



HAL
open science

A novel rare c.-39C>T mutation in the PROS1 5'UTR causing PS deficiency by creating a new upstream translation initiation codon

Sylvie Labrousche-Colomer, Omar Soukariéh, Carole Proust, Christine Mouton, Yoann Huguenin, Maguelonne Roux, Céline Besse, Anne Boland, Robert Olaso, Joël Constans, et al.

► To cite this version:

Sylvie Labrousche-Colomer, Omar Soukariéh, Carole Proust, Christine Mouton, Yoann Huguenin, et al.. A novel rare c.-39C>T mutation in the PROS1 5'UTR causing PS deficiency by creating a new upstream translation initiation codon. *Clinical Science*, 2020, 134 (10), pp.1181-1190. 10.1042/CS20200403 . inserm-03014711

HAL Id: inserm-03014711

<https://inserm.hal.science/inserm-03014711v1>

Submitted on 19 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A novel rare c.-39C>T mutation in the PROS1 5'UTR causing PS deficiency by creating a new upstream translation initiation codon.

Sylvie Labrouche-Colomer^{1,2*}, MD, PhD, Omar Soukariéh^{3*}, PhD, Carole Proust³, Christine Mouton¹, MD, Yoann Huguenin⁴, MD, Maguelonne Roux⁵, Céline Besse⁶, Anne Boland⁶, PhD, Robert Olasso⁶, PhD, Joël Constans¹, MD, PhD, Jean-François Deleuze^{6,7}, Pierre-Emmanuel Morange⁸, MD, PhD, Béatrice Jaspard-Vinassa² PhD, David-Alexandre Trégouët³, PhD on behalf of the GenMed consortium

¹ CHU de Bordeaux, Laboratoire d'Hématologie, Pessac, France.

² INSERM UMR 1034, Biology of Cardiovascular Disease, University of Bordeaux, Pessac, France.

³ INSERM UMR 1219, Bordeaux Population Health Research Center, University of Bordeaux, Bordeaux, France

⁴ CHU de Bordeaux, Service de pédiatrie médicale, Bordeaux, France.

⁵ Human Evolutionary Genetics Unit, Institut Pasteur, UMR2000, CNRS, Paris 75015, France

⁶ Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry, France.

⁷ Centre d'Etude du Polymorphisme Humain, Fondation Jean Dausset, Paris, France

⁸ C2VN INSERM UMR 1263, INRA, Aix-Marseille University, Marseille, France

*These two authors equally contributed to the work

Correspondence to : Dr David-Alexandre Trégouët, INSERM U1219, Bordeaux Population Health Research Center, 33076 Bordeaux, France ; david-alexandre.tregouet@inserm.fr; david-alexandre.tregouet@u-bordeaux.fr:

Abstract

Autosomal dominant inherited Protein S deficiency (PSD) (MIM 612336) is a rare disorder caused by rare mutations, mainly located in the coding sequence of the structural *PROS1* gene, and associated with an increased risk of venous thromboembolism. To identify the molecular defect underlying PSD observed in an extended French pedigree with 7 PSD affected members in who no candidate deleterious *PROS1* mutation was detected by Sanger sequencing of *PROS1* exons and their flanking intronic regions or via a MLPA approach, a whole genome sequencing strategy was adopted. This led to the identification of a never reported C to T substitution at c.-39 from the natural ATG codon of the *PROS1* gene that completely segregates with PSD in the whole family. This substitution ACG→ATG creates a new start codon upstream of the main ATG. We experimentally demonstrated in HeLa cells that the variant generates a novel overlapping upstream open reading frame (uORF) and inhibits the translation of the wild type Protein S. This work describes the first example of 5'UTR *PROS1* mutation causing PSD through the creation of an uORF, a mutation that is not predicted to be deleterious by standard annotation softwares, and emphasizes the need for better exploring such type of non coding variations in clinical genomics

Keywords

Protein S deficiency, venous thrombosis, whole genome sequencing, Open Reading Frame, mutation.

Clinical Perspectives

1: Inherited Protein S deficiency (PSD) is a severe disorder associated with high risk of venous thrombosis. This study aims at characterizing the molecular defect underlying unexplained PSD observed in 7 affected relatives in who clinical sequencing diagnosis was unfruitful.

2: We identify the first non coding mutation causing PSD via the creation of a new upstream translation initiation codon that generates an upstream open reading frame (uORF)

3: This work illustrates the importance of better exploring uORF-creating variants in the context of clinical sequencing

Introduction

Protein S (PS) is, with Protein C (PC) and Antithrombin (AT), one the three main natural inhibitors of the coagulation cascade and plays thus a key role in the control of blood clot formation. PS is a vitamin-K dependent glycoprotein that acts as a cofactor for PC to inactivate factors Va (FVa) and VIIIa (FVIIIa) and to limit thrombin generation via direct interactions with factor Xa (FXa) and FVa (1). In human plasma, PS circulates both under a free and active form (~40%) and an inactive form (~60%) when complexed with C4b-binding protein. Generally, PS plasma concentration can be characterized by antigen measurements of the free and total PS levels or by PS activity. This led to the definition of three clinical subtypes of PS deficiencies: i) Type I refers to deficiency of both free and total PS as well as decreased PS activity, ii) Type II is defined by normal plasma levels but decreased PS activity while iii) Type III shows decreased free PS plasma levels and decreased PS activity but normal total PS plasma levels. Type I and Type III PS deficiencies account for ~95% of PS deficiencies

and are considered to be the heterogeneous clinical expression of the same molecular defect (2).

Autosomal dominant inherited PS deficiency (PSD) (MIM 612336) is a rare disorder caused by private or rare mutations in the structural *PROS1* gene. Previous works have shown that complete or partial PSD was associated with increased risk of venous thrombosis (VT) (3–7). However, a substantial proportion of PSD cases remain without any definitive molecular diagnosis (8–11), up to 40% according to a recent estimate (8). The majority of identified PSD causing *PROS1* mutations are located in the coding regions of the gene or in their flanking sequences and are mainly missense or nonsense single nucleotide variations (12). Some intronic splice variants (13–15) and structural variants (16–18) have also been reported. Note that, compared to exonic variations, very few variations have been described in the promoter region or the 5'UTR of the *PROS1* gene (19–23) and even fewer have been functionally characterized. To our knowledge, the c.-168C>T is the sole *PROS1* promoter variation that has been experimentally demonstrated to cause inherited PSD by affecting the core binding site of Sp1 transcription factor (22) and no PSD causing mutation has yet been described in the 5'UTR region of *PROS1*.

We here describe a novel mutation in the 5'UTR region of the *PROS1* gene that causes PSD in a family with multiple relatives affected with VT. The mutation is a C to T substitution at c.-39 from the natural ATG codon and creates a novel ATG sequence upstream of the main open reading frame that alters the translation machinery of the protein.

Materials and Methods

Recruitment of the family

The studied family was ascertained through a proband addressed to the specialized clinical hematology laboratory of the Pellegrin Hospital (Bordeaux) for a thrombophilia screening in the context of an objectively documented first VT episode early in life (20 years old).

Standard thrombophilia screening revealed a Type I PSD (Fig 1; Supplementary Table 1).

However, no candidate deleterious *PROS1* mutation nor large genomic rearrangement were detected by Sanger sequencing of the 15 *PROS1* exons and of their flanking intronic regions (24) or by multiplex ligation dependent probe amplification (MLPA) using Salsa MLPA kit (P112 ; MRC-Holland, Amsterdam, the Netherlands), respectively. As the proband reported at least one first degree relative with VT, family members were invited to participate to a genetic exploration upon informed consent agreement according to the Helsinki declaration. Inherited Type I or Type III PSD was diagnosed in 6 additional relatives (Fig 1), a seventh relative (individual IV-4) had very low PS activity but, without information on free and total PS levels, Type I, Type II or Type III PSD could not be objectively distinguished. All PSD patients with DNA available were negative for the presence of *PROS1* deleterious coding mutation and for the F5 Leiden (FV Arg506Gln, rs6025), the F2 G20210 (rs1799963) mutations and any other major biological risk factors for venous thrombosis including antithrombin-, PC-deficiency or anti-phospholipid antibodies.

Phenotypic protein S deficiency was diagnosed by measuring APC-cofactor activity, free (FPS) and/or total (TPS) proteins S levels. The cofactor activity of protein S was measured by an activated partial thromboplastin time-based clotting assay using the Staclot protein S kit (Diagnostic Stago, Asnieres, France) or by the IL test protein S kit (Instrument Laboratory Company, Milano, Italy). The concentration of total and free PS antigen were determined by an enzyme-linked immunosorbent assay (Asserachrom total or free protein S; Diagnostica Stago). In normal conditions, TPS levels range from 70% to 150% while normal FPS levels are

generally greater than 60% in men, and 50% in pre-menopause women (>55% in menopause women). In the related PSD patients studied here, FPS levels ranged from 18% to 38% and TPS from 51% to 70% (Fig 1). APC-cofactor activity varied between 27% and 32%. Detailed clinical and biological information on family members are given in Supplementary Table 1.

All family members participating to this study signed a written informed consent for genetic investigations, according to the Helsinki declaration.

Whole genome sequencing

With the aim of identifying the hypothesized culprit mutation causing the observed familial PSD, we sequenced the whole genome of 5 PSD patients (individuals III-2, III-8, IV-2, IV-4, IV-5) and 2 unaffected relatives (individuals IV-1 and IV-6) (Fig 1).

Whole genome sequencing was performed at the Centre National de Recherche en Génomique Humaine (CNRGH, Institut de Biologie François Jacob, Evry, FRANCE). After a complete quality control, 1µg of genomic DNA was used for each sample to prepare a library for whole genome sequencing, using the Illumina TruSeq DNA PCR-Free Library Preparation Kit, according to the manufacturer's instructions. After normalisation and quality control, qualified libraries were sequenced on a HiSeqX5 instrument from Illumina (Illumina Inc., CA, USA) using a paired-end 150 bp reads strategy. One lane of HiSeqX5 flow cell was used per sample specific library in order to reach an average sequencing depth of 30x for each sequenced individual. Sequence quality parameters have been assessed throughout the sequencing run and standard bioinformatics analysis of sequencing data was based on the Illumina pipeline to generate FASTQ file for each sample. FastQ sequences were aligned on

human genome hg37 using the BWA-mem program (25). Variant calling was performed using the GATK HaplotypeCaller (GenomeAnalysisTK-v3.3-0, <https://software.broadinstitute.org/gatk/documentation/article.php?id=4148>) tool followed by recalibration. Single nucleotide variants (SNVs) that succeeded the "PASS" filter were then annotated using Annovar (26).

To comply with the autosomal dominant mode of inheritance of the observed PSD, we selected as candidate culprit PSD causing variant any rare variant that has either never been reported in public database or at a very low allele frequency (<1%) and that were present at the heterozygous states in the 5 whole genome sequenced patients but not present in 2 healthy relatives. Interrogated public databases were dbSNP, GnomAD (<https://gnomad.broadinstitute.org/>), Ensembl (<https://www.ensembl.org/index.html>), 1000 genomes (<https://www.internationalgenome.org/>), ExAC (<http://exac.broadinstitute.org/>) and FrEx (<http://lysine.univ-brest.fr/FrExAC/>).

Sanger sequencing validation

Sanger sequencing validation of the identified candidate variant was carried out using primers designed to span all the putative transcription factor binding sites described by de Wolf (27) (Supplementary Table 2). PCR amplification of genomic DNA was performed using Phusion™ Green Hot Start II High Fidelity DNA Polymerase with GC Buffer in presence of 10% of DMSO (ThermoFisher). Big dye sequencing chemistry was used to sequence the PCR products in both directions using the ABI 3500xL Genetic analyser (Applied Biosystems, Foster City, CA) and the sequence were analyzed by the SeqScape Software (Applied Biosystems, Foster City, CA).

Plasmid constructs

To assess the impact of the *PROS1* c.-39C>T variation, several expression vectors were investigated (Fig 2B). The human *PROS1* cDNA (GenBank NM_000313.4) was PCR-amplified from HepG2 cDNA and cloned into pcDNA3.1/myc-His(-) plasmid (Invitrogen). The PCR amplification of *PROS1* from c.-44 to c.2028 was performed using Phusion High Fidelity (HF) DNA Polymerase (ThermoFisher) and primer pairs designed to be either complementary to the wild-type sequence or carrying the c.-39C>T variant on the forward primer only (Supplementary Table 3). After restriction enzyme digestion, the wild-type and mutant PCR products were cloned into pcDNA3.1 to fuse the full-length *PROS1* coding sequence in-frame with a myc-his tag, generating the wild-type (WT *PROS1*) and the mutant (Natural mutant *PROS1*) *PROS1* vectors, respectively (Fig 2B). To analyze the translation potential of the novel upstream start codon and overlapping uORF generated by the c.-39 C>T variant, two additional expression vectors were constructed: (i) a truncated mutant *PROS1* vector was constructed by PCR-amplification of *PROS1* cDNA from c.-44 to c.110 (cloning of the overlapping uORF sequence) using mutagenic primers (Supplementary Table 3) fused in frame with the myc-His tag into pcDNA3.1; (ii) an elongated mutant *PROS1* vector (Elongated mutant *PROS1*) was obtained from the Natural mutant *PROS1* in which the new stop codon induced by the c.-39T allelic variant was suppressed by creating a 1-bp deletion (c.111delT) through rapid-site-directed-mutagenesis using phosphorylated non-overlapping primers (Supplementary Table 3) and Phusion HF DNA Polymerase (ThermoFisher), according to the manufacturer's recommendations. The latter construct was used here as an experimental control to avoid any ambiguity about the detection of the small-size expected fusion protein

from the new uORF by western blot. All the recombinant plasmids were verified by Sanger sequencing (Genewiz).

Cell culture and transfection

HeLa cells were cultured in RPMI medium (Gibco-Invitrogen) supplemented with 10% fetal calf serum (Gibco-Invitrogen), 1% penicillin/streptomycin and 1% HEPES. Cells were maintained in T75 flasks in a humidified atmosphere of 95% and 5% CO₂ at 37°C. Twenty four hours before the transfection, 6-well plates of HeLa cells were prepared with 4x10⁵ cells/well. Transfections were performed with jetPRIME[®] reagent (Polyplus Transfection) according to the manufacturer's recommendations, with 0.5-1.5 µg/mL of each expression plasmid or empty pcDNA[™]3.1/myc-His(-) vector at a cell confluence between 60 and 80%. In some experiments, 0,5 µg of pcDNA3.1/myc-His/lacZ was included in each transfection to control the transfection efficiency. Cells were harvested and lysed 48 hours after transfection to extract total RNA and protein.

RNA isolation and RT-qPCR analysis

Total RNA was isolated from the transfected HeLa cells using the RNeasy mini kit (Qiagen) following the manufacturer's instructions. Reverse transcription of 2 µg of total RNA to single- stranded cDNA was performed using High Capacity cDNA kit (Applied Biosystems). To quantify PROS1 transcripts, qPCR were carried out on the generated cDNA (diluted at 1/5) using PowerUp SYBR master mix (ThermoFisher Scientific) in a final volume of 10 µl and 40 cycles of amplification on QuantStudio3 Real-Time PCR System (ThermoFisher). Results were analyzed with the QuantStudio design and analysis software. Transcript levels were normalized to the reference RPL32 gene. Relative quantification of *PROS1* in different

samples was conducted according to threshold cycle (Ct) value based on the $\Delta\Delta\text{CT}$ method. All the primer pairs used in this study (Supplementary Table 4) have a reaction efficiency between 90% and 110% and melting curves were analyzed to check the specificity of each RT-qPCR reaction.

Protein preparation and western blot analysis

Protein extracts from the transfected HeLa cells were prepared by using 140 μL per well of RIPA buffer containing protease inhibitors. After clarification of the extracts by centrifugation, protein concentration was determined by the BCA method (Pierce™ BCA Protein Assay Kit). Equal amounts (50 μg) of total protein suspended in loading buffer were separated by SDS-PAGE using 8% or 4-20% gradient gels and transferred onto PVDF membranes (Immobilon-P or Immobilon PSQ, Merck Millipore) for immunostaining. Then, membrane was divided into two halves and probed with either monoclonal anti Myc Tag antibody (Merck Millipore) or anti- α -tubulin antibody (Sigma-Aldrich). After incubation with goat anti-mouse IgG Alexa Fluor 700 (ThermoFisher), blots were simultaneously scanned on Odyssey Infrared Imaging System (Li-Cor Biosciences) in the 700 channel. α -Tubulin was used as control.

Statistical data analysis

Differential mRNA levels analysis according to experimental conditions were performed with ANOVA followed by Tukey's HSD post-hoc tests for multiple comparisons. A statistical threshold of $p < 0.05$ was used to declare statistical significance.

Results

With the aim of completely ruling out the possibility of PSD causing mutations in the structural gene, we first focused on genetic variations mapping the *PROS1* locus, downstream and upstream of the coding sequence. We thus identified a never reported c.-39C>T (variant in the *PROS1* 5'UTR (Human Genome Variation Society nomenclature: 5'UTR cDNA : NM_000313.3(*PROS1*):c.-39C>T ; gDNA: Chr3(GRCh37):g.93692632G>A) present at the heterozygous state in all 5 PSD sequenced patients and absent in the two sequenced healthy relatives. The rare T allele results into a novel upstream translation start codon (uATG) generating a novel overlapping upstream open reading frame (uORF) of 51 codons out-of-frame with the normal ORF (Fig 2A).

Because of the broad role of uORFs on downstream translation (28,29), we hypothesized that the detected variant could be at the origin of the observed PSD. To validate this hypothesis, we first confirmed the presence of the variant in the five sequenced patients by Sanger sequencing (Supplementary Figure 1) and genotyped it in one additional affected related (individual IV-7) for whom DNA was available. The absence of the the variant in the two healthy relatives with available DNA was confirmed by Sanger sequencing as well.

Then, to experimentally assess the impact of the variant on *PROS1* translation, we first generated 2 myc-His tagged molecular constructs (Fig 2B), the full length *PROS1* constructs containing either the wild-type ("WT") or the mutated -39T allele ("Natural mutant"). We observed that transfecting HeLa cells with the wild type or the natural mutant was homogeneously and significantly associated ($p < 10^{-3}$) with increased *PROS1* mRNA levels as compared to controls (Fig 3A and Supplementary Figure 2) without altering the expression of endogenous *PROS1* (Fig 3B). Interestingly, the overexpression was not significantly different

($p = 0.38$) between WT and Natural mutant. These first results suggest that the c.-39C>T variant does not impact the transcription machinery of the *PROS1* transgene.

We further investigated the synthesis of the fusion-myc-His construct proteins and confirmed that PROS1 protein was produced in a dose-dependent manner by the wild type plasmid while no PROS1 protein was detected with the natural mutant plasmid (Fig 4A). To better understand the underlying mechanisms, we generated the "truncated mutant" construct, containing the overlapping uORF in frame with the tag (Fig 2B). Our aim was to evaluate the ability of the novel upstream start codon to initiate the translation of an expected 8.3kDa fusion protein (<https://www.snapgene.com/snapgene-viewer/>). We failed to evidence any fusion protein synthesis in this molecular weight range (Fig 4B) while the expression of PROS1 transgene was confirmed by qPCR (Supplementary Table 5). To go further, we generated an elongated mutant plasmid (see Materials and Methods) allowing the translation of a longer tagged protein from the identified upstream start codon. Western blot analyses showed the presence of a protein at the expected 81.9 kDa weight (Figure 4C), suggesting that a translation from the uAUG and driven by the uORF was initiated. However, the level of expression of the latter tended to be lower than that of the wild type protein, despite similar mRNA levels (Supplementary Table 5). This second set of experiments suggest that the c.-39C>T variant creates a new functional uAUG codon that is able to initiate translation process while inhibiting the translation of the normal form of the Protein S from the main AUG.

Discussion

We here report the first case of inherited PSD due to a never reported mutation in the *PROS1* promoter region that creates a premature ATG codon out-of-frame relative to the natural ATG codon, thus generating a premature stop codon.

Following a negative target gene sequencing of the *PROS1* exonic sequences and large genomic rearrangement, a whole genome sequencing strategy was adopted to identify the molecular defect responsible for inherited PSD in a French pedigree with 8 PSD affected individuals and 2 unaffected relatives. This led to the identification of the novel c.-39C>T variant, with the rare T allele generating a novel uAUG start codon and a novel overlapping uORF of 51 amino acids. Even if the flanking sequence of the new ATG does not completely match to the consensus predicted Kozak sequence ((G/A)NNAUGG), it seems similar to the one flanking the natural *PROS1* ATG, suggesting that the new ATG could be identified by the translational machinery and thus used in the cells (30,31). Several works have previously demonstrated that uAUG and uORF are *cis*-regulatory elements in 5'UTR that are able to control protein expression by altering the translation efficiency and could subsequently be associated with risk of diseases (32–35). Many previously ignored uORFs are now known to act as major post-transcriptional regulatory elements or to be translated to produce bioactive peptides or proteins (36,37).

Interestingly, the relative repression of uORF-containing mRNA was shown to be more pronounced when the uORF overlaps the translation starting site of the CDS (28). Our experimental data confirm that the c.-39T allele alters the translation of the main ORF leading to low Protein S production consistent with the observed PSD in carriers of the mutation. The c.-39T allele is expected to produce a shorter protein of 51 amino acids with no homology to wild-type Protein S. However, we were not able to detect it through our WB

experiments. This prevented us from assessing whether the resulting protein is not stable, or not produced because of a possible degradation of the new mRNA by the nonsense mediated decay (NMD). Unfortunately, we were not able to collect mRNA materials from patients, the main limitation of this study, which prevented us from investigating the aforementioned NMD hypothesis. Would mRNA have been available, we could have validated the presence of the expected uORF in carriers of the identified mutation.

Importantly, none of the popular bioinformatics tools used to annotate variants such as PolyPhen, Sift, CADD (38), TraP (39) or RegulomeDB (40) were able to predict any deleteriousness for this variant. The PreTIS software (41) dedicated to 5'UTR variants predicted a "moderate" impact of the variant while the DeepSea suite's predictions (42,43) were neither supportive for a strong functional impact of the variant.

The c.-39C>T variant was further screened in a sample of 150 PS-deficient patients with thromboembolic disease available at the specialized clinical hematology laboratory of the Pellegrin Hospital (Bordeaux) and in sample of 200 VT patients with normal PS plasma levels and sequenced for their whole genome at the Laboratory of Haematology of the La Timone Hospital (Marseille) (bioRxiv 2020.03.29.012054; doi:

<https://doi.org/10.1101/2020.03.29.012054>). No carrier was identified. Nevertheless, in the light of the present finding, it would be interesting to systematically investigate whether such type of non coding mutations that create overlapping uORF in *PROS1* could explain PSD cases that have been so far molecularly unsolved.

While the identified c.-39C>T mutation segregates with PSD, it is also important to emphasize that all but one family members with PSD and complete clinical information experienced one or more VT events, with age of onset of first VT varying between 16 and 38

years old. A part from the PSD patient with missing clinical information (individual IV-4), the sole affected patient carrying the disease mutation that has not (yet) developed a thrombotic event is the youngest one. She is currently 16 yrs old and her free PS levels were 18%. Free PS levels in other relatives with combined PSD & VT ranged from 22% to 38% while such values ranged from 88% to 125% in healthy family members. As no thrombotic event was observed in non PSD relatives, these observations are in agreement with previous works suggesting genetically induced low free PS levels as a strong risk factor for VT. Indeed, Pintao et al observed that FPS levels lower than 33% was associated with an Odds Ratio (OR) for VT of ~5 (7). In a retrospective family, Lijfering et al. observed an increased OR for VT of ~6 in relatives with FPS levels lower than 41% compared to relatives in the upper quartile of FPS distribution. Similarly, Alhenc-Gelas et al (3) proposed to define a ~30%-40% cut-off value for FPS levels as a way to identify individuals at high VT risk. All these proposed thresholds were consistent with the FPS levels observed in our studied family. As suggested by Alhenc-Gelas et al, the associated VT risk and the threshold value for FPS levels may be variable and may depend on the characteristics of the underlying causative variant. As an example, Suchon et al. demonstrated that the *PROS1* Heerlen mutation was associated with an OR for VT of ~6 but a moderate decrease in free PS levels (~ 70%)(44).

Of note, this family project was set up in the context of the GenMed LABORatory of Excellence (<http://www.genmed.fr/index.php/fr/>), one of the objectives of which is to propose WGS in Mendelian disorders for which whole exome sequencing and target gene sequencing has been inconclusive. Even though WGS is now more and more affordable, it must be stressed that a less expensive deep sequencing of the *PROS1* locus including

regulatory regions would have likely permit the identification of the culprit mutation. This illustrates the importance of exploring regulatory regions of known disease gene when searching for molecular diagnostics, a result that goes far beyond PSD but shall be generalized to other inherited (hematological) disorders.

Authors ' contributions

Clinical investigations of patients and of their relatives were conducted by SLC, CM, YH and JC. Whole sequencing was performed by CB, AB and RO. MR was in charge of the bioinformatics analysis of the whole genome sequenced data. Experimental investigations were conducted by CP, OS and BJV. The research study was designed by JFD, PEM, BJV and DAT. The manuscript was drafted by SLC, BJV and DAT and further reviewed by all co-authors.

Acknowledgments

O.S and M.R were financially supported by the GENMED Laboratory of Excellence on Medical Genomics [ANR-10-LABX-0013]. DA T was supported by the «EPIDEMIOLOGIE-VTE» Senior Chair from the Initiative of Excellence of the University of Bordeaux. This work was partially supported by the GENMED Laboratory of Excellence on Medical Genomics (ANR-10-LABX-0013) and the French Clinical Research Infrastructure Network on Venous Thrombo-Embolism (F-CRIN INNOVTE), two research programs managed by the National Research Agency (ANR) as part of the French Investment for the Future.

Figure Legends

Figure 1: Pedigree of the studied family affected with Protein S deficiency

Figure 2: Schematic representation of the identified *PROS1* c.-39C>T variant

(A) *PROS1* wild-type and c.-39C>T genomic sequences. Positions of the identified variant (c.-39), as well as the first nucleotide on the main ORF (+1) and of the generated stop codon (+111) are indicated above the WT sequence. Brackets indicate the main ORF and new uORF codons with the respective amino acid sequences. (B) The different molecular constructs used to investigate the variant's effect *in vitro*, as described under materials and methods.

Figure 3: Impact of the *PROS1* c.-39C>T variant on *PROS1* RNA expression

mRNA expression levels were evaluated after transfection of HeLa cell lines with 1 µg of vectors encoding full-length WT *PROS1* or Natural mutant *PROS1*. As control, cells were transfected with the empty pcDNA3.1/*myc*-His plasmid alone. After 48 hours of culture, mRNA expression was evaluated by Quantitative Reverse Transcription PCR (RT-qPCR) using primer pairs targeting either (A) total *PROS1* expression (ex1fw,ex2rv primers) or (B) only endogenous *PROS1* (3UTRfw,3UTRrv primers) detected, independently from transfected *PROS1* transcripts, thanks to the use of primers directed against 3UTR sequences absent from the “transfected constructs” (see Supplementary Table 4).. Results were normalized to RPL32 mRNA and analyzed by using the comparative Ct Method ($\Delta\Delta C_t$ Method). $-\Delta\Delta C_t$ Values are shown on y-axis.

Figure 4: Impact of the *PROS1* c.-39C>T variant on the Protein S expression

(A) Protein S expression was evaluated in HeLa cell lines transfected with 1, 1.5 and 2 µg of either WT, Natural mutant *PROS1* constructs or empty pcDNA3.1/*myc*-His, as control (see Materials & Methods). Note that the wild-type construct directs expression of full-length Protein S in a dose-

dependent manner, whereas this expression is blocked in the Natural mutant-PROS1 encoding the novel uAUG.(B) Protein expression was evaluated in HeLa cell lines transfected with 1, 1.5 and 2µg of either WT, natural mutant PROS1, truncated mutant PROS1 or empty pcDNA3.1/myc-His, used as control. The wild-type construct directs expression of the full-length Protein S. However, we have not detected any protein associated with the mutation (natural and truncated mutant PROS1 encoding the novel uAUG).

(C) Protein S expression was evaluated in HeLa cell lines transfected with 1µg of vectors encoding WT PROS1, Natural mutant PROS1, Elongated mutant PROS1 or empty vector control. To control the transfection efficiency, 0,5 µg of pcDNA3.1/myc-His/lacZ was included in each transfection. Note that the wild-type construct directs expression of full-length protein S. Interestingly, we highlighted the presence of a larger mutant protein with the elongated mutant PROS1 construct, probably resulting from the use of the new AUG.

REFERENCES

1. Castoldi E, Hackeng TM. Regulation of coagulation by protein S. *Curr Opin Hematol.* sept 2008;15(5):529- 36.
2. Zöller B, García de Frutos P, Dahlbäck B. Evaluation of the relationship between protein S and C4b-binding protein isoforms in hereditary protein S deficiency demonstrating type I and type III deficiencies to be phenotypic variants of the same genetic disease. *Blood.* 15 juin 1995;85(12):3524- 31.
3. Alhenc-Gelas M, Plu-Bureau G, Horellou MH, Rauch A, Suchon P, GEHT genetic thrombophilia group. PROS1 genotype phenotype relationships in a large cohort of adults with suspicion of inherited quantitative protein S deficiency. *Thromb Haemost. mars 2016;115(3):570- 9.*
4. Alhenc-Gelas M, Canonico M, Morange PE, Emmerich J, Geht Genetic Thrombophilia Group. Protein S inherited qualitative deficiency: novel mutations and phenotypic influence. *J Thromb Haemost.* déc 2010;8(12):2718- 26.

5. Lijfering WM, Mulder R, ten Kate MK, Veeger NJGM, Mulder AB, van der Meer J. Clinical relevance of decreased free protein S levels: results from a retrospective family cohort study involving 1143 relatives. *Blood*. 5 févr 2009;113(6):1225- 30.
6. Pegelow CH, Ledford M, Young JN, Zilleruelo G. Severe protein S deficiency in a newborn. *Pediatrics*. avr 1992;89(4 Pt 1):674- 6.
7. Pintao MC, Ribeiro DD, Bezemer ID, Garcia AA, de Visser MCH, Doggen CJM, et al. Protein S levels and the risk of venous thrombosis: results from the MEGA case-control study. *Blood*. 31 oct 2013;122(18):3210- 9.
8. Alhenc-Gelas M, Plu-Bureau G, Horellou MH, Rauch A, Suchon P, GEHT genetic thrombophilia group. PROS1 genotype phenotype relationships in a large cohort of adults with suspicion of inherited quantitative protein S deficiency. *Thromb Haemost*. mars 2016;115(3):570- 9.
9. Hermida J, Faioni EM, Mannucci PM. Poor relationship between phenotypes of protein S deficiency and mutations in the protein S alpha gene. *Thromb Haemost*. déc 1999;82(6):1634- 8.
10. Iwaki T, Mastushita T, Kobayashi T, Yamamoto Y, Nomura Y, Kagami K, et al. DNA sequence analysis of protein S deficiency--identification of four point mutations in twelve Japanese subjects. *Semin Thromb Hemost*. 2001;27(2):155- 60.
11. Biguzzi E, Razzari C, Lane DA, Castaman G, Cappellari A, Bucciarelli P, et al. Molecular diversity and thrombotic risk in protein S deficiency: the PROSIT study. *Hum Mutat*. mars 2005;25(3):259- 69.
12. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*. 2017;136(6):665- 77.
13. Choi J, Kim H-J, Chang MH, Choi J-R, Yoo J-H. A rare splicing mutation in the PROS1 gene of a Korean patient with type I hereditary protein S deficiency. *Ann Clin Lab Sci*. 2011;41(4):397- 400.
14. Menezes J, Ventura C, Costa JM, Parreira E, Romão L, Gonçalves J. PROS1 novel splice-site variant decreases protein S expression in patients from two families with thrombotic disease. *Clin Case Rep*. 2017;5(12):2062- 5.
15. Wang X, Tang N, Wang X, Lu Y, Yang J. PROS1 IVS10+5G>A mutation causes hereditary protein S deficiency in a Chinese patient with pulmonary embolism and venous thromboembolism. *Thromb Res*. 2019;174:1- 4.
16. Hurtado B, Nadal M, Margarit E, Sánchez A, Abasolo N, García N, et al. First case of protein S deficiency due to a translocation t(3;21)(q11.2;q22). *Thromb Haemost*. mai 2009;101(5):977- 9.

17. Lind-Halldén C, Dahlen A, Hillarp A, Zöller B, Dahlbäck B, Halldén C. Small and large PROS1 deletions but no other types of rearrangements detected in patients with protein S deficiency. *Thromb Haemost.* juill 2012;108(1):94- 100.
18. Seo JY, Lee K-O, Kim S-H, Oh D, Kim D-K, Kim H-J. The genomic architecture of the PROS1 gene underlying large tandem duplication mutation that causes thrombophilia from hereditary protein S deficiency. *Gene.* 1 sept 2014;547(2):295- 9.
19. Espinosa-Parrilla Y, Morell M, Borrell M, Souto JC, Fontcuberta J, Estivill X, et al. Optimization of a simple and rapid single-strand conformation analysis for detection of mutations in the PROS1 gene: identification of seven novel mutations and three novel, apparently neutral, variants. *Hum Mutat.* 2000;15(5):463- 73.
20. Hurtado B, Muñoz X, Mulero MC, Navarro G, Domènech P, García de Frutos P, et al. Functional characterization of twelve natural PROS1 mutations associated with anticoagulant protein S deficiency. *Haematologica.* avr 2008;93(4):574- 80.
21. Li L, Wu X, Wu W, Ding Q, Cai X, Wang X. Clinical Manifestation and Mutation Spectrum of 53 Unrelated Pedigrees with Protein S Deficiency in China. *Thromb Haemost.* mars 2019;119(3):449- 60.
22. Sanda N, Fujimori Y, Kashiwagi T, Takagi A, Murate T, Mizutani E, et al. An Sp1 binding site mutation of the PROS1 promoter in a patient with protein S deficiency. *Br J Haematol.* sept 2007;138(5):663- 5.
23. Tang L, Jian X-R, Hamasaki N, Guo T, Wang H-F, Lu X, et al. Molecular basis of protein S deficiency in China. *Am J Hematol.* oct 2013;88(10):899- 905.
24. Labrousche S, Reboul M-P, Guérin V, Vergnes C, Freyburger G. Protein C and protein S assessment in hospital laboratories: which strategy and what role for DNA sequencing? *Blood Coagul Fibrinolysis.* sept 2003;14(6):531- 8.
25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 15 juill 2009;25(14):1754- 60.
26. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* sept 2010;38(16):e164.
27. de Wolf CJF, Cupers RMJ, Bertina RM, Vos HL. The constitutive expression of anticoagulant protein S is regulated through multiple binding sites for Sp1 and Sp3 transcription factors in the protein S gene promoter. *J Biol Chem.* 30 juin 2006;281(26):17635- 43.
28. Johnstone TG, Bazzini AA, Giraldez AJ. Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J.* 1 avr 2016;35(7):706- 23.
29. Lin Y, May GE, Kready H, Nazzaro L, Mao M, Spealman P, et al. Impacts of uORF codon identity and position on translation regulation. *Nucleic Acids Res.* 26 2019;47(17):9358- 67.

30. Kozak M. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J.* 1 mai 1997;16(9):2482- 92.
31. Kozak M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci USA.* nov 1990;87(21):8301- 5.
32. Dvir S, Velten L, Sharon E, Zeevi D, Carey LB, Weinberger A, et al. Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proc Natl Acad Sci USA.* 23 juill 2013;110(30):E2792-2801.
33. Orr MW, Mao Y, Storz G, Qian S-B. Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res.* 20 2020;48(3):1029- 42.
34. Jones L, Goode L, Davila E, Brown A, McCarthy DM, Sharma N, et al. Translational effects and coding potential of an upstream open reading frame associated with DOPA Responsive Dystonia. *Biochim Biophys Acta Mol Basis Dis.* 2017;1863(6):1171- 82.
35. von Bohlen AE, Böhm J, Pop R, Johnson DS, Tolmie J, Stücker R, et al. A mutation creating an upstream initiation codon in the SOX9 5' UTR causes acampomelic campomelic dysplasia. *Mol Genet Genomic Med.* mai 2017;5(3):261- 8.
36. Jones L, Goode L, Davila E, Brown A, McCarthy DM, Sharma N, et al. Translational effects and coding potential of an upstream open reading frame associated with DOPA Responsive Dystonia. *Biochim Biophys Acta Mol Basis Dis.* 2017;1863(6):1171- 82.
37. Orr MW, Mao Y, Storz G, Qian S-B. Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res.* 29 août 2019;
38. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 8 janv 2019;47(D1):D886- 94.
39. Gelfman S, Wang Q, McSweeney KM, Ren Z, La Carpia F, Halvorsen M, et al. Annotating pathogenic non-coding variants in genic regions. *Nat Commun.* 09 2017;8(1):236.
40. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* sept 2012;22(9):1790- 7.
41. Reuter K, Biehl A, Koch L, Helms V. PreTIS: A Tool to Predict Non-canonical 5' UTR Translational Initiation Sites in Human and Mouse. *PLoS Comput Biol.* oct 2016;12(10):e1005170.
42. Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat Genet.* août 2018;50(8):1171- 9.
43. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods.* oct 2015;12(10):931- 4.

44. Suchon P, Germain M, Delluc A, Smadja D, Jouven X, Gyorgy B, et al. Protein S Heerlen mutation heterozygosity is associated with venous thrombosis risk. *Sci Rep.* 04 2017;7:45507.