



OPEN

Functional assays of non-canonical splice-site variants in inherited retinal dystrophies genes

Ana Rodríguez-Muñoz^{1,2,7}, Alessandro Liquori^{3,4,7}, Belén García-Bohorquez^{1,2,5}, Teresa Jaijo^{1,2,5,6}, Elena Aller^{1,2,5,6}, José M. Millán^{1,2,5,8}✉ & Gema García-García^{1,2,5,8}

Inherited retinal dystrophies are a group of disorders characterized by the progressive degeneration of photoreceptors leading to loss of the visual function and eventually to legal blindness. Although next generation sequencing (NGS) has revolutionized the molecular diagnosis of these diseases, the pathogenicity of some mutations casts doubts. After the screening of 208 patients with a panel of 117 genes, we obtained 383 variants that were analysed in silico with bioinformatic prediction programs. Based on the results of these tools, we selected 15 variants for their functional assessment. Therefore, we carried out minigene assays to unveil whether they could affect the splicing of the corresponding gene. As a whole, seven variants were found to induce aberrant splicing in the following genes: *BEST1*, *CACNA2D4*, *PRCD*, *RIMS1*, *FSCN2*, *MERTK* and *MAK*. This study shows the efficacy of a workflow, based on the association of the Minimum Allele Frequency, family co-segregation, in silico predictions and in vitro assays to determine the effect of potential splice site variants identified by DNA-based NGS. These findings improve the molecular diagnosis of inherited retinal dystrophies and will allow some patients to benefit from the upcoming gene-based therapeutic strategies.

Inherited retinal dystrophies (IRD) encompass a clinically and genetically heterogeneous group of disorders characterized by the progressive degeneration of rod and cone photoreceptors, leading to loss of the visual function and legal blindness. Clinical hallmarks of IRD include blurry vision, photophobia, colour vision abnormalities, visual field loss and nyctalopia¹. To date, 271 genes related to IRD have been identified (<https://sph.uth.edu/retnet/sum-dis.htm#A-genes>, accessed August 2021).

The identification of the disease-causing mutations is crucial for patient genetic counselling and to benefit from clinical trials or new treatments. The next-generation sequencing (NGS) has revolutionized the molecular diagnosis of these genetically heterogeneous diseases and the percentage of genetically diagnosed patients has increased dramatically^{2–4}. On the other hand, NGS unveils a high number of variants per patient, many of which have a low allele frequency in the population and, therefore, could be involved in rare diseases. In silico predictions are crucial to evaluate the pathogenic effects of these variants, including missense, silent and putative splicing changes. Nevertheless, in most of the cases, the clinical significance of them remains unknown.

In agreement with the American College of Medical Genetics and Genomics (ACMG) standards, functional studies are a powerful tool in support of pathogenicity assessment⁵. In contrast to protein function assays, characterization of potential splicing variants is more feasible. However, most of the genes involved in IRD are only expressed in non-accessible tissues and/or mRNA from patients is often unavailable. In these cases, in vitro minigene splicing assays are a good alternative to assess the impact of variants at the mRNA level⁶.

In recent years it has been shown that the use of antisense oligonucleotides (AONs) is a promising therapeutic approach to correct aberrant splicing. AONs are short single stranded molecules designed to target and modulate (pre-)mRNA splicing. AONs can be designed to bind directly splice sites or cis-regulatory elements in order to produce an exon-skipping or specifically mask altered splice-sites involved in aberrant splicing. This strategy has been shown to be useful as a therapy for different diseases, including retinal dystrophies^{7–10}. Of

¹Molecular, Cellular and Genomics Biomedicine Research Group, Instituto de Investigación Sanitaria La Fe (IIS La Fe), Valencia, Spain. ²Unidad Mixta de Enfermedades Raras La Fe-Centro de Investigación Príncipe Felipe, Valencia, Spain. ³Hematology Research Group, Instituto de Investigación Sanitaria La Fe, Valencia, Spain. ⁴Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain. ⁵Biomedical Research Network for Rare Diseases, CIBERER, Madrid, Spain. ⁶Genetics Unit, Hospital Universitario y Politécnico La Fe, Valencia, Spain. ⁷These authors contributed equally: Ana Rodríguez-Muñoz and Alessandro Liquori. ⁸These authors jointly supervised this work: José M. Millán and Gema García-García. ✉email: millan_jos@gva.es

Gene	Mutation	SS	Position	HSF			MaxEnt			SpliceAI ^a
				Wt	Mut	Variation (%)	Wt	Mut	Variation (%)	
<i>ABCA4</i>	c.2972G>T	D	E, 54/132	64.62	91.45	+41.52	0.41	8.05	+2063.41	DG: 0.09 (2)
<i>ABCA4</i>	c.6148G>C	A	E, 1/135	87.36	83.20	-4.76	12.46	9.41	-24.48	AL: 0.00 (0)
<i>BEST1</i>	c.637_639del	A	E, 1_3/78	93.39	91.03	-2.53	-	-	-	AL: 1.00 (3)
				75.49	91.03	+20.59	11.00	11.68	+6.18	AG: 0.93 (6)
<i>CACNA2D4</i>	c.2153-12_2155del	A	I, -12_-1	93.15	57.22	-38.57	8.53	-2.70	-131.65	-
			E, 1_3/94							
<i>CNNM4</i>	c.2039C>T	D	E, 91/182	55.92	82.75	+47.98	1.89	5.86	+409.52	DG: 0.01 (-2)
<i>FSCN2</i>	c.1105G>A	D	E, 122/122	-	76.95	-	-	9.04	-	DG: 0.80 (20)
				89.94	79.36	-11.76	9.10	4.04	-55.60	DL:0.47 (0)
<i>IDH3B</i>	c.532A>G	A	E, 1/134	88.49	91.62	+3.54	6.89	7.89	+14.51	AG: 0.01 (-8)
<i>MAK</i>	c.755A>G	A	E, 92/168	59.74	88.69	+48.46	-3.58	5.16	+244.13	AG: 0.68 (-1)
				81.74	81.69	-0.06	3.75	-2.2	-158.67	AL: 0.34 (-7)
<i>MERTK</i>	c.1450G>A	D	E, 154/154	-	82.74	-	-	5.74	-	DG: 0.43 (16)
				-	80.63	-	-	3.44	-	
				90.80	80.22	-11.65	8.94	3.45	-61.41	DL: 0.67 (0)
<i>PRCD</i>	c.74+5G>C	D	I, +5	79.79	67.78	-15.05	4.27	-3.52	-182.44	DL: 0.96 (-5)
<i>PRPF8</i>	c.434+3G>A	D	I, +3	85.58	86.23	+0.76	4.74	7.83	+65.19	DG: 0.00 (3)
<i>PRPF31</i>	c.182C>G	D	E, 5/61	62.94	89.77	+42.63	-1.98	6.29	+417.68	DG: 0.00 (-1)
<i>RHO</i>	c.316G>A	A	E, 316/361	64.15	93.10	+45.13	-1.61	6.34	+493.79	AG: 0.00 (2)
<i>RIMS1</i>	c.2544+4A>G	D	I, +4	82.83	74.49	-10.07	4.44	-1.23	-127.70	DL: 0.29 (-4)
<i>RPGRIP1</i>	c.930+3A>G	D	I, +3	96.23	85.39	-11.26	8.40	4.72	-43.81	-

Table 1. Results of the prediction programmes for the selected variants. A acceptor, AG acceptor gain, AL acceptor loss, D donor, DG donor gain, DL donor loss, E exon, I intron, Mut mutant, SS splice-site, Wt wildtype, *ABCA4*: NM_000350.2, *BEST1*: NM_004183.4, *CACNA2D4*: NM_172364.4, *CNNM4*: NM_020184.3, *FSCN2*: NM_012418.4, *IDH3B*: NM_001258384.2, *MAK*: NM_005906.5, *MERTK*: NM_006343.2, *PRCD*: NM_001077620.3, *PRPF8*: NM_006445.4, *PRPF31*: NM_015629.4, *RHO*: NM_000539.3, *RIMS1*: NM_014989.6, *RPGRIP1*: NM_020366.4 ^aDistance in bp. -Gene not available in the consulted website.

note, AONs has been used to correct the aberrant pre-mRNA splicing produced by deep-intronic mutations in *USH2A* or *CEP290*¹⁰⁻¹³.

In our cohort of 208 IRD patients, we identified 15 putative non-canonical splicing mutations as a result of the sequencing of 117 genes. The aim of this study was to predict the pathogenic nature of these variants and to assess their effect on the splicing process by minigene assays.

Results

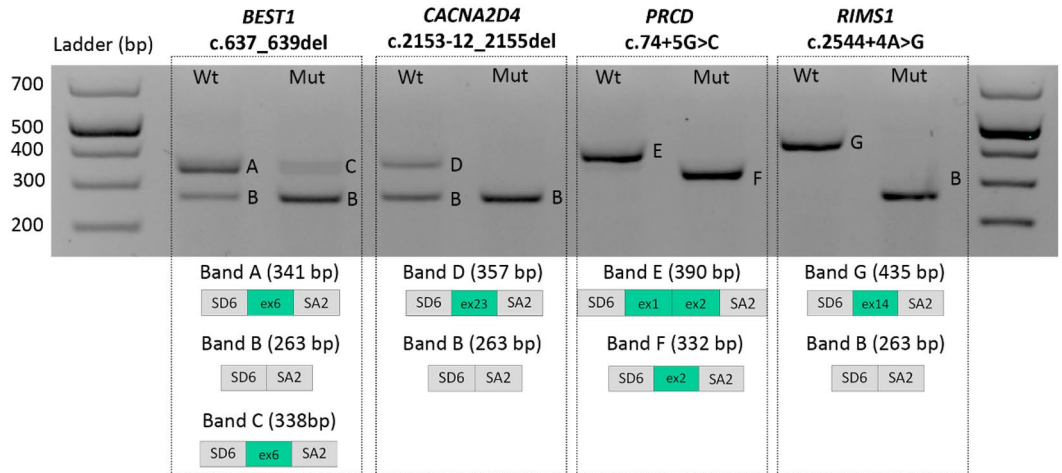
Genetic variants. Once variants with a MAF > 0.01 were discarded, 383 intronic and exonic variants identified in the IRD genes were studied with bioinformatic tools (HSF and MaxEnt) to analyze their effect on splicing. Variants were selected for further functional assays according to the criteria mentioned in Material and Methods section and independently from their involvement on the phenotype of the individual (i.e., any variant found in a IRD gene with a putative splicing effect). Table 1 shows the results obtained for the HSF, MaxEnt and SpliceAI predictors for each of the selected variants.

Minigene assays. Minigene assays were performed on 15 variants found in 15 patients, respectively. Results of this analysis are presented in Fig. 1 and Supplementary Fig. 1. Overall, seven variants were found to have an aberrant splicing compared with wildtype minigene: four by the loss of the natural acceptor or donor splice site (*BEST1* c.637_639del, *CACNA2D4* c.2153-12_2155del, *PRCD* c.74+5G>C, *RIMS1* c.2544+4A>G, Fig. 1A); two by the activation of a cryptic donor splice site (*FSCN2* c.1105G>A, *MERTK* c.1450G>A, Fig. 1B); and one by the creation of a new acceptor splice site (*MAK* c.755A>G, Fig. 1C). Normal splicing patterns were observed in mutated minigenes of the remaining eight cases (Supplementary Fig. 1).

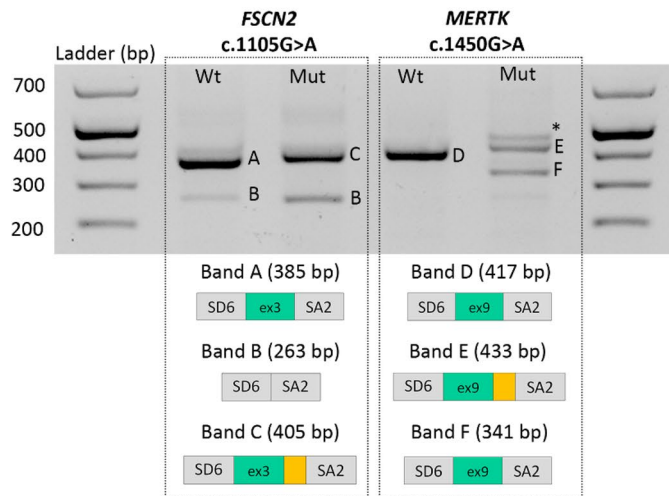
Among the four variants altering the acceptor splice sites (3'ss) recognition (Table 1), two were deletions (*BEST1* c.637_639del, *CACNA2D4* c.2153-12_2155del), and two substitutions occur in the first position of the exon (*ABCA4* c.6148G>C, *IDH3B* c.532A>G).

The *BEST1* c.637_639del mutation involves the deletion of the first three nucleotides of exon 6. The results of the SpliceAI predictor for this mutation indicate that two events would occur: the loss of the natural acceptor splice site and the activation of a cryptic acceptor splice site. These predictions were confirmed by minigene findings. In contrast, the HSF and MaxEnt algorithms did not correctly predict these two events. Taking into account the result of minigene assay, this deletion in exon 6 would not alter the reading pattern, likely resulting in the mutated protein p.(Glu212_Gln238del). However, the mutated construct revealed a major aberrant transcript corresponding to the exon 6 skipping (band B) and a faint product corresponding to the deletion of the first

A Exon skipping



B Cryptic donor site activation



C New acceptor site



Figure 1. Representation of the amplified products are depicted below the agarose gel. Grey boxes represent pSPL3 resident exons SD6 and SA2 and green boxes exons from the respective genes mutated. Yellow box represents retained intronic sequence. The asterisk indicates band considered as heteroduplex formation.

three bp of the exon 6, as a consequence of the cryptic acceptor site activation (Fig. 1A, band C). Therefore, if the major aberrant transcript was translated, it would encode for the p.(Glu212_Gln238del) protein, with a loss of 26 amino acids with respect to the wildtype protein. According to the NCBI Conserved Domains predictor, the bestrophin domain would not disappear in the mutated protein. However, according to the dbPTM database, phosphorylation at the residue 227 of the BEST1 protein would be lost.

In the case of the c.2153-12_2155del (*CACNA2D4*) variant, the mutated construct generated an aberrant transcript corresponding to the skipping of exon 23 (Fig. 1A, band B). The mutation deleted the natural acceptor site. This would result in the loss of the reading pattern and could lead to mRNA degradation mediated by terminator

mutations (known as nonsense-mediated mRNA decay, NMD), as supported by the predictor NMDEsc. If NMD would not occur, the mutated protein p.(Cys718Trpfs*19) would be produced, without the VGCC-alpha2 domain.

No splicing effect was observed for *ABCA4* c.6148G>C and *IDH3B* c.532A>G.

When variants altering donor splice sites (5'ss) were considered, the six studied variants consisted of four intronic changes, localized at position + 3 (*PRPF8* c.434 + 3G>A, *RPGRIP1* c.930 + 3A>G), + 4 (*RIMS1* c.2544 + 4A>G) and + 5 (*PRCD* c.74 + 5G>C); and two variants localized at the last position of an exon, affecting the G consensus nucleotide at the position - 1 of 5'ss motifs (*FSCN2* c.1105G>A, *MERTK* c.1450G>A) (Table 1).

Among the intronic variants, only the mutations c.2544 + 4A>G in *RIMS1* and c.74 + 5G>C in *PRCD* induced aberrant splicing in minigenes assays, leading to the complete skipping of the closest exon.

Mutant minigene of c.74 + 5G>C (*PRCD*) variant showed one transcript corresponding to the skipping of exon 1 (Fig. 1A, band F). This mutation avoided the recognition of the natural donor splice site leading to a deletion of the first exon of *PRCD*.

The mutated construct of the c.2544 + 4A>G (*RIMS1*) variant generated an aberrant transcript corresponding to an exon skipping (Fig. 1A, band B). Skipping of exon 14 of *RIMS1* would result in the loss of the reading pattern and NMD-mediated mRNA degradation, as predicted by the NMDEsc tool. If NMD would not occur, the mutated protein p.(Ser791Argfs*4) would be produced without the highly conserved domains C2A_RIM1alpha and C2B_RIM1alpha.

As expected from the in silico predictors, both substitutions located at exonic position - 1 induced the loss of the natural 5'ss and the activation of different cryptic donor splice sites.

The *FSCN2* c.1105G>A variant led to the exclusive use of an intronic cryptic 5'ss (HSF: 76.95, MaxEnt: 9.04), leading to the out-of-frame inclusion of the first 20 nucleotides of *FSCN2* intron 3 in the mature transcript. The processing of the mutated minigene generated two aberrant transcripts as shown in Fig. 1B: band C, in which an insertion of 20 intronic bp was observed, corresponding to donor cryptic site activation CAGgtactg (HSF: 76.95, MaxEnt: 9.04), and band B, corresponding to the loss of exon 3. The new mutated r.1105_1106ins[1105 + 1_1105 + 20] transcript (band C), according to the NMDEsc predictor, would not be degraded and would give rise to the p.(Lys370Glufs*50) protein, lacking one of the four fascin domains. On the other hand, the transcript r.984_1106del would be susceptible to degradation according to the NMDEsc predictor. If NMD would not occur, the mutated protein p.(Ser329Glnfs*84) would be produced without two conserved fascin domains.

The *MERTK* c.1450G>A mutation activated, on the one hand, a downstream cryptic 5'ss (HSF: 82.74, MaxEnt: 5.74), leading to the out-of-frame inclusion of the first 16 bp of *MERTK* intron 9, and, on the other hand, an exonic cryptic 5'ss (HSF: 80.63, MaxEnt: 3.44), leading to the loss of the last 76 nucleotides of *MERTK* exon 9. Mutant minigene showed two aberrant transcripts (Fig. 1B): band E, in which an insertion of 16 intronic bp was observed, as result of the donor cryptic site activation CCAgtaagg (HSF: 82.74, MaxEnt: 5.74), and band F, in which a deletion of 76 bp in the 5' end of exon 9 was observed, corresponding to the donor cryptic site activation ACAgtgagg (HSF: 80.63, MaxEnt: 3.44). Both transcript r.1450_1451ins[1450 + 1_1450 + 16] and transcript r.1374_1450del, would be susceptible to degradation, according to the NMDEsc predictor. If NMD would not occur, the mutated proteins p.(Trp485Glufs*63) and p.(Val459Leufs*60), respectively, would be produced. In both cases the protein tyrosine kinase domain (588–852 aa) would be absent.

Exonic mutations that were distant both from the 5' and 3'ss could activate or create a new splice site and result in splicing defects. We studied three variants whose in silico results predict a new donor site (*ABCA4* c.2972G>T, *CNNM4* c.2039C>T, *PRPF31* c.182C>G) and two new acceptor site (*MAK* c.755A>G, *RHO* c.316G>A) (Table 1). According with the SpliceAI predictor, only the *MAK* c.755A>G variation showed defects in splicing, that creates a new acceptor splice site 92 nucleotides downstream of the exon junction (Fig. 1C). Although, exon 8 canonical acceptor splice site of this gene was not recognized in both wild-type and mutated constructions, missplicing of mutated minigene should be interpreted from the comparison with the wildtype construct. Indeed, a cryptic acceptor splice site in c.760-bp position was activated in the case of wild-type minigene instead of the natural one. However, this was different to acceptor splice observed in mutated construction, where a novel site was activated as a result of the c.755A>G change and according to in silico analysis. Therefore, a 5-bp difference was found between mutated and wild-type minigenes and this difference might increase to 92-bp when distance between aberrant and canonical acceptor splice sites is considered (Fig. 1C, band B). The mutation generated a new acceptor site tcttattcccagTG (HSF: 88.69, MaxEnt: 5.16). The aberrant r.664_755del transcript, according to the NMDEsc predictor, would be susceptible to degradation. If NMD would not occur, it would give rise to the mutated protein p.(Ser222Cysfs*3), with the protein kinase domain (4–284 aa) truncated.

Discussion

Mutations affecting pre-mRNA processing are not insignificant in the population. To date, an 8.7% out of the total number of variants published in HGMD Professional 2020.4 (<http://www.hgmd.cf.ac.uk/ac/index.php>, accessed April, 2021) (26,610/306,768), are disease-causing mutations affecting splicing. However, this estimation still seems conservative because mutations in coding regions are usually considered as missense, nonsense, or silent, which may lead to misclassification of splicing mutations and underestimation of their number¹⁴.

It is difficult to predict the consequences of missense, synonymous and intronic variants, to discriminate neutral variants from those with pathogenic effect. Variants that putatively affect the splicing mechanism are usually considered pathogenic based on the conservation of their canonical splicing site, their absence in control population, co-segregation with the disease in families and the results of bioinformatics predictors¹⁵. In this work, HSF and MaxEnt algorithms have been used to predict the effect on splicing of the variants identified in our cohort. The interpretation of the in silico data and the establishment of thresholds were chosen based on previous work¹⁶. Following these criteria, 15 changes were selected that could potentially affect this mechanism and were further studied with the SpliceAI predictor (Table 1). Of note, of the three in silico predictors used,

SpliceAI was the only one that predicted all the results observed in the in vitro assays. However, these results are only computational predictions, so according to the ACMG criteria, functional studies are required to confirm the effect on mRNA processing. In vitro assays using minigenes have been found to be a useful approach to determine the effect of these variants when genes have a restricted expression profile and the specific tissue sample is difficult to obtain¹⁷.

In contrast to variants affecting the canonical splice sites, which in approximately two-thirds of cases lead to exon skipping, the effect of NCSS variants can vary greatly depending on the sequence alteration and its environment¹⁸. It is acknowledged that not all the positions of 5'ss motif contribute equally to the basepairing with the splicing factors. Excluding the invariant +1 and +2 positions, the positions -1 and +5 would be the most important^{18,19}, as shown also in our minigene assays.

Bestrophin-1 is a channel protein that conducts anions across the EPR membrane²⁰. Alteration of the primary sequence of this protein would have a negative impact on ion conductance, consistent with the decreased or absent response in the EOG described in patients with Best's disease²¹. Histopathological studies of the retina of a patient with Best's disease and the p.(Tyr227Asn) mutation suggest that this mutation could result in erroneous transport of bestrophin-1 to its destination in the cell²². Subsequently, Doumanov et al. demonstrated that the Tyr227AspTrpIle motif is essential in the proper delivery of bestrophin-1 to the basolateral membrane of EPR cells²³. Therefore, we consider that exon 6 of *BEST1* is essential in the function of the protein and the c.637_639del variant would be potentially pathogenic.

The *ABCA4* c.6148G>C and *IDH3B* c.532A>G variants, although they did not meet the impact threshold established for the predictors, were studied in vitro because they were located in the first position of the exon. As expected, none of them was shown to alter the correct processing of the pre-mRNA, in agreement with the in silico results. However, the *ABCA4* c.6148G>C was studied because there is no consensus on the clinical interpretation of this variant in the literature. This was first described by Allikmets et al. as c.6034G>C p.(Val2012Leu), identified in heterozygosity, in two independent families diagnosed with STGD²⁴, and has traditionally been considered as a pathology-causing mutation^{2,25}. However, an exhaustive literature review by Cornelis et al. classifies this variant of uncertain clinical significance²⁶. Furthermore, in many of the cases in which the p.(Val2050Leu) variant has been identified, the second mutation has not been identified and/or has been identified in resolved patients with mutations in other genes and RP phenotype^{27–30}. For all these reasons, we believe that this variant should be interpreted with caution, especially in the context of genetic and reproductive counselling.

In this study we found a variant in *FSCN2* that potentially alters the splicing of the gene likely leading to a truncated protein. The encoded protein Fascin2 is an actin cross-linking protein that is mainly expressed in the photoreceptors^{31,32}. *FSCN2* was firstly associated to autosomal dominant retinitis pigmentosa (RP) in the Japanese population³³. However, truncating mutations that did not co-segregate with the disease in a Spanish family and in a Chinese cohort cast doubts about the real involvement of *FSCN2* in RP^{34,35}.

Based on the literature, the majority of NCSS variants studied in this work (n = 4, 60%) affected the splice donor site. According to Sangermano et al., the longer intronic consensus of the splice acceptor sites (-14 to +1) in which the pyrimidine to pyrimidine substitutions have virtually low effect on the predicted splice strengths, may explain why variants that affect the splice acceptor site are generally less perturbative¹⁴.

In agreement with the results obtained in this work, independently from their type, all the variants suspected to be pathological should be analyzed by bioinformatics tools, and when high score variations are observed, such variants should be studied by in vitro assays, to avoid both false positive and false negative results.

The splicing events observed by minigenes do not always reflect the actual process in the affected tissues. The size restriction of the constructs represents a well-known limitation to fully assess the effect of variants in non-canonical splicing sites. In addition, in a few cases this issue can also affect the recognition of canonical splice sites; such as in MAK minigenes where the chimeric genomic context created between pSPL3 vector and insert sequences led to the recognition of a cryptic splice site instead of the canonical one. In other cases (e.g., *FSCN2* minigenes), a faint transcript with a canonical skipped exon is observed in wildtype minigene. Therefore, as shown here, it is important to study variants at non-canonical splicing sites in the context of their upstream or downstream introns and exons, as their effect is often influenced by other intronic cis-regulatory elements or affects exons located at considerable distances. Another drawback of the in vitro splicing assay is that HEK293 cells do not always mimic splicing defects that might occur in the patient's retina. Indeed, there are studies indicating that specific splicing factors are present in the retina^{36,37}. Despite this, minigenes are a good approach to determine the pathogenic nature of splicing variants when it is difficult to obtain RNA from patient tissue, as in the case of IRD genes, which most of them present a restricted expression profile associated with photoreceptors and RPE cells.

Interestingly, this study has allowed to reclassify, as mutations affecting splicing, three mutations previously categorized as missense: *FSCN2* c.1105G>A, *MAK* c.755A>G and *MERTK* c.1450G>A; and two deletions: *BEST1* c.637_639del and *CACNA2D4* c.2153-12_2155del. Determining the disease mechanism of splice-site variants will provide new opportunities for therapeutic approaches for patients carrying these variants. Antisense oligonucleotide-mediated exon-skipping is a hopeful therapeutic approach, as encouraging results were obtained in the already completed phase I/II clinical trials (<https://clinicaltrials.gov/ct2/show/NCT03140969>) for the most frequent deep-intronic mutation in *CEP290* that cause Leber congenital amaurosis^{12,13}.

This study demonstrates the efficacy of an integrated approach, constituted by MAF, co-segregation, in silico predictions and in vitro assays to determine the effect of potential splice site variants identified by high-throughput DNA sequencing. In this report, we have studied 15 different changes identified in IRD genes by the minigene strategy, confirming the pathogenicity of seven of them. Predictions by splice-site algorithms are crucial to evaluate the pathogenic effects of putative splicing variants and for the clinical interpretation of previously misclassified missense and splice variants. These findings improve the diagnosis of IRD patients who could benefit from the upcoming gene-based therapeutic strategies.

Materials and methods

Patients and molecular diagnosis. Genomic DNA was isolated from 208 patients diagnosed with non-syndromic inherited retinal dystrophies (NS-IRD) from 124 unrelated pedigrees. Written informed consent was obtained from all participants or their legal guardians. This study was approved by the Hospital La Fe Ethics Committee in agreement with the Declaration of Helsinki.

Genetic testing, data analysis and variant interpretation were performed as described previously⁴. Briefly, NGS analysis was performed by target enrichment with a custom panel of 117 IRD-associated genes (Agilent Technologies, Santa Clara, CA, USA) and libraries were sequenced on an Illumina MiSeq platform (Illumina, San Diego, CA, USA). Data analysis was performed using SureCall software (Agilent Technologies, Santa Clara, CA, USA) and variant annotation employing wANNOVAR. Variants were classified according to standards of the American College of Medical Genetics and Genomics (ACMG)⁵.

In silico analysis and selection of variants. To evaluate the effect of the identified variants in the splicing process and recognition of donor and acceptor sites, in silico studies were first performed. Two different algorithms: Human Splicing Finder (HSF) and MaxEnt, included in the splicing predictor tool HSF 3.1 (<http://www.umd.be/HSF/index.html>)³⁸ were used for this purpose. Candidate variants for the functional assay were selected according to the cut-offs defined by Desmet et al.³⁸: 5' or 3' splice sites scores > 3 for MaxEnt and > 65 for HSF were combined with a 30% and 10% score variation between wild-type and mutated sequences, respectively. In addition, mutations at position +4 and at position +3 or +5 with a score variation of 7% and 14%, respectively, were included. Furthermore, variants located in the first position of the exon and matching one or none of the above criteria were selected: *ABCA4*: c.6148G>C, *BEST1*: c.637_639del and *IDH3B*: c.532A>G. Mutations located at canonical splicing sites were not selected for the minigene assay. The mutations selected for the in vitro assay were also analyzed using the new generation predictor SpliceAI³⁹ included in the MobiDetails platform (<https://mobidetails.iurc.montp.inserm.fr/MD/>). Variants were analysed independently from their putative association with the phenotype of the patient. Clinical diagnosis and genotype of patients carrying the variants selected for the minigene assay is detailed in Supplementary Table 1. Co-segregation analysis was performed when DNA from family member was available (Supplementary Fig. 2).

Minigene assay. Minigenes were used to assess the impact of selected variants on the mRNA processing.

For all the selected variants, involved exons and their flanking intronic sequences (~ ± 200 bp for each side) were amplified from the patients' DNA using the High Fidelity Phusion polymerase (Thermo Fisher Scientific Waltham, MA, USA) with specific primers containing an additional XhoI and NheI restriction site (Supplementary Table 2). Amplicons were inserted between the XhoI and NheI restriction sites using T4 DNA ligase (Thermo Fisher Scientific) into the pSPL3 exon-trapping vector (kindly provided by Dr. I. Botillo and Dr. S. Tuffery-Giraud). *Escherichia coli* competent cells (XL1-Blue Supercompetent Cells, Agilent Technologies, Santa Clara, CA, USA) were transformed with the plasmid construction by heat shock. pSPL3 vector has not been developed for the study variants within first and last exons, since these are preceded and followed by UTRs, respectively. Therefore, to facilitate the recognition of *PRCD* exon 1 by spliceosome, we modified by directed mutagenesis (QuikChange II XL Site-Directed Mutagenesis Kit) a decoy acceptor splice site identified in silico within this exon obtaining a stronger acceptor splice site, as endorsed by in silico and functional analyses (i.e. wild-type minigene). To this aim, we modified 4 positions (as shown in the figure) in order to increase the acceptor site prediction score from 88.43 to 96.30 for HSF and 7.39–12.6 for MaxEnt (Supplementary Table 3 and Supplementary Fig. 3). The sequence of each minigene construction was verified by direct sequencing.

Subsequently, one microgram of each wild-type and mutated construct were transfected with the Lipofectamine™ 3000 reagents (Thermo Fisher Scientific) into HEK293 cells grown in six-well plates, according to manufacturer's instructions. All the transfection experiments in HEK293 cells were performed in duplicate. Twenty-four hours after transfection, cells were harvested and total RNA was isolated by the RNeasy Mini kit (Qiagen, Hilden, Germany). Then, cDNA synthesis was carried out using the PrimeScript RT Reagent Kit (TaKaRa, Kusatsu, Japan) according to the manufacturer's recommendations. The cDNA was amplified using the pSPL3-specific SD6 forward and SA2 reverse primers³⁴. PCR products were visualized on a 1% agarose gel and bands were purified with the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) and directly sequenced in both senses by Sanger method (BigDye Terminator kit v3.1, Applied Biosystems).

Interpretation of minigenes assay. We used the Name Checker tool of Mutalyzer 2.032 (<https://www.mutalyzer.nl/name-checker>) for the annotation of the transcripts resulting from the minigene assays together with their protein products. The NMDEscPredictor bioinformatics predictor (<https://nmdprediction.shinyapps.io/nmdescpredictor/>) was also used to test whether the newly mutated transcript could be degraded by the Nonsense Mediated Decay (NMD) pathway. In addition, the domains of the affected proteins were consulted in the NCBI Conserved domains database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?>) and the possible post-transcriptional modifications of the protein in the cBioPortal database (<https://www.cbioportal.org/>).

References

- Berger, W., Kloeckener-Gruissem, B. & Neidhardt, J. The molecular basis of human retinal and vitreoretinal diseases. *Prog. Retin. Eye Res.* **29**, 335–375 (2010).
- Carss, K. *et al.* Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. *Am. J. Hum. Genet.* **100**, 75–90 (2017).
- Sanchis-Juan, A. *et al.* Complex structural variants in Mendelian disorders: Identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Med.* **10**, 95 (2018).
- Rodríguez-Muñoz, A. *et al.* Expanding the clinical and molecular heterogeneity of nonsyndromic inherited retinal dystrophies. *J. Mol. Diagnostics* **22**, 532–543 (2020).
- Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–423 (2015).
- Bauwens, M. *et al.* ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: Novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. *Genet. Med.* **21**, 1761–1771 (2019).
- Siva, K., Covello, G. & Denti, M. A. Exon-skipping antisense oligonucleotides to correct missplicing in neurogenetic diseases. *Nucleic Acid Ther.* **24**, 69–86 (2014).
- Bremer, J. *et al.* Antisense oligonucleotide-mediated exon skipping as a systemic therapeutic approach for recessive dystrophic epidermolysis bullosa. *Mol. Ther. Nucleic Acids.* **5**, e379 (2016).
- Matsuo, M. Antisense oligonucleotide-mediated exon-skipping therapies: precision medicine spreading from Duchenne muscular dystrophy. *JMA J.* **4**, 232–240 (2021).
- Slijkerman, R. W. *et al.* Antisense oligonucleotide-based splice correction for USH2A-associated retinal degeneration caused by a frequent deep-intronic mutation. *Mol. Ther. Nucleic Acids* **5**, e381 (2016).
- Liquori, A. *et al.* Whole USH2A gene sequencing identifies several new deep intronic mutations. *Hum. Mutat.* **37**, 184–193 (2016).
- Cideciyan, A. V. *et al.* Durable vision improvement after a single treatment with antisense oligonucleotide sepoparsen: A case report. *Nat. Med.* **27**, 785–789 (2021).
- Cideciyan, A. V. *et al.* Effect of an intravitreal antisense oligonucleotide on vision in Leber congenital amaurosis due to a photoreceptor cilium defect. *Nat. Med.* **25**, 225 (2019).
- Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* **3**, 285–298 (2002).
- Sangermano, R. *et al.* ABCA4 midgenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. *Genome Res.* **28**, 100–110 (2018).
- Desmet, F.-O. *et al.* Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67–e67 (2009).
- Roepman, R. & Wolfrum, U. Protein networks and complexes in photoreceptor cilia. *Subcell. Biochem.* **43**, 209–235 (2007).
- Carmel, I., Tal, S., Vig, I. & Ast, G. Comparative analysis detects dependencies among the 5' splice-site positions. *RNA* **10**, 828–840 (2004).
- Le Guédard-Méruze, S. *et al.* Sequence contexts that determine the pathogenicity of base substitutions at position +3 of donor splice-sites. *Hum. Mutat.* **30**, 1329–1339 (2009).
- Sun, H., Tsunenari, T., Yau, K.-W. & Nathans, J. The vitelliform macular dystrophy protein defines a new family of chloride channels. *Proc. Natl. Acad. Sci. USA* **99**, 4008–4013 (2002).
- Weleber, R. G. Fast and slow oscillations of the electro-oculogram in best's macular dystrophy and retinitis pigmentosa. *Arch. Ophthalmol.* **107**, 530 (1989).
- Mullins, R. F., Oh, K. T., Heffron, E., Hageman, G. S. & Stone, E. M. Late development of vitelliform lesions and flecks in a patient with best disease: Clinicopathologic correlation. *Arch. Ophthalmol. (Chicago, Ill. 1960)*. **123**, 1588–1594 (2005).
- Doumanov, J. A. *et al.* Disease-causing mutations in BEST1 gene are associated with altered sorting of bestrophin-1 protein. *Int. J. Mol. Sci.* **14**, 15121–15140 (2013).
- Allikmets, R. *et al.* A photoreceptor cell-specific ATP-binding transporter gene (ABCR) is mutated in recessive Stargardt macular dystrophy. *Nat. Genet.* **15**, 236–246 (1997).
- Corton, M. *et al.* Exome sequencing of index patients with retinal dystrophies as a tool for molecular diagnosis. *PLoS ONE* **8**, 1–6 (2013).
- Cornelis, S. S. *et al.* In silico functional meta-analysis of 5962 ABCA4 variants in 3928 retinal dystrophy cases. *Hum. Mutat.* **38**, 400–408 (2017).
- Briggs, C. E. *et al.* Mutations in ABCR (ABCA4) in patients with Stargardt macular degeneration or cone-rod degeneration. *Investig. Ophthalmol. Vis. Sci.* **42**, 2229–2236 (2001).
- Poloschek, C. M. *et al.* ABCA4 and ROM1: Implications for modification of the PRPH2-associated macular dystrophy phenotype. *Invest. Ophthalmol. Vis. Sci.* **51**, 4253–4265 (2010).
- Haer-Wigman, L. *et al.* Diagnostic exome sequencing in 266 Dutch patients with visual impairment. *Eur. J. Hum. Genet.* **25**, 591–599 (2017).
- Jespersgaard, C. *et al.* Molecular genetic analysis using targeted NGS analysis of 677 individuals with retinal dystrophy. *Sci. Rep.* **9**, 1–7 (2019).
- Saishin, Y. *et al.* Isolation of a cDNA encoding a photoreceptor cell-specific actin-bundling protein: Retinal fascin. *FEBS Lett.* **414**, 381–386 (1997).
- Tubb, B. E. *et al.* Characterization of human retinal fascin gene (FSCN2) at 17q25: Close physical linkage of fascin and cytoplasmic actin genes. *Genomics* **65**, 146–156 (2000).
- Wada, Y. *et al.* Mutation of human retinal fascin gene (FSCN2) causes autosomal dominant retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **42**, 2395–2400 (2001).
- Gamundi, M. J. *et al.* Sequence variations in the retinal fascin FSCN2 gene in a Spanish population with autosomal dominant retinitis pigmentosa or macular degeneration. *Mol. Vis.* **11**, 922–928 (2005).
- Zhang, Q., Li, S., Xiao, X., Jia, X. & Guo, X. The 208delG mutation in FSCN2 does not associate with retinal degeneration in Chinese individuals. *Investig. Ophthalmol. Vis. Sci.* **48**, 530 (2007).
- Murphy, D., Cieply, B., Carstens, R., Ramamurthy, V. & Stoilov, P. The Musashi 1 controls the splicing of photoreceptor-specific exons in the vertebrate retina. *PLoS Genet.* **12**, e1006256 (2016).
- Parfitt, D. A. *et al.* Identification and correction of mechanisms underlying inherited blindness in human iPSC-derived optic cups. *Cell Stem Cell* **18**, 769–781 (2016).
- Desmet, F. O. Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**(9), e67 (2009).
- Jaganathan, K. *et al.* Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548.e24 (2019).

Author contributions

Conceptualization, G.G.G., E.A. and J.M.M.; methodology, A.R.M. and B.G.B.; formal analysis, A.R.M., G.G.G. and A.L.; writing—original draft preparation, A.R.M.; writing—review & editing, supervision, A.R.M., B.G.B., A.L., T.J. and G.G.G.; funding acquisition, J.M.M.

Funding

This study has been funded by the Project PI19/00303. Health Research Institute Carlos III (ISCIII; Spanish Ministry of Health and Innovation) and the Regional Government of the Valencian Community (PROMETEU/2018/135), all partially supported by with partial support from the European Regional Development Fund (ERDF). ARM is recipient of a Río Hortega contract (CM18/00199) from the ISCIII, BGB is recipient of a predoctoral contract (ACIF/2019/252) from the Government of the Valencian Community and GGG has a postdoctoral contract from CIBERER. A.L. is recipient of the APOSTD/2021/212 Generalitat Valenciana post-doctoral fellowship.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-03925-1>.

Correspondence and requests for materials should be addressed to J.M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022