (100)
Haplotype frequencies		
CGTC	0.385	0.500
TAAT	0.375	0.327
CAAT	0.240	0.173
		**p=0.005
PROCR c.4600A>G (p.Ser219Gly)		
Genotype		
AA	41 (78.9)	21 (80.8)
AG	11 (21.1)	5 (19.2)
	52 (100)	26 (100)
Allele frequencies		
A	0.894	0.904
G	0.106	0.096
		***p=0.713

*SNP sequence as follows: -1657C/T, -1644A/G, -1479A/T, -141T/C

**Chi-square test (O,E) -2-tailed, df=2, for CGTC haplotype between controls and probands

***Chi-square test (O,E) -2-tailed, df=2, for alleles between controls and probands

1

Table 3 Analysis of promoter haplotypes on 12 mutations (*PROC*) - 25 alleles of affected individuals (proband) from unrelated families.

Haplotype	Mutated Allele (chr)												Normal allele
	-13G	c.313C p.Cys105Arg	c.595T p.Arg199X	c.629T p.Pro210Leu	c.633_634delGC p.Leu212Hisfs*2	c.644A p.Gly215Glu	c.659A p.Arg220Gln	c.715A p.Gly239Arg	c.1015A p.Val339Met	c.1106T p.Pro369Leu	c.1180T p.Arg394Trp	c.1332C p.Trp444Cys	
CGTC			9 (36%)	1 (4%)					2 (8%)	1 (4%)	1 (4%)		16 (30.8%)
CAAT		1 (4%)				1 (4%)	1 (4%)	1 (4%)					19 (36.5%)
TAAT	4 (16%)				2 (8%)							1 (4%)	17 (32.7%)

*SNP sequence as follows: -1657C/T, -1644A/G, -1479A/T, -141T/C.

2
3
4

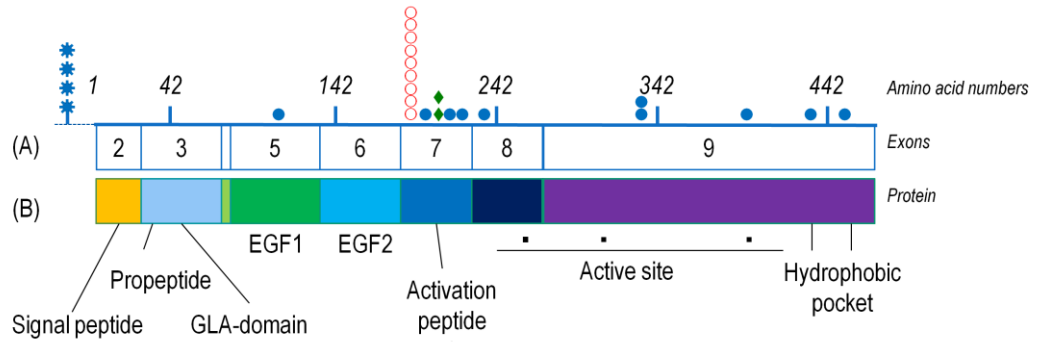
Table 4 Distribution of mutations by type and *PROC/PROCR* polymorphisms in affected individuals and its correlation with PC levels and thrombotic events.

I- Promotor, nonsense and small deletion													
Families ID (n)	Exon	Type of mutation	Nucleotide change	Amino acid change	Domain	Already listed in HGMD	<i>PROC</i> Haplotype	<i>PROCR</i> c.4600 A>G p.S209G	PC Activity (%)	PC Antigen (%)	Mean age of onset	kindreds vs patients with TE	
PC1, PC5, PC16, PC18 (4)	-	Promotor	-13A>G (-1533A>G)	-	HNF- 3 binding	Yes ²⁸	TAAT	AA, AG	52 (38-70)	51 (41-69)	26	8/2	
PC3, PC4, PC7, PC8, PC10, PC13, PC15, PC21, PC25, (9)	7	Nonsense	c.595C>T	p.Arg199X (157)	Linker peptide (dipeptide Lys198/Arg199)	Yes ²⁹	CGTC	AA, AG	51 (43-66); (27; 26)*	51 (44-59); (29; 32)*	29	18/11	
PC2, PC17 (2)	7	Small deletion	c.633_634delGC	p.Leu212Hisfs*2 (170)	Activation peptide	Yes ³⁰	TAAT	AA	43 (36-49); 25*	50 (45-55); 24*	19	3/1	
II- Missense mutations with their pathogenicity prediction													
Families ID (n)	Exon	Nucleotide change	Amino acid change	PolyPhen-2 prediction (score)	SIFT prediction (score)	MutationTaster (prob.)							
PC19 (1)	5	c.313T>C	p.Cys105Arg (63)	Probably damaging (0.999)	Damaging (0.00)	Disease causing (0.9999)	No	CAAT	AA	59 (48-74)	60 (50-76)	42	3/1
PC9 (1)	7	c.629C>T	p.Pro210Leu (168)	Probably damaging (0.966)	Damaging (0.05)	Disease causing (5.6937e-5)	Yes ²	CGTC	AA	63 (57-60)	56 (53-59)	26	4/1
PC22 (1)	7	c.644G>A	p.Gly215Glu (173) ^ξ	Probably damaging (1.000)	Damaging (0.00)	Disease causing (0.9999)	Yes ²⁹	CAAT	AA	56 ^a /45.5 ^c	77	47	1/1
PC14 (1)	7	c.659G>A	p.Arg220Gln (178)	Benign (0.357)	Tolerated (0.35)	Disease causing (7.3306e-6)	Yes ³²	CAAT	AA	60 (58-64)/20**	64 (61-66)/19**	35	5/2
PC11 (1)	8	c.715G>A	p.Gly239Arg (197)	Probably damaging (1.000)	Damaging (0.00)	Disease causing (0.9999)	Yes ³³	CAAT	AG	60 (54-68)	64 (60-70)	56	1/1
PC6, PC26 (2)	9	c.1015G>A	p.Val339Met (297)	Probably damaging (1.000)	Damaging (0.00)	Disease causing (0.9999)	Yes ³⁴	CGTC	AA	37 (37-48)	44 (38-55)	25	4/2
PC20 (1)	9	c.1106C>T	p.Pro369Leu (327)	Probably damaging (0.998)	Tolerated (0.16)	Polymorphism (0.9999)	Yes ²	CGTC	AG	AO	AO	20	1/1
PC24 (1)	9	c.1180C>T	P.Arg394Trp (352) ^ξ	Probably damaging (1.000)	Damaging (0.02)	Polymorphism (0.9999)	Yes ²	CGTC	AA	62 ^a /51.6 ^c	80	22	1/1
PC23 (1)	9	c.1332G>C	p.Trp444Cys (402)	Probably damaging (1.000)	Damaging (0.00)	Disease causing (0.9999)	Yes ³¹	TAAT	AA	51(42-59)/21*	48(41-55)/26*	19	3/2

Mutations are reported in two different nomenclature forms: as suggested by HGVS and as reported by Foster *et al* (1985) in *italics*. The patients under 17 years old, and also the homozygous subject, by their differences in values were excluded from mean ranges and are reported separately. TE – thrombotic events; ^ξ PC type II deficiency^a amidolytic assay; ^c coagulant assay; * individuals < 17 years; ** homozygous patient. SNP sequence on *PROC* haplotype as follows: -1657C/T, -1644A/G, -1479A/T, -1411T/C.

PolyPhen-2 is a tool that predicts the possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. Scores are evaluated as 0.000 (most probably benign) to 1.000 (most probably damaging). *SIFT* is used to predict the effect of sequence changes on the protein's function, based on homology search and the physical properties of amino acids - scores range from 0 to 1, the amino acid substitution is predicted damaging if the score is <= 0.05, and tolerated if the score is > 0.05. *MutationTaster* uses a Bayes classifier to calculate probabilities if the alteration in the sequence is a disease mutation or a harmless polymorphism. A probability close to 1 indicates a high security of prediction.

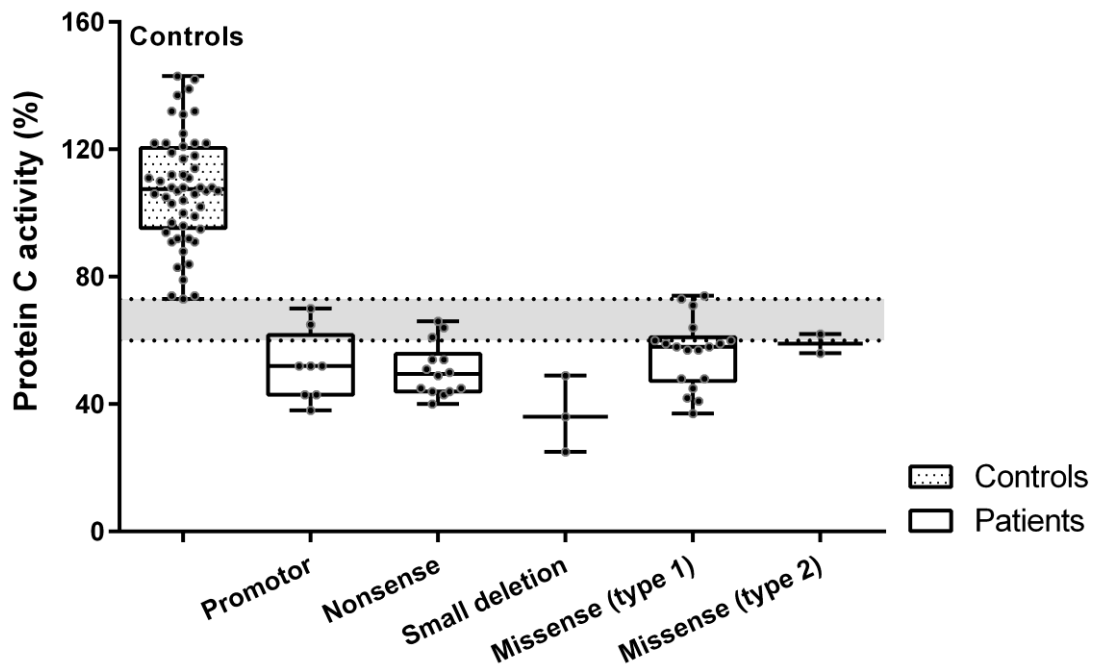
1 Fig. 1.



2

3

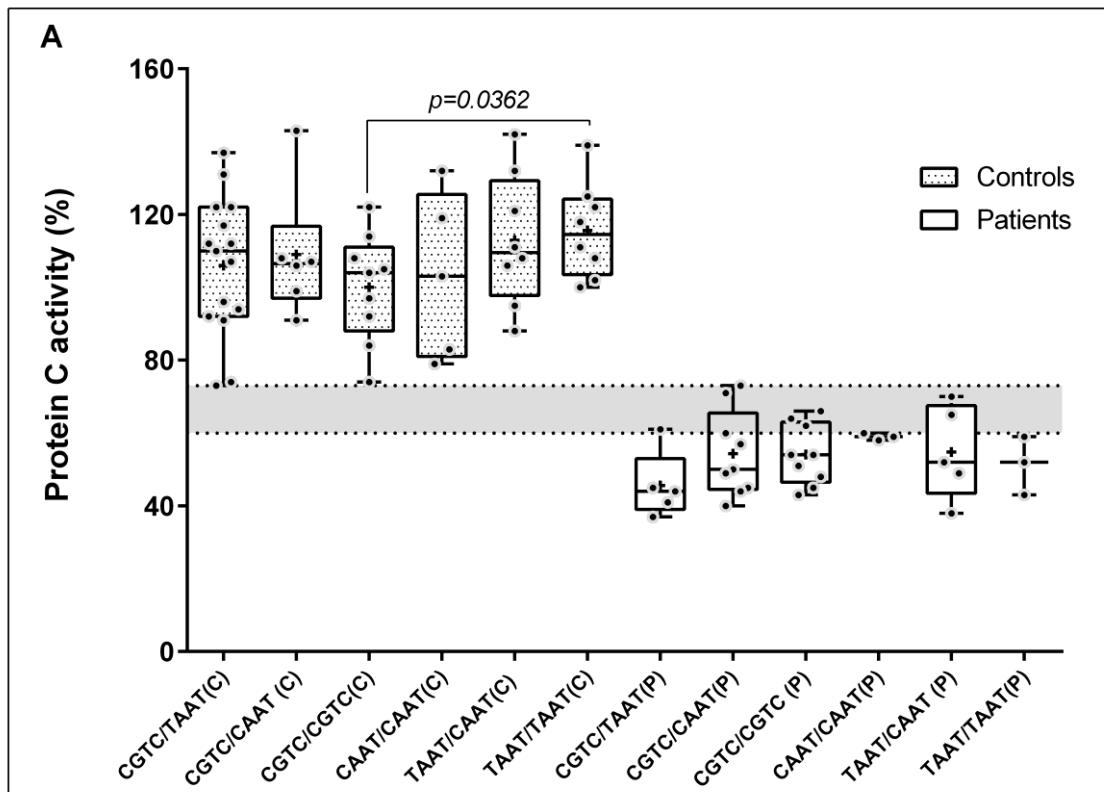
1 Fig. 2



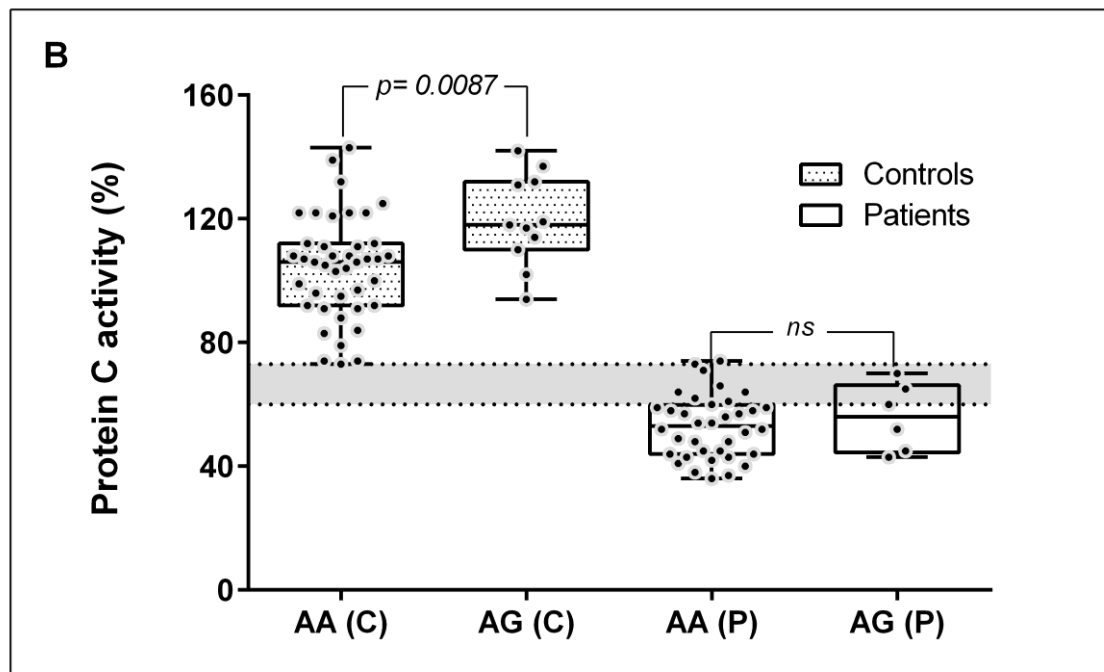
2
3

1 Fig. 3

2



3



4

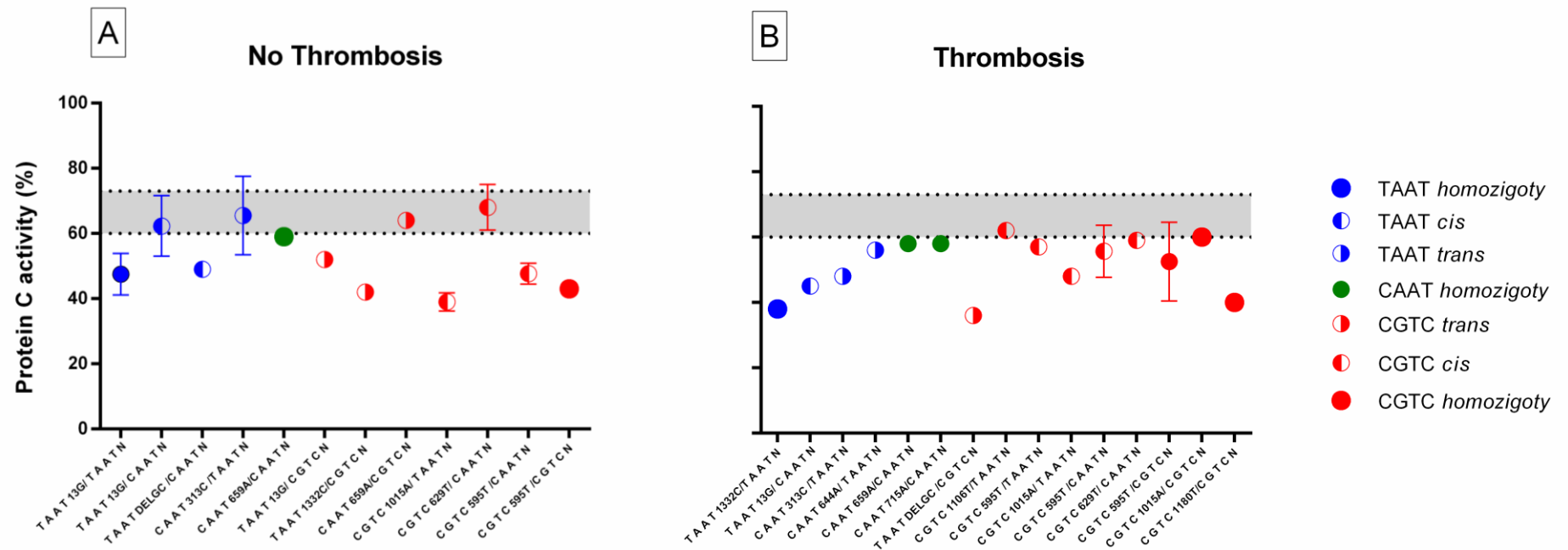
5

6

1 Fig. 4

2

3



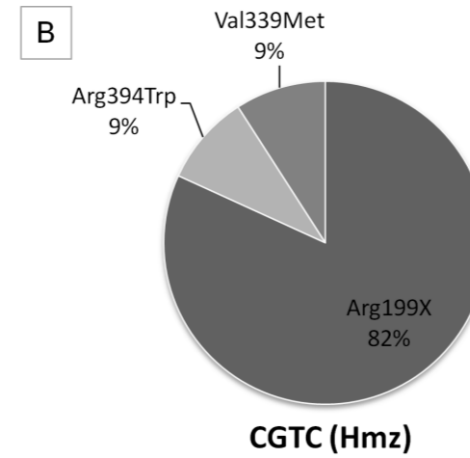
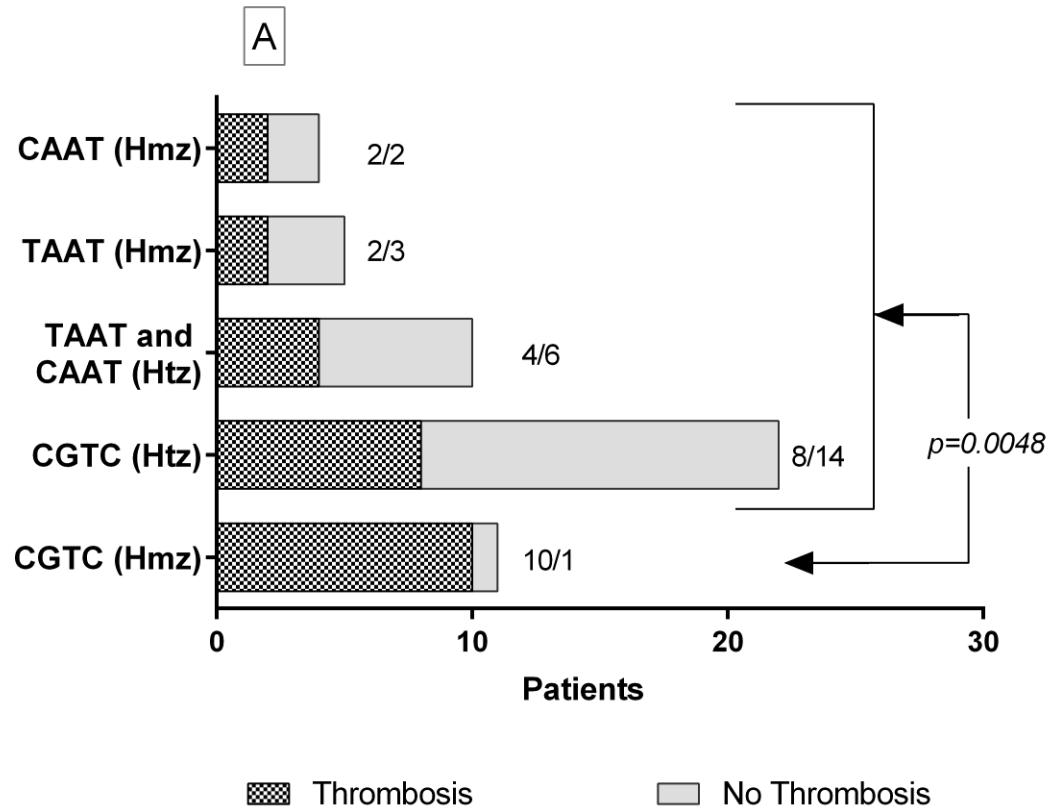
4

5

6

1 Fig. 5

2



3

4

1

2 **Figure legends:**

3

4 **Fig. 1.** – Twelve different mutations were identified in 25 families: promotor * -13A>G; nonsense ○ p.Arg199X;
5 small deletion ◆ p.Leu212Hisfs*2 and missense ● p.Cys105Arg (not previously described); p.Pro210Leu;
6 p.Gly215Glu; p.Arg220Gln; p.Gly239Arg; p.Val339Met; p.Pro369Leu; p.Arg394Trp and p.Trp444Cys. The
7 scheme represents their location through of *PROC* gene (A) and correspondence of functional or structural
8 domains of protein (B).

- 1 **Fig. 2 - Box-Whisker and dot plots of PC activity levels versus type of mutations.** PC activity levels
- 2 according to the different types of mutation, with each dot representing a single case. The grey zone limits the
- 3 levels of PC activity 60%-73%.

1 **Fig. 3 - Box-Whisker and dot plots of PC activity levels within *PROC* promoter haplotype combinations (A)**
2 **and *PROCR* c. 4600 A>G (p.S219G) genotypes (B).** A - PC activity of controls and patients are sorted by 6
3 haplotypes combinations and compared to CGTC/CGTC by t-test. B- PC activity of controls and patients are
4 sorted by c. 4600 AA and AG genotypes and compared by t-test. The grey zone limits the levels of PC activity
5 60%-73%. Significance level was 0.05. If not indicated otherwise comparison was statistically non-significant. (C),
6 controls; (P) patient.

1 **Fig. 4 - PC activity levels in different mutations and promoter *PROC* haplotype combinations– distribution**
2 **by group of patients with thrombosis versus group of patients without thrombosis.** Patients with PC
3 activity levels between 60%-73% (grey zone) were mostly included in the group without thrombosis; conversely in
4 the group of patients with thrombosis nearly all had PC levels below 60% and the majority was homozygotes for
5 CGTC haplotype.

1 **Fig. 5 – Comparison of thrombotic events in patients relative to promoter PROC haplotypes.**

- 2 (A) The number of thrombotic events of patients homozygous for haplotype CGTC had a significantly difference
3 with the other haplotypes when compared by Fisher's exact test; (B) The majority of patients homozygous for
4 CGTG (82%) had the nonsense mutation c.595C>T (p.Arg199X).
5 Hmz- homozygous; Htz- heterozygous