Investigation of Aberrant Splicing Induced by *AIPL1* Variations as a Cause of Leber Congenital Amaurosis

James Bellingham,[1] Alice E. Davidson,[1] Jonathan Aboshiha,[1,2] Francesca Simonelli,[3] James W. Bainbridge,[1,2] Michel Michaelides,[1,2] and Jacqueline van der Spuy[1]

[1]UCL Institute of Ophthalmology, London, United Kingdom
[2]Moorfields Eye Hospital NHS Foundation Trust, London, United Kingdom
[3]Eye Clinic, Multidisciplinary Department of Medical, Surgical and Dental Sciences, Second University of Naples, Naples, Italy

Correspondence: Jacqueline van der Spuy, UCL Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, UK;
j.spuy@ucl.ac.uk.

Submitted: September 1, 2015
Accepted: November 5, 2015

Citation: Bellingham J, Davidson AE, Aboshiha J, et al. Investigation of aberrant splicing induced by *AIPL1* variations as a cause of Leber congenital amaurosis. *Invest Ophthalmol Vis Sci*. 2015;56:7784–7793.
DOI:10.1167/iovs.15-18092

Purpose. Biallelic mutations in *AIPL1* cause Leber congenital amaurosis (LCA), a devastating retinal degeneration characterized by the loss or severe impairment of vision within the first few years of life. *AIPL1* is highly polymorphic with more than 50 mutations and many more polymorphisms of uncertain pathogenicity identified. As such, it can be difficult to assign disease association of *AIPL1* variations. In this study, we investigate suspected disease-associated *AIPL1* variations, including nonsynonymous missense and intronic variants to validate their pathogenicity.

Methods. *AIPL1* minigenes harboring missense and intronic variations were constructed by amplification of genomic fragments of the human *AIPL1* gene. In vitro splice assays were performed to identify the resultant *AIPL1* transcripts.

Results. We show that all nine of the suspected disease-associated *AIPL1* variations investigated induced aberrant pre-mRNA splicing of the *AIPL1* gene, and our study is the first to show that *AIPL1* missense mutations alter *AIPL1* splicing. We reveal that the presumed rare benign variant c.784G>A [p.(G262S)] alters in vitro *AIPL1* splicing, thereby validating the disease-association and clarifying the underlying disease mechanism. We also reveal that in-phase exon skipping occurs normally at a low frequency in the retina, but arises abundantly as a consequence of specific *AIPL1* variations, suggesting a tolerance threshold for the expression of these alternative transcripts in the retina normally, which is exceeded in LCA.

Conclusions. Our data confirm the disease-association of the *AIPL1* variations investigated and reveal for the first time that aberrant splicing of *AIPL1* is an underlying mechanism of disease in LCA.

Keywords: molecular chaperone, minigene, splice mutations

Leber congenital amaurosis (LCA) is the most severe inherited retinal degeneration and the most common cause of congenital blindness.[1] Leber congenital amaurosis is diagnosed in infancy and is characterized by a severely reduced or nonrecordable ERG. Leber congenital amaurosisis is genetically heterogeneous and typically inherited in an autosomal recessive manner with more than 19 genes reported to be involved (Retinal Information Network), including *AIPL1*.[2]

The *AIPL1* gene is comprised of six exons coding for the Aryl Hydrocarbon Receptor Interacting Protein-Like 1. AIPL1 is a photoreceptor-specific molecular cochaperone that interacts specifically with the molecular chaperone HSP90 to modulate the stability and assembly of the HSP90 substrate, retinal cGMP phosphodiesterase (PDE6).[3–5]

Presently, the Human Gene Mutation Database (HGMD) records 52 *AIPL1* mutations.[6] Several of these *AIPL1* mutations are known benign single nucleotide polymorphisms (SNPs), while several others can be reclassified as benign rare variants in the light of recent evidence, including the detection of these sequence variations in the homozygous state in unaffected control populations.[7,8] A number of *AIPL1* nonsynonymous missense sequence variations reported as LCA-associated, but not recorded in the HGMD, are of uncertain pathogenicity.[7] Moreover, several hundred *AIPL1* sequence variations of different functional classes are reported throughout the *AIPL1* gene in the Single Nucleotide Polymorphism Database (dbSNP). Ascribing the disease-causing status of LCA-associated *AIPL1* sequence variations is therefore a challenge confounded by the polymorphic nature of the gene. Confirmation of disease-associated *AIPL1* variations is important for patient diagnosis, counselling, and potential treatments, with the main therapeutic strategy currently focusing on *AIPL1*-targeted gene replacement therapy.

Thorough in vitro investigations are necessary to differentiate true disease-associated variations from rare polymorphisms. The functional consequences of only a handful of missense and nonsense changes in the *AIPL1* coding sequence have been investigated by introducing the variation in the cDNA sequence and testing the impact on protein function.[4,9–14] These assays have primarily investigated the consequences of *AIPL1* variations on domain-mediated AIPL1 protein interactions. The AIPL1 carboxy-terminal tetratricopeptide repeat (TPR) domain mediates the interaction with the molecular chaperone HSP90.[4] The AIPL1 amino-terminal FK506 binding protein (FKBP)-like domain interacts directly with a farnesyl motif in vitro and is therefore predicted to

interact with the farnesylated alpha subunit of retinal PDE6.[14] A number of in vitro functional studies investigating AIPL1 domain-dependent interactions,[4,9–11,14] have proven useful in validating disease-causing *AIPL1* mutations and understanding the role of AIPL1 as an HSP90 cochaperone for retinal PDE6.

In contrast, the effects of noncoding variations in *AIPL1* are unknown, and missplicing of *AIPL1* as an underlying disease mechanism has not been experimentally investigated. In this study, we investigated missense and intronic *AIPL1* variations identified in LCA patients that are predicted to alter *AIPL1* pre-mRNA splicing. We determined the outcome of the variations on *AIPL1* splicing and confirm that aberrant alternative transcription of *AIPL1* could be an underlying cause of LCA.

Materials and Methods

AIPL1 Sequence Variations and Nomenclature

AIPL1 sequence variations detected in previously genotyped LCA patients were investigated in this study.[7,15,16] The *AIPL1* sequence variations investigated were selected on the basis that the disease-causing status was either unknown or uncertain,[7] and on the basis that we had previously performed clinical investigations of the patients harboring the *AIPL1* variations.[7,16] The study was conducted in accordance with the tenets of the Declaration of Helsinki and approved by the Moorfields and Whittington Hospitals' local ethics committee. Ensembl and UCSC Genome Browsers were used for analyses of the Human GRCh37/hg19 (February 2009) and GRCh38/hg38 (December 2013) assemblies. The *AIPL1* cDNAs are numbered according to the Ensembl Transcript ENST00000381129 (RefSeq NM_014336, NP_055151). Complementary DNA nucleotide numbering uses +1 as the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. The coordinates for the genomic *AIPL1* sequence used are chr17:6,327,059-6,338,519 (hg19) and chr17:6,423,737-6,435,185 (hg38). Nomenclature was according to HGVS standards.[17]

AIPL1 In Silico Analysis

AIPL1 variations were analyzed using three in silico software prediction programmes: SIFT (Sorting Intolerant From Tolerant; in the public domain, http://sift.jcvi.org),[18] PolyPhen-2 (Polymorphism Phenotyping v2; in the public domain, http://genetics.bwh.harvard.edu/pph/index.html),[19] and BLOSUM62 (Blocks Substitution Matrix; in the public domain, http://www.uky.edu/Classes/BIO/520/BIO520WWW/blosum62.htm). In silico analysis of splice-site confidence levels was performed with the NetGene2 server (in the public domain, http://www.cbs.dtu.dk/services/NetGene2/)[20,21] and the Berkeley Drosophila Genome Project Neural Network splice site prediction algorithm NNSPLICE (in the public domain, http://www.fruitfly.org/seq_tools/splice.html)[22] using default settings.

Construction of Minigenes

Primers were designed to amplify genomic fragments of the human *AIPL1* gene. Restriction sites were added to the 5′end of the primers such that the amplified product could be directionally cloned into the pBK-CMV expression vector (Agilent Technologies, Waldbronn, Germany). The 5′ end of each minigene was cloned into the *Bmt*I (*Nhe*I) site, thereby deleting the *lac* promoter, while the corresponding 3′ ends were cloned into either the *Xho*I or *Not*I site.

Three *AIPL1* minigenes were constructed. Minigene 1 encompasses exons 1 through 4 and excludes the start codon to lessen the potential effects of nonsense-mediated decay (NMD). This minigene was amplified in two fragments that were cloned sequentially into pBK-CMV. Fragment 1 primers were designed to amplify a 1418 bp fragment and tagged with either *Nhe*I or *Sal*I: AIPL1_Ex1_1F, 5′-gctagcGATGCCGCTCTGCTCCTGAACGTGGAAG-3′ and AIPL1_Int2_1R, 5′-gtcgacACTCCTGCTGGTCATAGCCCTGCTCC-3.′ Fragment 2 primers were designed to amplify a 2126 bp fragment and were tagged with either *Sal*I or *Not*I: AIPL1_Int2_1F, 5′-gtcgacCAGCACTGCCAGGACACCAAAGCGACTCTCTTGG-3′ and AIPL1_Ex4_1R, 5′-gcggccgcTTGGTCTGCAGGTTCCTTAGGCAGATGATGG-3.′ The *Sal*I sites introduce an artificial junction to allow deletion of approximately 4.7 kb of the central region of intron 2 of *AIPL1*.

Minigene 2 encompasses exons 3-5. Primers were designed to amplify an 1892 bp fragment and were tagged with either *Bmt*I or *Xho*I: AIPL1_Ex3_1F, 5′-gctagcCACACGGGGGTCTACCCCATCCTATCC-3′ and AIPL1_Ex5_1R, 5′-ctcgagCTGGGTGGTGCCGGAGAATATCACTGGTGTGC-3.′

Minigene 3 encompasses exons 4-6. Primers were designed to amplify a 1693 bp fragment and were tagged with either *Bmt*I or *Xho*I: AIPL1_Int3_1F, 5′-gctagcGGGGTCCCTGCCTCACTGACCTGCAGC-3′ and AIPL1_Ex6_3UTR_1R, 5′-ctcgagACCAGAAGTGACCAGGCCACTTGCTCC-3.′

Genomic *AIPL1* fragments were PCR amplified using Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Hitchin, UK) in 25-μL reactions containing 10 pmol of each primer (Sigma-Aldrich, Dorset, UK) and 10 ng of genomic DNA, with the following cycling conditions: *Td* = 98°C for 30 seconds, then 30 cycles of *Td* = 98°C for 10 seconds, *Ta* = 72°C for 10 seconds, and *Te* = 72°C for 60 seconds. Polymerase chain reaction products were resolved on a 1.2% agarose gel (TAE buffer). Amplicons were excised and gel purified (QIAquick Gel Extraction Kit; QIAGEN, Crawley, UK), prior to cloning into pSC-Bamp/kan (StrataClone Blunt PCR Cloning Kit; Agilent Technologies). Sequence identity and integrity was confirmed by direct Sanger sequencing. Where patient gDNA was not readily available, mutations were engineered into w/t alleles cloned in the pSB-Bamp/kan vector using site-directed mutagenesis (SDM; Q5 Site-Directed Mutagenesis Kit; New England Biolabs). Primers were designed using NEBaseChanger software (in the public domain, http://nebasechanger.neb.com), and SDM PCRs undertaken using suggested conditions. For each mutation introduced by SDM, the entire allele was sequenced to confirm mutation and sequence integrity. Sequence verified *AIPL1* fragments were cloned into pBK-CMV using standard protocols (enzymes supplied by New England Biolabs). Briefly, after digestion with appropriate enzyme pairs, digestion products were separated on a 1.2% agarose gel and relevant fragments excised, gel purified and ligations prepared (T7 DNA Ligase). Ligations were transformed into DH5α competent cells (New England Biolabs). Completed expression constructs were sequence checked to confirm mutational identity.

In Vitro Splice Assays

Minigenes were transfected into HEK293 cells plated at 2.5×10^5 cells/well in 6-well plates with TransIT-LT1 Transfection Reagent (Cambridge BioScience for Mirus Bio LLC, Cambridge, UK) following the manufacturer's instructions. Twenty-four hours post transfection, cells were rinsed twice with PBS and total RNA extracted (RNeasy Mini Kit and QIAshredder, QIAGEN) with on-column DNAse treatment (Promega, Hampshire, UK). Poly-T primed cDNA was synthesized in 20-μL volume from 2.5 μg of total RNA (Tetro cDNA Synthesis Kit, Bioline Reagents Ltd., London, UK). Splice products were

TABLE 1. In Silico Analysis of *AIPL1* Variations

Nucleotide Change	Protein Change	dbSNP	Genomic Position Hg38	Genomic Position Hg19	ExAC Allele Count (Homo)	SIFT (Tolerance Index 0 to 1)	PolyPhen-2 (HumVar Score 0 to 1)	Blosum62 Score (−4 to 11)
c.97_104dupGTGATCTT	p.(F35Lfs*2)	NA	17:6,434,090_6,434,098	17:6,337,410_6,337,418	NI	NA	NA	NA
c.98_99insTGATCTTG	p.(I34Dfs*10)	NA	17:6,434,097_6,434,096	17:6,337,417_6,337,416	NI	NA	NA	NA
c.276+1G>A	NA	rs150097891	17:6,433,918	17:6,337,238	2/112,214 (0)	NA	NA	NA
c.276+2T>C	NA	NA	17:6,433,917	17:6,337,237	1/111,348 (0)	NA	NA	NA
c.277-2A>G	NA	rs140808549	17:6,428,508	17:6,331,828	2/122,492 (0)	NA	NA	NA
c.465G>T	p.(Q155H)	NA	17:6,428,318	17:6,331,638	3/121,960 (0)	DAM (0.01)	POS (0.702)	0
c.642G>C	p.(K214N)	NA	17:6,233,521	17:6,330,201	1/122,766 (0)	DAM (0)	PRD (0.954)	0
c.784G>A	p.(G262S)	rs142326926	17:6,426,615	17:6,329,935	6/117,098 (0)	TOL (0.18)	BENIGN (0.437)	0
c.785-10_786del	NA	NA	17:6,425,840_?	17:6,329,160_?	NI	NA	NA	NA

The cDNA is numbered according to the longest *AIPL1* transcript ENST00000381129, protein ID ENSP00000370521. ExAC allele count: Allele frequency of the *AIPL1* variations in a reference data set derived from 61,486 unrelated individuals sequenced as part of various disease-specific and population genetic studies, including the 1000 Genomes Project and the NHLBI-GO Exome Sequencing Project (ESP) in addition to many other large scale sequencing studies; excludes cases of severe pediatric disease. SIFT results: tolerance index ≥ 0.05 = tolerated (TOL), tolerance index < 0.05 = damaging (DAM). PolyPhen-2: benign, possibly damaging (POS), probably damaging (PRD). Blosum62 substitution matrix score: positive numbers = likely to be tolerated, negative numbers = unlikely to be tolerated. NA, not applicable; NI, not identified.

amplified from 1 μL of cDNA using the following primer pairs: minigene 1, AIPL1_Ex1_1F and AIPL1_Ex4_1R - expected w/t amplicon = 652 bp; minigene 2, AIPL1_Ex3_1F and AIPL1_Ex5_1R - expected w/t amplicon = 797 bp; minigene 3, AIPL1_Int3_1F and AIPL1_Ex6_3UTR_1R - expected w/t amplicon = 520 bp. Polymerase chain reaction conditions were similar to those used for minigene generation except that the extension time was reduced to 10s/cycle. Polymerase chain reaction products were resolved on 2% to 3% agarose gels. Amplicons were excised, gel purified, cloned, and sequenced as above. All sequence analysis was conducted using MacVector 11.1.2 (MacVector, Inc., Cary, NC, USA).

RNA-Seq Data Analysis

Publically available RNA-seq data from three healthy human adult retinas[23] was aligned against the hg19 reference genome. Aligned reads were visualized using the Broad Institute Integrated Genomics Viewer (IGV-2.3.40; Cambridge, MA, USA).[24,25] A Sashimi plot of *AIPL1* mRNA sequencing reads from each of the three retinas aligned to the reference genome (human hg19) was generated to visualize differentially spliced exons in normal retina. The transcript abundance was estimated using the FPKM (fragments per kilo bases of exons for per million mapped reads) normalization method.

RESULTS

In Silico Analysis of AIPL1 Variations

In silico analysis of the *AIPL1* variations investigated indicates that all are very rare (allele frequency < 1/10,000) and have not been observed in the homozygous state in the current Exome Aggregation Consortium (ExAC) data set (Table 1). Of the three missense substitutions, the c.784G>A [p.(G262S)] substitution appears to be the least damaging, being SIFT tolerant and PolyPhen-2 benign. This is consistent with previous in vitro studies that do not report a deficit in [p.(G262S)] function compared with the wild-type protein.[11] SIFT determines that the c.465G>T [p.(G155H)] and c.642G>C [p.(K214N)] missense substitutions are damaging, and neither are considered benign by PolyPhen-2 and Blosum62. In silico analysis of the effect on splicing indicates that all 9 *AIPL1* variations have the potential to result in some form of misspliced *AIPL1* transcripts (Table 2). All *AIPL1* variations are predicted to reduce or abolish native splice site recognition leading to exon skipping, or in the case of c.276+2T>C and c.785-10_786del, the creation of potential cryptic splice sites.

In Vitro Splicing Assay

We confirmed experimentally that *AIPL1* could not be amplified from total RNA purified from whole blood samples from unaffected individuals (data not shown). Therefore, three *AIPL1* minigenes were constructed to investigate the splice effects of the *AIPL1* variations (Figs. 1A, 1B). The identity of the indicated PCR fragments (Fig. 1C, bands 1–24) was determined by sequence analysis (Supplementary Figs. S2–S22). It can be seen from Figure 1C that the three w/t constructs produced the cleanest PCR fragment profiles with the fewest unique fragments, typically a strong splice product of the expected size for minigene 1 (band 1), minigene 2 (band 16), and minigene 3 (band 19). For w/t minigene 1, in addition to the expected PCR fragment (band 1), a smaller secondary splice product (band 2) was also amplified (see below). All of the minigenes carrying nucleotide variations exhibited PCR fragment profiles that differed from the corresponding w/t

Table 2. *AIPL1* Splice Site Predictions

Mutation	Context	NetGene2 w/t	NetGene2 mut	NNSPLICE w/t	NNSPLICE mut	Prediction(s)
c.97_104dup [p.(F35Lfs*2)]	intron 1 / EXON 2 w/t: ggtgactagGTGATCTT--------TCATTTCCGC mut: ggtgactagGTGATCTT**GTGATCTT**TCATTTCCGC	0.43	0.20	n/a	n/a	(1) Splice acceptor-site recognized, normal splicing. (2) Reduction in splice acceptor-site recognition, exon 2 skipped.
c.98_99insTGATCTTG [p.(I34Dfs*10)]	intron 1 / EXON 2 w/t: ggtgactagGT--------GATCTTTCATTTCC mut: ggtgactagGT**TGATCTTG**GATCTTTCATTTCC	0.43	0.17	n/a	n/a	(1) Splice acceptor-site recognized, normal splicing. (2) Reduction in splice acceptor-site recognition, exon 2 skipped.
c.276+1G>A	EXON 2 / intron 2 w/t: TGGTGCGACACCATCgtaagtagg mut: TGGTGCGACACCATC**a**taagtagg	0.96 (H)	n/a	0.92	n/a	(1) Splice donor-site abolished, exon 2 skipped.
c.276+2T>C	EXON 2 / intron 2 w/t: TGGTGCGACACCATCgtaagtagg mut: TGGTGCGACACCATCg**c**aagtagg	0.96 (H)	0.06 (0.71 @ c.276+5)	0.92	n/a (0.44 @ c.276+5)	(1) Splice donor-site abolished, exon 2 skipped. (2) Cryptic splice donor-site 5 bp into intron 2 activated?
c.277-2A>G	intron 2 / EXON 3 w/t: tccccacagCACACGGGGGTCTAC mut: tccccac**g**gCACACGGGGGTCTAC	0.95 (H)	n/a	0.87	n/a	(1) Splice acceptor-site abolished, exon 3 skipped.
c.465G>T [p.(Q155H)]	EXON 3 / intron 3 w/t: ATCGAGCTGCTGCAGgtggggctg mut: ATCGAGCTGCTGCA**T**gtggggctg	0.88	n/a	n/a	n/a	(1) Splice donor-site not recognized, exon 3 skipped.
c.642G>C [p.(K214N)]	EXON 4 / intron 4 w/t: AACCTGCAGACCAAGgtcagaggc mut: AACCTGCAGACCAA**C**gtcagaggc	0.94 (H)	0.55	0.81	n/a	(1) Splice donor-site recognized, normal splicing. (2) Reduction in splice site recognition, exon 4 skipped.
c.784G>A [p.(G262S)]	EXON 5 / intron 5 w/t: TCCGGCACCACCCAGgtgcgcggg mut: TCCGGCACCACCCA**A**gtgcgcggg	0.97 (H)	0.71	0.88	0.42	(1) Splice donor-site recognized, normal splicing. (2) Reduction in splice site recognition, exon 5 skipped.
c.785-10_786del	intron 5 / EXON 6 w/t: atgctccctgctccccacagGCATCGTGAAGGCCT mut: atgctccctg**********ATCGTGAAGGCCT	1.00 (H)	n/a (0.34 @ c.796)	0.98	n/a	(1) Splice acceptor-site abolished, intron 5 retained. (2) Cryptic splice acceptor-site 11 bp into exon 6 activated?

Variations are bold and underlined; asterisks (*) indicate deleted nucleotides; dashes (-) maintain nucleotide alignment for insertions; n/a indicates that a splice site recognition score was not returned. The cDNA is numbered according to the longest *AIPL1* transcript ENST00000381129. NetGene2 server (in the public domain, http://www.cbs.dtu.dk/services/NetGene2/) cutoff values used for confidence: highly confident donor sites (H), 95.0%; nearly all true donor sites, 50.0%; highly confident acceptor sites (H), 95.0%; nearly all true acceptor sites, 20.0%. NNSPLICE server (in the public domain, http://www.fruitfly.org/seq_tools/splice.html) donor-site and acceptor-site cutoff score = 0.40.

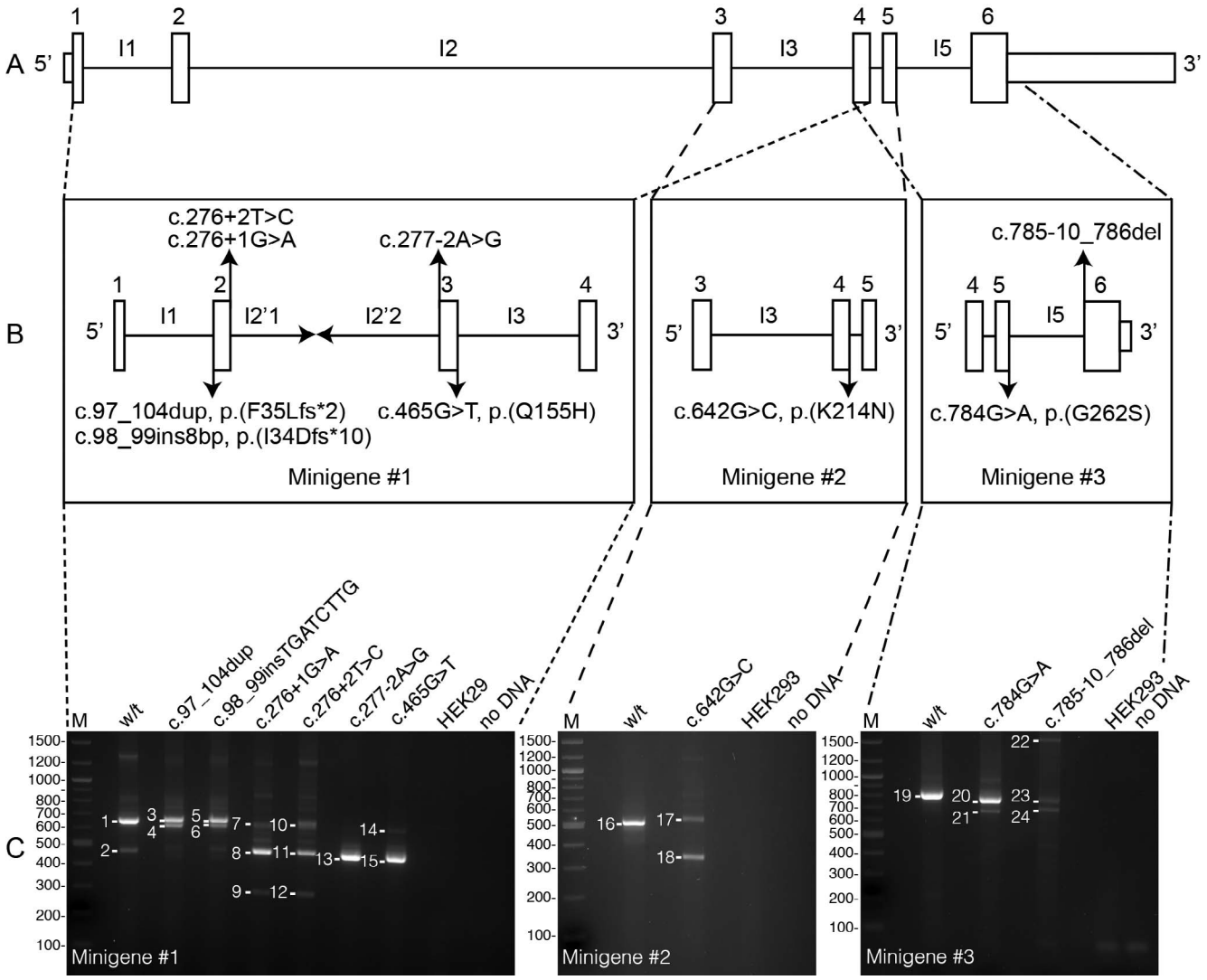

Figure 1. Transcript analysis of *AIPL1* variations mapping within exons or introns close to or within intron/exon boundaries in LCA patients. (**A**) *AIPL1* genomic structure of the longest *AIPL1* transcript (RefSeq NM_014336). The exons (1–6) and introns (I1–I5) are numbered from the 5′ to the 3′ untranslated region (UTR). (**B**) Minigene 1 encompasses the entire *AIPL1* sequence from exon 1 to 4 amplified from patient gDNA, with the exception of a large central part of intron 2. Minigene 2 encompasses the entire *AIPL1* sequence from exon 3 to 5. Minigene 3 encompasses the entire *AIPL1* sequence from exon 4 to 6 as well as a small region of the 3′ UTR. The corresponding minigenes from the wild-type allele were used as the control for transcript analysis. *Arrows* indicate the positions of the *AIPL1* variations under investigation. (**C**) Transcript analysis. *AIPL1* cDNA transcripts obtained from the minigene assay were PCR amplified and resolved on 2% to 3% agarose gels. Sanger sequencing was used to confirm the sequence identity and integrity of the resultant amplicons. The molecular weight markers are in base pairs, and the bands analysed are numbered (see Table 3).

minigene. Polymerase chain reaction products were selected for sequence analysis on the basis of size similarity to the w/t product(s) and abundance. High molecular weight bands greater than the expected size of the w/t product and present in all samples, including the w/t control, were excluded from analysis. The results of the in vitro splicing assay are summarised in Table 3, which lists the band size (bp), a detailed description of the corresponding observed splicing effect (also see Supplementary Figs. S2–S22) and the predicted protein product. The resultant splice products and their predicted protein products are also depicted schematically in Figures 2 and 3, respectively.

All three w/t constructs yield normally spliced RNA products as their most abundant splice product. Interestingly, w/t minigene 1 produces a second 463 bp splice product that lacks exon 3 and translates to p.H93_Q155del (ΔEx3). The p.H93_Q155del (ΔEx3) product encompasses an in-frame deletion of 62 amino acids from the FKBP-like domain of AIPL1 encompassing the region proposed to be critical for the interaction of AIPL1 with the farnesyl moiety. We did not observe exon skipping with the w/t minigene 2 and 3.

From Table 3, it can be seen that all the *AIPL1* variations investigated (with the exception of c.97_104dupGTGATCTT and c.98_99insTGATCTTG) affect splicing such that the native splice-site is not recognized, with either exon skipping, activation of cryptic splice-sites, or intron retention occurring. Partial recognition of the native splice-site in c.97_104dupGTGATCTT and c.98_99insTGATCTTG leads to premature translation termination resulting in p.F35Lfs*2 and p.I34Dfs*10, respectively. Both mutations also induce a frame-shifted prematurely terminated transcript coding for p.V33Sfs*57. The p.F35Lfs*2, p.I34Dfs*10 and p.V33Sfs*57 transcripts are all predicted to be subject to exon junction complex (EJC) nonsense-mediated mRNA decay (NMD).

Table 3. *AIPL1* Minigene Splice Products

Mutation (Minigene #)	Band #	Size, bp	Observed Splicing Effect	Predicted Protein	Suppl. Fig.
w/t (1)	1	652	w/t	w/t	n/a
	2	463	Exon 3 skipped	p.H93_Q155del (ΔEx3)	S2
c.97_104dup (1)	3	670	None - duplication retained (frame-shift)	p.F35Lfs*2	S3
	4	606	Native intron 1 splice acceptor-site not recognized, cryptic splice acceptor-site at c.142 activated initial 46 bp of exon 2 skipped (frameshift)	pV33Sfs*57	S4
c.98_99insTGATCTTG (1)	5	670	None - insertion retained (frame-shift)	p.I34Dfs*10	S5
	6	606	Native intron 1 splice acceptor-site not recognized, cryptic splice acceptor-site at c.142 activated - initial 46 bp of exon 2 skipped (frameshift)	pV33Sfs*57	S6
c.276+1G>A (1)	7	617	Native intron 2 splice donor-site not recognized - exon 2 skipped, cryptic 145 bp exon 3a inserted between exons 3 and 4 (frame-shift)	p.V33_I92del V156Efs*50	S7a & S7b
	8	472	Native intron 2 splice donor-site not recognized - exon 2 skipped	p.V33_I92del (ΔEx2)	S8
	9	283	Native intron 2 splice donor-site not recognized - exons 2 and 3 skipped	p. V33_Q155del (ΔEx2+3)	S9
c.276+2T>C (1)	10	656	Native intron 2 splice donor-site not recognized, cryptic splice donor-site 5 bp into intron 2 at c.276+5 activated - 4 bp of intron 2 retained (frame-shift)	p.H93Afs*66	S10
	11	472	Native intron 2 splice donor-site not recognized - exon 2 skipped	p.V33_I92del (ΔEx2)	S11
	12	283	Native intron 2 splice donor-site not recognized - exons 2 and 3 skipped	p. V33_Q155del (ΔEx2+3)	S12
c.277-2A>G (1)	13	463	Native intron 2 splice acceptor-site not recognized - exon 3 skipped	p.H93_Q155del (ΔEx3)	S13
c.465G>T (1)	14	439	Native intron 3 splice donor-site not recognized, cryptic splice donor-site at c.442 activated - last 24 bp of exon 3 skipped	p.V148_Q155del	S14
	15	463	Native intron 3 splice donor-site not recognized - exon 3 skipped	p.H93_Q155del (ΔEx3)	S15
w/t (2)	16	520	w/t	w/t	n/a
c.642G>C (2)	17	499	Native intron 4 splice donor-site not recognized, cryptic splice donor-site 22 bp into intron 4 at c.642+22 activated - 21 bp of intron 4 retained	p.K214N_E215insVRGRWPG	S16
	18	343	Native intron 4 splice donor-site not recognized - exon 4 skipped	p.V156_K214del (ΔEx4)	S17
w/t (3)	19	797	w/t	w/t	n/a
c.784G>A (3)	20	757	Native intron 5 splice donor-site not recognized, cryptic splice donor-site c.745 activated - last 40 bp of exon 5 skipped (frame-shift)	p.V249Afs*3	S18
	21	655	Native intron 5 splice donor-site not recognized - exon 5 skipped (frame-shift)	p.E215Afs*3	S19
c.785-10_786del (3)	22	1569	Native intron 5 splice acceptor-site not recognized - intron 5 not spliced out (frame-shift)	p.I263Afs*9	S20
	23	755	Native intron 5 splice acceptor-site not recognized, cryptic acceptor-site at c.826 activated - initial 42 bp of exon 6 skipped	p.G262_A275del	S21
	24	687	Native intron 5 splice acceptor-site not recognized, cryptic acceptor-site at c.894 activated - initial 110 bp of exon 6 skipped (frame-shift, read through into 3′UTR)	p.G262Efs*109	S22

Variations at the 5′ end of intron 2 (c.276+1G>A and c.276+2T>C) induce skipping of exon 2, leading to in-frame deletion of 60 residues from the FKBP-like domain, p.V33_I92-del (ΔEx2), and disruption of the putative farnesyl binding motif. Interestingly, skipping of exon 2 induced by c.276+1G>A and c.276+2T>C also couples with the native skipping of exon 3 to create a large 122-residue in-frame deletion of almost the entire FKBP-like domain, p.V33_Q155del (ΔEx2+3). The c.276+1G>A and c.276+2T>C variations also induce aberrant transcripts coding for p.V33_I92delV156Efs*50 and p.H93Afs*66, respectively, both of which are predicted to be degraded by NMD.

Interestingly, the c.277-2A>G variation at the 3′ end of intron 2 induces only a single aberrant transcript coding for p.H93_Q155del (ΔEx3). Similarly, the most abundant splice product induced by the c.465G>T variation skips exon 3 producing p.H93_Q155del (ΔEx3). The c.465G>T variation additionally induces transcription of a longer minor aberrant

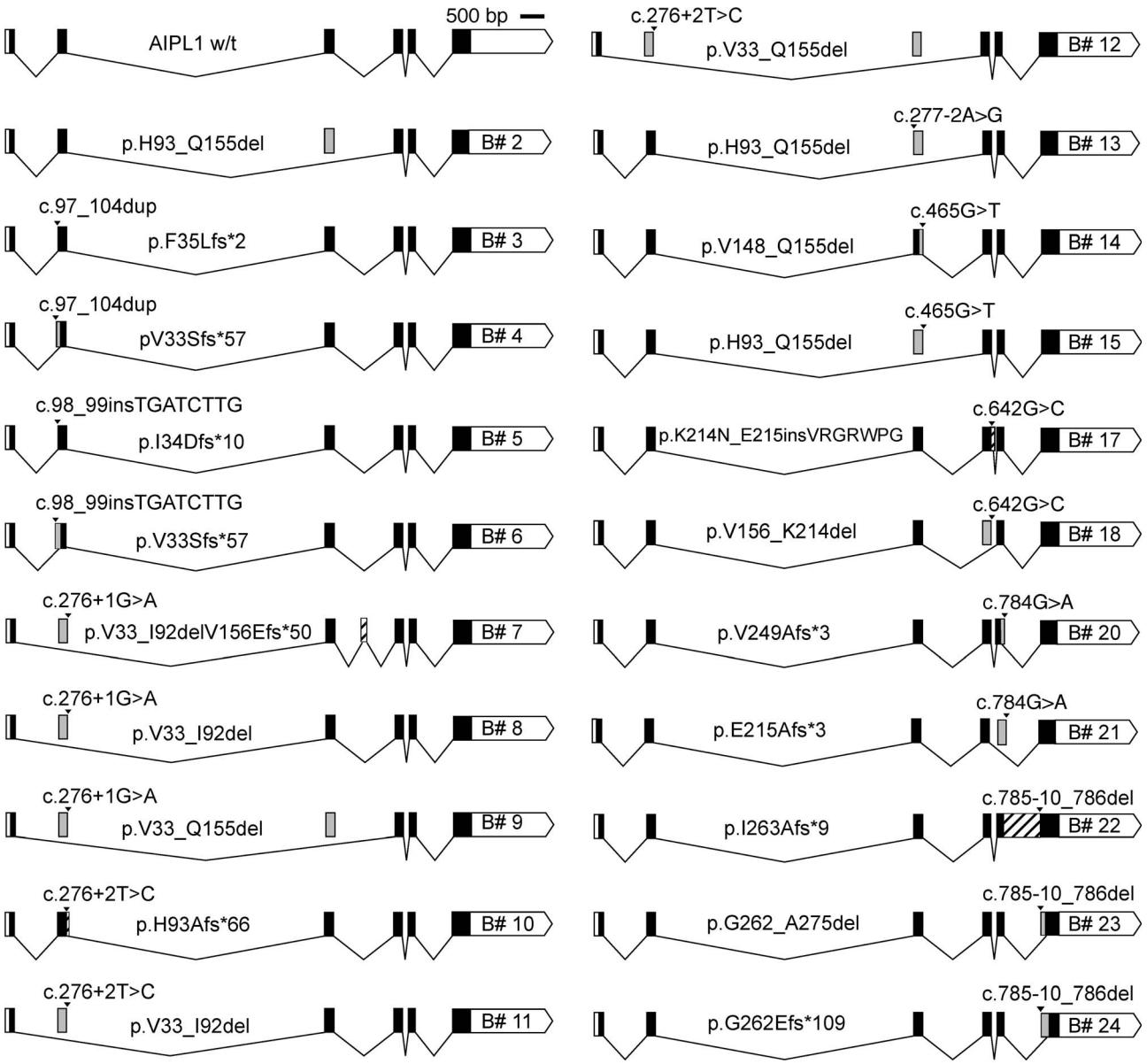

Figure 2. Splice site products identified by transcript analysis of *AIPL1* variations. The genomic structure of the longest *AIPL1* transcript (RefSeq NM_014336) is shown for w/t *AIPL1*. The transcripts are numbered according to the band number (B#2 to B#24) of the corresponding PCR amplicon in Figure 1C. The *AIPL1* variations are demarcated by an *arrowhead* according to their position in the *AIPL1* gene. The resultant predicted AIPL1 protein encoded by each transcript is shown above each transcript. Skipped coding sequence (deletions) and retained intron sequence (insertions) are demarcated by *gray* and *hatched bars*, respectively. The sequence data for each transcript is shown in Supplementary Figures S2 through S22. *Scale bar*: 500 bp.

splice product leading to an in-frame deletion of 8 residues in the FKBP-like domain at its C-terminus, p.V148_Q155del.

The c.642G>C *AIPL1* variation generates two aberrant splice products from minigene 2. The smaller 343 bp product skips exon 4 leading to the in-frame deletion of the last nine residues of the FKBP-like domain and the first TPR motif of the TPR domain, p.V156_K214del (ΔEx4). The larger 499 bp splice product results in the in-frame insertion of seven amino acids in the loop region connecting the first and second TPR motifs in the TPR domain, p.K214N_E215insVRGRWPG.

Two aberrant splice products are observed for the c.784G>A *AIPL1* variation in minigene 3, both of which result in a frame-shift and premature translation termination (p.V249Afs*3 and p.E215Afs*3). The less abundant 655 bp transcript coding for p.E215Afs*3 is predicted to be cleared by NMD. The more abundant 757 bp transcript producing p.V249Afs*3 may escape NMD resulting in C-terminal truncation of AIPL1 including half of the TPR domain and the entire polyproline-rich domain (PRD).

The c.785-10_786del mutation in minigene 3 generates three aberrant splice products that lead to premature termination (p.I263Afs*9), a 14 amino acid deletion in the TPR domain (p.G262_A275del), and a frame-shift read-through into the 3′UTR (p.G262Efs*109), respectively.

Alternative Transcription of AIPL1 in Normal Retina

Analysis of the RNA-seq dataset publically available through Illumina's Human BodyMap 2.0 project confirmed that *AIPL1* is

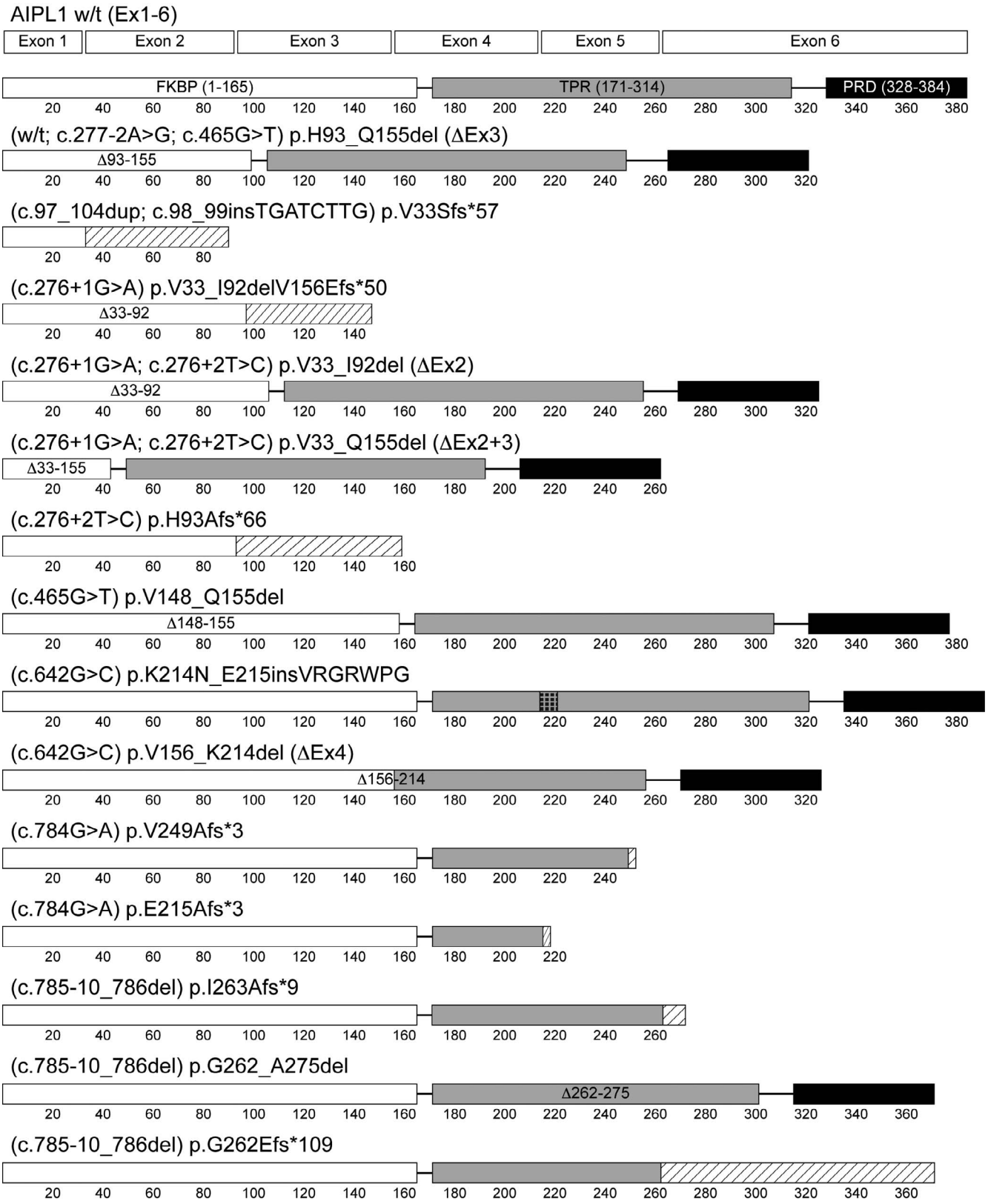

Figure 3. AIPL1 protein isoforms encoded by alternative *AIPL1* transcripts. *White bars*: AIPL1 amino-terminal FKBP-like domain (residues 1–165). The putative farensyl binding motif encompasses residues 89 to 147. *Gray bars*: carboxy-terminal TPR domain (residues 171–314). The TPR domain encompasses three consecutive TPR motifs, TPR1 (178–211), TPR2 (230–263), and TPR3 (264–297). Each TPR motif consists of a pair of antiparallel α helices (helix A and helix B), and the contiguous series of antiparallel α helices pack against one another to form a chaperone binding channel. *Black bars*: primate-specific polyproline-rich domain (PRD) (residues 328–384). *Hatched bars*: inclusion of nonnative AIPL1 sequence induced by a frame-shift variation. *Bar filled with small squares*: in-frame AIPL1 insertion. The transcripts encoding the splice products p.V33Sfs*57, p.I34Dfs*10, p.F35Lfs*2, p.V33_I92delV156Efs*50, p.H93Afs*66, and p.E215Afs*3 are predicted to be degraded by NMD.

Investigative Ophthalmology & Visual Science

expressed specifically in the retina and pineal gland, and is not expressed in any of the 16 tissue types included in the dataset, including blood, brain, and testes. Annotation of *AIPL1* on the Ensembl database (in the public domain, http://www.ensembl.org) shows 11 alternative *AIPL1* transcripts, one of which is subject to NMD (AIPL1-007; transcript ID: ENST00000381128). The normally spliced 6 exon wild-type transcript is annotated as AIPL1-001 (transcript ID: ENST00000381129), while a splice variant lacking exon 3 is annotated as AIPL1-002 (transcript ID: ENST00000250087). However, the relative abundance and physiological relevance of these alternative transcripts is unknown. Analyses of RNA-seq data from three retina-derived datasets[23] confirmed that alternative splicing of the *AIPL1* gene occurs in normal retina. Quantitative multisample visualization of the three independent mRNA sequencing reads aligned to gene annotations (Sashimi plot) using IGV-2.3.40 (Broad Institute)[24,25] revealed that differential splicing of the *AIPL1* gene normally includes skipping of exon 2 (ΔEx2), exon 3 (ΔEx3), and exon 4 (ΔEx4; Supplementary Fig. S1). The transcript abundance from all three datasets was estimated at 87% (95% confidence interval [CI] = ± 0.7) for the full-length transcript (NM_014336.3), 9.7% (95% CI = ± 1.0) for the ΔEx3 transcript (NM_001033054.1), and 3.4% (95% CI = ± 0.4) for the ΔEx2 transcript (NM_001033055.1). Skipping of exon 4 (ΔEx4) appears to be a rare event and was detected in only one of the three independent datasets with the greatest coverage. Of note, the two splice products detected from w/t minigene 1 in the in vitro splicing assay correspond to the two most frequently observed RNA-seq transcripts.

Discussion

This study describes the detailed investigation of uncharacterized exonic and intronic *AIPL1* variations predicted to alter the normal splicing of the *AIPL1* gene in LCA patients.[7,16] A caveat of our findings is the limitations inherent in the in vitro approach, however in the absence of patient derived RNA or protein, the in vitro minigene splicing assays have proven vitally important for validating the theoretical predictions and supporting the clinical findings. Our data demonstrates that all nine *AIPL1* mutations investigated cause aberrant pre-mRNA splicing to produce transcripts predicted to be degraded by NMD and/or encode functionally deficient protein isoforms. Our findings thus confirm that aberrant alternative transcription of the *AIPL1* gene may be an underlying cause of LCA.

Our study is the first to show that the *AIPL1* missense mutations c.465G>T [p.(Q155H)], c.642G>C [p.(K214N)] and c.784G>A [p.(G262S)] alter *AIPL1* splicing, a novel finding as alternative transcription from missense mutations is frequently overlooked as a potential disease mechanism. The *AIPL1* variation c.784G>A [p.(G262S)] is considered a possible rare benign variant.[7,11] However, our in vitro splice assay revealed that transcription of c.784G>A [p.(G262S)] yields only p.E215Afs*3 and p.V249Afs*3 variants and no p.G262S. While the p.E215Afs*3 transcript is expected to be degraded by NMD, the more abundant transcript producing p.V249Afs*3 may escape NMD, resulting in truncation of the chaperone-interacting TPR domain and the loss of the PRD. Our new findings support the prediction that c.784G>A [p.(G262S)] is a rare loss-of-function disease-associated mutation. Similarly, both c.465G>T [p.(Q155H)] and c.642G>C [p.(K214N)] missense mutations alter *AIPL1* transcription in vitro yielding aberrant splice products that disrupt the domain organization of the AIPL1 protein. In all cases, correctly spliced transcripts were not detected. Thus, a complex disease mechanism acting through the disruption of protein function by aberrant splicing rather than a simple amino acid substitution exists for these three missense mutations.

Our analysis also reveals that alternative transcription of *AIPL1* occurs normally in the retina, producing transcripts that lack exon 2, 3, or 4 as a result of deletion of these 0-0 phase exons. Interestingly, several of the disease-associated *AIPL1* variations investigated in this study also produced transcripts lacking exon 2, 3, or 4. Both c.276+1G>A and c.276+2T>C induced in-frame skipping of exon 2 to produce p.V33_I92del (ΔEx2), c.277-2A>G and c.465G>T [p.(Q155H)] induced in-frame skipping of exon 3 to produce p.H93_Q155del (ΔEx3), and c.642G>C [p.(K214N)] induced skipping of exon 4 to produce p.V156_K214del (ΔEx4). Notably, in-frame skipping of exon 3 was also detected as a minor transcript from wild-type *AIPL1* in the in vitro splice assay, and a low abundance of the transcript was also detected in vivo in normal retina. Therefore the splice variants c.277-2A>G and c.465G>T [p.(Q155H)] shift the relative abundance of this transcript in retina. Interestingly, the c.465G>T [p.(Q155H)] mutation was recently identified as a homozygous mutation in all affected members of a consanguineous family diagnosed with autosomal recessive retinal degeneration, with heterozygous carriers being unaffected.[15] It is therefore possible that there is a tolerance threshold for expression of this naturally occurring alternatively spliced *AIPL1* transcript, and that the expression of this transcript in the homozygous state may be associated with a less severe disease phenotype. Similarly, there may normally be a tolerance threshold for expression of the p.V33_I92del (ΔEx2) and p.V156_K214del (ΔEx4) alternative transcripts.

Analyses of the predicted impact of aberrant splicing on the AIPL1 protein suggest that all of the *AIPL1* variations investigated in this study are invariably loss-of-function mutations. To summarize (Fig. 3), alternative aberrant transcripts of *AIPL1* that shift the open reading frame and lead to early premature termination are predicted to be degraded by NMD and are loss-of-function mutations. However, alternate aberrant transcripts of *AIPL1* that involve in-frame exon skipping, and therefore in-frame deletion of specific domains within the AIPL1 protein escape NMD, and are likely to exhibit functional deficits associated with the domain-specific deletion. In-frame skipping of exon 2 or exon 3 disrupts the FKBP-like domain of AIPL1 leaving the TPR domain intact, whilst skipping of exon 4 primarily disrupts the TPR domain leaving the FKBP-like domain intact. The observation that skipping of exons 2, 3, and 4 normally occurs in the retina, albeit a rare occurrence, suggests that low levels of expression of alternative *AIPL1* transcripts lacking exons 2, 3, or 4 are normally tolerated in the retina. However, our data suggest that the exclusive or abundant expression of alternative *AIPL1* transcripts lacking exons 2, 3, or 4 as a result of splice site variations lead to pathogenesis. Variants affecting the splicing of exons 5 and 6 of *AIPL1*, which are out-of-phase, result in truncation or disruption of the TPR domain and PRD leaving the FKBP-like domain intact. Because the PRD and a significant proportion of the TPR domain are encoded by exon 6, many of the resultant transcripts likely escape NMD to express a faulty protein unable to interact with the molecular chaperones.

In conclusion, our data has identified aberrant transcription of *AIPL1* as a potential underlying cause of LCA. The analysis has been important in solving the molecular mechanism of disease in LCA patients harboring these variations and has increased our understanding of the role of aberrant RNA processing as a cause of LCA associated with variations in *AIPL1*. This opens up the possibility that a number of previously uncharacterized *AIPL1* variations of unknown pathogenicity may cause aberrant splicing of *AIPL1*. Moreover, unidentified variations in the promoter, untranslated regions,

cis-acting elements or regulatory elements in the introns or exons, such as splicing enhancers and silencers, may affect gene expression, RNA stability, or splicing and contribute to disease. It is therefore possible that the prevalence of LCA attributed to *AIPL1* mutations may be higher than previously estimated. Confirmation of this proposal would require further investigation and screening of *AIPL1* in patients with retinal degeneration and LCA, in combination with in-depth investigations of the effect of newly identified *AIPL1* variations at the transcript and protein level.

Acknowledgments

The authors thank Vincent Plagnol (UCL Genetics Institute, London, UK) for assistance with the RNA-seq data analysis. They also thank Alison Hardcastle (UCL Institute of Ophthalmology, London, UK) for critical review of the manuscript.

Supported by grants from Moorfields Eye Hospital Special Trustees (London, UK), Rosetrees Trust (Edgware, Middlesex, UK), Fight for Sight (London, UK), Retinitis Pigmentosa Fighting Blindness (Buckingham, Buckinghamshire, UK).

Disclosure: **J. Bellingham**, None; **A.E. Davidson**, None; **J. Aboshiha**, None; **F. Simonelli**, None; **J.W. Bainbridge**, None; **M. Michaelides**, None; **J. van der Spuy**, None

References

1. Weleber RG, Francis PJ, Trzupek KM, Beattie C. Leber Congenital Amaurosis. 2004 [cited 19/12/2014&rsqb. In: *GeneReviews [Internet]*. Seattle (WA) University of Washington, Seattle. Available at: http://www.ncbi.nlm.nih.gov/books/NBK1298/. Accessed November 27, 2015.
2. Sohocki MM, Bowne SJ, Sullivan LS, et al. Mutations in a new photoreceptor-pineal gene on 17p cause Leber congenital amaurosis. *Nat Genet*. 2000;24:79–83.
3. van der Spuy J, Kim JH, Yu YS, et al. The expression of the Leber congenital amaurosis protein AIPL1 coincides with rod and cone photoreceptor development. *Invest Ophthalmol Vis Sci*. 2003;44:5396–5403.
4. Hidalgo-de-Quintana J, Evans RJ, Cheetham ME, van der Spuy J. The LCA protein AIPL1 functions as part of a chaperone heterocomplex. *Invest Ophthalmol Vis Sci*. 2008;49:2878–2887.
5. Kolandaivelu S, Huang J, Hurley JB, Ramamurthy V. AIPL1, a protein associated with childhood blindness, interacts with alpha-subunit of rod phosphodiesterase (PDE6) and is essential for its proper assembly. *J Biol Chem*. 2009;284:30853–30861.
6. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet*. 2014;133:1–9.
7. Tan MH, Mackay DS, Cowing J, et al. Leber congenital amaurosis associated with AIPL1: challenges in ascribing disease causation, clinical findings, and implications for gene therapy. *PLoS One*. 2012;7:e32330.
8. Stone EM. Leber congenital amaurosis - a model for efficient genetic testing of heterogeneous disorders: LXIV Edward Jackson Memorial Lecture. *Am J Ophthalmol*. 2007;144:791–811.
9. Ramamurthy V, Roberts M, van den Akker F, Niemi G, Reh TA, Hurley JB. AIPL1, a protein implicated in Leber's congenital amaurosis, interacts with and aids in processing of farnesylated proteins. *Proc Natl Acad Sci U S A*. 2003;100:12630–12635.
10. Kanaya K, Sohocki MM, Kamitani T. Abolished interaction of NUB1 with mutant AIPL1 involved in Leber congenital amaurosis. *Biochem Biophys Res Commun*. 2004;317:768–7673.
11. van der Spuy J, Cheetham ME. The Leber congenital amaurosis protein AIPL1 modulates the nuclear translocation of NUB1 and suppresses inclusion formation by NUB1 fragments. *J Biol Chem*. 2004;279:48038–48047.
12. Gallon VA, Wilkie SE, Deery EC, et al. Purification, characterisation and intracellular localisation of aryl hydrocarbon interacting protein-like 1 (AIPL1) and effects of mutations associated with inherited retinal dystrophies. *Biochim Biophys Acta*. 2004;1690:141–149.
13. Bett JS, Kanuga N, Richet E, et al. The inherited blindness protein AIPL1 regulates the ubiquitin-like FAT10 pathway. *PLoS One*. 2012;7:e30866.
14. Majumder A, Gopalakrishna KN, Cheguru P, Gakhar L, Artemyev NO. Interaction of aryl hydrocarbon receptor-interacting protein-like 1 with the farnesyl moiety. *J Biol Chem*. 2013;288:21320–21328.
15. Li D, Jin C, Jiao X, et al. *AIPL1* implicated in the pathogenesis of two cases of autosomal recessive retinal degeneration. *Mol Vis*. 2014;20:1–14.
16. Aboshiha J, Dubis AM, van der Spuy J, et al. Preserved outer retina in AIPL1 Leber's congenital amaurosis: implications for gene therapy. *Ophthalmology*. 2015;122:862–864.
17. den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat*. 2000;15:7–12.
18. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucl Acids Res*. 2003;31:3812–3814.
19. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Meth*. 2010;7:248–249.
20. Brunak S, Engelbrecht J, Knudsen S. Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol*. 1991;220:49–65.
21. Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. *Nucl Acids Res*. 1996;24:3439–3452.
22. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997;4:311–323.
23. Farkas MH, Grant GR, White JA, Sousa ME, Consugar MB, Pierce EA. Transcriptome analysis of the human retina identify unprecedented transcript diversity and 3.5 Mb of novel transcribed sequence via significant alternative splicing and novel genes. *BMC Genomics*. 2013;14:486.
24. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotech*. 2011;29:24–26.
25. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14:178–192.