Article

A Deep Exon Cryptic Splice Site Promotes Aberrant Intron Retention in a Von Willebrand Disease Patient

John G. Conboy

Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; jgconboy@lbl.gov

Abstract: A translationally silent single nucleotide mutation in exon 44 (E44) of the von Willebrand factor (VWF) gene is associated with inefficient removal of intron 44 in a von Willebrand disease (VWD) patient. This intron retention (IR) event was previously attributed to reordered E44 secondary structure that sequesters the normal splice donor site. We propose an alternative mechanism: the mutation introduces a cryptic splice donor site that interferes with the function of the annotated site to favor IR. We evaluated both models using minigene splicing reporters engineered to vary in secondary structure and/or cryptic splice site content. Analysis of splicing efficiency in transfected K562 cells suggested that the mutation-generated cryptic splice site in E44 was sufficient to induce substantial IR. Mutations predicted to vary secondary structure at the annotated site also had modest effects on IR and shifted the balance of residual splicing between the cryptic site and annotated site, supporting competition among the sites. Further studies demonstrated that introduction of cryptic splice donor motifs at other positions in E44 did not promote IR, indicating that interference with the annotated site is context dependent. We conclude that mutant deep exon splice sites can interfere with proper splicing by inducing IR.

Keywords: von Willebrand disease; intron retention; cryptic splice site; pre-mRNA splicing; aberrant splicing

1. Introduction

Aberrant splicing is a common cause of human genetic disease with various studies estimating that 15–60% of disease-causing mutations disrupt pre-mRNA splicing ([1,2]; reviewed in [3]). Moreover, 5–20% of cancer-predisposing mutations adversely affect splicing [4]. Mechanistically, mutations can create or destroy splice site motifs, alter enhancer or silencer elements that promote or inhibit recognition of splice sites, or alter the function of splicing regulatory proteins ([5–8] and references therein). Mutations can also alter RNA secondary structure to modulate accessibility of these motifs or the interactions of these motifs with relevant RNA-binding proteins [9–14]. Aberrant effects on the transcriptome can manifest as whole or partial exon skipping, exonification of intron sequences, or whole or partial intron retention. In some cases, even synonymous single nucleotide exonic mutations that do not impact known regulatory motifs can adversely affect pre-mRNA splicing patterns. Distinguishing the mechanisms by which such mutations disrupt splicing requires detailed studies that will ultimately improve therapeutic strategies.

Intron retention (IR) is a major form of alternative splicing that regulates expression of many genes during normal differentiation [15–17], cellular aging [18], and cellular response to physiological signals [19–21]. Aberrant IR, on the other hand, contributes to dysfunctional gene expression in many cancers [22,23] and in other diseases [24–30]. At the level of individual transcripts, mutations that promote aberrant intron retention associated with human disease have received relatively little attention. Recently, intron retention was reported in a patient with a bleeding disorder known as type I von Willebrand disease (VWD), caused by deficiency of von Willebrand factor (VWF) protein. A translationally silent C→T mutation in exon 44 (E44) of the VWF gene in that patient, which substituted

one glycine codon for another, is associated with altered splicing [31]. Although the mutation maps deep in the exon, 86 nt upstream of the splice donor site, it nevertheless acts from a distance to cause retention of downstream intron 44 [31]. Computer modeling suggested that the mutant E44 RNA adopts an aberrant secondary structure, sequestering the annotated 5′ splice site in a double-stranded configuration that renders intron 44 splicing inefficient [31]. Retained intron 44 sequences contain premature termination codons that would prevent synthesis of full-length protein from the mutant allele, consistent with the observed VWF deficiency phenotype. However, no experimental testing of the secondary structure model has been reported.

The current study examined an alternative model for intron retention in this VWF patient, based on the observation that the mutation generates a strong 5′ splice site motif deep within E44. The presence of cryptic splice site motifs near annotated exon/intron boundaries has been shown in several contexts to inhibit use of neighboring annotated sites, leading to intron retention [32–34]. Here, using a minigene splicing reporter that reproduces VWF intron 44 retention, we engineered numerous mutations designed to distinguish whether altered secondary structure or introduction of a new splice site motif is the primary determinant of intron retention. Collectively the data suggest not only that secondary structure can impact intron retention but also that the internal E44 splice site is a strong inducer of IR independent of secondary structure considerations.

2. Results

VWF E44 normally possesses a fairly weak 5′ splice site, AGT/gtaggt, which has a score of 3.31 according to a commonly used algorithm for estimating splice site strength [35]. Inspection of the E44 sequence revealed the presence of a very weak potential 5′ splice site motif (AAG/gcgagt; score = 2.72) deep in the exon. This motif would likely be nonfunctional in normal individuals and, indeed, there is no evidence for splicing at that site. However, the patient's C→T mutation created a GT dinucleotide that greatly strengthens this motif to a near-consensus splice site (AAG/gtgagt; score = 10.67). Weak 5′ splice sites at exon/intron borders are often susceptible to regulation, and there is precedence for the ability of cryptic exon splice site motifs to promote downstream intron retention [34]. We therefore hypothesized that the patient's mutation induces intron retention not only by altering RNA secondary structure but also via creation of a strong internal 5′ splice site.

To investigate the mechanism(s) responsible for IR in the VWF gene, we generated minigene splicing reporters with which splicing phenotype could be assessed following transfection into K562 erythroleukemia cells. SF3B1-mut35 is a splicing reporter used to study intron 4 retention in the SF3B1 gene [32]. This construct was modified to remove all of its natural IR-promoting elements (decoy exons), so that retention is observed only if heterologous IR-promoting elements are cloned into the intron.

Two series of VWF splicing reporters were engineered into the mut35 base construct. VWFwt-short was generated by inserting into mut35, at the site previously occupied by decoy exon 3e [32], a 314 nt region of normal VWF genomic sequence spanning exon 44 and short flanking intron sequences (Figure 1A). A variant reporter containing the patient's mutation, VWFm1-short, was altered at the indicated site. These reporters were designed to test whether intron retention might occur via the decoy exon mechanism, which would require only sequences proximal to E44. A second set of longer reporters was constructed to test whether IR might require the context of full-length intron 44 (Figure 1B). The long reporters contained 2.65 kb of VWF genomic sequence extending from the distal portion of intron 43 to a proximal region of intron 45.
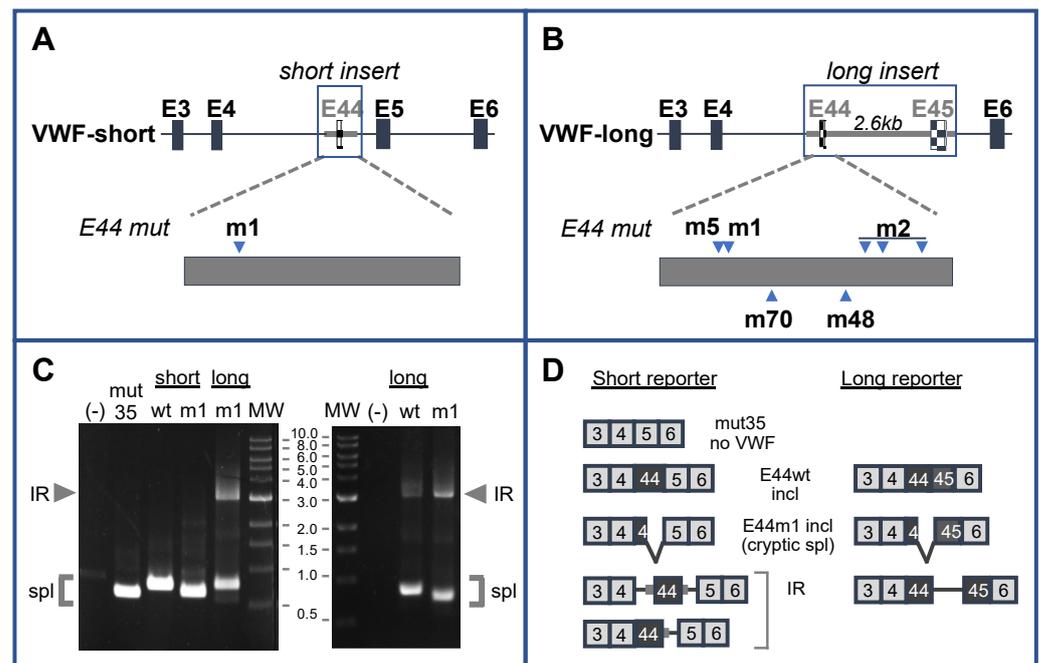
Figure 1. Engineering patient-specific retention of VWF intron 44 in minigene splicing reporters. (**A**) Structure of a short reporter. Boxed region represents VWF genomic sequence. The position of the patient's mutation (m1) is indicated. (**B**) Structure of a long reporter. Boxed region represents VWF genomic sequence. The positions of various engineered mutations are indicated. (**C**) Gel analysis of spliced products in K562 cells generated from a parent construct lacking VWF sequences (mut35); short reporters containing normal (wt) or patient (m1) sequences for VWF E44 plus short flanking intronic sequences as shown in (**A**); long constructs containing normal (wt) or patient sequences (m1) as diagrammed in (**B**). MW standards (in kb) are shown. (**D**) Expected spliced products from short (left) and long (right) reporters. Numbered boxes indicate the exons spliced together in each transcript. IR was detected in transcripts derived from long reporters with the m1 mutation; potential IR products generated from the short reporter were not observed.

Gel analysis of spliced products amplified from transfected K562 cells is shown in Figure 1C, and the structures of these products deduced after sequencing are depicted in Figure 1D. As expected, the base construct mut35 did not exhibit IR. Similarly, neither of the VWF short reporters, containing E44wt or E44m1 sequences, yielded any evidence for IR. A positive control containing the OGT decoy exon, processed in parallel and analyzed on a separate gel, did show substantial IR (data not shown). The absence of IR transcripts derived from these short VWF constructs suggests that E44 does not function as a decoy exon to promote IR. In contrast, the longer reporters exhibited low retention of VWF intron 44 in VWFwt, and substantially greater IR in the m1 variant (Figure 1C). Regarding the spliced products, VWFwt short produced a larger band than mut35 due to the inclusion of VWF E44. Interestingly, the spliced m1 transcript was slightly smaller than its wild-type counterpart due to the use of the m1 cryptic 5′ splice site in E44. These results not only confirmed the potential of m1 to be recognized as a functional splice site but also suggested that it can suppress splicing at the annotated site, perhaps due to reduced accessibility of the latter. Notably, splicing at the m1 splice site was not reported in the VWF patient, likely because the aberrant splice would have altered the translational reading frame to induce nonsense-mediated decay.

Having reproduced the VWF patient's IR phenotype, we then generated a series of splicing reporters to explore the mechanism(s) by which mutation m1 promotes IR. These reporters were designed to alter either of the features proposed to induce IR: (a) secondary structure at the annotated E44-5′ss that might impact its ability to interact with the splicing

machinery [31] and/or (b) aberrant internal 5′ splice site motif(s) that might compete with the annotated 5′ss for interaction with that machinery (e.g., [34]).

The predicted secondary structures for normal E44wt and patient E44m1 are shown in Figure 2. These were re-drawn from [31] and newly annotated to facilitate discussion of various secondary structure features. In normal E44, the weak internal 5′ss motif is located in the left arm of stem 1, with the annotated E44-5′ss predicted to be downstream of stem 2 in a relatively open (single-stranded) conformation (upper left). In the patient, reordered folding would incorporate the cryptic internal 5′ss, m1, into the right arm of new stem 3, with the annotated 5′ss sequestered in new stem 5 (upper right). Because the stability of these structures are not dramatically different [31], it seems plausible that the two might exist in some equilibrium within normal cells and that E44 mutations could alter their relative abundance to impact IR.
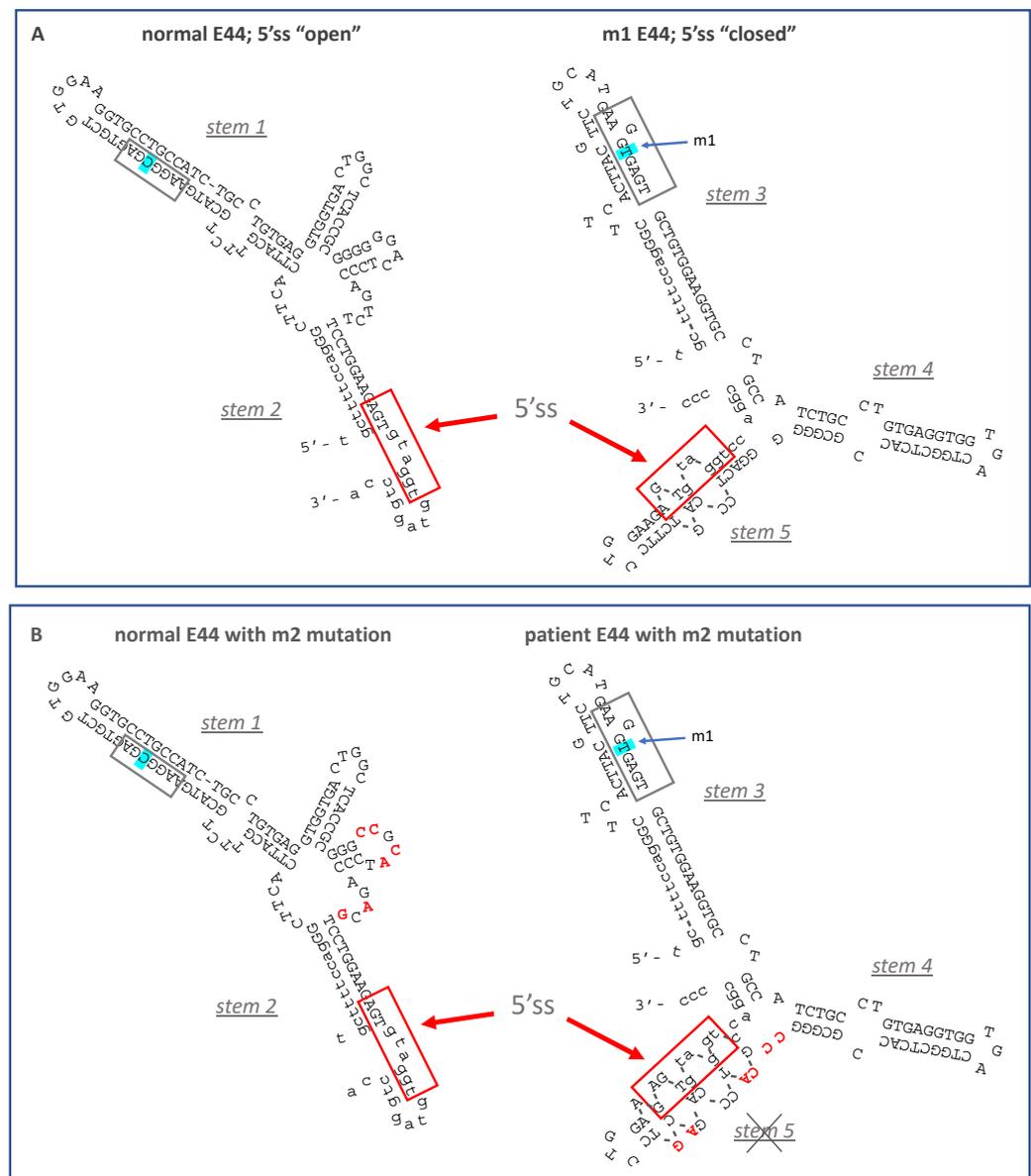

Figure 2. (**A**) Predicted structure of wild-type E44 ("open 5′ss") and patient E44 ("closed 5′ss"), redrawn from [31]. 5′ splice site motifs are boxed, and annotation of stem structures has been added for clarity. (**B**) Superimposition of m2 mutations (red text) on normal "open" and patient "closed" conformations for E44 shown above. M2 mutations are predicted to disrupt stem 5, increasing accessibility of the 5′ss.

Reporter VWFm2 was designed to test the hypothesis that secondary structure at the annotated site is a major driver of intron 44 retention. VWFm2 contains six nucleotide substitutions expected to disrupt stem 5, thereby reducing the potential for secondary structure at the annotated 5′ splice site (Figure 2B, right). According to this model, an open conformation should greatly reduce IR, whether or not cryptic site m1 is present. Analysis of splicing patterns in K562 cells transfected with VWFm2 and relevant control reporters yielded two important findings (Figure 3). First, although IR in VWFwt was quite modest, the proportion of retention product relative to spliced product generated from VWFm2 was even less (compare lanes wt and m2). Second, retention was greatly increased in double mutant VWFm1m2. This result indicates that the cryptic site can induce retention even when the annotated site is predicted to be accessible. Notably, although accessibility of the annotated site was not measured directly, that m2 favors a more open conformation is suggested by the finding that residual splicing switched back from the cryptic site in m1 (see Figure 1C) to the annotated site in m1m2. These results support the hypothesis that secondary structure at the annotated 5′ss of E44 does impact intron retention, but the presence of an internal cryptic 5′ss also strongly induces IR.
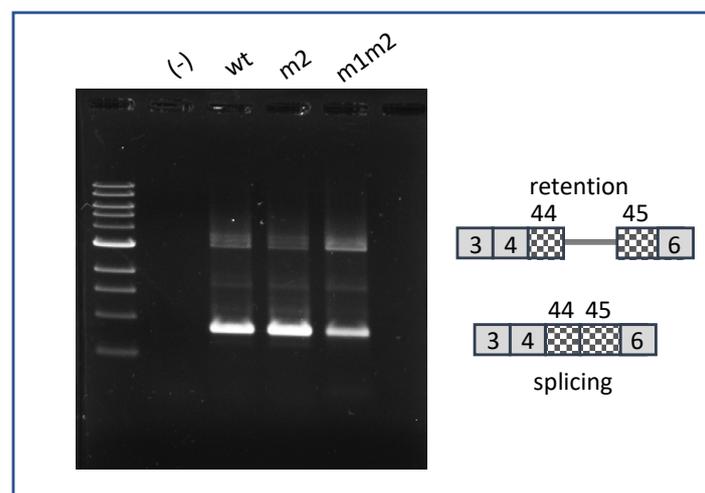

Figure 3. Effects of m2 mutations on IR. Lane (-), untransfected cells. Other lanes represent RT-PCR products from cells transfected with VWFwt, VWFm2, and VWFm1m2. Diagrams at **right** indicate the structures of the spliced products.

Next, we generated a construct aimed at testing the hypothesis that inaccessibility of the annotated site alone can promote IR. Mutation m5 was predicted to mimic m1-mediated stabilization of stem 3, given the identical ΔG value of the two structures (Figure 4, left). m5 was therefore expected to favor the closed conformation associated with sequestration of E44-5′ss. However, in contrast to m1, m5 occurs in the context of a very weak 5′ splice site motif (AAC/gcgagt; score = 1.78). Splicing analysis in K562 cells revealed that m5 yielded much less retention product relative to spliced product compared to m1, but slightly more retention product than VWFwt (Figure 4, right). This result is consistent with the interpretation that sequestration of E44-5′ss, in the absence of a competing 5′ splice site, increases intron retention only to a minor extent.

For completeness, we also tested the double mutant m1m5, which has features similar to m1, i.e., a strong 5′ splice site motif (slightly weaker than m1), and strong stabilization of stem 3. When tested in K562 cells, m1m5 exhibited strong IR (Figure 4, right), but it also switched back to the use of the annotated 5′ splice site for residual spliced products. This result is difficult to interpret in detail, since competition among splice sites is complex and may depend on poorly characterized accessory factors [36,37]. It is possible that the m1m5 cryptic site failed to compete with the annotated site due to the fact of its slightly weaker splice site motif (Figure 4, left), to its potential for greater sequestration in a hyperstabilized

stem 3, or both. Nevertheless, taken together, these results suggest that both RNA folding and splice site competition contribute to intron 44 retention but that splice site competition is likely the major determinant.
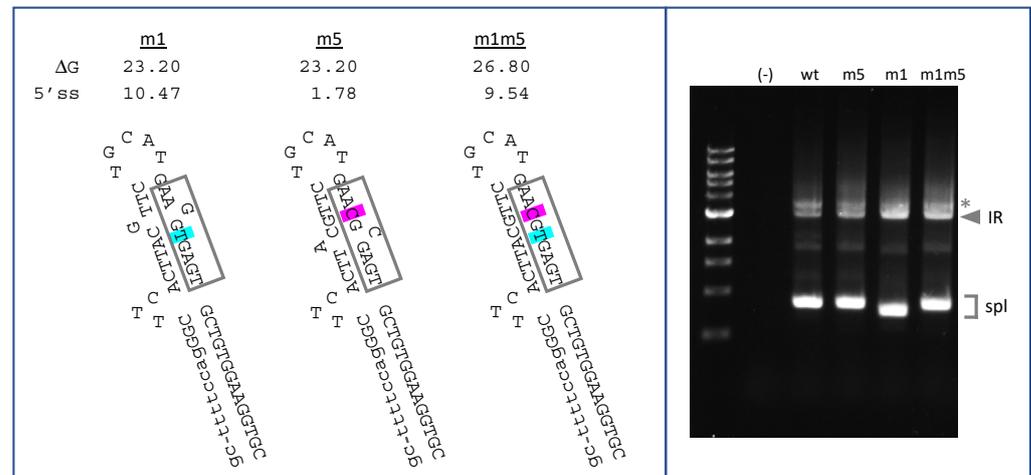

Figure 4. Effects of mutation m5 on IR predicted RNA structure and splice site strength (**left**) and splicing (**right**). For the gel analysis of splicing, lane (-) represents untransfected K562. Other lanes were transfected with VWFwt, VWFm5, VWFm1, and VWFm1m5 as indicated. * Indicates PCR artifact.

Finally, we asked whether the ability of a strong internal 5′ss motif to promote IR is position dependent. We generated splicing reporters containing strong 5′ splice site motifs, identical to that of m1 (AAG/gtgagt), at two different locations within E44. One was 48 nt upstream of the annotated E44-5′ss (m48) and the other was 70 nt upstream (m70). The splicing phenotypes of these pre-mRNAs are shown in Figure 5. As before, construct VWFm1 exhibited much higher IR than did VWFwt. However, neither VWFm48 nor VWF70 displayed IR above the wild-type background. That mutations m48 and m70 did generate functional 5′ splice sites was confirmed by sequence analysis showing that the E44 spliced products in these mutants primarily utilized the internal cryptic sites. These results suggest that an internal 5′ss motif must reside in an appropriate exonic context in order to interfere with splicing at the annotated site to yield intron retention.
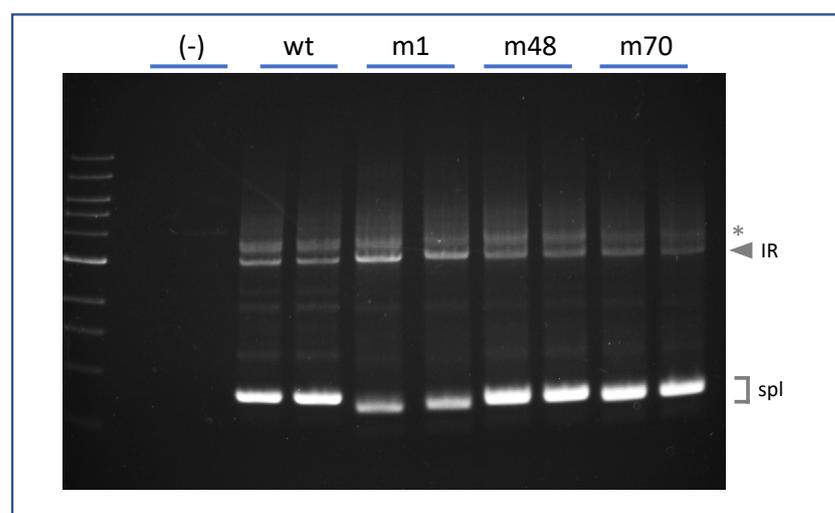

Figure 5. Position-dependence of IR stimulation by cryptic 5′ splice sites in E44. Analysis of splicing phenotype for E44 variants. Lanes (-) represent untransfected K562. Other lanes were transfected with VWFwt, VWFm1, VWFm48, and VWFm70. * Indicates PCR artifact.

3. Discussion

Von Willebrand disease is one of the most common inherited bleeding disorders, caused by any one of hundreds of mutations in the VWF gene [38,39]. The mutation studied here, although present at low frequency among the populations surveyed (dbSNP: rs900907976; minor allele frequency = 0.000026 (7/264690, TOPMED)), has an unusual impact on pre-mRNA splicing by acting at a distance to promote retention of the downstream intron [31]. Understanding how a "deep exon" mutation can promote IR may be important for the design of therapies for patients with this and other mechanistically related diseases.

Two major hypotheses have been proposed to explain how the m1 mutation interferes with function of the annotated 5′ splice site. First, the splice site sequestration model predicts that mutation m1 would alter long range RNA folding across exon 44 so as to increase base-pairing at the annotated 5′ splice site. According to this model, RNA–RNA interactions promote IR by sequestering that splice site away from the splicing machinery [31]. This model was originally based on computer modeling, but no experimental testing was reported [31]. Second, the splice site competition model posits that function of the annotated site is compromised by the presence of the cryptic 5′ splice site created by m1. Productive splicing in this model would be silenced by protein–protein interactions between complexes bound at the decoy site within the exon and at the annotated site at the normal exon/intron boundary. As discussed below, these models were experimentally tested by analysis of the behavior of splicing reporters altered at either of these features.

One instructive finding was the difference in splicing behavior between mutants m1 and m5. Both were predicted to stabilize stem 3 (Figure 4) to the same extent, favoring the conformation that sequesters the annotated 5′ splice site; however, they differ greatly with regard to strength of the internal splice site motif. Mutant m1 possesses a strong splice site motif and exhibits high IR. In contrast, m5 has a very weak motif and exhibits much lower IR. This experiment indicated that sequestration alone is insufficient to induce substantial IR. A second critical comparison involved constructs m1 and m1m2. These have identical strong deep exon splice site motifs, but different sequestration potentials at the annotated site. Most importantly, both exhibited strong IR, supporting the interpretation that the cryptic site is sufficient to induce IR in this context. Taken together, these results indicate that secondary structure does impact splice site choice and IR; however, IR occurs even with a more open configuration at the annotated site, as long as an upstream cryptic site is present in an appropriate sequence context as in m1m2.

A potential limitation of these interpretations is that secondary structure has not been probed directly. However, using splice site usage as a proxy for accessibility, we observed that residual splicing switched from the cryptic site in m1 to the annotated site in m1m2, consistent with a more accessible conformation of the annotated site in the double mutant. Interestingly, the double mutant m1m5 shifts residual splicing in the opposite direction back to the annotated 5′ splice site. Although folding models suggest that m1 and m1m5 have equivalent potential to sequester the annotated site, the strong cryptic site in m1m5 may be disfavored for residual splicing, either because it is slightly weaker than m1 or because m1m5 itself becomes partially sequestered in stem 3 (Figure 4). A caveat to these interpretations is that secondary structure has a complex role in alternative splicing [12], and the predicted stability of stem loop structures may not be as critical as specific structures in determining RNA functionality [40].

Another limitation to this study is that we have not directly measured spliceosomal binding to the cryptic site. For that reason, the possibility that a non-spliceosomal complex may bind at the cryptic site region to induce IR cannot be formally excluded. However, several observations support involvement of a spliceosomal component. First, cryptic site m1 does operate as a functional splice site as shown by sequence analysis of residual spliced products. Second, variant substrates bearing a strong cryptic splice site were associated with strong IR (m1, m1m2, m1m5), while those having a weak cryptic site were correlated with low IR (wt, m5). Finally, there is precedence for the idea that cryptic splice sites can interfere with annotated splice sites to promote IR [34].

In the Drosophila P element gene, binding of spliceosomal components to an exonic decoy splice site modulates retention of the downstream intron in a tissue-specific manner [34]. The splice site itself is necessary but not sufficient for regulation, with auxiliary RNA binding proteins being required to promote IR [41–43]. Binding of auxiliary protein factors can, in the presence of an exonic cryptic splice site, also inhibit productive splicing at the downstream authentic 5′ splice site to alter the balance of splice site choice [36,44,45]. In decoy exon-mediated intron retention, it has been proposed that splice sites of noncoding decoy exons interact nonproductively with intron-terminal splice sites to inhibit intron excision [32]. A common feature of decoy-mediated IR is the presence of multiple splice sites near the decoy exon/intron boundary [32,33], implying that competition among splice sites plays a role. Even more distally, deep intron splice site motifs, defined via their binding to U2AF65 in cross-linking experiments, appear capable of competing with recognition of splice sites of neighboring downstream exons [46].

Given these precedents, we propose that the splice site introduced by the VWF patient's mutation interacts in a nonproductive manner with the annotated 5′ splice site located 86 nt downstream. If the downstream intron 44 is recognized by an intron definition mechanism [47], then blocking its 5′ splice site could yield IR. Not well understood yet is what neighboring sequence features constitute a permissive environment for induction of IR. Regardless, it is interesting to speculate that blocking the cryptic splice site with antisense oligonucleotides might reactivate the annotated 5′ splice site to ameliorate the VWF protein deficiency.

More globally, relatively little is known about the incidence of IR due to the deep exon mutations and what might be their contribution to human disease. Methodological and conceptual considerations may have limited the identification of such mutations in previous exome and RNA-seq studies. Detecting association between deep exon SNVs and intron retention events requires reads of sufficient length to overlap both features, sequencing depth sufficient to identify allele-specific association of such features, and analytical strategies focused on finding these associations. Earlier work has, in fact, revealed that many exonic SNVs are associated with IR [23], but this analysis was focused on exon sequences near (within 30 nt of) exon–intron junctions. IR-associated exonic SNVs in that study mapped predominantly to the last nucleotide of the exon, and these likely act via direct interference with 5′ splice site recognition. A few IR-associated mutations were found outside of the splice site motif, but their mechanism of action was not explored other than to determine that they were not enriched in known splicing silencer or enhancer motifs [23]. Other studies have reported numerous deep exon mutations that create new 5′ splice site motifs and are associated with aberrant splicing [48]. Again, these were detected by finding short RNA-seq reads that contain mutations linked to abnormal (cryptic) exon/exon junctions; however, whether these mutations might also promote IR was not studied. Systematic analysis of deep exon mutations associated with IR might be enabled by acquisition of more long RNA-seq reads that overlap deep exon sequences and exon/intron boundaries. All of these analyses would likely benefit from application of algorithms that optimize detection of retained introns from short RNA-seq data [49–51] and long RNA-seq data [52]. Finally, mutations that create novel 5′ splice site motifs and are associated with IR can also occur in downstream intron sequences [53].

4. Materials and Methods

Construction of splicing reporters: Reporters diagrammed in Figure 1 were constructed as follows by modification of an SF3B1-based minigene, mut35, in which intron 4 retention was suppressed due to the removal of all decoy exons [32]. The first-generation reporter (short) was made by amplifying a 314 nt region of VWF gene spanning exon 44 and proximal intron sequences using the following primers:

forward primer:	5′-ACAGCTTGCAACAGGGATGAAAATGCCCAGACCAGTG-3′
reverse primer:	5′-GCACAGTTTAGAGTCGATGAAACCAAGGTCAACGCTG-3′

The lower case regions provided overlap sequences to facilitate fusion [54] of this fragment into mut35 at the normal position of decoy exon 3e [32].

The second wild-type reporter (long) encompassed a 2.65 kb region of the VWF gene spanning across a terminal portion of intron 43, exon 44, intron 44, exon 45, and a proximal region of intron 45.

forward primer: 5′-GACCTTGGTTTCATCTCATTCCCTGCCCTTTG-3′
reverse primer: 5′-ACAATGGAGAACCATAGCGTTGGGGTGGCAAG-3′

The fragment amplified with these primers was cloned into VWFwt short using In-Fusion methods [54] with 15 nt in lowercase sequence representing overlap with the ends of the linearized vector. To accommodate this larger insert, VWFshort was amplified using primers that omitted ~1.5 kb of SF3B1 sequence (spanning the distal 0.2 kb region of intron 4 to the proximal 1.3 kb region of intron 5).

Mutations in VWF E44 were introduced by amplifying the wild-type VWF reporter with primers carrying the desired mutations. The various 5′ splice site motifs (boxed) and their immediate flanking sequences are shown below. Construct m1 displays the patient's sequence. Construct m5 was mutated to stabilize stem 3 equivalently to m1 but without generating a strong 5′ splice site motif (see Figure 4). m1m5 contained both the m1 and m5 mutations. Constructs m48 and m70 represent mutations at different positions in E44 that introduce strong 5′ splice site motifs identical to that of m1. Slashes (/) indicate the site of splicing. Shaded nucleotides differ from the wild-type sequence. The m1 splice site from the VWF patient was 86 nt upstream from the annotated site, while the m48 and m70 were 48 nt and 70 nt, respectively, upstream of the annotated site.

m1   CATG AAG/GTGAGT GCTGTGGAAGGTGCCTG
m5   CATG AAC/GCGAGT GCTGTGGAAGGTGCCTG
m1m5   CATG AAC/GTGAGT GCTGTGGAAGGTGCCTG
m48   CTGT AAG/GTGAGT ACTGGCTCACCGCGGGG
m70   GTGG AAG/GTGAGT GCCATCTGCCTGTGAGG

Construct m2 contained 6 nucleotide changes near the annotated 5′ splice site, designed to disrupt secondary structure in this region. The linear sequence of the annotated 5′ splice site region for the wild-type ("closed conformation") and m2 ('open conformation') constructs is given below. Predicted secondary structures are shown in Figure 2.

wt   GGGGGGACTCCCAGTCTTCCTGGAAGAGT/gtaggt
m2   GGGCCGCATCCCAGACGTCCTGGAAGAGT/gtaggt

Analysis of splicing: Splicing reporters were transfected into K562 using FuGENE according to the manufacturer's instructions (Promega; Madison, WI, USA). RNA was isolated after 48 h using RNeasy columns as per instructions from the manufacturer (Qiagen; Germantown, MD, USA) with the inclusion of a DNase step to minimize contamination by genomic DNA. RNA was reverse transcribed with Superscript III (Invitrogen) into cDNA using the BGH reverse primer in the vector (5′-tagaaggcacagtcgagg-3′). Spliced products were amplified using a forward primer in exon 3 (5′-catcatctacgagtttgcttgg-3′) and a reverse primer in the vector (5′-atttaggtgacactatagaatagggc-3′). This strategy amplified minigene-derived transcripts but not endogenous mRNA as confirmed using RNA from untransfected or empty vector-transfected cells. When assaying IR products, PCR reaction conditions were adjusted to allow for amplification of DNA bands ≥ 3 kb in length (denaturation at 95 °C for 20 s, annealing at 56 °C for 10 s, extension at 70 °C for 2 min 30 s; 35 cycles) using KOD polymerase in the presence of betaine. PCR products were analyzed on either 2% agarose gels or 4.5% acrylamide gels. Identity of the PCR products was confirmed by DNA sequencing, and all splicing reporters were assayed at least three

times. The splicing phenotype of test constructs relative to control constructs was highly reproducible, despite inevitable variations in baseline intron retention.

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

References

1. Krawczak, M.; Reiss, J.; Cooper, D.N. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Hum. Genet.* **1992**, *90*, 41–54. [CrossRef] [PubMed]
2. Lopez-Bigas, N.; Audit, B.; Ouzounis, C.; Parra, G.; Guigo, R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* **2005**, *579*, 1900–1903. [CrossRef]
3. Lord, J.; Baralle, D. Splicing in the Diagnosis of Rare Disease: Advances and Challenges. *Front. Genet.* **2021**, *12*, 689892. [CrossRef] [PubMed]
4. Casadei, S.; Gulsuner, S.; Shirts, B.H.; Mandell, J.B.; Kortbawi, H.M.; Norquist, B.S.; Swisher, E.M.; Lee, M.K.; Goldberg, Y.; O'Connor, R.; et al. Characterization of splice-altering mutations in inherited predisposition to cancer. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 26798–26807. [CrossRef]
5. Cartegni, L.; Chew, S.L.; Krainer, A.R. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* **2002**, *3*, 285–298. [CrossRef] [PubMed]
6. Lee, Y.; Rio, D.C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* **2015**, *84*, 291–323. [CrossRef]
7. Hershberger, C.E.; Moyer, D.C.; Adema, V.; Kerr, C.M.; Walter, W.; Hutter, S.; Meggendorfer, M.; Baer, C.; Kern, W.; Nadarajah, N.; et al. Complex landscape of alternative splicing in myeloid neoplasms. *Leukemia* **2021**, *35*, 1108–1120. [CrossRef]
8. Bonnal, S.C.; Lopez-Oreja, I.; Valcarcel, J. Roles and mechanisms of alternative splicing in cancer—Implications for care. *Nat. Rev. Clin. Oncol.* **2020**, *17*, 457–474. [CrossRef]
9. Shepard, P.J.; Hertel, K.J. Conserved RNA secondary structures promote alternative splicing. *RNA* **2008**, *14*, 1463–1469. [CrossRef] [PubMed]
10. Grover, A.; Houlden, H.; Baker, M.; Adamson, J.; Lewis, J.; Prihar, G.; Pickering-Brown, S.; Duff, K.; Hutton, M. 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J. Biol. Chem.* **1999**, *274*, 15134–15143. [CrossRef]
11. Varani, L.; Hasegawa, M.; Spillantini, M.G.; Smith, M.J.; Murrell, J.R.; Ghetti, B.; Klug, A.; Goedert, M.; Varani, G. Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8229–8234. [CrossRef] [PubMed]
12. Bartys, N.; Kierzek, R.; Lisowiec-Wachnicka, J. The regulation properties of RNA secondary structure in alternative splicing. *Biochim. Biophys. Acta. Gene. Regul. Mech.* **2019**, *1862*, 194401. [CrossRef] [PubMed]
13. Shilo, A.; Tosto, F.A.; Rausch, J.W.; Le Grice, S.F.J.; Misteli, T. Interplay of primary sequence, position and secondary RNA structure determines alternative splicing of LMNA in a pre-mature aging syndrome. *Nucleic Acids Res.* **2019**, *47*, 5922–5935. [CrossRef]
14. Taliaferro, J.M.; Lambert, N.J.; Sudmant, P.H.; Dominguez, D.; Merkin, J.J.; Alexis, M.S.; Bazile, C.; Burge, C.B. RNA Sequence Context Effects Measured In Vitro Predict In Vivo Protein Binding and Regulation. *Mol. Cell* **2016**, *64*, 294–306. [CrossRef]
15. Jacob, A.G.; Smith, C.W.J. Intron retention as a component of regulated gene expression programs. *Hum. Genet.* **2017**, *136*, 1043–1057. [CrossRef] [PubMed]
16. Monteuuis, G.; Wong, J.J.L.; Bailey, C.G.; Schmitz, U.; Rasko, J.E.J. The changing paradigm of intron retention: Regulation, ramifications and recipes. *Nucleic Acids Res.* **2019**, *47*, 11497–11513. [CrossRef]
17. Grabski, D.F.; Broseus, L.; Kumari, B.; Rekosh, D.; Hammarskjold, M.L.; Ritchie, W. Intron retention and its impact on gene expression and protein diversity: A review and a practical guide. *Wiley Interdiscip. Rev. RNA* **2021**, *12*, e1631. [CrossRef]
18. Yao, J.; Ding, D.; Li, X.; Shen, T.; Fu, H.; Zhong, H.; Wei, G.; Ni, T. Prevalent intron retention fine-tunes gene expression and contributes to cellular senescence. *Aging Cell* **2020**, *19*, e13276. [CrossRef]
19. Boothby, T.C.; Zipper, R.S.; van der Weele, C.M.; Wolniak, S.M. Removal of retained introns regulates translation in the rapidly developing gametophyte of Marsilea vestita. *Dev. Cell* **2013**, *24*, 517–529. [CrossRef]

20. Mauger, O.; Scheiffele, P. Beyond proteome diversity: Alternative splicing as a regulator of neuronal transcript dynamics. *Curr. Opin. Neurobiol.* **2017**, *45*, 162–168. [CrossRef]
21. Ninomiya, K.; Kataoka, N.; Hagiwara, M. Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor. *J. Cell Biol.* **2011**, *195*, 27–40. [CrossRef] [PubMed]
22. Dvinge, H.; Bradley, R.K. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* **2015**, *7*, 45. [CrossRef]
23. Jung, H.; Lee, D.; Lee, J.; Park, D.; Kim, Y.J.; Park, W.-Y.; Hong, D.; Park, P.J.; Lee, E. Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* **2015**, *47*, 1242–1248. [CrossRef] [PubMed]
24. Sznajder, L.J.; Thomas, J.D.; Carrell, E.M.; Reid, T.; McFarland, K.N.; Cleary, J.D.; Oliveira, R.; Nutter, C.A.; Bhatt, K.; Sobczak, K.; et al. Intron retention induced by microsatellite expansions as a disease biomarker. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 4234–4239. [CrossRef]
25. Inoue, D.; Polaski, J.T.; Taylor, J.; Castel, P.; Chen, S.; Kobayashi, S.; Hogg, S.J.; Hayashi, Y.; Pineda, J.M.B.; El Marabti, E.; et al. Minor intron retention drives clonal hematopoietic disorders and diverse cancer predisposition. *Nat. Genet.* **2021**, *53*, 707–718. [CrossRef]
26. Wang, Q.; Conlon, E.G.; Manley, J.L.; Rio, D.C. Widespread intron retention impairs protein homeostasis in C9orf72 ALS brains. *Genome Res.* **2020**, *30*, 1705–1715. [CrossRef] [PubMed]
27. Li, H.D.; Funk, C.C.; McFarland, K.; Dammer, E.B.; Allen, M.; Carrasquillo, M.M.; Levites, Y.; Chakrabarty, P.; Burgess, J.D.; Wang, X.; et al. Integrative functional genomic analysis of intron retention in human and mouse brain with Alzheimer's disease. *Alzheimers Dement.* **2021**, *17*, 984–1004. [CrossRef] [PubMed]
28. Luisier, R.; Tyzack, G.E.; Hall, C.E.; Mitchell, J.S.; Devine, H.; Taha, D.M.; Malik, B.; Meyer, I.; Greensmith, L.; Newcombe, J.; et al. Intron retention and nuclear loss of SFPQ are molecular hallmarks of ALS. *Nat. Commun.* **2018**, *9*, 2010. [CrossRef] [PubMed]
29. Jangi, M.; Sharp, P.A. Building robust transcriptomes with master splicing factors. *Cell* **2014**, *159*, 487–498. [CrossRef]
30. Humphrey, J.; Birsa, N.; Milioto, C.; McLaughlin, M.; Ule, A.M.; Robaldo, D.; Eberle, A.B.; Krauchi, R.; Bentham, M.; Brown, A.L.; et al. FUS ALS-causative mutations impair FUS autoregulation and splicing factor networks through intron retention. *Nucleic Acids Res.* **2020**, *48*, 6889–6905. [CrossRef] [PubMed]
31. Yadegari, H.; Biswas, A.; Akhter, M.S.; Driesen, J.; Ivaskevicius, V.; Marquardt, N.; Oldenburg, J. Intron retention resulting from a silent mutation in the VWF gene that structurally influences the 5′ splice site. *Blood* **2016**, *128*, 2144–2152. [CrossRef]
32. Parra, M.; Booth, B.W.; Weiszmann, R.; Yee, B.; Yeo, G.W.; Brown, J.B.; Celniker, S.E.; Conboy, J.G. An important class of intron retention events in human erythroblasts is regulated by cryptic exons proposed to function as splicing decoys. *RNA* **2018**, *24*, 1255–1265. [CrossRef] [PubMed]
33. Pirnie, S.P.; Osman, A.; Zhu, Y.; Carmichael, G.G. An Ultraconserved Element (UCE) controls homeostatic splicing of ARGLU1 mRNA. *Nucleic Acids Res.* **2017**, *45*, 3473–3486. [CrossRef] [PubMed]
34. Siebel, C.W.; Fresco, L.D.; Rio, D.C. The mechanism of somatic inhibition of Drosophila P-element pre-mRNA splicing: Multiprotein complexes at an exon pseudo-5′ splice site control U1 snRNP binding. *Genes. Dev.* **1992**, *6*, 1386–1401. [CrossRef] [PubMed]
35. Yeo, G.; Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **2004**, *11*, 377–394. [CrossRef]
36. Yu, Y.; Maroney, P.A.; Denker, J.A.; Zhang, X.H.; Dybkov, O.; Luhrmann, R.; Jankowsky, E.; Chasin, L.A.; Nilsen, T.W. Dynamic regulation of alternative splicing by silencers that modulate 5′ splice site competition. *Cell* **2008**, *135*, 1224–1236. [CrossRef] [PubMed]
37. Roca, X.; Krainer, A.R.; Eperon, I.C. Pick one, but be quick: 5′ splice sites and the problems of too many choices. *Genes. Dev.* **2013**, *27*, 129–144. [CrossRef] [PubMed]
38. Borras, N.; Orriols, G.; Batlle, J.; Perez-Rodriguez, A.; Fidalgo, T.; Martinho, P.; Lopez-Fernandez, M.F.; Rodriguez-Trillo, A.; Loures, E.; Parra, R.; et al. Unraveling the effect of silent, intronic and missense mutations on VWF splicing: Contribution of next generation sequencing in the study of mRNA. *Haematologica* **2019**, *104*, 587–598. [CrossRef]
39. de Jong, A.; Eikenboom, J. Von Willebrand disease mutation spectrum and associated mutation mechanisms. *Thromb. Res.* **2017**, *159*, 65–75. [CrossRef]
40. Bao, C.; Zhu, M.; Nykonchuk, I.; Wakabayashi, H.; Mathews, D.H.; Ermolenko, D.N. Specific length and structure rather than high thermodynamic stability enable regulatory mRNA stem-loops to pause translation. *bioRxiv* **2021**. [CrossRef]
41. Horan, L.; Yasuhara, J.C.; Kohlstaedt, L.A.; Rio, D.C. Biochemical identification of new proteins involved in splicing repression at the Drosophila P-element exonic splicing silencer. *Genes. Dev.* **2015**, *29*, 2298–2311. [CrossRef] [PubMed]
42. Labourier, E.; Adams, M.D.; Rio, D.C. Modulation of P-element pre-mRNA splicing by a direct interaction between PSI and U1 snRNP 70K protein. *Mol. Cell* **2001**, *8*, 363–373. [CrossRef]
43. Siebel, C.W.; Kanaar, R.; Rio, D.C. Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes Dev.* **1994**, *8*, 1713–1725. [CrossRef]
44. Nilsen, T.W.; Graveley, B.R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **2010**, *463*, 457–463. [CrossRef] [PubMed]
45. Ohe, K.; Miyajima, S.; Tanaka, T.; Hamaguchi, Y.; Harada, Y.; Horita, Y.; Beppu, Y.; Ito, F.; Yamasaki, T.; Terai, H.; et al. HMGA1a Induces Alternative Splicing of the Estrogen Receptor-alphalpha Gene by Trapping U1 snRNP to an Upstream Pseudo-5′ Splice Site. *Front. Mol. Biosci.* **2018**, *5*, 52. [CrossRef] [PubMed]

46. Shao, C.; Yang, B.; Wu, T.; Huang, J.; Tang, P.; Zhou, Y.; Zhou, J.; Qiu, J.; Jiang, L.; Li, H.; et al. Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. *Nat. Struct. Mol. Biol.* **2014**, *21*, 997–1005. [CrossRef]

47. De Conti, L.; Baralle, M.; Buratti, E. Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* **2013**, *4*, 49–60. [CrossRef] [PubMed]

48. Shiraishi, Y.; Kataoka, K.; Chiba, K.; Okada, A.; Kogure, Y.; Tanaka, H.; Ogawa, S.; Miyano, S. A comprehensive characterization of cis-acting splicing-associated variants in human cancer. *Genome Res.* **2018**, *28*, 1111–1125. [CrossRef]

49. Pimentel, H.; Conboy, J.G.; Pacher, L. Keep Me Around: Intron Retention Detection and Analysis. *arXiv* **2015**, arXiv:1510.00696.

50. Middleton, R.; Gao, D.; Thomas, A.; Singh, B.; Au, A.; Wong, J.J.; Bomane, A.; Cosson, B.; Eyras, E.; Rasko, J.E.; et al. IRFinder: Assessing the impact of intron retention on mammalian gene expression. *Genome Biol.* **2017**, *18*, 51. [CrossRef]

51. Ni, T.; Yang, W.; Han, M.; Zhang, Y.; Shen, T.; Nie, H.; Zhou, Z.; Dai, Y.; Yang, Y.; Liu, P.; et al. Global intron retention mediated gene regulation during CD4+ T cell activation. *Nucleic Acids Res.* **2016**, *44*, 6817–6829. [CrossRef] [PubMed]

52. Tang, A.D.; Soulette, C.M.; van Baren, M.J.; Hart, K.; Hrabeta-Robinson, E.; Wu, C.J.; Brooks, A.N. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat. Commun.* **2020**, *11*, 1438. [CrossRef] [PubMed]

53. Varela, P.; Caldas, M.M.; Pesquero, J.B. Novel GLA Mutation Promotes Intron Inclusion Leading to Fabry Disease. *Front. Genet.* **2019**, *10*, 783. [CrossRef]

54. Gibson, D.G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol.* **2011**, *498*, 349–361. [PubMed]