



HAL
open science

The challenge of genetically unresolved haemophilia A patients: Interest of the combination of whole F8 gene sequencing and functional assays

Fanny Lassalle, Yohann Jourdy, Loubna Jouan, Laura Swystun, Julie Gauthier, Christophe Zawadzki, Jenny Goudemand, Sophie Susen, Georges-etienne Rivard, David Lillicrap

► To cite this version:

Fanny Lassalle, Yohann Jourdy, Loubna Jouan, Laura Swystun, Julie Gauthier, et al.. The challenge of genetically unresolved haemophilia A patients: Interest of the combination of whole F8 gene sequencing and functional assays. *Haemophilia*, 2020, 26 (6), pp.1056-1063. 10.1111/hae.14179 . hal-04372004

HAL Id: hal-04372004

<https://hal.science/hal-04372004v1>

Submitted on 4 Jan 2024

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.

The challenge of genetically unresolved haemophilia A patients: Interest of the combination of whole *F8* gene sequencing and functional assays

Fanny Lassalle, Yohann Jourdy, Loubna Jouan, Laura Swystun, Julie Gauthier, Christophe Zawadzki, Jenny Goudemand, Sophie Susen, Georges-Etienne Rivard, David Lillicrap

Affiliations

- ¹CHU Lille, Institut d'Hématologie - Transfusion, Pôle de Biologie Pathologie Génétique, Lille, France.
- ²Univ Lille, Inserm, U1011 - EGID, Institut Pasteur de Lille, Lille, France.
- ³Service d'hématologie biologique, Centre de Biologie et Pathologie Est, Hospices Civils de Lyon, France.
- ⁴EA 4609 Hémostase et Cancer, Université Claude Bernard Lyon 1, Lyon, France.
- ⁵Integrated Centre for Pediatric Clinical Genomics, CHU Sainte Justine, Montreal, Canada.
- ⁶Department of Pathology and Molecular Medicine, Queen's University, Kingston, Canada.
- ⁷Molecular Diagnostic Laboratory and Division of Medical Genetics, Department of Pediatrics, CHU Sainte Justine, Montreal, Canada.
- ⁸Service d'hématologie-oncologie, CHU Sainte Justine, Montreal, Canada.

Abstract

Background

The causative variant remains unidentified in 2%–5% of haemophilia A (HA) patients despite an exhaustive sequencing of the full *F8* coding sequence, splice consensus sequences, 5'/3' untranslated regions and copy number variant (CNV) analysis. Next-generation sequencing (NGS) has provided significant improvements for a complete *F8* analysis.

Aim

The aim of this study was to identify and characterize pathogenic non-coding variants in *F8* of 15 French and Canadian HA patients genetically unresolved, through the use of NGS, mRNA sequencing and functional confirmation of aberrant splicing.

Methods

We sequenced the entire *F8* gene using an NGS capture method. We analysed *F8* mRNA in order to detect aberrant transcripts. The pathogenic effect of candidate intronic variants was further confirmed using a minigene assay.

Results

After bioinformatic analysis, 11 deep intronic variants were identified in 13 patients (8 new variants and 3 previously reported). Three variants were confirmed to be likely pathogenic with the presence of an aberrant transcript during mRNA analysis and minigene assay. We also found a small intronic deletion in 6 patients, recently described as causing mild HA.

Conclusion

With this comprehensive work combining NGS and functional assays, we report new deep intronic variants that cause HA through splicing alteration mechanism. Functional analyses are critical to confirm the pathogenic effect of these variants and will be invaluable in the future to study the large number of variants of uncertain significance that may affect splicing that will be found in the human genome.

1 INTRODUCTION

Haemophilia A (HA; MIN: 306700), the most common severe inherited bleeding disorder, is caused by a deficiency and/or dysfunction of coagulation factor VIII (FVIII). According to the recommendations of the International Society on Thrombosis and Haemostasis (ISTH), the clinical severity of HA depends on the residual factor VIII activity (FVIII:C) and is classified into 3 levels: severe (FVIII:C < 1 IU/dl), moderate (1 IU/dl ≤ FVIII:C ≤ 5 IU/dl) or mild (5 IU/dl < FVIII:C < 40 IU/dl).¹ The FVIII is encoded by *F8* (MIN: 300841) which spans 186 kb of the long arm of the X chromosome (Xq28) and includes 26 exons.² A wide variety of pathogenic variants have been described in all HA phenotypes in the factor VIII variant database (<http://f8-db.eahad.org/>; accessed May 2020) that documents > 3000 different *F8* variants. Identification of the molecular defect responsible for HA is important for both enhancing haemophilia patient care and identifying female carriers with the possibility of prenatal diagnosis in cases of a male foetus and severe HA.

Despite an exhaustive screening of all *F8* exons, consensus splice sequences and 5'/3' untranslated regions (UTR), recurrent rearrangements such as inversions affecting introns 1 and 22, and copy number variant analysis, a causative variant is still missing in 2%–5% of HA patients.^{3–6} In these patients, a deep intronic variant or a genomic rearrangement disrupting the *F8* gene could be causal. Several *F8* deep intronic variants have been previously reported to create de novo acceptor splice sites (ASS) or donor splice sites (DSS), increase the strength of a pre-existing cryptic ASS or DSS, promote *Alu* exonization or reduce *F8* RNA expression.^{7–11} Consequently, several groups have developed whole *F8* gene sequencing approaches using next-generation sequencing (NGS) in order to identify deep intronic variants and to resolve the mutant genotype in patients in whom conventional genetic approaches failed.^{12, 13} However, the influence of deep intronic variants is challenging to evaluate and functional studies are often necessary to generate conclusive evidence of variant pathogenicity.

This study aimed to identify *F8* deep intronic variants in genetically unresolved French and Canadian HA patients using *F8* whole gene sequencing and to characterize these variants using functional analyses.

2 MATERIAL AND METHODS

2.1 Patient

Fifteen male HA patients (9 French and 6 Canadian) were enrolled in this study. Two patients had severe HA, 4 had moderate HA (FVIII:C 1%–3%), and 9 patients had mild HA (FVIII:C 8%–38%; Table 1). No *F8* candidate pathogenic variant was found after direct sequencing of *F8* coding region, including the flanking intronic sequences and 5'/3' UTR. Multiplex ligation-dependent probe amplification assay (MRC Holland) was performed, and all patients had a normal allele copy number for *F8* exons. Both intron 1 and intron 22 inversion screening were negative in the 2 severe HA patients. von Willebrand disease (including type 2N) and combined FV and FVIII deficiency have been previously excluded. Written informed

consent for genetic analyses was signed by all patients or parents if the patient was a minor. Clinical investigations have been conducted according to the principles expressed in the Declaration of Helsinki.

TABLE 1. Intronic candidate variants detected with NGS for the 15 patients

ID	HA severity	Family history	FVIII:C	Candidate variant	Intron	MAF GnomAD	In silico prediction on splice sites	mRNA on sequencing	Minigene assay	ACMG classification (consequence of the variant)
P1	Severe	No	<1%	c.265+1116C>G	2	Not described	de novo DSS (MaxEnt 4.9)	No transcript	No impact	Uncertain significance
				c.1010-443A > G	7	Not described	de novo DSS (MaxEnt 5.2; HSF 82.3)			Uncertain significance
P2	Severe	Unknown	<1%	c.5219+7731G>T	14	Not described	No effect	No transcript	No impact	Uncertain significance
P3	Moderate	No	1%-2%	c.787+1967C>T	6	0.04%	No effect	No transcript	No impact	Uncertain significance
P4	Moderate	No	1%	c.5999-27A > G	18	Not described	No effect	Aberrant	Aberrant	Likely pathogenic (exon 19)

ID	HA severity	Family history	FVIII:C	Candidate variant	Intron	MAF GnomAD	In silico prediction on splice sites	mRNA on sequencing	Minigene assay	ACMG classification (consequence of the variant)
P5	Moderate	Yes	3%	No variant				Normal		skipping)
P6	Moderate	No	3%	No variant				No transcript		
P7	Mild	No	29%	c.1753-750G > A	11	0.01%	No effect	Unavailable	No impact	Uncertain significance
P8	Mild	Yes	8%	c.2113+1153G>C	13	Not described	No effect	Aberrant	Uninterpretable	Likely pathogenic (exonization of intronic sequence)
P9	Mild	Yes	38%	c.2113+461_2113+473del	13	Described by Jourdy <i>et al</i> (Am J Hum Genet, 2018)				Pathogenic
P10	Mild	No	31%	c.2113+461_2113+473del	13	Described by Jourdy <i>et al</i> (Am J Hum Genet, 2018)				Pathogenic
P11	Mild	Yes	30%	c.2113+461_2113+473del	13	Described by Jourdy <i>et al</i> (Am J Hum Genet, 2018)				Pathogenic

ID	HA severity	Family history	FVIII:C	Candidate variant	Intron	MAF GnomAD	In silico prediction on splice sites	mRNA sequencing	Minigene assay	ACMG classification (consequence of the variant)	
P12	Mild	No	28%	c.2113+463_2113+473del	13	Described by Jourdy <i>et al</i> (J Thromb Haemost, 2020)					Pathogenic
P13	Mild	No	35%	c.2113+463_2113+473del	13	Described by Jourdy <i>et al</i> (J Thromb Haemost, 2020)					Pathogenic
P14	Mild	Unknown	30%	c.2113+465_2113+473del	13	Described by Jourdy <i>et al</i> (J Thromb Haemost, 2020)					Pathogenic
P15	Mild	Yes	11%	c.5220-8563A > G	14	Not described	de novo ASS (MaxEnt 8.0; HSF 81.1)	Aberrant	Aberrant	Likely pathogenic (exonization of intronic sequence)	

No candidate variant was found for patients 5 and 6.

- Abbreviations: ACMG, American College of Medical Genetics and Genomics classification; ASS, acceptor splice site; DSS, donor splice site; HA, haemophilia A; MAF, minor allele frequency; P, patient.

2.2 Next-generation sequencing

The whole *F8* (NM_000132.3) gene with the complete 5' and 3' UTR and an additional 2500 bp upstream sequence (ChrX:155,026,925-154,831,631; GRCh38/hg38) was analysed using NGS. The SureSelect^{XT} Target Enrichment System (Agilent) which is a capture-based protocol for library preparation was used, according to the manufacturer's instructions. An iterative approach was used, and 40-kb repetitive regions were excluded to maximize the target coverage, resulting in a total of 1630 probes for a total 156 kb covered region of 195 kb total target region size, corresponding to 80% of *F8*. Patients' blood was collected into 2 EDTA tubes. DNA was extracted using the QIAamp[®] DNA Blood Maxi kit (Qiagen) in France and Gentra[®] Puregene[®] Blood kit (Qiagen) in Canada, both according to the manufacturer's protocol. Three micrograms of genomic DNA were necessary for NGS analysis. After library preparation, NGS was performed on a MiSeq sequencer (Illumina Inc) with 150 bp paired-end reads. All identified *F8* variants were confirmed by Sanger sequencing.

2.3 Bioinformatic analyses

An in-house pipeline was used for bioinformatic analysis. Raw data were aligned to the human reference sequence of the X chromosome (GRCh38/hg38). Sequence processing, alignment and variant calling were conducted according to the Broad Institute Genome Analysis ToolKit (GATK v.3.6) best practices, and variant annotation was done with a modified version of ANNOVAR using publicly available databases and Qiagen Genome Trax. For variant trimming, only variants with a minor allele frequency (MAF) $\leq 1\%$ in the GnomAD database were retained. We also consulted the dbSNP, ClinVar and Intervar databases for clinical interpretation of previously described variants.

For in silico prediction, we used Alamut Visual[®] software version 2.15 (Interactive Biosoftware, Rouen, France) and its algorithms (MaxEnt and HumanSpliceFinder (HSF)) to predict the potential splice effect of the variants.

Variants were categorized according to the American College of Medical Genetics and Genomics classification (ACMG): pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign and benign.¹⁴

2.4 *F8* mRNA analysis

Patients' blood was collected into 2 PAXgene[®] blood RNA tubes. RNA was extracted from leucocytes using the PAXgene[®] Blood RNA kit (Qiagen) following the manufacturer's instructions. We followed the protocol established by El-Maarri et al for the *F8* mRNA analysis.¹⁵ After reverse transcription, the *F8* cDNA was amplified in 8 fragments using 2 rounds of nested PCR. The fragments were Sanger sequenced with the same primers as used for the nested PCR.

2.5 Minigene splicing reporter assay

The minigene assay was performed as previously reported.¹¹ Briefly, the exon of interest and its flanking intronic regions were amplified from patient or control DNA using PrimeSTAR GXL DNA polymerase (Clontech, Saint-Germain-en-Laye, France). Then, the PCR product

was inserted in the *NdeI* site of the pTB minigene¹⁶ using the In-fusion HD PCR cloning kit (Clontech, Mountain View, CA, USA). In the case of c.787+1957C>T, c.5220-8563A>G and c.5219+7731G>T, only the region of intron 6 or intron 14 containing these variants respectively was inserted in the pTB minigene. The correct sequences of wild-type (WT) and mutant plasmids were verified by DNA sequencing. The detailed vector constructs are summarized in Table S1. HeLa cells were transiently transfected with 1 µg of either WT or mutated plasmid using X-tremeGENE9 DNA (Roche Molecular Diagnostics, Meylan, France). Forty-eight hours after transfection, total RNA was extracted, and RT-PCR analysis was performed using specific primers located in the minigene exons surrounding the pTB cloning site. The PCR products were resolved on agarose gel and sequenced to identify splicing events. If more than one transcript was present, an additional cloning step using TOPO TA cloning kit (Life technologies, Carlsbad, CA, USA) was performed before the sequencing, to isolate the different transcripts. All experiments were run in triplicate.

3 RESULTS

3.1 Identification of deep intronic variants with NGS

NGS data analysis yielded a high global output (4.3 GB) with a mean of 2.3 million reads for each patient. On average, 98.6% of the targeted region was covered with > 30 reads. The coverage depth was between 150 and 400x. More than 90% of the reads had a quality value above 30. An average of 50 variants per patient was detected. Variants with MAF > 1% and variants that were misaligned in repetitive regions were excluded from the analysis. Eleven deep intronic variants (2 being recurrent) remained in 13 patients (Table 1). No candidate deep intronic variant was detected in 2 moderate HA patients (patients 5 and 6).

Short deletions in *F8* intron 13, c.2113+461_2113+473del, c.2113+463_2113+473del and c.2113+465_2113+473del which were previously described as pathogenic variants causing HA^{11,17} were found in 6 mild HA (patients 9-14), and consequently, no additional genetic analysis was performed in these patients. Among the 8 remaining variants, 2 were listed in GnomAd with a MAF < 1% (c.787+1967C>T at 0.04% and c.1753-750G > A at 0.01%, Table 1). The 6 other variants were not previously described (c.265+1116C>G, c.1010-443A > G, c.5219+7731G>T, c.5999-27A > G, c.2113+1153G>C, c.5220-8563A > G). The in silico analysis predicted that the variants c.265+1116C>G and c.1010-443A > G (patient 1) create a de novo DSS with a medium strength (MaxEnt = 4.9 and 5.2, respectively) and the variant c.5220-8563A > G (patient 15) creates a strong de novo ASS at c.5220-8562 (MaxEnt = 8, HSF = 81.1). No impact on splicing was predicted for the other variants (Table 1).

3.2 mRNA analysis and functional confirmation of the candidate variants

Except for patient 7 in whom RNA sample was not available, *F8* mRNA analysis was performed in all the patients in which no clearly pathogenic variants was found (patients 1-6, 8 and 15). Aberrant transcripts were observed in three cases (patients 4, 8 and 15).

RNA analysis revealed in patient 4, carrying the variant c.5999-27A > G, a unique transcript in which exon 19 was skipped (p.Gly2000_Lys2039delinsGlu) (Figure 1A). The deleterious impact on exon 19 splicing of c.5999-27A > G was confirmed in vitro. Two transcripts were found with the minigene experiment (363 and 246 bp, Figure 1B). The sequencing of each band showed that the 363-bp transcript corresponded to correct exon 19 splicing and the 246-

bp corresponded to the exon 19 skipping. A third very slight band appeared at 800bp on the mutated minigene, but we did not succeed to isolate it for characterization.

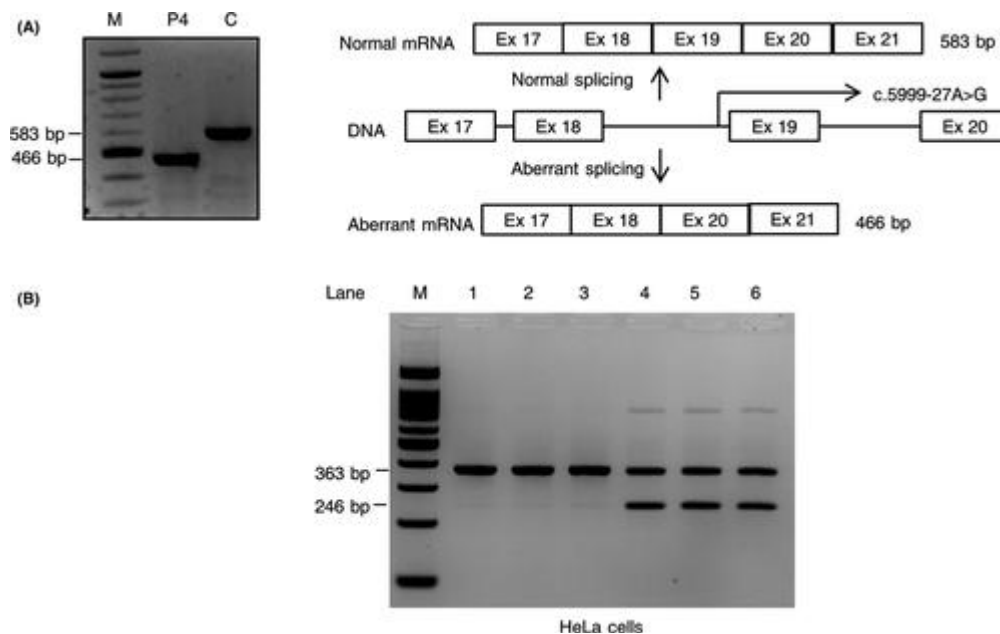


FIGURE 1

Open in figure viewerPowerPoint

Characterization of the deep variant c.5999-27A>G in intron 18. (A) RNA analysis performed in patient P4. A unique aberrant 466-bp transcript was observed when amplifying fragment spanning exons 17-21. Sequencing of this fragment revealed complete exon 19 skipping. (B) Minigene analysis of the c.5999-27A > G variant. HeLa cells were transiently transfected with WT (lanes 1-3) and mutated (lanes 4-6) splicing reporter constructs. Forty-eight hours after transfection, RNA products were analysed by RT-PCR (normal exon 19 splicing 363 bp; exon 19 skipping 246 bp). Abbreviations: C, control; M, marker size; WT, wild type; Ex, exon; bp, base pair

RNA analysis performed in patient 8, carrying the variant c.2113+1153G>C, showed an additional fragment corresponding to a 113-bp sequence from intron 13 (c.2113_2114ins2113+1144_2113+1257) inserted at the exon 13-14 junction (Figure 2). This insertion in the cDNA sequence causes a frameshift during translation and the generation of a premature stop codon (p.Gly705Valfs*55). This 113-bp sequence in intron 13 is naturally flanked by strong AG/GT splice sites (MaxEnt = 4.14 and 10.5, for AG ASS in position c.2113+1144 and GT DSS in position c.2113+1256, respectively). However, this pseudo-exon was spontaneously exonized in the WT minigene; thus, in vitro analysis was uninterpretable for the c.2113+1153G>C variant (Figure S2A).

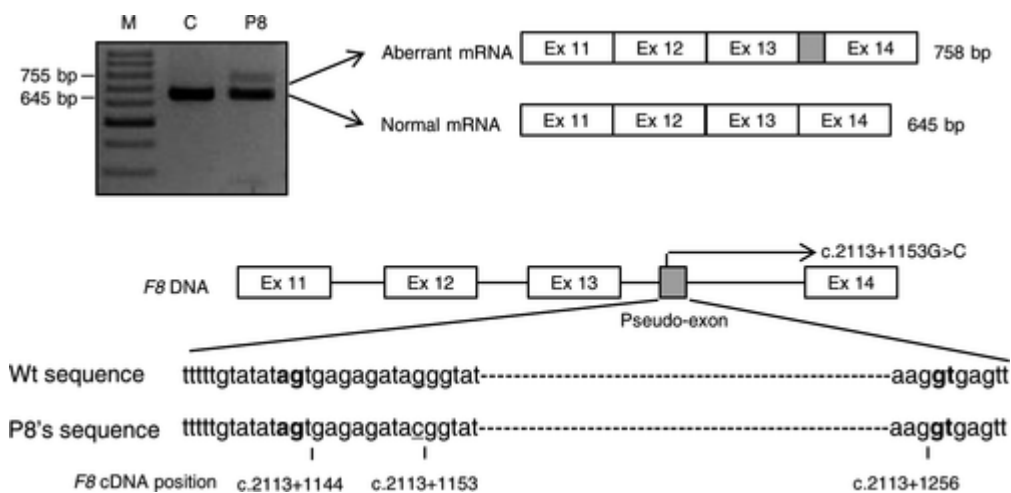


FIGURE 2

Open in figure viewerPowerPoint

Characterization of the deep variant c.2113+1153G>C in intron 13. Two mRNA fragments were observed when amplifying fragments spanning exons 11-14 (WT 645 bp; aberrant 758 bp). Sequencing of the 758-bp fragment found the insertion of intron 13 fragment of 113 bp (c.2113_2114ins2113+1144_2113+1256). Abbreviations: C, control; M, marker size; WT, wild type; Ex, exon; bp, base pair

RNA analysis performed in patient 15 also showed an additional fragment. The variant c.5220-8563A > G created a de novo ASS in intron 14 leading to the insertion of a 124-bp sequence from intron 14 (c.5219_5220ins5220-8562_5220-8439) at the exon 14-15 junction (Figure 3) and causing a frameshift and the generation of a premature stop codon (p.Arg1740Argfs*60). In vitro analysis confirmed the creation of a new ASS by the variant c.5220-8563A>G; however, this result should be interpreted with caution because aberrant splicing was observed for the WT minigene (Figure S2B).

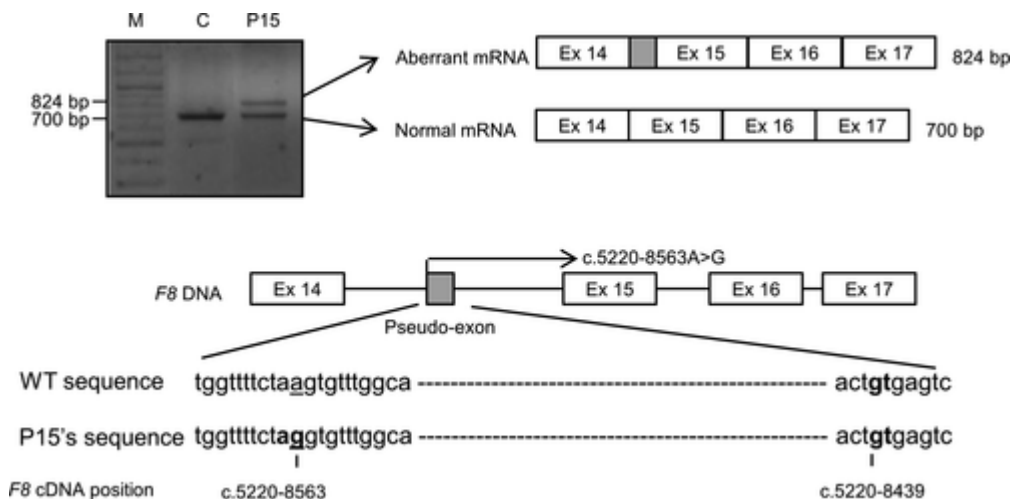


FIGURE 3

Open in figure viewerPowerPoint

Characterization of the deep variant c.5220-8563A>G in intron 14. Two mRNA fragments are observed for patient 15 (P15) when amplifying fragments spanning exons 14-17 (WT 700 bp; mutant 824bp). Sequencing of the 824-bp fragment found the insertion of intron 14 fragment of 124 bp (c.5219_5220ins5220-8562_5220-8439). Abbreviations: C, control; M, marker size; WT, wild type; Ex, exon; bp, base pair

Only a normal FVIII transcript was detected for patient 5 (with no candidate variant). Despite several tests, no *F8* transcript was detected after mRNA analysis in patients 1, 2, 3 and 6.

Minigene analysis was performed for the candidate variants found in patient 1 (c.265+1116C>G and c.1010-443A>G), patient 2 (c.5219+7731G>T), patient 3 (c.787+1967C>T) and patient 7 (c.1753-750G>A). No impact on splicing was found for all these variants (Figure 4).

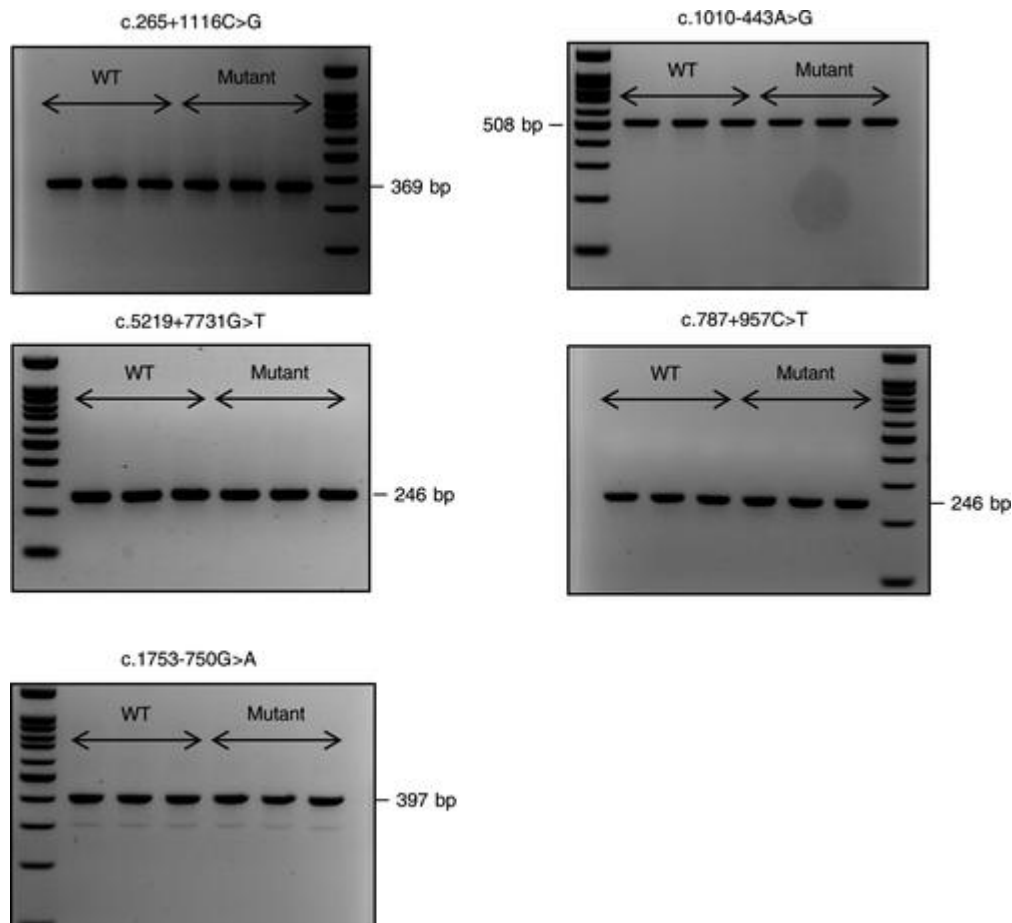


FIGURE 4

Open in figure viewerPowerPoint

Negative minigene results for candidate variants c.265+1116C>G, c.1010-443A>G, c.5219+7731G>T, c.787+1967C>T and c.1753-750G>A. HeLa cells were transiently transfected with WT and mutated splicing reporter constructs. Forty-eight hours after transfection, RNA products were analysed by RT-PCR. In each case, the same unique transcript was observed for both WT and mutant minigene. The 369-bp transcript observed for the c.265+1116C>G corresponded to normal exon 2 splicing, the 508-bp transcript observed for the c.1010-443A>G corresponded to normal exon 8 splicing, and the 397 bp transcript observed for the c.1753-750G>A corresponded to normal exon 12 splicing. In the case of c.5219+7731G>T and c.787+967C>T, the 246-bp transcript corresponded to the absence of intronic sequence retention. WT, wild type; bp, base pair

4 DISCUSSION

F8 deep intronic variants have been previously reported as HA-causing mutation.^{7-9, 11, 18} These variants either create de novo ASS or DSS (c.143+1567A>G, c.5587-93C>T,

and c.1537+325A>G) or increase the strength of a pre-existing ASS or DSS located close to the substitution (c.5999-277G>A, c.5998 + 941G>A, c.2113+461_2113+473del and c.2113+1152delA). However, the determination of the functional impact of variants found using this approach remains challenging. In this study, we used a combination of whole *F8* gene sequencing analysis and functional analyses (mRNA and minigene assay analysis) to identify disease causing deep intronic variants in genetically unresolved HA patients.

Using NGS whole *F8* sequencing, at least one candidate variant was identified in 13/15 HA patients with mild to severe HA. Short deletions in a poly(T)-track in *F8* intron 13, from 9 to 13 bp in size, were identified in two thirds of mild HA patients included in this study. This finding was in line with the recent study published by Jourdy *et al* where such deletions were found in about half of French mild HA patients in whom no genetic events were identified using conventional genetic testing approaches.¹¹ The authors showed that these deletions result in a 122-bp AluY insertion in *F8* transcripts. Recently, it was suggested that the poly(T)-track in *F8* intron 13 (c.2113+449_2113+473) is a deletion hotspot and these deletions should be investigated in all genetically unresolved mild HA whatever their geographic origin.¹⁷

Using our complete analysis protocol combining NGS and functional analyses, we identified and characterized the causative variant in 3 HA patients. In two cases (c.5220-8563A>G and c.2113+1153G>C), intronic variants led to the exonization of intronic sequence introducing premature stop codons in *F8* transcripts. In silico and functional analyses clearly showed that the c.5220-8563A>G creates a new ASS. However, due to the uninterpretable result obtained with the minigene assay for the c.2113+1153G>C, the molecular mechanism by which this variant led to the activation of the 113-bp pseudo-exon was not clear. Splicing prediction tools from Alamut Visual[®] identified 2 new splicing regulatory elements created by the variant c.2113+1153G>C: one is a SF2/ASF splicing factor site (SFS) with a score at 1.97 (threshold 1.956), and one is a SRp55 SFS with a score at 2.95 (threshold 2.676). Interestingly, the same intronic insertion has been described in a mild HA patient carrying the c.2113+1152delA variant.⁸ The fact that 2 mild HA patients are both carrying variants in the region c.2113+1152_2113+1153 and presenting the same exonization combining with the in silico splicing prediction supports the evidence of a splicing regulatory sequence in this region. For the last case, both mRNA and minigene analysis showed that the c.5999-27A>G variant leads to *F8* exon 19 skipping confirming the causality of this variant. As they are not quantitative methods, the mRNA analysis might be not sensitive enough to detect the normal transcript found with the minigene, if it is in very low quantity, explaining the moderate and not severe phenotype of patient 4.

The c.787+1967C>T and c.1753-750G>A variants were reported in the healthy population at very low frequency (MAF < 1%), and no impact on splicing was found in minigene analysis suggesting that these variants were more likely rare benign than pathogenic variants. The 3 remaining variants (c.265+1116C>G, c.1010-443A>G and c.5219+7731G>T) were not reported in the healthy population which is suggestive of their pathogenic effect but no impact was found in minigene analysis. However, it is possible that these variants reduce the *F8* mRNA levels and/or stability as it has been previously reported by Pezeshkpoor *et al*.⁹ Therefore, the causality of these variants remains uncertain.

F8 mRNA was undetectable in 4 patients. Two of these (patients 1 and 2) have a severe HA phenotype. A structural genomic rearrangement is hypothesized. Indeed, no

detectable *F8* mRNA has been previously described in a severe HA patient carrying a *F8* disrupting genomic rearrangement.¹⁹ Whole *F8* sequencing approaches are not able to detect *F8* structural variants, and thus, additional experimental strategies are needed to detect breakpoints like long-range PCR. The two remaining patients (patients 3 and 6) have a moderate HA phenotype. The lack of *F8* mRNA in non-severe HA patient has been previously reported in mild HA patient and remained unexplained.⁹

Finally, no candidate variant was found in 2 moderate HA patients (patients 5 and 6). In these patients, variant in repetitive regions, which were not explored with our NGS methodology, could be suspected. The implication of miRNA as molecular mechanism resulting in HA has also been proposed recently.²⁰

As we know, the in silico prediction efficacy for *F8* deep intronic variants is quite poor,^{8,9} and it is desirable to confirm pathogenic effects predicted by in silico analysis with functional assays evaluating alterations of splicing. Classically, the study of variant impact on gene splicing is performed by the analysis of mRNA. Liver sinusoidal endothelial cells (ECs) and lymphatic ECs are considered as the main sites of the biosynthesis of FVIII.²¹⁻²³ Because of the relative inaccessibility of this tissue in HA patients, *F8* mRNA studies are classically performed in leucocytes as they contain small amounts of ectopic *F8* mRNA transcripts.^{11, 15} These results obtained on 'illegitimate' *F8* mRNA might be tissue-specific and might not reflect the native splicing mechanism that occurs in FVIII-producing cells. However, no alternative transcript of the *F8* gene has yet to be described, and thus, *F8* ectopic mRNA analysis provides useful information on the splicing impact of genomic variants. Alternatively, the minigene approach used in this study provides an alternative testing strategy in the cases where an RNA sample is not available.^{24, 25}

RNA analysis is known to present a certain lack of sensitivity on some points. First, this technique does not use long-range PCR; therefore, an aberrant transcript including a full intron inclusion, with a very long sequence, would not be detected and cannot be excluded for patients with a normal transcript only. Even in the literature, we can find some results contradictions for the same *F8* variants, like for the variant c.222A>G for which Zimmermann *et al* found no impact²⁶ but Martorell *et al* found an exon 2 skipping,²⁷ or for the variant c.5587-93C>T for which Castaman *et al* found a pseudo-exon insertion⁸ but Pezeshkpoor *et al* did not detect any aberrant transcript.⁹ Then, even if the *F8* gene does not seem to be subjected to nonsense-mediated mRNA decay (NMD),^{27, 28} a biological process that degrades nonsense transcripts and could underestimate aberrant isoforms, we cannot exclude either that aberrant transcripts leading to a premature stop codon in the coding sequence could undergo to NMD anyway, preventing their detection or lowering their amount.

Unfortunately, this level of comprehensive combinatorial analysis will be challenging as a routine assay in a clinical setting. The application of functional tests, such as the minigene studies employed here, will not be feasible on large numbers of variants. In our study, we evaluated 8 minigene constructs and only 2 gave a clear result of pathogenicity, after the commitment of a significant amount of work. It seems likely that the only practical solution will be the development of enhanced in silico programs to assess functional significance through the incorporation of strategies employing artificial intelligence-based algorithms.

In conclusion, we elucidated the causing defect for 9 out of 15 genetically unresolved French and Canadian HA patients and we reported new deep intronic variants that can cause HA.

This comprehensive study is one of very few to combine NGS and mRNA sequencing as well as minigene assay to confirm the pathogenic effect of the detected variants. This work emphasized the important role of deep regulatory sequences in causing the disease and above all the difficulty of interpretation of these non-coding variants.

ACKNOWLEDGEMENTS

The study was supported by a fellowship grant for Loubna Jouan from CSL Behring Canada to La Fondation du CHU Sainte-Justine (Dr Rochelle Winikoff) and by a grant for 'Center of Excellence in Pediatric Hemostasis' from Bayer Canada to La Fondation du CHU Sainte-Justine (Dr Georges E Rivard). DL is the recipient of a Canada Research Chair in Molecular Hemostasis.

CONFLICT OF INTEREST

DL receives grant support from Bayer, Biomarin, Bioverativ, CSL Behring and Octapharma. LJ received a fellowship grant from CSL Behring. JG reports personal fees from LFB-Biotechnologies, Novo Nordisk, Shire-Takeda and Roche. SS reports grants from Carmat, CorWave and Stago; grants and personal fees from LFB-Biotechnologies, Roche, Sobi and Siemens Healthineers; and personal fees from Hemabiologics. The remaining authors stated that they had no interests which might be perceived as posing a conflict or bias.

AUTHOR CONTRIBUTIONS

FL, LJ and YJ performed experiments. FL, LJ, LS, YJ, JG, GER and DL designed experiments and interpreted the data. FL and YJ wrote the manuscript, and all authors edited the manuscript.

REFERENCES

- 1White G, Rosendaal F, Aledort L, Lusher J, Rothschild C, Ingerslev J. Definitions in hemophilia. Recommendation of the scientific subcommittee on factor VIII and factor IX of the scientific and standardization committee of the International Society on Thrombosis and Haemostasis. *Thromb Haemost.* 2001; **85**(3): 560.
- 2Bowen DJ. Haemophilia A and haemophilia B: molecular insights. *Mol Patho.* 2002; **55**(2): 127-144.
- 3Oldenburg J, Ananyeva NM, Saenko EL. Molecular basis of haemophilia A. *Haemophilia.* 2004; **10**: 133-139.
- 4Klopp N, Oldenburg J, Uen C, Schneppenheim R, Graw J. 11 hemophilia A patients without mutations in the factor VIII encoding gene. *Thromb Haemost.* 2002; **88**(2): 357-360.
- 5Peyvandi F, Kunicki T, Lillicrap D. Genetic sequence analysis of inherited bleeding diseases. *Blood.* 2013; **122**(20): 3423-3431.
- 6Swystun LL, James PD. Genetic diagnosis in hemophilia and von Willebrand disease. *Blood Rev.* 2017; **31**(1): 47-56.
- 7Inaba H, Koyama T, Shinozawa K, Amano K, Fukutake K. Identification and characterization of an adenine to guanine transition within intron 10 of the factor VIII gene as a causative mutation in a patient with mild haemophilia A. *Haemophilia.* 2013; **19**(1): 100-105.
- 8Castaman G, Giacomelli SH, Mancuso ME, et al. Deep intronic variations may cause mild hemophilia A: intronic mutations in mild hemophilia A. *J Thromb Haemost.* 2011; **9**(8): 1541-1548.

- 9Pezeshkpoor B, Zimmer N, Marquardt N, et al. Deep intronic ‘mutations’ cause hemophilia A: application of next generation sequencing in patients without detectable mutation in F8 cDNA. *J Thromb Haemost.* 2013; **11**(9): 1679-1687.
- 10Bagnall RD, Waseem N, Green PM, Giannelli F. Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. *Blood.* 2002; **99**(1): 168-174.
- 11Jourdy Y, Janin A, Fretigny M, et al. Recurrent F8 intronic deletion found in mild hemophilia A causes alu exonization. *Am J Hum Genet.* 2018; **102**(2): 199-206.
- 12Bach JE, Wolf B, Oldenburg J, Müller CR, Rost S. Identification of deep intronic variants in 15 haemophilia A patients by next generation sequencing of the whole factor VIII gene. *Thromb Haemost.* 2015; **114**(4): 757-767.
- 13Inaba H, Shinozawa K, Amano K, Fukutake K. Identification of deep intronic individual variants in patients with hemophilia A by next-generation sequencing of the whole factor VIII gene. *Res Pract Thromb Haemost.* 2017; **1**(2): 264-274.
- 14Richards 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): 405-423.
- 15El-maarri O, Herbiniaux U, Graw J, et al. Analysis of mRNA in hemophilia A patients with undetectable mutations reveals normal splicing in the factor VIII gene. *J Thromb Haemost.* 2005; **3**(2): 332-339.
- 16Baralle D, Baralle M. Splicing in action: assessing disease causing sequence changes. *J Med Genet.* 2005; **42**(10): 737-748.
- 17Jourdy Y, Frétigny M, Lassalle F, Lillcrap D, Négrier C, Vinciguerra C. The highly prevalent deletions in F8 intron 13 found in French mild hemophilia A patients result from both founder effect and recurrent de novo events. *J Thromb Haemost.* 2020; **18**(5): 1087-1093.
- 18Bagnall RD, Waseem NH, Green PM, Colvin B, Lee C, Giannelli F. Creation of a novel donor splice site in intron 1 of the factor VIII gene leads to activation of a 191 bp cryptic exon in two haemophilia A patients. *Br J Haematol.* 1999; **107**(4): 766-771.
- 19Pezeshkpoor B, Rost S, Oldenburg J, El-Maarri O. Identification of a third rearrangement at Xq28 that causes severe hemophilia A as a result of homologous recombination between inverted repeats. *J Thromb Haemost.* 2012; **10**(8): 1600-1608.
- 20Jankowska KI, McGill J, Pezeshkpoor B, Oldenburg J, Atreya CD, Sauna ZE. Clinical manifestation of hemophilia A in the absence of mutations in the F8 gene that encodes FVIII: role of microRNAs. *Transfusion.* 2020; **60**(2): 401-413.
- 21Shahani T, Covens K, Lavend’homme R, et al. Human liver sinusoidal endothelial cells but not hepatocytes contain factor VIII. *J Thromb Haemost.* 2014; **12**(1): 36-42.
- 22Fahs SA, Hille MT, Shi Q, Weiler H, Montgomery RR. A conditional knockout mouse model reveals endothelial cells as the principal and possibly exclusive source of plasma factor VIII. *Blood.* 2014; **123**(24): 3706-3713.
- 23Pan J, Dinh TT, Rajaraman A, et al. Patterns of expression of factor VIII and von Willebrand factor by endothelial cell subsets in vivo. *Blood.* 2016; **128**(1): 104-109.
- 24Jourdy Y, Fretigny M, Nougier C, Négrier C, Bozon D, Vinciguerra C. Splicing analysis of 26 F8 nucleotide variations using a minigene assay. *Haemophilia.* 2019; **25**(2): 306-315.
- 25Bach JE, Müller CR, Rost S. Mini-gene assays confirm the splicing effect of deep intronic variants in the factor VIII gene. *Thromb Haemost.* 2016; **115**(1): 222-224.
- 26Zimmermann MA, Gehrig A, Oldenburg J, Müller CR, Rost S. Analysis of F8 mRNA in haemophilia A patients with silent mutations or presumptive splice site mutations. *Haemophilia.* 2013; **19**(2): 310-317.
- 27Martorell L, Corrales I, Ramirez L, et al. Molecular characterization of ten F8 splicing mutations in RNA isolated from patient’s leucocytes: assessment of in silico prediction tools accuracy. *Haemophilia.* 2015; **21**(2): 249-257.

- 28David D, Santos IMA, Johnson K, Tuddenham EGD, McVey JH. Analysis of the consequences of premature termination codons within factor VIII coding sequences. *J Thromb Haemost.* 2003; **1**(1): 139-146.