Dissection of pleiotropic effects of variants in and adjacent to *F8* exon 19 and rescue of mRNA splicing and protein function

Silvia Lombardi,[1] Gabriele Leo,[1] Simone Merlin,[2] Antonia Follenzi,[2] John H. McVey,[3] Iva Maestri,[4] Francesco Bernardi,[1] Mirko Pinotti,[1,*] and Dario Balestra[1,*]

Summary

The pathogenic significance of nucleotide variants commonly relies on nucleotide position within the gene, with exonic changes generally attributed to quantitative or qualitative alteration of protein biosynthesis, secretion, activity, or clearance. However, these changes may exert pleiotropic effects on both protein biology and mRNA splicing due to the overlapping of the amino acid and splicing codes, thus shaping the disease phenotypes. Here, we focused on hemophilia A, in which the definition of *F8* variants' causative role and association to bleeding phenotypes is crucial for proper classification, genetic counseling, and management of affected individuals. We extensively characterized a large panel of hemophilia A-causing variants (n = 30) within *F8* exon 19 by combining and comparing *in silico* and recombinant expression analyses. We identified exonic variants with pleiotropic effects and dissected the altered protein features of all missense changes. Importantly, results from multiple prediction algorithms provided qualitative results, while recombinant assays allowed us to correctly infer the likely phenotype severity for 90% of variants.

Molecular characterization of pathogenic variants was also instrumental for the development of tailored correction approaches to rescue splicing affecting variants or missense changes impairing protein folding. A single engineered U1snRNA rescued mRNA splicing of nine different variants and the use of a chaperone-like drug resulted in improved factor VIII protein secretion for four missense variants. Overall, dissection of the molecular mechanisms of a large panel of HA variants allowed precise classification of HA-affected individuals and favored the development of personalized therapeutic approaches.

Introduction

Prediction and definition of the causative role of nucleotide variants are major aims of molecular genetics and biochemical studies. Hemophilia A (HA, HEMA [MIM: 306700]), a model disease in human genetics, is caused by the deficiency of coagulation factor VIII (FVIII) leading to bleeding episodes at soft tissues, joints, and muscles.[1]

FVIII is a large multi-domain protein encoded by *F8* (MIM: 300841) located at Xq28, and more than 3,000 unique variants have been reported to date in the European Association for Haemophilia and Allied Disorders (EAHAD) database (Factor VIII Gene [*F8*] Variant Database).[2] Together with the common intron 22 and intron 1 inversions (40%–45% and 1%–6% of severely affected individuals, respectively),[3,4] missense variants represent the most frequent cause of HA (~45%).[5] Based on residual plasma FVIII activity levels, HA-affected individuals are classified as 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), and the treatment regimen is established accordingly.[6]

The pathogenic significance of missense variants in relation to classification of affected individuals is commonly attributed to quantitative or qualitative changes in protein expression, with alteration of protein biosynthesis, secretion, activity, or clearance.[7–11] Nevertheless, exonic variants may have detrimental effects on mRNA splicing because the amino acid coding sequence overlaps with an intricate set of splicing regulatory elements controlling the exon fate.[12] As a consequence, the combination of mRNA splicing and protein alterations shapes a significant but still underestimated proportion of HA phenotypes, as previously demonstrated for the relatively frequent HA-causing variant (c.6046C>T [p.Arg2016Trp]) located in *F8* exon 19.[13,14]

In this context, particularly in consideration of the high number of reported variants, computational tools are commonly exploited to predict the combined effects on mRNA splicing and protein function.[15,16] However, predictions are not completely accurate,[17,18] which makes the experimental investigation to validate and refine the *in silico* output extremely important.

Here, we combined and compared *in silico* prediction tools and *in vitro* analyses to characterize, both at the mRNA and protein levels, HA-causing variants located within *F8* exon 19, chosen as a model in this study. As a matter of fact, exon 19 belongs to those (8 out of 26) that are poorly defined, as witnessed by its weak 3′ss, which points toward the presence of functional regulatory

[1]Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara 44121, Italy; [2]Department of Health Sciences, University of Piemonte Orientale, Novara 28100, Italy; [3]School of Bioscience and Medicine, University of Surrey, Guildford GU2 7XH, UK; [4]Department of Experimental and Diagnostic Medicine, University of Ferrara, Ferrara 44123, Italy
*Correspondence: pnm@unife.it (M.P.), blsdra@unife.it (D.B.)
https://doi.org/10.1016/j.ajhg.2021.06.012.

elements promoting proper exon inclusion. These features, and the paradigmatic example of the *F8* exon 19 c.6046C>T (p.Arg2016Trp) variant that elicited a pleiotropic effect,[14] prompted us to select an extended panel of HA-causing exon 19 variants (n = 23) as models to systematically dissect their differential impact on mRNA splicing and protein.

Results from multiple prediction algorithms provided only qualitative results, while the recombinant assays allowed to correctly infer the phenotype severity for 90% of variants, with the identification of exonic variants affecting splicing processing, protein features, or both. The resulting *F8* variants' classification in relation to HA coagulation phenotypes demonstrated the high reliability of *in vitro* assays. Further, the knowledge of the mechanisms underlying the altered phenotypes led to developing tailored correction approaches acting at mRNA (U1snRNA variants) or protein (chaperone-like compounds) level.

Material and methods

The sequence of all primers and detailed methods are provided as supplemental material (Tables S1–S3).

Nomenclature

The investigated variants were reported in the EAHAD Factor VIII Gene (*F8*) Variant and CDC Hemophilia A Mutation Project database. They are reported according to the HGVS nomenclature, with numbering starting at the A (+1) nucleotide of the AUG (codon 1) translation initiation codon. Intronic variants are numbered according to their position relative to the 5′ss. Reference sequences are GenBank: NG_011403 (LRG_555), NM_000132.4, and NP_000123.1.

Bioinformatic tools

The splice sites' strength of *F8* exon 19 was predicted by using the Splice Site Prediction tool.[19] The presence of exonic splicing regulatory motifs was predicted by exploiting the HOT-SKIP tool.[20] The prediction of splicing motifs and their binding proteins was conducted by exploiting the bioinformatic tool SpliceAid.[21]

The impact of exonic changes on FVIII was predicted by exploiting the REVEL tool.[22]

Construction of expression vectors

To generate the *F8* exon 19 minigene (pF8), the genomic region of human *F8* spanning from c.5999−436 to c.6115+381 was amplified from genomic DNA of a normal subject and cloned into the pTB expression vector by exploiting the *Nde*I restriction sites within primers. Exon 19 variants were introduced by site-directed mutagenesis. The expression vectors for the modified U7snRNAs (pU7^{a}, pU7^{b}) were generated as previously reported.[23]

The expression plasmids for the modified U1snRNAs (pU1$^{A\text{-}D}$) were created by replacing the sequence between the BglII and *Xba*I restriction sites with an engineered U1 cassette generated by PCR. Exonic changes were introduced by site-specific mutagenesis into the codon-optimized B-less human FVIII cDNA.[24] All constructs were validated by DNA sequencing.

Splicing assays

Human embryonic kidney (HEK) 293T cells were transiently transfected with pF8 minigenes alone or with a molar excess (1.5×) of the pU1/pU7 plasmids. Total RNA was isolated 24 h post-transfection, reverse-transcribed (RT) with random hexamers, and amplified with minigene-specific primers.[25] Amplified fragments were analyzed by agarose gel and denaturing capillary electrophoresis approaches. All RT-PCR fragments were validated by sequencing.

Pull-down of RNA-binding proteins

RNA biotinylated oligonucleotides were incubated with HeLa nuclear extract and RNA-protein complexes were isolated with streptavidin-coated magnetic beads. Proteins were analyzed through western blotting with monoclonal antibodies.

Protein expression studies

HEK293T cells were transiently transfected in 24-well plates with pFVIII vectors with Lipofectamine 2000, as per manufacturer's instructions. Culture medium was replaced 4 h post-transfection with fresh OptiMEM with or without sodium phenylbutyrate (Sigma-Aldrich). Media were harvested 48 h after transfection, centrifuged for 5 min at 3,000 × *g*, and stored at −20°C. Secreted FVIII levels and cofactor FVIII activity were determined in cell media through a commercial ELISA kit (F8C-EIA, Affinity Biologicals) and Biophen Factor VIII:C Chromogenic assay (Hyphen Biomed), respectively. Secreted FVIII antigen and activity levels were expressed as percent of the wild-type recombinant FVIII (rFVIII) and specific activity was calculated as the ratio between activity and protein levels.

Statistical analysis

Data analysis was performed by unpaired t test.

Results

In silico tools predicted detrimental effects for all *F8* exon 19 variants

The SpliceAid computational tool predicted splicing regulatory elements recognized by RNA-binding proteins that would promote proper exon 19 inclusion (Figure 1). To experimentally dissect their functional relevance, we expressed the wild-type *F8* minigene and exploited an antisense approach based on modified U7snRNAs to mask two 20–nucleotide-long sequences of exon 19. Whereas the minigene alone showed correct processing, consistent with what was observed in liver samples (Figure S1), its co-expression with the antisense U7snRNAs resulted in almost complete exon 19 skipping (Figure 1, inset). These findings indicated the presence of functional splicing regulatory elements and prompted us to investigate the impact of all reported exonic variants on both splicing and protein function by *in silico* prediction tools (Table 1). In particular, we exploited the HOT-SKIP algorithm to identify variants affecting pre-mRNA splicing, which were identified by a score < -5, and the REVEL tool to infer their impact on FVIII protein based on the predicted pathogenicity score (severe, score

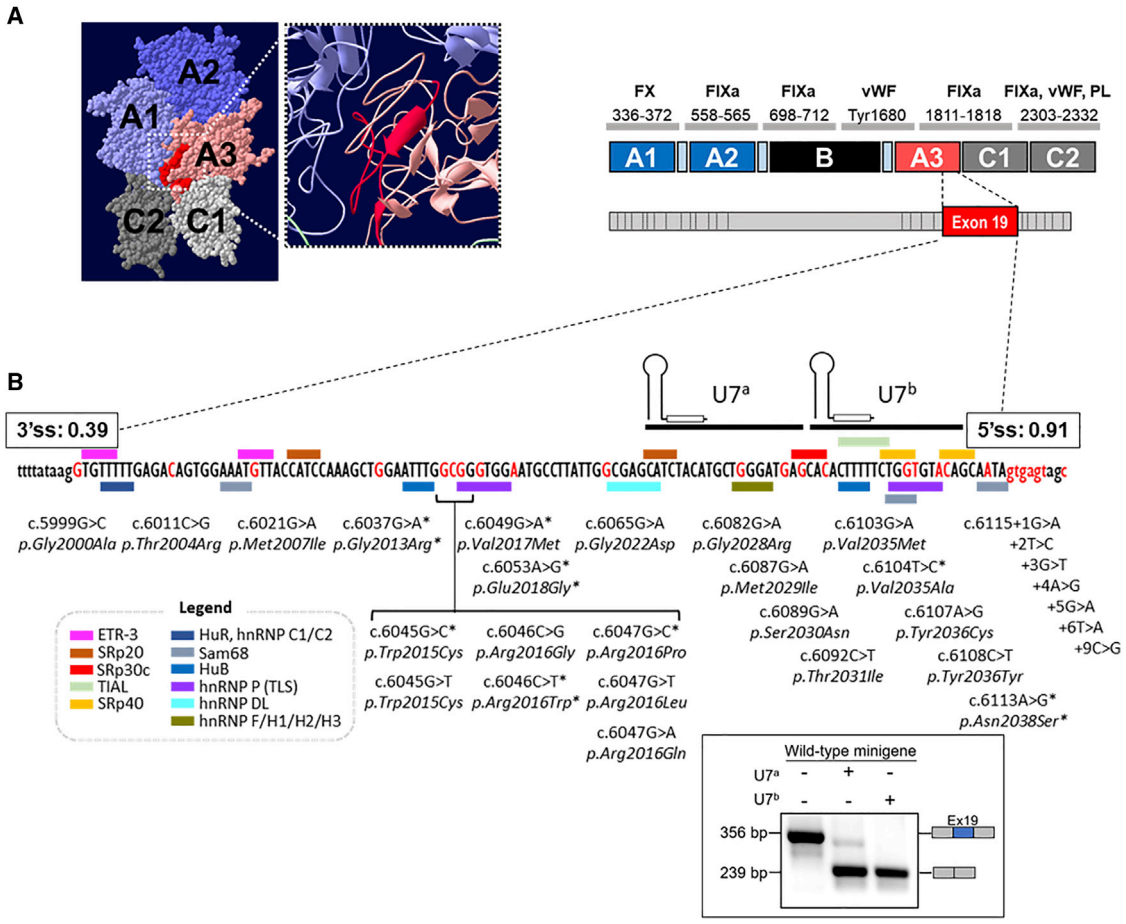


Figure 1. ***F8*** **exon 19 contains dense splicing regulatory elements**
(A) Crystal structure (PDB: 2R7E) and schematic representation of FVIII. The regions of FVIII interacting with other coagulation factors are reported: factor X (FX), activated factor IX (FIXa), von Willebrand factor (vWF), and phospholipids (PL). The involved residues are also shown. Grey bars are not in scale.
(B) Sequence of *F8* exon 19, with exonic and intronic nucleotides indicated by capital and lowercase letters, respectively. With the sequence we report the predicted binding sites for splicing factors (predicted through SpliceAid tool), the binding sites of the modified U7snRNAs, and the investigated variants (reported as nucleotide and amino acid changes). All variants previously analyzed[14] are indicated by asterisks. Exon 19 is shown in red. The splicing pattern resulting from co-expression of modified U7snRNAs ($U7^a$, $U7^b$) is also shown (inset). Scores of the splice sites, predicted through the Splice Site Prediction tool, are reported at the exon boundaries.

0.8–1; moderate, score 0.6–0.8; mild, score 0.4–0.6; negligible, score < 0.4).

The majority of exonic variants (17 out of 23) were predicted to have from weak to strong impact on FVIII protein, whereas four variants (p.Gly2000Ala, p.Arg2016Gly, p.Arg2016Trp, p.Glu2018Gly) were expected to affect both splicing and protein features. In addition, two exonic variants with predicted null (c.6108C>T [p.Tyr2036Tyr]) or negligible (p.Asn2038Ser) effects on protein were candidates for altered mRNA splicing. Seven intronic variants occurring at the 5′ end (+1 to +9) of intron 19 were also included in the study as HA-causing variants affecting splicing *by default*, mostly associated with severe HA.

Overall, the integrated bioinformatics analysis provided a plausible interpretation of the molecular mechanisms underlying the pathogenic effect of all submitted variants, albeit with some discrepancies (Table 1, asterisks) when compared to the reported phenotypes of the HA-affected individuals.

Minigene expression studies indicated a differential impact of exon 19 variants on splicing

Due to the unavailability of samples from affected individuals, splicing impairment was assessed through the expression of minigene variants (Figure 2A). Most of the exonic changes (16 out of 23) did not affect or slightly reduced (p.Gly2028Arg, 83.5% ± 14.5%; p.Met2029Ile, 94.5% ± 6.2%) exon 19 inclusion compared to the wild-type. In contrast, three exonic variants (p.Gly2000Ala, p.Arg2016Gly, and c.6108C>T [p.Tyr2036Tyr]) caused significant exon skipping (58.9% ± 4.9%, 7.5% ± 4.1%, and 51.9% ± 11.3% of exon inclusion, respectively), as we previously showed for four additional variants (p.Gly2013Arg, p.Arg2016Trp, p.Glu2018Gly, and p.Asn2038Ser).[14] In

Table 1. Bioinformatic prediction of the effects of *F8* exon 19 variants

		Subjects' data (EAHAD database)			*In silico* prediction		Predicted impact		
					Splicing	Protein			
Variant	Protein change	Subjects (n)	FVIII:C[a] (%)	Severity[a]	HOT-SKIP[b]	REVEL[c]	on splicing[d]	on protein[e]	Inferred pathogenic mechanism
c.5999G>C	p.Gly2000Ala	3	21	mild*	–	0.871	X	XXX	*severe impact on protein and splicing
c.6011C>G	p.Thr2004Arg	1	NR	NR	−0.22	0.964	–	XXX	severe impact on protein
c.6021G>A	p.Met2007Ile	2	14–30	mild*	1.58	0.934	–	XXX	*severe impact on protein
c.6037G>A[f]	p.Gly2013Arg[f]	2	<1	severe	0.01	0.976	–	XXX	severe impact on protein
c.6045G>T	p.Trp2015Cys	4	<1	severe	−2	0.945	–	XXX	severe impact on protein
c.6045G>C[f]	p.Trp2015Cys	1	NR	NR	0	0.945	–	XXX	severe impact on protein
c.6046C>G	p.Arg2016Gly	1	NR	moderate	−22	0.846	X	XXX	severe impact on protein and splicing
c.6046C>T[f]	p.Arg2016Trp[f]	100	1	moderate	−17	0.822	X	XXX	severe impact on protein and splicing
c.6047G>A	p.Arg2016Gln	1	38	mild	2	0.668	–	XX	moderate impact on protein
c.6047G>T	p.Arg2016Leu	4	18	mild	3.2	0.55	–	X	mild impact on protein
c.6047G>C[f]	p.Arg2016Pro	2	<1	severe	4.67	0.788	–	XX	moderate impact on protein
c.6049G>A[f]	p.Val2017Met	3	44.5	mild	0.83	0.658	–	XX	moderate impact on protein
c.6053A>G[f]	p.Glu2018Gly[f]	9	2	moderate	−9.8	0.869	X	XXX	severe impact on protein and splicing
c.6065G>A	p.Gly2022Asp	4	1	moderate	0.32	0.946	–	XXX	severe impact on protein
c.6082G>A	p.Gly2028Arg	10	14	mild*	4	0.945	–	XXX	*severe impact on protein
c.6087G>A	p.Met2029Ile	2	6–32	mild*	0.57	0.89	–	XXX	*severe impact on protein
c.6089G>A	p.Ser2030Asn	63	27	mild	−0.13	0.573	–	X	mild impact on protein
c.6092C>T	p.Thr2031Ile	1	NR	mild	−3	0.746	–	XX	moderate impact on protein
c.6103G>A	p.Val2035Met	5	5	moderate	−3.33	0.936	–	XXX	severe impact on protein
c.6104T>C[f]	p.Val2035Ala	38	11	mild*	1	0.955	–	XXX	*severe impact on protein
c.6107A>G	p.Tyr2036Cys	1	3	moderate	−0.47	0.878	–	XXX	severe impact on protein
c.6108C>T	p.Tyr2036Tyr	1	NR	mild	−7	–	X	–	impact on splicing
c.6113A>G[f]	p.Asn2038Ser[f]	19	10	mild	–	0.254	X	–	impact on splicing
c.6115+1G>A	–	1	<1	severe	–	–	X	–	impact on splicing
c.6115+2T>C	–	2	<1	severe	–	–	X	–	impact on splicing
c.6115+3G>T	–	3	<1	severe	–	–	X	–	impact on splicing

(*Continued on next page*)

Table 1. *Continued*

Variant	Protein change	Subjects' data (EAHAD database)			*In silico* prediction		Predicted impact		Inferred pathogenic mechanism
		Subjects (n)	FVIII:C[a] (%)	Severity[a]	Splicing HOT-SKIP[b]	Protein REVEL[c]	on splicing[d]	on protein[e]	
c.6115+4A>G	–	1	NR	severe	–	–	X	–	impact on splicing
c.6115+5G>A	–	2	<1–1	severe	–	–	X	–	impact on splicing
c.6115+6T>A	–	1	<1	severe	–	–	X	–	impact on splicing
c.6115+9C>G	–	1	23	mild	–	–	X	–	impact on splicing

Asterisk indicates major discrepancies between reported patients' phenotype and predicted severity. NR, not reported.
[a]For variants with n patients ≥ 3, the median FVIII:C (%) and predicted severity are reported.
[b]The reported values correspond to the difference between wild-type and mutated ESS/ESE ratios according to HOT-SKIP (see web resources).
[c]See REVEL in web resources.
[d]Splicing-affecting variants were identified by values ≤ −5 (arbitrary threshold).
[e]The REVEL pathogenicity score was arbitrarily defined as severe (XXX; values 0.8–1), moderate (XX; 0.6–0.8), mild (X; 0.4–0.6), or negligible (<0.4).
[f]Variant investigated also in Donadon et al.[14]

contrast, all intronic variants resulted in complete exon skipping, except for the c.6115+9C>G variant that showed residual exon 19 inclusion levels (48.5% ± 15.1%).

To further detail the molecular mechanisms underlying exon skipping, we selected the p.Arg2016Gly variant as a model due to its strong impact on splicing. We extended the analysis to the p.Arg2016Trp variant, located at the same nucleotide position but associated with a mild effect on splicing, and to the p.Glu2018Gly variant, reported to strongly alter exon 19 inclusion. Bioinformatic prediction by the SpliceAid tool identified the heterogeneous nuclear ribonucleoproteins (hnRNP) H and F as candidates for specific binding to the variant sequences. Notably, RNA pull-down assays validated the prediction and demonstrated the increased hnRNP F/H binding on all exonic variants compared to the wild-type counterpart, with a preference for the p.Arg2016Gly (3.3-fold) and p.Glu2018Gly (4.3-fold) variants (Figure 2B).

Overall, the splicing analysis validated the bioinformatic prediction demonstrating the derangement of the splicing process for some missense and all intronic variants. In addition, *in vitro* studies provided further details of the molecular mechanisms responsible for exon skipping for three exonic variants (p.Arg2016Gly, p.Arg2016Leu, and p.Glu2018Gly).

Recombinant FVIII expression studies indicated a differential impact of amino acid changes on protein

We experimentally evaluated the impact of missense variants on FVIII protein features by transient expression of B-domain deleted and codon-optimized FVIII in HEK293T cells. As shown in Figure 2C, data led us to classify variants into four groups based on rFVIII protein secretion and activity levels.

The first group (p.Thr2004Arg, p.Trp2015Cys, and p.Arg2016Pro) was characterized by low or undetectable levels of secreted rFVIII protein and functional activity in conditioned media. A second group showed strongly (≤ 5%, p.Gly2022Asp and p.Tyr2036Cys) or moderately (5%–10%, p.Arg2016Gly, p.Arg2016Leu, and p.Val2035Met) reduced activity associated with either strongly (p.Arg2016Gly, p.Arg2016Leu, p.Val2035Met, and p.Tyr2036Cys) or moderately (60%; p.Gly2022Asp) reduced secreted protein levels. The third group of variants (p.Arg2016Gln, p.Val2017Met, p.Gly2028Arg, p.Met2029Ile, p.Ser2030Asn, and p.Thr2031Ile) was associated with a modest impact on rFVIII expression, as shown by the reduced but appreciable levels of rFVIII protein (25%–70% of wild-type) and activity (10%–50%). Finally, the fourth group comprises the p.Gly2000Ala variant, which showed no impact on rFVIII protein, with secretion (104% ± 9.7%) and activity (96.1% ± 6.0%) levels comparable to those of wild-type rFVIII.

Interestingly, analysis of rFVIII-specific activities (Figure 2C, values within boxes) further detailed the defect type caused by each missense variant. In particular, specific

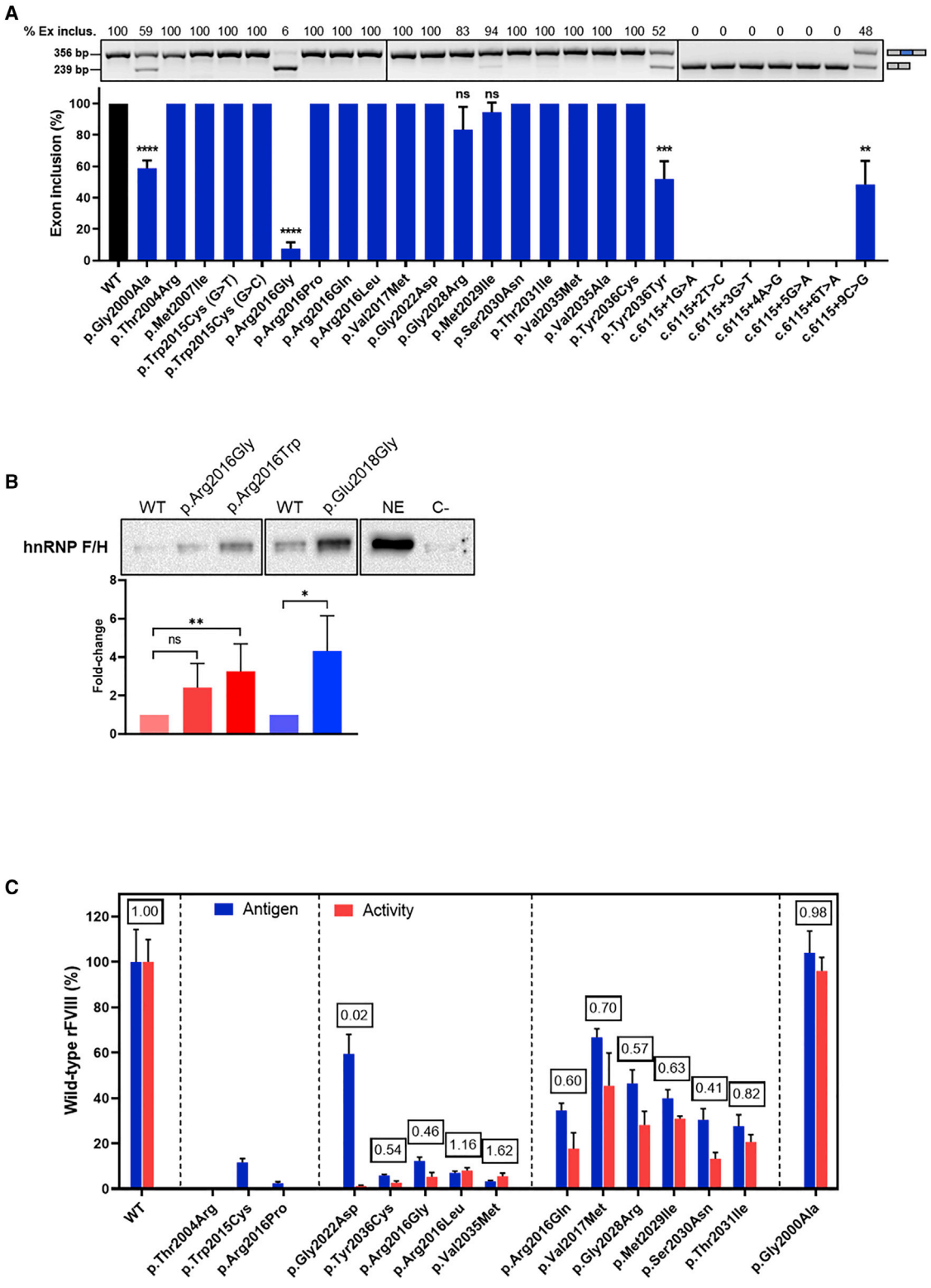


Figure 2. Characterization of the impact of exon 19 variants on splicing and protein

(A) Exon 19 splicing assays. Minigenes were transfected in HEK293T cells and the splicing pattern evaluated by RT-PCR followed by agarose gel electrophoresis. The levels of exon inclusion, expressed as percentage of all transcripts, were estimated by densitometric analysis of bands (ImageJ).

(B) Western blot analysis of pull-down experiments. Results are presented as mean ± standard deviation (SD) of three independent experiments and a representative blot is shown (NE, nuclear extract input; C-, naked beads). Histograms report fold changes over the wild-type.

(legend continued on next page)

activities close to that of the wild-type protein (=1) pointed toward a primary secretion defect, as for variants such as p.Arg2016Leu and p.Val2035Met. Conversely, low rFVIII-specific activity indicated a detrimental impact on FVIII activity rather than on secretion, as observed for the p.Gly2022Asp variant.

In comparison to the bioinformatic prediction, through recombinant expression studies we demonstrated the specific rFVIII features of each missense variant, resulting in the alteration of protein activity and/or secretion.

Combined effects of protein and splicing mechanisms

Data from *in vitro* experiments allowed us to infer the residual FVIII activity levels and to assign a predicted phenotype to each variant (Table 2). This prediction was in agreement with the reported patients' phenotype for >89% (17 out of 19) of the exonic and all intronic variants. The p.Arg2016Gly gave a conflicting result since it has been reported in one affected individual with moderate HA, whereas our experimental system showed a severe output. In comparison, *in silico* prediction correctly classified a lower percentage of variants (25%–65%), even when the most stringent ranges were chosen (Figure 3).

Overall, we identified variants with major effects on (1) splicing (p.Gly2000Ala, c.6108C>T [p.Tyr2036Tyr], and p.Asn2038Ser), (2) FVIII protein secretion/activity (p.Thr 2004Arg, p.Trp2015Cys, p.Arg2016Gln, p.Arg2016Leu, p.Arg2016Pro, p.Val2017Mey, p.Gly2022Asp, p.Ser2030-Asn, p.Thr2031Ile, p.Val2035Mer, and p.Tyr2036Cys), or (3) both (p.Gly2013Arg, p.Arg2016Gly, p.Arg2016Trp, p.Glu2018Gly, p.Gly2028Arg, and p.Met2029Ile).

Tailored correction approaches can rescue multiple FVIII variants

RNA-based approaches

Based on previous findings,[26–29] all exon 19 splicing-affecting variants were challenged by a correction approach based on modified U1snRNAs (U1s). First, we performed a preliminary evaluation on the c.6115+4A>G variant, previously associated with complete exon skipping. In particular, we tested one complementary ($U1^{A}$) and three exon-specific (ExSpeU1; $U1^{B}$, $U1^{C}$, $U1^{D}$) U1s (Figure S2), with the latter designed to target less-conserved intronic sequences to ensure higher target specificity and lower off-target effects. In co-transfection experiments, three out of four U1s ($U1^{A}$, $U1^{B}$, $U1^{C}$) efficiently rescued exon 19 definition (91.2% ± 3.9%, 75.2% ± 3.1%, and 37.1% ± 2.5% of correct transcript, respectively). We selected the best performing ExSpeU1 ($U1^{B}$) to challenge all the other splicing affecting variants. Co-expression of $U1^{B}$ efficiently rescued all exonic as well as three intronic variants, with an exon inclusion increase ranging from +18% to +68% (Figure 4A, upper panel, Figure S3). In contrast, splicing analysis for the c.6115+1G>A, +2T>C, and +5G>A variants revealed that the $U1^{B}$ induced the usage of a cryptic 5′ss located 13 nucleotides downstream of the natural one, thus resulting in partial intron 19 retention (Figure 4A, lower panel). In the attempt to rescue splicing of the c.6115+5 variant, and based on previous findings,[30] we exploited a combined approach in which the $U1^{B}$ was co-delivered with an engineered U6snRNA designed to fully complement the c.6115+5G>A variant ($U6^{ATT}$). A modified U6snRNA, designed to complement the wild-type 5′ss ($U6^{ACT}$), was used as control (Figure S4A). Even though the delivery of either U6s alone did not exert any effect on splicing outcome, the co-delivery of $U1^{B}$ and $U6^{ATT}$ partially rescued the c.6115+5G>A variant (from 0% to 3%) (Figures S4B and S4C).

Chaperone-based approaches

Missense variants can impair protein folding and intracellular processing, which may be ameliorated by small compounds with chaperone-like activity.[31,32] Therefore, we focused on those missense variants (p.Arg2016Leu, p.Val2035Met, and p.Thr2030Ile) with a specific activity close (>0.8) to that observed for the wild-type rFVIII, and on the p.Ser2030Asn due to its frequency (63 reported patients). We challenged the selected variants with the sodium phenyl-butyrate (NaPBA) compound, previously shown to restore secretion of coagulation factor IX missense variants.[33] Treatment of transfected cells with 2 mM NaPBA resulted in a ∼4-fold increase in antigen and activity levels for the wild-type rFVIII (up to 396%) and the p.Arg2016Leu (up to 45%) and p.Val2035Met (up to 25%) variants. The p.Ser2030Asn and p.Thr2030Ile variants showed 5-fold (up to 37%) and 2-fold (up to 46%) increase, respectively (Figure 4B). Notably, NaPBA treatment improved the secretion of rFVIII variants but neither improved nor worsened the specific activity.

Overall, our data showed that tailored correction approaches, based on engineered U1snRNA or chaperone-like compounds, are effective in rescuing multiple FVIII variants, and further validated the predicted impact of missense changes on FVIII expression.

Discussion

Definition of the pathogenic role of nucleotide variants is a major aim of molecular genetics and biochemical studies.[34–36] This is particularly relevant in the hemophilia A field, where the knowledge of the pathogenic

(C) Expression of rFVIII missense variants. rFVIII antigen and activity levels in media from transiently transfected HEK293T cells were measured by ELISA and chromogenic assays, respectively. Results are expressed as percentage of wild-type rFVIII. Missense variants are grouped as specified in the results section. Specific activity (values within boxes) was calculated as the ratio between activity and antigen levels. Results are presented as mean ± SD of three independent experiments. ns, not significant; *$p < 0.1$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

Table 2. Overview of *F8* exon 19 variants and their *in vitro* characterization

		Subjects' data (EAHAD database)			*In vitro* expression data					
Variant	**Protein change**	**Subjects (n)**	**FVIII:C[a] (%)**	**Severity[a]**	**Exon inclusion (%)**	**FVIII:Ag (%)**	**FVIII:Act[b] (%)**	**Inferred activity[c]**	**Inferred severity**	**Concordance[d]**
c.5999G>C	p.Gly2000Ala	3	21	mild	58.9 ± 4.9	104.0 ± 9.7	96.1 ± 6.0	48.7–65.1	mild	+
c.6011C>G	p.Thr2004Arg	1	NR	NR	100	ND	ND	<1	severe	N/A[e]
c.6021G>A	p.Met2007Ile	2	14–30	mild	100	*	*	N/A	N/A	N/A
c.6037G>A[f]	p.Gly2013Arg[f]	2	<1	severe	41 ± 3	7.0 ± 0.9	8.4 ± 0.8	2.9–4.0	moderate	–
c.6045G>T	p.Trp2015Cys	4	<1	severe	100	11.8 ± 1.7	ND	<1	severe	++
c.6045G>C[g]	p.Trp2015Cys	1	NR	NR	100	11.8 ± 1.7	ND	<1	severe	N/A[e]
c.6046C>G	p.Arg2016Gly	1	NR	moderate	7.5 ± 4.1	12.4 ± 1.6	5.3 ± 1.9	0.1–0.8	severe	–
c.6046C>T[f]	p.Arg2016Trp[f]	100	1	moderate	70 ± 5	11.0 ± 0.4	6.0 ± 2.9	2.0–6.7	mod-mil	+
c.6047G>T	p.Arg2016Leu	4	18	mild	100	7.1 ± 0.7	8.1 ± 1.3	6.8–9.4	mild	+
c.6047G>C[g]	p.Arg2016Pro	2	<1	severe	100	2.5 ± 0.6	ND	<1	severe	++
c.6047G>A	p.Arg2016Gln	1	38	mild	100	34.6 ± 3.2	17.8 ± 7.0	10.8–24.8	mild	+
c.6049G>A[g]	p.Val2017Met	3	45	mild	100	66.8 ± 3.7	45.5 ± 14.4	31.1–59.9	mild	++
c.6053A>G[f]	p.Glu2018Gly[f]	9	2	moderate	28 ± 2	69.0 ± 18.1	19.4 ± 2.3	4.4–6.5	mod-mil	+
c.6065G>A	p.Gly2022Asp	4	1	moderate	100	59.5 ± 8.5	1.2 ± 0.4	0.8–1.6	sev-mod	++
c.6082G>A	p.Gly2028Arg	10	14	mild	83.5 ± 14.5	46.4 ± 6.0	28.2 ± 6.0	15.3–33.5	mild	+
c.6087G>A	p.Met2029Ile	2	6–32	mild	94.5 ± 6.2	40.0 ± 3.7	30.9 ± 1.1	26.3–32.2	mild	++
c.6089G>A	p.Ser2030Asn	63	27	mild	100	31.4 ± 7.6	13.4 ± 2.6	10.8–16.0	mild	+
c.6092C>T	p.Thr2031Ile	1	NR	mild	100	27.6 ± 5.1	20.6 ± 3.2	17.4–23.8	mild	++
c.6103G>A	p.Val2035Met	5	5	moderate	100	3.3 ± 0.4	5.6 ± 1.2	4.4–6.8	mod-mil	++
c.6104T>C[g]	p.Val2035Ala	38	11	mild	100	*	*	N/A	N/A	N/A
c.6107A>G	p.Tyr2036Cys	1	3	moderate	100	5.9 ± 0.6	2.7 ± 0.8	1.9–3.5	moderate	++
c.6108C>T	p.Tyr2036Tyr	1	NR	mild	51.9 ± 11.3	**	**	40.6–63.2	mild	++
c.6113A>G[f]	p.Asn2038Ser[f]	19	10	mild	26 ± 2	NR	99.6 ± 12.5	20.9–31.4	mild	+
c.6115+1G>A	–	1	<1	severe	ND	**	**	<1	severe	++
c.6115+2T>C	–	2	<1	severe	ND	**	**	<1	severe	++
c.6115+3G>T	–	3	<1	severe	ND	**	**	<1	severe	++
c.6115+4A>G	–	1	NR	severe	ND	**	**	<1	severe	++
c.6115+5G>A	–	2	<1–1	severe	ND	**	**	<1	severe	++

(*Continued on next page*)

Table 2. *Continued*

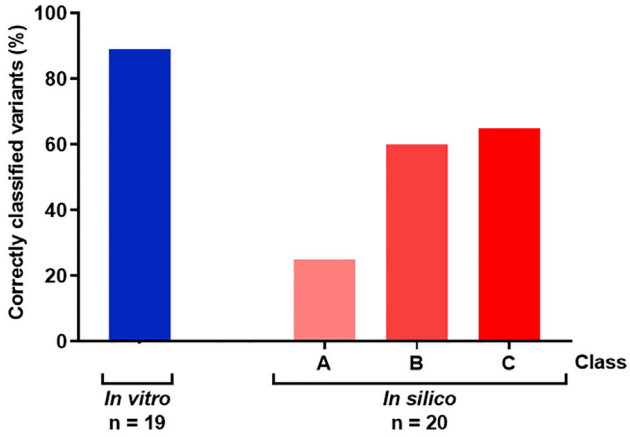
Variant	Protein change	Subjects' data (EAHAD database)			*In vitro* expression data				
		Subjects (n)	FVIII:C[a] (%)	Severity[a]	Exon inclusion (%)	FVIII:Ag (%)	FVIII:Act[b] (%)	Inferred activity[c]	Inferred severity	Concordance[d]
c.6115+6T>A	–	1	<1	severe	ND	**	**	<1	severe	++
c.6115+9C>G	–	1	23	mild	48.5 ± 15.1	**	**	33.4-63.6	mild	+

NR, not reported; ND, not detectable; N/A, not applicable; asterisk (*) indicates the missense variant was not tested due to experimental issues; double asterisk (**) indicates The protein variant was not produced because it is equivalent to the wild-type rFVIII.
[a]For variants with n patients ≥ 3, the median FVIII:C (%) and predicted severity are reported.
[b]The activity of recombinant variants was assessed by chromogenic assay.
[c]Inferred activity was calculated as (Exon inclusion*FVIII:Act)/100.
[d]++ when phenotype and FVIII activity match, + with only one parameter, – no matches
[e]Concordance was not assessable due to the lack of information on patients' severity.
[f]*In vitro* data from Donadon et al.[14]
[g]The splicing pattern has been tested also in Donadon et al.[14]

mechanisms is crucial for proper classification, genetic counseling, and management of affected individuals and also in the perspective of designing targeted therapeutic strategies.[37]

Commonly, nucleotide position within the gene structure dictates the predicted pathogenic effect, with exonic changes generally attributed to quantitative or qualitative alteration of protein biosynthesis, secretion, activity, or clearance. However, exonic variants may exert pleiotropic effects and impair mRNA splicing due to the overlapping of the amino acid and splicing codes,[38] thus shaping intricate forms of disease phenotypes. This combined effect has previously been demonstrated in different genomic contexts, including coagulation factors *F9* (MIM: 300746) and *F8* (MIM: 300841).[14,39,40]

However, the systematic characterization at the splicing and protein levels of an entire exonic sequence has been hampered to date by the need for extensive *in vitro* studies. To address this issue, we took advantage of bioinformatics tools and experimental assays to deeply characterize a large panel of *F8* variants. In particular, we chose as model the *F8* exon 19, coding for amino acids of the FVIII A3 domain mainly involved in the interaction of FVIIIa with FIXa, and in which several missense variants have been described. Importantly, *F8* exon 19 is poorly defined and harbors several putative splicing regulatory elements, suggesting that missense variants may have a combined effect on both splicing process and protein features. The exploitation of two bioinformatics tools, combining results from multiple prediction algorithms to give more reliable results,[20,22,41] provided a qualitative prediction of the effects of exon 19 variants. Nevertheless, the lack of experimentally determined thresholds associated with defined patients' severity limited the predictive ability of these bioinformatic tools. In particular, three protein variants (c.6046C>G [p.Arg2016Gly], c.6046C>T [p.Arg2016Trp], and c.6053A>G [p.Glu2018Gly]) and one synonymous codon (c.6108C>T [p.Tyr2036Tyr]) were predicted by the HOT-SKIP tool to strongly impair exon 19 definition, in addition to the two variants affecting the 3′ss (c.5999G>C [p.Gly2000Ala]) and 5′ss (c.6113A>G [p.Asn 2038Ser]), considered to impair splicing by default. Interestingly, the *in vitro* splicing assays confirmed the prediction for these variants but, as previously reported,[14] also revealed that one additional missense change was associated with significant exon skipping (c.6037G>A [p.Gly2013Arg]). All intronic variants were associated with complete exon skipping, except the c.6115+9C>G change, which does not fall within the 5′ss consensus sequence and thus less prone to impair splicing. Notably, the extent of exon skipping did not correlate with the HOT-SKIP score, highlighting that bioinformatic tools are best suited for qualitative rather than quantitative prediction. Furthermore, pull-down studies provided preliminary but significant insights on the molecular mechanism leading to altered splicing, where variants c.6046C>G (p.Arg2016Gly), c.6046C>T



Class	Severe	Moderate	Mild/Negligible
A	0.8-1	0.6-0.8	<0.6
B	0.9-1	0.75-0.9	<0.75
C	0.95-1	0.75-0.95	<0.75

Figure 3. Evaluation of the prediction power of *in vitro* and *in silico* analyses

In vitro expressed variants were considered as correctly classified if the inferred severity (class) corresponded to that reported in the EAHAD and CHAMP databases (see Table 1 and 2). Based on REVEL tool recommendation and the predicted "pathogenicity" scores, we arbitrarily set three classes (A–C) with increasing degree of stringency to infer the impact of protein changes and thus the associated severe, moderate, mild, or negligible phenotypes. See Table 1 for REVEL scores.

(p.Arg2016Trp), and the c.6053A>G (p.Glu2018Gly) create an exonic splicing silencer recognized by the hnRNP F/H class of proteins. Notably, the binding efficiency of hnRNP F/H proteins was paralleled by different levels of exon skipping for each variant, thus pointing toward the creation of a splicing silencer motif with different strengths.

Overall, *in vitro* splicing assays detailed the detrimental effects of different exonic variants (c.5999G>C [p.Gly2000Ala], c.6037G>A [p.Gly2013Arg], c.6046C>G [p.Arg2016Gly], c.6046C>T [p.Arg2016Trp], c.6053A>G [p.Glu2018Gly], c.6082G>A [p.Gly2028Arg], c.6087G>A [p.Met2029Ile], c.6108C>T [p.Tyr2036Tyr], c.6113A>G [p.Asn2038Ser])[14] and preliminarily pinpointed the molecular mechanisms behind exon skipping for three of them. Our data indicate that, despite our increasing knowledge of molecular mechanisms and the generation of better prediction algorithms,[17,18] the experimental investigation of nucleotide variants still represents the best approach to dissect the pathological mechanisms.

The need for integration of *in silico* predictions with experimental studies was clearly demonstrated in the context of FVIII protein evaluation, where the bioinformatics tools are not able to predict the impact of missense variants on protein secretion and activity. Here, the REVEL tool, which combines 13 different algorithms, provided qualitative information on the possible severity of most variants, but incorrectly classified the p.Gly2000Ala as deleterious. Indeed, recombinant expression studies revealed that this change was well tolerated by rFVIII, as demonstrated by secretion and activity levels comparable to those observed for the wild-type rFVIII, and suggested the altered splicing as the main pathogenic mechanism for this variant.

In vitro characterization of all missense changes allowed discrimination of the secretion defects, indicated by specific activity close to 1, from functional alterations, indicated by a low specific activity. It is worth noting that the specific activity represents the ratio between the activity and antigen levels, so a value close to 1 indicates a completely functional protein. Analysis of the paradigmatic c.6065G>A (p.Gly2022Asp) variant, associated with antigen levels in excess over the observed activity in both our experimental system and in one reported affected individual,[42] clearly demonstrates that this variant causes the synthesis of a secreted but dysfunctional rFVIII protein. Conversely, expression of the c.6047G>T (p.Arg2016Leu) and c.6103G>A (p.Val2035Met) variants showed low but comparable antigen and activity levels, suggesting the synthesis of a functional rFVIII protein with impaired secretion.

Overall, the *in vitro* characterization allowed us to evaluate residual activity for each variant, the key indicator of patients' phenotype and bleeding severity. This allowed the classification of exon 19 variants as severe (p.Thr2004Arg, p.Trp2015Cys, p.Arg2016Gly, and p.Arg2016Pro), moderate (p.Gly2013Arg, p.Arg2016Trp, p.Glu2018Gly, p.Val2035Met, and p.Tyr2036Cys), or mild (p.Gly2000Ala, p.Arg2016Gln, p.Arg2016Leu, p.Val2017Met, p.Gly2028Arg, p.Met2029Ile, p.Ser2030Asn, p.Thr2031Ile, c.6108C>T [p.Tyr2036Tyr], and p.Asn2038Ser). The reported residual plasma FVIII levels associated with variants in the HA databases have to be considered with caution due to several variables (such as assay protocols and washout period) that might influence them. Notwithstanding, most of the *in vitro* predicted severities (17/19) are in agreement with the reported patients' phenotypes.

Conversely, the *in silico* approach proved to be useful for qualitative indications but revealed reduced ability (60% at maximum) to classify *F8* missense variants even with different threshold scores adjustments, thus highlighting the need for experimental validation of bioinformatic predictions.

It is worth noting that the dissection of the molecular mechanisms of pathological variants can be exploited for therapeutic purposes, with the design of tailored correction approaches acting on mRNA splicing or protein folding from a therapeutic perspective.

Here, a modified U1snRNA was able to efficiently restore proper exon 19 definition for the majority variants

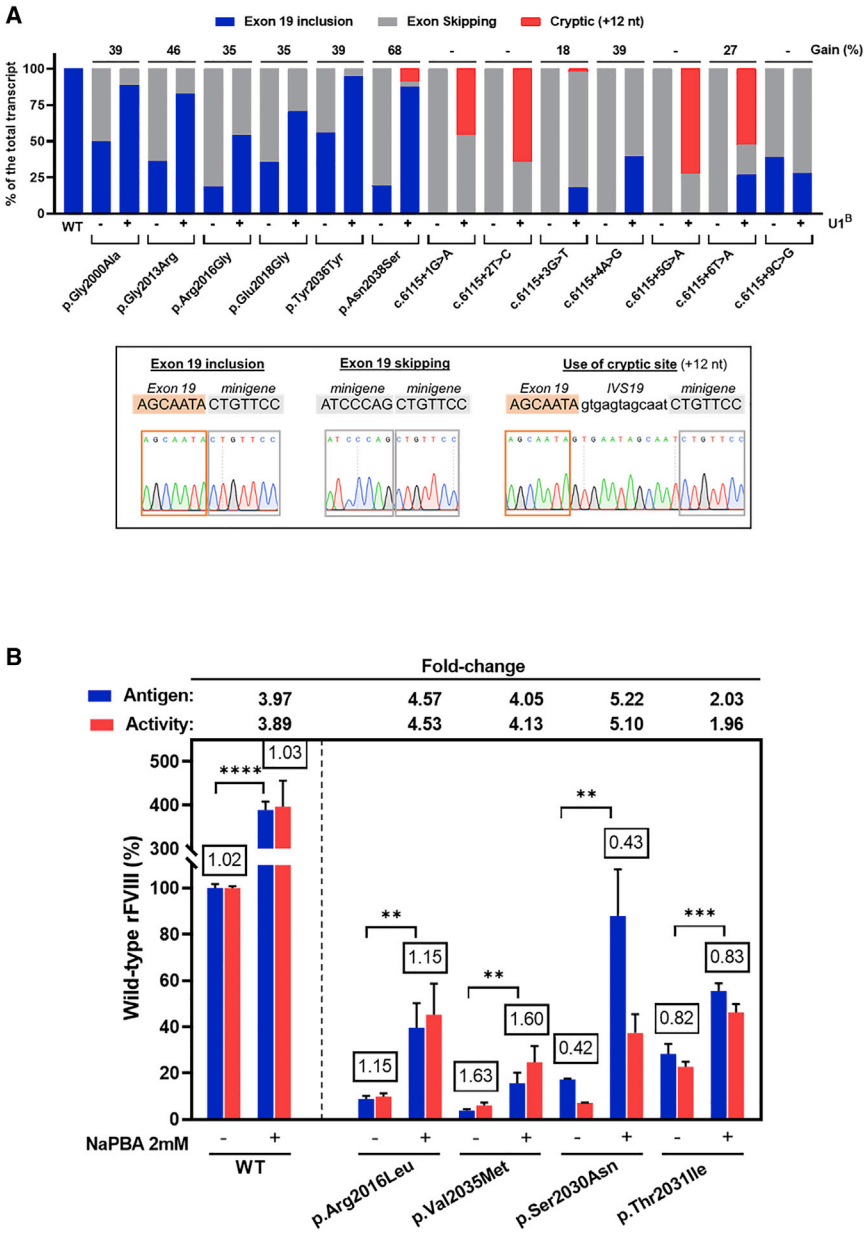


Figure 4. Targeted correction of exon 19 variants

(A) Analysis of splicing patterns in HEK293T cells expressing minigenes variants alone or in combination with $U1^B$. Results have been obtained by analysis of peaks from denaturing capillary electrophoresis of fluorescently labeled RT-PCR products (Figure S3). The amount of each transcript is represented as percent of the total. Gain was calculated as the difference between the percent of correct transcript (exon 19 inclusion) in treated (+) and not treated (−) groups. Electropherograms of each detected transcript is provided (lower panel).

(B) FVIII antigen and activity levels, measured respectively by ELISA and chromogenic assays, in media from transiently transfected HEK293T cells with and without the addition of NaPBA. Results are expressed as percentage of untreated wild-type rFVIII. Specific activity (values within boxes) was calculated as the activity/antigen ratio. Results are presented as mean ± SD of three independent experiments. The increase, expressed as fold-change compared to the untreated cells, is reported above. **p < 0.01; ***p < 0.001; ****p < 0.0001.

affecting splicing. In particular, all exonic and three intronic variants were efficiently rescued by a unique modified U1snRNA, demonstrating its potential applicability to multiple and different types of variants. Moreover, as previously reported,[30] the combination of engineered U1 and U6, both involved in the recognition of the 5′ss, can rescue splicing variants at +5 site insensitive to the U1 treatment. Altogether, these data provided insights into the molecular mechanism underlying the splicing defect, resulting from loss of exon definition due to nucleotide changes that can be counteracted by engineered U1/U6snRNA forcing its recognition.

Importantly, if translated at the protein level, the U1snRNA-mediated rescue of the intronic c.6115+3G>T, +4A>G, +6T>A variants, as well as the synonymous c.6108C>T (p.Tyr2036Tyr) change, would result in the production of wild-type FVIII, whereas the correction of c.5999G>C (p.Gly2000Ala) and c.6113A>G (p.Asn2038Ser) variants would result in the synthesis of fully functional and efficiently secreted FVIII protein variants. We are aware that the highly heterogeneous pattern of variants associated with hemophilia A (3,052 unique variants in 10,144 individual cases) limits the number of *F8* variants that could be treated by translation of the $U1^B$ to the clinic. However, these data further extend the U1snRNA potential, of particular interest for diseases with a highly represented splicing change to be targeted for therapy, as we showed in the paradigmatic examples of spinal muscular atrophy (SMA1 [MIM: 253300])[43] and familial dysautonomia (FD, HSAN3 [MIM: 223900]).[44]

On the other hand, a tailored correction approach based on the chaperone-like compound NaPBA, in the attempt to ameliorate protein folding and secretion, was exploited for the c.6047G>T (p.Arg2016Leu), c.6092C>T (p.Thr2031Ile), and c.6103G>A (p.Val2035Met) variants, characterized by functional FVIII with impaired secretion. Notably, treatment with NaPBA, which is an approved drug with different applications,[45,46] resulted in an increase (~4-fold compared to untreated cells) of secreted rFVIII levels that, if translated to affected individuals, would ameliorate the phenotype severity.

In conclusion, we extensively characterized a large panel of HA-causing variants by combining *in silico* and *in vitro* analysis and provided evidence for the pleiotropic effects of several exonic changes. Our data reveal that the currently available bioinformatics tools, in particular those combining results from multiple algorithms, can provide important information on the effects of nucleotide variants on splicing and protein features but still suffers from major drawbacks for quantitative prediction, thus limiting their predictive goals, particularly in genetic diseases characterized by a wide variation of clinical phenotypes with a modest variation in residual gene/protein function. We, therefore, suggest caution during variants classification mainly based on nucleotide location or bioinformatics prediction and highlight the importance of experimental characterization to dissect the molecular mechanisms underlying HA, which might pave the way for the development of individualized therapeutic strategies, also translatable to other genetic diseases.

Data and code availability

The published article includes all data generated or analyzed during this study.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.06.012.

Acknowledgments

D.B. received a 2017 Early Career Investigation Award from Bayer, which supported the study. D.B., S.L., G.L., M.P., and F.B. were also supported by the University of Ferrara. We would like to thank Professor Franco Pagani, from the International Center for Genetic Engineering and Biotechnology (ICGEB, Trieste, Italy), for providing the pTB and pSP64-U7 plasmids.

Declaration of interests

M.P. is the inventor of a patent (PCT/IB2011/054573) on modified U1snRNAs. All other authors declare no competing interests.

Received: March 16, 2021
Accepted: June 14, 2021
Published: July 8, 2021

Web resources

CDC Hemophilia A Mutation Project (CHAMP), https://www.cdc.gov/ncbddd/hemophilia/champs.html
European Association for Haemophilia and Allied Disorders (EAHAD) database (Factor VIII Gene (F8) Variant Database), https://f8-db.eahad.org/
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
HOT-SKIP, https://hot-skip.img.cas.cz/
OMIM, https://www.omim.org/
RCSB Protein Data Bank, http://www.rcsb.org/pdb/home/home.do
REVEL, https://sites.google.com/site/revelgenomics/
Splice Site Prediction by Neural Network (NNSPLICE v.0.9), https://www.fruitfly.org/seq_tools/splice.html
SpliceAid, http://www.introni.it/splicing.html

References

1. Peyvandi, F., Garagiola, I., and Young, G. (2016). The past and future of haemophilia: diagnosis, treatments, and its complications. Lancet *388*, 187–197.
2. McVey, J.H., Rallapalli, P.M., Kemball-Cook, G., Hampshire, D.J., Giansily-Blaizot, M., Gomez, K., Perkins, S.J., and Ludlam, C.A. (2020). The European Association for Haemophilia and Allied Disorders (EAHAD) Coagulation Factor Variant Databases: Important resources for haemostasis clinicians and researchers. Haemophilia *26*, 306–313.
3. Lakich, D., Kazazian, H.H., Jr., Antonarakis, S.E., and Gitschier, J. (1993). Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat. Genet. *5*, 236–241.
4. Bagnall, R.D., Waseem, N., Green, P.M., and Giannelli, F. (2002). Recurrent inversion breaking intron 1 of the factor VIII gene is a frequent cause of severe hemophilia A. Blood *99*, 168–174.
5. Jourdy, Y., Nougier, C., Roualdes, O., Fretigny, M., Durand, B., Negrier, C., and Vinciguerra, C. (2016). Characterization of five associations of F8 missense mutations containing FVIII B domain mutations. Haemophilia *22*, 583–589.
6. White, G.C., 2nd, Rosendaal, F., Aledort, L.M., Lusher, J.M., Rothschild, C., Ingerslev, J.; and Factor VIII and Factor IX Subcommittee (2001). 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. *85*, 560, 560.
7. Summers, R.J., Meeks, S.L., Healey, J.F., Brown, H.C., Parker, E.T., Kempton, C.L., Doering, C.B., and Lollar, P. (2011). Factor VIII A3 domain substitution N1922S results in hemophilia A due to domain-specific misfolding and hyposecretion of functional protein. Blood *117*, 3190–3198.
8. Wei, W., Zheng, C., Zhu, M., Zhu, X., Yang, R., Misra, S., and Zhang, B. (2017). Missense mutations near the N-glycosylation site of the A2 domain lead to various intracellular trafficking defects in coagulation factor VIII. Sci. Rep. *7*, 45033.
9. Gilbert, G.E., Novakovic, V.A., Kaufman, R.J., Miao, H., and Pipe, S.W. (2012). Conservative mutations in the C2 domains of factor VIII and factor V alter phospholipid binding and cofactor activity. Blood *120*, 1923–1932.
10. Jacquemin, M., Lavend'homme, R., Benhida, A., Vanzieleghem, B., d'Oiron, R., Lavergne, J.-M., Brackmann, H.H., Schwaab, R., VandenDriessche, T., Chuah, M.K.L., et al. (2000). A novel cause of mild/moderate hemophilia A: mutations scattered in the factor VIII C1 domain reduce factor VIII binding to von Willebrand factor. Blood *96*, 958–965.
11. Pipe, S.W., Eickhorst, A.N., McKinley, S.H., Saenko, E.L., and Kaufman, R.J. (1999). Mild hemophilia A caused by increased rate of factor VIII A2 subunit dissociation: evidence for nonproteolytic inactivation of factor VIIIa in vivo. Blood *93*, 176–183.
12. Falanga, A., Stojanović, O., Kiffer-Moreira, T., Pinto, S., Millán, J.L., Vlahoviček, K., and Baralle, M. (2014). Exonic splicing

signals impose constraints upon the evolution of enzymatic activity. Nucleic Acids Res. *42*, 5790–5798.
13. Castaman, G. (2018). Hemophilia A: different phenotypes may be explained by multiple and variable effects of the causative mutation in the *F8* gene. Haematologica *103*, 195–196.
14. Donadon, I., McVey, J.H., Garagiola, I., Branchini, A., Mortarino, M., Peyvandi, F., Bernardi, F., and Pinotti, M. (2018). Clustered *F8* missense mutations cause hemophilia A by combined alteration of splicing and protein biosynthesis and activity. Haematologica *103*, 344–350.
15. Margaglione, M., Castaman, G., Morfini, M., Rocino, A., Santagostino, E., Tagariello, G., Tagliaferri, A.R., Zanon, E., Bicocchi, M.P., Castaldo, G., et al.; AICE-Genetics Study Group (2008). The Italian AICE-Genetics hemophilia A database: results and correlation with clinical phenotype. Haematologica *93*, 722–728.
16. Ni, Y., and Eng, C. (2011). Response to Bayley: Functional Study Informs Bioinformatic Analysis. Am. J. Hum. Genet. *88*, 676.
17. Martorell, L., Corrales, I., Ramirez, L., Parra, R., Raya, A., Barquinero, J., and Vidal, F. (2015). Molecular characterization of ten F8 splicing mutations in RNA isolated from patient's leucocytes: assessment of in silico prediction tools accuracy. Haemophilia *21*, 249–257.
18. Jourdy, Y., Fretigny, M., Nougier, C., Négrier, C., Bozon, D., and Vinciguerra, C. (2019). Splicing analysis of 26 F8 nucleotide variations using a minigene assay. Haemophilia *25*, 306–315.
19. Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. J. Comput. Biol. *4*, 311–323.
20. Raponi, M., Kralovicova, J., Copson, E., Divina, P., Eccles, D., Johnson, P., Baralle, D., and Vorechovsky, I. (2011). Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. Hum. Mutat. *32*, 436–444.
21. Piva, F., Giulietti, M., Nocchi, L., and Principato, G. (2009). SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. Bioinformatics *25*, 1211–1213.
22. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am. J. Hum. Genet. *99*, 877–885.
23. Scalet, D., Sacchetto, C., Bernardi, F., Pinotti, M., van de Graaf, S.F.J., and Balestra, D. (2018). The somatic FAH C.1061C>A change counteracts the frequent FAH c.1062+5G>A mutation and permits U1snRNA-based splicing correction. J. Hum. Genet. *63*, 683–686.
24. Ward, N.J., Buckley, S.M.K., Waddington, S.N., Vandendriessche, T., Chuah, M.K.L., Nathwani, A.C., McIntosh, J., Tuddenham, E.G.D., Kinnon, C., Thrasher, A.J., and McVey, J.H. (2011). Codon optimization of human factor VIII cDNAs leads to high-level expression. Blood *117*, 798–807.
25. Scalet, D., Maestri, I., Branchini, A., Bernardi, F., Pinotti, M., and Balestra, D. (2019). Disease-causing variants of the conserved +2T of 5¿ splice sites can be rescued by engineered U1snRNAs. Hum. Mutat. *40*, 48–52.
26. Donadon, I., Bussani, E., Riccardi, F., Licastro, D., Romano, G., Pianigiani, G., Pinotti, M., Konstantinova, P., Evers, M., Lin, S., et al. (2019). Rescue of spinal muscular atrophy mouse models with AAV9-Exon-specific U1 snRNA. Nucleic Acids Res. *47*, 7618–7632.
27. Balestra, D., Giorgio, D., Bizzotto, M., Fazzari, M., Ben Zeev, B., Pinotti, M., Landsberger, N., and Frasca, A. (2019). Splicing Mutations Impairing CDKL5 Expression and Activity Can be Efficiently Rescued by U1snRNA-Based Therapy. Int. J. Mol. Sci. *20*, 4130.
28. Balestra, D., Scalet, D., Ferrarese, M., Lombardi, S., Ziliotto, N., C Croes, C., Petersen, N., Bosma, P., Riccardi, F., Pagani, F., et al. (2020). A Compensatory U1snRNA Partially Rescues FAH Splicing and Protein Expression in a Splicing-Defective Mouse Model of Tyrosinemia Type I. Int. J. Mol. Sci. *21*, 2136.
29. Balestra, D., Maestri, I., Branchini, A., Ferrarese, M., Bernardi, F., and Pinotti, M. (2019). An Altered Splicing Registry Explains the Differential ExSpeU1-Mediated Rescue of Splicing Mutations Causing Haemophilia A. Front. Genet. *10*, 974, 974.
30. Schmid, F., Hiller, T., Korner, G., Glaus, E., Berger, W., and Neidhardt, J. (2013). A gene therapeutic approach to correct splice defects with modified U1 and U6 snRNPs. Hum. Gene Ther. *24*, 97–104.
31. Chaudhuri, T.K., and Paul, S. (2006). Protein-misfolding diseases and chaperone-based therapeutic approaches. FEBS J. *273*, 1331–1349.
32. Muntau, A.C., Leandro, J., Staudigl, M., Mayer, F., and Gersting, S.W. (2014). Innovative strategies to treat protein misfolding in inborn errors of metabolism: pharmacological chaperones and proteostasis regulators. J. Inherit. Metab. Dis. *37*, 505–523.
33. Pignani, S., Todaro, A., Ferrarese, M., Marchi, S., Lombardi, S., Balestra, D., Pinton, P., Bernardi, F., Pinotti, M., and Branchini, A. (2018). The chaperone-like sodium phenylbutyrate improves factor IX intracellular trafficking and activity impaired by the frequent p.R294Q mutation. J. Thromb. Haemost. *16*, 2035–2043.
34. Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797.
35. Hopf, T.A., Ingraham, J.B., Poelwijk, F.J., Schärfe, C.P.I., Springer, M., Sander, C., and Marks, D.S. (2017). Mutation effects predicted from sequence co-variation. Nat. Biotechnol. *35*, 128–135.
36. Rowlands, C.F., Baralle, D., and Ellingford, J.M. (2019). Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing. Cells *8*, 8.
37. Buratti, E., Baralle, M., and Baralle, F.E. (2006). Defective splicing, disease and therapy: searching for master checkpoints in exon definition. Nucleic Acids Res. *34*, 3494–3510.
38. Pagani, F., and Baralle, F.E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. Nat. Rev. Genet. *5*, 389–396.
39. Balestra, D., Barbon, E., Scalet, D., Cavallari, N., Perrone, D., Zanibellato, S., Bernardi, F., and Pinotti, M. (2015). Regulation of a strong F9 cryptic 5'ss by intrinsic elements and by combination of tailored U1snRNAs with antisense oligonucleotides. Hum. Mol. Genet. *24*, 4809–4816.
40. Tajnik, M., Rogalska, M.E., Bussani, E., Barbon, E., Balestra, D., Pinotti, M., and Pagani, F. (2016). Molecular Basis and Therapeutic Strategies to Rescue Factor IX Variants That Affect Splicing and Protein Function. PLoS Genet. *12*, e1006082, e1006082.
41. Gyulkhandanyan, A., Rezaie, A.R., Roumenina, L., Lagarde, N., Fremeaux-Bacchi, V., Miteva, M.A., and Villoutreix, B.O.

(2020). Analysis of protein missense alterations by combining sequence- and structure-based methods. Mol. Genet. Genomic Med. *8*, e1166.

42. Liu, M., Murphy, M.E., and Thompson, A.R. (1998). A domain mutations in 65 haemophilia A families and molecular modelling of dysfunctional factor VIII proteins. Br. J. Haematol. *103*, 1051–1060.
43. Dal Mas, A., Rogalska, M.E., Bussani, E., and Pagani, F. (2015). Improvement of SMN2 pre-mRNA processing mediated by exon-specific U1 small nuclear RNA. Am. J. Hum. Genet. *96*, 93–103.
44. Anderson, S.L., Coli, R., Daly, I.W., Kichula, E.A., Rork, M.J., Volpi, S.A., Ekstein, J., and Rubin, B.Y. (2001). Familial dysautonomia is caused by mutations of the IKAP gene. Am. J. Hum. Genet. *68*, 753–758.
45. Lee, B., Rhead, W., Diaz, G.A., Scharschmidt, B.F., Mian, A., Shchelochkov, O., Marier, J.F., Beliveau, M., Mauney, J., Dickinson, K., et al. (2010). Phase 2 comparison of a novel ammonia scavenging agent with sodium phenylbutyrate in patients with urea cycle disorders: safety, pharmacokinetics and ammonia control. Mol. Genet. Metab. *100*, 221–228.
46. Zubarioglu, T., Dede, E., Cigdem, H., Kiykim, E., Cansever, M.S., and Aktuglu-Zeybek, C. (2020). Impact of sodium phenylbutyrate treatment in acute management of maple syrup urine disease attacks: a single-center experience. J. Pediatr. Endocrinol. Metab. *34*, 121–126.