EUROPEAN HEMATOLOGY ASSOCIATION
haematologica
Journal of the European Hematology Association

Unraveling the effect of silent, intronic and missense mutations on VWF splicing: contribution of next generation sequencing in the study of mRNA

by Nina Borràs, Gerard Orriols, Javier Batlle, Almudena Pérez-Rodríguez, Teresa Fidalgo, Patricia Martinho, María Fernanda López-Fernández, Ángela Rodríguez-Trillo, Esther Lourés, Rafael Parra, Carme Altisent, Ana Rosa Cid, Santiago Bonanad, Noelia Cabrera, Andrés Moret, María Eva Mingot-Castellano, Nira Navarro, Rocío Pérez-Montes, Sally Marcellini, Ana Moreto, Sonia Herrero, Inmaculada Soto, Núria Fernández-Mosteirín, Víctor Jiménez-Yuste, Nieves Alonso, Aurora de Andrés-Jacob, Emilia Fontanes, Rosa Campos, María José Paloma, Nuria Bermejo, Ruben Berrueco, José Mateo, Karmele Arribalzaga, Pascual Marco, Ángeles Palomo, Nerea Castro Quismondo, Belén Iñigo, María del Mar Nieto, Rosa Vidal, María Paz Martínez, Reyes Aguinaco, Jesús María Tenorio, María Ferreiro, Javier García-Frade, Ana María Rodríguez-Huerta, Jorge Cuesta, Ramón Rodríguez-González, Faustino García-Candel, Manuela Dobón, Carlos Aguilar, Francisco Vidal, and Irene Corrales

Haematologica 2018 [Epub ahead of print]

Citation: Nina Borràs, Gerard Orriols, Javier Batlle, Almudena Pérez-Rodríguez, Teresa Fidalgo, Patricia Martinho, María Fernanda López-Fernández, Ángela Rodríguez-Trillo, Esther Lourés, Rafael Parra, Carme Altisent, Ana Rosa Cid, Santiago Bonanad, Noelia Cabrera, Andrés Moret, María Eva Mingot-Castellano, Nira Navarro, Rocío Pérez-Montes, Sally Marcellini, Ana Moreto, Sonia Herrero, Inmaculada Soto, Núria Fernández-Mosteirín, Víctor Jiménez-Yuste, Nieves Alonso, Aurora de Andrés-Jacob, Emilia Fontanes, Rosa Campos, María José Paloma, Nuria Bermejo, Ruben Berrueco, José Mateo, Karmele Arribalzaga, Pascual Marco, Ángeles Palomo, Nerea Castro Quismondo, Belén Iñigo, María del Mar Nieto, Rosa Vidal, María Paz Martínez, Reyes Aguinaco, Jesús María Tenorio, María Ferreiro, Javier García-Frade, Ana María Rodríguez-Huerta, Jorge Cuesta, Ramón Rodríguez-González, Faustino García-Candel, Manuela Dobón, Carlos Aguilar, Francisco Vidal, and Irene Corrales. Unraveling the effect of silent, intronic and missense mutations on VWF splicing: contribution of next generation sequencing in the study of mRNA. Haematologica. 2018; 103:xxx doi:10.3324/haematol.2018.203166

Unraveling the effect of silent, intronic and missense mutations on *VWF* splicing: contribution of next generation sequencing in the study of mRNA

Nina Borràs,[1,2] Gerard Orriols,[1] Javier Batlle,[3] Almudena Pérez-Rodríguez,[3] Teresa Fidalgo,[4] Patricia Martinho,[4] María Fernanda López-Fernández,[3] Ángela Rodríguez-Trillo,[3] Esther Lourés,[3] Rafael Parra,[1,2] Carme Altisent,[2] Ana Rosa Cid,[5] Santiago Bonanad,[5] Noelia Cabrera,[5] Andrés Moret,[5] María Eva Mingot-Castellano,[6] Nira Navarro,[7] Rocío Pérez-Montes,[8] Sally Marcellini,[9] Ana Moreto,[10] Sonia Herrero,[11] Inmaculada Soto,[12] Núria Fernández-Mosteirín,[13] Víctor Jiménez-Yuste,[14] Nieves Alonso,[15] Aurora de Andrés-Jacob,[16] Emilia Fontanes,[17] Rosa Campos,[18] María José Paloma,[19] Nuria Bermejo,[20] Ruben Berrueco,[21] José Mateo,[22] Karmele Arribalzaga,[23] Pascual Marco,[24] Ángeles Palomo,[25] Nerea Castro Quismondo,[26] Belén Iñigo,[27] María del Mar Nieto,[28] Rosa Vidal,[29] María Paz Martínez,[30] Reyes Aguinaco,[31] Jesús María Tenorio,[32] María Ferreiro,[33] Javier García-Frade,[34] Ana María Rodríguez-Huerta,[35] Jorge Cuesta,[36] Ramón Rodríguez-González,[37] Faustino García-Candel,[38] Manuela Dobón,[39] Carlos Aguilar,[40] Francisco Vidal,[1,2,41] and Irene Corrales[1,2.]

[1]Banc de Sang i Teixits, Barcelona, Spain; [2]Institut de Recerca Vall d'Hebron - Universitat Autònoma de Barcelona (VHIR-UAB), Barcelona, Spain; [3]Complexo Hospitalario Universitario A Coruña, INIBIC, A Coruña, Spain; [4]Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal; [5]Hospital Universitario y Politécnico La Fe, Valencia, Spain; [6]Hospital Regional Universitario de Málaga, Málaga, Spain; [7]Hospital Universitario Dr. Negrín, Las Palmas de Gran Canaria, Spain; [8]Hospital Universitario Marqués de Valdecilla, Santander, Spain; [9]Salud Castilla y León, Segovia, Spain; [10]Hospital Universitario Cruces, Barakaldo, Spain; [11]Hospital Universitario de Guadalajara, Guadalajara, Spain; [12]Hospital Universitario Central de Asturias, Oviedo, Spain; [13]Hospital Universitario Miguel Servet, Zaragoza, Spain; [14]Hospital Universitario La Paz, Madrid, Spain; [15]Hospital Infanta Cristina, Badajoz, Spain; [16]Complexo Hospitalario Universitario Santiago de Compostela, Spain; [17]Hospital Universitario Lucus Augusti, Lugo, Spain; [18]Hospital Jerez de la Frontera, Cádiz, Spain; [19]Hospital Virgen del Camino, Pamplona, Spain; [20]Hospital San Pedro de Alcántara, Cáceres, Spain; [21]Hospital Sant Joan de Deu, Barcelona, Spain; [22]Hospital Sta Creu i St Pau, Barcelona, Spain; [23]Hospital Universitario Fundación de Alcorcón, Madrid, Spain; [24]Hospital General de Alicante, Alicante, Spain; [25]Hospital Regional Universitario Carlos Haya, Málaga, Spain; [26]Hospital Universitario 12 de Octubre, Madrid, Spain; [27]Hospital Clínico San Carlos, Madrid, Spain; [28]Complejo Hospitalario de Jaén, Jaén, Spain; [29]Fundación Jiménez Díaz, Madrid, Spain; [30]Hospital Nuestra Sra. de Sonsoles de Ávila, Ávila, Spain; [31]Hospital Joan XXIII, Tarragona, Spain; [32]Hospital Ramón y Cajal, Madrid, Spain; [33]Hospital Montecelo, Pontevedra, Spain; [34]Hospital Río Hortega,

Valladolid, Spain; [35]Hospital Gregorio Marañón, Madrid, Spain; [36]Hospital Virgen de la Salud, Toledo, Spain; [37]Hospital Severo Ochoa, Madrid, Spain; [38]Hospital Universitario Virgen Arrixaca, Murcia, Spain; [39]Hospital Lozano Blesa, Zaragoza, Spain; [40]Hospital Santa Bárbara, Soria, Spain; [41]CIBER de Enfermedades Cardiovasculares.

Statement of equal authors' contribution: IC and FV contributed equally to this study.

Running head: Effect of *VWF* mutations on mRNA splicing

Contact information for correspondence:
Irene Corrales Insa, PhD. Banc de Sang i Teixits, Barcelona, Spain. E-mail: icorrales@bst.cat
Francisco Vidal Pérez, PhD. Banc de Sang i Teixits, Barcelona, Spain. E-mail: fvidal@bst.cat

Word Count:
Abstract word count: 248

Text word count: 4049

Number of tables: 2

Number of figures: 5

Supplemental files: 1

Acknowledgments

We are indebted to Baxalta US Inc., now a part of Shire, for support of the PCM-EVW-ES (Grant H13-000845). This study was also supported by the Spanish Ministry of the Economy and Competitiveness (MINECO, *Ministerio de Economía y Competitividad)*, Instituto de Salud Carlos III (ISCIII) (PI12/01494, PI15/01643 and RD12/0042/0053). We are very grateful for the kind collaboration of the participating patients and their families. CIBERCV is an initiative of ISCIII, co-financed by the European Regional Development Fund (ERDF), "A way to build Europe".

ABSTRACT

Large studies in von Willebrand disease patients, including Spanish and Portuguese registries, led to identification of >250 different mutations. It is a challenge to determine the pathogenic effect of potential splice site mutations on *VWF* mRNA. This study aimed to elucidate the true effects of 18 mutations on *VWF* mRNA processing, investigate the contribution of next-generation sequencing to *in vivo* mRNA study in von Willebrand disease, and compare the findings with *in silico* prediction. RNA extracted from patient platelets and leukocytes was amplified by RT-PCR and sequenced using Sanger and next generation sequencing techniques. Eight mutations affected *VWF* splicing: c.1533+1G>A, c.5664+2T>C and c.546G>A (p.=) prompted exon skipping; c.3223-7_3236dup and c.7082-2A>G resulted in activation of cryptic sites; c.3379+1G>A and c.7473G>A (p.=) demonstrated both molecular pathogenic mechanisms simultaneously; and the p.Cys370Tyr missense mutation generated two aberrant transcripts. Of note, the complete effect of 3 mutations was provided by next generation sequencing alone because of low expression of the aberrant transcripts. In the remaining 10 mutations, no effect was elucidated in the experiments. However, the differential findings obtained in platelets and leukocytes provided substantial evidence that 4 of these would have an effect on VWF levels. In this first report using next generation sequencing technology to unravel the effects of *VWF* mutations on splicing, the technique yielded valuable information. Our data bring to light the importance of studying the effect of synonymous and missense mutations on *VWF* splicing to improve the current knowledge of the molecular mechanisms behind von Willebrand disease.

ClinicalTrials.gov identifier: NCT02869074

INTRODUCTION

Von Willebrand disease (VWD), the most common congenital bleeding disorder, is caused by a genetic defect in the von Willebrand factor gene (*VWF*).[1] *VWF* mutational analysis can be valuable for diagnosing and investigating the molecular etiology of VWD, as was seen in the Spanish (PCM-EVW-ES project) and Portuguese cohort of VWD patients.[2-4] One interesting challenge in this condition is to elucidate the pathogenic mechanism of *VWF* mutations. *In silico* analysis is considered a suitable supporting tool to predict the pathogenicity of the variants identified.[5, 6] However, functional studies remain essential to unequivocally determine their deleterious effect.[7]

Functional studies can be performed by analyzing the potential effect of splice site mutations (PSSM) in RNA. In addition to splice site consensus sequence mutations, deep intronic, missense, and synonymous mutations can also disturb splicing. Along this line, 25% of synonymous mutations positioned at exon-intron boundaries result in altered splicing, which, in itself, can cause disease, modify the severity of the disease phenotype, or be linked with disease susceptibility.[8] In *VWF*, heterozygotes for PSSM may be associated with mild forms of VWD type 1 or be phenotypically silent, but when 2 such mutations are found in different alleles, the phenotype is associated with VWD type 3.[9]

Almost all related studies of splicing effects have examined peripheral blood platelets, as *VWF* is exclusively expressed in these cells and endothelial cells.[10, 11] Platelets are anucleated, but they contain small amounts of translationally active megakaryocytic mRNA.[12,13] In contrast, the amount of mRNA obtained from leukocytes is higher and contains mRNA transcripts for genes that are not normally expressed in these cells, known as ''ectopic transcripts''. Their analysis has been used to investigate mutations in several inherited disorders,[14,15] as they facilitate the study of mRNA of those genes expressed in hard to reach tissues.[16] PSSM can affect mRNA reorganization and introduce premature termination codons (PTCs) into open reading frames, a common cause of genetic disorders. Most nonsense transcripts are recognized and degraded by nonsense-mediated mRNA decay (NMD),[17] a degradation pathway to control synthesis of truncated proteins.[18] The efficiency of NMD varies between cell types; hence, the use of RNA from platelets and leukocytes for *in vivo* study of *VWF* PSSM offers complementary results, particularly when NMD occurs in the allele carrying the mutation in platelets, as we reported.[7]

Since its development, next-generation sequencing (NGS) has been increasingly used in molecular genetics to identify mutations causing disease. However, few groups have explored its potential for analyzing splicing variants following RT-PCR.[19, 20] In this new scenario, the

procedure we previously described to analyze the effects of PSSM in *VWF*[7] has been optimized and adapted to an NGS-based technique to investigate its value in this field. Our main objective was to elucidate the true effects of 18 selected mutations (intronic, synonymous, delins, and missense) on mRNA processing and their genotype/phenotype correspondence by analysis of leukocytes and platelets from clinical samples. Finally, the *in vivo* effects of the mutations were compared with the *in silico* predictions.

METHODS

Patients

We studied 15 patients diagnosed with different types of VWD, 5 from Complexo Hospitalario Universitario A Coruña, 8 from Hospital Universitari Vall d'Hebron (HUVH), and 2 from Centro Hospitalar e Universitário de Coimbra. Samples from 4 healthy individuals were used as controls. The study was performed according to the guidelines of the Declaration of Helsinki and was approved by the local Research Ethics Committee. All participants provided written informed consent.

Splice site prediction software

The predicted impact of potential splice site mutations was analyzed with NetGene2[21] and the splicing prediction module of Alamut Visual v.2.6.1 software (Interactive Biosoftware, Rouen, France), which integrates data from 3 methods: Splice Site Prediction by Neural Network (NNSplice), MaxEntScan, and Human Splicing Finder (HSF).

Platelet and leukocyte separation and RNA isolation

Leukocyte and platelet RNA from patients and controls was isolated from 10 mL of peripheral blood collected in EDTA tubes, as previously described.[7]

VWF mRNA amplification

After RNA isolation, cDNA was synthesized using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's recommendations. The region including the mutation was amplified by Platinum Taq DNA Polymerase (Thermo Fisher Scientific) in leukocyte and platelet cDNA (Online Supplementary Methods and Table S1). PCR products were separated on 1% agarose gel and visualized by SYBR Safe DNA Gel Stain (Thermo Fisher Scientific).

Sanger sequencing and analysis

PCR products were sequenced as previously described.[22] However, multiple-band PCR products were previously agarose-purified using the MiniElute Gel Extraction kit (Qiagen). The sequences obtained were assembled and aligned against the consensus wild-type (WT) *VWF* mRNA sequence (GenBank NM_000552) using SeqScape v2.7 software (Thermo Fisher Scientific).

Next-generation sequencing and analysis

PCR amplicons obtained per patient were equimolarly mixed in a single tube in a total amount of 250 ng. Subsequently, the libraries were fragmented and the adapter and barcodes were ligated using the NGSgo protocol (GenDX, Utrecht, Netherlands) following the manufacturer's recommendations. Resulting libraries were combined and sequenced on a MiSeq platform (Illumina, San Diego, CA).

After sequencing, barcoded sequences were demultiplexed and analyzed individually. The paired sequence files (fastq format) were used as input for analysis with the CLC Genomic Workbench v.11 software (Qiagen, Aarhus, Denmark) (Online Supplementary Methods and Figure S1-S2).

RESULTS

An in-depth study was performed in PCM-EVW-ES,[3] the Portuguese cohort,[4] and HUVH patients to select previously undescribed *VWF* mutations and mutations with an unknown or controversial pathogenic mechanism. Eighteen mutations (15 patients) were selected: 8 intronic (4 in canonical, GT and AG, splice site sequences), 5 synonymous, 2 missense, and 3 delins. The patients' phenotypic and molecular data are summarized in Table 1 and Online Supplementary Table S2. Total mRNA was obtained from platelets and leukocytes of all patients, with the exception of patients UMP08 and UMP14, in whom platelet RNA isolation failed due to blood lysis. All mutations were analyzed by both Sanger sequencing and NGS, and results were compared to the predictions generated by *in silico* analysis (Online Supplementary Table S3).

Mutations in canonical splice site sequences

The **c.1533+1G>A** mutation (intron 13) was identified in a type 3 VWD carrier (UMP01). The exon 11-15 region was analyzed with a specific primer pair to avoid amplification of a prevalent alternative-splicing product (skipping of exons 14 and 15) in *VWF* from leukocytes.[7] Whereas only the expected PCR product was observed in platelets, leukocyte amplification resulted in 4 PCR bands. Sanger analysis of leukocytes showed 2 mRNA aberrant transcripts: 1) lacking exon 13; and 2) lacking exon 13 and 14 (Online Supplementary Figure S3). By means of NGS it was determined that 19% were transcripts without exon 13, 45% transcripts without exons 13 and 14, and an additional aberrant transcript without exon 14 was detected in 8%

transcripts (Online Supplementary Table S4). By both techniques, no effect was visible in platelets due to NMD.

The **c.3379+1G>A** mutation (intron 25) was identified in a type 1H VWD patient (UMP02). We designed 2 new primers to analyze the exon 22-26 region. Amplification of leukocyte cDNA yielded 2 bands, one with the expected size and a smaller band, whereas in platelets only the expected PCR product was observed (Online Supplementary Figure S4). Sanger analysis of leukocytes showed that c.3379+1G>A caused exon 25 skipping, leading to a frameshift at position 1075 and adding 88 aberrant amino acids before a PTC was encountered (p.Pro1075ValfsTer88). On NGS, however, 2 aberrant transcripts were detected: the major one, which lacked exon 25, was detected in 43% of all reads, and the minor one, which resulted from activation of a cryptic donor splice site (DSS) 31 nucleotides upstream from the WT-DSS, was found in 1.4% of reads (p.Pro1117ValfsTer88) (Table 2 and Online Supplementary Table S4). In platelets, however, the mutated allele had undergone NMD, which precluded observation of aberrant transcripts.

The **c.5664+2T>C** mutation was identified in a type 1 VWD patient (UMP03), *in trans* with the p.Leu2407Pro mutation. To analyze the PSSM in intron 33, the exon 30-35 region was amplified. The PCR product had the expected size in platelets, whereas an additional smaller one was observed in leukocytes (Figure 1). Sanger sequencing of the leukocyte PCR product confirmed exon 33 skipping (p.Gly1874AlafsTer32) (Table 2). NGS detected 2 aberrant transcripts: the major one (37% of reads) showed exon 33 skipping, and the minor (2% of reads) showed exons 33 and 34 skipping. Moreover, analysis of the p.Leu2407Pro mutation confirmed that NMD of the allele carrying the c.5664+2T>C mutation had occurred in platelets.

Candidate intronic mutations

The **c.7081+6G>T** mutation, identified in a type 1 VWD patient (UMP04), generated a new GT dinucleotide in intron 41 (Online Supplementary Table S3). The exon 38-43 region showed the expected PCR product in both cell types (772 bp). To confirm these results, two informative SNPs (rs216321 in exon 20 and rs216902 in exon 35) were genotyped. Both were in heterozygous state, indicating that the allele carrying the mutation was unaffected by NMD.

The **c.7730-56C>T** mutation (intron 45) was identified in a type 1 VWD patient (UMP05). The exon 43-49 region, examined by Sanger and NGS, showed no visible effect on mRNA splicing. To confirm the presence of the allele carrying the intronic mutation, two informative SNPs were analyzed: rs216321 and rs1800380. Surprisingly, the sequence obtained in both cell types

indicated that only one allele was expressed. To further explore these results, we tested whether intron 45 was retained within the mature mRNA (Online Supplementary Methods) and performed complete sequencing of *VWF* cDNA from leukocytes. Nonetheless, no changes were observed (data not shown).

The **c.7730-4C>G** mutation (intron 45) was identified in a type 2A/2M patient (UMP06), combined with the p.Arg1374Cys mutation. The HSF predicted activation of a cryptic intronic acceptor splice site (ASS), but in the exon 43-49 region, splicing was not affected by the mutation. To confirm the presence of the allele carrying c.7730-4C>G, the p.Arg1374Cys mutation was analyzed, and both nucleotides were seen in platelets and leukocytes, suggesting expression of both alleles. We then performed the same experiment as was done in patient UMP05 to test intron 45 retention, but no changes in *VWF* cDNA were observed (data not shown).

The **c.8254-5T>G** mutation (intron 51) was found in a type 1 VWD patient (UMP07). The exon 49-52 region was amplified, and the expected 546-bp size was observed in both cell types. Sequence analysis by the 2 techniques revealed no changes in *VWF* cDNA. Interestingly, this patient's 2 children, heterozygous for the c.8254-5T>G mutation, had bleeding symptoms. Hence, we performed complete sequencing of *VWF* cDNA from leukocytes. However, no changes were identified.

Synonymous mutations

The mutations **c.546G>A** (exon 6) and **c.4866C>T** (exon 28) were identified *in trans* in a type 1 VWD patient (UMP08). These mutations were studied only in leukocyte RNA. To investigate c.546G>A, the exon 4-7 region was amplified, which resulted in the expected band and a slightly diffuse, smaller band. Sanger sequencing detected the c.546G>A change (Figure 2). However, NGS not only detected the nucleotide change, but also recognized a loss of 125 nucleotides corresponding to exon 6 (p.Thr179ProfsTer31) in small 2.3% of reads (Online Supplementary Table S4). The predicted impact of the c.546G>A mutation was discrepant and only two *in silico* algorithms predicted an effect on mRNA splicing (Online Supplementary Table S3) that was confirmed *in vivo* even though with a slight effect. To study c.4866C>T, the exon 25-31 region was amplified and sequenced, but no effect on mRNA processing was seen, as was predicted by *in silico* tools. Both nucleotides were identified in heterozygous state, indicating the presence of both alleles. Only offspring carrying the c.546G>A mutation (1 of 3 children) had lower levels of VWF:Ag and VWF:RCo compared to the others. This may be explained by the mutation and/or other factors not explored in the current study.

The **c.3291C>T** mutation (exon 25), was identified in patient UMP09 (female), as well as a novel mutation in *F8* (c.1346C>G, p.Ala430Gly). Of note, her brother, who was not included in this study, also had these mutations. As both siblings experienced bleeding, and the HSF prediction regarding c.3291C>T interpreted creation of an exonic splicing silencer site that could potentially alter splicing, we hypothesized that this synonymous mutation could have an effect on *VWF* mRNA processing. Hence, the exon 22-26 region was amplified, but no change in the *VWF* cDNA sequence was observed. In addition, we tested whether intron 25 was retained within mature *VWF* mRNA (Online Supplementary Methods), but no differences compared to control leukocyte cDNA were found (data not shown). Finally, as both siblings presented similar FVIII:C levels, we investigated X chromosome inactivation in patient UMP09. The results demonstrated skewed inactivation of the WT X chromosome in this patient (Online Supplementary Figure S5).

Missense mutations

The mutation **c.1109G>A** (p.Cys370Tyr, exon 9) was identified in a type 3 VWD carrier (UMP10). The exon 6-12 region was investigated. Whereas the expected size was observed in platelets, amplification from leukocytes resulted in 2 additional bands of ~650 and ~500 bp (Figure 3). Sanger sequencing of leukocyte mRNA confirmed that the 650-bp band corresponded to a transcript lacking exon 9. NGS of leukocyte mRNA revealed 3 different transcripts: the WT, a transcript skipping exon 9 (p.Glu333AlafsTer87) (13% of reads), and a transcript skipping exons 8 and 9 (p.Ser292ThrfsTer87) (7% of reads)(Table 2). Sequencing of PCR products from platelets showed no changes, suggesting that the mutated allele had experienced NMD.

Duplication mutation

The **c.3223-7_3236dup** mutation, which generated an in-frame tandem duplication of 21 nucleotides, including 7 nucleotides from intron 24 and the first 14 from exon 25, was detected in a type 1 VWD patient (UMP11). *S*tudy of this mutation by the 2 techniques demonstrated a change in the native ASS of intron 24 in both cell types, which resulted in maintaining the 21-bp insert corresponding to 7 aberrant amino acids in mature *VWF* mRNA (p.Pro1079_Tyr1080insLeuGlnValAspProGluPro) (Table 2 and Online Supplementary Figure S6). Furthermore, NGS showed that the aberrant mRNA transcript was present in ~14% of reads in both cell types (Online Supplementary Table S4).

Combined potential splice site mutations in individual patients

The **c.7082-2A>G** mutation (intron 41) was identified in a type 3 VWD patient (UMP12) combined *in trans* with the **p.Leu150Pro** (c.449C>T, exon 5) mutation. To study c.7082-2A>G, the exon 38-43 region was amplified in platelets and leukocytes, which yielded a band of the expected size. However, Sanger and NGS sequencing of the PCR products revealed activation of a cryptic ASS within exon 42. This led to deletion of the 7 initial nucleotides of exon 42 (r.7082_7088del), as was predicted by the 4 algorithms, and resulted in a frameshift and a PTC (p.Ala2361GlyfsTer40) (Figure 4). Additionally, NGS showed that the aberrant transcript was present in 24% of reads. On sequencing of platelet amplicons, there were no changes, a finding highly suggestive of NMD. As type 3 VWD is characterized by absent or non-functional expression of both *VWF* alleles, we postulated that the p.Leu150Pro missense mutation could have an effect on *VWF* mRNA. Nonetheless, no sequence change was identified. These results were confirmed by analysis of 2 informative SNPs, rs1800375 and rs1800376, which were found in heterozygous state in leukocyte transcripts and homozygous state in platelets.

The **c.[3426T>C; 3485_3486delinsTG]** mutations (exon 26) were found in a type 2A VWD patient (UMP13), combined *in trans* with p.Cys2773Ser. *In silico* analysis predicted no impact on the splice site for c. 3485_3486delinsTG, but 2 of the 4 algorithms predicted that c.3426T>C would have an effect on splicing. The exon 25-28 region was investigated, but no effect was found on mRNA processing. Study of the p.Cys2773Ser mutation and 2 informative SNPs (rs2228317 and rs4021576) confirmed that the 2 alleles were present, suggesting that exon 26 mutations do not have an effect on *VWF* splicing.

The mutations **c.7473G>A** (p.=, exon 43), and **c.6699_6702dup** (p.Cys2235Argfs8Ter, exon 38) were identified in *trans* in a severe type 1 VWD patient (UMP14) and studied only in leukocyte RNA. For c.7473G>A, we amplified exons 41-45, and 2 bands emerged: one with the expected size, and a smaller band of □□350 bp (Figure 5). Sequencing analysis by both techniques revealed 2 aberrant transcripts: one lacked exon 43 (p.Val2430GlyfsTer335) and the other showed activation of a cryptic DSS within exon 43, leading to deletion of its 4 final nucleotides (p.Ser2479AlafsTer23) (Table 2). Study of c.6699_6702dup in leukocyte mRNA showed no splicing changes, but the duplication introduced 4 nucleotides that changed the reading frame and generated a PTC (p.Cys2235ArgfsTer8).

The **c.546G>A** (p.=, exon 6) and **c.7082-2A>G** (intron 41) mutations *in trans* with **c.8155+3G>C** (intron 50) were detected in a type 3 VWD patient (UMP15). This patient had been described in our previous article, but at that time we were only able to identify the effect of c.8155+3G>C (p.Gly2706ValfsTer24) on splicing. On reanalysis of this case with NGS, we were able to

characterize the effect of c.546G>A (p.Thr179ProfsTer31) and c.7082-2A>G (p.Ala2361GlyfsTer40) *in cis* in leukocyte mRNA. The results are consistent with those observed in patients UMP08 and UMP12, who harbored each mutation individually. In addition, in leukocytes, transcripts resulting from c.546G>A were detected in 2% of reads, transcripts from c.7082-2A>G in 6% and transcripts from c.8155+3G>C in 95% of reads. In platelet analyses, NGS detected transcripts without exon 6 resulting from the c.546G>A mutation in 11% of the total reads but no aberrant transcript derived from the mutation c.7082-2A>G indicating that NMD rate could differ between 5' and 3' regions of mRNA. These results indicate the c.[546G>A; 7082-2A>G] allele is under-represented compared to the allele carrying the c.8155+3G>C mutation, showing that this allele underwent NMD in both cell types, as was previously hypothesized.[7]

DISCUSSION

This study reports the results of in-depth analysis of 18 PSSM in samples from VWD patients. The novel and robust procedure used, combining two-step RT-PCR and NGS sequencing, was faster and more sensitive than the method used in our previous article.[7] This new approach provides several advantages, such as allele-specific individual sequences, higher sensitivity to detect transcript variants present at low copy numbers, and simplified sample preparation. Moreover, the NGS data on total reads obtained for each transcript is additional information that can provide an approximation of the expression levels of each *VWF* mRNA. Studies are currently ongoing to confirm that the relative expression of each transcript obtained by NGS is comparable to that provided by real-time RT-PCR.

The proven usefulness of leukocyte analysis to interpret mutations masked by NMD in platelets was seen in relation to the **c.3379+1G>A** mutation, which had been previously investigated by RT-PCR in platelets.[23] In that study, the true pathogenic effect of this mutation could not be determined due to NMD. In the present study, using leukocyte mRNA and NGS technology, we found that the mutation induces two aberrant transcripts.

Similarly, the pathogenic mechanism of 7 PSSM (c.1533+1G>A, c.5664+2T>C, c.7082-2A>G, c.546G>A, c.7437G>A, and p.Cys370Tyr) was determined in leukocytes, and only the **c.3223-7_3236dup** mutation effect was observed in both cell types. Because this mutation is located in the D3 domain implicated in multimerization, it is possible that the pathogenic mechanism leading to type 1 VWD may be related to impaired VWF secretion due to intracellular retention, as has been described for other D3 mutations such as p.Cys1130Phe.[24] To confirm this hypothesis, expression studies by means of heterologous cell lines or blood outgrowth endothelial cells (BOECs) will remain the gold standard.

As would be expected, mutations located in the consensus splicing sequence affected mRNA processing. These included **c.1533+1G>A, c.5664+2T>C,** leading to exon skipping, and **c.7082-2A>G,** activating a cryptic splice site. Both these molecular effects have been reported previously in splicing mutations causing VWD.[7, 25, 26] The c.7082-2A>G *in trans* with p.Leu150Pro leads to the development of type 3 VWD. The p.Leu150Pro mutation does not affect splicing. However, other mutations located in the propeptide, such as p.Asp141Tyr, have been identified in type 3 patients,[27] and, as has been described in these cases, we suspect that p.Leu150Pro may compromise propeptide folding and affect intracellular survival and the capacity to mediate multimerization. The c.1533+1G>A mutation in leukocytes generated 3 aberrant transcripts. One of them, which leads to the inframe deletion of exons 13-14, should

have been detected in platelets since it would not generate a PTC and would not be degraded. However, because the mutation is located in a region with weak splicing signal sequences, as previously demonstrated,[7, 25, 26] we suggest that the results obtained in leukocytes could be an artifact that would not occur in the natural cellular type of VWF expression. Based on these observations, we propose that the real effect of c.1533+1G>A is the skipping of exon 13, which predicts a frameshift at position 478 and addition of 138 aberrant amino acids before a PTC is encountered (p.Gly478AlafsTer138).

Two synonymous mutations included in our study, **c.546G>A** and **c.7437G>A**, affect splicing of *VWF* mRNA. To our knowledge, only 3 synonymous mutations affecting *VWF* mRNA processing have been reported: c.7056C>T,[28] c.7464C>T,[29] and c.3390C>T.[26] Of note, the results of the present study have almost doubled the number of reported mutations of this type causing VWD. The pathogenic effect of **c.546G>A** could only be documented by NGS; it eluded Sanger detection because of the low expression in leukocytes and platelets. The **c.7437G>A** mutation located at the last nucleotide of exon 43 was found in a patient with severe type 1 VWD (UMP14) *in trans* with c.6699_6702dup (p.Cys2235ArgTer8). This synonymous mutation produced 2 splicing variants. First, creation of a new DSS 4 nucleotides upstream of the WT-DSS in exon 43 that generated a PTC in exon 44 and would lead to NMD in platelets; and second, exon 43 skipping and generation of a PTC in exon 52. In this latter case, we would expect that NMD had been abolished in platelets, since this cellular mechanism is not effective when a PTC is encountered within 50 bp of the last exon-exon junction.[30]

Study of the **p.Cys370Tyr** (c.1109G>A in exon 9) mutation, resulted in generation of 2 *VWF* transcripts, one of them identified only by NGS because of their low expression. Of note, only one other missense mutation, p.Gly1108Arg, has been reported to affect *VWF* splicing. Thus, these 2 mutations support the concept that missense mutations on the last exonic nucleotides can also have repercussions on the splicing process.[31] Interestingly, the pathogenic effect of this mutation is the same as that described for c.1109+2T>C (intron 9).[32] In addition to the patient reported here, the p.Cys370Tyr mutation was identified in 6 additional related patients included in the PCM-EVW-ES: in homozygous state in a type 3 VWD patient and in heterozygosis in 4 type 3 carriers and 1 patient with type 1 VWD, based on their phenotype levels. Certain genetic modifiers of VWF levels[33] and inter-individual variability in NMD efficiency between patients carrying identical mutations may lead to differences in the disease severity and clinical phenotype.[34]

Lastly, study of c.7081+6G>T, c.7730-4C>G, c.7730-56C>T, c.8254-5T>G, c.3291C>T, c.4866C>T and c.[3426T>C; 3485_3486delinsTG] showed no visible effect on mRNA processing, although this does not necessarily mean that they have no effect on splicing. For instance, mRNA SNP analysis in patient UMP05 with **c.7730-56C>T** showed an absence of 1 allele in leukocytes and platelets, suggesting that the allele may have experienced NMD in both cell types or lack expression due to a mutation that was not detected by our sequencing protocol. Of particular note, deep intronic mutations such as c.6599-20A>T have been found to cause VWD.[35] In patient UMP07 carrying the **c.8254-5T>G** mutation**,** no informative SNP was identified and we were unable to determine whether there was a lack of expression of 1 allele, which could explain cosegregation of the mutation in the family with bleeding symptoms.

Study of **c.[3426T>C; 3485_3486delinsTG]** and **c.7730-4C>G** showed no visible effect on splicing. However, these mutations have been identified in homozygous state in a type 3 patient (c.[3426T>C; 3485_3486delinsTG])[3] and in a severe type 1 VWD patient (c.7730-4C>G)[4] with VWF:Ag at 7% and VWF:RCo at 5%, combined with the heterozygous p.Ala631Val (previously reported in a healthy control).[36] Based on these findings, there is substantial evidence that these mutations would have an effect on VWF levels. Therefore, to unequivocally determine the potential deleterious effect of these variants, functional studies remain essential. These studies are traditionally undergoing by *in vitro* analyses performed using heterologous cell lines (COS7, AtT-20 and HEK293). However, the advent of the possibility of obtaining BOECs from patients represents a valuable alternative, since this is the functional expression site of *VWF*.[37] Moreover, BOECs allows protein expression and mRNA studies simultaneously.

In silico algorithms used to assess the impact of mutations on splicing are more sensitive and accurate in determining the putative effect of intronic mutations than that of synonymous or missense mutations, such as c.546G>A and p.Cys370Tyr. Therefore, the results here should not be considered definitive, and as with all analytical approaches, should form one aspect of a wider investigation.[5]

In conclusion, we present an extensive study reporting the effect of 18 candidate mutations on *VWF* mRNA processing. *In vivo* mRNA studies incorporating NGS technology together with traditional sequencing enabled us to determine the pathogenic effect of 8 PSSM (44%). Our study emphasizes the importance of examining selected mutations, including synonymous and missense mutations, to determine their pathogenic role in splicing. Taken together, our results add to the current knowledge about the molecular events leading to VWD.

REFERENCES

1. Rodeghiero F, Castaman G, Dini E. Epidemiological investigation of the prevalence of von Willebrand's disease. Blood. 1987;69(2):454-459.
2. Batlle J, Perez-Rodriguez A, Corrales I, et al. Molecular and clinical profile of von Willebrand disease in Spain (PCM-EVW-ES): Proposal for a new diagnostic paradigm. Thromb Haemost. 2016;115(1):40-50.
3. Borras N, Batlle J, Perez-Rodriguez A, et al. Molecular and clinical profile of von Willebrand disease in Spain (PCM-EVW-ES): comprehensive genetic analysis by next-generation sequencing of 480 patients. Haematologica. 2017;102(12):2005-2014.
4. Fidalgo T, Salvado R, Corrales I, et al. Genotype-phenotype correlation in a cohort of Portuguese patients comprising the entire spectrum of VWD types: impact of NGS. Thromb Haemost. 2016;116(1):17-31.
5. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-424.
6. Wang QY, Song J, Gibbs RA, Boerwinkle E, Dong JF, Yu FL. Characterizing polymorphisms and allelic diversity of von Willebrand factor gene in the 1000 Genomes. J Thromb Haemost. 2013;11(2):261-269.
7. Corrales I, Ramirez L, Altisent C, Parra R, Vidal F. The study of the effect of splicing mutations in von Willebrand factor using RNA isolated from patients' platelets and leukocytes. J Thromb Haemost. 2011;9(4):679-688.
8. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Re Genet. 2007;8(10):749-761.
9. Eikenboom JC, Reitsma PH, Peerlinck KM, Briët E. Recessive inheritance of von Willebrand's disease type I. Lancet. 1993;341(8851):982-986.
10. Jaffe EA, Hoyer LW, Nachman RL. Synthesis of von Willebrand factor by cultured human endothelial cells. Proc Natl Acad Sci U S A. 1974;71(5):1906-1909.
11. Sporn LA, Chavin SI, Marder VJ, Wagner DD. Biosynthesis of von Willebrand protein by human megakaryocytes. J Clin Invest. 1985;76(3):1102-1106.
12. Kieffer N, Guichard J, Farcet JP, Vainchenker W, Breton-Gorius J. Biosynthesis of major platelet proteins in human blood platelets. Eur J Biochem. 1987;164(1):189-195.
13. Fink L, Holschermann H, Kwapiszewska G, et al. Characterization of platelet-specific mRNA by real-time PCR after laser-assisted microdissection. Thromb Haemost. 2003;90(4):749-756.
14. McVey JH, Boswell EJ, Takamiya O, et al. Exclusion of the first EGF domain of factor VII by a splice site mutation causes lethal factor VII deficiency. Blood. 1998;92(3):920-926.
15. David D, Santos IM, Johnson K, Tuddenham EG, McVey JH. Analysis of the consequences of premature termination codons within factor VIII coding sequences. J Thromb Haemost. 2003;1(1):139-146.
16. Roberts RG, Bentley DR, Bobrow M. Infidelity in the structure of ectopic transcripts: a novel exon in lymphocyte dystrophin transcripts. Hum Mutat. 1993;2(4):293-299.
17. Byers PH. Killing the messenger: new insights into nonsense-mediated mRNA decay. J Clin Invest. 2002;109(1):3-6.
18. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A. 2003;100(1):189-192.
19. Martorell L, Luce E, Vazquez JL, et al. Advanced cell-based modeling of the royal disease: characterization of the mutated F9 mRNA. J Thromb Haemost. 2017;15(11):2188-2197.
20. Hojny J, Zemankova P, Lhota F, et al. Multiplex PCR and NGS-based identification of mRNA splicing variants: Analysis of BRCA1 splicing pattern as a model. Gene. 2017;637:41-49.
21. Brunak S, Engelbrecht J, Knudsen S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol. 1991;220(1):49-65.
22. Corrales I, Ramirez L, Altisent C, Parra R, Vidal F. Rapid molecular diagnosis of von Willebrand disease by direct sequencing. Detection of 12 novel putative mutations in VWF gene. Thromb Haemost. 2009;101(3):570-576.

23. Nesbitt IM, Hampton KK, Preston FE, Peake IR, Goodeve AC. A common splice site mutation is shared by two families with different type 2N von Willebrand disease mutations. Thromb Haemost. 1999;82(3):1061-1064.
24. Tjernberg P, Vos HL, Castaman G, Bertina RM, Eikenboom JC. Dimerization and multimerization defects of von Willebrand factor due to mutated cysteine residues. J Thromb Haemost. 2004;2(2):257-265.
25. Gallinaro L, Sartorello F, Pontara E, et al. Combined partial exon skipping and cryptic splice site activation as a new molecular mechanism for recessive type 1 von Willebrand disease. Thromb Haemost. 2006;96(6):711-716.
26. Pagliari MT, Baronciani L, Garcia Oya I, et al. A synonymous (c.3390C>T) or a splice-site (c.3380-2A>G) mutation causes exon 26 skipping in four patients with von Willebrand disease (2A/IIE). J Thromb Haemost. 2013;11(7):1251-1259.
27. Baronciani L, Federici AB, Cozzi G, et al. Expression studies of missense mutations p.D141Y, p.C275S located in the propeptide of von Willebrand factor in patients with type 3 von Willebrand disease. Haemophilia. 2008;14(3):549-555.
28. Daidone V, Gallinaro L, Grazia Cattini M, et al. An apparently silent nucleotide substitution (c.7056C>T) in the von Willebrand factor gene is responsible for type 1 von Willebrand disease. Haematologica. 2011;96(6):881-887.
29. Yadegari H, Biswas A, Akhter MS, et al. Intron retention resulting from a silent mutation in the VWF gene that structurally influences the 5' splice site. Blood. 2016;128(17):2144-2152.
30. Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol. 2004;5(2):89-99.
31. James PD, O'Brien LA, Hegadorn CA, et al. A novel type 2A von Willebrand factor mutation located at the last nucleotide of exon 26 (3538G>A) causes skipping of 2 nonadjacent exons. Blood. 2004;104(9):2739-2745.
32. Castaman G, Plate M, Giacomelli SH, Rodeghiero F, Duga S. Alterations of mRNA processing and stability as a pathogenic mechanism in von Willebrand factor quantitative deficiencies. J Thromb Haemost. 2010;8(12):2736-2742.
33. James PD, Lillicrap D. von Willebrand disease: clinical and laboratory lessons learned from the large von Willebrand disease studies. Am J Hematol. 2012;87 Suppl 1:S4-11.
34. Miller JN, Pearce DA. Nonsense-mediated decay in genetic disease: friend or foe? Muta Res Rev Mutat Res. 2014;762:52-64.
35. Hawke L, Bowman ML, Poon MC, Scully MF, Rivard GE, James PD. Characterization of aberrant splicing of von Willebrand factor in von Willebrand disease: an underrecognized mechanism. Blood. 2016;128(4):584-593.
36. Bellissimo DB, Christopherson PA, Flood VH, et al. VWF mutations and new sequence variations identified in healthy controls are more frequent in the African-American population. Blood. 2012;119(9):2135-2140.
37. Wang JW, Bouwens EA, Pintao MC, et al. Analysis of the storage and secretion of von Willebrand factor in blood outgrowth endothelial cells derived from patients with von Willebrand disease. Blood. 2013;121(14):2762-2772.

TABLES

Table 1. Laboratory and molecular data of VWD patients

Patients code	VWD	VWF:Ag	VWF:RCo	FVIII:C	NT change	AA change	Exon	Intron	Domain
UMP01	3 carrier	70	70.5	111	**c.1533+1G>A**	-	-	13	intronic
UMP02	1H	36	49	101	**c.3379+1G>A** *	-	-	25	intronic
					c.-2627C>T *	-	-	-	upstream
UMP03	1	26	21.2	64.8	**c.5664+2T>C** *	-	-	33	intronic
					c.7220T>C *	p.Leu2407Pro	42	-	C2
UMP04	1	50	52	47	**c.7081+6G>T**	-	-	41	intronic
UMP05	1	27	32	72	**c.7730-56C>T**	-	-	45	intronic
UMP06	2A/2M	12	4	24	c.4120C>T *	p.Arg1374Cys	28		A1
					c.7730-4C>G *	-	-	45	intronic
UMP07	1	37	28	60	**c.8254-5T>G**	-	-	51	intronic
UMP08	1	42.4	41.7	73.4	**c.546G>A** †	p.Ser182	6	-	D1
					c.4866C>T †	p.Asp1622	28	-	A1-A3
UMP09 ‡	1	51	49.8	71	**c.3291C>T**	p.Cys1097	25	-	D3
UMP10	3 carrier	46.4	45	105	**c.1109G>A**	p.Cys370Tyr	9	-	D1
UMP11	1	31	24	57	**c.3223-7_3236dup**	p.Pro1079_Tyr1080ins LeuGlnValAspProGlu Pro	25	-	D3
UMP12	3	3.7	5	6.7	**c.449T>C** †	p.Leu150Pro	5	-	D1
					c.7082-2A>G †	-	-	41	intronic
UMP13	2A	117	6.5	90	**c.3426T>C** §	p.Cys1142	26	-	D3
					c.3485_3486delinsTG §	p.Pro1162Leu	26	-	D3
					c.8318G>C †	p.Cys2773Ser	52	-	CK
UMP14	1	6	6	4,5	**c.6699_6702dup** †	p.Cys2235ArgTer8	38	-	D4
					c.7437G>A †	p.Ser2479	43	-	C3
UMP15||	3	2	<1	5	c.546G>A §	p.Ser182	6	-	D1
					c.7082-2A>G §	-	-	41	intronic
					c.8155+3G>C †	-	-	50	intronic

All mutations were identified in heterozygous state. In bold, mutations selected to study their effect on *VWF* mRNA**.** The subtype 1H (historical) refers to patients previously diagnosed as type 1 VWD that, at the time of enrollment at the PCM-EVW-ES project, show a slight decrease or even a normal VWF plasma levels. AA indicates amino acid; NT, nucleotide and; VWD, von Willebrand disease. *The allelic phase (cis/trans) of mutations could not be determined since no informative relatives are available. †Mutations *in trans*. ‡Patient with an additional mutation in the F8 gene. §Mutations *in cis*. ||Patient previously studied at mRNA level by Corrales et al.[7]

Table 2. Effect of VWF mutations on mRNA from leukocytes and platelets

Mutation type	NT change	AA change	Exon	Intron	Leukocyte effect	Platelet effect	Protein prediction
Intronic	**c.1533+1G>A**	-	-	13	Exon 13 skipping (r.1433_1533del)	NMD	p.Gly478AlafsTer138
	c.3379+1G>A*	-	-	25	Exon 25 skipping (r.3223_3379del)	NMD	p.Pro1075ValfsTer88
					Activation of a cryptic site at +126 nt in exon 25 (r.3349_3379del) ‡	NMD	p.Pro1117ValfsTer88
	c.5664+2T>C	-	-	33	Exon 33 skipping (r.5621_5664del)	NMD	p.Gly1874AlafsTer32
					Exon 33+34 skipping (r.5621_5842del) ‡	NMD	p.Phe1875_Cys1948del
	c.7081+6G>T	-	-	41	*No visible effect*	*No visible effect*	-
	c.7082-2A>G †	-	-	41	Activation of a cryptic site +7 nt in exon 42 (r.7082_7088del)	NMD	p.Ala2361GlyfsTer40
	c.7730-4C>G	-	-	45	*No visible effect*	*No visible effect*	-
	c.7730-56C>T	-	-	45	*No visible effect*	*No visible effect*	-
	c.8254-5T>G	-	-	51	*No visible effect*	*No visible effect*	-
Synonymous	**c.546G>A** †	p.Ser182	6	-	Exon 6 skipping (r.533_657del) ‡	NA	p.Thr179ProfsTer31
	c.3291C>T	p.Cys1097	25	-	*No visible effect*	*No visible effect*	-
	c.3426T>C	p.Cys1142	26	-	*No visible effect*	*No visible effect*	-
	c.4866C>T	p.Asp1622	28	-	*No visible effect*	NA	-
	c.7437G>A	p.Ser2479	43	-	Exon 43 skipping (r.7288_7437)	NA	p.Val2430GlyfsTer335
					Activation of a cryptic site at +146 nt in exon 43 (r.7434_7437del)	NA	p.Ser2479AlafsTer23
Missense	**c.449T>C**	p.Leu150Pro	5	-	*No visible effect*	*No visible effect*	-
	c.1109G>A	p.Cys370Tyr	9	-	Exon 9 skipping (r.998_1109del)	NMD	p.Glu333AlafsTer87
					Exon 8+9 skipping (r.875_1109del)	NMD	p.Ser292ThrfsTer87
DelIns	**c.3485_3486delinsTG**	p.Pro1162Leu	26	-	*No visible effect*	*No visible effect*	-
	c.3223-7_3236dup	p.Pro1079_Tyr1080insLeu GlnValAspProGluPro	25	-	(r.3477_3478instttgcaggtggaccccgagcc)	(r.3477_3478instttgcaggtggaccccgagcc)	p.Pro1079_Tyr1080insLeuGlnValAspProGluPro
	c.6699_6702dup	p.Cys2235ArgTer8	38	-	*No visible effect*	*NMD*	p.Cys2235ArgTer8

AA indicates amino acid; NA, not available; NMD, nonsense-mediated decay and; NT, nucleotide. *Previous functional studies by Nesbitt *et al.*[23] †Previous functional studies by Corrales *et al.*[7] ‡Leukocyte effect observed in a really low % of reads by NGS.

FIGURE LEGENDS

Figure 1. Analysis of the c.5664+2T>C mutation in patient UMP03. **A)** RT-PCR products amplified with primers located in exons 30 and 35 in leukocyte (L) and platelet (P) RNA, separated on 1% agarose gel. **B)** Traditional Sanger sequencing of PCR products from patient leukocytes (L) and platelets (P), and control leukocytes (L). Analysis of the band 2 from patient leukocyte on agarose gel demonstrates exon 33 skipping. In platelets, only the allele without c.5664+2T>C mutation could be amplified. **C)** NGS of PCR products from leukocytes showed exon 33 skipping, indicated by arrows depict aberrant transcripts. However, exon 33-34 skipping was also detected, but in a really low of transcripts. **D)** Schematic representation of the mutation in genomic DNA and its effect on the *VWF* mRNA sequence. M indicates a 100-bp DNA ladder.

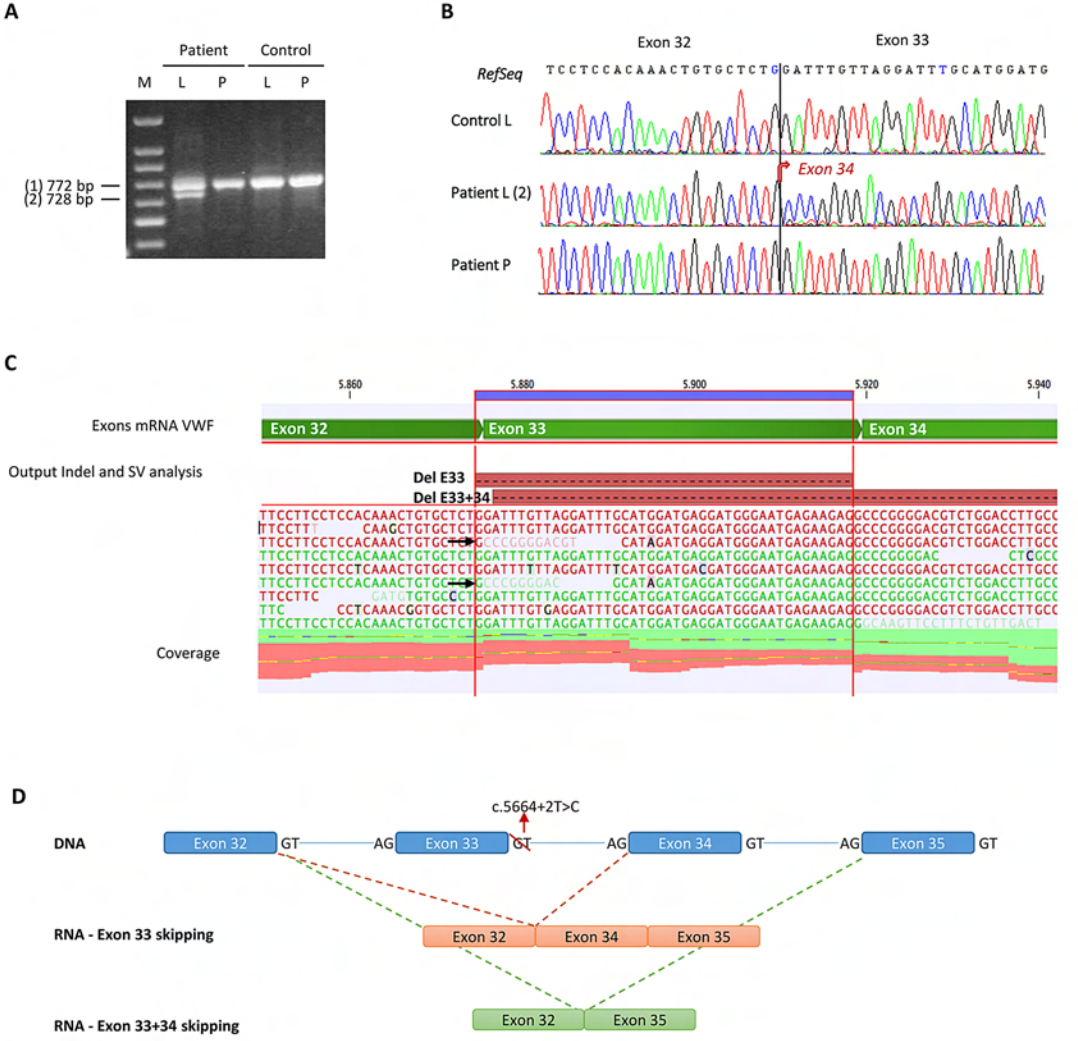
Figure 2. Analysis of the c.546G>A (p.=) mutation, located at nucleotide 14 from the beginning of exon 6, in patient UMP08. **A)** RT-PCR products amplified with primers located in exons 4 and 7 in leukocyte RNA (L), separated on 1% agarose gel. **B)** Traditional Sanger sequencing of PCR product from patient L (1) showed the single nucleotide change. **C)** NGS of PCR products from leukocytes identified the single nucleotide variant, as well as exon 6 skipping. Arrows show aberrant transcripts. **D)** Schematic representation of the mutation in genomic DNA and its effect on the *VWF* mRNA sequence. M indicates a 100-bp DNA ladder.

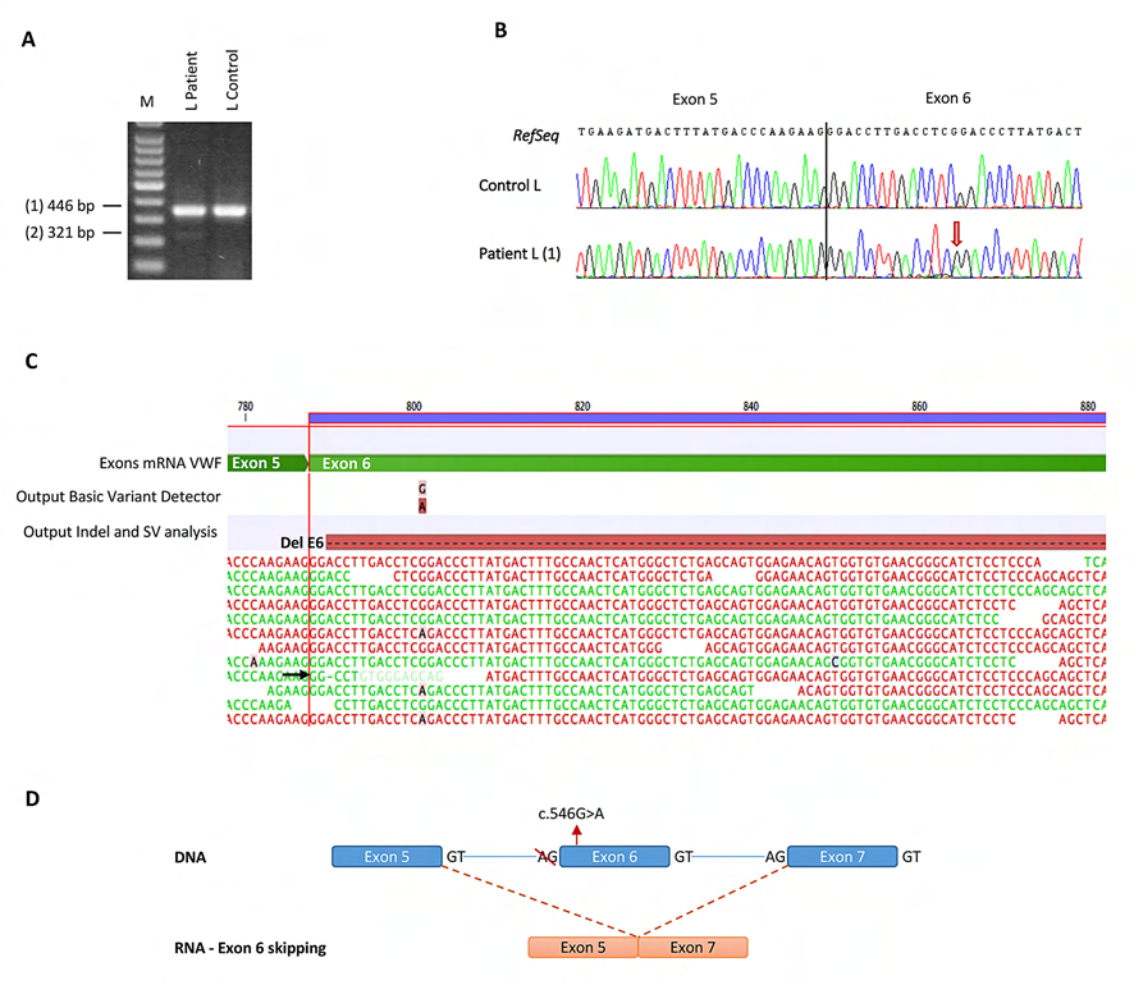
Figure 3. Analysis of the c.1109G>A (p.Cys370Tyr) mutation in patient UMP10, located in the last nucleotide of exon 9. **A)** RT-PCR products amplified with primers located in exons 6 and 12 in leukocyte (L) and platelet (P) RNA, separated on 1% agarose gel. **B)** Traditional Sanger sequencing of PCR product from patient L (2) showed exon 9 skipping. The L (3) band could not be purified in the agarose gel due to low concentration, thus it was not analyzed by Sanger. In platelets, only the non-mutated allele could be amplified. **C)** NGS of PCR products from leukocytes identified transcripts lacking exon 9, as well as transcripts lacking exons 8 and 9. **D)** Schematic representation of the mutation in genomic DNA and its effect on the *VWF* mRNA sequence. M indicates a 100-bp DNA ladder.

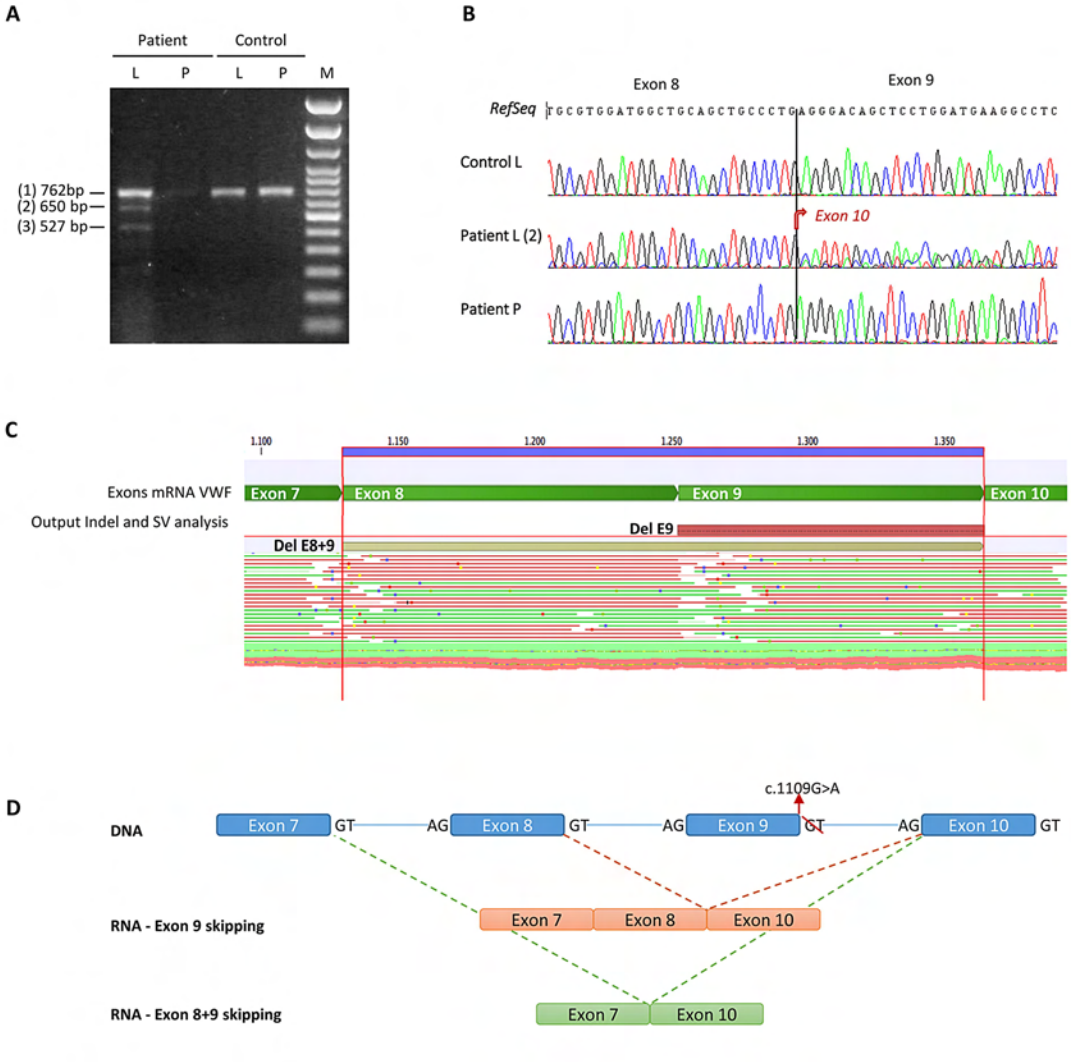
Figure 4. Analysis of the c.7082-2A>G mutation in patient UMP12. Agarose gel electrophoresis results of RT-PCR amplification of exon 38 to 43 using RNA from leukocytes and platelets were the same as those of healthy controls (data not shown). **A)** Traditional Sanger sequencing of PCR product from patient leukocyte (L) demonstrate activation of a cryptic splice site 7 nucleotides downstream of the native splice site within exon 42. In patient platelet (P), only the allele carrying the p.Leu150Pro mutation could be amplified. **B)** NGS of PCR products from

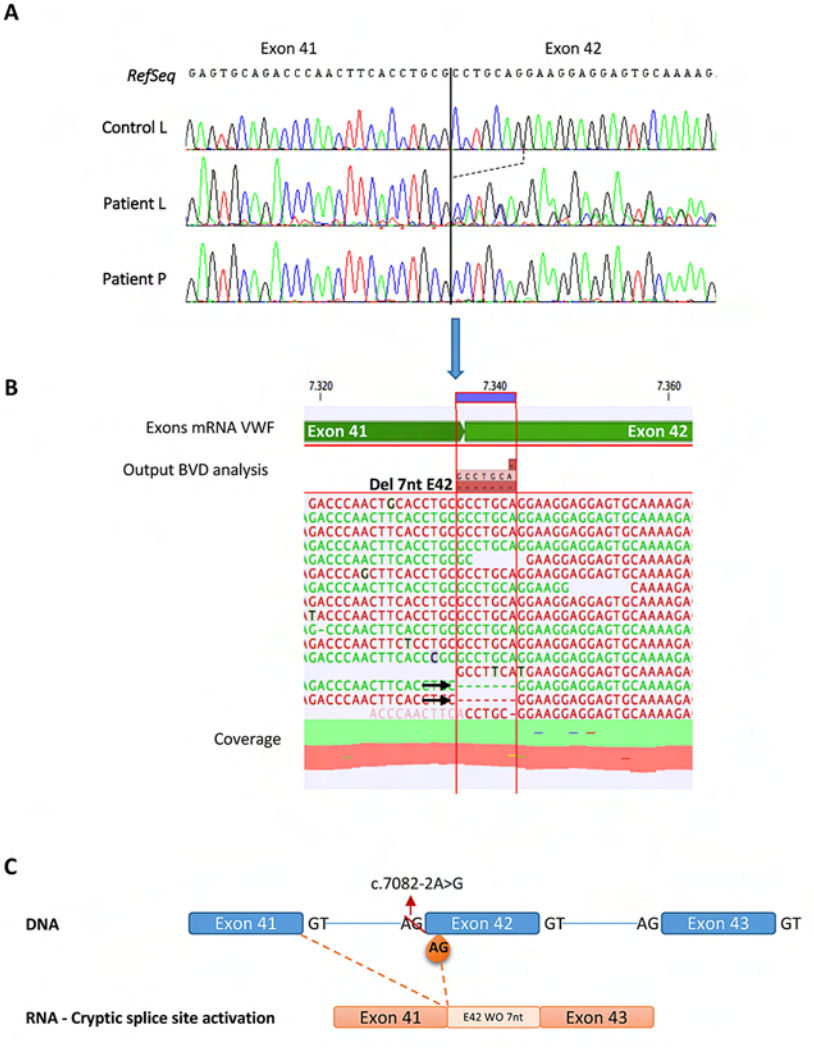
leukocytes showed deletion of the 7 initial nucleotides of exon 42. Arrows show aberrant transcripts. **C)** Schematic representation of the mutation in genomic DNA and its effect on the *VWF* mRNA sequence. WO indicates without.

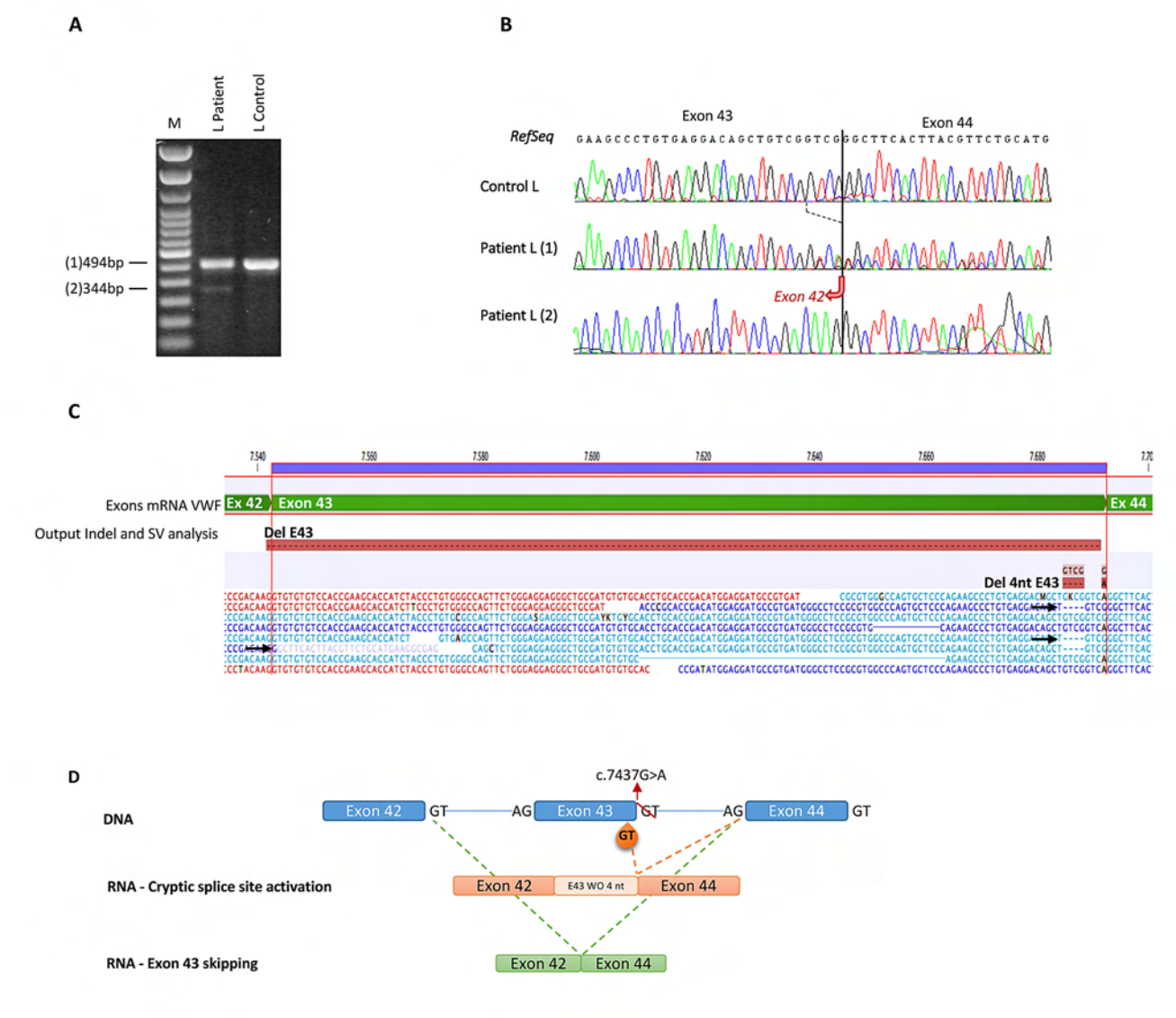
Figure 5. Analysis of the c.7437G>A (p.=) mutation in patient UMP14, located in the last nucleotide of exon 43. **A)** RT-PCR products amplified with primers located in exons 41 and 45 in leukocyte RNA (L), separated on 1% agarose gel. **B)** Traditional Sanger sequencing of PCR product from patient leukocyte (L) shows two aberrant transcripts: activation of a cryptic splice site—4 nucleotides upstream to WT-DSS—in exon 43 (1), and exon 43 skipping (2). **C)** NGS of PCR products gave the same results than Sanger sequencing. **D)** Schematic representation of the mutation in genomic DNA and its effect on the *VWF* mRNA sequence. M indicates a 100-bp DNA ladder and; WO, without.


A
Patient
Control
M
L
P
L
P
(1) 772 bp
(2) 728 bp
B
Exon 32
Exon 33
RefSeq
TCCTCCACAAACTGTGCTCTGGATTTGTTAGGATTTGCATGGATG
Control L
Exon 34
Patient L (2)
Patient P
C
5.860
5.880
5.900
5.920
5.940
Exons mRNA VWF
Exon 32
Exon 33
Exon 34
Output Indel and SV analysis
Del E33
Del E33+34
Coverage
D
c.5664+2T>C
DNA
Exon 32
GT
AG
Exon 33
GT
AG
Exon 34
GT
AG
Exon 35
GT
RNA - Exon 33 skipping
Exon 32
Exon 34
Exon 35
RNA - Exon 33+34 skipping
Exon 32
Exon 35


A
M
L Patient
L Control
(1) 446 bp
(2) 321 bp
B
Exon 5
Exon 6
RefSeq
TGAAGATGACTTTATGACCCAAGAAGGGACCTTGACCTCGGACCCTTATGACT
Control L
Patient L (1)
C
780
800
820
840
860
880
Exons mRNA VWF
Exon 5
Exon 6
Output Basic Variant Detector
G
A
Output Indel and SV analysis
Del E6
D
c.546G>A
DNA
Exon 5
GT
AG
Exon 6
GT
AG
Exon 7
GT
RNA - Exon 6 skipping
Exon 5
Exon 7


A
Patient
Control
L
P
L
P
M
(1) 762bp
(2) 650 bp
(3) 527 bp
B
Exon 8
Exon 9
RefSeq
TGCGTGGATGGCTGCAGCTGCCCTGAGGGACAGCTCCTGGATGAAGGCCTC
Control L
Exon 10
Patient L (2)
Patient P
C
1.100
1.150
1.200
1.250
1.300
1.350
Exons mRNA VWF
Exon 7
Exon 8
Exon 9
Exon 10
Output Indel and SV analysis
Del E9
Del E8+9
D
c.1109G>A
DNA
Exon 7
GT
AG
Exon 8
GT
AG
Exon 9
GT
AG
Exon 10
GT
RNA - Exon 9 skipping
Exon 7
Exon 8
Exon 10
RNA - Exon 8+9 skipping
Exon 7
Exon 10


A
Exon 41
Exon 42
RefSeq
GAGTGCAGACCCAACTTCACCTGCGCCTGCAGGAAGGAGGAGTGCAAAAG
Control L
Patient L
Patient P
B
7.320
7.340
7.360
Exons mRNA VWF
Exon 41
Exon 42
Output BVD analysis
Del 7nt E42
GCCTGCA
Coverage
C
c.7082-2A>G
DNA
Exon 41
GT
AG
Exon 42
GT
AG
Exon 43
GT
AG
RNA - Cryptic splice site activation
Exon 41
E42 WO 7nt
Exon 43


A
M
L Patient
L Control
(1)494bp
(2)344bp
B
Exon 43
Exon 44
RefSeq
GAAGCCCTGTGAGGACAGCTGTCGGTCGGGCTTCACTTACGTTCTGCATG
Control L
Patient L (1)
Exon 42
Patient L (2)
C
Exons mRNA VWF
Ex 42
Exon 43
Ex 44
Output Indel and SV analysis
Del E43
Del 4nt E43
D
c.7437G>A
DNA
Exon 42
GT
AG
Exon 43
GT
AG
Exon 44
GT
GT
RNA - Cryptic splice site activation
Exon 42
E43 WO 4 nt
Exon 44
RNA - Exon 43 skipping
Exon 42
Exon 44

SUPPLEMENTARY DATA

SUPPLEMENTARY METHODS

VWF mRNA amplification

The primers used have been reported previously,[1] except for those specifically designed to analyze certain mutations (Online Supplementary Table S1). The PCR mix contained 1X PCR Buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 2 U Platinum Taq DNA polymerase, 0.75 µM of each primer, and cDNA in a total volume of 25 µL. After initial denaturation at 94ºC for 3 min, 38 cycles of 94ºC for 20 sec, 62ºC for 30 sec, and 72ºC for 1 min were performed, followed by a final extension at 72ºC for 3 min.

CLC Bio Data analysis parameters

Primer sequences and low-quality regions were trimmed and aligned against the WT *VWF* mRNA sequence with specific parameter settings. The Basic Variant Detection algorithm was applied for single nucleotide variant-calling and the Indels and Structural Variants tool was performed to identify insertions and deletions produced by alternative splicing due to mutations. Analytical settings of both algorithms are described in Online Supplementary Figure S1.

Finally, a specific probe was designed to each pre-characterized aberrant transcript and used to perform a second alignment from the original dataset. The results indicate the number of reads that aligned to each mRNA probe.

Intron retention of exon 25

To determine whether intron 25 was retained within the mature mRNA of patient UMP09, RT-PCR reactions using 2 allele-specific primer combinations were performed. In the first primer pair, the forward was designed at the junction of exon 24-25 (VWFR24/25-1) and the reverse primer was designed inside intron 25 (VWF_E25-2). In the second primer pair, the forward primer targeted intron 25 (VWF_E26-1) and the reverse was located across the junction of exon 26-27 (VWFR26/27-2).

Intron retention of exon 45

To determine whether intron 45 was retained within the mature mRNA of patients UMP05 and UMP06, RT-PCR reactions using 2 allele-specific primers combinations were performed. In the first primer pair, the forward was designed at the junction of exon 44-45 (VWFR44/45-1C) and the reverse primer was designed inside intron 45 (VWF_E45-2). In the second primer pair, the

forward primer targeted intron 45 (VWF_E46-1) and the reverse was located across the junction of exon 47-48 (VWFR46/47-2).

X-chromosome inactivation

The X-chromosome inactivation profile studies were performed on DNA extracted from peripheral blood. The study was done using HpaII digestion. Genomic DNA samples were digested with the methylation-sensitive enzyme HpaII (Invitrogen) and subjected to PCR amplification of the highly polymorphic CAG repeat sequence in the first exon of the human androgen receptor gene (HUMARA), with specific fluorescent primers. The primer sequences are shown in Online Supplementary Table S1. The PCR products, both before and after HpaII digestion, were electrophoresed on an automated DNA sequencer (ABI 3130XL; Applied Biosystems, Warrington, United Kingdom) and analyzed by GeneMapper software (Applied Biosystems).[2]

SUPPLEMENTARY TABLES

Table S1. Specific primers used for amplification and sequencing by Sanger of *VWF* RNA.

Forward Primer	Location	Reverse Primer	Location	Purpose	Product size
VWRNA1-1 CTACATAACAGCAAGACAGTCC	Exon 1	**VWRNA1-2** GGTGCTCTTCAGAAGCTGG	Exon 7	Full-length VWF amplification	836 bp
VWRNA2-1 GCAGCTCATGCAACATCTCC	Exon 6	**VWRNA2-2** CCTTGTGAAACTGAAGCATGG	Exon 12	Full-length VWF amplification	762 bp
VWRNA3-1 CCATTGTCATTGAGACTGTCCAC	Exon 11	**VWRNA3-2** CCTGACCTGCCGCTCTCT	Exon 16	Full-length VWF amplification	744 bp
VWRNA4-1 GCGTGGCGCGAGCCAGG	Exon 15	**VWRNA4-2** GTAACCCTGGGACCTTTCG	Exon 21	Full-length VWF amplification	796 bp
VWRNA5-1 CATGGCCCACTACCTCACC	Exon 20	**VWRNA5-2** GGTGGTGACCTGGAGGAC	Exon 25	Full-length VWF amplification	752 bp
VWRNA6-1 TCCTGTGAGTCCATTGGGG	Exon 25	**VWRNA6-2** GGAAGCGACCGTCAGAGC	Exon 28	Full-length VWF amplification	750 bp
VWRNA7-1 CTTTGTGGTGGACATGATGG	Exon 28	**VWRNA7-2** CCTCTCTGACCACAGCTTC	Exon 28	Full-length VWF amplification	876 bp
VWRNA8-1 CCTGCAGCGGGTGCGAG	Exon 28	**VWRNA8-2** CCTGGTCACGGACGTCTC	Exon 31	Full-length VWF amplification	728 bp
VWRNA9-1 AAATCGGGGATGCCTTGGG	Exon 30	**VWRNA9-2** GGAGGTGACGGTGAATGGG	Exon 35	Full-length VWF amplification	772 bp
VWRNA10-1 GGAGGTGATTCTCCATAATGG	Exon 35	**VWRNA10-2** CTGCGTTGTGCAGTTGACC	Exon 39	Full-length VWF amplification	930 bp
VWRNA11-1 GTGAGCATGGCTGTCCCC	Exon 38	**VWRNA11-2** GGATGCCGTGATGGGCCT	Exon 43	Full-length VWF amplification	753 bp
VWRNA12-1 GTGTGTCCACCGAAGCACC	Exon 43	**VWRNA12-2** CCCTTTGATGAACACAAGTGT	Exon 49	Full-length VWF amplification	813 bp
VWRNA13-1 CGTGATGAGACGCTCCAGG	Exon 49	**VWRNA13-2** CCTCTGCATGTTCTGCTCTT	Exon 52	Full-length VWF amplification	546 bp
VWRNA3-1 CCATTGTCATTGAGACTGTCCA	Exon 11	**VWRNA_E15-2** CGTCGTAGCGGCAGTTCC	Exon 15	primers to avoid alternative splicing in leukocytes	564 bp
VWFR4-1 GCCTCTCCGTGTATCTTGG	Exon 4	**VWRNA1-2B** GGTGCTCTTCAGAAGCTGG	Exon 7	Exon 5-6	446 bp
VWF22-1 GGTCCTGAAGCAGACATACC	Exon 22	**VWF26-2A** CTCCCGGAGATTCCTCTCC	Exon 26	Exon 25 complete	466 bp
VWFR24/25-1 GACTGCAACAAGCTGGTGG	Junction exon 24-25	**VWF_E25-2** ATCCAGTCCCTACTAACACT	Intron 25	Intron 25 retention	301 bp
VWF_E26-1 ACATAGCAAGACCCCATCTG	Intron 25	**VWFR26/27-2** CATCCAGGATTTTCCCTGG	Junction exon 26-27	Intron 25 retention	401 bp
VWFR43/44-1 GGACAGCTGTCGGTCGGG	Junction exon 43-44	**VWF_E45-2** CAGGAGCCAAAAGTGGAAAG	Intron 45	Intron 45 retention	381 bp
VWF_E46-1 CGACCGATACAGGAGGGAG	Intron 45	**VWF47/48-2** ATTTTCTTCCTTGTAACCCAGG	Junction exon 47-48	Intron 45 retention	295 bp
VWFR41/42-1 CCCAACTTCACCTGCGCC	Junction exon 41-42	**VWFR44/45-2** TGGGAGCCGACACTCTTCC	Junction exon 44-45	Exon 43 complete	494 bp

At the top of the table, primers previously designed by Corrales et al.[1] At the bottom of the table, primers newly designed to analyze potential splice site mutations included in this study.

Table S2. Additional laboratory and clinical data of VWD patients

Patients code	VWD	Hospital	Registry code	Age	Sex	ABO	Multimeric pattern	BS	Family history
UMP01	3 carrier	CHUC	P64's mother	37	F	A	ND	0	Yes
UMP02	1H	CHUAC	C01P034F17	32	M	O+	NORMAL	2	Yes
UMP03	1	CHUAC	C01P080	40	M	O+	ND	1	NA
UMP04	1	HUVH	-	38	F	O+	ND	C, D	No
UMP05	1	CHUAC	C01P006F03	39	F	O+	NORMAL	11	No
UMP06	2A/2M	CHUAC	C01P024F12	33	F	O	SMEAR	10	No
UMP07	1	HUVH	-	47	M	O+	ND	B	Yes
UMP08	1	HUVH	-	42	M	O+	ND	B	Yes
UMP09	1	HUVH	-	20	F	ND	ND	B, C	Yes
UMP10	3 carrier	HUVH	-	45	F	O	ND	A, C, E	Yes
UMP11	1	CHUAC	C01P051F11	36	M	O+	NORMAL	7	Yes
UMP12	3	HUVH	C03P021F25	15	M	A	ND	10	Yes
UMP13	2A	HUVH	C03P048F231	15	F	O+	↓HMWM	0	Yes
UMP14	1	CHUC	P12	10	M	ND	ND	16	Yes
UMP15	3	HUVH	-	25	F	O	ND	B, E	Yes

The subtype 1H (historical) refers to patients previously diagnosed as type 1 VWD that, at the time of enrollment at the PCM-EVW-ES project, show a slight decrease or even a normal VWF plasma levels. Bleeding scores calculated with the International Society on Thrombosis and Haemostasis bleeding assessment tools (ISTH-BAT) were used to assess bleeding in patients included in the Spanish (PCM-EVW-ES) and Portuguese registries. In the remaining patients, symptoms were assigned letters: A, easy bruising; B, epistaxis; C, prolonged bleeding from wounds; D, menorrhagia; and E, postoperative bleeding. BS indicates bleeding score; CHUAC, Complexo Hospitalario Universitario A Coruña, Spain; CHUC, Centro Hospitalar e Universitário de Coimbra, Portugal; F, female; HUVH, Hospital Universitari Vall d'Hebron, Barcelona, Spain; HMWM, high molecular weight multimers; M, male; NA; not available; ND, not determined and; VWD, von Willebrand disease.

Table S3. Summary of the *in silico* analysis of PSSM

Mutation Type	NT change	AA change	Exon	Intron	Affected splice site	Adjacent nucleotides	NetGene2		Neural Network Splice		MaxEnt		HSF		Score	HSF Interpretation
							wt	var	wt	var	wt	var	wt	var		
Intronic	**c.1533+1G>A**	-	-	13	DSS intron 13	GAAGgtaggt>GAAGataggt	1	Native DSS destroyed	0.98	Native DSS destroyed	10.29	2.1	90.71	63.87	4/4	Alteration of the WT donor site, most probably affecting splicing.
	c.3379+1G>A	-	-	25	DSS intron 25	TGCC**g**tgag>TGCC**a**tgag	0.95	Native DSS destroyed	0.90	Native DSS destroyed	8.56	0.38	81.32	54.48	4/4	Alteration of the WT donor site, most probably affecting splicing.
	c.5664+2T>C	-	-	33	DSS intron 33	AGGg**t**aagt>AGGg**c**aagt	0.83	Native DSS destroyed	1	Native DSS destroyed	10.45	2.7	94.19	67.35	4/4	Alteration of the WT donor site, most probably affecting splicing.
	c.7081+6G>T	-	-	41	DSS intron 41	taagg**c**ctc>taag**t**cctc	1	1	1	1	9.63	11.37	91.56	93.49	0/4	No significant splicing motif alteration detected.
	c.7082-2A>G	-	-	41	ASS intron 41	cctc**a**gCCT>cctc**g**gCCT	NP	Native ASS destroyed	0.69	Native ASS destroyed	6.68	-1.27	88.28	59.34	4/4	Alteration of the WT donor site, most probably affecting splicing.
	c.7730-4C>G	-	-	45	ASS intron 45	tcccc**c**agA>tccc**g**cagA	1	1	0.99	0.99	11.71	13.00	91.73	91.11	0/4	Activation of an intronic cryptic acceptor site. Potential alteration of splicing.
	c.7730-56C>T	-	-	45	ASS intron 45	tcgc**c**ggtc>tcgc**t**ggtc	1	1	0.99	0.99	6.54	7.02	87.43	88.69	0/4	Creation of an intronic ESE site. Probably no impact on splicing.
	c.8155+3G>C	-	-	50	DSS intron 50	ATgt**g**agtg>ATgt**c**agtg	0.54	Native DSS destroyed	0.94	Native DSS destroyed	7.83	2.02	85.46	78.65	3/4	Alteration of the WT donor site, most probably affecting splicing.
	c.8254-5T>G	-	-	51	ASS intron 51	ttctt**g**cag>ttct**g**gcag	1	0.96	0.99	0.94	12.47	10.09	92.5	88.8	0/4	No significant splicing motif alteration detected.
Synonymous	**c.546G>A**	p.Ser182	6	-	New ASS	CCTC**G**GACC>CCTC**A**GACC	0.71	0.33	0.90	0.88	NP	NP	57.31	86.26	2/4	Activation of an exonic cryptic acceptor site, with presence of one or more cryptic branch point(s). Potential alteration of splicing.
	c.3291C>T	p.Cys1097	25	-	New DSS	ACTG**C**GCCT>ACTG**T**GCCT	0.95	0.95	0.90	0.90	NP	NP	NP	NP	0/4	Creation of an exonic ESS site. Potential alteration of splicing.
	c.3426T>C	p.Cys1142	26	-	Alteration GT	AGTG**T**GAGT>AGTG**C**GAGT	0.67	0.67	1	1	6.96	-0.79	82.16	55.33	2/4	Alteration of an exonic ESE site. Potential alteration of splicing.
	c.4866C>T	p.Asp1622	28	-	Not affected	GAGA**C**ATCC>GAGA**T**ATCC	0.96	0.96	0.97	0.97	NP	NP	75.04	75.88	0/4	Alteration of an exonic ESE site.Potential alteration of splicing.
	c.7437G>A	p.Ser2479	43	-	DSS intron 43	GGTC**G**gtgag>GGTC**A**gtgag	1	0.95	0.98	0.63	3.52	11.11	92.11	81.53	3/4	Alteration of the WT donor site, most probably affecting splicing.
Missense	**c.449T>C**	p.Leu150Pro	5	-	Not affected	CTGCTGTCA>CTGC**C**GTCA	1	1	0.98	0.98	NP	NP	85.99	86,64	0/4	No significant splicing motif alteration detected.
	c.1109G>A	p.Cys370Tyr	9	-	DSS intron 9	ACCT**G**gtaa>ACCT**A**gtaa	NP	NP	0.31	Native DSS destroyed	5.67	-2.74	81.00	70.42	3/4	Alteration of the WT donor site, most probably affecting splicing.
DelIns	**c.3485_3486delinsTG**	p.Pro1162Leu	26	-	Not affected	GAGC**CA**CTGG>GAGC**TG**CTGG	0.67	0.83	1	1	6.69	7.09	71.11	71.95	1/4	Creation of an exonic ESS site. Potential alteration of splicing.
	c.3223-7_3236dup	p.Pro1079_Tyr1080insLeuGlnValAspProGluPro	25	-	New ASS	cagG-------- TGGA>cagG**tttgcagG**TGGA	0.95	0,95 + New ASS (0.81)	0.73	0,59 + New ASS	3.09	4.86	15.42	84.46	4/4	Activation of an exonic cryptic acceptor site. Potential alteration of splicing.
	c.6699_6702dup	p.Cys2235Argfs*8	38	-	Not affected	AGGC----TGTT>AGGC**AGGC**TGTT	0,93	0,93	0,97	0,98	4,3	4,27	82,59	83,23	0/4	Alteration of an exonic ESE site.Potential alteration of splicing.

The *in silico* global score is based on the number of *in silico* algorithms that predicted a splicing effect. *In silico* analysis of mutations included in the PCM-ES-EVW registry was performed by ALAMUT. *In silico* analysis of the mutations not included in the PCM-EVW-ES registry was performed individually under the established parameters of each algorithm, as well as NetGene2 (unavailable in ALAMUT). For HSF and MaxEnt, the sequence analyzed is 50 nucleotides in length. For NetGene2 and Neural Network Splice, the sequence analyzed is the exon length plus 100 intronic nucleotides at the exon ends. In the HSF software, if the WT score is above the threshold (65) and the score variation (between WT and Mutant) is below -10%, the mutation is considered to break the splice site. In the other case, if the WT score is below the threshold and the score variation is above +10% the mutation is considered to create a new splice site. In the MaxEnt software, if the WT score is above the threshold (3) and the score variation (between WT and Mutant) is below -30% the mutation is considered to break the splice site. In the other case, if the WT score is below the threshold and the score variation is above +30% the mutation is considered to create a new splice site. AA indicates amino acid; ASS, acceptor splice site; DSS, donor splice site; ESS, exonic splicing silencer; ESE, exonic splicing enhancer; HSF, Human Splicing Finder; MaxEnt, Maximum Entropy; NP not predicted; NT, nucleotide and; PSSM, potential splice site mutation.

Table S4. Analysis of mRNA aberrant transcript expression by next-generation sequencing

Patient code	Mutation	mRNA probes	Leukocytes		Platelets	
			% Patient Reads	% Control Reads	% Patient Reads	% Control Reads
UMP01	c.1533+1G>A	WT	**28 %**	100 %	**100 %**	100 %
		Exon 13 skipping (r.1433_1533del)	**19 %**	0 %	**0 %**	0 %
		Exon 13+14 skipping (r.1433_1729del)	**45 %**	0 %	**0 %**	0 %
		Exon 14 skipping (r.1534_1729del)	**8 %**	0 %	**0 %**	0 %
UMP02	c.3379+1G>A	WT	**56 %**	100 %	**99 %**	100 %
		Exon 25 Skipping (r.3223_3379del)	**43 %**	0 %	**1 %**	0 %
		Activation of a cryptic site at +126 nt in exon 25 (r.3349_3379del) ‡	**1.4 %**	0 %	**0 %**	0 %
UMP03	c.5664+2T>C	WT	**61 %**	100 %	**97 %**	100 %
		Exon 33 skipping (r.5621_5664del)	**37 %**	0 %	**0 %**	0 %
		Exon 33+34 skipping (r.5621_5842del) ‡	**2 %**	0 %	**3 %**	0 %
UMP08	c.546G>A	WT	**98 %**	99 %	**NA**	-
		Exon 6 skipping (r.533_657del) ‡	**2.3 %**	1 %	**NA**	-
UMP10	p.Cys370Tyr	WT	**79 %**	100 %	**100 %**	100 %
		Exon 9 skipping (r.998_1109del)	**13 %**	0 %	**0 %**	0 %
		Exon 8+9 skipping (r.875_1109del)	**7 %**	0 %	**0 %**	0 %
UMP11	p.Pro1079_Tyr1080insLeuGlnValAspProGluPro	WT	**86 %**	98 %	**87 %**	100 %
		p.Pro1079_Tyr1080insLeuGlnValAspProGluPro	**14 %**	2 %	**13 %**	0 %
UMP12	c.7082-2A>G	WT	**76 %**	100 %	**100 %**	100 %
		Activation of a cryptic site at +7 nt in exon 42 (r.7082_7088del)	**24 %**	0 %	**0 %**	0 %
UMP14	c.7437G>A	WT	**58 %**	99 %	**NA**	-
		Exon 43 skipping (r.7288_7437)	**12 %**	0 %	**NA**	-
		Activation of a cryptic splice site at +146 nt in exon 43 (r.7434_7437del)	**30 %**	1 %	**NA**	-
UMP15	c.546G>A*	Exon 6 skipping (r.533_657del)	**3 %**	100 %	**11 %**	100 %
	c.7082-2A>G*	Activation of a cryptic site at +7 nt in exon 42 (r.7082_7088del)	**6 %**	0 %	**0 %**	0 %
	c.8155+3G>C †	Exon 50 skipping (r.8116_8155del)	**95 %**	0 %	**95 %**	0 %

The values correspond to the percentage of the total reads aligned to the aberrant mRNA probes, setting as 100% the sum of reads aligned against the aberrant mRNA probes and reads aligned against the wild-type mRNA probes. NA indicates not available; and WT, wild-type. *Mutations in cis. †Mutations in trans. ‡Transcripts with low % of reads in leukocytes.

SUPPLEMENTARY FIGURES

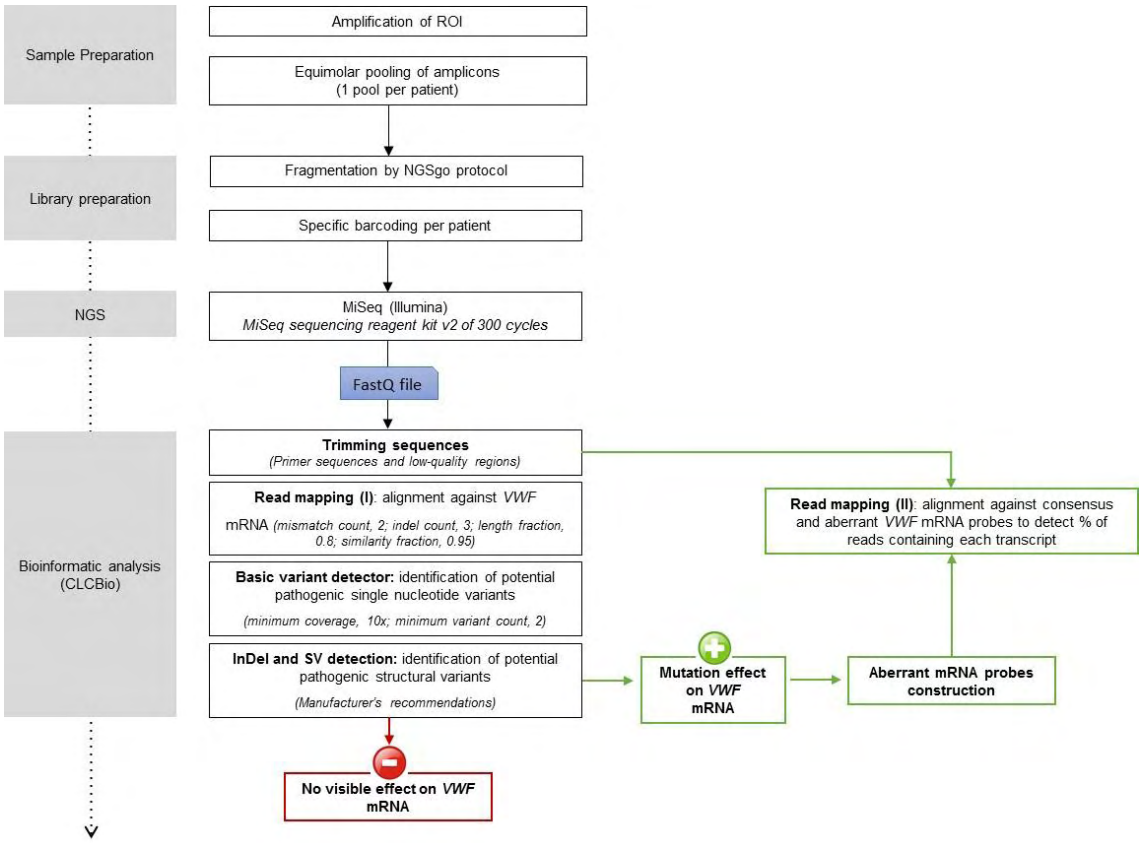


Figure S1. Flow chart used to analyze the effect of the mutation on mRNA using next-generation sequencing. ROI indicates regions of interest; NGS, next generation sequencing and; SV, structural variants.

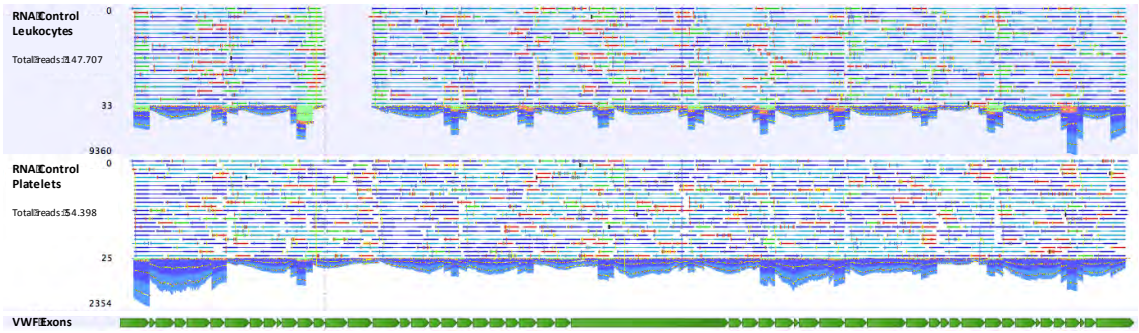


Figure S2. Schematic representation of alignment of NGS reads (from complete amplification of *VWF* mRNA from control platelets and leukocytes) to the *VWF* mRNA reference sequence. Because leukocytes show a predominant alternative splicing resulting in deletion of exon 14 (its start is shown with a discontinuous line) and exon 15, no reads were identified and mapped in this region. Except for exons 14 and 15 in leukocytes, all regions are presented with a minimum coverage of 10X in leukocytes and platelets.

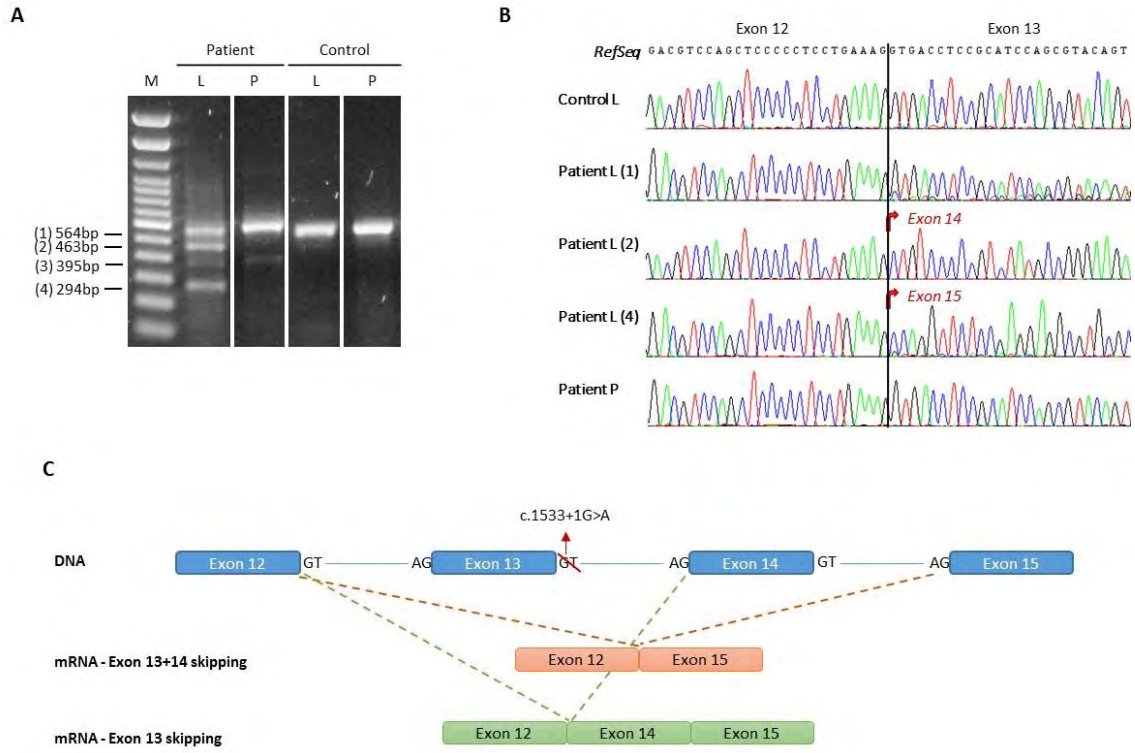


Figure S3. Analysis of the c.1533+1G>A mutation in patient UMP01. **A)** RT-PCR products amplified with primers located in exons 11 and 15 in RNA from leukocytes (L) and platelets (P) and separated on 1% agarose gel. Grouping of images from different parts of the same gel. **B)** Traditional Sanger sequencing of PCR product from patient L (1) showed the WT mRNA sequence, L (2) demonstrated exon 13 skipping and L (4) demonstrated exon 13 and 14 skipping. The L (3) band could not be purified in the agarose gel due to low concentration, thus it was not analyzed by Sanger. Platelets showed the WT mRNA sequence, indicating that the mutated allele had been degraded by NMD. **C)** Schematic representation of the mutation in genomic DNA and its effect on the *VWF* mRNA sequence. M indicates a 100-bp DNA ladder.

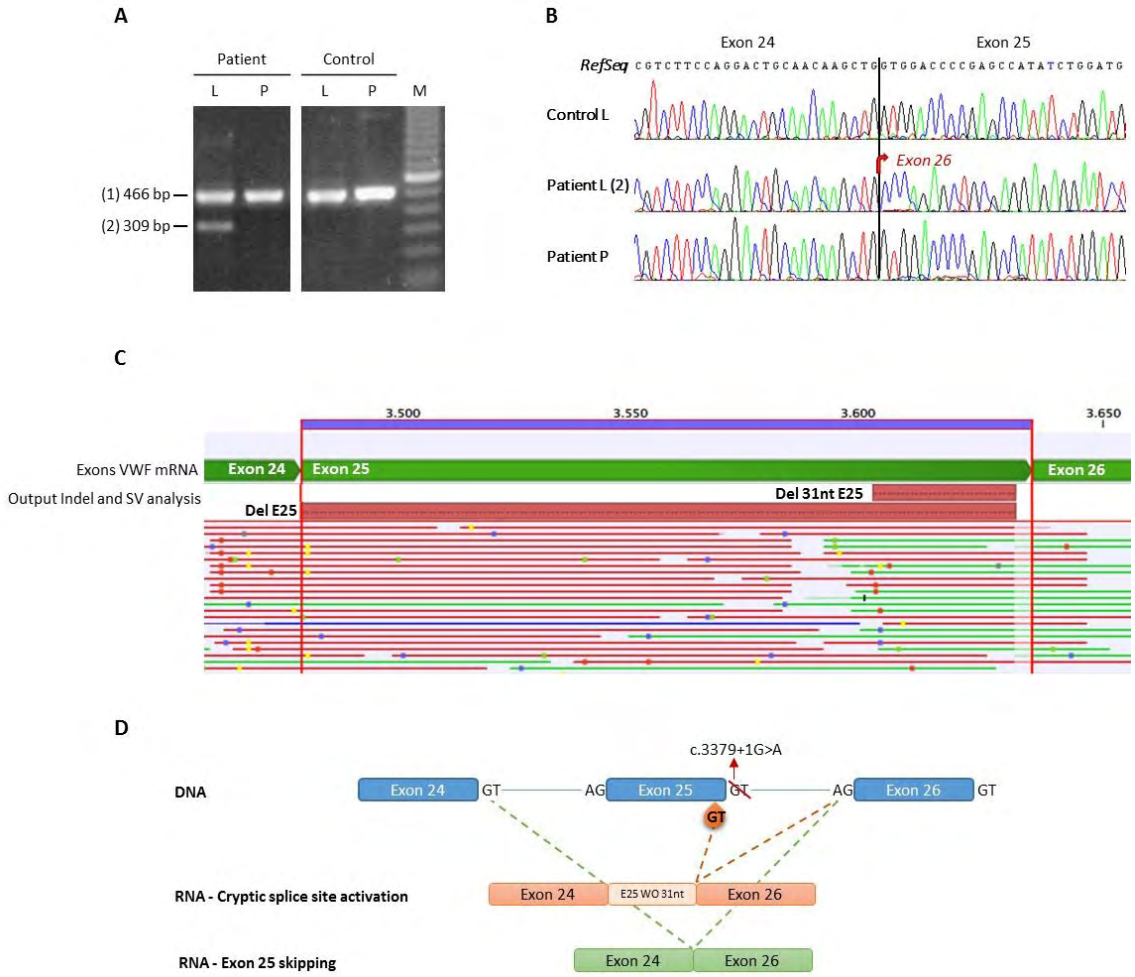


Figure S4. Analysis of the c.3379+1G>A mutation in patient UMP02. **A)** RT-PCR products amplified with primers located in exons 22 and 26 in RNA from leukocytes (L) and platelets (P) and separated on 1% agarose gel. Grouping of images from different parts of the same gel. **B)** Traditional Sanger sequencing of PCR product from patient L (2) showed exon 25 skipping. No changes were seen in the mRNA sequence in platelets, indicating that the mutated allele had been degraded by NMD. **C)** NGS of PCR products obtained from patient leukocytes identified two splicing variants: exon 25 skipping and activation of a cryptic splice site—DSS 31 nt upstream to WT-DSS—in exon 25. This last aberrant transcript was identified in a really low of transcripts. **D)** Schematic representation of the mutation in genomic DNA and its effect on the *VWF* mRNA sequence. M indicates 100-bp DNA ladder and; WO, without.

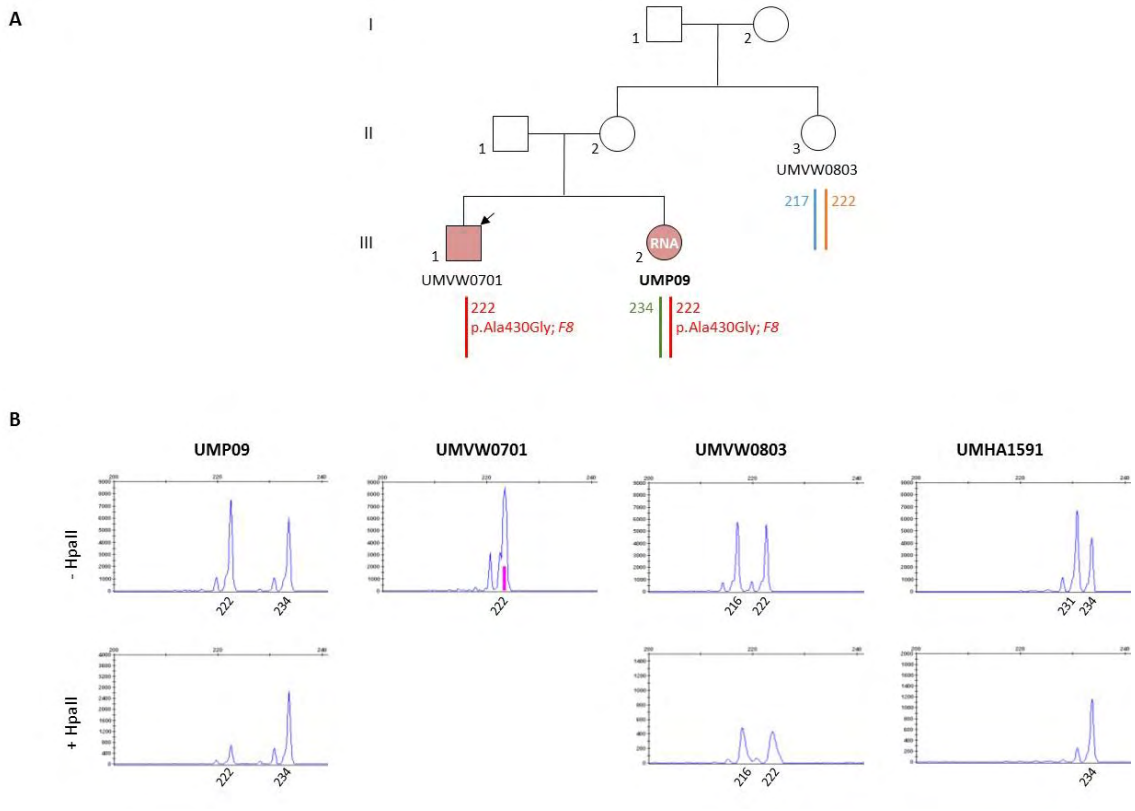


Figure S5. A) Pedigree of patient UMP09. The siblings, UMP09 and UMVW0701, showed the p.Ala430Gly mutation in the F8 gene in the X chromosome with the 222 peak. The arrow shows the index case. **B)** In patient UMP09, included in this study, her maternally-derived (UMVW0803) allele (222) was completely digested by the methylation-sensitive restriction enzyme Hpa II and, therefore, was not amplified by PCR. The remaining peak (234) is the paternally-derived allele representing the inactive, methylated X chromosome that resisted cleavage by Hpa II and was successfully amplified. Thus, this female patient shows complete skewing of X inactivation corresponding to the WT F8 gene. The UMHA1591 sample is a positive control for skewing of X inactivation.

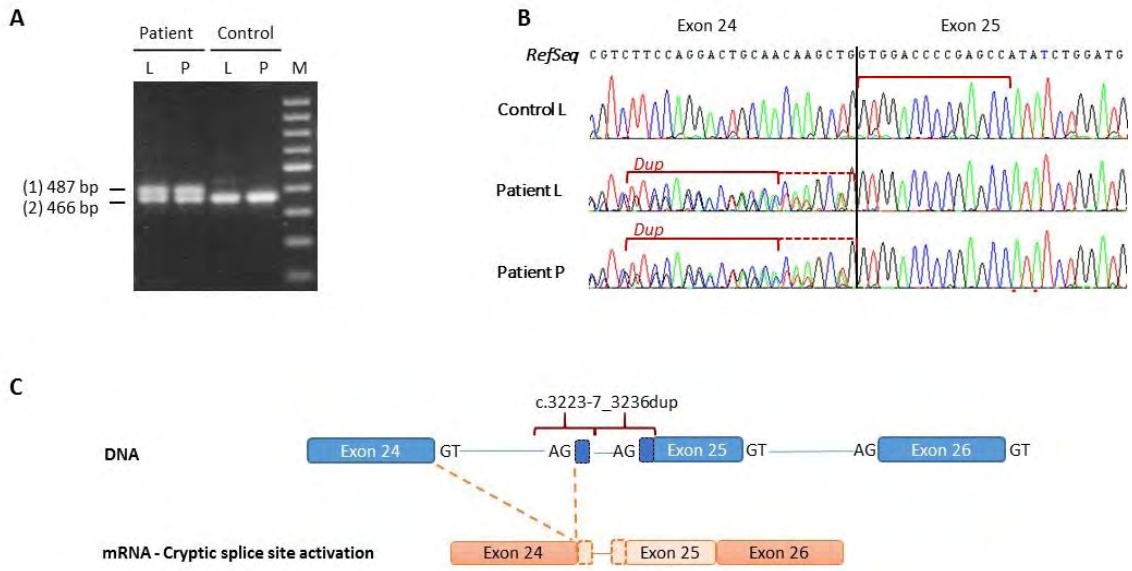


Figure S6. Analysis of the c.3223-7_3236dup mutation in patient UMP11. **A)** RT-PCR products amplified with primers located in exons 22 and 26 in RNA from leukocytes (L) and platelets (P) and separated on 1% agarose gel. **B)** Traditional Sanger sequencing of PCR product from patient L and P showed insertion of 21 nucleotides. Same results were observed by NGS (data not shown). **C)** Schematic representation of the mutation in genomic DNA and its effect on the VWF mRNA sequence. M, 100-bp DNA ladder.

REFERENCES

1. Corrales I, Ramirez L, Altisent C, Parra R, Vidal F. The study of the effect of splicing mutations in von Willebrand factor using RNA isolated from patients' platelets and leukocytes. J Thromb Haemost. 2011;9(4):679-688.

2. Favier R, Lavergne JM, Costa JM, et al. Unbalanced X-chromosome inactivation with a novel FVIII gene mutation resulting in severe hemophilia A in a female. Blood. 2000;96(13):4373-4375.