Deletions and Functional Assessment of Exonic Variants affecting Splicing in Genes underlying Inherited Retinal Dystrophies

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Elena Buena Atienza

aus Palencia, Spanien

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	27.11.2020
Stellvertretender Dekan:	Prof. Dr. József Fortágh
1. Berichterstatter:	Prof. Dr. Bernd Wissinger
2. Berichterstatter:	Prof. Dr. Thomas Lahaye

Erklärung

Ich erkläre hiermit, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

Deletions and Functional Assessment of Exonic Variants affecting Splicing in Genes underlying Inherited Retinal Dystrophies

selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich erkläre, dass die Richtlinien zur Sicherung guter wissenschaftlicher Praxis der Universität Tübingen (Beschluss des Senats vom 25.5.2000) beachtet wurden. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

Tübingen, den …… ……

Datum Unterschrift

A mis padres

Table of Contents

1 Summary 1

2 Zusammenfassung 2

3 Introduction 3

3.1 The retina 3

3.2 Inherited retinal dystrophies 4

3.2.1 Retinitis Pigmentosa 4

3.2.1.1 Autosomal Recessive Retinitis Pigmentosa 5

3.2.2 Cone Dysfunction Disorders 6

3.2.2.1 Blue Cone Monochromacy 7

3.3 Diagostics of Mendelian genetic disorders 10

3.3.1 Missing heritability 11

3.3.2 Variant interpretation challenge 13

3.4 Splicing 15

3.4.1 Splicing mechanism 15

3.4.2 Splicing regulation 17

3.4.3 Splicing and disease 18

4 Aims of Research 22

5 List of Publications 23

6 Results and Discussion 24

6.1 Publication I: A 73,128 bp *de novo* deletion encompassing the *OPN1LW/OPN1MW* gene cluster in sporadic Blue Cone Monochromacy: a case report **..... 24**

6.2 Unpublished I: A 142 bp intra-exonic deletion in both *OPN1LW* and *OPN1MW* underlying Blue Cone Monochromacy **..... 29**

6.3 Publication II: *De novo* intrachromosomal gene conversion from *OPN1MW* to *OPN1LW* in the male germline results in Blue Cone Monochromacy **..... 38**

6.3.1 Gene conversion event from *OPN1LW* to *OPN1MW* in family BCM72 38

6.3.2 Haplotype-dependent splicing regulation of *OPN1LW/MW* exon 3 44

6.4 Publication III: Retinitis pigmentosa: impact of different *Pde6a* point mutations on the disease phenotype **..... 55**

6.5 Unpublished II: Development of a multiplexed splicing assay **69**
6.5.1 Introduction 69
6.5.2 Results 71
6.5.3 Discussion 77
6.6 Publication IV: A comprehensive reference for exon 3 splicing defects of the human *OPN1LW* and *OPN1MW* opsin genes by means of a parallelized minigene splicing assay **79**
6.7 Unpublished III: Putative splicing factors binding *cis*-regulatory elements within exon 3 of *OPN1LW/MW* haplotypes **86**
6.7.1 Introduction 86
6.7.2 Results 87
6.7.3 Discussion 95
7 General discussion 100
8 Conclusions 105
9 Outlook 106
10 References 108
11 Abbreviations 121
12 Personal Contribution 123
13 Appendix 125
13.1 Publications 125
13.1.1 Publication I 125
13.1.2 Publication II 132
13.1.3 Publication III 152
13.1.4 Publication IV 173
13.1.5 Publication V 216
13.2 Supplementary Material 255
13.2.1 Supplementary material Introduction and Results 255
13.2.2 Supplementary material Section 6.2 255
13.2.3 Supplementary material Section 6.4 256
13.2.4 Supplementary material Section 6.7 256
14 Acknowledgements 260

1 Summary

The present dissertation comprises research about the molecular etiology of certain Mendelian disorders characterized by retinal malfunction and visual impairment. I performed a detailed investigation of *de novo* mutation events entailing genomic deletions or the conversion to pathogenic haplotypes in exon 3 of *OPN1LW* and *OPN1MW* underlying the occurrence of Blue Cone Monochromacy. Yet, the focus of this thesis goes beyond variant identification, including functional characterization to assess putative pathogenic variants and establish robust genotype-phenotype correlation. Exonic variants were assessed at the transcript level – more precisely, attending to the splicing efficiency in terms of exon retention or exon skipping. Namely, the exonic variant c.1684C>T/p.Arg562Trp in *Pde6a* induces in-frame exon skipping in >60% of murine retinal transcripts. The homologous missense mutation in a human RP patient exerts a second pathomechanism at the protein level, reducing the enzymatic activity of PDE6A. Comparisons with analogous mutations in ortholog and paralog genes underscore the potential influence of synonymous variants in splicing. In this sense, other than single exonic variants, I retrospectively characterized the effect of haplotypes confined to exon 3 of *OPN1LW* and *OPN1MW* in Blue Cone Monochromacy patients on splicing efficiency. Nine out of twelve haplotypes individually assessed by a semi-quantitative minigene splicing assay resulted in a fraction of ≥20% of aberrantly spliced transcripts. To explore the full breadth of exon 3 haplotype induced splicing defects I developed a parallelized minigene assay leveraging the newest sequencing technologies to quantify for more than 200 haplotypes both the extent of splicing defect exerted by each haplotype and the overall effect of each exonic variant within the haplotype. These experiments showed that c.532A>G and c.538T>G in exon 3 of *OPN1LW/MW* are the two variants with the highest impact on exon retention during transcript splicing. An RNA-pulldown assay including the haplotype region identified the hnRNPF splicing factor as a putative candidate, which binds to guanosine triplets created by the variants c.532A>G and c.538T>G.

2 Zusammenfassung

Diese Dissertation beschäftigt sich mit der Erforschung der molekularen Ätiologie bestimmter mendelscher Erkrankungen, die durch Netzhautfunktionsstörungen und Sehstörungen gekennzeichnet sind. Ich führte eine detaillierte Untersuchung von *de novo* Mutationsereignisse durch, die zu genomischen Deletionen oder zur Umwandlung in pathogene Haplotypen in Exon 3 von *OPN1LW* und *OPN1MW* führen und dadurch eine Blauzapfenmonochromasie verursachen. Der Fokus dieser Arbeit geht jedoch über die Identifizierung von genetischen Varianten hinaus und fokussiert sich auf die funktionale Charakterisierung putative pathogene (Splicing-) Varianten und der Etablierung einer robusten Genotyp-Phänotyp-Korrelation. Der Schwerpunkt der Untersuchungen galt der Wirkung exonische Varianten auf die Transkriptprozessierung, namentlich auf die Retention bzw. das Skipping des betroffenen Exons. So induziert die exonische Variante c.1684C>T/p.Arg562Trp im murinen *Pde6a*-Gen ein In-Frame-Exon-Skipping in >60% der Transkripte in der Mäusenetzhaut. Die homologe Missense-Mutation bei RP-Patienten verursacht zudem eine verminderte enzymatische Aktivität der PDE6A auf Proteinebene. Vergleiche zwischen analogen Mutationen in Ortholog- und Paralog-Genen belegen den potenziellen Einfluss synonymer Varianten auf den Spleißprozess. In diesem Sinne habe ich retrospektiv den Effekt von seltenen Variantenkombinationen (d.h. Haplotypen) in Exon 3 von *OPN1LW* und *OPN1MW* bei Blauzapfenmonochromasie Patienten auf ihre Wirkung bzgl. Transkriptsplicing untersucht. In neun von zwölf untersuchten Haplotypen, die individuell mittels eines semi-quantitativen Minigene Spleiß-*Assay* analysiert wurden, wurde ein Anteil von ≥20% anomal gespleißter Transkripte nachgewiesen. Um den Einfluß von Exon 3 Haplotypen auf das Splicing systematisch zu studieren, wurde in dieser Arbeit ein parallelisierte Minigen-Assay entwickelt und damit für mehr als 200 Haplotypen sowohl das Ausmaß des von jedem Haplotyp verursachten Spleißdefekts als auch den Gesamteffekt jeder exonischen Variante innerhalb des Haplotyps bestimmt. Die Varianten c.532A>G und c.538T>G in Exon 3 von *OPN1LW/MW* zeigten die größten Auswirkung auf die Exon-Retention. Mittels eines differentiellen RNA-Pulldown-Assays konnte der hnRNPF-Splicing-Faktor als möglicher Kandidat identifziert werden, der an Guanosin-Tripletts bindet, die durch die Varianten c.532A>G und c.538T>G gebildet werden.

3 Introduction

3.1 The retina

The human retina is the 0.5 mm thin layer at the back of the eye in charge of sensing and the **light stimuli** and converting these into electrical signals that are then transmitted to the brain and interpreted as visual images (Kolb, 1995a). Three layers of neuronal cell bodies alternate with plexiform layers with integrated axons and dendrites for synaptic contacts. The design of the retina is conserved across vertebrates; the outer nuclear layer comprises rods and cones; the inner nuclear layer comprises bipolar, horizontal and amacrine cells, and the ganglion cell layer mainly consists of ganglion cells but also displaced amacrine cells in several species (Kolb, 1995a; Hoon et al., 2014). Rods and cones, the two types of photoreceptors in the retina, are at the head of the **phototransduction** cascade; they sense and convert the light into changes of the membrane potential, and in turn modulate the quantity of neurotransmitter release. **Rods** are approximately 20 times more abundant than cones in the human retina (Osterberg, 1937). About 17,500 **cones** are packed at the center of a depression within the macula. This is due to the displacement of the inner retinal layers in an area of 0.3 mm in diameter in the central rod-free fovea which enables to reduce blur from light scattering, constituting the spot of highest resolution in the eye (Kolb, 1995b). Cones are optimized for daytime achromatic and chromatic vision, or luminance and colour, respectively, while rods exhibit a superior sensitivity and are specialized for vision under dim-light conditions with limiting photon flux (Schultze, 1866; Sabesan et al., 2016; Ingram et al., 2016). The light-sensitive **photopigments** are embedded in the outer segment membranes of photoreceptors and consist of an **opsin**, a 7-transmembrane G protein-coupled receptor class protein, and a covalently attached **chromophore**, the vitamine A derivate 11-*cis*-retinal. Whereas rods express a uniform photopigment, rhodopsin with at absoption maximum at λ=~500nm, the three types of cone photoreceptors present in the human retina are distinguished by the expression of photopigments with distinct light absorption maxima at λ=564 nm (**long-**wavelength sensitive cone photopigment in L-cones), at λ=533 nm (**middle-**wavelength sensitive cone photopigment in M-cones, and at λ=437 nm (**short-**wavelength sensitive cone photopigment in S-cones) (Kolb, 1995c). The

differences in the absorption spectrum relate to differences in the amino acid sequence of the opsin moiety of the photopigments and its physico-chemical interaction with the chromophore. Absorption of a **photon** by 11-*cis*-retinal triggers its photoisomerization to all-*trans*-retinal, which causes the opsin to undergo a series of conformational changes (Wald, 1968) that results in activation of a **signalling cascade** involving a G-protein (transducin), a cyclic guanosine monophosphate (cGMP)-specific phosphodiesterase and a cyclic nucleotide-gated channel (CNG channel). Thereby, excitation of the photopigments causes the decrease in cGMP levels, a subsequent closure of CNG channels and membrane hyperpolarization (Karpen et al. 1988). This change in **membrane potential** decreases glutamate release of the photoreceptor synapse and serves as excitation signal for sign-preserving OFF- or sign-reverting ON-bipolar cells (Fu, 1995).

3.2 Inherited retinal dystrophies

Inherited retinal dystrophies (IRDs) are a **clinically** highly **heterogenous** group of diseases which overall affect approximately 1 in 3000 people (Bundey and Crews, 1984; Chiang and Trzupek, 2015). IRDs can be generally classified into syndromic and non-syndromic, congenital or degenerative, primarily affecting rods or cones, and therefore impairing predominantly peripheral or central vision, respectively. The complexity of IRDs is documented by its high **genetic diversity**, overlapping phenotypes between distinct diseases and shared etiology gene locus for distinct clinical entities. So far, mutations in 271 genes causing IRDs have been identified, as listed in the Retinal Information Network database (RetNet, https://sph.uth.edu/retnet/; last accessed in June 2020). In the following, I will introduce the clinical and genetic features of two IRD groups: Retinitis Pigmentosa and Cone Dysfunction Disorders.

3.2.1 Retinitis Pigmentosa

The prevalence of Retinitis Pigmentosa (RP) is estimated to range between 1 in 3000 and 1 in 7000 individuals (Haim et al., 2002) worldwide. RP is primarily caused by a defect in **rod** photoreceptors where outer segments are severely diminished. At advanced stages, impairment of cones and central vision occurs, being RP a **progressive** disease (Berson, 1993). RP patients suffer from impaired dim-light vision which develops into night blindness (nyctalopia) and the progressive peripheral vision loss leads to construction of visual fields; known as **tunnel** vision. A characteristic RP

fundus presents bony spicule pigmentation and thinning of vessels (Verbakel et al., 2018). In some patients cataracts and macular edema are found associated. RP is characterized by inheritance and genetic **heterogeneity** (Verbakel et al., 2018). Non-syndromic RP can be inherited as an autosomal-dominant, autosomal recessive, X-linked, and in rare instances as uniparental (Rivolta et al., 2002) or digenic inherited trait (Kajiwara et al., 1994). Unknown inheritance is largely common for simplex cases, representing sporadic cases the majority (~40-50%) of RP cases. Autosomal recessive RP (arRP) accounts for ~20-30% of cases. Autosomal dominant RP (adRP) explains about 10-25% of cases whereas X-linked RP is found in approximately 6-18% of RP cases depending on the study and population (Boughman et al., 1980; Fishman, 1978; Haim, 2002; Hu, 1982; Jay, 1982).

3.2.1.1 Autosomal Recessive Retinitis Pigmentosa (arRP)

Autosomal recessive forms account for approximately 20% of all non-syndromic RP cases (Daiger et al., 2007) and can account for up to ~60% of simplex cases (Bravo-Gil et al., 2017). The onset and course of retinal degeneration is highly variable and may range from early onset in childhood (Dvir et al., 2010) to late onset in the adulthood (Ávila-Fernández et al., 2011; Arno et al., 2017). Hitherto, mutations in 63 genes have been identified to cause arRP (RetNet; last accessed in June 2020). Most identified genes underlying arRP account for a modest or small percentage of cases. The *USH2A* gene is the most frequently mutated gene in arRP with up to 10% of all cases (Seyedahmadi et al., 2004). Whilst **Usherin** has a structural function, several other arRP-causing genes encode for key proteins of the phototransduction cascade, such as the alpha (A) and beta (B) subunits of the ***rod cGMP-gated channel****, CNGA1* (Dryja et al., 1995) and *CNGB1* (Bareil et al., 2001), respectively, or the two catalytic subunits of the ***rod cGMP Phosphodiesterase 6***; *PDE6A* (Huang et al., 1995) and *PDE6B* genes (McLaughlin et al., 1993), respectively. Mutations in *PDE6A* and *PDE6B* account for 4% of all arRP cases each (Dryja et al., 1999; McLaughlin et al., 1995) whereas mutations in the gene encoding for the inhibitory gamma subunit of the rod PDE6 (*PDE6G*) have been less frequently described (Dvir et al., 2010). Under physiological conditions, rod PDE6 catalyzes **cGMP to GMP hydrolysis** in rod photoreceptor outer segments (ROS) upon light stimulation (Stryer, 1986). The reduction in cGMP concentration induces the cyclic-nucleotide-gated (CNG) cation channels to close (Karpen et al., 1988), resulting in membrane **hyperpolarization** and cease of glutamate release at the rod synaptic terminal (see 3.1). When PDE6 activity

is lacking (e.g. by the presence of biallelic mutations in either *PDE6A* or *PDE6B*), ROS cGMP concentration rises above physiologic levels and produces a constant influx of sodium and calcium cations through CNG channels which – in chronic state – is toxic and results in the death of **photoreceptor cells** (Farber and Lolley, 1974). Mice strains with naturally-occurring mutations in *Pde6b* (Bowes et al., 1990); for instance the *rd1* (Keeler, 1924) and the *rd10* (Chang et al., 2007), have been extensively used to model human *PDE6B*-associated RP and study the molecular and cellular mechanism of rod degeneration. Similarly, for *PDE6A*-associated RP, mutagenesis-induced *Pde6a* **mouse models** have been generated including the *nmf282* strain (Sakamoto et al., 2009) carrying a missense mutation fully homologous to the c.2053G>A/p.Val685Met mutation in the human *PDE6A* identified in RP patients (Corton et al., 2010) and a model of the latter mutation in compound heterozygous state with an ultra-rare variant (see Results and Discussion, Section 6.4).

3.2.2 Cone Dysfunction Disorders

Cone dysfunction disorders, sometimes coined as cone dysfunction syndromes (Aboshiha et al., 2016), are **non-syndromic** inherited forms of retinal disorders caused by the lack or reduced function of all or certain types of **cone** photoreceptorers and characterized by reduced visual acuity, colour vision defects, photophobia and frequently nystagmus (Michaelides et al., 2004). Cone dysfunction disorders are usually present from birth and show a **stationary** profile in terms of functional impairment of **central vision** and preserved normal rod responses. High variability is encountered for colour vision abnormalities, nystagmus and photophobia. Cone dysfunction disorders represent a **heterogeneous** group that can be subcategorized into autosomal recessive inherited forms including complete and incomplete forms of achromatopsia, oligocone trichromacy and bradyopsia (Michaelides et al., 2004; Aboshiha et al., 2016), and X-linked recessive inherited forms including Bornholm Eye Disease (BED) and Blue Cone Monochromacy (BCM). While patients with BCM are severely visually handicapped (see below), BED patients do present with residual cone function and dichromacy (protanopia or deuteranopia) and myopia (Haim et al., 1988; McClements et al., 2013). The **genetic etiology** of BCM and BED – collectively termed X-linked cone dysfunction (XLCD) disorders – lies within the same genetic locus: the cluster of the human opsin 1, long- and middle-wavelength sensitive genes (*OPN1LW* and *OPN1MW,* respectively), as described below (Nathans et al., 1986; Nathans et al., 1989).

3.2.2.1 Blue Cone Monochromacy

BCM is characterized by the **absence** of **red** and **green** cone function; relying vision exclusively on blue cones and rods (Alpern et al., 1965). Central vision is impaired with varying degrees of colour vision disturbances among affected individuals based on residual cone function (Nagy and Boynton, 1979; Smith et al., 1983) and/or due to illumination conditions during the test. BCM patients present with **reduced visual acuity** (Green, 1972), which hinders basic tasks such as facial recognition or reading, often accompanied by light sensitivity (**photophobia**), uncontrolled eye movements (**nystagmus**) and association with short-sightedness (**myopia**) albeit not always present (Aboshiha et al., 2016). Although mainly stable, some cases with progressive deterioation of visual function have been reported in young adult males (Ayyagari et al., 1999; Kellner et al., 2004; Michaelides et al., 2005). Full-field electroretinography (ERG) recordings in BCM shows reduced or completely absent responses under photopic (light-adapted) conditions whereas scotopic (dark-adapted) responses are preserved (Kellner and Foerster, 1992). Diverse colour vision tests are available. The Berson plate test (Berson et al., 1983) relies on the tritan discrimination of BCM subjects to distinguish these from achromats.

BCM is an **X-linked** recessive congenital disorder that affects approximately 1 in 100,000 males and has a theoretical frequency of 1 in 10 billion females, (Sharpe et al., 1999). BCM affects males due to their X-chromosome **hemizygosity** whereas the inactivation of one X-chromosome that occurs randomly in females usually compensates for a monoallelic mutation on one of the X-chromosomes. However, severely skewed X-inactivation in female carriers can lead to clinically manifest BCM (Frederiksen et al., 2013). BCM is caused by mutations in the **human cone opsin gene cluster** (Nathans et al., 1989). The *OPN1LW* and *OPN1MW* genes encode for the apo-proteins of the human long-wavelength and middle-wavelength sensitive cone photoreceptor pigments, respectively (Nathans et al., 1986). The resulting apo-proteins in conjunction with the covalently linked chromophore *cis*-retinal absorb light with a maximum peak sensitivity at 564 nm and 533 nm , respectively (Kolb, 1995c). The *OPN1LW/MW* gene cluster is located at the tip of the long (q) arm of the X-chromosome at Xq28. Arranged in a **head-to-tail tandem array** (Vollrath et al., 1988), the *OPN1LW* gene copy occupies the centromeric (or proximal) position followed (and separated by ~24 kilobases (kb) of intergenic sequence) by one or several *OPN1MW* gene copies telomeric (or distal) to the *OPN1LW* gene copy. Both *OPN1LW* and

OPN1MW genes comprise six exons (1,092 bp of translated coding sequence) or 1258 bp and 1239 bp cDNA sequence including the 5' and 3' untranslated region (UTR), and five introns with a total length of 14,000 bp and 12,036 bp, for the *OPN1LW* and *OPN1MW* genes, respectively (Sharpe et al., 1999). *OPN1LW* and *OPN1MW* genes share an **identity of 98%** both for coding and non-coding nucleotide sequence and 96% in amino-acid sequence. The existence of separate *OPN1LW* and *OPN1MW* genes presumably originated from a single ancestral gene by a gene **duplication** event in the Old World primate lineage ~35 million years ago (MYA) after the divergence of New and Old World monkeys (Nathans, 1987) as an adaptation to detect mature fruits against the green leaf foliage (Mollon, 1989).

High variability has been observed for the number of *OPN1MW* gene copies among subjects with normal colour vision. A study on human retinal *OPN1LW/MW* transcripts showed that only the **two most proximal** gene copies in the array – regardless of its sequence nature, are expressed above the limit of detection (Winderickx et al., 1992). An upstream regulatory element, the **locus control region** (LCR), governs expression of the *OPN1LW/MW* gene cluster in a mutually exclusive and distance-dependent manner by interacting with the promoter of either the *OPN1LW* or a *OPN1MW* gene copy (Peng and Chen, 2011; Wang et al., 1992; Winderickx et al., 1992). In a single cone the *OPN1LW* and *OPN1MW* gene promoters compete for binding the LCR which determines which type of opsin is expressed and thus, **cone identity** (Smallwood et al., 2002).

Genetic heterogeneity is characteristic of BCM and XLCD disorders (Nathans et al., 1993) yet to a lesser extent if compared with other IRDs. Two main categories of mutation (Figure 1) causing BCM were already described in the seminal paper of Nathans and colleagues (Nathans et al., 1989). Genomic **deletions** encompassing the LCR and/or parts of or the entire *OPN1LW/MW* gene cluster (Fig. 1A) constitute the major cause of BCM (30-40%, depending on the population) (Nathans et al., 1989, 1993; Ayyagari et al., 1999; Kellner et al., 2004; Gardner et al., 2014). The second most frequent mutational pathway (Fig. 1A) comprises a **two-step** mechanism: first, unequal homologous recombination results in the formation of a gene cluster reduced to a single *OPN1LW* or a single *OPN1LW/MW* which is subsequently inactivated by the **c.607T>C/p.Cys203Arg** point mutation in exon 4 (Nathans et al., 1989, 1993).

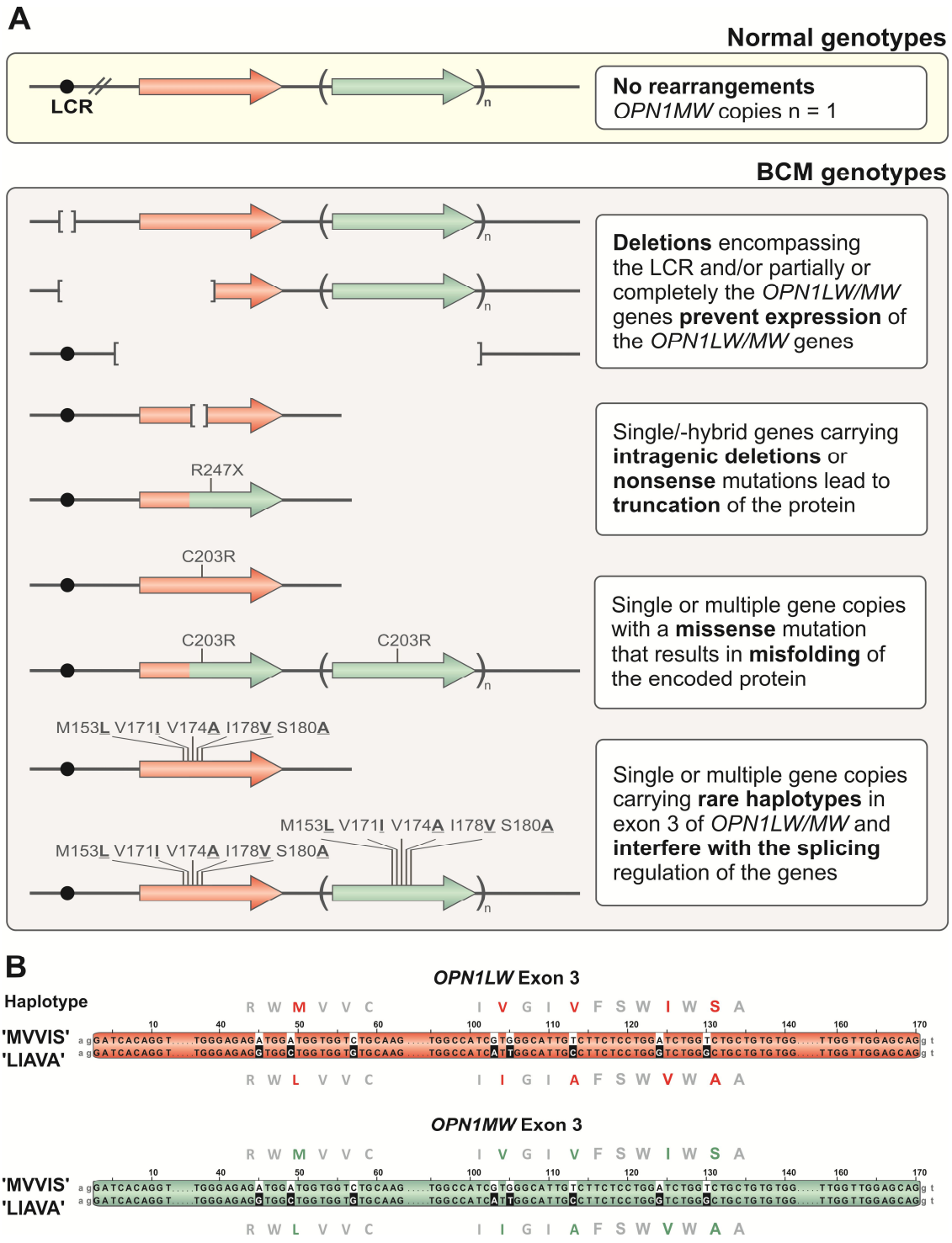


Figure 1. A) Causative known mutations underlying BCM. The top box shows the genotype of an unaffected individual without rearrangements and with ≥ 1 *OPN1MW* copies. The bottom box summarizes the mutational mechanisms found in BCM subjects: LCR deletions, deletions encompassing the *OPN1LW/*MW genes, intragenic deletions, nonsense and missense mutations, and rare haplotypes in exon 3 of *OPN1LW/MW*. **B) Exon 3 of *OPN1LW* and *OPN1MW*** showing the haplotype 'MVVIS' differing at c.(453A > G; 457A > C; 465C > G; 511G > A; 513G > T; c.521T>C; 532A > G; 538T > G) from the haplotype 'LIAVA' (referring to the amino acid change). Exonic variants are spread out along exon 3 and the two outermost varians are located ≥40 bp away from the splice sites. Figure 1A has been adapted from an original drawing of Prof. Bernd Wissinger.

In addition, Reyniers and collegues reported a BCM family where affected males carry the c.607T>C/p.Cys203Arg in both *OPN1LW* and *OPN1MW* copies (Reyniers et al., 1995). Similarly a family with a missense mutation, c.529T>C/p.Trp177Arg in all gene copies within the *OPN1LW/MW* gene array has been reported (Gardner et al. 2010). Both missense mutations, the c.607T>C/p.Cys203Arg as well as the c.529T>C p.Trp177Arg has been shown to disrupt folding of the mutant protein. A rare nonsense mutation, c.739C>T/p.Arg247Ter in exon 4 leading to truncation of the protein (Nathans et al., 1993) and an intragenic exon 4 deletion mutation (Ladekjær-Mikkelsen et al., 1996) have been described in subjects with gene clusters reduced to a single gene copies. The third most frequent mutational pathway is the occurence of combinations of single nucleotide polymorphisms (SNPs) in exon 3 which are non-deleterious individually, but associated with XLCD disorders if present in a certain rare combination, i.e. haplotype (Nathans et al., 1989). This **exon 3 haplotype** (Fig. 1B) comprising the following SNP variants: c.(453A > G; 457A > C; 465C > G; 511G > A; 513G > T; 521C; 532A > G; 538T > G) and thus, enconding for a cone photopigment protein with the amino acid exchanges: p.[(=); M153**L**; (=); V171**I**; A174**A**; I178**V**; S180**A**]; abbreviated as **'LIAVA'** (referring to the amino acid changes), has been repeatedly observed in subjects with BCM and BED (Crognale et al., 2004; Mizrahi-Meissonnier et al., 2010; McClements et al., 2013).

Not until recently, a study involving protanope males has shed light into the pathomechanism underlying the *OPN1LW* exon 3 haplotypes. The 'LIAVA' haplotype induces a **splicing defect** through skipping of exon 3 as shown by an *in vitro* assay (Ueyama et al., 2012). The splicing mechanism, its regulation as well as its impact on Mendelian diseases will be introduced in Section 3.4.

3.3 Diagostics of Mendelian genetic disorders

Next-generation sequencing (NGS) approaches are well-suited to enbale and fasten the comprehensive screening of larger sets of genes from several individuals for causative gene(s) and/or variant(s) identification. The advent of NGS has led to **high rates** of genetically diagnosed cases of IRDs. Due to its cost-effectiveness, NGS-based gene panels and whole exome sequencing (WES) are currently widely used for genetic diagnostics in IRDs. Targeted sequencing is based on the enrichment of a single or multiple defined genomic regions (e.g. genes of interest or coding exons), whereas whole genome sequencing (WGS) allows to explore the entire genome

including non-coding and regulatory regions. The choice of a particular genetic test scheme or NGS approach is determined by several factors (i.e. clinical and genetic knowledge of a given condition, the number of known causative genes, accessibility to familiar samples and inheritance pattern among others). Yet, NGS-based technologies relying on **short-read** sequencing are not suitable to detect copy number variation (CNV), or to cover repetitive sequences and regions within gaps of the genome assembly.

The genetic testing currently performed for BCM is **target-specific, unique and distinct** from the testing approach for other IRDs. While for instance known achromatopsia-causing genes can be scanned by NGS-based methods, BCM mutational screening has not benefited from NGS so far. Short-read sequencing and applied algorithms are not sufficient to unveil the complexity and discern between copies of the human *OPN1LW/MW* gene cluster. A meticulous unique opsin cluster-specific genotyping strategy customized for the *OPN1LW/MW* gene array based on Polymerase Chain Reaction (PCR), restriction fragment length polymorphism (RFLP) and Sanger sequencing (for details see Appendix 13.2.1, Supplementary Material; Supplementary Figure S1) is continuously optimized (Cideciyan et al., 2013) to screen patients with XLCD disorders or male patients with a clinically compatible phenotype at the Molecular Genetics Laboratory at the Centre for Ophthalmology, University of Tübingen. Such an optimized and specific test is unavailable at most other diagnostic and research centers. Still, this test does not fully resolve the complexity of the *OPN1LW/MW* gene cluster. This genotyping strategy was applied for this thesis (Sections 6.1, 6.2, 6.3 and 6.6) together with additional genotyping tools established to decipher specific unsolved cases.

3.3.1 Missing heritability

A **sensitivity** of 50-70% is achieved with targeted NGS and WES in IRDs (Eisenberger et al., 2013; Glöckle et al., 2014; Weisschuh et al., 2016). Besides patients presenting with acquired disease phenocopying IRDs, or mutations in as yet unknown disease genes, there is growing evidence that there is a decent fraction of still to be discovered *hidden* mutations (in known disease genes) underlying IRDs (Bauwens et al., 2019; Zeitz et al., 2019). For instance, by means of the custom screening protocol (see above) developed for BCM, about 11% of the BCM patients of the Tübingen cohort have no known mutation assigned yet (Dr. Susanne Kohl,

personal communication). Such *hidden* mutations could be difficult to detect or to interpret as disease-causing (Girirajan, 2017). **Structural** genomic alterations of intermediate to large size have probably been miscaptured (Chaisson et al., 2019). The so-called *dark matter* of the genome includes **non-coding** variants such as deep-intronic mutations (Bauwens et al., 2015) and regulatory variants. Moreover, **synonymous** variants in the coding sequence are usually neglected, although they may have a pathogenic impact, for instance by causing a splciing defect. In addition, ***de novo*** germline and somatic mutations are troublesome to detect.

There are two main categories of limitations most likely underlying the missing hereditability *problem*: i) methodology, and ii) interpretation. First, the methodological caveats can be i.e. associated with the limitations of short-reads and the low precision and recall to detect structural variants, or with the inappropriate application of existing methodologies (WGS, RNASeq, long or linked-read sequencing). Secondly, the greatest challenge is the limitation or uncertainty in the interpretation of variants. Even if detected, disease-causing variants might be overlooked or misinterpreted during prioritization and lead to inconclusive or wrong classification.

Novel methodologies are being implemented to tackle these limitations. For IRDs, **CNV detection** algorithms are now being implemented into the data analyis of NGS panels (Ellingford et al., 2017), or assessed separately by quantitative real-time PCR and microarray-based comparative genomic hybridization (Mayer et al., 2017; Van Cauwenbergh et al., 2017). Other recent technologies (and their combination) may bear the potential to further increase diagnostic sensitivity: bioinformatic algorithms based on linked reads have shown to identify difficult structural variants such as large inversions (Eslami Rasekh et al., 2017). In another instance, the combination of long-read and RNA sequencing successfully enabled the identification of a deletion in a patient with neoplasia and cardiac myxomata negative in prior targeted NGS and WGS analysis (Merker et al., 2018). These cases underscore the necessity of combining or applying several different and/or appropiate methodologies. Typically, variant assessment is performed from the genomic perspective without being able to consider for variant effects at the **transcript level**. Transcriptome analysis as a complementary tool for genomics in the diagnosis of Mendelian disorders has been shown to be useful with a success rate of 10% (Kremer et al., 2017) and 35% (Cummings et al., 2017) for mitochondrial and muscle disorders, respectively. Notably, RNASeq can aid to interpret coding and non-coding variants, detecting expression

outliers, mono-allelic expression of rare heterozygous variants, somatic variants and aberrant splicing by direct sequencing of newly spliced isoforms. RNASeq-assisted diagnosis is favoured for rare diseases for which **biopsies** from the affected tissue are available, i.e. muscle, liver, kidney and skin among others (Cummings et al., 2017). However, RNA from patients is often not available or the gene(s) of interest is only expressed in a tissue inaccessible for routine biopsy (e.g. retina for IRDs). In such cases, **reprogramming** easily accessible patient cells (from skin punches, blood or hair follicles) into induced Pluripotent Stem Cells (iPSCs) and their differentiation into cells expressing the gene of interest may overcome these limitations (Chuang et al., 2017; Parfitt et al., 2016). Techniques to evaluate the impact of variants on transcripts of unaccessible tissues in a higher throughput are needed.

3.3.2 Variant interpretation challenge

As sequencing technologies improve, variant discovery rates accelerate while resources to aid interpretation evolve only moderately. Assessment of variant pathogenicity currently relies mostly on the following resources or criteria: allele frequency in population datasets, literature search, amino acid conservation and algorithms to predict the impact on the protein function and on transcript splicing. Shortcomings are apparent, such as subjective selection and weighting of criteria for variant interpretation and contradictory results. In order to overcome this problem, the American College of Medical Genetics and Genomics (ACMG) together with the Association for Molecular Pathology (AMP) and the College of American Pathologists have conveyed variant classification guidelines applicable for Mendelian genes (Richards et al., 2015). This rule-based scoring system set common criteria for **standardization** in the categorization of variants (Richards et al., 2015). A certain variant is assigned a weighted score from the five-tier system: (i) pathogenic (score = 7), (ii) likely pathogenic (score = 6), (iii) uncertain significance (scores = 3, 4 and 5), (iv) likely benign (score = 2), and (v) benign (score = 1), (scores in brackets from Karbassi et al., 2016). The lines of evidence include different criteria with their own limitations and/or challenges: (1) **segregation** analysis; confirmed paternity and maternity can be weighted for *de novo* variants, identified variant may segregate because it is in linkage disequilibrium (LD) with the true pathogenic variant (2) **co-occurrence** of a variant with a pathogenic mutation – depending on phasing – only for dominant and X-linked in males but not if the condition has multiple genetic contributors (Nykamp et al., 2017), (3) **population and disease-specific databases**

such as dbSNP, 1000 Genomes, Exome Aggregation Consortium (ExAC), Genome Aggregation Database (gnomAD), Leiden Open Variation Database (LOVD) and Human Gene Mutation Database (HGMD) from which the allele frequency is retrieved inevitably contain misclassified variants; the single threshold of allele frequency used, 5% of the population, does not reflect the variable likelihood of a variant being benign, since this is higher as its observed frequency becomes higher (Nykamp et al., 2017); and founder mutations or variants in mutational hotspots have a higher allele frequency. (4) **Literature** search which has caveats due to single observations, nomenclature inconsistence and redundant data from the same families and finally, (5) ***in silico*** tools to assess nucleotide conservation and predict damaging consequences of coding variants to the resultant protein are considered as minor evidence for instance: PolyPhen-2 (Adzhubei et al., 2010), MutationTaster (Schwarz et al., 2010), SIFT (Kumar et al., 2009), or to predict effects on the the processing of the pre-mRNA such as Human Splicing Finder (HSF) (Desmet et al., 2009), NNSplice (Reese et al., 1997), SROOGLE (Schwartz et al., 2009).

Guidelines for interpreting variants putatively leading to **splicing defects** (Caminsky et al., 2014) have been extended to include non-canonical splice site mutations in both coding and non-coding regions (Ohno et al., 2018). Integration of several different *in silico* tools is recommended provided they complement each other in terms of the analyzed feature(s). Notably, synonymous exonic variants are classified as "**Variants of Uncertain Significance**" (VUS) provided that prediction tools suggest a possible impact on splicing (Richards et al., 2015) which renders a lavish number of variants awaiting characterization at the RNA level. **Functional** data obtained from the direct analysis of tissue biopsies from the patient or from an animal model, or, alternatively, by well-established *in vitro* assays that reflect the biological function, is considered as strong evidence (Richards et al., 2015), albeit the latter one is not sufficient if no other supporting classifiers are met. Comparison of *in silico* tools and *in vitro* minigene assays showed some discrepancies for certain variants (Sharma et al., 2014). Notwithstanding, when combining *in silico* tools with an *in vitro* minigene assay, ~5% of all identified VUS in the autosomal dominant cardiomyopathy genes *LMNA* and *MYBPC3*, were shown to alter splicing (Ito et al., 2017). Recently, 23 VUS found in 26 patients with IRDs were validated as non-canonical splicing mutations by means of *in vitro* minigene splicing assays (Soens et al., 2017).

3.4 Splicing

3.4.1 Splicing mechanism

The **spliceosome** is a multi-megadalton large and dynamic ribonucleoprotein complex (Wahl and Lührmann, 2015a; Zhang et al., 2017) that carries out the pre-mRNA splicing reaction in eukaryotic cells, mostly exclusively confined to its nucleus (Brody and Abelson, 1985). The spliceosome **excises** the intronic non-coding sequence (Gilbert, 1978) from pre-mRNA molecules, and subsequently, joins the adjacent coding regions, the exons (Gilbert, 1978), to generate functional mature mRNAs (Fig. 2A) translated into proteins by the ribosomes in the cytoplasm or untranslated long non-coding RNAs (Papasaikas and Valcárcel, 2016). Intron excision is preceded by spliceosome recognition and extensive base pairing at the exon-intron junction; between spliceosomal components and the pre-mRNA and within sequences of the pre-mRNA substrate. Consensus sequences, notably the **5' splice site** (5'ss) and **3' splice site** (3'ss) with invariable 'GU' and 'AG' dinucleotides (Fig. 2A and Fig. 1; Appendix 13.1.5), respectively, and the **branch point** and the **polypyrimidine tract** (PPT) which are located in the introns are critical (Reed and Maniatis, 1985).

Two consecutive transesterification reactions mediate intron excision and exon ligation (Fig. 2B). First, a nucleophilic attack is carried out by the 2'OH group of a branch-point adenosine residue on the phosphate group of the 5' terminal nucleotide of the intron. The resultant 2'-5' phosphodiester bond forms a intermediate product called lariat (Ruskin et al., 1984). In a second step, the free 3'OH group of the terminal nucleotide of the upstream exon reacts with the phosphate group of the phosphodiesterase bond between 3' terminal nucleotide of the intron and the 5' terminal nucleotide of the downstream exon which results in the joining of the two exons, and the release of the intron lariat (Scotti and Swanson, 2015; Wahl and Lührmann, 2015a). The spliceosome assembles *de novo* for each pre-mRNA splicing event in a stepwise manner. The major U2-type and the minor U12-type spliceosome are responsible for the splicing of U2-type introns (mainly with GU-AG as 5' and 3' terminal dinucleotides; Wahl and Lührmann, 2015b), which account for more than 99.5% of introns, and the minor U12-type introns (mainly with AU-AC as 5' und 3' terminal nucleotides), respectively (Turunen et al., 2013). The major spliceosomal building blocks, the U1, U2, U4, U5, and U6 **small nuclear ribonucleoproteins** (snRNPs), small nuclear RNA-protein complexes (Krainer and Maniatis, 1985) that assemble on their substrate

targets during the course of pre-mRNA **transcription** (Herzel et al., 2017) *via* interplay with the RNA polymerase II (Zeng and Berget, 2000).

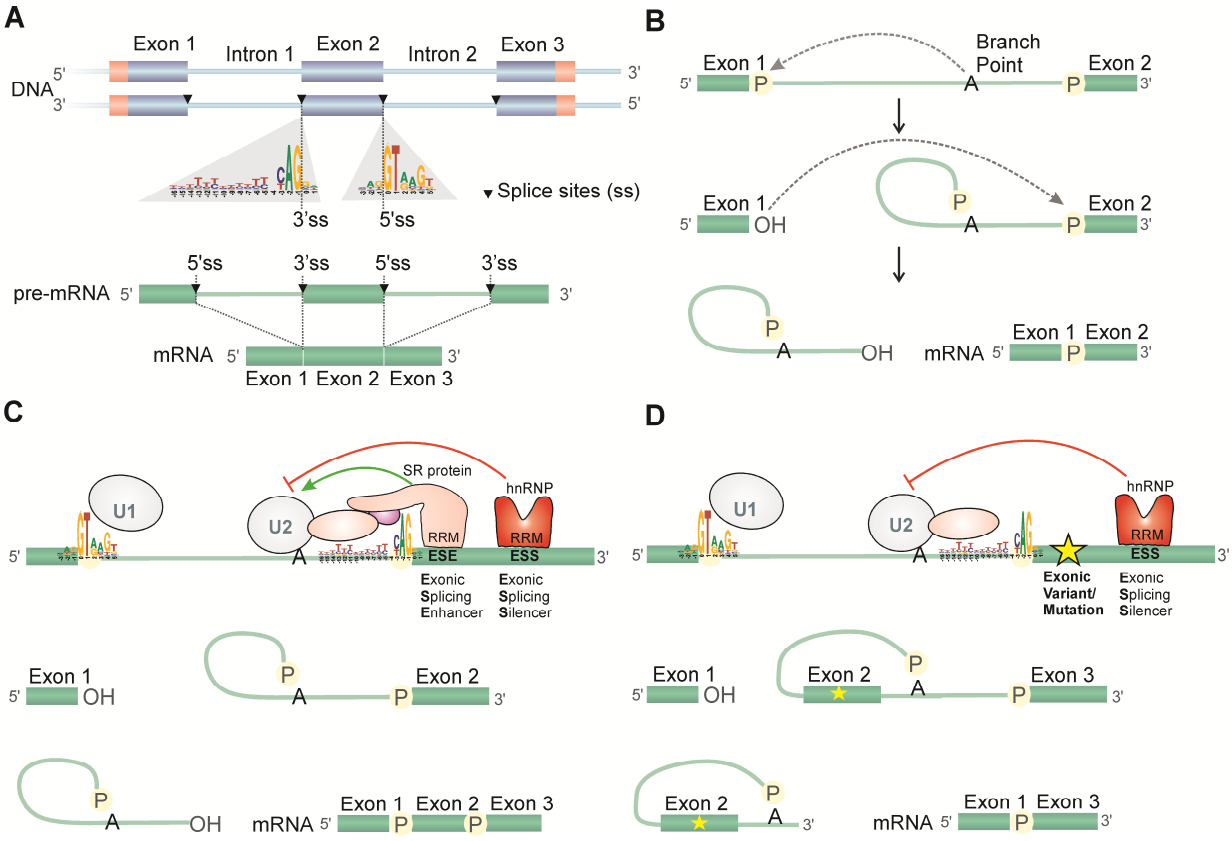


Figure 2. Splicing reaction. **A**) Exons are joined together at the splice splices while introns are excised out from the pre-mRNA molecule to form a mature mRNA during the splicing reaction (Scotti and Swanson, 2015). The consensus sequences of the 5' splice site (5'ss) and 3' splice site (3'ss) were generated using Weblogo (Crooks et al., 2004). For detailed branch point and the polypyrimidine tract (PPT) sequences see our review (Fig. 1; Weisschuh et al., 2020; Appendix 13.1.5; Publication V). **B**) Two main steps result in the joining of the exons and the excision of the intronic sequence in between as a lariat structure (Scotti and Swanson, 2015; Wahl and Lührmann, 2015a). **C**) Core spliceosomal proteins (i.e. U1 and U2 snRNAs) and splicing factors (SR and hnRNP proteins) influence the coordination of the splicing reaction by the recognition of canonical (splice sites, branch point, PPT) and non-canonical (i.e. Exonic Splicing Enhancers, ESE or Silencers, ESS) *cis*-elements (Scotti and Swanson, 2015; Wahl and Lührmann, 2015b). **D**) Exonic variants within *cis*-regulatory elements might impair the splicing reaction (Wahl and Lührmann, 2015c); a mutation that disrupts an ESE leading to skipping of exon 2 is illustrated. This is an own generated illustration that summarizes and adapts content from the literature (Cartegni et al., 2002; Scotti and Swanson, 2015; Wahl and Lührmann, 2015a, 2015b, 2015c).

The U1 snRNPs recognize the 5'ss (Roca et al., 2013) and the auxiliary proteins Splicing Factor 1 (SF1; Krainer and Maniatis, 1985) and U2AF (Ruskin et al., 1988) bind the branch point, the PPT and the 3'ss. The U1 and U2 snRNPs interact reciprocally to conceive the pre-spliceosome which recruits the pre-assembled snRNP U4, U5 and U6 (Matera and Wang, 2014). Multiple rearrangements, conformational

changes and RNA helicases are required to generate a catalytically active complex. Upon release of U4 and U1 snRNPs, the active complex undertakes the first catalytic step. The second catalytic step between a 3'OH of a free exon and the intron–exon lariat intermediate is preceded by rearrangements of the complex (Zhang et al., 2017). The **intron lariat** and **spliced exons** constitute the post-spliceosomal complex. Finally, the U2, U5 and U6 snRNPs are liberated to be used in a new splicing reaction (Matera and Wang, 2014).

3.4.2 Splicing regulation

Splicing reactions are subjected to fine regulation ensuring efficiency, fidelity and flexibility. The RNA polymerase II processivity is faster in introns than in exons (Jonkers and Lis, 2015). Transcription of (human) introns expends a greatest deal of time and cost. Thus, transcription and removal of introns with single base pair **accuracy** must be crucial if this apparently *inefficient* mechanism has been kept (Matera and Wang, 2014; Papasaikas and Valcárcel, 2016). Chromatin may facilitate the exon-intron arrangement by wrapping approximately the average length of a human internal exon in a nucleosome (Herzel et al., 2017). Moreover, co-transcriptional splicing ensures a reduction of exposed splice sites on the nascent RNAs (Herzel et al., 2017), although introns can be post-transcriptionally spliced upon signal recognition (Braunschweig et al., 2013).

The spliceosome reliably **distinguishes** the *bona fide* splice sites from *decoy* sites with consensus sequences similar to the authentic sites. Illegitimate splice sites within large introns are temporally used during recursive splicing (Georgomanolis et al., 2016). Besides faithful recognition, the spliceosome concurrently guarantees the **pliability** in splice site selection required for alternative splicing. Regulation of splicing leverages further mechanisms beyond the recognition of 5′ and 3′ ss, the branch site and PPT (De Conti et al., 2013). Plenty of regulatory sequence elements within the pre-mRNA substrate recruit and bind concomitantly and consecutively different composites of ***trans*-acting** splicing regulatory factors. Such redundancy contributes to a higher fidelity during constitutive splicing. The malleability for **alternative splicing** (Gallego-Paez et al., 2017) is usually tuned by expression of splice factors in a stimuli-, development-, or tissue-dependent fashion (Kornblihtt et al., 2013). The RNA-protein interaction is mediated through splicing factors which are RNA-binding proteins

(RBPs) with domains displaying RNA-recognition motifs (**RRMs**) (Lunde et al., 2007; Ray et al., 2013).

There are two major groups of *trans*-acting RNA-binding factors: serine/ariginine-rich (**SR**) proteins (Liu et al., 1998) and heterogenous nuclear ribonucleoproteins (**hnRNPs**) (Gallinaro et al., 1981). While SR proteins recruit the splicing machinery to splice sites, hnRNPs prevent elements of the spliceosome to recognize particular splice sites (Fig. 2C). The *cis*-acting regulatory sequences can be classified attending to their location into **exonic** and **intronic** regulatory sequences of splicing. Likewise, *cis*-acting elements can be classified according to their activities into splicing **enhancers** or **silencers**, distinguishing further into exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE) and intronic splicing silencer (ISS) sequences (Cartegni et al., 2002). This thesis focuses on *cis*-acting exonic elements. SR proteins mostly bind to ESE motifs (Liu et al., 1998; Jobbins et al., 2018) whereas hnRNPs recognize ESS to inhibit splicing. Both SR and hnRNPs interfere with the binding of U1 and U2 snRNPs to the substrates at the splice sites (Wahl and Lührmann, 2015b) during or after assembly of the spliceosomal machinery (Fig. 2C). For instance, multiple ESEs are involved in the inclusion of the alternative terminal ORF15 exon in the retina (Vervoort et al., 2000) and promote extensive diversity of alternatively spliced *RPGR* transcripts (Hong and Li, 2002).

3.4.3 Splicing and disease

Genetic variations inducing **aberrant** pre-mRNA splicing are frequently involved in hereditary diseases in humans. Approximately 25-30% of disease-causing mutations are estimated to affect pre-mRNA splicing (López-Bigas et al., 2005; Sterne-Weiler et al., 2011). Given the paucity in variant interpretation and the lack of systematic analysis of intronic variants, these numbers may still be an underestimation. However, the frequency of disease causing mutations affecting pre-mRNA splicing can vary considerably among genes (Wang and Cooper, 2007; Soukarieh et al., 2016). Mutations affecting the functionality of ***trans*-acting factors** of the splicing machinery are detrimental for the expression and proper splicing of a plethora of RNA substrates and thus, many of these mutations are most likely lethal (Singh and Cooper, 2012). An imbalance in the stoichiometry of recruited splicing factors can occur due to repetitive elements, i.e. RNA transcribed from a microsatellite expansion can result in a RNA gain of function and major modifications in the availability or stability of a given splicing

factor (Daguenet et al., 2015). Variants affect ***cis*-acting regulatory elements** by compromising the specificity or the fidelity of splice site choice, by activating cryptic splice sites (Daguenet et al., 2015), by reversing the activity of the recruited *trans*-acting factor or by impairing its recruitment (Wahl and Lührmann, 2015c) as illustrated in Fig. 2D. Mutations altering the canonical dinucleotides at the 5′ and 3′ ss do almost exclusively cause aberrant splicing (Roca et al., 2013; Weisschuh et al., 2020; Appendix 13.1.5). Although most exonic variants affect conserved positions at the 5′ and 3′ ss (i.e. 9 and 14 bases, respectively), about 30% of exonic variants are predicted to involve ESE loss and/or ESS gain (Mort et al., 2014). Mutations in exons affecting splicing might be nonsense, missense or silent (Cartegni et al., 2002).

Identification of variants altering *cis*-regulatory elements of splicing can be achieved by **direct transcript** analysis derived from patient's samples carrying a putative splicing mutation. Alternatively in the absence of RNA samples form the patient, a so-called minigene splicing assay may be used to assess the effect of variants on splicing (Cooper, 2005). The exon-trapping vector pSPL3, originally developed for identification of coding regions, is a commomly used construct for minigene splicing assays (Burn et al., 1995). For assessing the impact of an exonic variant, the exon harbouring the variant, the flanking introns, and optimally, the contiguous flanking exons to provide for endogenous donor and acceptor splice sites are inserted into the construct. **Minigenes** represent a reliable and convenient choice to assess the effect of VUS on splicing (Gaildrat et al., 2010) and to validate *in silico* predictions (Desviat et al., 2012). Straightforward results are obtained when i.e. correct exon inclusion is observed for the minigene carrying the wild-type allele but aberrantly spliced transcripts for the mutant constructs (Desviat et al., 2012), yet different outcomes might be found depending on basal inclusion levels of a given exon.

Mutations disrupting *cis*-acting regulatory elements of splicing and the counterpart *trans*-acting splicing factors have been well studied for the autosomal recessive inherited neuromuscular disorder **Spinal Muscular Atrophy** (SMA) which is a paradigm for disruption of exon definition and splicing regulation. Mutations in the *Survival Motor Neuron 1* gene (*SMN1*) resulting in an insufficient amount of the SMN protein and lead to SMA. *SMN1* and its paralogue, *SMN2*, share 99% homology and lie within a 500 kb inverted duplication (Monani et al., 1999). Most of the *SMN2* transcripts are unstable due to skipping of exon 7 as a consequence of the c.840C>T **synonymous** variant (Lorson et al., 1999) which disrupts an SF2/ASF-dependent

ESE (Cartegni and Krainer, 2002) and creates a new ESS binding site for hnRNPA1 (Kashima et al., 2007). Correction of the exon skipping by antisense oligonucleotides or the expression of a sequence-adapted U1 snRNA to promote exon definition and mask silencers could be used as an indirect therapy for SMA by reactivating *SMN2*, the latent paralog of *SMN1* (Dal Mas et al., 2015; Rogalska et al., 2016).

Another well-studied example is **Cystic Fibrosis** (CF). Missplicing of *Cystic Fibrosis Transmembrane conductance Regulator* gene (*CFTR*) transcripts is observed in ~12% of all described mutations (Aissat et al., 2013). Notably, a quarter of all **synonymous** variants in *CFTR* exon 12 are promoting exon skipping (Pagani et al., 2003a, 2005) which results in a non-functional CFTR protein associated with CF phenotypes. The overlapping splicing regulatory elements which constitute the so-called Composite Exonic Regulatory Element of Splicing (CERES) entail virtually every nucleotide of *CFTR* exon 12 (Haque et al., 2010). In addition, high susceptibility for missplicing is linked to **weak definition** of the 3'ss of exon 9. As a consequence, 74% of single nucleotide variants (SNVs) in this exon cause alteration of splicing (Pagani et al., 2003b). Weak *cis*-regulatory motifs seem to increase the vulnerability of *CFTR* to exon skipping (Aissat et al., 2013).

Variants in the *BRCA1* and *BRCA2* genes confer high risk to develop Hereditary Breast and Ovarian **Cancer**, accounting for ~16% of the familial risk of breast cancer (Stratton and Rahman, 2008). Fifty percent of identified *BRCA1/2* variants are classified as VUS, which may represent a **reservoir** of putative splicing mutations. In a series of studies, aberrant splicing was observed in 54% of the tested 154 *BRCA1/2* variants (Sanz et al., 2010; Acedo et al., 2012, 2015; Fraile-Bethencourt et al., 2017, 2018), of which ~75% were classified as pathogenic or likely pathogenic (Fraile-Bethencourt et al., 2018). A G-to-T transversion in exon 18 of *BRCA1* at the same position as the aforementined c.840C>T variant in *SMN2* (+6 from the acceptor splice site) formally generates a **stop codon**. However, the c.840C>T variant creates a silencer-rich sequence which leads to an in-frame exon skipping (Goina et al., 2008). Thereby, aberrant splicing of *BRCA1* results in the removal of 26 amino acid residues from a highly conserved motif (Mazoyer et al., 1998).

Splicing deficiency and splicing defects also play an important role in IRDs. Intriguingly genes for *trans*-acting core spliceosomal proteins are an important category of **IRD disease genes**; namely (hypomorphic) mutations PRPF31, PRPF8, PRPF6, PRPF3,

PAP-1 and SNRNP200/BRR2, have been associated with adRP (Weidenhammer et al., 1997; Martínez-Gimeno et al., 2003; Waseem et al., 2007).

On the pre-mRNA substrate, a myriad of variants abolishing *bona fide* donor and acceptor splice sites in nearly all IRD genes have been reported. Intronic non-canonical splice site variants are located a few base pairs upstream (at the acceptor site, i.e. -14 to -3) or downstream of exons (at the donor site, i.e. +3 to +9). Exonic non-canonical splice site variants affect the first and two last nucleotides of an exon. Variants at *cis*-regulatory splicing elements other than the splice sites are one of the current challenges (Cremers et al., 2018). Both non-synonymous (Bellingham et al., 2015; Becirovic et al., 2016) and synonymous variants (Braun et al., 2013) as well as intronic variants located far away from the 5' and 3' ss, referred to as deep-intronic mutations (den Hollander et al., 2006; Mayer et al., 2016) have shown pathogenic potential by means of aberrant splicing of IRD genes. Misspliced transcripts with skipped exons, included pseudoexons or partially retained introns may lead to alteration of the reading frame and premature termination of the protein or in-frame deletions affecting essential protein domains (Weisschuh et al., 2020).

Extensively investigated for splicing mutations, *ABCA4* is an IRD gene with 50 exons. Splice site mutations accounted for 13.5% of *ABCA4* rare variants identified (n = 171) in 335 **Stargardt patients** (Schulz et al., 2017). Whereas most *ABCA4* alleles represented missense (59.6%) and nonsense and frameshift mutations (15.8%), synonymous variants accounted for 8.8% (Schulz et al., 2017). Non-canonical splice site variants represent 10% of all reported *ABCA4* alleles associated with Stargardt; being 75% at the donor site (Cornelis et al., 2017). Deep-intronic variants were found in 67% of Stargardt patients (n = 36) with one or no variant in *trans* (Sangermano et al., 2019). Exonic variants that altered splicing represented 28% (n = 28) of all putative splice-disrupting variants in 335 Stargardt patients (Schulz et al., 2017).

The main sub-cateogories of splicing mutations and the consequences have been illustrated with examples of mutations in IRDs genes attending to the splicing defect exerted and/or the position within the gene (Weisschuh et al., 2020; Appendix 13.1.5). An unprecedented example within IRDs is the exon skipping induced by the simultaneous presence of certain exonic variants (in *cis*; haplotypes) within exon 3 of *OPN1LW* in individuals with a protanopia (Ueyama et al., 2012). Investigating the misregulation of *OPN1LW/MW* exon 3 splicing a central topic of this thesis.

4 Aims of Research

i) To identify disease-causing mutations in unsolved BCM/XLCD cases and to elucidate the underlying mutational mechanisms at the *OPN1LW/MW* gene cluster

ii) To investigate splicing regulation confered by exonic motifs of IRD genes; to characterize either individual synonymous and non-synonymous mutations as well as haplotypes that potentially alter *cis*-regulatory elements of splicing with a pathogenic effect on the retinal phenotype

iii) To develop an assay with increased throughput compared with individual minigene splicings assay which should enable the screening of hundreds of putative exonic spliceogenic variants. The objective was to design an *in vitro* minigene assay in the form of a multiplexed assay leveraging massively parallel sequencing technologies

iv) To apply the established multiplexed assay for prospective assessment of exonic variants in the *OPN1LW/MW* genes, and specifically, to provide unbiased assessment and interpretation of the splicing outcome from 256 different exon 3 haplotypes

5 List of Publications

I **Elena Buena-Atienza**, Fadi Nasser, Susanne Kohl, and Bernd Wissinger
A 73,128 bp *de novo* deletion encompassing the *OPN1LW/OPN1MW* gene cluster in sporadic Blue Cone Monochromacy: a case report.
BMC Medical Genetics (2018) 19 (Appendix 13.1.1)

II **Elena Buena-Atienza**, Klaus Rüther, Britta Baumann, Richard Bergholz, David Birch, Elfride De Baere, Helene Dollfus, Marie T. Greally, Peter Gustavsson, Christian P. Hamel, John R. Heckenlively, Bart P. Leroy, Astrid S. Plomp, Jan Willem R. Pott, Katherine Rose, Thomas Rosenberg, Zornitza Stark, Joke B. G. M. Verheij, Richard Weleber, Ditta Zobor, Nicole Weisschuh, Susanne Kohl, and Bernd Wissinger. *De novo* intrachromosomal gene conversion from *OPN1MW* to *OPN1LW* in the male germline results in Blue Cone Monochromacy.
Scientific Reports (2016) 6, 28253 (Appendix 13.1.2)

III Vithiyanjali Sothilingam, Marina Garcia Garrido, Kangwei Jiao, **Elena Buena-Atienza**, Ayse Sahaboglu, Dragana Trifunović, Sukirthini Balendran, Tanja Koepfli, Regine Mühlfriedel, Christian Schön, Martin Biel, Angelique Heckmann, Susanne C. Beck, Stylianos Michalakis, Bernd Wissinger, Mathias W. Seeliger, and François Paquet-Durand. Retinitis pigmentosa: impact of different Pde6a point mutations on the disease phenotype.
Human Molecular Genetics (2015) 24, 5486–5499 (Appendix 13.1.3)

IV **Elena Buena-Atienza**, Marius Cosmin Codrea, Stefan Czemmel, Britta Baumann, Susanne Kohl, Sven Poths, Sven Nahnsen, Bernd Wissinger
A comprehensive reference for exon 3 splicing defects of the human *OPN1LW* and *OPN1MW* opsin genes by means of a parallelized minigene splicing assay
Manuscript in preparation (Appendix 13.1.4)

V Nicole Weisschuh, **Elena Buena-Atienza**, Bernd Wissinger
Splicing Mutations in Inherited Retinal Diseases
Manuscript accepted (Appendix 13.1.5)

6 Results and Discussion

6.1 Publication I: A 73,128 bp *de novo* deletion encompassing the *OPN1LW/OPN1MW* gene cluster in sporadic Blue Cone Monochromacy: a case report

This publication reports the genetic basis underlying a sporadic BCM case. The molecular genetic analysis revealed the presence of a novel genomic deletion of >73 kb encompassing the LCR, the entire *OPN1LW* and parts of the *OPN1MW* gene in the single affected male. Segregation analysis confirmed that this deletion is a *de novo* mutation which had occurred in the mother's germline. This study underscores the importance to consider mutations in the *OPN1LW/MW* gene cluster in male patients with cone dysfunction features regardless of the absence of disease history in the family.

While the patient (BCM17-III:2; Buena-Atienza et al., 2018) was initially diagnosed with **achromatopsia**, a detailed clinical evaluation of the retinal function provided evidence compatible with a clinical diagnosis of a classical **BCM** phenotype (Fig. 1; Appendix 13.1.1; Publication I). The genetic analysis (Appendix 13.2.1; Supp. Fig. S1) showed that the unaffected maternal grandfather (I:1) and mother (II:1) had an intact *OPN1LW/MW* gene cluster, yet the patient (III:2, Fig. 2; Appendix 13.1.1, Publication I) beared a large **novel** deletion spanning the LCR and the adjacent sequence including the entire *OPN1LW* and the first three exons of *OPN1MW*. Due to the **loss of the LCR** and any intact *OPN1LW* or *OPN1MW*, the deletion NC_000023.11:g.154,118,184_154,191,311del is considered a loss-of-function mutation (Nathans et al., 1989) responsible for the phenotype in this patient (BCM17-III:2). Microsatellite marker analysis demonstrated transmission of the *OPN1LW/MW* bearing X-chromosome segment from the grandfather (BCM17-I:1) to the patient (BCM17-III:2; Fig. 2a; Appendix 13.1.1, Publication I). We proposed a ***de novo*** origin of the deletion in the mother's (BCM17-II:1) germline because the deletion was not detected in neither the grandfather's (BCM17-I:1) nor the mother's (BCM17-II:1) blood DNA (Fig. 2c; Appendix 13.1.1, Publication I). A post-zygotically *de novo* origin in the patient cannot be formally ruled out. However, genotype and phenotype derived from peripheral blood DNA and retinal examinations, respectively, correlate well indicating that the mutation is not exclusively present in a specific tissue, as expected for a

mutation of somatic origin. The deletion is most likely constitutively present in the patient (BCM17-III:2), including gamete cells, and therefore is transmittable to the patient's offspring. Alternatively, mosaicism with a high fraction of hemizygous mutant cells due to an occurrence of the *de novo* event during the first cell divisions may be considered (Acuna-Hidalgo et al., 2016).

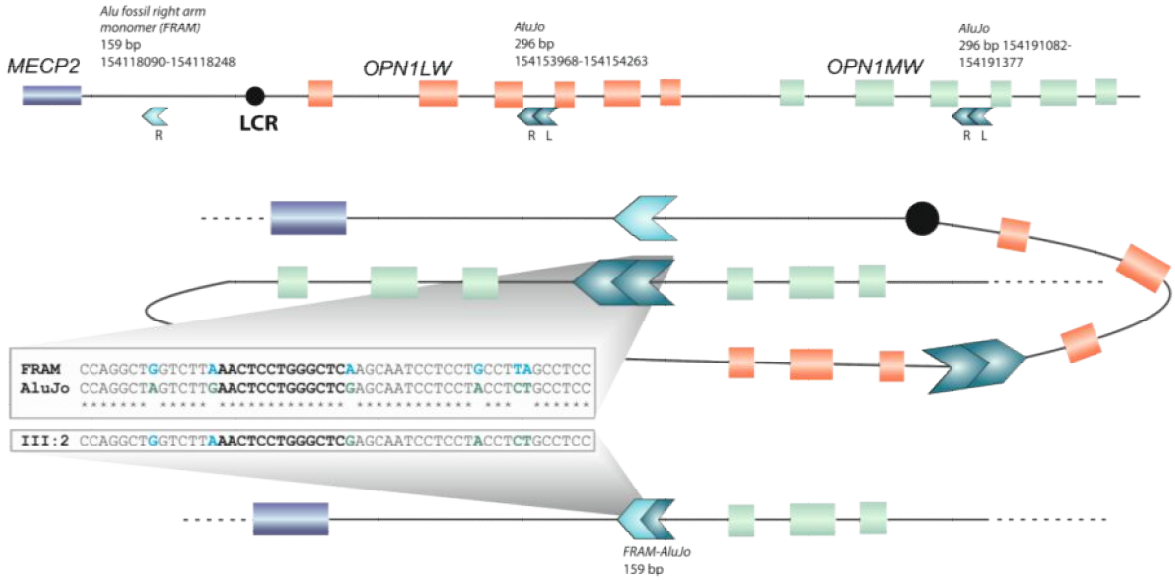


Figure 3. Genomic sequence context of the 73,128 bp *de novo* deletion found in BCM17-III:2. In the upper panel the Alu fossil right arm monomer (FRAM) is depicted as a light blue arrow in the intergenic region between *MECP2* (dark blue) and the LCR (black filled circle). *AluJo* right and left arms in intron 3 are represented by dark green arrows, in the same direction as FRAM (telomeric to centromeric). Exons of the *OPN1LW* and *OPN1MW* genes are depicted by red and green boxes, respectively. In the middle panel a scheme of the local architecture and the sequences likely involved in the recombination event is shown. An intrachromatid intrachromosomal (intrachromosomal between sister chromatids or interchromosomal could also be possible) scenario of non-allelic homologous recombination (NAHR) mediated by two Alu elements is depicted. The longest stretch homology (13 bp) is shown in bold black letters and specific nucleotides originating from FRAM (blue) and *AluJo* (dark green) are highlighted. Asterisks show perfect sequence matches between the two Alu elements. At the bottom, the hybrid FRAM-*AluJo* element, product of the recombination, and the *OPN1LW/MW* gene cluster of BCM17-III:2 lacking 73,128 bp of sequence, are shown.

The novel junction created by the deletion is embedded within two **Alu elements** (Fig. 2d; Appendix 13.1.1, Publication I). FRAM and *AluJo,* the most ancient Alu-related element and the oldest Alu subfamily, respectively, most likely promoted the occurrence of the deletion by the misaligning the longest homologous stretch (Liu et al., 2009) between them (Fig. 3): a 13-bp **microhomology** region. More specifically, one may assume pairing of the FRAM element with the left arm of the *AluJo* element followed by recombination within the 13 bp given that the right arm of the *AluJo* has only an 11 bp stretch of continuous homology and overall the FRAM element has

fewer mismatches with the left arm of the *AluJo* repeat. Both the FRAM element and the *AluJo* repeat are oriented in a telomer-to-centromer direction; the left arm of the *AluJo* occupying the most distal position in *OPN1MW* intron 3. The resultant hybrid Alu element present in the patient's (BCM17-III:2) DNA includes the stretch of 81 bp of the original FRAM element (centromeric to the *OPN1LW/MW*), the common 13 bp region and 65 bp of the most telomeric part of the left arm of *AluJo*, yielding a 159 bp hybrid Alu-element in the size of a monomeric Alu element (Fig. 3).

In general, there are two major mechanisms to repair DNA double strand breaks (DSBs): rejoining strands without a template by **non-homologous end-joining** (NHEJ) or, alternatively, using a homologous sequence on both ends as template during recombination (homologous recombination) (Gu et al., 2008). Recombination between two homologous sequences that are not alleles is known as **non-allelic homologous recombination** (NAHR). Misalignment with non-allelic copies of high identity in meiosis can result in *de novo* genomic rearrengements in the germline (Gu et al., 2008). Microhomology between Alu elements can mediate misalignment and recombination, resulting in deletions and duplications or deletions and inversions (Wong et al., 2018) depending on the orientation of the elements and whether the crossover occurs between different chromosomes or within a single chromosome, respectively (Batzer and Deininger, 2002; Kim et al., 2016; Liu et al., 2009).

Alu-microhomology can induce NHEJ in its close proximity by fusing partially homologous Alu elements, a mechanism known as homeology-induced NHEJ (HI-NHEJ) (White et al., 2015). The **chimeric** Alu-element observed at the novel deletion junction in the patient BCM17-III:2 (Fig.3) resembles partially fused Alu elements typical of HI-NHEJ (Morales et al., 2015). However, the length of the monomeric Alu donor (FRAM) and the fact that FRAM must have aligned with the left arm of *AluJo* (instead of the right one) speaks against a random rejoining of parts of the two Alu elements. Moreover, the **microhomology** between the FRAM and the *AluJo* element is not in the proximity but coincides with the deletion **breakpoint.** This makes **NAHR** the most plausible underlying mechanism. Hence, NAHR could have occured either intra-chromosomally or between different X-chromosomes during meiosis I in primary oocytes. In X-linked recessive Duchenne muscular dystrophy SNVs originate more likely during spermatogenesis and deletions and duplications during oogenesis (Grimm et al., 1994). A maternal bias for deletions in autism due to a meiotic hotspot has been observed (Duyzend et al., 2016).

CNVs are structural variants leading to genomic imbalance; for example deletions or duplications. The **clinical impact** of CNVs has been underscored by Truty and colleagues. A total of 1810 deletions and 1034 duplications across 384 disease-causing genes were found in ~2% of the individuals referred for a NGS-based genetic test (n = 143,515) and in ~10% of those with a clinically significant result (Truty et al., 2019). Intragenic CNVs represented ~8% of all pathogenic variants associated with pediatric and rare disorders whereas neurological disorders presented the highest prevalence (35%) of pathogenic CNVs. The majority of genes carried a single CNV whereas 8% of the genes had ≥ 15 CNVs which accounted for 70% of all intragenic CNVs; suggesting for a biased distribution of CNVs within the investigated genes (Truty et al., 2019).

About 2,500 structural variants per human genome were estimated to differ from the reference genome (The 1000 Genomes Project Consortium, 2015) using WGS. A recent study with multiple sequencing platforms elevated the figure by more than 10-fold (Chaisson et al., 2019). A global **frequency** estimate for *de novo* CNV formation regardless of genomic size is currently unknown albeit it has been estimated as 1 in 50 individuals (Veltman and Brunner, 2012). CNVs of >100kb have been estimated to occur *de novo* at a rate of 1.2×10^{-2} events per generation (Itsara et al., 2010). The insufficient detection power by most techniques used until now suggests that CNVs have been underestimated and represent a major source for missing heritability.

NAHR events often take place between repetitive elements (Cardoso et al., 2016). Low-copy repeats such as **segmental duplications** of 1-100 kb of genomic size with >90-95% of sequence identity are associated with Alu elements (Chen et al., 2014). Taking up > 10% of the human genome, these elements are high-copy repeats of the class of short interspersed nuclear elements known to be drivers for the formation of segmental duplications (Deininger, 2011). Segmental duplications are hotspots for NAHR events and predispose to CNV mutations. Likewise, some NAHR events use repetitive elements such as Alu elements as homology substrates, as in the deletion event presented here. Due to the shorter homology length, the latter ones are probably less frequent than those mediated by larger homologous sequences between low-copy repeats. Thus, the density of the two kinds of repetitive elements might play a role in CNV formation.

The density of Alu elements varies greatly among chromosomes and correlates with GC content. Seventy-five percent of genes are associated with Alu elements (Grover et al., 2004). The involvement of Alu elements in the origin of disease-causing deletions has been extensively reported (Cook et al., 2013). Cancer genes are often affected by Alu-mediated deletions; for instance, germline mutations in *MSH2* and *MLH1* that cause hereditary nonpolyposis colorectal cancer (van der Klift et al., 2005; Wang et al., 2003) or the *VHL* tumor suppressor gene associated with Von Hippel-Lindau disease for which 90% of germline deletions (n = 66) were located in Alu elements (Franke et al., 2009). Seventy-five percent of the deletions in *FANCA* associated with Fanconi anemia were reported to be Alu-mediated (Flynn et al., 2014). Alu-mediatedd NAHR often underlies rare adRP-associated large deletions in *PRPF31* – on the chromosome (19) with the highest density in Alu elements (Rose et al., 2011).

The density of Alu subfamilies also varies among chromosomes, being the oldest Alu elements more abundant in the X-chromosome (Grover et al., 2004). Within the X-chromosome, the highest Alu content is found at the end of Xp, comprising ~29% of the pseudo-autosomal region 1 (Ross et al., 2005). The local architecture of **Xq28** (~3Mb) is prone to rearrangements and harbours a number of segmental duplications. Deletions resulting from recombination between inverted repeats, Alu-mediated or non-homologous recombination in *FVIII* underlie severe hemophilia (Pezeshkpoor et al., 2012; Rossetti et al., 2004; Vidal et al., 2002) or the Rett syndrome when affecting *MECP2* – upstream of the *OPN1LW/MW* loci. The highest GC content in Xq28 is found between the *OPN1LW/MW* and the *G6PD* loci, representing an Alu-rich region as well (Chen et al., 1996).

Indeed, highly repetitive sequences have been reported to be involved in the origin of complex rearrangements at the *OPN1LW/MW* gene cluster in two BCM cases reported by Wang and colleagues (Wang et al., 2016). A 3-kb deletion frequently observed in BCM patients from the US (Yatsenko et al., 2016; Prof. Bernd Wissinger, *personal communication*) and the second BCM case carries a deletion which ocurred likely due to Alu-mediated NAHR with 17-bp microhomology with 8 contiguous nucleotides (Katagiri et al., 2018) homologous to the sequence at the deletion junction in BCM17-III:2 (Fig. 2b; Appendix 13.1.1, Publication I).

6.2 Unpublished I: A 142 bp intra-exonic deletion in both *OPN1LW* and *OPN1MW* underlying Blue Cone Monochromacy

A novel intragenic deletion was identified in exon 5 of both *OPN1LW* and *OPN1MW* gene copies in a male affected by BCM. The presented findings reveal the genetic basis presumably underlying this sporadic case of BCM. The putative origin for the mutation(s) identified in the *OPN1LW/MW* gene cluster is discussed.

I have investigated the family BCM262 from Spain with a single male affected with cone dysfunction (see Fig. 4A for a pedigree showing relevant family members). Similar to the **sporadic** patient in BCM17 presented in Section 6.1, the index patient (BCM262-III:2) had a primary clinical diagnosis of achromatopsia. Although no other family members had a visual impairment as severe as the index, a variety of ocular symptoms had been observed within the family. The patient's sister presents with nystagmus and strabismus since birth. However, in contrast to her brother, she has a good visual acuity of 0.8 and 0.6 in the right and left eye, respectively, no colour vision abnormalities and performs normal in ERG recordings. The patient's deceased maternal grandfather (Fig.4, BCM262-I:2, †) was a car driver with normal visual acuity but presented with a certain degree of colour blindness as reported by the family. The maternal great uncle had low visual acuity with high myopia yet family members did not report on his colour vision performance. Although the latter two family members may well presented with *OPN1LW/MW*-related phenotypes, lack of opthalmological examination records and samples for genetic investigation did not allow concluding on a potential relationship with the retinal disease and its cause in the index.

Upon genotyping subject BCM262-III:2 (Fig.4B and Appendix 13.2.1; Supp. Fig. S1), a novel **intragenic deletion** mutation in the human *OPN1LW/MW* gene cluster was identified. Strikingly, the exact same 142-bp deletion was simultaneously present in exon 5 of *OPN1LW* and *OPN1MW* gene copies (NM_020061.5:c.807_948del and NM_000513.2:c.807_948del, respectively), since only shortened exon 5 PCR products were detected by direct amplification from genomic DNA (amplifies exon 5 from both *OPN1LW* and *OPN1MW*), as well as by nested re-amplification of exon 5 following Long Distance Polymerase Chain Reaction (LD-PCR) specific for either proximal *OPN1LW* or distal *OPN1MW* gene copies (Fig. 4).

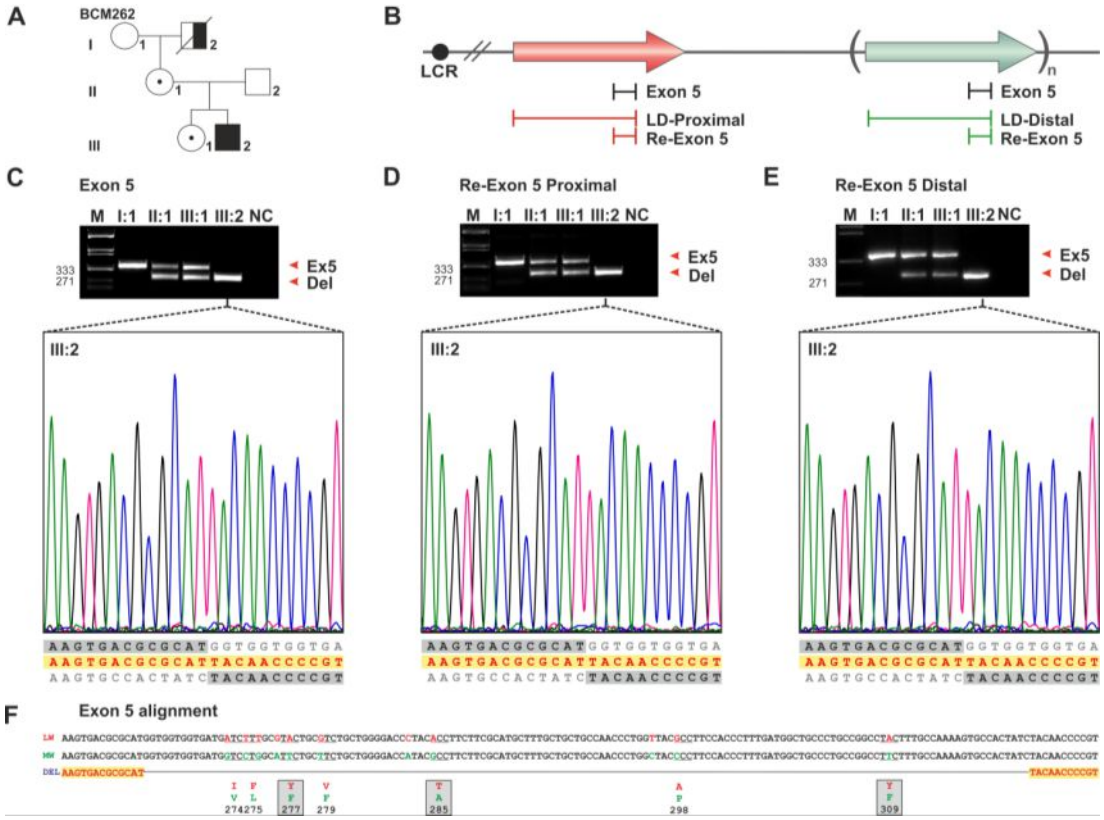


Figure 4. Identification of intraexonic deletions in exon 5 of *OPN1LW/MW* in case BCM262. **A)** Pedigree of family BCM262 showing the index male patient in the third generation (III) by a black square (BCM262-III:2) and close relatives. **B)** Genotyping *OPN1LW/MW* exon 5 by either direct amplification of exon 5 (black) or using LD- PCR amplify specifically the proximal or *OPN1LW* (red) and distal or *OPN1MW*(s) (green) gene copies, followed by a nested PCR that re-amplifies exon 5 from each LD-PCR product. **C-E)** Agarose gels showing PCR products for I:1, II:2, III:1 obtained by direct amplification of exon 5 from *OPN1LW/MW* simultaneously (C), and re-amplification upon LD-PCR of *OPN1LW* (D) or *OPN1MW*(s) (E). Sanger sequencing traces of the amplicons derived from III:2 of each PCR (C-E) highlighted in yellow and the reference sequence (grey), showing the breakpoints of the identified deletion. **F)** The sequence region with the deletion aligned to the reference sequence of exon 5 of *OPN1LW* and *OPN1MW*, showing specific polymorphisms of *OPN1LW* (red) and *OPN1MW* (green) and the respective differences in amino acid sequence below. Key Tyr277Phe (Y277F), Thr285Ala (T285A) and Tyr309Phe (Y309F) that contribute importantly wavelength sensitivity are squared in grey. LCR: Locus Control Region, LD: Long-Distance, Re-Exon 5: re-amplification of exon 5; M: Marker ladder, Ex5: Exon 5, Del: Deletion, NC: Negative Control, LW: *OPN1MW* and MW: *OPN1MW*.

Thus, the patient BCM262-III:2 has no single *OPN1LW/MW* gene copy with an intact exon 5. Sanger sequencing of direct and nested (from proximal- and distal-specific LD-PCR) exon 5 amplification products (Fig. 4C-E) revealed a 142 bp deletion within exon 5 which causes a **frameshift** (p.Met269Ilefs*10) and the introduction of a premature termination codon (PTC). Shortened and normal length exon 5 PCR products (for both proximal and distal *OPN1LW/MW* gene copies) were amplified with DNA from the patient's mother (BCM262-II:1) and the sister (BCM262-III:1) by direct

and LD-PCRs (Fig. 4C-E, BCM262-II:1 and BCM262-III:1). Thus, both are carriers of the NM_020061.5:c.807_948del deletion in *OPN1LW* and at least one of their *OPN1MW* copies harbours the NM_000513.2:c.807_948del mutation – on the X-chromosome transmitted from the mother to both siblings. With the available technology it is not possible to determine whether in these female carriers none, only one or all *OPN1MW* gene copies in *cis* with the deletion-carrying *OPN1LW* gene copy are mutated. Moreover, in principle it is not possible to distinguish intact *OPN1MW* gene copies in the deletion-bearing X-chromosome from those amplified from the non-mutated X-chromosome. Notably, no deletion was detected in any of the PCRs in the patient's grandmother (BCM262-I:1) (Fig. 4C-E), indicating that the deletion-bearing X-chromosome stems most likely from the **maternal grandfather** (BCM262-I:2).

Although no sample for genotyping was available from the deceased maternal grandfather (BCM262-I:2), his history of colour vision deficiency but normal visual acuity suggests that he likely carried the exon 5 deletion in either *OPN1LW* or the *OPN1MW,* and thus, a genotype compatible with *simple* **dichromacy**. In this scenario, gene conversion in the germline of the patient's grandfather (BCM262-I:2) may have transferred the mutation from *OPN1LW* to the most upstream *OPN1MW* or *vice versa*, with the consequence of a genotype underlying the BCM phenotype in the grandson (BCM262-III:2). In an alternative yet less likely gene conversion scenario, a *OPN1MW* gene copy downstream of the two first expressed *OPN1LW/MW* gene copies in the germ cells of either the patient's grandfather or mother could have plausibly acted as the donor copy and the most upstream *OPN1MW* copy as the recipient copy of the NM_000513.2:c.807_948del.

The direction of gene conversion could not be established from the experimental data; all 11 **paralog-specific** SNVs in exon 5 which can putatively distinguish *OPN1LW* from *OPN1MW* lie within the deleted sequence (Fig. 4F). Moreover, searching for polymorphisms in exon 5-flanking introns in phase with the deletion mutation did not allow to differentiate the nature of a donor copy from a past gene conversion event due to the lack of sequence differences.

Previously, intragenic deletions have only been reported in BCM subjects with a cone opsin gene cluster reduced to a single *OPN1LW* gene or single *OPN1LW/MW* hybrid gene (Gardner et al., 2009; Ladekjær-Mikkelsen et al., 1996). A 1.5 kb intragenic deletion removing the entire exon 4 and flanking intronic sequences and resulting in a

truncated opsin due to a frameshift and a PTC at position 266 has been reported in a Danish BCM family (Ladekjær-Mikkelsen et al., 1996). Secondly, a 1.2 kb deletion encompassing 919 bp of intron 1 and 288 bp of exon 2 has been reported by Gardner and colleagues in a British BCM family. This mutant allele is not predicted to alter the reading frame and may give rise to transcripts with an in-frame deletion that would lack important functional and structural elements of the native opsin protein (Gardner et al., 2009). Notably, the latter intragenic deletion may be a by-product of an unequal homologous recombination between *OPN1LW* and *OPN1MW* gene copies since all exons distal to the deletion are derived from an *OPN1MW* gene copy (Gardner et al., 2009).

The identified NM_020061.5:c.807_948del and NM_000513.2:c.807_948del mutations have never been reported in the literature. These findings are interesting due to the fact that the same novel and **non-recurrent** mutation is found in multiple gene copies of the *OPN1LW/MW* cluster in one BCM subject. The inactivating p.Cys203Arg point mutation in exon 4 has been repeatedly reported to occur simultaneously in multiple gene copies of the *OPN1LW/MW* cluster, and suggested to be a result of a founder effect (Jagla et al., 2002). In addition, the p.Trp177Arg point mutation (Gardner et al., 2010) and the *interchange* exon 3 haplotypes (Gardner et al., 2014) have been found simultaneously in multiple gene copies in cases with severe cone dysfunction disorders (see also Section 6.3). Conversely, an intragenic deletion has never been found in **multiple copies** of the *OPN1LW/MW* gene cluster. An independent occurrence of a unique deletion at the exact same site in each of the distinct gene copies in the same subject is very unlikely. It may rather be explained by a two-step process: first, the occurence of the deletion in one of the *OPN1LW/MW* gene copies, and secondly, the *spread* of the deletion into the non-mutant gene copy/copies by gene conversion.

In contrast to the molecular *fingerprint* found in the breakpoints of the deletion described in the previous chapter (Section 6.1), no repetitive elements were found by RepeatMasker in exon 5 of *OPN1LW/MW*. Whereas a global alignment revealed the exact breakpoints of the deletion, by performing a ClustalO (Sievers et al., 2011) local alignment, 11 bases upstream of the 5' deletion breakpoint were misaligned to the non-deleted exon 5 reference sequence spanning within the deletion. A 3-bp (-GTG-) **microhomology** stretch was found 8 bp upstream of the 3’ deletion breakpoint (Fig. 5A).

By means of *in silico* prediction with a model of expectation maximization (Bailey and Elkan, 1994), I searched for novel motifs (MEME, version 5.0.5). The candidate motif with the highest predictive score was found twice in the input sequence bearing the deletion (BCM262-III:2A). One motif (5'-ACGCGC-3') was found 2 bp upstream of the 5' deletion breakpoint. The second motif was found in the minus strand (5'-ACGGGG-3') 5 bases downstream of the 3' deletion breakpoint. Whereas most likely this does not represent a recurrent motif, the sequence on the same strand (5'-CCCCGT-3') at the 3' deletion breakpoint could represent a putative inverted repeat not fully complementary to that at the 5' deletion breakpoint which could potentially lead to a cruciform nearby the deletion (Fig. 5B).

Inspecting the sequence, I found two 11-bp **imperfect inverted repeats** in the neighbouring sequence of the deletion breakpoints (Fig. 5C). The first motif, 5'-GGCAGAGAAGG-3', is located 13 bp upstream of the 5’ deletion breakpoint. The second motif on the same strand is located within the deleted sequence; 15 bp upstream of the 3’ deletion breakpoint. Whereas the motif sequence in *OPN1MW* exon 5 (5'-CCTTCTTTGCC-3') has 10 out of 11 bases of complementary to the first motif, the same motif in *OPN1LW* has 9 complementary bases (5'-CCTACTTTGCC-3'). In both cases, intra-strand mispairing would include six Watson and Crick base pairings of G:C between two repeat motifs (Fig. 5C).

Non-B-DNA conformations were evaluated with QGRS Mapper (Kikin et al., 2006). Exon 5 of either *OPN1LW* or *OPN1MW* was queried for potential **quadruplex** forming G-rich sequences with default settings and including overlaps. For *OPN1LW* two putative sequences were found with a G-score of 9 while a total of 10 putative sequences were found for *OPN1MW* with G-scores ranging from 6 to 21. When the minimum size of the loop was restricted to 3 bp (instead of ‘0’ as default), no motifs were retrieved for *OPN1LW*. However, two sequences of 29 and 20 bp with a G-score of 12 and 21, respectively, were retrieved for *OPN1MW*.

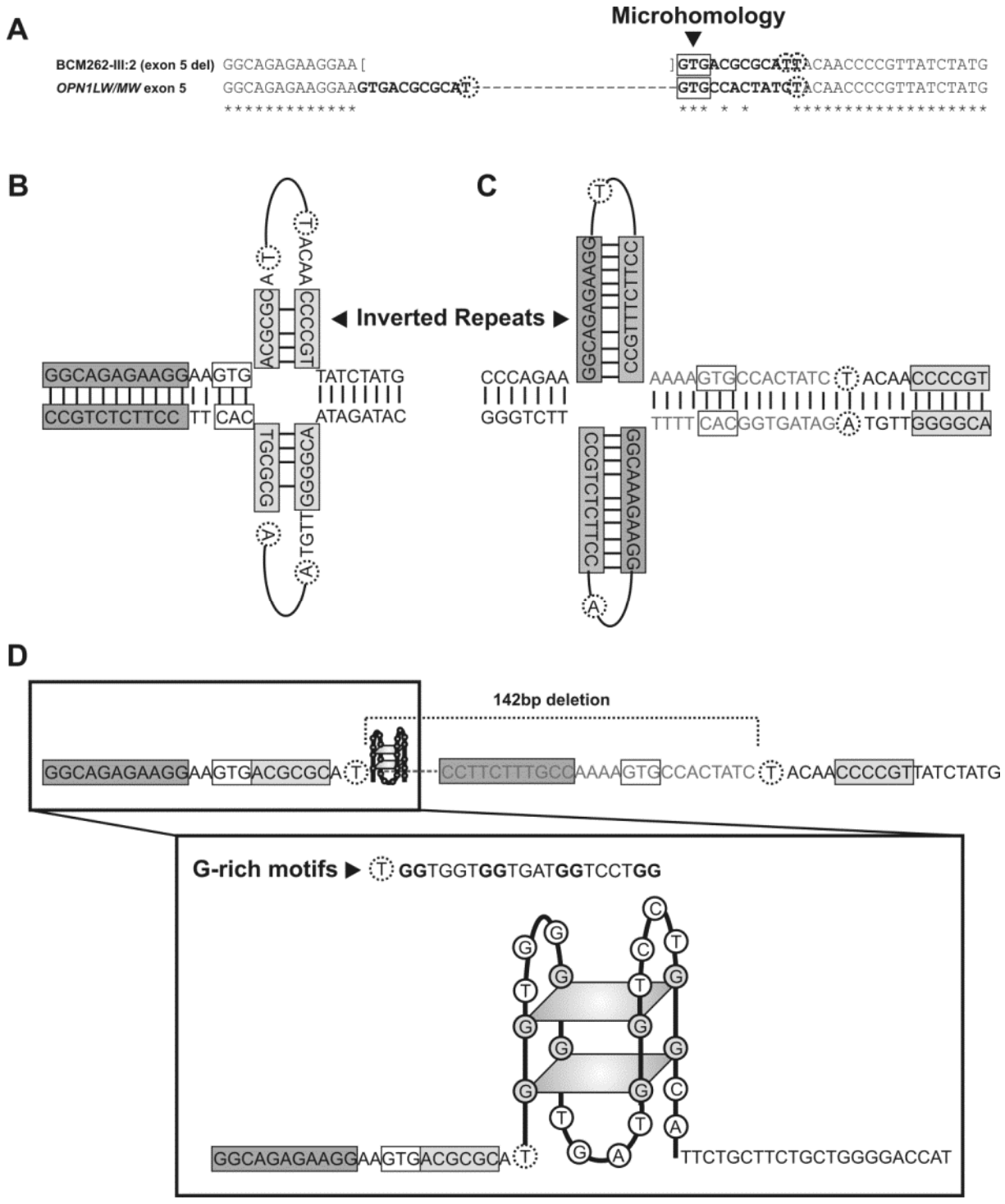


Figure 5. Genomic architecture features found in the breakpoints vicinity of the deletion found in BCM262-III:2. A) Microhomology stretch of 3 bp discoverd by local alignment with ClustalO (Sievers et al., 2011); 11 bases upstream of the 5' deletion breakpoint (a dashed circle highlights the “T” at the 5' and 3’ deletion breakpoints) were misaligned to the non-deleted exon 5 reference sequence spanning within the deletion. **B)** Imperfect inverted repeats (grey boxes) and **C)** imperfect inverted motifs (MEME) and **D)** quadruplex forming G-rich motifs (QGRS Mapper) G-rich motifs in bold letters. Dashed circles highlight the “T” at the 5' and 3’ deletion breakpoints.

The first predicted motif (5'-**GG**AAGTGACGCGCAT|**GG**T**GG**TGGTGAT**GG**-3'; G-rich sequences in bold letters and a boundary line representing the 5' breakpoint of the

deletion) overlaps the 5' deletion breakpoint. The second predicted sequence (5'-│**GG**TGGT**GG**TGAT**GG**TCCT**GG**-3'; G-rich sequences in bold letters and a boundary line representing the 5' breakpoint of the deletion; Fig. 5D), is predicted with the highest score and starts just downstream of the 5' deletion breakpoint. In addition, a G-quadruplex was predicted 71 bp downstream of 5' breakpoint of the deletion in the minus strand only in exon 5 sequence of *OPN1MW* (5'-G**CC**AA**CCC**TGGCTA**CCCC**TTCCA**CCC**TTT-3'; bases complementary to the G-rich sequences in the minus strand in bold letters) by G4Hunter (Brázda et al., 2019).

The breakpoint regions of the deletions presented in this subject might be associated with at least one of the following three the genomic architecture features: (1) 3 bp of microhomology 8 bp upstream of each breakpoint, (2) two pairs of imperfect inverted repeats either at each side of the breakpoint or with only one repeat flanking the breakpoint and the other repeat embedded in the deletion, with 4 and 6 bp of continuous complementarity, respectively and (3) G-rich sequences able to form G-tetrads spanning at least one breakpoint. However, not enough evidence is provided to prioritize the significance of each of the architecture features in relation to the other types in order to conclude which mechanism most likely underlies the deletion.

Microhomology was found in 80% of deletion breakpoints (n = 30) in a systematic study of genomic architecture nearby CNVs (Vissers et al., 2009). CNVs with short strechtes of microhomology could be indicative of NHEJ or microhomology-mediated break-induced replication (Hastings et al., 2009).

The two 11-bp motifs of **inverted repeats** – 140 bp appart – may fold into an intra-strand complementary hairpin. This conformation in each strand could render a cruciform of non-B DNA structure. Likewise, the novel 6-mer motifs – with 149 bp in between – could lead to a cruciform structure with a looped DNA bearing the 142-bp deleted sequence. Whereas the longest pair of repeats might be more stable, repeats flanking the breakpoint seem more plausible to contribute to the deletion formation due to the location with respect to the breakpoints. Hairpin loop inducing deletions by intra-strand complementarity of single-stranded DNA (ssDNA) and loop degradation was proposed for a 17-bp intraexonic deletion in exon 47 of the dystrophin gene in a case of X-linked Duchenne muscular dystrophy (Robinson et al., 1997).

The predicted oligo$(G)_n$ tracts could potentially induce the formation of **tetraplexes**. Intramolecular G-quadruplexes could be formed by the G-rich sequences on the same

strand or between two strands. Deletions have been shown to be enriched at G4 motifs in Single-Molecule Real-Time (SMRT) sequencing (Guiblet et al., 2018).

In spite of *OPN1LW/MW* being a NAHR-prone gene cluster, no evidence of hybrid genes, large blocks of homology (other than the duplicated genes in the array themselves) or repetitive elements (i.e. Alu-mediated) were found at the breakpoints of the deletion in this case. Lack of a molecular scar for classical NHEJ and the fact that microhomology is found adjacent to the breakpoint could point out to alternative NHEJ mechanisms such as microhomology-mediated end joining (Ottaviani et al., 2014). Yet, the microhomology does not overlap the breakpoint junction.

Mechanisms other than meiotic recombination processes might have been involved (Carvalho and Lupski, 2016). Non-B DNA structures different than the right-handed Watson and Crick B-form; for instance, cruciforms or G-quadruplexes, could lead to stalling of replication forks in mitotic microhomology-mediated mechanisms (Vissers et al., 2009). Such microhomology or other replication-based mechanisms frequently underlie non-recurrent deletions (Vissers et al., 2009) and have been found a major contributor of CNV formation in IRD genes (Van Schil et al., 2018).

A **second mutational event** could have transferred the deletion from distal to proximal or *vice versa*. Interestingly, two of the sequence features found in the breakpoints suggested that the (first) occurrence of the deletion in the *OPN1MW* gene copy is more likely. Intringuingly, gene conversion events might *spread* insertions or deletions as well. It has been observed that a sequence identity of at least 92% is a prerequisite for non-allelic gene conversion. Thus, the minimum gene conversion tract transferred from the donor to the recipient gene copy might have involved at least 2 kb of sequence. Indeed, in a genome-wide analysis in primates, Assis and Kondrashow found hundreds of sites with evidence for adjustments in length between paralogous sequences, most likely explained by non-allelic gene conversion. Particularly, deletions in paralogous sequences were >3 times more frequent compared with insertions (Assis and Kondrashov, 2012). While such gene conversion events involve mostly small deletions or insertions (indels) of <10 bp, there are also common human variation examples due to gene conversion events involving larger length difference between donor and recipient. For example, a ~10 kb CNV *shuffle* – which can occur in either directionality (proximal to distal, and *vice versa*) – between paralogous

sequences in the inverted repeat regions (IR3) on the Y chromosome (Shi et al., 2018).

In human disease, a recurrent 55-bp deletion in exon 9 of the glucocerebrosidase (GBA) gene in patients with Gaucher disease has been proposed to result from a gene conversion event originating from the *GBAP1* pseudogene with 96% homology at the coding sequence (Tayebi et al., 2003; Walley and Harris, 1993). Analogously, gene conversion between the iduronate-2-sulfatase gene *IDS* underlying Hunter disease (Bunge et al., 1998) and its neighboring pseudogene *IDS-2* in inverted orientation on the Xq28 including the occurence of deletions of exons 4-7 (~10kb) with remnant sequences from *IDS-2* has been reported by two groups (Birot et al., 1999; Li et al., 1999). *De novo* intraexonic deletions in *MECP2,* located upstream of *OPN1LW*, have been found associated with Rett syndrome (Bunyan and Robinson, 2008; Todorov et al., 2012).

A study searching for signatures of gene conversion revealed a truncated **Chi-element** (-TGGTGG-) to be enriched at converted tracts (Chuzhanova et al., 2009). A Chi-like element has been previously proposed to underlie gene conversion in exon 3 of *OPN1LW/MW* (Winderickx et al., 1993). This motif (-T│GGTGGTGG-; boundary exemplifies the deletion breakpoint) is found at the junction of the deletion in BCM262-III:2. Moreover, it has been shown that motifs able to adopt **non-B DNA** structures are overrepresented in converted tracts and thus, together with the action of recombination hotspots induce mutational gene conversion (Chuzhanova et al., 2009).

In summary, we have identified a novel non-recurrent pathogenic structural variant in both *OPN1LW* and *OPN1MW* underlying an X-linked BCM case. Several signatures which can promote non-recurrent CNVs have been found in the genomic architechture nearby the deletion breakpoints. A first origin in one gene copy (most likely in *OPN1MW*) and subsequent transfer by gene conversion to the non-mutated gene copy in a second step is proposed.

6.3 Publication II: *De novo* intrachromosomal gene conversion from *OPN1MW* to *OPN1LW* in the male germline results in Blue Cone Monochromacy

A *de novo* event of non-allelic gene conversion was identified in the lineage of a single family, supported by the evidence of phenotypically distinct cone dysfunction disorders in grandfather and grandson. The data supports that a strongly pathogenic exon 3 haplotype in *OPN1LW* originated from a non-reciprocal transfer of a sequence stretch from its paralogous gene *OPN1MW*. The converted and donor haplotypes induce high quantities of aberrantly spliced *OPN1LW* and *OPN1MW* transcripts, respectively. Longtime but evolutionary recent gene conversions have presumably led to intringuing haplotype diversity in exon 3. *Cis*-regulatory elements of splicing in exon 3 have been thus shaped, as shown by the quantitatively diverse splicing outcomes *in vitro*.

6.3.1 Gene conversion event from OPN1LW to OPN1MW in family BCM72

In contrast to the previous case (Section 6.2) in which the unavailability of samples hindered the confirmation of gene conversion underlying the origin of an intragenic 142 bp deletion in both *OPN1LW* and *OPN1MW* in subject BCM262-III:1, we have obtained sound evidence from a **second family (BCM72)** that proves a recent gene conversion event at the *OPN1LW/MW* gene cluster (Appendix 13.1.2, Publication II). In this case, an exon 3 haplotype inducing nearly complete exon skipping is *transferred* from *OPN1MW* to *OPN1LW*. The unaffected patient's mother (BCM72-II:2) carried the pathogenic haplotype 'LIAVA' in both *OPN1LW* and *OPN1MW* genes in the X-chromosome transmitted to the patient (BCM72-III:1). However, the grandfather's (BCM72-I:1) transmitting X-chromosome harboured the 'LIAVA' haplotype exclusively in a *OPN1MW* gene copy, while the 'LIAVS' haplotype was present in the *OPN1LW* gene copy (Fig.4; Appendix 13.1.2, Publication II).

Gene conversion entails the **unidirectional** exchange of short regions between homologous or paralogous sequences or gene copies (Dumont, 2015; Hallast et al., 2005). Based on the relative genomic and chromosomal location of donor and recipient sequences, two forms of gene conversion events can be distinguished: (1) **allelic gene conversion** – also referred to as interallelic, intralocus or intragenic gene conversion (Halldorsson et al., 2016; Wiehe et al., 2000; Zangenberg et al., 1995) – usually involving two alleles of the same gene (in *trans*) and thus occuring

interchromosomally, and (2) **non-allelic gene conversion** – often named intergenic, interlocus or ectopic gene conversion – involving homologous sequences between two distinct genes or genomic loci (Arguello and Connallon, 2011; Assis and Kondrashov, 2012; Dumont and Eichler, 2013; Trombetta et al., 2016).

Paralogous genes originating from segmental duplications frequently belong to the latter class. In non-allelic gene conversion between **paralogues** located on the same chromosome, both intra- or inter-chromosomal events may occur yet intrachromosomal events are more frequently reported to transfer disease-causing mutations (Casola et al., 2012). Interlocus gene conversion may be neutral, as reported in an single event involving *BRCA1* and its pseudogene (Tessereau et al., 2015). The event identified in the germline of BCM72-I:1 (Appendix 13.1.2, Publication II) transferred sequence from a **non-functional donor**, the *OPN1MW* gene copy inactivated by a deleterious rare exon 3 haplotype, to a (at least partially) **functional acceptor** gene copy: *OPN1LW*. This event led to a detrimental outcome for the offspring that inherited a converted *OPN1LW* gene. It has been proposed that gene conversion accounts for 1% of *de novo* disease alleles as tested using 3,196 genes from HGMD (Casola et al., 2012) and for 3% of the divergence between paralogous genes (Dennis et al., 2017; Dumont, 2015).

Detection succeeds if first, paralog-specific variants are included in the transferred sequence and are flanked by sequence of the original sequence, and second, non-mendelian segregation of such variants from parents to offspring is observed. The haploid nature of X-linked loci in males enables direct comparison of sequences. The ancestral haplotype ‘LIAVS’ was found only in the acceptor gene, *OPN1LW* of the grandfather’s X-chromosome (Fig. 4; Appendix 13.1.2, Publication II). The analysis of **allelic haplotypes** (phasing) from females is critical. Remarkably, this revealed that the patient’s mother carried the converted ‘LIAVA’ haplotype in the *OPN1LW* gene of one X-chromosome. Upon confirmation of the grandfather’s transmitting X-chromosome, the possibility of the grandmother (BCM72-I:2) being a carrier of the transmitted X-chromosome to the patient (BCM72-III:1) can be ruled out. These findings strongly support that the gene conversion event had happened upon transmission of the X-chromosome from the grandfather to the mother, i.e. in the **grandfather’s germline**, and furthermore excludes allelic or interchromosomal non-allelic gene conversion. The most likely scenario to explain these findings is a single non-allelic, intrachromosomal event. The homologous sequence in the **sister**

chromatid of the downstream *OPN1MW* gene may have been the donor template; for instance, by means of synthesis-dependent strand annealing (Ira et al., 2006; McMahill et al., 2007).

The occurence of a single crossover recombination event was ruled out because no chimeric products or changes in copy number were detected in the lineage. Formally, it is very difficult to distinguish recombination events with double crossover from a gene conversion event. An allelic crossover event between *OPN1LW* gene copies in the mother's germline would involve recombination between the outermost non-recombinant variants. An allelic **double crossover** between *OPN1LW* and *OPN1MW*, namely crossover at both sides of a double Holliday junction (Haber et al., 2004), involving exchange of the DNA strands, could in principle result in such a converted product with flanking regions of the original gene. This hypothesis could be accepted only if the following criteria are met: (1) occurred without change in copy number, (2) the grandmother (mother of BCM72-I:2) would have carried the haplotype ‘LIAIA’ in the *OPN1LW* gene (of the X-chromosome in which BCM72-I:2 carries the ‘LIAIS’ haplotype in the *OPN1LW* gene copies) and (3) the first crossover happened exactly in the region of 5 bp (in between c.532A>G and c.538T>G) between the haplotype ‘LIAVS’ (grandfathers’ X-chromosome) and the previously assumed ‘LIAIA’ (grandmothers’ X-chromosome). However, no evidence supported this hypothesis.

Furthermore, reciprocal crossing usually involves larger regions than **non-crossover** gene conversions, the latter being typically less than 1 kb of tract length (Chen et al., 2007; Mansai et al., 2011). The mean tract length of gene conversions is estimated to be <300 bp (Jeffreys and May, 2004; Odenthal-Hesse et al., 2014; Reiter et al., 1998). The **maximal converted tract** in BCM72 was delimited to 1,297 bp (Fig. 5; Publication II). Most likely the converted tract length lies below 1 kb. This large window in length can be attributed to the scarcity of paralogous-specific polymorphic sites at this genomic region. Introns are less divergent than exons between the *OPN1LW* and *OPN1MW* genes; with >99% identity in the maximal converted intronic sequence.

Table 1. Allele frequencies of exon 3 variants of *OPN1LW* and *OPN1MW*(s) genes in public databases and published reports

Gene[#]	Nucleotide change[§]	Amino acid change[$]	UCSC Coordinate(X:)	Minor Allele	MAF[⊖]	Males [⊖⊖]	dbSNP	MAF[‡]	Allele Number[¶]	dbSNP	MAF[‡]	Allele Number[¶]	rsID	MAF[‡]	Allele Number[¶]	Flag	MAF <0.05
					Winderickx		1000G			ExAC			gnomAD				
OPN1 LW	c.453G>A	Arg151=	153418456	A	0.189	74	rs949430	0.049*	3775	rs75719797	0.0034	73319	rs949430	0.0028	171250	SD	3
	c.457C>A	Leu153Met	153418460	A	0.189	74	rs713	-*		rs78407189	0.0013	73368	rs713	0.0022	171620	SD	2
	c.465G>C	Val155=	153418468	C	0.176	74	rs713614	0.071	3775	rs72616469	0.0372	73646	rs731614	0.0306	171468	SD	2
	c.511G>A	Val171Ile[&]	153418514	A	0.135	74	rs5986963	0.198	3775	rs79866052	0.1448	80200	rs5986963	0.1377***	184436	SD	0
	c.513G>T	Val171Ile[&]	153418516	T	0.135	74	rs5986964	0.199	3775	rs55676401	0.1452	80192	rs5986964	0.1383***	184522	SD	0
	c.521C>T	Ala174Val	153418524	T	0.054	74	rs14989767	0.135	3775	rs149897670	0.1011	79674	rs149897670	0.0925	180910	SD	0
	c.532A>G	Ile178Val	153418535	G	0.054	74	rs14500967	0.037	3775	rs145009674	0.0221	79514	rs145009674	0.0170	181881	SD	3
	c.538T>G	Ser180Ala	153418541	G	0.378	74	rs949431	0.17	3775	rs949431	0.2053	77022	rs949431	0.2200	165866	SD	0
OPN1 MW	c.453G>A	Arg151=	153455586	G	0.110	52		-*			-*			0.0009	20461		1
	c.457C>A	Leu153Met	153455590	C	0.110	52		-*			-*			0.0016**	5766**		1
	c.465G>C	Val155=	153455598	G	0.159	52		0.2	3775		-*		rs368356511	0.0428	18989		1
	c.511G>A	Val171Ile[&]	153455644	A	0.000	52		0.03	3775		-*		rs782545619	0.0012***	13779		3
	c.513G>T	Val171Ile[&]	153455646	T	0.000	52		0.03	3775		-*		rs782563380	0.0011***	13577		3
	c.521C>T	Ala174Val	153455654	T	0.378	52		0.51	3775	rs372044027	0.0488	82	rs372044027	0.0421	12419		2
	c.532A>G	Ile178Val	153455665	G	0.378	52	rs5742604	0.47	3775		0.4600	699	rs375538821	0.0285	11600		1
	c.538G>T	Ala180Ser	153455671	T	0.060	52		-*			-*		rs782648959	0.0742	9890		0
OPN1 MW2	c.453G>A	Arg151=	153492704	G	0.110	52		-*			-*		rs782680594	0.0072	45635	SD	1
	c.457C>A	Leu153Met	153492708	C	0.110	52		-*			-*		rs782285971	0.0043	45273	SD	1
	c.465G>C	Val155=	153492716	G	0.159	52	rs371039245	0.197	3775	rs371039245	0.1695	1628	rs371039245	0.1337	43189	SD	0
	c.511G>A	Val171Ile[&]	153492762	A	0.000	52		-*			-*		rs782395114	0.0053***	40682	SD	2
	c.513G>T	Val171Ile[&]	153492764	T	0.000	52		-*			-*		rs782024115	0.0051***	40648	SD	2
	c.521C>T	Ala174Val	153492772	T	0.378	52	rs142076459	0.494	3775	rs142076459	0.4300	872	rs142076459	0.3817	30027	SD	0
	c.532A>G	Ile178Val	153492783	G	0.378	52		0.470	3775	rs146189239	0.5379	699	rs5742604	0.3644	37551	SD	0
	c.538G>T	Ala180Ser	153492789	T	0.060	52		-*			-*		rs139163406	0.0455	37010	SD	1

1000G: dataset from 1000 Genomes Project; ExAC: Exome Aggregation Consortium; gnomAD v2.1: The genome Aggregation Database entails exome and whole-genome data; [#]*OPN1LW* (Ensembl Gene ID: ENSG00000102076; UCSC Coordinates X:153409699-153424508; GRCh37/hg19), *OPN1MW* (Ensembl Gene ID: ENSG00000147380; UCSC Coordinates X:153448108-153461634; GRCh37/hg19) and *OPN1MW2* (Ensembl Gene ID: ENSG00000166160; UCSC Coordinates X:153485226-153498756; GRCh37/hg19); [§]Nucleotide changes that represent homologous variants among *OPN1LW* and *OPN1MW* gene copies given as the corresponding coding DNA sequence nomenclature; [$] Amino acid sequence change resulting from the respective nucleotide change, =: synonymous variant, [&]The variants c.511G>A and c.513G>T are in strong LD and located within the same codon, [‡]MAF: minor allele frequency (MAF), [⊖]Rarer allele frequency determined from normal colour vision males (Winderickx et al., 1993); [⊖⊖]Total number of male individuals studied; restricted to arrays with one or several *OPN1MW* gene copies with either the same haplotype or differing in only one polymorphism (Winderickx et al., 1993); [¶]Total count of alleles; *Low or not available frequencies due to non-unique short reads misplaced or collapsed for one single locus, **Frequency and total number of alleles derived only from genomes as exome data provided no high confidence genotype;***Variant annotation dubious; SD: warning flag indicating a variant falls in a segmental duplication region. Databases include homozygous, heterozygous and hemizygous alleles from females and males. Winderickx and colleagues considered only hemizygous alleles from male subjects.

This **sequence erosion** between the red and green opsin genes has been noticed in human, chimpanzee, baboon and gibbon species (Hiwatashi et al., 2011; Kuma et al., 1988; Zhou and Li, 1996); being the intra-species sequence similarity between paralogous genes higher than the similarity to orthologs in related species even if duplication and divergence occurred prior speciation. This prevents individual genes in the cluster from accumulating mutations and thus, these genes evolve depedent on one another which is known as **concerted evolution** (Bagnall et al., 2005). This fact underscores long-term gene conversion between *OPN1LW* and *OPN1MW* in humans, apes and other Old World primates (Fawcett and Innan, 2011; Shyue et al., 1994; Verrelli et al., 2008; Zhao et al., 1998). Maintenance of sequence diversity in exon 5 responds to a **selection pressure** to preserve differential spectral sensitivity (Shyue et al., 1994; Zhao et al., 1998); specifically, the Tyr277 and Thr285 residues in the long-wavelength sensitive (LWS) opsin, and the Phe277 and Ala285 in the middle-wavelength sensitive (MWS) opsin which account for ~7 and ~14 nm difference in peak spectral sensitivity, respectively (Deeb et al., 1992). Combinations of the latter are found in hybrid genes as a result of unequal homologous recombination (Merbs and Nathans, 1992). The c.538T>G/p.Ser180Ala substitution results in a ~5 nm shift (Merbs and Nathans, 1992). The c.538G is the minor allele in *OPN1LW* and the ancestral allele in *OPN1MW* (Table 1).

Precisely, the **c.538T>G/p.Ser180Ala** is the only variant by which the original ('LIAVS') and the converted ('LIAVA') exon 3 haplotypes in *OPN1LW* of the grandfather (BCM72-I:1) and patient (BCM72-III:1), respectively, differ (Fig. 4; Publication II). According to the MAFs documented by Winderickx, 1000G, ExAC and gnomAD (Table 1), c.538T>G is a **high-frequency** polymorphism in *OPN1LW*. The latter is shared between *OPN1LW* and *OPN1MW* in the population (Asenjo et al., 1994; Winderickx et al., 1993). Actually, all six non-synonymous SNPs found to be shared between *OPN1LW* and *OPN1MW* genes in a total of 75 normal colour vision and 58 colour-vision deficient Caucasian males were located in exon 3 (Winderickx et al., 1993) constituting a variety of haplotypes. The c.511G>A and c.513G>T seem to be mostly found within *OPN1LW* (Table 1). Due to the non-unique short reads and the fact that Winderickx considered only hemizygous alleles, certain SNPs may be below the minor allele frequency (MAF) threshold of 0.05 in the databases (Table 1), but are common in Winderickx et al., 1993. According to the latter study, most of the

haplotypes we found in our study can be considered **rare** (Fig. 1b and Table 2; Appendix 13.1.2, Publication II). Further discussion will follow in 6.3.2.

Upon comparison of ~5 kb from intron 2 to the 5‘ part of exon 5 of *OPN1LW* genes of 236 human males with undefined colour vision, 9 out of 14 variants in ~0.5 kb of coding sequence were in exon 3 (Verrelli and Tishkoff, 2004). An atypical haplotype structure highlighted by only sporadic significant LD between pairs of variants in and around exon 3 was found, suggesting frequent recombination or gene conversion (Verrelli and Tishkoff, 2004). The latter supported prior observations of **partial LD** between variants at 5' and 3' parts of exon 3, and between the 3' part of exon 3 and exons 4 and 5 (Winderickx et al., 1993).

The shifts in the absorption maximum of the cone pigment(s) generated by the amino acid changes encoded by exon 3 polymorphisms do not contribute to adaptive selection. However, the **unusual haplotype structure** in exon 3 is not compatible with a model of neutrality (Verrelli and Tishkoff, 2004). With the exception of the Ser180Ala variant, no other polymorphic sites are found in exon 3 of the multi-allelic single LWS/MWS opsin gene in New World primates, indicating that haplotype diversity in exon 3 acquired its non-neutral character after gene duplication. Indeed, the overall divergence among polymorphic X-linked opsin genes in New World primates is ~3.5% less than between the duplicated genes of Old World primates (Hunt et al., 1998). While the Chi-element in exon 3 – thought to trigger gene conversion – is conserved across primates (Winderickx et al., 1993), the unusual haplotype structure in exon 3 is not present in chimpanzees (Verrelli et al., 2008) or gibbons (Hiwatashi et al., 2011). The *OPN1MW* gene is **copy-number polymorphic** in humans (Nathans et al., 1986) and a significantly higher frequency of multiple-gene arrays was found in human compared with chimpanzees (Terao et al., 2005). Moreover, the frequency of dichromacy – repeatedly caused by unequal homologous recombination between *OPN1LW* and *OPN1MW* gene copies (Nathans et al. 1986) – is higher in humans compared to macaques or chimpanzees (Kawamura et al., 2012). Dichromacy or anomalous trichromacy do not represent a severe handicap for humans at least in agricultural societies. Verrelli and Tishkoff proposed two mechanisms underlying the non-neutrality observed for *OPN1LW* exon 3: (1) **alternative functional layers** within the cone photoreceptors other than the spectral sensitivity may be involved, and (2) **haplotypes rather than individual** variants should be considered for functional investigations.

6.3.2 Haplotype-dependent splicing regulation of OPN1LW/MW exon 3

Splicing might be compromised in *OPN1LW* (Ueyama et al., 2012); for instance, by the rare exon 3 converted haplotype 'LIAVA' in BCM72-III:1 (Fig.5; Publication II) which induces exon 3 skipping. Pathogenic **exon 3 haplotypes** in both *OPN1LW* and *OPN1MW* may originate by recurrent *de novo* mutations, allelic or unequal homologous recombination, or most frequently by gene conversion (Hurles, 2005).

Cis-regulatory elements of splicing in exon 3 are disrupted in a haplotype-dependent manner as indicated by our analysis with sixteen **BCM/XLCD** families. In total, twenty-four affected subjects with intact arrays of single, single-hybrid or multiple arrays and without common mutations in *OPN1LW/MW* (Fig.1) other than exon 3 *interchange haplotypes* were included (Fig. 1a; Appendix 13.1.2, Publication II). Most of the haplotypes found in these subjects are **rare** (Fig. 1b and Table 2; Publication II and Winderickx et al., 1993). In order to assess the efficiency in transcript processing, a series of twelve **minigene** constructs reproducing the observed haplotypes were expressed in Human Embryonic Kidney 293 cells that harbour the SV40 T-antigen (HEK293T; American Type Culture Collection CRL-11268). Sanger sequencing validated the identity of correctly and aberrantly spliced products from the Reverse Transcriptase (RT)-PCR (Fig. 6A). Moreover, we determined semi-quantitatively the ratio of incorrectly to correctly spliced transcripts by implementing a fluorescence-based fragment analysis (Table 2 and Fig. 2; Publication II). In particular, the haplotype 'LIAVA' resulted in the complete absence of correctly spliced transcripts (Fig.2; Publication II). The majority of aberrantly spliced transcripts showed skipping of the entire exon 3 (Fig. 6B), most likely resulting in a **frameshift** that leads to a PTC in exon 4. Thus, the derived opsin protein is predicted to be truncated. The resultant transcript might be degraded by nonsense-mediated decay (NMD). For certain haplotypes inducing aberrant splicing, there was in addition a minor aberrant transcript species in which exon 3 and parts of exon 2 were lacking (Fig. 6C) which also results in a PTC.

Aberrant splicing levels of ≥20% were found for 75% of the haplotypes characterized. Four haplotypes yielded levels of less than 10% of correctly spliced products. By **pairwise comparison** of the splicing competence among *single-mutant* haplotypes, i.e. those haplotypes which differ by only one polymorphism, we concluded that the c.532A>G/p.Ile178Val has the greatest **single base effect** (~50%) on exon 3 inclusion levels. For instance, Haplotype 1 ('LIAVA') with 0% exon 3 inclusion compared to

Haplotype 8 ('LIAIA') with 40.75% exon 3 inclusion, or, Haplotype 4 ('LVAVA') with 6.7% exon 3 inclusion compared to Haplotype 5 ('LVAIA') with 79.4% exon 3 inclusion (Table 2; Publication II). If c.532A>G/p.Ile178Val and c.521C>T/p.Ile174Val are simultaneously present (Haplotype 9, 'MVVVA$^{c.465G'}$ and Haplotype 10, 'MVVVA$^{c.465C'}$; Fig. 1 and Table 2; Appendix 13.1.2, Publication II), in comparison to Haplotype 12 ('MVAVA') exon 3 inclusion levels rise from 50 to 80%. However, for each variant the effect exterted in exon inclusion inferred from the difference in exon inclusion is not maintained across all pairwise comparisons, suggesting that the effect of individual variants within the exon 3 haplotyes is not simply addittive.

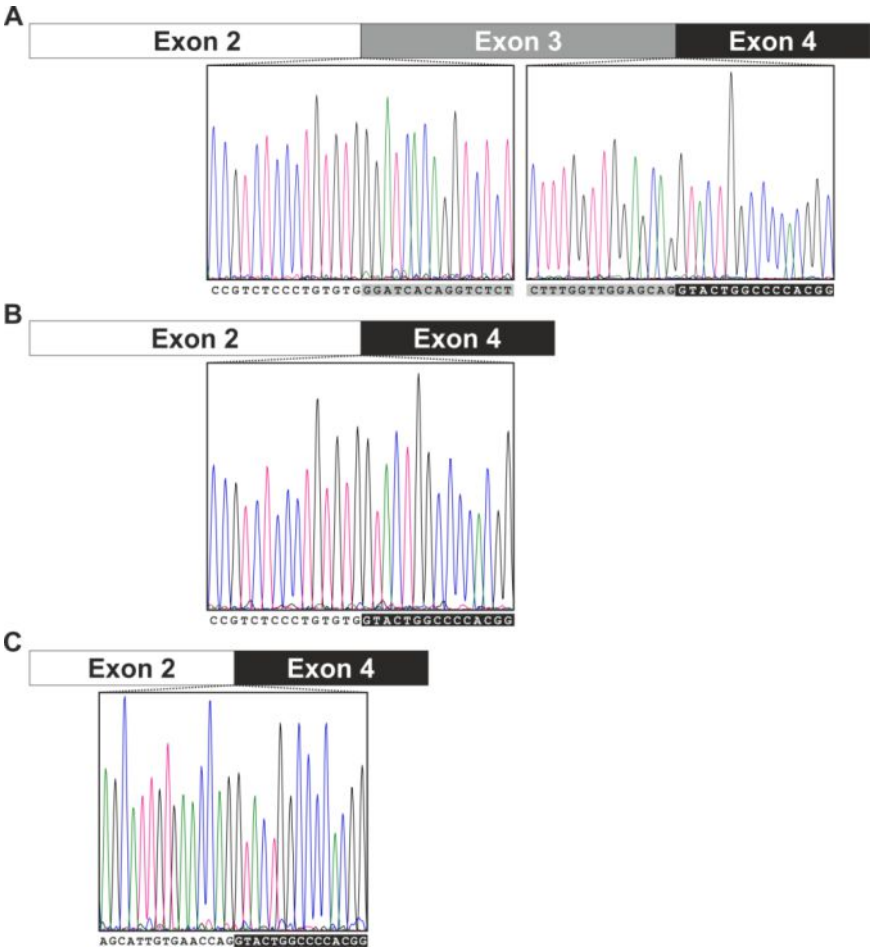


Figure 6. Sequencing of exon junctions of Reverse Transcriptase (RT)-PCR products derived from correctly and aberrantly spliced transcripts upon transient expression of *OPN1LW/MW* minigenes. A) Diagram of the correctly spliced product with exon 3 inclusion illustrated by traces of Sanger sequencing with intact exon-exon junctions between exon 2, exon 3 and exon 4. **B)** Aberrantly spliced product showing skipping of the entire exon 3, as indicated by the sequence of the exon 2-exon 4 junction. **C)** A second aberrantly spliced product detected in minor levels lacking exon 3 and partially sequence of exon 2, most likely as a result of using a cryptic splice donor site within exon 2.

Since exon 3 skipping results in non-functional transcripts, such mutations or haplotypes – if fully penetrant – may cause colour vison deficiencies or even more severe cone dysfunction disorders.

Any preceding event to protein function – transcription, splicing, translation – can have an impact on the ultimate synthesized proteins and therefore, potentially constrain protein evolution (Warnecke et al., 2009). Balancing selection has shaped haplotype-specific alternative splicing in TAP2, a protein involved in contact and transport of antigenic peptides (Cagliani et al., 2011; Qu et al., 2007). Splicing regulatory sequences may influence codon usage and therefore, these sequences may maintain a balanced dual role with respect to functions (Parmley and Hurst, 2007). Nucleotide changes may imply a conflict between splicing and protein fitness (Falanga et al., 2014) or, may have pathogenic effects at multiple functional layers (Donadon et al., 2018). The effect of exon divergence on splicing modulation following gene duplication is yet poorly understood (Zhang et al., 2009).

The c.538T>G/p.Ser180Ala could represent such an example where improvement of protein function and reduced splicing competence are kept in balance. The **dual effect** of sequence alteration has been highlighted for the exon 3 haplotype ‘LVAVA’ which results in a toxic photopigment in mice, presumably due to defects in folding or trafficking of the protein (Greenwald et al., 2017) in addition to a splicing defect with only 7% of correctly spliced transcripts (Fig.2b; Publication II). Such low levels either do not suffice (*hypomorphic* mutation) or, as proposed by Greenwald and colleagues, are followed by a second deleterious effect at the protein level which prevents functionality of the photopigment (resembling a *null* mutation). Hitherto, it remains elusive whether additional haplotypes, i.e. ‘MVAVA’, which results in a less pronounced splicing defect (53% correct transcripts; see Fig. 2 in Publication II), do also impair protein function.

In our study we observed certain phenotypic variability among patients with rare exon 3 *OPN1LW/MW* haplotypes, as exemplified by the assigned clinical subtypes. Most likely this reflects the inherent genotypic diversity in this group. Due to the small size of the cohort, **genotype-phenotype** correlations are limited. While variability is mostly found in refraction, visual acuity and colour vision performance, agreement in ERG is found, being scotopic responses essentially normal and photopic responses considerably reduced yet quantification is limited (Supplementary Table S1; Appendix 13.1.2, Publication II). As a matter of fact, phenotypic variability of BCM/XLCD and dichromatic subjects has been highlighted across and within mutational categories (Aboshiha et al., 2016; Carroll et al., 2009, 2012; Gardner et al., 2014, 2016; Patterson et al., 2016; Wagner-Schuman et al., 2010).

The genotypes of our group can be divided into two main categories: (1) single gene and (2) multi-gene *OPN1LW/MW* arrays with a total number of ≥ 2 gene copies. Subjects from other studies are included for genotype-phenotype correlations.

Single gene arrays can comprise either a single *OPN1LW* gene copy, a single *OPN1MW* or single-hybrid genes. A total of 5 subjects with a diagnosis of BCM or XLCD from two families presented single arrays with the 'LIAVA' haplotype in either single *OPN1MW* gene (N = 1; N: number of subjects) or in a single-hybrid gene configuration (N = 4) in our study (Supplementary Table S1; Appendix 13.1.2, Publication II). When tested, colour vision and photopic responses were impaired. All presented with nystagmus yet neither photophobia nor progression. A BCM/XLCD subject with impaired colour vision carrying a single *OPN1LW* gene with the 'LIAVA' haplotype was reported by Gardner and colleagues. Mizrahi-Meissonnier and colleagues reported one XLCD family carrying a single-hybrid gene with the 'LIAVA' haplotype and presenting with either protan or impaired colour vision performance (N = 2; see Table 2). The 'LIAVA' haplotype inactivating single genes resembles a *null* mutation. Among the subjects of 3 families carrying either an *OPN1LW* single gene (N = 2) or a single-hybrid gene (N = 2) with the 'LVAVA' haplotype, diagnose varied from BCM with impaired colour vision to cone-rod dystrophy (CRD) with protan defect (Supplementary Table S1;, Publication II and Table 2). Within the same genotype categories, others studies reported subjects with varying phenotypes from impaired to deutan colour vision performance (Table 2). Although single gene arrays with the 'LVAVA' haplotype are associated with an XLCD/BCM diagnose, the phenotypes of these subjects are more variable than those with 'LIAVA'-bearing single gene arrays.

The 'LVAVA' haplotype results in minor levels of exon 3 inclusion (<10%) and thus is classified as strongly deleterious. An additional derived pathomechanism (as proposed by Greendwald et al., 2017) might account for the phenotypic variability observed among subjects carrying single genes carrying the 'LVAVA' haplotype.

Table 2. Subjects with single or single-hybrid *OPN1LW/MW* gene arrays

Nº	Family ID - Patient ID	Gene	Exon 3 Haplotype		Phenotype			Reference
					Colour vision	Clinical diagnosis	Progression	
1	BCM66-16407	*OPN1LW*	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA	Impaired	BCM	No	This study
	BCM194-25474				Normal	BCM	No[†]	
	F16 - S18 -				Impaired	BCM/XLCD	No	Gardner et al., 2014
	10-0059 (=JC_0347)				Deutan	Deutan/CD	Yes	Greenwald et al., 2017; Carroll et al., 2012
2	MOL0250 - MOL0250-III:2 (= JC_0118)	*OPN1LW*	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G	LIAVS	Errors	XLCD[β]	ND	Mizrahi-Meissonnier et al., 2010; Carroll et al., 2012
	MOL0250 - MOL0250-IV:3				Errors	XLCD[β]	Yes	Mizrahi-Meissonnier et al., 2010
	MOL0267 - MOL0267-III:1				Deutan >Protan	XLCD[β]	ND	
3	F18 - S21	*OPN1LW*	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	Impaired	BCM/XLCD	Yes	Gardner et al., 2014
4	BCM93-19164	*OPN1MW*	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	Impaired	XLCD	No	This study
5	11-0115 (=JC_0564)	*OPN1MW*	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA	Protan	Protan/CD	Yes	Greenwald et al., 2017; Carroll et al., 2012
6	BCM73-16953	Hybrid	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	Impaired	BCM	No	This study
	BCM73-17481				ND	BCM	No	
	BCM73-20537				ND	XLCD	No	
	BCM73-20770				ND	BCM	No	
	MOL0057 - MOL0057-III:1				Impaired	XLCD	Yes	Mizrahi-Meissonnier et al., 2010
	MOL0057-MOL0057-III:3				Protan	XLCD	ND	
7	BCM112-23518	Hybrid	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA	Protan	CRD	Slow	This study
	BCM112-22852							
	F13 - S19				Impaired	BCM/XLCD	No	Gardner et al., 2014
	F17 - S20				Impaired	BCM/XLCD	No	

[†]progression due to high myopia; [β]retained photopic ERG responses. ND: not determined. Adapted from Supplementary Table S1 (Publication II; Appendix 13.1.2)

Within the second main genotype category of our study, that is, *OPN1LW/MW* gene arrays with **multiple genes**, the clearest impact of different exon 3 haplotypes on the phenotype is provided by the BCM72 family (Fig.3; Publication II and Table 3). While the 'LIAVA' haplotype present in the *OPN1MW* donor gene in BCM72-I:1 and in the converted *OPN1LW* gene in BCM72-III:3 exclusively produces aberrant processed transcripts, the 'LIAVS' haplotype present in the grandfather's (carried by the *OPN1LW* gene copy) yielded ~20% of correctly spliced transcripts (Fig. 5; Publication II). This level seemed to be tolerated since the grandfather had a phenotype of deuteranopia (a colour vision anomaly caused by non-functional M-cones) but had otherwise normal visual function (Publication II). In contrast, the grandson with the converted 'LIAVA' haplotype in the *OPN1LW* gene copy instead of 'LIAVS', presented with a by far more severely impaired vision. The two distinct phenotypes in BCM72-I:1 and BCM72:III:1 result from a hypomorphic variant and a loss-of-function mutation, respectively. In agreement, affected subjects from two families (Nº 2, N = 3; Table 2) carrying a single *OPN1LW* gene array with the 'LIAVS' haplotype presented with non-extinguished ERG photopic responses and relatively well preserved cone function. Intra-familiar variability at middle ages was observed at the fundus, varying from nominal macular findings to foveal atrophy (Mizrahi-Meissonnier et al., 2010). Also, a macula dystrophy was observed in BCM72-I:1 (Fig. 3b; Publication II and Table 3).

Within the multi-gene *OPN1LW/MW* arrays, a **first subcategory** of subjects with **identical haplotypes** in all gene copies or the gene copies at the two most proximal positions in the array (which are most likely expressed) is distinguished. Considering subjects of our study and others (Gardner et al., 2014), a total of 7 subjects fall in this subcategory with 'LIAVA' + 'LIAVA' (N = 3), 'LVAVA' + 'LVAVA' (N = 2) and 'MIAVA $^{c.465C}$' + 'MIAVA $^{c.465C}$' (N = 2) and a phenotype compatible with BCM. Progression was reported for subjects with the 'LVAVA' + 'LVAVA' genotype; BCM126-20616 (Publication II and Table 3) and F21 - S26 subject (Table 3) with reduced central vision from the second decade of life (Gardner et al., 2014). It remains unknown whether the degeneration in 'LVAVA'-bearing patients is related to toxicity exerted by the opsin protein carrying 'LVAVA' (Greenwald et al., 2017). A **second subcategory** of multi-gene arrays would comprise configurations of 'LIAVA' + 'MIAVA$^{c.465G}$' (N = 3) or 'LIAVA' + 'MIAVA$^{c.465C}$' (N = 2) with all *OPN1MW* gene copies carring the same haplotype (Table 3; our study and Gardner and colleagues). All subjects were compatible with a BCM diagnose and made errors in the colour vision test. All

haplotypes from the first and second subcategories (except in BCM72-I:1) at the two most proximal positions are considered deleterious in terms of splicing competence (<10% of correctly spliced products).

Table 3. Subjects bearing *OPN1LW/MW* multi-gene arrays

Nº	Patient ID - Family ID	n *	*OPN1LW*	*OPN1MW(s)*				Phenotype			Ref.
			Exon 3 Haplotype	Exon 3 Haplotype 1		Exon 3 Additional Haplotypes		Colour Vision	Diag-nose	Pr.	
1	BCM72-16874 (BCM72-I:1)	4	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G	LIAVS	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.521C>T; c.532A>G; c.538T>G	MVVVA c.465C	Deutan	MD§	Yes	This study
2	BCM72-17075 (BCM72-III:1)	4	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.521C>T; c.532A>G; c.538T>G	MVVVA c.465C	Protan	BCM	Yes	This study
3	F14 - S16 F14 - S17	3	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.521C>T; c.532A>G; c.538T>G	MVVVA c.465C	Errors	BCM/ XLCD	No	Gardner et al., 2014
4	BCM126-20616	2	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA			Impaired	BCM	Yes	This study
5	F21 - S26	3	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA	c.521C>T; c.532A>G; c.538T>G	MVVVA c.465C	ND	BCM /XLCD	Yes	Gardner et al., 2014
6	F20 - S24 F20 - S25	3	c.511G>A; c.513G>T; c.532A>G; c.538T>G	MIAVA c.465C	c.511G>A; c.513G>T; c.532A>G; c.538T>G	MIAVA c.465C	c.521C>T; c.532A>G; c.538T>G	MVVVA c.465C	ND	BCM/X LCD	No	Gardner et al., 2014
7	ZD379-19194 ZD379-19195	3	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.511G>A; c.513G>T; c.532A>G; c.538T>G	MIAVA c.465C	c.511G>A; c.513G>T; c.532A>G; c.538T>G	MIAVA c.465C	ND	BCM	No ND	This study
8	BCM101-19818 F19 - S22 F19 - S23	3	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	MIAVA c.465G	c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	MIAVA c.465G	Impaied ND Errors	BCM BCM /XLCD	No No ND	This study Gardner et al., 2014

*n: number of total copies in the array; Pr.: Progression; Ref.: Reference study; c.465?: the genotype for the position c.465C/G has not been specified; †progression due to high myopia; #BCM/XLIA: (Non-classical) BCM or X-linked incomplete achromatopsia; ¶incongruencies in the genotyping of the same patient between the two studies; §MD: Late-onset Macular Dystrophy (in addition to deuteranopia). ND: not determined. Adapted from Supplementary Table S1 (Publication II; Appendix 13.1.2)

Table 3 (continued). Subjects bearing *OPN1LW/MW* multi-gene arrays

Nº	Patient ID - Family ID	n *	*OPN1LW*		*OPN1MW(s)*				Phenotype			Ref.
			Exon 3 Haplotype		Exon 3 Haplotype 1		Exon 3 Additional Haplotypes		Colour Vision	Diag-nose	Pr.	
9	MOL0152-III:1 MOL0152-III:2 MOL0152-III:3	2	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.465(?); c.521C>T; c.532A>G; c.538T>G	MVVVA c.465?			ND	BCM	ND	Mizrahi-Meissonnier et al., 2010
10	11-0271 (=MF)[¶]	2	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.465(?); c.521C>T; c.532A>G; c.538T>G	MVVVA c.465?			Mainly Protan defect	BCM/ XLIA[#]	ND	Crognale et al., 2004; Greenwald et al., 2017
11	BCM51-12359	3	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.465C>G; c.521C>T; c.532A>G; c.538T>G	MVVVA c.465G	c.465C>G; c.521C>T; c.532A>G; c.538T>G	MVVVA c.465G	Protan	XLCD	ND	This study
12	Protan-opic-Family 2	3	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	LIAVA	c.465(?); c.521C>T; c.532A>G; c.538T>G	MVVVA c.465?	c.465(?); c.521C>T; c.532A>G; c.538T>G	MVVVA c.465?	Protan	BED	ND	McClements et al., 2013
13	ZD547-4544	2	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA	c.532A>G; c.538T>G	MVAVA			Protan	XLCD	No	This study
14	Protan-opic-Family 1-Minne-sota	4	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA	c.538T>G	MVAIA	c.465(?); c.521C>T; c.532A>G; c.538T>G	MVVVA c.465?	Protan	BED	ND	Young et al., 2004; McClements et al., 2013
15	BCM133-20960 BCM133-20961 BCM133-23364	2	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA	c.453A>G; c.457A>C; c.465C>G; c.538T>G	LVAIA			Errors ND ND	BCM	No No[†] ND	This study
16	Deuteran-opic Family 1 - original BED	4	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	LVAVA	c.465(?); c.521C>T; c.532A>G; c.538T>G	Hybrid MVVVA c.465?	c.538T>G	MVAIA	Deutan	BED	ND	Young et al., 2004; McClements et al., 2013
17	BCM98-19713	2	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.538T>G	LIAIA	c.538T>G	MVAIA			Errors	XLCD	No	This study
18	BCM142-21958	3	c.453A>G; c.457A>C; c.465C>G; c.556C>T	LVAISS	c.538T>G	MVAIA	c.453A>G; c.457A>C; c.465C>G; c.538T>G	LVAIA	Impaired	BCM	Yes	This study

*n: number of total copies in the array; Pr.: Progression; Ref.: Reference study; c.465?: the genotype for the position c.465C/G has not been specified; [†]progression due to high myopia; [#]BCM/XLIA: (Non-classical) BCM or X-linked incomplete achromatopsia; [¶]incongruencies in the genotyping of the same patient between the two studies; [§]MD: Late-onset Macula Dystrophy (in addition to deuteranopia). ND: not determined. Adapted from Supplementary Table S1 (Publication II; Appendix 13.1.2)

A **third subcategory** of genotypes is composed by subjects with either 'LIAVA' or 'LVAVA' haplotypes in the first gene plus haplotype(s) with an intermediate to minor effect on splicing in the second and more downstream copies in the array. For instance, the combination 'LIAVA' + 'MVVVA$^{c.465C/G}$' (N = 6) was found associated with a protan defect; diagnose varied from XLCD, BCM, incomplete achromatopisa to BED (our study and others; see Table 3). A non-classical BCM phenotype with retained M-cone function was found in the subject diagnosed with incomplete achromatopsia (Crognale et al., 2004). The most recent genotype of this subject (Greenwald et al., 2017) is shown in Table 3 (Nº 10). Residual M-cone function may stem from the low expression of a downstream functional *OPN1MW* gene copy carrying one of the 'MVVVA' haplotypes. Thus, this category may represent XLCD cases with dichromacy resembling the phenotype of BED. Adaptive optics retinal imaging performed for a 26-year-old deuteranope with a normal *OPN1LW* and an inactivating 'LIAVA' haplotype in the *OPN1MW* gene revealed an incomplete mosaic lacking one third of his cones and complete loss of M-cone function (Carroll et al., 2004; Mizrahi-Meissonnier et al., 2010). Unlike dichromats with two functional gene copies, dichromats with one of the gene copies carrying a gene inactivated by 'LIAVA' likely developed M-cones in the fovea, which secondarily degenerate and eventually die (Carroll et al., 2004). Yet, it remains uncertain whether these M-cones are truely lost, or remained invisible due to shortening of outer segments.

The 'MVVVA' haplotype, considered the most ancestral haplotype (Winderickx et al., 1993), in either the second or third gene copy but always in *OPN1MW* (often with a more detrimental haplotype in *OPN1LW*) was found in a total of 3 subjects (2 families) in our study (Appendix 13.1.2, Publication II and Table 3) and in overall 12 subjects (8 families) in other studies (Gardner et al., 2014; Mizrahi-Meissonnier et al., 2010; Crognale et al., 2004; Greenwald et al., 2017; McClements et al., 2013; Young et al., 2004).

The combination of 'LVAVA' plus an additional haplotype (N = 6) of our study and others (Table 3) was more diverse, having the second gene in the array either 'MVAVA', 'MVAIA', 'LVAIA' or 'MVVVA' haplotypes. These subjects presented with dichromacy or had errors in the colour vision test and were diagnosed with either XLCD/BCM (our study) or BED (Young et al., 2004; McClements et al., 2013). The subject ZD547- 4544 of our study was genotyped with 'LVAVA' + 'MVAVA' (Nº 13; Table 3). Having the highest visual acuity among the entire cohort and a fundus with a

myopic optic nerve, this subject mainly presented a protan defect. Recently, both ‘LVAVA’ and ‘MVAVA’ haplotypes have been found in subjects with high myopia while retaining normal visual acuity and colour vision that showed progression and macular dystrophy at middle ages (Orosz et al., 2017). Conversely, no progression was reported for the 46-year-old subject ZD547-4544 (Supp. Table. S1; Appendix 13.1.2, Publication II). However, it remains unknown whether the genotype is the same to that of our subject due to the lack of specific LD-PCRs (Orosz et al., 2017).

Finally, two subjects of our study (BCM98-19713; Nº 17 and BCM142-21958; Nº 18 in Table 3, continued) with a multi-gene array do not fall in any of the aforementioned subcategories. BCM98-19713 showed reduced photopic responses, yet not as strong as in other subjects, which may be explained by his genotype of a two-copy gene array with ‘LIAIA’ and ‘MVAIA’ haplotypes in the proximal and distal gene copy, respectively. These haplotypes result in moderate and almost negligible levels of aberrantly spliced transcripts, respectively (Table 2; Appendix 13.1.2, Publication II). Moreover, this subject performed normal on the saturated D-15 colour vision test, yet confusions in the desaturated version of the test were apparent (Supp. Table S1; Publication II). Therefore, a different pathomechanism might be trigged by these haplotypes, or alternatively, further yet not detected mutations are found in these subjects. Namely, the BCM142-21958 subject carries a c.556C>T/p.P186S missense mutation in exon 3 of *OPN1LW* (Publication II). Yet, it remains hitherto as a VUS.

The formation and integrity of the cone mosaic seems to depend on mutational type and the relative number of cones (L:M) expressing the non-functional photopigment (Patterson et al., 2016). Can the great phenotypic variability be somewhat attributed to differential expression (within an individual) and variable expressivity (among individuals) of *OPN1LW* and *OPN1MW* genes? The relative proportion of cones expressing the LWS pigment varied from 52.7 to 94.3% in males with normal colour vision (Hofer et al., 2005). A stochastic process, by which LCR binds either promoter of *OPN1LW* or *OPN1MW*, dictates the photopigment expressed (Smallwood et al., 2002; Winderickx et al., 1992). Additionally, it has been postulated that modifiers (McMahon et al., 2004) or regulatory sequences (Gunther et al., 2008) of either *OPN1LW* and *OPN1MW* may modulate the relative composition of the cone mosaic or the mutation itself.

The converted haplotype 'LIAVA' differed solely by one variant from the haplotype in the acceptor copy (BCM72-I:1). Interestingly, the first subcategory of multi-gene arrays described above, with identical haplotypes: 'LIAVA' + 'LIAVA', 'LVAVA' + 'LVAVA' or 'MIAVA $^{c.465C}$' + 'MIAVA $^{c.465C}$' at the two most proximal positions (Table 3) suggests that gene conversion could have occurred between *OPN1LW* and *OPN1MW*. The arrays of the second subcategory with the 'LIAVA' haplotype in the *OPN1LW* copy have haplotypes which differ in two and three variants from 'LIAVA' in the downstream *OPN1MW* gene copies: either 'MIAVA$^{c.465G}$' or 'MIAVA$^{c.465C}$', respectively. Interestingly, these three-copy arrays have the identical haplotype at the two most distal *OPN1MW* gene copies. Complete exon 3 conservation among all gene copies in the array was only found in subject BCM126- 20616 (Nº 4; Table 3). The other arrays in which not all copies are identical but differ from one to three variants could theoretically have underwent conversion of partial stretches or recombination and represent transitions which have not been completely homogenized.

In short, we have presented genotype and phenotype data from sixteen different BCM/XLCD families and characterized the output of twelve different exon 3 haplotypes observed in the human *OPN1LW/MW* gene cluster of these families. Attending to the splicing efficiency retained by each exon 3 haplotypes on an implemented semi-quantification minigene-based assay, we could establish three different functional categories (strongly deleterious, intermediate and minor defective). Correlation between the functional splicing output and the genotype-phenotype data was observed and used for categorization. Remarkably, correlation and evidence to pinpoint the origin of an intrachromosomal gene conversion event that was detected after the course of two generations of a single pedigree is presented.

6.4 Publication III: Retinitis pigmentosa: impact of different *Pde6a* point mutations on the disease phenotype

While in Section 6.3 combinations of exonic variants, i.e. haplotypes, were investigated, herein I describe the functional characterization of a single exonic non-synonymous variant that interferes with splicing regulation in a context-dependent manner. An important finding of this study is that the c.1684C>T/p.Arg562Trp missense mutation in the murine *Pde6a* induces aberrant splicing by in-frame skipping of exon 13. This chapter also includes related unpublished data on the effect of c.1684C>T on splicing of the human *PDE6A* gene and the effect of c.1678C>T on splicing of the murine *Pde6b* gene.

The c.1684C>T/p.Arg562Trp variant in exon 13 of *PDE6A* had been found in **compound heterozygous** state with the c.2053G>A/p.Val685Met mutation in exon 17 of *PDE6A* in a patient with retinal dystrophy. The c.2053G>A/p.Val685Met transition is considered pathogenic because it has been associated with RP (Corton et al., 2010), has a MAF of 0.0000426; a total of 5 counts for the alternate allele (ExAC), and the homologous mutation has been shown to cause retinal degeneration in mice (Sakamoto et al., 2009). In contrast, the variant c.1684C>T/p.Arg562Trp has never been described in the literature nor found in a large series of RP patients (n = 514) and controls (n = 504) (Prof. Bernd Wissinger, *personal communication*).

By means of an enzymatic assay that relies on the recombinant expression of a PDE5/6 chimeric protein in SF9 insect cells (Grau et al., 2011; Muradov et al., 2003), Dr. Tanja Koepfli showed that c.1684C>T/p.Arg562Trp results in a mutant phosphodiesterase with reduced cGMP **hydrolysis activity** of ~10% of that of the wildtype protein (Fig.1; Appendix 13.1.3, Publication III). In order to model and investigate the disease *in vivo*, a **knock-in mouse** mutant for *Pde6a*:c.1684C>T/p.Arg562Trp was generated and subsequently crossbred with a mouse line carrying the *Pde6a*:c.2053G>A/p.Val685Met to model the compound heterozygous state, homologous to the patients' genotype (p.Val685Met/ p.Arg562Trp). A targeting vector comprising a neomycin cassette was generated and used to introduce the *Pde6a*:c.1684C>T/p.Arg562Trp mutation in mouse embryonic stem cells (mESC) by homologous recombination (Sothilingam et al., 2015). Recombinant mESC clones were used to produce chimeric mice, and the neomycin

selection cassete removed from the locus by Cre recombinase-mediated excision through mating with a C57BL/6 Cre deleter line. Crossbreeding led to homozygous *Pde6a*:c.1684C>T/p.Arg562Trp knock-in animals. Removing the neomycin cassette flanked by two *loxP* sequences left a single *loxP* site and remnant sequences from the targeting vector construction residing in intron 12 of *Pde6a*. Allelic quantification of the mutation in retinal cDNA of heterozygous (Fig. 2A; Appendix 13.1.3, Publication III) performed by Dr. Sukirthini Braun-Balendran revealed an **allelic imbalance**, namely an overrepresentation of the wildtype allele. Further qualitative cDNA analysis in homozygous knock-in mutant mice (Fig. 2B; Publication III) revelead two transcript species: first, normal spliced transcripts and secondly, transcripts lacking exon 13 (Supp. Fig.S1; Appendix 13.1.3, Publication III). Approximately 2/3 of retinal transcripts derived from the knock-in allele are **mis-spliced**, according to pyrosequencing results (Fig. 2A and Fig. 2B; Appendix 13.1.3, Publication III) further supported by allelic cloning experiments. The skipping of exon 13 causes an in-frame deletion that is eventually translated into a shortened, most likely non-functional PDE6A protein lacking 36 amino acid residues (p.541_576del) at the N-terminal part of the catalytic domain including a metal binding domain (Supp. Fig. S5; Publication III). Correctly spliced transcripts still harbour the missense mutation and are expected to yield a PDE6A protein with reduced cGMP-hydrolysis activity (Fig. 1; Publication III). The total residual functional activity of PDE6 in homozygous mice carrying the p.Arg562Trp is estimated to be about 3% of the normal level.

In order to elucidate the underlying cause of the *Pde6a* splicing defect observed in knock-in mutant mice, we envisaged three feasible scenarios which could putatively account for the generation of aberrantly spliced transcripts: (1) the most straightforward explanation is that the c.1684C>T/p.Arg562Trp mutant allele *per se* leads to misplicing. (2) The second possibility could rely on the presence of the foreign sequence introduced in the flanking intron, including the *loxP* sequence, which may have altered intronic splicing regulatory sequences. (3) A third explanation could be the result of a combined or interactive effect of the two sequence alterations, i.e. the c.1684C>T/p.Arg562Trp mutation and the remnant sequences in intron 12. The three possibilities were tested by means of *in vitro* minigene assays (Fig. 7). Other than the *in vivo* characterization by the mouse model, **minigenes** easily allowed to dissect the allelic impact of altered *cis*-regulatory sequences on splicing. I assessed the splicing efficiency of four pSPL3-based minigene *Pde6a* constructs (Supp. Fig.S2; Appendix

13.1.3, Publication III) representing the possible combinations of wildtype or mutant and modified or unmodified sequences in exon 13 and intron 12: i) the wildtype allele c.1684C, ii) the c.1684T mutant allele, iii) the wildtype allele c.1684C flanked by the *loxP* site and targeting vector remnants, and iv) the c.1684T mutant allele flanked by the *loxP* site and targeting vector remnants.

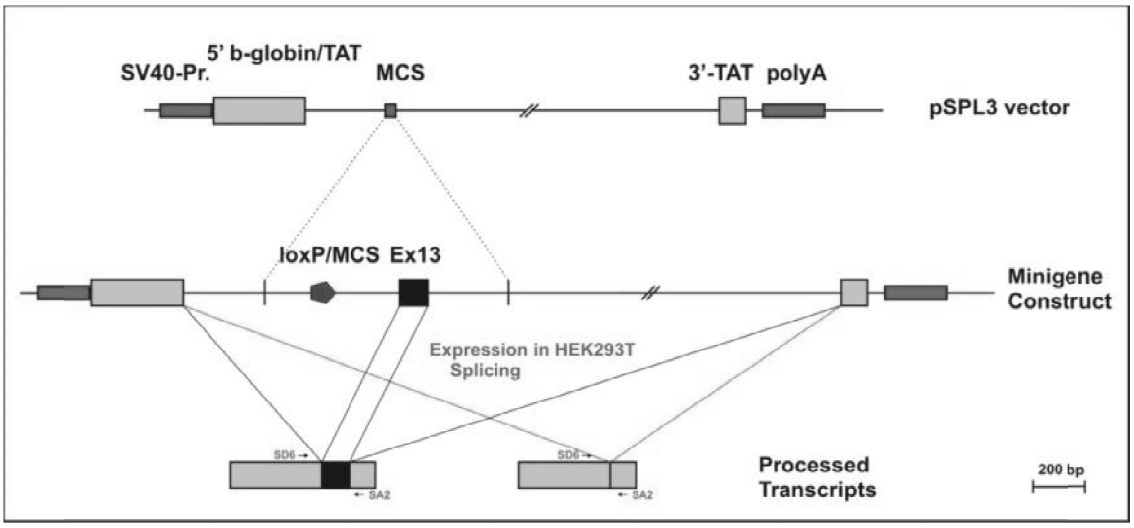


Figure 7. Sequence map of the pSPL3 vector and the expected splicing outcomes from *Pde6a* minigene constructs. Constructs refer to Material and Methods; Publication III (Sothilingam et al., 2015). Thick boxes represent exons of the pSPL3 vector (grey; 5' b-globin/TAT and 3'-TAT) or introduced by cloning (black) into the multiple cloning site (MCS) for the splicing assay (Ex13; *Pde6a* exon 13). Regulatory sequences for proper expression of vector-bearing sequences (SV40-Promoter and polyA signal) are depicted by thin dark grey boxes. The *loxP* site and remnants of the targeting construct (a grey polygon) between the 5' b-globin/TAT exon and the exon 13 (Ex13) were included in two out of the four scenarios tested *versus* two minigene constructs lacking such exogenous sequence. The location of primers SA2 und SD6 used for RT-PCR of transcripts is indicated by arrows above and below the boxes that represent the processed differentially spliced transcripts. Exon 13 inclusion is depicted showing exon 13 as a black box in between the SD6 and SA2 primers and exon 13 skipping event is denoted by the lack of the black box between the SD6 and SA2 primers. Figure credits: Prof. Bernd Wissinger.

The outcome of the minigene assays clearly demonstrated that **c.1684C>T** is able to **induce skipping** of exon 13 in the mouse *Pde6a per se* (Supp. Fig.S3; Appendix 13.1.3, Publication III) and that the foreign sequence remnants in intron 12 do not influence the inclusion of *Pde6a* exon 13. Bionformatic analysis with the *in silico* tool SROOGLE (Schwartz et al., 2009) predict that the c.1684C>T transition introduces an **ESS site** (from hexamer -TGGCGG- to -TGGTGG-; alleles at position c.1684 underlined). Moreover, SpliceAid 2 (Piva et al., 2012) predicts that the RRM of hnRNPP (also named TLS or FUS) recognizes the newly created motif (-UGGUGG-) and could presumably inhibit splicing.

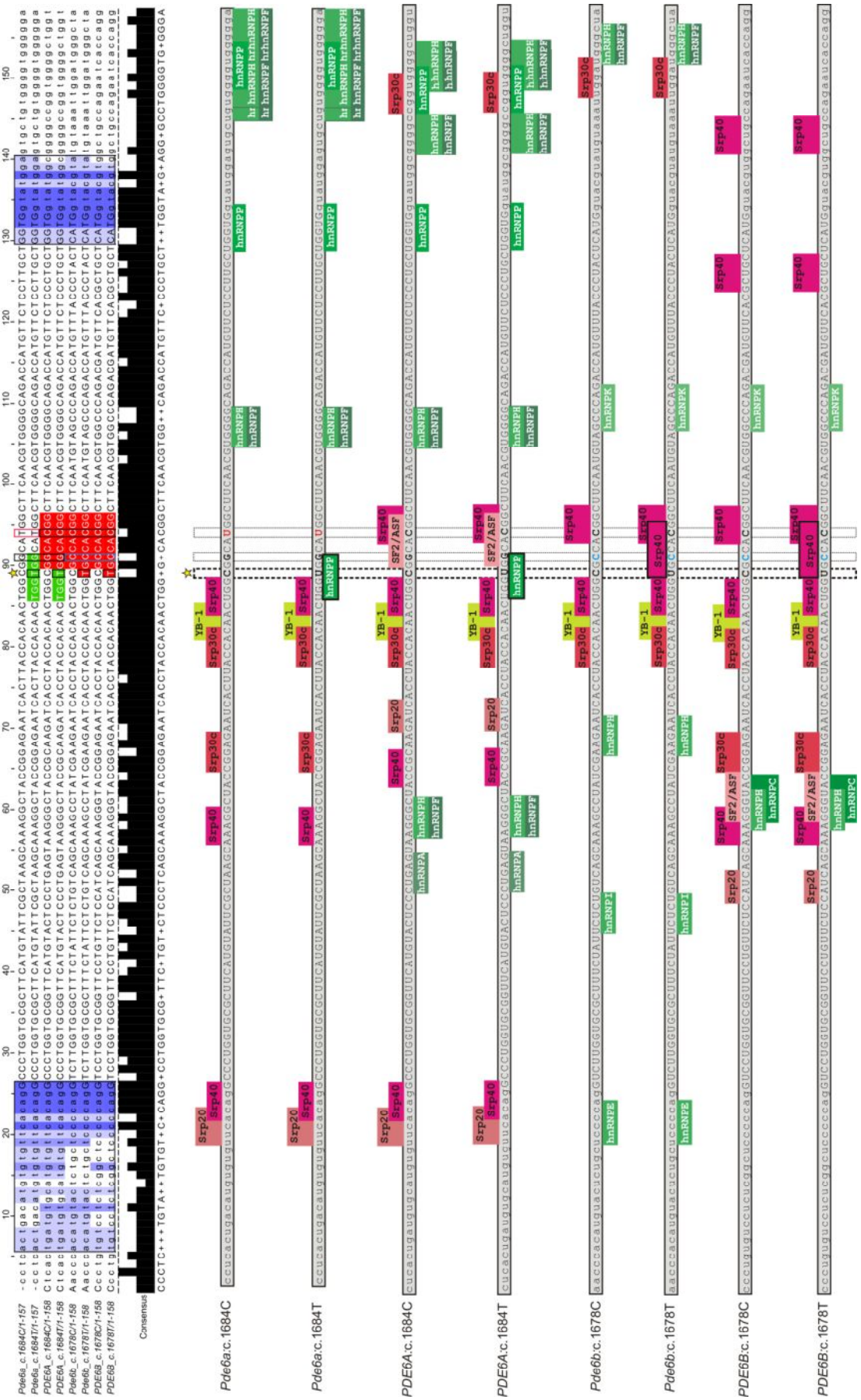


Figure 8 (figure legend in the next page; see also Fig. 7; Appendix 13.1.5).

Figure 8. Multiple sequence alignment of exon 13 in *Pde6a*, *PDE6A*, *Pde6b* and *PDE6B* genes. Either the wildtype or the mutant alleles are shown at positions c.1684 and c.1678 of *Pde6a*/*PDE6A* and *Pde6b*/*PDE6B*, respectively (position indicated by a star above the alignment). Approximately 25 bp of intronic sequences flanking exon 13 are shown at each splice site. The consensus is high at the splice sites (black bars below the alignment) showing full conservation among all 4 genes in 4 bases at the acceptor site (+1,-1,-2 and -3) and 5 bases at the donor site (-2 to +3) which are shown in dark blue (acceptor on the left, donor on the right). High consensus is found in a block of ~35 bp flanking the mutation with few variants. The c.1684/c.1678 site is marked with a star symbol. The mutation c.1684C>T creates an ESS (-TGGTGG-; depicted in green) in both *Pde6a* and *PDE6A*. The homologous mutation c.1678C>T in *Pde6b* and *PDE6B* is embedded in an ESE-rich region (depicted in red). In *Pde6a* and *PDE6A* the c.1686G synonymous variant is part of the ESS (-TGGTGG-) two bases downstream from the mutation. The homologous site carries the c.1680C (ortholog-specific) variant in both *Pde6b* and *PDE6B* which prevents the creation of an ESS in the presence of the c.1678T mutation (-TGGTGC-). Five bases downstream of the mutation, the c.1689T variant is only present in *Pde6a*. The c.1689C in the human ortholog *PDE6A* as well as the c.1683C of the murine *Pde6b* and human *PDE6B* at the homologous position create an ESE site (depicted in red), which is not created for *Pde6a*. At the bottom panel the main splicing factors predicted (SpliceAid 2 (Piva et al., 2012) and Human Splicing Finder (Desmet et al., 2009) to bind to each of the pre-mRNA of exon 13 with enhancer (depicted as boxes above the sequence) or silencer (depicted as boxes below the sequence) activities are shown. The hnRNPP splicing factor is predicted to bind the ESS (-UGGUGG-) created by the c.1684C>T mutation exclusively in exon 13 of *Pde6a* and *PDE6A*. In *PDE6A* next to the c.1684C>T mutation where hnRNPP is predicted to bind, two additional SR splicing factors are predicted: SRp40 and SF2/ASF which were not predicted in the *Pde6a* mutant sequence. The homologous mutation in the ortholog human and murine genes is predicted to recruit a SRp40 splicing factor creating an ESE-rich landing pad for enhancer splicing factors. Although splice site consensus is conserved, the predicted splicing factors binding to the intronic sequences differ substantially; for example, an intronic ESS-rich sequence flanking the donor site creates motifs for multiple hnRNPs in *Pde6a* and, to a lower extent in *PDE6A*. Multiple sequence alignment was performed with ClustalO (Sievers et al., 2011) and visualized with Jalview (Waterhouse et al., 2009).

Notably, hnRNPP is expressed in human and mouse retina (FANTOM5; The Human Protein Atlas website and Mouse Genome Informatics). Nevertheless, it is currently unknwon to which extent species-specific differences in expression levels of hnRNPP or other splicing factor have an impact on splicing efficiency. The quantitative differences observed in the minigene assays with murine *Pde6a* constructs transfected in either murine or human cell lines (HEK293T and 661W, respectively; Supp. Fig. S4, Appendix 13.1.3, Publication III) could result from differences in the affinity and/or the expression levels of core or auxiliary splicing factors. Congruent results were achieved for murine *Pde6a* minigenes spliced in murine cone photoreceptor-derived cells, yet leaky misplicing was observed for the wildtype c.1684C allele of the murine *Pde6a* minigene construct in human HEK293T cells (Supp. Fig. S3; Appendix 13.1.3, Publication III).

Notwithstanding, cross-species incongruencies in exon 13 inclusion of the murine *Pde6a* and its **human ortholog** *PDE6A* may not only stem from differences in stoichiometry and expression of *trans*-splicing factors, but also from differences in *cis*-elements in the nucleotide sequence of gene ortholologues which may influence the recruitment of splicing factors. The amino acid sequence in the vicinity of p.Arg562 is fully conserved between murine and human, yet a synonymous T-to-C transition is present 5 bp downstream of c.1684C>T mutation site in the human *PDE6A* (3 bp downstream of the created ESS which is highlighted in green in the sequence alignment; Fig. 8). This difference in the human sequence may create an **ESE site** for SRp40/SRSF5 as predicted by SpliceAid 2 (highlighted in red in the sequence alignment; Fig. 8), absent in the murine *Pde6a*.

The sequence context in *PDE6A* may favour exon 13 inclusion, counteracting the c.1684C>T mutation, and thus might explain why human *PDE6A* minigenes harbouring the c.1684C>T mutation yielded much lower levels of mis-spliced transcripts in comparison with the corresponding murine *Pde6a* minigene construct (Fig. 9 and Appendix 13.2.3, Publication III; Supp. Fig. S1).

Wild-type and mutant minigene constructs with the human sequence of *PDE6A* for heterologous expression in HEK293T cells were generated (Appendix 13.2.3; Supplementary material Section 6.4) as described for the *Pde6a* minigenes (Material and Methods section, Publication III). Both the 371-bp and 263-bp RT-PCR amplicons corresponding to correctly are aberrantly spliced *PDE6A* transcripts, respectively, are detected for minigenes carrying the c.1684C>T mutation (Fig. 9). The c.1684C allele results in complete exon 13 inclusion (Fig. 9).

Splicing motifs according to SpliceAid 2 predictions are well conserved at the region close to the 5'ss donor (3' terminal part of exon 13) as well as within the exonic core between *Pde6a* and its human ortholog, except for the hexamer sequences adjacent to the c.1684C>T in the middle of exon 13 (Fig. 8).

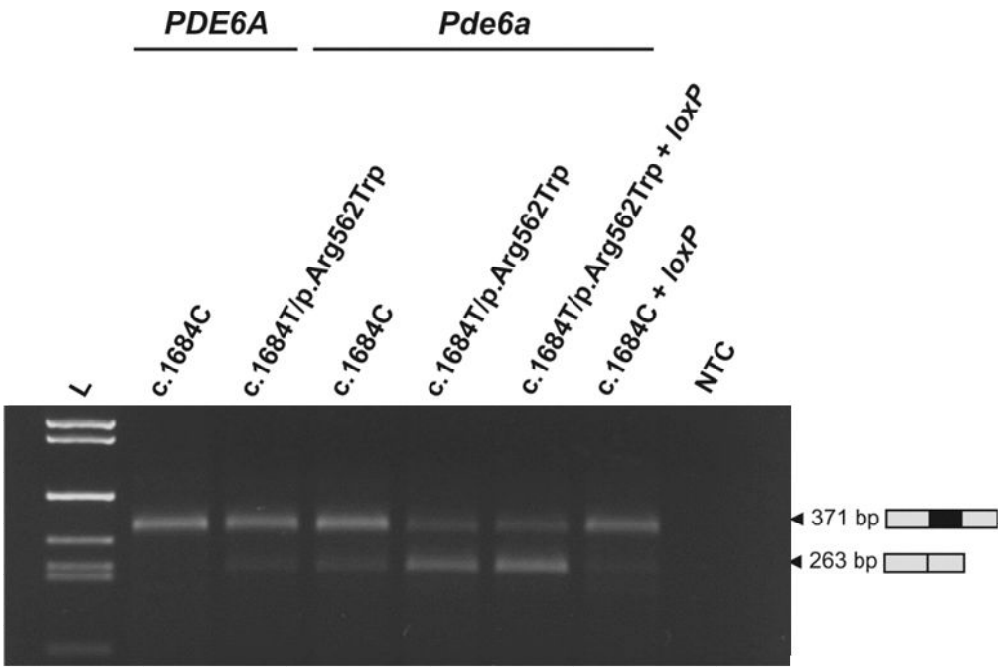


Figure 9. RT-PCR products from *PDE6A* and *Pde6a* minigene constructs generated to assess the effect of the c.1684C>T/p.Arg562Trp mutation. RT-PCR products obtained with RNA isolated from HEK293T cells transfected with either human *PDE6A* or murine *Pde6a* exon 13 minigene constructs were separated on a 2% agarose gel. The cDNA with retained exon 13 (spliced to the flanking *tat* exons of the vector) results in a RT-PCR product of 371 bp which runs above the 333 bp band of the ladder ("L": KEB ladder use as size standard). The cDNA fragments with skipped exon 13 render a PCR product of 263 bp in size which runs at virtually the same height of the 271 bp band of the ladder ("L"). The human minigene constructs carrying either the wildtype ("c.1684C") or the mutant ("c.1684C>T/p.Arg562Trp") allele are shown below the line labeled as "*PDE6A*". The four mouse minigene constructs are shown below the "*Pde6a*" label; the wildtype ("c.1684C"), the mutant ("c.1684C>T/p.Arg562Trp"), the mutant including the *loxP* site and remnants of the targeting vector in intron 12 ("c.1684C>T/p.Arg562Trp + *loxP*"), and the wildtype combined with the *loxP* site and remnants of the targeting vector in intron 12 ("c.1684C + *loxP*"). Controls: no RNA – no RNA added to reverse transcription reaction; NTC – no template control in the PCR.

Furthermore, intra-species differences in splicing *robustness* across exon 13 of murine **paralogous genes** (*Pde6a* and *Pde6b*) carrying mutations at homologous positions (c.1684C>T/p.Arg562Trp and c.1678C>T/p.Arg560Cys) were noticed. The amino acid sequence between *Pde6a* and *Pde6b* in this region is highly conserved. The c.1678C>T/p.Arg560Cys mutation in *Pde6b* in *rd10* mice (Chang et al., 2007) is a C-to-T transition in the analogous codon to the c.1684C>T/p.Arg562Trp mutation. Under the hypothesis that the analogous mutation could likewise compromise splicing, we analysed the integrity of *Pde6b rd10*-derived retinal transcripts (Fig. 10).

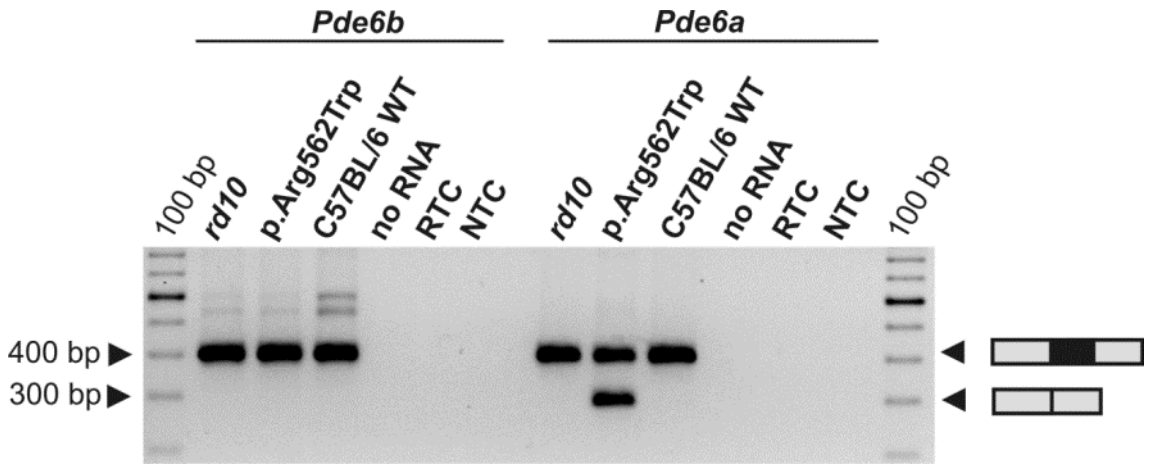


Figure 10. Qualitative assessment of *Pde6a* and *Pde6b* retinal transcripts in mice homozygous for either c.1684C>T/p.Arg562Trp (p.Arg562Trp) or c.1678C>T/p.Arg560Cys (*rd10*) and wild-type mice (C57BL/6 WT). The ~400-bp amplicon represents correctly spliced transcripts including exon 13. The amplicon 108-bp shorter than the latter, observed from *Pde6a* retinal transcripts of p.Arg562Trp animals, corresponds to transcripts lacking exon 13 (depicted as a filled black box between flanking exons in grey). Negative controls (no RNA: no input RNA included for primer hybridization for the cDNA synthesis, RTC: only master mix for reverse transcription, and NTC: no template control for the PCR amplification). The 100-bp ladder is shown in each side of the agarose gel. For supplementary methodology applied, see Appendix 13.2.2.

Surprisingly, we did not detect any mis-spliced *Pde6b* transcripts encompassing exon 13 in the *rd10* mutant. *Pde6a* and *Pde6b* sequences differ in the vicinity of the homologous mutations only at two sites: (1) a T-to-C transition (c.1689T in *Pde6a* and c.1683C in *Pde6b*) five bases downstream of the mutation(s) potentially reconstructing an **ESE** (see Fig. 8 in this chapter and Fig. 7 in our review; Weisschuh et al., 2020; Appendix 13.1.5; Publication V), and (2) two bases downstream of the homologous mutation(s) a G-to-C transversion (c.1686G in *Pde6a* and c.1680C in *Pde6b*); which **prevents the creation of the ESS** motif (-TGGTGG-). Conversely, the hexamer -TGGTGC- (c.1680C allele underlined) in *rd10* mutant *Pde6b* has not been attributed to act as enhancer or inhibitor (regardless of the mutation c.1678C>T*)*. In turn, when both c.1678C>T mutation (*rd10*) and the c.1680G>C ortholog-specific variant are together (in *cis*), these are embedded in an ESE motif, which is inexistent in the wildtype (c.1678C) of both murine and human gene ortologues *Pde6b* and *PDE6B* (Fig. 8). Moreover, the aforementioned ESE is not created by the homologous sequence in either wildtype or mutant (c.1684) of both murine and human gene orthologues *Pde6a* and *PDE6A*. In other words, the mutation c.1678C>T creates that particular **ESE** exclusively in the sequence context of *Pde6b* and *PDE6B* (Fig. 8;10-11) promoting thereby inclusion of exon 13.

In contrast, the opposite effect – exon 13 skipping – is exerted by the c.1684C>T mutation in both murine *Pde6a* and human *PDE6A.* An ESE-rich sequence context surrounding the *rd10* mutation (Fig. 8 and Fig. 11) might explain the disparity of splicing robustness observed between virtually *equivalent* mutations of two paralog genes in the same species. These findings underscore the importance of context-dependent sequences in the final overall interpretation of splicing mutations, and the generally underestimated impact of synonymous variants.

Discrepancies in the splicing efficacy of the same minigene constructs between 661W and HEK293T cell lines could be subjected to tissue- or species-specific regulatory mechanisms. The wildtype *Pde6a* allele results in a very similar splicing outcome when expressed in the murine retina or when harboured by mouse minigenes and expressed in 661W mouse photoreceptor-derive cells (same species, same or at least similar tissue/cell type), yet different to that of mouse minigenes expressed in human embryonic kidney-derived HEK293T cells. This *mismatch* may explain leaky exon skipping of the wildtype allele c.1684C harboured by the murine *Pde6a* minigene expressed in HEK293T (Supp. Fig.S4; Appendix 13.1.3, Publication III). If one would assume that the underlying crucial factor are only tissue- or photoreceptor-specific splicing regulation, a similar pattern of leaky exon skipping would be expected upon expression of the human wildtype *PDE6A* minigene in HEK293T. Contrary to this assumption, the human wildtype *PDE6A* minigene is efficiently and correctly spliced in HEK293T cells (Fig. 9). Thus, the disparity seems to be mainly determined by the species-dependent *cis*-regulatory sequence and the counterpart *trans*-acting splicing factors recruited. One of the most likely determining elements is the *cis*-sequence of an enhancer in the proximity of the c.1684C>T mutation in the human *PDE6A* sequence context which is absent in the murine *Pde6a* sequence due to the synonymous variant at c.1689 (Fig. 8).

The 5’ ss sequence of exon 13 in the murine *Pde6a* has 6 bases complementary out of the 11 bases of the U1 snRNA in comparison to 5 bases for the human *PDE6A*. Thus, *robustness* of splice **donor** site in exon 13 is similar for both gene orthologues and most likely do not largely account for the the effect observed. Of note, the stoichiometry of splicing factors and the affinity or specificity to certain binding sites may vary within species. By comparing retinal transcripts carrying homologous mutations of *Pde6a* and *Pde6b* one could ignore species-dependent and tissue-

specific differences (Fig. 10-11). Consequently, the *cis*-sequence context surrounding the mutation is presumably determinating the overall discordant splicing outcome.

We could then extrapolate the results from the mouse retina to the murine minigenes transfected in mouse cells and to the human minigenes in human cells.

Exon 13 skipping leads to *Pde6a* transcripts bearing an **in-frame deletion**. These transcripts are translated in a shortened Pde6a protein lacking a stretch of 36 amino acid residues (p.541_576del). The deleted segment encompasses the N-terminus of the catalytic domain (PFAM: 00233), including residues known to be essential for the enzymatic function of the human PDE6A (UniProt: P16499). The mutation is embedded in a metal ion binding motif His-Asn-Trp-Arg-His at p.559-563 that interact with zinc ions (He et al., 2000). The metal ion binding motif at p.559-563 is highly conserved from mammals to fish (Supp. Fig.S5; Appendix 13.1.3; Publication III).

The c.1684C>T mutation could be considered **hypomorphic** in comparison for example with a full gene deletion (*null* allele) since not every *Pde6a* transcript is aberrantly spliced. About 1/3 of the transcripts include exon 13 and result in a mutant Pde6a protein carrying the p.Arg562Trp substitution which decreases the catalytical activity of the mutant protein. Thus, the mutation exerts a **dual effect** on splicing and protein function. Nevertheless, residual levels of phosphodiesterase activity are retained in p.Arg562Trp mutants. In agreement, the phenotype observed in homozygous and heterozygous mice carrying the mutation is not as severe as that of other missense mutations such as the c.2053G>A/p.Val685Met mutation. The **rod degeneration** in homozygous mice for c.1684C>T/p.Arg562Trp starts about 4 days later than homozygous mice for c.2053G>A/p.Val685Met and with 2 days offset (later) from the intermediate compound heterozygous mice with the latter two mutations; see Fig, 4E, J, O); Appendix 13.1.3; Publication III (Sothilingam et al., 2015). Homozygous mice for c.1684C>T/p.Arg562Trp showed a less milder phenotype than those carrying the c.2009A>G/p.Asp670Gly (Sakamoto et al., 2009) in both alleles which exhibited the slowest rod degeneration profile (Sothilingam et al., 2015).

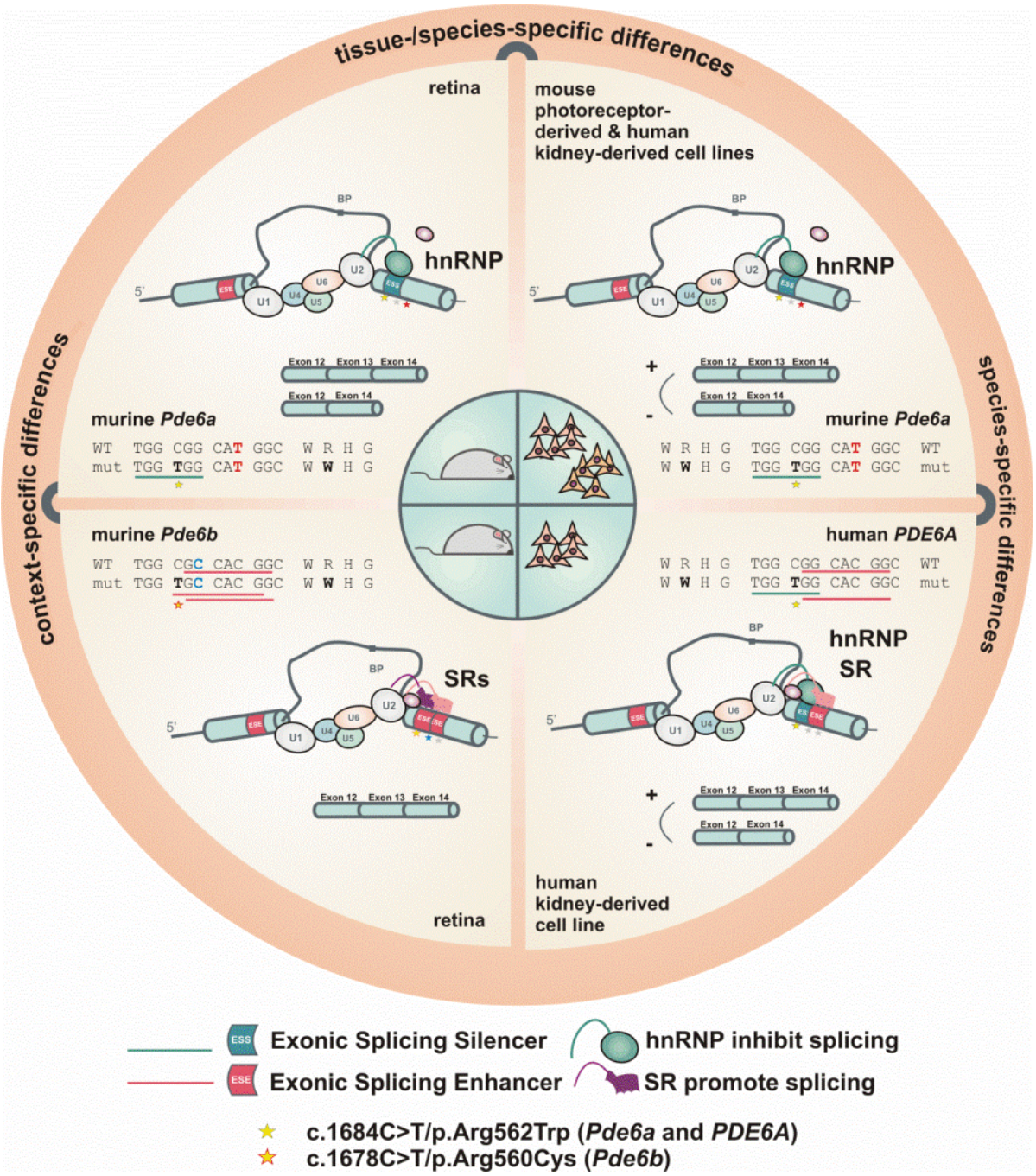


Figure 11. *Cis*-regulatory elements of splicing in exon 13 of *Pde6/PDE6*. The discrepant effect of the analogous mutations, c.1684C>T/p.Arg562Trp in *Pde6a* and c.1678C>T/p.Arg560Cys in *Pde6b* (black bold letters) on *in vivo* splicing in the murine retina putatively involve *cis*-regulatory sequence elements (upper left and bottom left, respectively). The c.1684C>T/p.Arg562Trp mutation in exon 13 of *Pde6a* is predicted to create an Exonic Splicing Silencer (ESS) depicted by the green line spanning a hexamer including the mutation. The c.1678C>T/p.Arg560Cys mutation in *Pde6b* is embedded within a hexamer predicted to act as an Exonic Splicing Enhancer (ESE), which is shown as a red line. Note the difference in the context sequence (two synonymous variants; in red and blue bold letters) in the vicinity of the homologous positions c.1684 and c.1678 in *Pde6a* and *Pde6b.* One of the synonymous variants creates a second ESE hexamer in both wildtype and mutant *Pde6b* (see also Fig. 8). ESEs motifs suggest binding of SR proteins to exon 13 of *Pde6b*, thereby increasing exon recognition by interactions with the core spliceosomal proteins. Complete exon retention is observed for both wildtype c.1678C and c.1678T *Pde6a* alleles (Fig. 10). *Figure legend continues in next page*.

Figure 11 (legend continued). The ESS in *Pde6a* is created by the c.1684T mutant allele in conjunction with the downstream synonymous variant. The ESS putatively recruits hnRNP proteins inhibiting *Pde6a* exon recognition, and resulting in partial exon skipping as detected by retinal transcript analysis of homozygous *Pde6a*:c.1684C>T/p.Arg562Trp mutant mice (Appendix 13.1.3; Publication III; Fig.2B). In comparison, *Pde6a* wildtype and c.1684C>T mutant minigene constructs were expressed in HEK293T and 661w cells (upper right). In this scenario, the wildtype c.1684C construct yielded a certain extent of exon skipping in HEK293T, likely due to tissue-/species-specific differences in splicing regulation. Alongside, a comparison of minigene-based heterologous assays in HEK293T for human *PDE6A* sequence (bottom right) is shown. Note that exon 13 of the human *PDE6A* shares one synonymous variant with the murine *Pde6a* – which results in the creation of an ESS when the mutation c.1684T is present –, and one synonymous variant with the murine *Pde6b* – which creates an ESE. The mutation c.1684C>T/p.Arg562Trp induces a rather weak exon 13 skipping in *PDE6A*, as revealed by minigenes expressed in HEK293T cells (Fig. 9).

In conjunction with c.2053G>A/p.Val685Met, the c.1684C>T/p.Arg562Trp explains the retinal phenotype observed in the mouse model which mirrors that of the RP patient. A double pathomechanism is exerted by the c.1684C>T/p.Arg562Trp. Similarly, partial impairment of splicing and reduced catalytic activity is exerted by the c.2368G>A/p.Glu790Lys mutation found in compound heterozygous state in *PDE6C* in two siblings affected with achromatopsia (Chang et al., 2009).

Previously, a *Pde6b* missense mutation c.1814A>G/p.Asn605Ser, 19 bp upstream of the donor site, has been shown to result in out-of-frame exon 14 skipping in ~80% of retinal transcripts among other aberrantly spliced transcripts. Correctly spliced *Pde6b* transcripts were detected as low as 5% in the slow-degenerating *atrd3* mice (Muradov et al., 2012). The expression of the analogous already spliced transcript in the human cone *PDE6C,* p.Asn610Ser, did not impair photoreceptor function in transgenic *Xenopus laevis* (Muradov et al., 2012) suggesting that the Asn-to-Ser substitution is benign. The pathogenic effect is caused by a splicing defect in the murine *Pde6b* and it remains uncertain whether this effect is recapitulated in the human *PDE6C* gene.

Differences in splicing regulation among vertebrate species have been noted; with a more complex regulation in the primate lineage (Barbosa-Morais et al., 2012). Divergence of tissue- or organ-specific splicing might be in some instances less determining than **species-specific** splicing signatures. Indeed, splicing profiles are better conserved within a species than those of the same organ or cell type in different species. Disconcordance in 13 events of exon skipping between human and mouse were predominatly assigned to the divergence of *cis*-regulatory variants (Barbosa-Morais et al., 2012). While mouse *Smn* and human *SMN* exhibit a high degree of concordance in minigene-based assays (Gladman et al., 2010), skipping of exon 7 as

observed in the human *SMN2* is bypassed in porcine *Smn1* due to the lack of an ISS in intron 7 (Doktor et al., 2014). Disconcordance in *in vivo* liver splicing profiles of a mutation altering the 5' ss in patients affected with ornithine transcarbamylase deficiency with those of the *homologous* mouse model has been attributed to differences in the genomic context (Rivera-Barahona et al., 2015). The importance of sequence context is further highlighted by the action of compensatory variants. A synonymous variant in *MCAD* drives a compensatory effect by inactivating an ESS, thereby preventing exon skipping elicited by a missense mutation that inactivates an ESE in exon 5, which leads to medium-chain acyl-CoA dehydrogenase deficiency, a frequent defect in the β-oxidation of fatty acids in human mitochondria (Nielsen et al., 2007).

Inter-species differences in splice site recognition and splicing regulation as well as **tissue or cell-type specific** splicing patterns have been reported for numerous retinal expressed genes including IRD disease genes (Garanto et al., 2011). For instance, the deep intronic c.2991+1655AG mutation in *CEP290*, frequently observed in patients with Leber congenital amaurosis (LCA) patients induces the insertion of a cryptic exon in about 50% transcripts (den Hollander et al., 2006). Analysis of a humanized knock-in mouse model carrying >6 kb of human sequence, including the c.2991+1655AG mutation, revealed an unexpected splicing profile, particularly in the retina, and overall much lower levels (15%) of aberrantly spliced transcripts (Garanto et al., 2013). In contrast, non-human primate derived cell lines expressing *CEP290* minigenes recognized the cryptic exon and performed much closer to that of LCA patients (Garanto et al., 2015), congruent with a more sophisticated splicing machinery in recent evolutionary lineages (Barbosa-Morais et al., 2012). Strengthening splice sites of the cryptic exon allowed its recognition also in murine cells (Garanto et al., 2015).

Conversely, recognition of cryptic splice sites might be in particular cases specific to a cell type, regardless of the species. The aberrant splicing profile elicited by the IVS1-2A>G *BBS8* mutation expressed from murine minigenes or BACs with the full-length human *BBS8* in mouse photoreceptors was not recapitulated by murine minigenes expressed neither in HEK293T nor in inner neuron cells of the retina. In the latter two cell types, minigenes result in exon skipping regardless of carrying the wildtype or mutant allele. Thus, the basal splicing profile of the wild-type allele in *BBS8* is likely photoreceptor-specific. Seemingly, the use of a cryptic 3'ss downstream of normal

acceptor site is exclusive to photoreceptors which explains the confined phenotype of non-syndromic RP elicited by this mutation (Murphy et al., 2015).

Retina-specific *RPGR* transcripts have been described (Neidhardt et al., 2007); including the inclusion of an alternative exon (ORF15) which is exclusive to the retina (both in humans and mice) and harbour the majority of known disease-causing mutations in *RPGR*. Alternatively spliced genes represent 24% of all murine retinal genes, according to a transcriptome analysis at postnatal day 21 (Gamsiz et al., 2012). In comparison, a recent transcriptome of the adult normal human retina revealed ~80,000 novel **alternative splicing** events of which 2,000 are putatively **retina-specific** (Farkas et al., 2013). Skipping of one or two exons were the most common alternative splicing events (35%). Yet, only 15% of these alternatively spliced transcripts maintain their open reading frame (ORF) beyond the skipped exons (Farkas et al., 2013). Recently, ubiquitiously expressed genes, such as *Cep290*, *Cc2d2a*, *Ttc8* and *Prom1*, have been shown to harbour photoreceptor-specific “switch-like” exons whose inclusion is increased during the developmental stage prior to maturation of outer segments in the mouse retina (Murphy et al., 2016).

In short, a mouse model mimicking the phenotype driven by compound heterozygous mutations underlying human arRP prompted us to investigate the impact of c.1684C>T/p.Arg562Trp at the retinal transcript level in *Pde6a*. We dissected how aberrantly *Pde6a* spliced transcripts could be generated *in vitro*, in either HEK293T or 661w cell lines. We pursued to validate our findings by assessing the effect of analogous mutations in either human *PDE6A* minigenes or murine *Pde6b* (*rd10*) retinal transcripts. Moreover, a comparison to deduce the differences within the sequence context among paralog and ortholog genes encoding for PDE6 was presented. A twofold explanation of the quantitative impairment of PDE6A is driven by the c.1684C>T/p.Arg562Trp at post-transcriptional and protein level in both *Pde6a,* and most likely also in the human *PDE6A,* leading to a milder phenotype than animals homozygous for c.2053G>A/p.Val685Met. The experiments and predictions presented suggest that aberrant splicing in murine *Pde6a* is modulated to a different extent by the neighbouring synonymous variant(s) present in the human *PDE6A* or the murine *Pde6b* genes. This may impact the relative contribution of the two pathomechanisms driven by the mutation in the two species. The enzymatic activity of the full-length *PDE6A* might be the counterparty that is impaired to higher levels in humans.

6.5 Unpublished II: Development of a multiplexed splicing assay

6.5.1 Introduction

As described in Section 3.3.2, variant interpretation remains a major challenge. The only criteria that can be applied for all variants in a proactive manner are *in silic*o prediction tools (Weile and Roth, 2018). Hitherto, functional assays are mostly designed on a **one-by-one** basis rationale for single targeted mutations in the gene of interest and performed either side by side (in different tubes or wells) and/or at a different time point in independent experiments. Although supporting evidence can be attained, assessing *one variant at a time* represents a tedious bottleneck, entails great efforts to reduce assay variability and reproducibility and it is bias- and error-prone. Therefore, only few variants currently undergo functional assessment, limiting the overall impact of NGS-derived data on the medical management of patients. Accommodating the unceasing number of newly discovered VUS to this *post hoc* approach is impractical, expensive and time-consuming. Functional assays aiming to reduce the gap from variant sequencing to accurate interpretation and molecular diagnosis are required (Starita et al., 2017).

Mirroring the burgeoning advances that sequencing has achieved simply by **parallelization**, functional assays can similarly be performed by multiplexing several variants in single assays. *A priori* (in advance) testing of gene variants from diseased and healthy individuals simultaneously would generate a comprehensive resource for variant interpretation (Section 3.3.2). For variants deposited in databases, functional evidence adds value to the revision of the assigned interpretation and categorization *a posteriori*. Coupling NGS to well-established assays enables parallelization of screenings that query functionality or expression influenced by particular protein or gene features at a single-base pair resolution. **Multiplexed assays for variant effect** (MAVEs) are emerging as an eminent solution for the challenge of variant interpretation (Gasperini et al., 2016). MAVEs can be applied proactively and are foreseen to enable functional assays at an unprecedented scale (Weile and Roth, 2018). Pioneering papers on **massively parallel reporter assay** (MPRA) assaying the effect of variants on mammalian enhancers have been published as early as 2012 (Melnikov et al., 2012; Patwardhan et al., 2012). Thenceforward, multiple functional features including protein activity, e.g. *BRCA1* ligase and binding activity upon deep mutational scanning (Starita and Fields, 2015; Starita et al., 2015), the effect of all

9,595 possible missense variants on peroxisome proliferator-activated receptor γ (PPARγ) activity (Majithia et al., 2016), or the impact of 7,244 missense PTEN (Phosphatase and tensin homolog) variants on lipid phosphatase activity (Mighell et al., 2018), and regulatory elements, e.g. *cis*-regulatory elements of expression in the mouse retina (Shen et al., 2016), effects of CNV (Tai et al., 2016) or non-coding regulatory elements (Elkon and Agami, 2017) and its alteration by large deletions (Gasperini et al., 2017) have been tested by MPRAs/MAVEs.

As mutations affecting regulatory elements, including *cis*-regulatory elements of splicing, are difficult to predict *a priori*, MPRAs can be especially useful for screening putative spliceogenic mutations. Prior to the herein presented development of a multiplexed minigene splicing assay for *OPN1LW/MW* exon 3, the only reported **splicing assay** *in batch* had aimed to test the effect of all possible random hexamers (Ke et al., 2011) and the obtained data were used to generate a prediction algorithm. Yet, this work lacked the genomic context. In the meantime, a number of MPRA applications on splicing have been reported; the assessment of two million of short synthetic sequences in alternatively spliced constructs (Rosenberg et al., 2015), the impact of synonymous variants on exon inclusion (Bhagavatula et al., 2017; Mueller et al., 2015), the unbiased analysis of all possible substitutions on an alternatively spliced exon (Julien et al., 2016), the evaluation of ~5,000 exonic variants by a so-called Massively Parallel Splicing Assay (MaPSy) (Soemedi et al., 2017), and recently, the testing of 2,059 variants in 110 alternatively spliced exons by Variant exon sequencing (Vex-Seq) (Adamson et al., 2018) and the examination of 27,733 variants from 2,339 human exons by Multiplexed Functional Assay of Splicing using Sort-seq (MFASS) (Cheung et al., 2018). All approaches mentioned above use different design and methodology, yet all made use of short-read sequencing.

We designed a novel *in vitro* minigene assay to assess splicing competence of *OPN1LW/MW* exon 3 haplotypes (Section 6.3; Appendix 13.1.2) in parallel. We chose to use random molecular barcodes and to perform targeted mutagenesis restricted to two alternative bases per variant site to generate a library of exon 3 haplotypes (Supp. Fig. S1-S3; Appendix 13.1.4). In this section of the thesis, I summarize the **technical** efforts undertaken, including failures and successes until the generation of a pilot MPRA method that suits the particularities of our design and feasibility of application as further described in Section 6.6; Appendix 13.1.4, Publication IV and Appendix 13.1.5, Publication V.

6.5.2 Results

Input Library generation

Barcode and exon 3 haplotype libraries were generated in parallel. The barcode library was cloned downstream of exon 6 of the original minigene construct as described by Ueyama and colleagues (Ueyama et al., 2012) and subsequently merged with the exon 3 haplotype library (Supp. Fig. S1; Appendix 13.1.4, Publication IV). Hybridization of a 91-mer ultramer oligonucleotide including the 8N ('NNNNNNNN') barcode tag with the 21-mer seed oligonucleotide and the extension reaction to generate a double-stranded DNA (dsDNA) molecule were optimized by trying different elongation times (3, 5, 7, 10, 15, 30, 45, 60 and 90 min). A 60 min elongation time was sufficient for complete extension as confirmed by resistance to S1 nuclease digestion. The barcode pool of 91 bp dsDNA was cloned into the reference minigene plasmid by In-Fusion Cloning (IFC) and confirmed by *Sal*I (introduced in the 91-mer) digestion and Sanger sequencing. The IFC reaction was optimized to achieve high cloning efficiency by testing different molar ratios of vector to insert molecules. With a 1:20 ratio, a maximum yield of 3.5×10^6 transformants or colony forming units (cfu) per µg (cfu/µg) was obtained and 100% of clones were positive for the barcode insert as assessed by colony PCR screening (n = 30) and subsequent Sanger sequencing. Outgrowth of colonies in liquid culture was omitted to avoid out-competition but rather colonies were harvested from agar plates for plasmid preparation. The number of pooled colonies for the barcode library (n = 33,000) was approximately half of the expected number of theoretically possible barcodes.

Ultramer oligonucleotides synthesized with alternative bases at the variant sites were used to amplify two overlapping fragments of exon 3 and flanking sequence, which were subsequently used as template for an overlap extension PCR (Supp. Fig. S2; Appendix 13.1.4). A reduction in the amount of the plasmid template (carrying the Reference haplotype) to 5 ng and a *Dpn*I digestion cleaving at methylated sites ameliorated the background of non-mutagenized clones. Insertion of the exon 3 haplotype library into the barcode library by conventional cloning, i.e. ligation through cohesive ends turned out to be inefficient at three different vector:insert ratios tested. I leveraged the long overhangs from *Hind*III and *Afl*II to the end of the overlap extension amplicon of the exon 3 mutagenesis library (Materials and Methods; Appendix 13.1.4, Publication IV) to perform IFC. An optimized 1:12 molar ratio (vector:insert) which yielded 1.3×10^6 cfu/µg. Applying colony PCR and Sanger sequencing on a sample of

n = 60 colonies, I observed 95% of clones carrying a proper insert. For the resultant Input Library the number of pooled clones (n = 3,500) was one order of magnitude higher than the actual haplotype diversity (n = 256), and one order of magnitude less than the complexity of the barcode library (n = 33,000).

Massively parallel sequencing of the Input Library

"Mate-pair"-like short-read sequencing approach

Sequencing the Input Library aimed to read the barcode and the 'linked' exon 3 haplotype which are 2 kb apart in each single molecule. To accomplish this with short-read NGS technology, we first tried a "mate-pair"-like approach to join barcodes and exon 3 sequences in close vicinity. First, amplification of the library with 5' phosphorylated primers was performed followed by circularization with T4 ligase and ATP (1 mM final concentration). Second, inverse PCR bridging the site of ligation with tailed primers to introduce P5/P7 was performed to yield an amplicon size compatible with sequencing on the Illumina platform. A pilot experiment was performed with a mixture of two minigene constructs harbouring different known barcodes and exon 3 haplotypes. Since circularization of monomeric molecules is crucial for this strategy, ligation reactions were treated with Plasmid-Safe ATP-dependent DNase from Epicentre (0.5 units per reaction). This treatment ensures removal of dsDNA linear concatemers, but does not degrade closed circular dsDNA concatemers resulting from multimeric intermolecular ligation. Different concentrations of phosphorylated amplicons down to 0.5 ng/µl were tested in the ligation reaction. Diluted DNA concentrations resulted in reduced amounts of high-molecular concatemer ligation products, yet a dimeric circular dsDNA band remained visible. In an effort to purify monomeric circular molecules by electrophoretic separation and gel extraction, I first assessed the differential migration behaviour of circular and linear monomeric fragments on agarose gels. Digestion with *Bcl*I, a restriction enzyme with a single recognition site in the target fragment was used to differentiate between circular and linear topology; thus, digestion with *Bcl*I only linearizes circular molecules to its monomeric size but cuts linear molecules into two smaller fragments. Using this strategy, I could demonstrate that circular monomeric fragments migrate at an apparent size of 1 kb in the agarose gel and is converted into a fragment of the expected size of 2.2 kb for the linear molecule upon *Bcl*I digestion.

These gel-extracted circular monomers were then used as template for the inverse PCR. However, Sanger sequencing of cloned inverse PCR amplicons revealed a mixture of correctly and uncorrectly matched exon 3 haplotypes and cognate barcodes at a similar rate indicating that gel excision upon electrophoresis was insufficient for the purification of monomeric circular molecules. An alternative strategy for sequencing of the Input Library was therefore pursued.

Long-read SMRT sequencing

SMRT sequencing enables determination of the entire sequence of interest between exon 3 and the barcode in the Input Library. The Input Library was digested with *Hind*III and *Sal*I (Supp. Fig. S2; Appendix 13.1.4) and separated on a dye-free 1% agarose gel. Two wells were loaded with an aliquot of the digestion reaction, cut-off from the remainder of the gel and stained with Ethidium bromide to visualize and mark the location of the 2.2 kb fragment of interest. Re-aligning the stained and marked gel stripes with the unstained gel piece allowed excising the 2.2 kb from the unstained gel slice (~15 cm) avoiding UV-light illumination. The gel-extracted DNA was purified on silica-based columns and sent to the Max Planck Institute of Molecular Cell Biology and Genetics, Dresden for SMRTbell library preparation and sequencing on a PacBio RSII instrument.

Massively parallel sequencing of the Output Library

Paired-end (PE) sequencing using MiSeq (Illumina) was used for short-read sequencing. Quality control with DNA1000, DNA7500 or DNA High Sensitivity chips on an Agilent Bioanalyzer, as well as Qubit and KAPA quantification were performed either at the Molecular Genetics Laboratory or at the c.ATG Core Facility.

A variety of different protocols and modifications thereof were tested and optimized to establish a valid protocol for the succesful sequencing of the Output Library. Main challenges were i) the low sequence diversity of the Output Library, and ii) the length of the molecules at the upper border of the MiSeq instruments' specification.

Short-read Amplicon Sequencing with custom primers (Version 1.0)

A fragment library consisting of RT-PCR amplicons with attached Illumina grafting P5/P7 adaptors was sequenced with high performance liquid chromatography (HPLC)-purified custom amplicon-specific sequencing primers. As the Output Library has an inherent low diversity, PhiX174 bacteriophage DNA was spiked in at two

different concentrations – representing 70% and 40% of molecules per run, respectively. High proportion of PhiX174 DNA resulted in run abortion at the “Index Read” since the PhiX174 DNA is not indexed. Combining 20% of PhiX174 with 20% of an external WGS library with a relative large insert size, or 7% of PhiX174 with 16% of a WGS library and 17% of a metagenomic library yielded reads exclusively derived from external spiked-in libraries.

Short-read Amplicon Sequencing with offset random spacers and Illumina sequencing primers (Version 2.0)

The RT-PCR was carried out with primers having a gene-specific 3' portion and a 5' tail comprising the binding site for ”Read 1” and “Read 2” Illumina primers for sequencing followed by a set of 2, 4, 6 or 8 random “N” bases. Pairs of forward and reverse primers were designed and combined in the PCR so that each amplicon has the same length and a total of 10 random bases in order to compensate the low diversity of the library and improve cluster identification in the first 4 run cycles. The offset of 2-6 bases between the amplicons aimed to reduce the phasing and prephasing values. A modified run protocol of the “Chemistry.xml” file with 32 cycles instead of the default 26 amplification cycles (V2 kits) was used and adjusted by loading ~30% less than recommended concentration of 9 pM. Applying these modifications the sequencing run was undermined by overclustering. A reduction of ~54% in the loading concentration resulted in optimal cluster density, yet the low signal intensity hindered proper cluster detection. A second modification of the “Chemistry.xml” file was implemented by increasing the duration of the First Extension (4,000 ms) and Amplification (30,000 ms) steps resulting in overclustering.

Short-read Amplicon Sequencing with offset spacers with additional random bases, dual direction, P5-PE grafting primer and Illumina sequencing primers (Version 3.0)

In this approach, all PCR primer tails included four consecutive random bases directly downstream of the sequencing primer binding sites to ensure diversity for proper cluster detection. The size of the amplicons bearing or lacking exon 3 was 990 and 821 bp, respectively. All amplicons were increased by 16 bases of 8 random bases and 8 additional offset bases designed as to keep the diversity at a given site for sequencing (Supp. Table S2; Appendix 13.1.4, Publication IV). By this modification I pursued to increase the diversity throughout the insert length in order to improve the “Cluster Passing Filter” (Cluster PF) and the “Phasing/Prephasing” by amplifying the

library in both directions (see internal barcode; Supp. Fig. S3; Publication IV). Offset sequences design took into account avoidance of GC-rich sequences and 'patches' of very low diversity (i.e. homooligomeric sequences). The PE-P5 grafting primer which harbours the “GAUCT” sequence was used. The “Chemistry.file” with 32 cycles of amplification was further modified as follows: the duration of First Extension was increased to 15,000 ms while the Amplification step was reset to 15,000 ms. In this experiment, ~70% of the recommended 9 pM was loaded on a Nano flow cell (single lane; 2 tiles). A total of 325 cycles were extracted, called and scored. “Read 1”, “Index Read” and “Read 2” had 115, 8 and 202 cycles, respectively. While in “Read 1” 96% of bases had a quality score of 30 or higher (Q30; i.e. the estimated chance of base call error rate is 0.1%), the quality of “Read 2” had 10% less Q30 bases (Fig. 12).

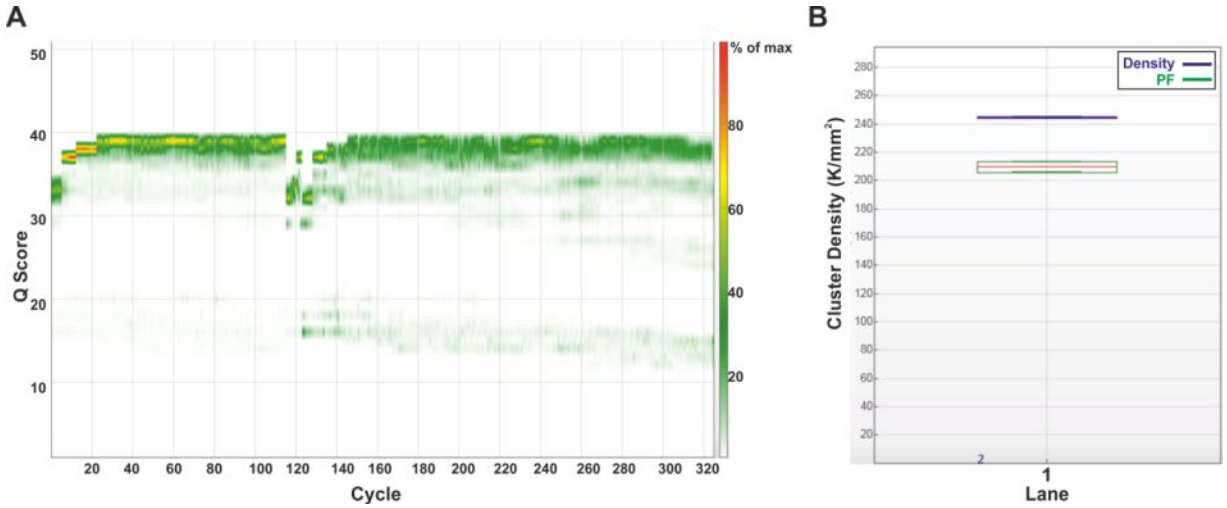


Figure 12. Snapshot of the “Sequencing Analysis Viewer” (Illumina) showing parameters of the successful sequencing run of the Output Library (*Version 3.0 protocol)*. A) The Q-score heat map shows Q scores per cycle. The colour bars to the right of the chart indicate the values in a scale of 0 to 100% of maximum value. A Q-score of Q30 indicates that a 0.1% chance of calling the wrong base (Q-20 corresponds to 1% and Q40 to 0.01%). A high percentage of reads had a score over Q30. The low intensity in “Read 2” (*not shown*) and subsequent lower Q-scores for “Read 2” is a known issue which might be exacerbated by the special conditions of this run. B) The Cluster density (K/mm^2) and the Cluster Passing Filter (Cluster PF) parameters are close to each other, indicating good clustering on the flow cell despite having a lower cluster density with a large size library (800-1000 bp) than a library of standard fragment size (200-600 bp).

The sequencing run yielded 664,730 paired-end reads passing filtering (=86% of total reads) and a total of 110 Megabases (Mb) of sequence (Fig. 6). Ninety-seven percent of filter passing reads were assigned to an I7 index from 701-TAAGGCGA, 702-CGTACTAG, 703-AGGCAGAA and 704-TCCTGAGC (Supp. Table S2; Appendix 13.1.4, Publication IV) and used for low-plex pooling of single-index samples to ensure signal in both colour channels during the “Index Read”. The coefficient of variation

across all indexes (CV=0.0818) indicates good representation of the pooled four replicates in an equimolar ratio.

Long-read SMRT Amplicon Sequencing

For SMRT amplicon sequencing the cDNA was amplified with KAPA HiFi HotStart DNA Polymerase (KAPA Library Amplification Kit) followed by an immediate purification optimized to a ratio of 0.78 vol. AMPure PB Beads to 1 vol. of RT-PCR products. As part of the SMRTbell chemistry from PacBio, the 46-nt long blunt adaptors (5' – pATCTCTCTCTTTTCCTCCTCCTCCGTTGTTGTTGTTGAGAGAGAT – 3') able to fold into a stem-loop structure were ligated to both ends of the amplicon. This results in structurally linear double-stranded molecules, yet have a circular topology (Travers et al., 2010). Electrophoretic mobility of these molecules resemble that of dsDNA, and the total gain in size (92 bp; 10% of the amplicon size) result in an almost undistinguishable profile and virtually the same expected size of SMRTbell templates and initial RT-PCR amplicon libraries. For the Input Library an increase in size of ~4% over the 2.2 kb insert was not detectable after SMRTbell preparation (Dr. Sylke Winkler, *personal communication*). Final recovery of the SMRTbell library was 68% (648 ng) of the input DNA used for the DNA repair reaction. Prior to sequencing, a complementary quality control was performed by digestion with *Kpn*I. The single *Kpn*I site ~100 bp from the 3' end of the RT-PCR amplicon enabled to detect a clear size difference of libraries before and after SMRTbell preparation and confirmed ligation of the adaptors. The SMRTbell library was loaded on diffusion mode at 2 pM on the Sequel at the MPI in Tübingen (Dr. Christa Lanz) according to the instructions from the binding calculator of SMRTLink v.5 based on the average size and concentration measured by Bioanalyzer and Qubit, respectively. An immobilization time of 120 min, a pre-extension time of 30 min and a movie time of 600 min were used. The output was 2.33 Gb. The productivity is measured by the fraction of wells on the SMRTcell that were loaded with a single molecule ("P1 productivity"), which was 7.3%. The "P0 productivity" refers to empty wells and was 88.5%, indicating underloading. The N50 polymerase read length was 56,250 bp and the average read length was 32,459 bp. The computation of circular consensus sequence (CCS) was performed with default parameters with SMRT Link Version 5.0.1.9585 at c.ATG (Jakob Matthes). CCS demultiplexing was performed by the "tailed" mode, filtering for Minimum Predicted Accuracy and Barcode quality were adjusted to 0.99 and ≥ 45, respectively.

6.5.3 Discussion

Generating a variant library for multiplexed functional assays with appropriate means to link the *output* readout to individual *input* variants requires the assessment of efficacy and efficiency at every single step. **High efficiencies** are needed for cloning, transformation, amplification and sequencing in order to keep the complexity, diversity and size of the library in balance. For instance, an increased length of homology arms (50-60 bp) for IFC during Input Library preparation improved cloning efficiency and yield of recombinants; in agreement with prior observations (Sleight et al., 2010). A critical point is the actual Input Library size, i.e. the number of colonies pooled for bulk plasmid DNA isolation which should be below the maximal **complexity** defined by the 4^8 = 65,536 theoretical possible barcodes. When the actual size of the library approaches its maximal theoretical size, the likelihood of picking clones with identical barcodes increases (Bystrykh et al., 2014).

Sequencing large molecules with an intact phasing of variants represents a challenge for short-read technologies. We aimed at developing a strategy to reduce the size of the molecules *via* circularization followed by inverse PCR which would bring the crucial sequence into close proximity in a smaller molecule. However, I realized a considerable proportion of intermolecular ligation products which result in illegitimate amplicons with interchanged barcodes indistinguishable from genuine PCR products. Gel purification of the desired circularized monomers did not result in satisfying results most likely due to co-purification of contaminant dimeric circular ligation products. Nowadays, this strategy might be improved with automatic size fractioning.

To circumvent this problem, we explored the concept of using a 20 bp sequence, upstream the internal barcode of our library, as a "Reflex" site for copying the barcode to the opposite site of the amplicon as described by Casbon and colleagues (Casbon et al., 2013). Although virtually inexistent levels of intermolecular rearrangement are claimed, this cannot be completely ruled out. Further alternative approaches such as T4 ligation in-emulsion to increase the rate of intramolecular ligation products or the more recent microfluidics-based encapsulation of long fragments into droplets for amplification, fragmentation and barcoding of single molecules of up to 10 kb, 30 kb (Redin et al., 2017) or even reaching 200 kb (Zheng et al., 2016) may diminish this problem, but were not implemented due to the need of further optimization or need for sophisticated equipment. I rather opted for long-read SMRT sequencing as other

reports published afterwards (D'Amore et al., 2017) to ensure **unequivocal phasing** of haplotypes and barcodes in the Input Library.

Despite the absence of intron 3 which reduces the amplicon size by 1,465 bp, sequencing of the Output Library was as challenging as for the Input Library. From the failure of the Library Preparation and Sequencing Protocol (LPSP) *Version 1.0* (LPSP V1), I learned that although 'spiked-in' external libraries were close (yet shorter) to the length of our library, these were able to out-compete the Output Library. Moreover, mixing with low or high proportion of external balanced libraries – only libraries of shorter average length were available – was insufficient to override the low diversity. We thus refrained from further using PhiX174 or any external libraries and custom sequencing primers. As shown in a previous report on amplicon sequencing (Wu et al., 2015), increasing the diversity by using primers with offset spacers and random bases in LPSP V2 improved the detection of clusters. Modification of the sequencing parameters such as altered loading concentration, bridge amplification efficiency, clustering and intensity are interrelated. Finally, LPSP V3 with increased diversity combined with the improved binding and linearization efficiency on the flow cell, together with the optimal loaded concentration and increased extension was found optimal and yielded useful sequencing data (Fig. 12).

The loading concentration for optimal cluster density was determined empirically, since lengthy clusters take up more space on the flow cell. The time for the First Extension was prolonged under the reasoning that synthesis of the second strand may not have been completed for ~1 kb fragments. Amplification time was not increased to avoid "overclustering". Since we increased First extension time and assumed that LPSP V3 bound more efficiently than LPSP V2 to the PE flow cell, we further reduced the loading concentration by 70% to ~2.7 pM. Accordingly, 245K cluster/mm^2 were detected, 30% of the ideal cluster density for a standard size library. These findings are in line with reports of **long amplicon** libraries on a MiSeq Nano flow cell (Burke and Darling, 2016).

6.6 Publication IV: A comprehensive reference for exon 3 splicing defects of the human *OPN1LW* and *OPN1MW* opsin genes by means of a parallelized minigene splicing assay

Having its origin in the minigene-based splicing assay of individual haplotypes (Section 6.3), a parallelized assay was developed (Section 6.5). Leveraging massive parallel sequencing technologies enabled to perform splicing assessment of several hundreds of exonic variants at once. The first pilot assay demonstrated the **feasibility** of the parallelized approach and provided a more comprehensive assessment of exon 3 *OPN1LW/MW* haplotyes with regard to splicing.

In silico tools such as HSF, EX-SKIP and HEXplorer are able to predict exon 3 haplotype-dependent deregulation, creation and/or loss of splicing enhancer and inhibitory elements. HEXplorer includes surrounding non-mutated context-sequence which represents an improvement, yet is still subjected to hexamer-based limitations (Erkelenz et al., 2014). In general, bioinformatic tools provide low predictability for SNVs other than those affecting the canonical splice sites (Soukarieh et al., 2016; Cummings et al., 2017; Kremer et al., 2017). Not surprisingly, overall predicted scores for whole haplotypes do not correlate well with outcomes achieved by functional minigene assays for individual haplotypes (Publication II), especially for those yielding mixtures of both correctly and aberrantly spliced transcripts. Lack of an appropriate **toolkit** for reliable **prediction** of the functional effect of known and newly observed *OPN1LW/MW* haplotypes hinders interpretation. An approach to determine and quantifiy the splicing outcome induced by every possible variation in a confined region with the highest diversity within exon 3 of *OPN1LW/MW* was pursued.

The work presented in this publication consisted in generating a batch of minigene constructs that would comprenhend **all possible combinations** of the defined common eight variants in exon 3 of *OPN1LW/MW*. This means that I targeted a total of 2^8=256 haplotypes (2 possible alleles per variant; 8 variants). For visualization of the variants along exon 3 see Supp.Fig.S1; Appendix 13.1.4 and Fig. 8 in Appendix 13.1.5, Publication V). The generation of such a collection of minigenes occurred in the form of a pool from the starting point and are only *separated* when transformed in bacteria and individually propagated in single clones. Ultimately these clones are pooled again prior plasmid isolation. The pool of minigene constructs is named after

"Input Library" because represents the input for the splicing assay. The minigenes in the pool are traceable, meaning that the haplotype sequence in exon 3 could be trailed once the *in vitro* splicing reaction in mammalian cells takes place, even if a particular combination of variants in exon 3 leads to skipping of the very same exon. Instead of scoring each variant with the relative abundance of reads spanning the tested exon between input and output libraries normalized to the wild-type allele (Ke et al., 2011; Julien et al., 2016; Soemedi et al., 2017; Ke et al., 2018), we applied a **barcoding** system integrated in the minigene backbone. For each molecule generated, this barcode enables tracing of the exon 3 haplotype from the original pool of minigenes in the Input Library all the way through up to its detection in the Output Library – the pool of transcripts derived from spliced minigenes – even if the exon 3 encoding the haplotype sequence itself is alternatively spliced because the barcode remains in the transcript molecule (Fig. 1, Supp. Fig. S2-3; Appendix 13.1.4, Publication IV). The recently published Vex-Seq method (Adamson et al., 2018) also exploited barcodes; yet these were synthesized within the oligonucleotide pool along the exon tested. In contrast, the barcodes of our assay are located downstream of the ORF of *OPN1LW/MW* to prevent interfering with splicing regulatory signals.

The barcode allowed to tell apart the original exon 3 sequence and provided the link with the splicing outcome (exon skipping and/or exon inclusion) associated to that particular exon 3 haplotype. This relies on the **uniqueness** of the barcodes; meaning that if the same barcode was found phased with two or more distinct haplotypes, this barcode was discarded. We found almost 2,000 unique barcodes in the Input Library. The other way around, if the same haplotype was found phased with one or more distinct barcodes, these barcodes were retained. Thus, over 1,500 unique barcodes were overall found tagging 248 targeted haplotypes. Due to the stochastic nature of the generation of the barcoded constructs we relied on the complexity of the barcode library in order to have enough unique barcodes per haplotype. This makes **redundancy** an intrinsic feature of our library. Multiple unique barcodes (median of 6) have been detected per haplotype in the Input Library stemming from originally different clones that carried the same haplotype in exon 3. Moreover, for each multi-barcoded targeted haplotype we relied on several observations; an average of 29 observations – circular consensus sequencing (CCS) reads – each of them derived from single molecules coming from the same or different minigene constructs with the same haplotype. We reason that this could control for a *single-construct* bias.

The outcome of the splicing assay for a total of **232** different *OPN1LW/MW* exon 3 haplotypes was determined as the fraction of correctly spliced transcripts. This is calculated based on the number of reads supporting the presence of exon 3 over the ones indicating the absence of exon 3 (Supp. Table S3; Fig. 3B; Appendix 13.1.4). High **reproducibility** was achieved by our assay with correlation efficiencies above 0.89 and reaching 0.97 among pairwise comparisons of the ratios obtained by the four technical replicates from the Output Library (Appendix 13.1.4, Publication IV), comparable or superior to previous published MPRAs (Adamson et al., 2018; Julien et al., 2016, Soemedi et al., 2017).

The **quantification** based on capillary electrophoresis carried out for the individually assessed haplotypes observed in XLCD patients (Buena-Atienza et al., 2016; Publication II) was already novel since other reports (Ueyama et al., 2012; Gardner et al., 2014) have not performed experimental quantitation of the products derived from differentially spliced transcripts. Concerning the reliability of the results from the parallelized assay, we compared our results to a set of individually assessed haplotypes in barcoded-minigenes extracted from the Input Library for quality control in the generation process (Publication IV) and considered this as the *truth* dataset. Our parallelized assay aims to categorize haplotypes by quantifying (usable scores) the *grey zone* of outcomes between complete exon 3 inclusion and full exon 3 skipping, which represent clearly non-pathogenic and pathogenic scenarios, respectively. The **categorization** (“minor defective”, “intermediate” or “strongly deleterious”) in terms of *splicing competence* determined by individual minigene assays yielded reasonable predictability with a sensitivity of 66.7% and a specificity of 71.4%. Although the outcome of the comparison between the individual and parallelized assays was positive; seemingly, the parallelized approach indicates higher exon skipping levels. Further improvements in sequencing of the Output Library may bring the discrepancies and the overall shift closer.

We observed a decent number of haplotypes yielding low ratios of exon inclusion; namely, more than the half of the haplotypes resulted in below 50% of transcripts being correctly spliced. About 1/4 of haplotypes rendered a mean of exon 3 rentention levels of less than 20%. A significant impairment is related to the **c.532G** or the **c.538G** alleles which were found to cause an overall reduction of **~38%** and **~26%** in the **exon 3 inclusion** levels, respectively (Fig. 3A; Appendix 13.1.4). Grouping haplotypes according to the genotype at the latter two variants resulted in highly

significant differences in the mean ratio for exon inclusion among all four groups except between the two groups harbouring either only the c.532G or only the c.538G allele. Remarkably, haplotypes carrying both c.532G and c.538G had the overall lowest ratio for exon inclusion (Fig. 3B; Appendix 13.1.4). By adding the c.511 variant to split these four groups, further grading of the cluster with the lowest scores (c.532G; c. 538G) was possible (Fig. 4; Appendix 13.1.4). Namely, haplotypes carrying c.511A; c.532G; c.538G showed a significantly lower ratio of exon inclusion in comparison to those with the background of c.511G; c.532G; c.538G. Although overall the presence of the c.521C>T variant leads to an increase in exon inclusion – as also suggested by individual minigene assays (Appendix 13.1.2, Publication II) – when individual haplotypes differing by either c.521T or c.521C allele were compared, the *positive modifier* effect of the c.521C>T variant was not statistically significant when the data from our MPRA assay was analyzed (Supp. Table S6; Appendix 13.1.4).

For every targeted site in exon 3, it was shown that an **accumulation** of such variants correlated with a decrease in exon inclusion (Fig. 5; Appendix 13.1.4). In an attempt to elucidate the mechanisms underlying the **interactions** within variants, I analysed whether these responded to an additive model, or otherwise to positive or negative epistasis. In the latter case of epistasis this would mean that the effect of the haplotype of two variants was greater or lower than the sum of the individual effect of each variant (Fig. 6; Supp. Table S7-8; Appendix 13.1.4). An important additive interaction is that of the c.532G and the c.538G alleles, which accounts for the highest impact in fitness among the double variants, as expected from the sum of the individual effect of each variant.

The underlying motivation of the parallelized assay was to provide predictive scores to aid interpretation in the disease context. We collected a sample of 1,175 fully resolved *OPN1LW/MW* exon 3 haplotypes of which 812 corresponded to the control group and 243 were found in the disease group (BCM and XLCD). For 85% of the haplotypes observed we could relate the sample fraction to the levels of exon 3 retention determined by the parallelized assay, finding that those with the lowest levels of exon retention were enriched in the disease group (Fig. 8; Appendix 13.1.4). These findings reinforce the potential of the parallelized assay and its utility as an interpretation resource.

Previous reports have individually assessed the splicing competence of a number of exon 3 *OPN1LW/MW* haplotypes from XLCD subjects and in some instances including haplotypes obtained by site-directed mutagenesis in a *one-by-one* fashion. The total number of haplotypes that each study assessed would represent 7% (Ueyama et al., 2012), 3% (Gardner et al., 2014) and 5% (Buena-Atienza et al., 2016; Publication II) of those herein presented. With parallelization we have succedded to increase the **capacity** of the splicing assay by almost 20-fold in the number of *OPN1LW/MW* exon 3 haplotypes tested and reduced the number of required transfections by 58-fold. Not only the sequencing-based quantification surpasses the fragment analysis used in our prior study (Publication II), but the whole process of plasmid generation, transfection and cDNA synthesis is superior in terms of **throughput**.

Yet, the total number of variants tested by our MPRAs has been only moderately boosted in comparison to other published MPRAs due to several reasons. Whereas the median **length** of a human exon is 134 bp (Savisaar and Hurst, 2017) minigene-based MPRAs hitherto have assayed exons of shorter length; 51 bp (Ke et al., 2018), 54 bp (Mueller et al., 2015), 63 bp (Julien et al., 2016), 68-97 bp (Adamson et al., 2018) or 100 bp at most (Cheung et al., 2019; Soemedi et al., 2017). In contrast, the 169-bp exon 3 of *OPN1LW/MW* herin tested is considerably larger. Likewise, the median length of a human intron is 1,567 bp (Savisaar and Hurst, 2017). Limited through the use of commercially synthesized oligonucleotides with a length of ~170 bp, previous MPRA approaches included only 15-55 bp of 3' and 5' endogenous flanking introns (Soemedi et al., 2017; Adamson et al., 2018; Cheung et al., 2019). Moreover, these MPRAs take advantage of minigene backbones of **heterologous** genes, for instance the globin gene (Adamson et al., 2018), artifically split the gene of the green fluorescent protein GFP with intron backbones of either *SMN1* or *DHFR* (Cheung et al., 2019), the Chinese hamster *dhfr* gene (Ke et al., 2018), or hybrid constructs including exon and intron sequences of adenovirus (pHMS81) and *ACTN1* (Soemedi et al., 2017). Other than the examples above, the *OPN1LW/MW* minigene library pool (Fig. 1; Appendix 13.1.4) relies on **native** donor and acceptor sites of adjacent exons to exon 3. By our mutagenesis approach we preserved the context of the entire cloned sequence: the complete cDNA sequence including five adjacent exons to exon 3 and the entire and unmodified intron 2 and intron 3 sequences of *OPN1LW*.

Hitherto minigene-based MPRAs have been adapted to existing short-length sequencing with few limitations to perform large-scale assays at the expense of loosing regulatory elements of biological relevance. We rather gave more importance to preserving the native sequence context and thereby paved way for a minigene-based MPRA able to accommodate testing of variants within large exons, multiple exons, and even intronic sequences. We incorporated SMRT sequencing of the Input Library and subsequently used a combination of SMRT and short-read sequencing for the Output Library (Supp. Fig. S3; Appendix 13.1.4). In our strategy, **long reads** enabled the generation of a reference sequence scaffold. We pursued mapping of **short reads** in a haplotype-wise and isoform-specific manner. Quantification based on the short reads was guided by the pre-assembled long reads. Hybrid strategies have previously shown to take advantage of the improved sensitivity and specificity of read mapping, phasing of variants and the detection and discriminaton of full-length transcript isoforms of long reads (Au et al., 2013; Tilgner et al., 2014; Ning et al., 2017) on one hand, and on the other hand, the high throughput and yield of short-read sequencing to complement for means of allele counting and basecalling accuracy. The latter strategy bears limitations which could be overcome by the imminent increasing throughput of long-read sequencing. For most reported MPRAs (Ke et al., 2011; Julien et al., 2016; Soemedi et al., 2017; Ke et al., 2018), short-read sequencing was applied for input and output libraries, enabling the calculation of enriching scores. Taking advantage of barcoding, we required only quantification of the Output Library and thus, avoided quantification across different types of libraries from different time points relative to the splicing reaction.

MPRAs screening for splicing-disrupting variants can be divided into two types of approaches; those focused on a certain **gene or exon of interest** (Julien et al., 2016; Mueller et al., 2015) as ours for *OPN1LW/MW* exon 3 and MRPAs accommodating **reported variants** either in databases (Adamson et al., 2018; Cheung et al., 2019) or those incorporating known disease-causing variants (Soemedi et al., 2017). Approximately 2,000 and 28,000 ExAC variants in the context of about 110 and 2,400 alternative exons were included in Vex-Seq and MFASS aproaches, respectively (Adamson et al., 2018; Cheung et al., 2019). In contrast, the *gene-of-interest* category of investigations includes base substitutions not listed in databases, aiming i) to obtain a comprehensive *a priori* characterization of each individual base and any substitution thereof, and ii) to obtain a map of crucial sites or substitutions which deepens insight

into the *language* of splicing for a certain gene or exon. Either known variable regions are specifically targeted by **site-directed** mutagenesis (Mueller et al., 2015 and our own study; Publication IV), or all possible single substitutions are generated by **saturated** mutagenesis (Julien et al., 2016; Ke et al., 2018). We included all possible combinations of haplotypes in a confined region with high variability and analysed interactions for double and triple variants (Appendix 13.1.4). At a smaller scale, similar analyses have been reported including dinucleotide substitutions and hexamers (Ke et al., 2018) or double-mutants (Julien et al., 2016). MPRAs concentrated on variants observed in individuals including non-pathogenic, putative and known pathogenic variants reported ~4% of variants that strongly impaired exon recognition (Cheung et al., 2019). Surprisingly, most (>80%) of such variants were located outside canonical splice sites (Cheung et al., 2019). Soemedi and colleagues reported that ~10% of the rare disease-associated missense and nonsense variants, and 3% of the common variants tested by MAPSy result in alteration of splicing (Soemedi et al., 2017).

In some contrast to these figures from survey studies interrogating known variants for a large number of exons/genes, scanning or saturation mutagenesis studies focusing on individual complete exons reported high susceptibility to sequence variation interfering with splicing competence. By testing a series of 5,560 mutant constructs of exon 5 of the human Wilms' tumor gene 1, Ke and colleagues found 70% of variants resulting in 2-fold absolute change of splicing enrichment scores, both increasing and reducing exon inclusion (Ke et al., 2018). Likewise, Julien and colleagues reported a high proportion (~60%) of splicing-altering variants in exon 6 of *FAS*. Moreover, 23% of all single and multiple variants in exon 7 of *SMN1* resulted in aberrant splicing (Mueller et al., 2015). Similarly, we showed that ~25% of the targeted *OPN1LW/MW* exon 3 haplotypes cause high levels of exon skipping. The apparent disproportion might be due to the selection of genes or exons known to harbour splicing mutations or an especially susceptible exon definition.

In conclusion, this study aims to assist the interpretation of each individual haplotype assessed within the parallelized assay and to facilitate the joint study of several variants and its sequence context. We have scattered the fitness of exon 3 towards splicing; nevertheless, only by assessing the variants that are commonly observed within the population and in rare haplotypes. Exon 3 seems to present robustness towards splicing aberrations. Moreover, we found that those haplotypes with the lowest correctly spliced levels of exon 3 are enriched in a disease population.

6.7 Unpublished III: Putative splicing factors binding *cis*-regulatory elements within exon 3 of *OPN1LW/MW* haplotypes

6.7.1 Introduction

Splicing factors are generally divided into SR and hnRNP proteins (Section 3.4.2) attending to the **enhancing or inhibitory** effect exerted upon binding to ESE/ISE or ESS/ISS, respectively. Certain splicing factors might not attend to the latter discrete classification, being able to either induce or inhibit the use splice sites according to their relative location (Zhou et al., 2014), surrounding sequence (Rimoldi et al., 2013) or co-interaction with other RBPs (Caceres et al., 1994; Huelga et al., 2012).

Bioinformatic tools are based on experimental approaches to map the **recognition** sites for binding of pre-mRNA splicing factors; for instance, by systematic evolution of ligands by exponential enrichment (SELEX) which consists on iterative enrichment cycles of ligands with a high **affinity** from a pool (Tuerk and Gold, 1990) or *in vivo* cross-linking immunoprecipitation (Ule et al., 2003), and threrefore build consensus motif sequences for each splicing factor. **Accessibility** of RRMs to *cis*-regulatory elements of splicing is constrained by RNA secondary structures (Hiller et al., 2007). A deep interaction assay including mutagenesis to study RNA structure (Taliaferro et al., 2016) revealed the importance of the context and showed mutation-dependent RNA secondary structure effects. The steady state equilibrium between double or single strandness at the splice factor binding site(s) might affect splicing.

We sought to further dissect the regulatory landscape of splicing by querying the cognate *trans*-acting factors that bind, recognize and coordinate *cis*-regulatory sequences within *OPN1LW/MW* exon 3. The identification of splicing factors involved in the regulation of *OPN1LW/MW* exon 3 splicing by experimental assays has never been reported. By means of an *in vitro* **RNA-pulldown** assay, we pursued to identify splicing factors that differentially bind to exon 3 of *OPN1LW/MW* in a haplotype-dependent manner. We identified hnRNPF as putative RNA-binding candidate protein with affinity for RNA probes encoding the haplotype ‘LIAVA’. We further pursued functional validation through silencing of the candidate splicing factor during transient expression of haplotype-harbouring minigenes that allow for testing retention or skipping of exon 3, leveraging the established splicing assay (Sections 6.4 and 6.5).

6.7.2 Results

A preliminary analysis with the *in silico* algorithms HSF (Desmet et al., 2009) and SpliceAid 2 (Piva et al., 2012) was performed to predict *cis*- and *trans*-elements putatively involved in splicing regulation of *OPN1LW/MW* exon 3 (Table 4).

Table 4 : Splicing factors predicted for *OPN1LW/MW* exon 3 variants by *in silico* tools

Splice Factor	HGNC Symbol*	Family	Motif†	Variant‡	*Cis*-element§	'LIAVA' effect¶	Algorithm#
9G8	SRSF7	SR-rich	AG**A**TGG	c.453G	ESE	Broken	HSF
			G**A**TGGT	c.457C	ESE	Broken	HSF
			GTGGGC	c.511A	ESE	Broken	HSF
			GAG**G**TGGA	c.453G	ESE	New	HSF
			GTGGAT	c.453G	ESE	New	HSF
SF2/ASF	SRSF1	SR-rich	GAGAG**G**T	c.453G	ESE	New	HSF
SRp40	SRSF5	SR-rich	AGAG**G**	c.453G	ESE	New	SpliceAid 2
			GCTGCT	c.538G	ESE	New	SpliceAid 2
SRp30c	SFRS9	SR-rich	TGG**A**T	c.457C	ESE	Broken	SpliceAid 2
			TGG**A**T	c.532G	ESE	Broken	SpliceAid 2
SC35	SRSF2	SR-rich	GGT**C**TGCA	c.465G	ESE	Broken	HSF
			GT**C**TGCAA	c.465G	ESE	Broken	HSF
YB-1	YBX1		GGT**C**TGC	c.465G	ESE	Broken	SpliceAid 2
			G**A**TCTG	c.532G	ESE	Broken	SpliceAid 2
			GG**T**CTGC	c.538G	ESE	Broken	SpliceAid 2
hnRNP P	FUS	hnRNP	TGGT**G**T	c.465G	ESS	New	SpliceAid 2
ETR-3	CELF2	CELF/BRUNOL	CATC**G**	c.511A	ESE	Broken	SpliceAid 2
SRp20	SRSF3	SR-rich	G**A**TC	c.532G	ESE	Broken	SpliceAid 2
			CATC**A**T	c.511A	ESE	New	SpliceAid 2
hnRNP A1	HNRNPA1	hnRNP	G**A**TGGA	c.453G	ESS	Broken	HSF
			G**A**TGGT	c.457C	ESS	Broken	HSF
			GAG**G**TG	c.453G	ESS	New	HSF
hnRNP H1	HNRNPH1	hnRNP	T**G**GGC	c.513T	ESS	Broken	SpliceAid 2
			TGG**G**T	c.532G	ESS	New	SpliceAid 2
			TGG**G**C	c.538G	ESS	New	SpliceAid 2
hnRNP H2	HNRNPH2	hnRNP	T**G**GGC	c.513T	ESS	Broken	SpliceAid 2
			TGG**G**T	c.532G	ESS	New	SpliceAid 2
			TGG**G**C	c.538G	ESS	New	SpliceAid 2
hnRNP H3	HNRNPH3	hnRNP	T**G**GGC	c.513T	ESS	Broken	SpliceAid 2
			TGG**G**T	c.532G	ESS	New	SpliceAid 2
			TGG**G**C	c.538G	ESS	New	SpliceAid 2
hnRNP F	HNRNPF	hnRNP	T**G**GGC	c.513T	ESS	Broken	SpliceAid 2
			TGG**G**C	c.538G	ESS	New	SpliceAid 2
KSRP	KHSRP	KH-type	TGG**G**T	c.532G	ESS	New	SpliceAid 2
SRp55	SRSF6	SR-rich	TGG**G**TC	c.532G	ESE	New	HSF

*HGNC: Human Genome Organisation (HUGO) Gene Nomenclature Committee; †Recognized motif sequence (variant of 'LIAVA' haplotype that creates or breaks the motif highlighted in bold); ‡Variant entailed in the *OPN1LW/MW* exon 3 haplotype 'LIAVA'; §*Cis*-acting regulatory elements: Exonic Splicing Enhancer (ESE) or Exonic Splicing Silencer (ESS); ¶Effect on the cis-regulatory element predicted to be recognized by a given splicing factor when 'LIAVA' haplotype is harboured in exon 3 instead of the Reference haplotype; #HSF:Human Splicing Finder (Desmet et al., 2009) or SpliceAid 2 (Piva et al., 2012) algorithms were used for bioinformatic predictions.

Multiple putative motifs for RBPs were retrieved for each of the variants that differ between the haplotype 'LIAVA' and the Reference exon 3 haplotypes (Table 4). The recruitment of certain splicing factors by ESE motifs such as 9G8 and SRp20 is **predicted** to be changed by different variants of the exon 3 haplotype so that new ESE sites are created and ESE sites are broken. For instance, the c.453G variant (in 'LIAVA') results in the creation of two new ESE sites for the same splicing factor 9G8 and abolishment of one ESE. Interestingly, one single variant may create and abolish both ESEs and ESSs for the very same splicing factor simultaneously, such as the c.453G for 9G8 and hnRNP A1, respectively. Many variants were predicted to individually exert mixed (inhibiting and enhancing) effects such as: c.453G, c.457C, c.511A and c.538G. In contrast, the c.532G was only predicted to be involved in either avoiding ESE sites or creating new ESS sites. Although the c.538G variant exhibited a similar profile, new ESE and ESS motifs were found **overlapped**. The variants c.465G and c.513T were associated with either only inhibiting or enhancing roles, respectively. These predictions are based on tetra- to octa-mers spanning one single variant site of the haplotype yet are not capable to include the sequence context – additional variants within the haplotype – into consideration.

In silico predictions to determine the RNA secondary structure of exon 3 sequences were made with RNA2DMut (Moss, 2017). RNA2DMut uses RNAfold and relies on the Boltzmann structure ensemble to calculate the ensemble diversity (ED) enabling to query every substitution spanning all positions in the input sequence (Moss, 2017). The ED metric is the average base pair distance between the set of structures. For RNA carrying the Reference haplotype the ED was 49.92 whereas for 'LIAVA' was 13.32; ~3.7x lower, suggesting the latter has a **tighter ensemble** and the Reference has a less defined structure or may show multiple conformations (Fig. 13). The predicted Minimum Free Energy (MFE) change in the Gibb's folding energy (ΔG) for the Reference and 'LIAVA' input sequence was -58 and -70.9 in kcal/mol, respectively. The average ED of all 507 possible single mutations with respect to the Reference exon 3 sequence was 52.45, very similar to the average ED of the all variants from 'LIAVA' (52.3). However, high ED values were retrieved for two variants of 'LIAVA'; c.532A>G and c.538T>G with 74.55 and 74.17, respectively. These variants retrieved within the nine (c.532A>G) and the eleven (c.538T>G) **highest values** of all 507 tested mutations (Appendix 13.2.4; Supp. Table S2). The associated predicted

centroid structure of the latter variants deviated greatly from the MFE structure, suggesting that these variants pertub the native structure (Fig. 13).

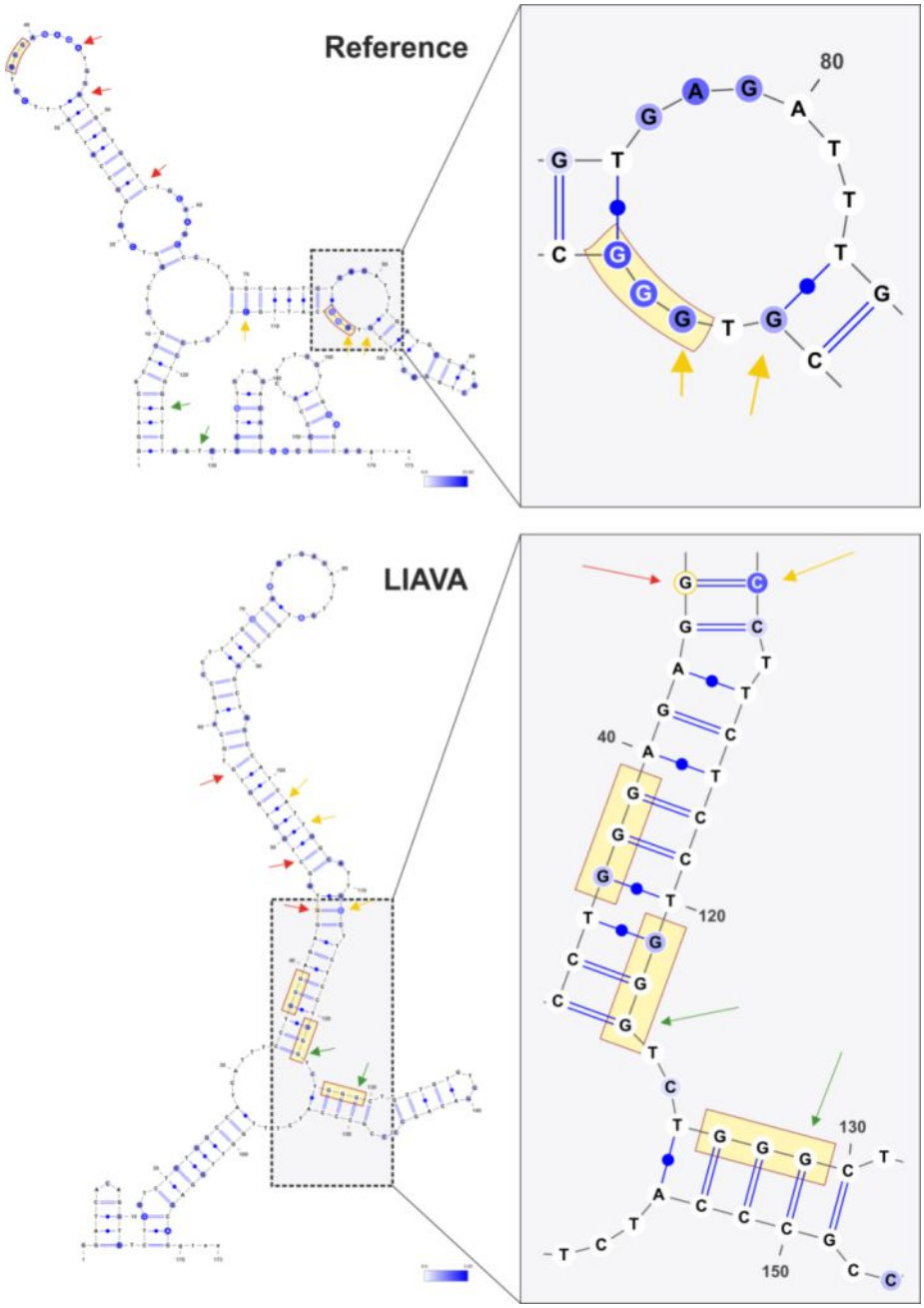


Figure 13. The 2D model of ensemble centroid structure for pre-mRNA molecules bearing Reference and 'LIAVA' exon 3 haplotypes. The 2D models generated by RNA2DMut (Moss, 2017) for the sequence (showed as DNA nucleotides) comprising the complete exon 3 of *OPN1LW/MW* for the Reference (top) and the 'LIAVA' (bottom) haplotypes**.** The variants at positions c.453 (44), c.457 (48), c.465 (56) are indicated with red arrows, c.511 (102), c.513 (104) and c.521 (112) with yellow arrows and c.532 (123) and (129) as green arrows. The magnified area for the Reference shows an open region including c.511G and c.513G whereas for the 'LIAVA' haplotype c.511A and c.513A seems to pair with the sequence upstream in exon 3 in between of polymorphisms c.453G and c.457C. The magnified region for 'LIAVA' shows that a tight region collapsing 3 different G-runs and forming a secondary structure stabilized by 11 G=C bonds, including c.532G, c.538G, c.465G and c.521C variants of the 'LIAVA' haplotype.

In our experimental approach we pursued to pinpoint RBPs which would **differentially** recognize and bind ESEs or ESSs present on the haplotype-encoded RNA baits through specific RRM motifs. Three **RNA baits** (Appendix, 13.2.4; Supp. Table S1) were designed: 1) a 109-mer recapitulating the deleterious 'LIAVA' haplotype, 2) a 109-mer with the reference haplotype, and 3) a 109-mer *scrambled* control of comparable length, devoid of high score binding sites and balanced for nucleotide composition. HEK293T cell lysates were incubated with RNA baits under competing conditions and **RNA-bound proteins** were purified applying biotin-streptavidin affinity purification. The enriched protein fraction was then subjected to liquid chromatography-coupled mass spectrometry for identification (see details in Appendix, Appendix 13.2.4 and Supp. Fig. S2). Intensities derived from the Reference and 'LIAVA' baits were normalized to the Scrambled RNA bait (Fig. 14).

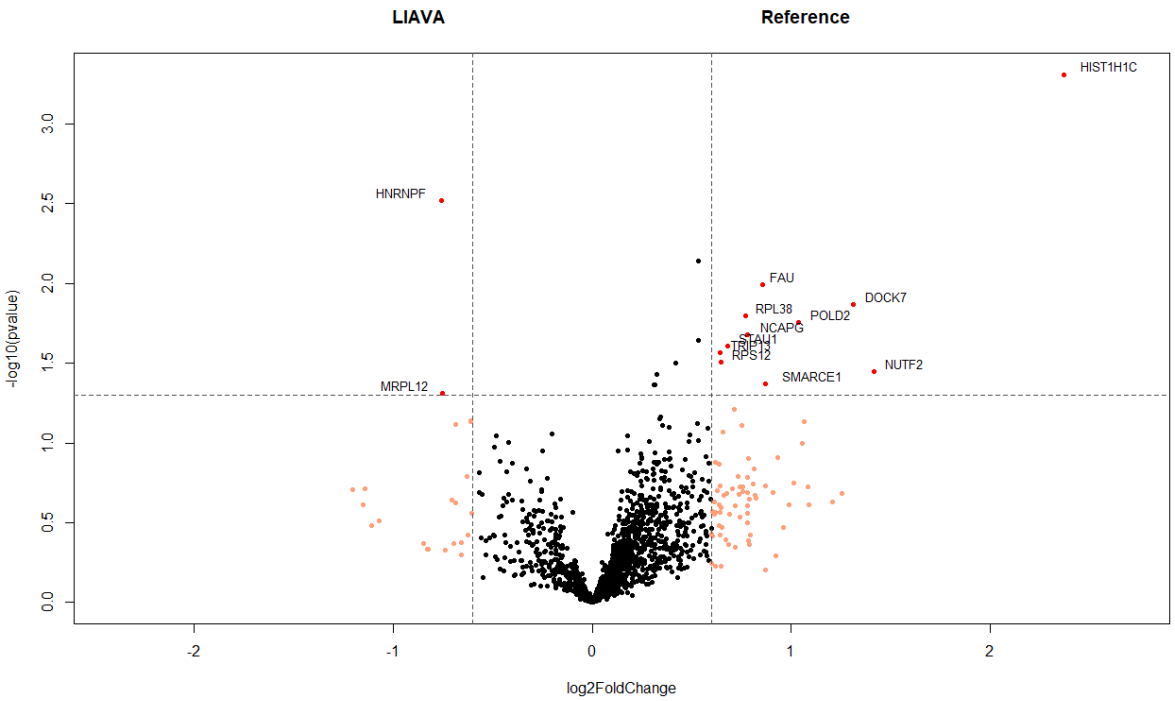


Figure 14. Volcano plot of protein enrichment in haplotype RNA-bound fractions. Statistical significance (y-axis) is represented as the minus logarithm (base 10) of the p-value and enrichment (x-axis) is depicted by the logarithm (base 2) of the fold change. Fold change is calculated as 2 to the power of the absolute mean difference that results from substraction of the normalized mean ratio of 'LIAVA' to the normalized mean ratio of the Reference. Hits with negative values left to the vertical dotted line and hits with positive values right to the vertical dotted line on the x-axis correspond to enriched proteins in the fraction of RNA-bound with 'LIAVA'- and Reference-RNA baits, respectively. Significantly enriched proteins are shown above the dotted horizontal line (Student's T-test; Interval Confidence (IC) 95%; n = 6). Points below the horizontal dotted line and right and left to the vertical dotted lines depict enriched proteins that are not statistically significant, and points which are in between the two vertical dotted lines are unspecific proteins or not enriched for a specific haplotype.

Upon six replicate experiments we observed two proteins enriched in the Reference-bait bound fraction, and two different proteins enriched in the 'LIAVA'-bait bound fraction (Table 5).

Table 5. Haplotype-dependent enriched proteins in the RNA-pulldown assay

Gene Name	Protein ID	Protein Name	Mean Ref*	Mean 'LIAVA'†	Mean Diff.‡	S.D.§	FC¶	Ratio Sig.#	P-value**
HIST1H1C	P16403	Histone H1.2	3.23	0.86	2.37	0.87	5.17	1.40E-19	0.000489
NOP10	Q9NPE3	H/ACA ribonucleoprotein complex subunit 3	0.88	-0.21	1.08	0.80	2.12	0.000103	0.188635
HNRNPF	P52597	Heterogeneous nuclear ribonucleoprotein F	0.64	1.40	-0.76	0.41	1.69	0.000436	0.003014
MRPL12	P52815	39S ribosomal protein L12, mitochondrial	0.38	1.14	-0.75	0.93	1.69	0.000463	0.048780

*Ratio of Mean Intensity: Reference to Scrambled; †Ratio of the Mean Intensity: 'LIAVA' to Scrambled; ‡Mean Difference Reference -'LIAVA' is the value that results of the substraction of the normalised mean of 'LIAVA' from the normalised mean of the Reference; §Standard Deviation of the Mean Differences; ¶Fold Change (2 to the power of the absolute Mean Difference) normalized to the Scramble bait; #Significance A of the Ratio Reference/'LIAVA' indicates significant outlier values relative to a certain population with a two-sided test and Benjamini-Hochberg FDR correction with a threshold value of 0.05; **P-value of a Student's T-test of the Mean Difference Reference-'LIAVA' (IC 95%; n = 6, being n the number of independent RNA-pulldown experiments).

Among those, three showed a statistically significant difference in enrichment between the normalized mean intensities of the Reference and 'LIAVA' bound fractions (Student's T-test; n = 6). Apart from Histone H1 enriched in the Reference and the mitochondrial 39S ribosomal protein L12 enriched in 'LIAVA', we found a ~1.7-fold enrichment of the known **splicing factor hnRNPF** in the 'LIAVA'-bait bound fraction (p-value = 0.003; Student's T-test; n = 6; alpha = 0.05). Given the known function of hnRNPF, we considered this protein as a prime candidate putatively involved in the haplotype-dependent regulation of exon 3 splicing.

The hnRNPF candidate remained significantly enriched in the 'LIAVA'-bait bound fraction when the analysis was repeated only taking unique peptides into consideration (p-value = 0.007; Student's T-test; n = 6; alpha = 0.05). Thus, the evidence from this mass spectrometry experiment for enrichment of hnRNPF in the 'LIAVA'-bait bound fraction was sound enough to perform further characterization. Both hnRNPF and the closely related hnRNPH were confirmed by Western blotting in HEK293T cell lysates as well as in eluted fractions of enriched proteins upon RNA pulldown. Although hnRNPH1 was significantly enriched in the 'LIAVA' bait bound fraction when only unique peptides were considered (Table 5); the difference of the

means between Reference and 'LIAVA' was not statistically significant for hnRNPH1 (p-value = 0.081; Student's T-test, Interval Confidence 95%; n = 6; Table 5).

Table 5. Enriched proteins in the RNA-pulldown assay with only unique peptides

Gene Name	Protein ID	Protein Name	Mean Diff*	Ratio Significance[¶]	P-value[#]
CNBP	P62633	Cellular nucleic acid-binding protein	0.03	0.82487	0.972
TIMM13	Q9Y5L4	Mitochondrial import inner membrane translocase subunit Tim13	0.28	0.25571	0.659
PTRH2	Q9Y3E5	Peptidyl-tRNA hydrolase 2, mitochondrial	-0.38	0.04335	0.387
CHTOP	Q9Y3Y2	Chromatin target of PRMT1 protein	0.62	0.00205	0.092
NHP2L1	P55769	NHP2-like protein 1	-0.47	0.01529	0.482
DHX36	Q9H2U1	ATP-dependent RNA helicase DHX36	-0.33	0.06997	0.508
OLA1	Q9NTK5	Obg-like ATPase 1	-0.32	0.07746	0.297
NIPSNAP1	Q9BPW8	Protein NipSnap homolog 1	0.38	0.09338	0.377
NPM3	O75607	Nucleoplasmin-3	0.09	0.96803	0.803
MRPL14	Q6P1L8	39S ribosomal protein L14, mitochondrial	-0.40	0.03547	0.435
SRSF2	Q01130	Serine/arginine-rich splicing factor 2	0.15	0.68404	0.643
SRPRB	Q9Y5M8	Signal recognition particle receptor subunit beta	0.28	0.26003	0.375
SRP14	P37108	Signal recognition particle 14 kDa protein	0.37	0.09918	0.073
ACAD11	Q709F0	Acyl-CoA dehydrogenase family member 11	-1.43	2.59E-11 (+)	0.052
MRPL12	P52815	39S ribosomal protein L12, mitochondrial	-0.90	1.43E-05 (+)	0.082
HNRNPH1	P31943	Heterogeneous nuclear ribonucleoprotein H	-0.68	0.00079 (+)	0.081
HIST1H2BL[‡]	Q99880[§]	Histone H2B[†]	-0.72	0.00042 (+)	0.564
HIST1H1C	P16403	Histone H1.2	2.36	4.40E-38 (+)	0.002 (+)
HNRNPF	P52597	Heterogeneous nuclear ribonucleoprotein F	-0.78	0.00015 (+)	0.007 (+)

R: Reference, L: LIAVA, S: Scrambled; *Mean Difference Reference -'LIAVA' is the value that results of the substraction of the normalised mean intensity of 'LIAVA' from the normalised mean intensity of the Reference; [¶]Significance A of the Ratio of the Intensity Means indicates significant outlier values ("+") relative to a certain population with a two-sided test and Benjamini-Hochberg FDR correction with a threshold value of 0.05 (default); [#]Paired Student's T-test of the difference between two mean intensities, p-value <0.05 ("+"), IC 95%, n = 6, being n the number of independent RNA-pulldown experiments; [†]Histone H2B type 1-L, Histone H2B type 1-M, Histone H2B type 1-N, Histone H2B type 1-H, Histone H2B type 2-F, Histone H2B type 1-C/E/F/G/I, Histone H2B type 1-D, Histone H2B type 1-K, Histone H2B type F-S, Histone H2B type 2-E, Histone H2B type 1-B, Histone H2B type 1-O, Histone H2B type 1-J, Histone H2B type 3-B; [‡]HIST1H2BL, HIST1H2BM, HIST1H2BN, HIST1H2BH, HIST2H2BF, HIST1H2BC, HIST1H2BD, HIST1H2BK, H2BFS, HIST2H2BE, HIST1H2BB, HIST1H2BO, HIST1H2BJ, HIST3H2BB; [§]Q99880, Q99879, Q99877, Q93079, Q5QNW6, P62807, P58876, O60814, P57053, Q16778, P33778, P23527, P06899, Q8N257, Q96A08, Q6DRA6, Q6DN03.

In order to validate a potential functional effect of the enriched binding of hnRNPF in the context of haplotype-specific *OPN1LW/MW* exon 3 splicing, I performed small interfering RNA (siRNA)-mediated depletion combined with heterologous minigene splicing reporter assays (Cooper, 2005) for *OPN1LW/MW* exon 3 (Section 6.3).

Single knockdown of hnRNPF with an efficiency of ~60% at the protein level revealed no (rescue) effect on the splicing of the 'LIAVA' minigene – an haplotype which induces complete exon 3 skipping in this heterologus splicing assay (Buena-Atienza et al., 2016). Due to the potential ability of hnRNPF/H to substitute or replace each other, I further performed single and **double knockdown** of hnRNPF and hnRNPH1. Single knockdowns of hnRNPH1 and hnRNPF achieved ~75% and 60%

efficiency, respectively, whereas the double knockdown achieved ~73% and ~80% silencing efficiency for both hnRNPH1 and hnRNPF, respectively (Fig. 15).

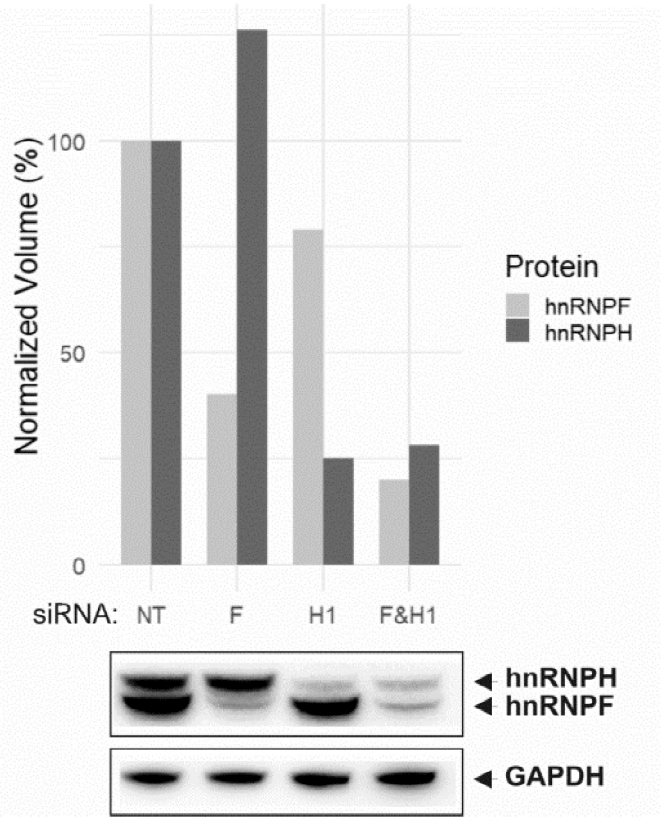


Figure 15. Efficacy of single and double siRNA-knockdowns of hnRNPF and hnRNPH1. Quantification of the bands detected by Western blot using an antibody against hnRNPF/H. The amount volume (%) of protein given as volume (%) was normalized to the GAPDH loading control for each siRNA treatment (NT:Non-Targeting, F: hnRNPF; H1: hnRNPH1 and F&H1: hnRNPF and hnRNPH1). The amount (in %) of protein obtained for each of the bands corresponding to hnRNPF and hnRNPH detected in the treatment condition with the Non-Targeting siRNA was considered as 100%. Calculated amounts of hnRNPH and hnRNPF for the different on-target siRNA treatment were: hnRNPF (125.8; 39.9), hnRNPH1 (24.8; 78.8) and double knockdown of hnRNPF and hnRNPH1 (27.3; 19.5), respectively.

Either single or double knockdowns performed in triplicates did not alter minigene-based *OPN1LW/MW* exon 3 inclusion or skipping levels as assessed by semi-quantitative PCR. Fluorescein Amidite (FAM)-labeled RT-PCR fragments were quantified (Appendix 13.2.4) based on capillary electrophoresis. A rescue effect to a certain level on 'LIAVA' haplotype splicing upon hnRNPF/H depletion, i.e. by observing an increase of exon 3 inclusion was not observed. Therefore, we reasoned that such a rescue might be more efficient or evident with other haplotypes being at

least partial splicing competent. I performed then minigene splicing assays applying a single construct harbouring a 'LMVVA' haplotype (c.453G; c.457C; c.465C; c.511A; c.513G; c.521T; c.532G; c.538G) which yields <10% of correctly spliced transcripts. This haplotype had been obtained through mutagenesis (Section 6.5). In addition, the Input Library pool of barcoded minigenes (Section 6.6) was used. These splicing assays were performed under conditons of single or double knockdown of hnRNPF/H.

Table 6. Splicing of exon 3 minigenes under conditions of hnRNPF/H depletion.

Exon 3 Haplotype*	siRNA Treatment†	Mean (%) Exon Inclusion‡	S.D. §	P-value¶
c.453G; c.457C; c.465C; c.511A; c.513G; c.521T; c.532G; c.538G	Non-Targeting	6.8485	0.7298	
	sihnRNPF	9.9063	4.1854	0.2794
	sihnRNPH1	6.0075	0.7736	0.4269
	sihnRNPF; sihnRNPH1	5.7959	1.7699	0.3549
>200 pool of haplotypes**	Non-Targeting	74.0649	0.8843	
	sihnRNPF	70.3766	3.1725	0.2557
	sihnRNPH1	67.7218	3.2066	0.1152
	sihnRNPF; sihnRNPH1	69.0628	0.4838	0.0150

*Exon 3 Haplotype encoded in the *OPN1LW/MW* minigene plasmid as described in Buena-Atienza et al., 2016; †siRNA-mediated knockdown of endogenous transcripts in HEK293T cells used for the minigene assays; ‡Mean % of RT-PCR products derived from correctly spliced transcripts including exon 3 as measured by FAM-based fluorescent semi-quantitative fragment analysis in triplicates; §Standard Deviation of the Mean % of exon inclusion over three replicates; ¶P-value of a Student's T-test of the Mean Difference in Exon Inclusion between Non-Targeting treatment and each of the targeted siRNA-mediated single or double knockdowns (IC 95%; n = 3); **Pool of minigene constructs (Input Library; see Section 6.5, Section 6.6 and Appendix 13.1.4, Publication IV).

The mean (%) of correctly spliced transcripts was calculated from triplicate experiments based on the area-under-curve analysis as described previously (Buena-Atienza et al. 2016). A two-sample Student’s T-test was used to calculate the confidence of 95% (alpha = 0.05) between the mean (%) correctly spliced levels detected in the Non-Targeting treatment and each of the three targeted knockdown treatments (Table 6). The double knockdown of hnRNPF and hnRNPH1 in cells co-transfected with the “Input Library” (Section 6.6) induced a statistically significant reduction of about 5% in the exon inclusion levels ($p<0.05$; Student’s T-test).

6.7.3 Discussion

Bioinformatic analysis using SpliceAid 2 and HSF yield a variety of RBPs predicted to be able to recognize regulatory motifs and thus, candidates to bind differentially to either the Reference or the 'LIAVA' haplotype (Table 4). Overall, for the variants present in the 'LIAVA' haplotype new ESE sites were predicted to be created for SRSF1, SRSF3, SRSF5, SRSF6 and SRSF7 whereas new ESS sites could be recognized by FUS, hnRNPA1, hnRNPH1, hnRNPH2, hnRNPH3 and hnRNPF. Broken sites by individual variants of the 'LIAVA' haplotype; for instance, SRSF2, SRSF3, SRSF7, SFRS9 and YBX1, for the enhancers, and hnRNPA1, hnRNPH1, hnRNPH2, hnRNPH3 and hnRNPF for the inhibitory splicing factors indicate putative binding to the Reference. Several of the latter splicing factors recognize both created and broken sites. Seemingly, the c.532G variant was only related to an inhibitory profile.

Unexpectedly, among the haplotype-specific enriched proteomic prey pool we only found a single well-known RBP, hnRNPF, with good candidacy for being involved in differential binding and splicing of *OPN1LW/MW* exon 3 haplotypes. Indeed, *in silico* tools had predicted both new and broken sites for hnRNPF (Table 4). hnRNPF is a member of the **hnRNP family** of RBPs known to participate in mRNA biogenesis (Piñol-Roma and Dreyfuss, 1993) and to play a role in splicing regulation defects underlying certain human diseases, for instance; SMA (Geuens et al., 2016), Breast Cancer (Goina et al., 2008) or Parkinson, where hnRNPF is involved in the haplotype-dependent regulation of *MAPT* exon 3 expression (Lai et al., 2017).

The 5' biotinylated RNA baits with a length of 109 nucleotides were designed to cover all common variant sites of the haplotype within exon 3 of *OPN1LW/MW* and flanking sequence (c.441-c.549). This length exceeds by large the usually length in RNA-pulldown experiments (~25nt RNA probes) and may have boosted the formation of stable secondary structures in our *in vitro* setting, thereby preventing the binding of additional putative *trans*-splicing factors (Goina et al., 2008; Rimoldi et al., 2013). Despite an almost negligible increase in GC content ('LIAVA': 56.88%; Reference: 55.05%) by two additional guanine bases, the RNA bait carrying 'LIAVA' was predicted to fold into the **thermodynamically most stable** secondary RNA structures (lowest MFE; Appendix 13.2.4; Supp. Table S1).

Both c.532A>G and c.538T>G variants as part of the ‘LIAVA’ haplotype create **guanine-rich elements** (also called G-runs, G-tracts or G-triplets, which are located in close proximity (UGG**G**UCUGG**G**; G-runs underlined, c.532A>G and c.538T>G in bold in the respective order). Notably, a major impact on splicing competence was observed in our *in vitro* multiplexed assay (Fig. 3B; Appendix 13.1.4, Publication IV) when the two latter variants were phased.

The candidate found to preferentially bind to the synthetic RNA carrying the ‘LIAVA’ haplotype – hnRNPF – is known to recognize **polyguanine** sequences (Matunis et al., 1994). The quasi-RNA recognition motifs of hnRNPF harbour β sheets with aromatic residues able to mediate intermolecular hydrophobic contacts and hydrogen bonds with motifs entailing the three consecutive guanine bases of target single-stranded RNA (ssRNA) (Dominguez et al., 2010; Wang and Yan, 2017). The modular structure of RBPs with several RRMs guarantee recognition of several motifs even if located far apart (Lunde et al., 2007).

However, the solely presence of isolated G-tracts might be insufficient for hnRNPF binding. An additonal G-tract (c.446-c.448) common to both ‘LIAVA’ and Reference haplotypes and a further G-tract (c.513-c.515) only present in the sequence of the Reference haplotype were found. hnRNPF was enriched for ‘LIAVA’ albeit present in all three RNA-bound fractions. Notwithstanding, a conglomeration of non-specific and specific G-runs in the 2D space structure was predicted only for *OPN1LW/MW* exon 3 harbouring the ‘LIAVA’ haplotype (Fig. 13) which may favour the binding effiiency of hnRNPF (Dominguez et al., 2010) through a 3-bp linker (Shamoo et al., 1995). In contrast, the two non-specific and specific G-runs are predicted far apart from each other in the 2D model of the Reference RNA (Fig. 13). Seemingly, the predicted base pairings in the secondary dimension of RNA sequences carrying either the Reference or the 'LIAVA' haplotype in exon 3 of *OPN1LW/MW* differ considerably (Fig. 13). The relative position of the created or abolished motifs to each other within the haplotype region in exon 3 was striking; an agglomerate of G-runs in 'LIAVA' that resemble a landing pad for protein recruitment. **RNA secondary structure** may also be influenced by other variants within the ‘LIAVA’ haplotype in conjunction with the G-run elements. For instance, when a shift in the conformation from a loop structure containing regulatory elements to a region with several consecutive base-pairing double-stranded RNA (Buratti and Baralle, 2004). An effect can be observed if the loop is dissolved and a RBP with an opposite activity recognizes double-stranded

RNA motifs (Ke et al., 2018). Indeed, RNA secondary structures have been shown to modulate the interaction of hnRNPA1-specific motifs in exon 12 of *Mag* (Zearfoss et al., 2013) and hnRNPH and SRSF2 in exon 6D of HIV-1 (Jablonski et al., 2008).

Applying the established heterologous minigene splicing assay (Section 6.3; Publication II), we assessed the effect on *OPN1LW/MW* exon 3 inclusion upon silencing of endogenous hnRNPF in HEK293T cells *in vitro*. However, **depletion of hnRNPF** yielded no significant difference in exon 3 inclusion levels in cells expressing the minigene harbouring the 'LMVVA' haplotype defined as c.453G; c.457C; c.465C; c.511A; c.513G; c.521T; c.532G; c.538G (Table 3). This haplotype carries all three G-runs (two shared with the 'LIAVA' haplotype that is not present in the Reference and one that is beared in the Reference haplotype but not in 'LIAVA'). The RNA2DMut model of ensemble centroid structure for 'MVVVA' and 'LMVVA' was virtually identical to that of 'LIAVA' (Supp. Fig. S3; Appendix 13.2.4).

The difference of the mean level of hnRNPH1 in the bound fractions obtained with the Reference RNA bait and the 'LIAVA' was not statistically significant. A stronger signal for hnRNPH than for hnRNPF was observed by immunodetection in RNA-bound eluted fractions which might have to do with different affinities for the RNA target and/or the antibody used. A higher expression of hnRNPH in HEK293T cells has been previously proposed (Alkan et al., 2006). siRNA-targeted downregulation of hnRNPF produced no off-target silencing effect on hnRNPH. Conversely, hnRNPH expression increased upon hnRNPF depletion (Fig. 8). A prior report hypothesized that this effect is due to loosening the autoregulation loop of exon 4 splicing of hnRNPH (Mauger et al., 2008). hnRNPF/H may have partially **overlapping** functions (Garneau et al., 2005) and are part in a fine-tuned interconnected regulatory mechanism in splicing (Huelga et al., 2012). A caveat is the fact that the amino acid sequence identity between hnRNPF and the two hnRNPH homologs (72 and 70% with hnRNPH1 and hnRNPH2, respectively) is highest within the three RRM domains (Mauger et al., 2008). As a matter of fact, hnRNPF/H are able to recognize **common RNA motifs** comprising G-runs such as UGGGA, UGGGC and UGGGU (Yoshida et al., 1999; Caputi and Zahler, 2001; Alkan et al., 2006; Dominguez and Allain, 2006; Wang et al., 2007; Russo et al., 2010; Giudice et al., 2016) although prediction tools (i.e. SpliceAid 2) only predict hnRNPH to bind to the G-run created by the c.532G (UGGGU) (Table 4).

This suggests that G-runs are *per se* weak consensus motifs, which might result in a competition for those binding sites *in vitro.* Additional *cis*-regulatory elements and *trans*-acting interacting partners might be necessary if hnRNPF- or hnRNPH-specific binding is mediated. Whereas hnRNPF belongs to a complex involved in neural-specific inclusion of the 18-bp c-*src* N1 exon in WERI-1 extracts (Min et al., 1995), hnRNPH did not show a specific binding pattern in retinal-derived cells (Caputi and Zahler, 2001), indicating that hnRNPF and hnRNPH functions are not necessarily redundant at least under certain conditions.

A stronger decrease in skipping was observed for alternatively spliced exons of the fibroblast growth factor receptor 2 (Mauger et al., 2008) and the acetylcholinesterase (Nazim et al., 2017) genes when both hnRNPF and hnRNPH were depleted instead of downregulation of either hnRNPF or hnRNPH alone. Similarly, we pursued siRNA-**knockdown** of both **hnRNPF and hnRNPH** in order to rule out potential compensatory or evolved synergistic effects (Wang and Cambi, 2009). However, no significant effect on exon inclusion of the aforementioned ‘LMVVA’ haplotype (Table 3) was observed. Interestingly, a net overall enhancing effect on exon inclusion upon hnRNPF silencing was detected as statistically significant with the Input Library (Section 6.6) pool of exon 3 haplotypes (Table 3). Although hnRNPs are generally regarded as inhibitors of splicing (Gallinaro et al., 1981), a systematic analysis of the transcriptome upon hnRNPF depletion showed an overall reduction in exon inclusion (Huelga et al., 2012). In fact, both inhibiting (Mauger et al., 2008) and enhancing (Wang et al., 2007; Talukdar et al., 2011) roles of hnRNPF/H in splicing have been reported. The position of the G-run with respect to the splice sites (Schaub et al., 2007) and the strength of the splice sites (Xiao et al., 2009) may influence hnRNPF/H activity, being mostly enhancing when binding to intronic G-runs – which are enriched nearby the splice sites (McCullough and Berget, 1997) – and inhibitory when acting on exons or pseudo-exons (Hai et al., 2008). The donor site at exon 2 of *OPN1LW/MW* (“UG│gu”) deviates from the highest consensus (“AG│gu”). Uracil is the second most common exonic base at -2 from the exon-intron boundary (“│”). Intronic G-triplets promote the usage of upstream 5’ss (McCullough and Berget, 2000). The G-runs might mask exon 3 as an intron and promote joining of exon 2 and exon 4. As a matter of fact, silencers were also created by G-tracts in introns (Sohail et al., 2014) and most surprisingly, distinct G-runs within a pseudoexon of the fibrinogen gamma-chain transcript have been show to result in antagonistic effects exerted by hnRNPF: the two

motifs nearby the splice sites act as silencers whereas the central G-run induces an enhancing role, supporting the dual and buffering activity of hnRNPF in splicing (Rimoldi et al., 2013). Hence, the presence of several G-runs in close proximity (as in exon 3 of *OPN1LW/MW*) which may exert **opposite effects** on the splicing competence of a particular exon may be difficult to dissect.

The moderate overall net reduction in exon inclusion observed with the pool of minigenes upon considerable hnRNPF/H depletion is not fully conclusive. Further **validation** is needed to decipher to what extent hnRNPF/H contribute to the splicing regulation of *OPN1LW/MW* exon 3. Although the achieved silencing efficiency was high (Fig. 8) and comparable to that in other reports (Huelga et al., 2012; Remenyi et al., 2016; Rimoldi et al., 2013; Talukdar et al., 2011; White et al., 2012), it remains elusive whether a higher efficiency of the simultaneous silencing of hnRNPF and hnRNPH would be required to trigger a stronger impact on splicing competence of *OPN1LW/MW* exon 3. Since hnRNPH1 and hnRNPH2 share an identity of ~80% and 96% in nucleotide and amino acid sequence, respectively (Mauger et al., 2008), further optimization of the triple knockdown with siRNAs targeting all three paralogous genes hnRNPF, hnRNPH1 and hnRNPH2 is required.

From this experimental study on *trans*-acting factors participating in *OPN1LW/MW* exon 3 splicing and mis-splicing, the following recommendations are proposed for future studies: i) *in vitro* transcription of the full-length pre-mRNA transcript (Marín-Béjar and Huarte, 2015) to generate a bait reminiscent of the endogenous pre-mRNA in terms of ensuring sequence space for accommodating splicing factors; only 11 nt upstream to the terminal 3’ end were available in the used synthetic RNA baits, ii) stable isotope labelling of amino acids in cell culture (SILAC) for joint purification of RNA-bound proteins for two different RNA baits to improve accuracy, reproducibility and relative quantification (Jazurek et al., 2016), iii) validation of mass spectrometry candidates *in vitro* by electrophoresis mobility shift assays (EMSA) and *in vivo* cross-linking of RNA and RBPs fused with fluorescent proteins (Strein et al., 2014), iv) Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 knockout (Fei et al., 2017) instead of siRNA knockdown, and v) if a *trans*-acting splicing factor is validated by knockdown or knockout experiments, the effect of its transient re-expression (and expected rescue of the knockdown/knockout effect) may be explored albeit the noise of non-transfected cells is a higher pitfall for over-expression (Cooper, 2005).

7 General discussion

The results presented in this thesis elucidated novel and unexpected aspects on the molecular genetic basis and pathomechanisms underlying BCM and related XLCD disorders. Thus, these will most likely improve the sensitivity and accuracy of the diagnosis of the latter conditions in the future. Extended pedigrees with multiple affected males are frequently recorded in patients with BCM. This and the recurrent observation of some frequent founder mutations demonstrate inheritance of disease-causing mutations for many generations. Yet, little is known on the rate and causes of *de novo* mutations and the underlying mechanisms which results in simultaneous mutations in both *OPN1LW* and *OPN1MW*.

Herein, I report the **simultaneous** presence of a novel intraexonic **deletion** in both *OPN1LW* and *OPN1MW* of a BCM subject, which putatively originated as a *de novo* event in a two-step mechanism (Section 6.2). While 30-100 SNVs occurr *de novo* per generation (Acuna-Hidalgo et al., 2016), *de novo* CNVs occur in only 1 out of 50 individuals (Veltman and Brunner, 2012). Moreover, I was able to identify and map a novel *de novo* >70 kb deletion encompassing large parts of the *OPN1LW/MW* gene cluster in a sporadic BCM patient (Publication 1; Section 6.1). This deletion likely resulted from NAHR between Alu elements; a mechanism frequently associated with segmental duplications (Jurka et al., 2004).

Molecular diagnostics of BCM subjects is challenging since the sum of copy number of *OPN1LW* and *OPN1MW* genes within the gene cluster varies in the population and the sequence identity between *OPN1LW* and *OPN1MW* is high. Short-read WGS and WES yield inconclusive or incorrect results and SNP arrays only partially detect CNVs (Bujakowska et al., 2017). Sanger sequencing of **gene-specific** amplicons is currently essential for accurate genotyping, especially for pinpointing breakpoints (Publication I; Section 6.1), and phasing differential SNVs within haplotypes belonging to either *OPN1LW* or *OPN1MW* (Publication II; Section 6.3). A technical caveat is still the discrimination among **phased alleles** from different distal *OPN1MW*(s) copies in *cis* or in different X-chromosomes in *trans* for females (Section 6.2). Barcoding of fragmented LD-PCR amplicons (Gould et al., 2018) and molecular inversion probes (Nuttle et al., 2013) circumvent some of the drawbacks for genotyping segmental duplications.

Segmental duplications including pairs of genes and pseudogenes are also prone to non-reciprocal recombination, a mechanism called non-allelic gene conversion (Hallast et al., 2005; Dumont and Eichler, 2013; Hartasánchez et al., 2014; Dumont, 2015). The occurence of **segmental duplications** is an efficient mechanism in evolution since it allows accumulation of novel variants and thus functional diversification of duplicated genes that can be fixated by natural selection. An example for this mechanism is the presence of distinct *OPN1LW* and *OPN1MW* genes in Old World primates and their diversification in terms of spectral senstitivty through acquisiation of few crucial amino acid changes (p.180, p.230, p.277 and p.285). The opposite **balancing** force shaping segmental duplications is non-allelic gene conversion which acts by homogenizing sequence and simultaneously conferring haplotype diversity (Verrelli and Tishkoff, 2004).

Being a known polymorphism shared between *OPN1LW* and *OPN1MW*, I showed that the c.538T>G variant is shuffled as a result of a single **gene conversion** event by evidence of phased haplotypes within the lineage of a single family (Publication II; Section 6.3) corroborating previous population-based evidence (Winderickx et al., 1993). I could further demonstrate a **functional** difference between the ancestral and the converted haplotypes in terms of the fraction of aberrantly spliced transcripts, and thus explain the difference in presented **phenotype** (Publication II; Section 6.3). The **splicing defect** induced by rare exon 3 haplotypes is a rather novel pathomechanism underlying *OPN1LW/MW*-linked disorders. In this thesis I have compiled the exon 3 haplotypes and their in-array composition in a large series of patients with severe XLCD and performed splicing assessment of a dozen haplotypes by a semi-quantification method, thereby corroborating and further extending the work by Ueyama and coworkers (Publication II; Section 6.3).

Exonic variants other than the canonical splice site mutations are generally underestimated in terms of burden for splicing precision and accuracy. Notably, 15-20% of exonic fourfold degenerate bases have been estimated to belong to functional ESEs (Savisaar and Hurst, 2018). Ortholog- or paralog-specific synonymous variants can have an influence on splicing competence when along analogous variants among ortholog or paralog genes (Section 6.4). A higher correlation between sequence conservation and effect on splicing was found when only **synonymous** variants were considered (Adamson et al., 2018), suggesting that regulation at silent sites is not neutral to evolution. Most likely, splicing regulation is not independent from selective

pressure on variants altering the protein sequence (Mueller et al., 2015; Pagani et al., 2005). Selection on protein-coding regions converge with selection forces to maintain splicing competence as documented by the fact that ~10% of rare pathogenic missense and nonsense variants disrupt splicing (Soemedi et al., 2017).

Indeed, a **pleiotropic** effect was observed for the c.1684C>T/p.Arg562Trp mutation in *Pde6a* which compromised splicing efficiency to considerable reduced levels while remaining correctly spliced *Pde6a* transcripts code for mutant proteins, showing impaired phosphodiesterase activity (Publication III; Section 6.4). Although two synonymous variants (c.453A>G and c.465C>G) are comprised the *OPN1LW/MW* exon 3 haplotypes resulting in exon skipping, the variants with the highest impact on splicing are the non-synonymous **c.532A>G/p.Ile178Val** and **c.538T>G/p.Ala180Ser** variants (Appendix 13.1.4, Publication IV). The latter was shown to shift the absorbance maximum of the resultant photopigment (Merbs and Nathans, 1992) and may be under counteracting balancing evolutionary forces resulting in the maintenance of partially deleterious alleles at relatively high frequencies. Haplotypes that have a clear detrimental outcome on exon 3 splicing and/or protein function may undergo negative selection and thus kept at low frequencies (i.e. 'LIAVA'), while that of individual SNPs can still be relatively high because they only exert subtle effects individually. *Cis*-regulatory mutations may contribute to phenotypic trait evolution (Wray, 2007). Evolutionary forces behind the splicing regulatory sequences may shape the compromise of splicing and protein function in humans. In Publication III we reported a missense mutation in *PDE6A* impairing both functional layers. This dual deleterious effect may account for the observation of this variant being ultra-rare or private in the population.

Notwithstanding, my investigations on both examples, the *PDE6A/B* human and *Pde6a/b* murine exon 13, and the human *OPN1LW/MW* exon 3 haplotypes, emphasize that splicing is a highly sequence **context-dependent** mechanism where surrounding variants with no apparent effect on splicing *per se* can interact with the putative splicing-disrupting variant and modulate the final splicing outcome. It remains to be determined whether this complexity in the *splicing code* holds true for the majority of genes and exons in mammalian species. The complex regulatory network underlying splicing of exon 3 *OPN1LW/MW* (Publication II and Publication IV) denotes features resembling the CERES; elements with overlapping inhibiting and enhancing activities (Haque et al., 2010; Tammaro et al., 2014). Whereas the consequence of

exon 3 splicing defect is an out-of-frame event which presumably leads to NMD of the aberrantly spliced *OPN1LW/MW* transcripts, the c.1684C>T leads to in-frame deletion in a fraction of transcripts, and these render shortened Pde6a proteins.

Consequently, a **milder phenotype** was observed for homozygous mice harbouring the c.1684C>T/p.Arg562Trp mutation in *Pde6a* (non-penetrant splicing defect) in comparison to the homozygous c.2053G>A/p.Val685Met mutant or animals from an intercross of both lines (compound heterozygotes) which showed more severe and faster photoreceptor degeneration (Publication III). Likewise, the more **severe phenotype** in subjects harbouring the 'LIAVA' haplotype in *OPN1LW/MW* genes well correlate with the fully penetrant exon 3 skipping defect, representing a *null* allele (Publication II). 'LIAVA'-opsin expressing cones in XLCD patients lack outer segments and thereby functionality as photoreceptors. Yet they retain viability and inner segment structure, as measured by high resolution *in vivo* imaging, and thus may still be targeted and recovered by gene therapy (Patterson et al., 2018). In contrast, patients harbouring the 'LVAVA' haplotype present with a submosaic of residual normal functioning cones. Still, the integrity of foveal layering is less preserved in patients bearing rare exon 3 haplotypes compared to patients with the p.Cys203Arg mutation (Carroll et al., 2012).

To characterize underlying pathomechanisms, a **reductionist** approach was followed (Yang et al., 2015): genotyping affected and unaffected individuals, variant identification, and use of model systems such as animal models (Publication III), cells (Publication II, Publication III) and *in vitro* assays (Section 6.7). Considering ~9 billion possible single substitutions and an estimated number of 30-100 *de novo* SNVs per genome – with a mutation rate of 1-1.8 x 10^{-8} per nucleotide per generation (Acuna-Hidalgo et al., 2016; Rahbari et al., 2016) –, characterization of every variant is relevant for assessment on pathogenicity potential and categorization. Switching from a retrospective to a **prospective** perspective requires high throughput functional characterization. This may be feasible through the implementation of MAVEs which can be used to train machine learning algorithms (Gray et al., 2018), as it has been shown for classification of variants creating cryptic splice sites (Lee et al., 2017). Nonetheless, MAVEs output and *a priori* assessment by means of *artificial* bioassays do not provide standalone evidence for variant categorization and should be understood as a resource to support clinical evidence (Nykamp et al., 2017).

Taking together, our individual and parallelized minigene splicing assays demonstrated that haplotypes with **c.532G and c.538G** lead to the highest increase in aberrantly spliced *OPN1LW/MW* transcripts (Publication II and Publication III), produced the tightest RNA secondary structures (Fig. 13) and both generated G-triplets in close proximity to each other to which the splicing factor hnRNPF preferentially binds *in vitro* (Section 6.7). Guanine-rich motifs interact with U1 snRNP directly (McCullough and Berget, 2000) and might be less tolerated close to the splice donor site (c.532G and c.538G are located towards the 3' end of exon 3). G-triplets generated by c.532G and c.538G may mask *OPN1LW/MW* exon 3 as a small intron where these are often found (≤200-bp introns) (McCullough and Berget, 1997).

Recently, the impact of splicing-disrupting variants on gene expression and the impact of the natural variation on splicing have acquired high interest. Genetic variation associated with alternative splicing can be detected by splicing quantitative trait locus (sQTL) analysis of population-based genotyping and RNASeq data, and eventually can lead to the identification of variants affecting *cis*-regulatory elements of splicing (Park et al., 2018). These may frequently manifest as expression quantitative trait locus (eQTLs) in *cis* (*cis*-eQTLs) and – in the context of disease – may modulate disease expression as proposed for exon recognition in regions associated with late-onset neurological disorders (Ramasamy et al., 2014). Applying MPRAs can resolve the causal variant(s) when sQTL are associated with defined haplotypes (Tewhey et al., 2016). Allele-specific splicing analyses interrogating coding SNPs may be used; yet, these are inconclusive if the coding SNP is located within the alternatively spliced region (Park et al., 2018).

SMRT sequencing of individual transcriptomes and barcode-based MPRAs elude the latter drawbacks. MAVEs can also aid to dissect the multiple layers of misregulation implicated in pathogenicity. A multitude of factors including but not limited to chromatin structure (i.e. DNA methylation) and the effect on RNA Polymerase II, *cis*-acting elements (Publication II-IV), *trans*-acting RBPs (Section 6.7) which are at the same time constrained by composite effects, sequence context modulation, distance between motifs, RNA secondary structure, mRNA accessibility, tissue- and cell-specific quantitative expression levels of RBPs and splice site strength, act in concert to determine exon fate (Baralle and Baralle, 2018).

8 Conclusions

i) Germline *de novo* mutations in the human *OPN1LW/MW* gene cluster should be considered in sporadic patients with BCM-like phenotype.

ii) Alu-mediated NAHR can cause deletions at the human *OPN1LW/MW* gene cluster.

iii) Homogenization of the exon 3 sequence between *OPN1LW* and *OPN1MW* gene copies by gene conversion can cause the transfer of pathogenic haplotypes.

iv) The actual combination of exonic variants as part of haplotypes; for instance, the 'LIAVA' and derivative rare haplotypes in the *OPN1LW* and *OPN1MW* genes, and the interactions among variants can have a major impact on splicing.

v) Incomplete penetrant splicing defects may still be pathogenic and contribute to the phenotypic variability caused by a defined disease gene – as exemplified by the milder phenotype of mice homozygous for c.1684C>T/p.Arg562Trp in *Pde6a*, and by the exon 3 haplotype-dependent disease severity in patients with *OPN1LW/MW*-linked cone dysfunction.

vi) The deleterious effect of exonic SNVs on splicing is context-dependent: synonymous variants close to the c.1684C>T/p.Arg562Trp mutation in the murine and the human *PDE6A gene*, or to the c.1678C>T/p.Arg560Cys mutation in *Pde6b* modulate the final splicing efficiency.

vii) A comprehensive dataset on the splicing competence of almost all possible *OPN1LW/MW* exon 3 (defined) haplotypes was established to aid *a priori* interpretation of variants in XLCD subjects. A total of 264 minigenes; including 20 barcoded and 12 non-barcoded minigenes assayed individually, as well as 232 minigenes assayed in a parallel format have been relatively quantified in terms of exon 3 inclusion or skipping.

viii) c.532 and c.538 are found to represent *cis*-regulatory elements of *OPN1LW/MW* splicing with the largest allelic impact on exon 3 inclusion.

ix) Preliminary evidence for haplotype-dependent preferential binding of RBPs *in vitro* might suggest a putative role of G-triplets (mapped with the c.532A>G and c.538T>G variants) in splicing modulation of *OPN1LW/MW* exon 3.

9 Outlook

Given the pace in technological improvement and development of suited algorithms (Merker et al., 2018; Vollger et al., 2019), I envisage that in the near future massive parallel sequencing will be applicable to determine the structure and sequence (composition) of the *OPN1LW/MW* gene cluster in routine genetic testing. High fidelity **long reads** from PCR-free enriched and/or genome sequencing (i.e. with read lengths ≥50kb) will enable unambiguous phasing of gene copy-specific haplotypes including their composition in subjects pre- and post-gene conversion events (Sections 6.2 and 6.3) together with **single-cell** gamete DNA sequencing. Ultra-long reads of >100kb (Jain et al., 2018) would enable to resolve the contiguity of the *OPN1LW* and *OPN1MW*(s) gene copies and mutations therein, which is needed for proper genotyping of subjects with *OPN1LW/MW*-linked disorders.

The establishment and technological implementation of the parallelized splicing assay (Sections 6.5 and 6.6) opens several avenues for future research. The mutation spectrum may be further expanded, either targeting additional exonic and intronic variants or in an unbiased approach of **random mutagenesis** by means of degenerate oligonucleotide synthesis or by spiked or doped oligonucleotide mutagenesis (Isalan, 2006). **Multiple barcodes** of random bases could be alternated with constant regions in between allowing for a higher complexity and number of differences in sequence space. In order to increase the counts per haplotype from long reads, the number of passes per molecule can be reduced by concatenation of the amplicons. In addition to accurate CCS long reads, nanopore long-read sequencing could be used for deeper **higher throughput** sequencing. It remains yet uncertain whether such an approach would reduce ambiguous mapping. To circumvent PCR biases, duplicates and chimeras, the Output Library pool of cDNAs can be now sequenced directly without amplification.

The MAVE (Section 6.6) eases comparative splicing assessment in **multiple cell lines** including cell lines of retinal origin. Ultimately, it would be most relevant to apply the MAVE *in vivo* in the retinae of animal models to characterize splicing in the **native cellular context** (Shen et al., 2016).

A variety of tools and approaches involving lentiviral vectors (Inoue et al., 2017), adeno-associated virus integration site 1 (AAVS1)-specific (Cheung et al., 2019) or

CRISPR/Cas9-mediated (Findlay et al., 2014; Gasperini et al., 2017; Klein et al., 2018) genome integration of the MAVE may help to reduce dosage bias and pleiotropic effects. The **endogenous** genomic context includes the possibility to study long-distance interactions (Singh et al., 2015) and provides an additional layer of regulation in the processing of *OPN1LW* and *OPN1MW* transcripts.

Long-read genome sequencing combined with direct RNA (Workman et al., 2018) or full-length cDNA targeted sequencing (Dougherty et al., 2018) of *OPN1LW*- and *OPN1MW*-specific transcripts from bulk RNA can reveal exon connectivity and transcript splicing diversity of duplicated genes. The later strategy can be applied on iPSCs-derived **photoreceptor** progenitor cells (Sangermano et al., 2016; Albert et al., 2018) or optic cups (Parfitt et al., 2016) from subjects carrying a single or multiple *OPN1LW/MW* gene copies with rare exon 3 haplotypes. To gain insight on the quantitative transcript profiles of exon 3 haplotypes that do not induce full exon skipping, fluid partitioning systems can be leveraged to sequence **single cells** expressing either *OPN1LW* or *OPN1MW* from a stem cell-derived cone population – pre-sorted with biomarkers for the two latter cone opsin types (Welby et al., 2017).

Finally, a **multi-disciplinary strategy** including advances in genomic engineering, stem cell differentiation and multiplexed splicing assays (Section 6.6) could achieve accurate characterization of splicing outcomes prospectively and at a higher throughput; without the need of establishing iPSCs from every patient. For instance, to leverage the potential of stem cell-derived cone photoreceptors, CRISPR/Cas9-based integration of a barcoded MAVE splicing reporter in the genome of such stem cell-derived cone precursors would provide cell-specific native context for splicing. Stem cell lines with well-characterized *OPN1LW/MW* genomic structure may be engineered to integrate the MAVE splicing barcoded library reporters specifically to *OPN1LW* and *OPN1MW* genomic locations. Upon differentiation into cone (precursor) cells, long- and short-read sequencing can be perform on either bulk RNA (barcoding strategy can differentiate *OPN1LW*- or *OPN1MW*-specific MAVEs) or with cDNA from pre-sorted single cells. In the latter scenario, experiments with MAVE-engineered stem cell-derived cones depleted for certain splicing factor(s) *versus* non-treated may decipher haplotypes sensitive to a given splicing factor (Sections 6.6 and 6.7). Moreover, one can envisage the application of the MAVE assay to screen for splice-modulation molecules with potential therapeutic impact.

10 References

Aboshiha, J., Dubis, A.M., Carroll, J., Hardcastle, A.J., and Michaelides, M. (2016). The cone dysfunction syndromes. Br. J. Ophthalmol. *100*, 115–121.

Acedo, A., Sanz, D.J., Durán, M., Infante, M., Pérez-Cabornero, L., Miner, C., and Velasco, E.A. (2012). Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. Breast Cancer Res. *14*, R87.

Acedo, A., Hernández-Moro, C., Curiel-García, Á., Díez-Gómez, B., and Velasco, E.A. (2015). Functional Classification of BRCA2 DNA Variants by Splicing Assays in a Large Minigene with 9 Exons. Hum. Mutat. *36*, 210–221.

Acuna-Hidalgo, R., Veltman, J.A., and Hoischen, A. (2016). New insights into the generation and role of de novo mutations in health and disease. Genome Biol. *17*, 241.

Adamson, S.I., Zhan, L., and Graveley, B.R. (2018). Vex-seq: high-throughput identification of the impact of genetic variation on pre-mRNA splicing efficiency. Genome Biol. *19*, 71.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.

Agca, C., Boldt, K., Gubler, A., Meneau, I., Corpet, A., Samardzija, M., Stucki, M., Ueffing, M., and Grimm, C. (2015). Expression of leukemia inhibitory factor in Müller glia cells is regulated by a redox-dependent mRNA stability mechanism. BMC Biol. *13*.

Aissat, A., Becdelièvre, A. de, Golmard, L., Vasseur, C., Costa, C., Chaoui, A., Martin, N., Costes, B., Goossens, M., Girodon, E., et al. (2013). Combined Computational–Experimental Analyses of CFTR Exon Strength Uncover Predictability of Exon-Skipping Level. Hum. Mutat. *34*, 873–881.

Albert, S., Garanto, A., Sangermano, R., Khan, M., Bax, N.M., Hoyng, C.B., Zernant, J., Lee, W., Allikmets, R., Collin, R.W.J., et al. (2018). Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. Am. J. Hum. Genet. *102*, 517–527.

Alkan, S.A., Martincic, K., and Milcarek, C. (2006). The hnRNPs F and H2 bind to similar sequences to influence gene expression. Biochem. J. *393*, 361–371.

Alpern, M., Lee, G.B., and Spivey, B.E. (1965). π1 Cone Monochromatism. Arch. Ophthalmol. *74*, 334–337.

Arguello, J.R., and Connallon, T. (2011). Gene Duplication and Ectopic Gene Conversion in Drosophila. Genes (Basel) 2, 131–151.

Arno, G., Carss, K.J., Hull, S., Zihni, C., Robson, A.G., Fiorentino, A., Black, G., Hall, G., Ingram, S., Gillespie, R., et al. (2017). Biallelic Mutation of ARHGEF18, Involved in the Determination of Epithelial Apicobasal Polarity, Causes Adult-Onset Retinal Degeneration. Am. J. Hum. Genet. *100*, 334–342.

Asenjo, A.B., Rim, J., and Oprian, D.D. (1994). Molecular determinants of human red/green color discrimination. Neuron *12*, 1131–1138.

Assis, R., and Kondrashov, A.S. (2012). A Strong Deletion Bias in Nonallelic Gene Conversion. PLOS Genetics *8*, e1002508.

Au, K.F., Sebastiano, V., Afshar, P.T., Durruthy, J.D., Lee, L., Williams, B.A., Bakel, H. van, Schadt, E.E., Reijo-Pera, R.A., Underwood, J.G., et al. (2013). Characterization of the human ESC transcriptome by hybrid sequencing. Proc. Natl. Acad. Sci. U.S.A. *110*, E4821–E4830.

Ávila-Fernández, A., Cortón, M., López-Molina, M.I., Martín-Garrido, E., Cantalapiedra, D., Fernández-Sánchez, R., Blanco-Kelly, F., Riveiro-Álvarez, R., Tatu, S.D., Trujillo-Tiebas, M.J., et al. (2011). Late Onset Retinitis Pigmentosa. Ophthalmology *118*, 2523–2524.

Ayyagari, R., Kakuk, L.E., Coats, C.L., Bingham, E.L., Toda, Y., Felius, J., and Sieving, P.A. (1999). Bilateral macular atrophy in blue cone monochromacy (BCM) with loss of the locus control region (LCR) and part of the red pigment gene. Mol. Vis. *5*, 13.

Bagnall, R.D., Ayres, K.L., Green, P.M., and Giannelli, F. (2005). Gene conversion and evolution of Xq28 duplicons involved in recurring inversions causing severe hemophilia A. Genome Res. *15*, 214–223.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. *2*, 28–36.

Baralle, M., and Baralle, F.E. (2018). The splicing code. Biosystems *164*, 39–48.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science *338*, 1587–1593.

Bareil, C., Hamel, C.P., Delague, V., Arnaud, B., Demaille, J., and Claustres, M. (2001). Segregation of a mutation in CNGB1 encoding the beta-subunit of the rod cGMP-gated channel in a family with autosomal recessive retinitis pigmentosa. Hum. Genet. *108*, 328–334.

Batzer, M.A., and Deininger, P.L. (2002). Alu repeats and human genomic diversity. Nat. Rev. Genet. *3*, 370–379.

Bauwens, M., De Zaeytijd, J., Weisschuh, N., Kohl, S., Meire, F., Dahan, K., Depasse, F., De Jaegere, S., De Ravel, T., De Rademaeker, M., et al. (2015). An augmented ABCA4 screen targeting noncoding regions reveals a deep intronic founder variant in Belgian Stargardt patients. Hum. Mutat. *36*, 39–42.

Bauwens, M., Garanto, A., Sangermano, R., Naessens, S., Weisschuh, N., Zaeytijd, J.D., Khan, M., Sadler, F., Balikova, I., Cauwenbergh, C.V., et al. (2019). 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. 1.

Becirovic, E., Böhm, S., Nguyen, O.N.P., Riedmayr, L.M., Koch, M.A., Schulze, E., Kohl, S., Borsch, O., Santos-Ferreira, T., Ader, M., et al. (2016). In Vivo Analysis of Disease-Associated Point Mutations Unveils Profound Differences in mRNA Splicing of Peripherin-2 in Rod and Cone Photoreceptors. PLOS Genet. *12*, e1005811.

Bellingham, J., Davidson, A.E., Aboshiha, J., Simonelli, F., Bainbridge, J.W., Michaelides, M., and Spuy, J. van der (2015). Investigation of Aberrant Splicing Induced by AIPL1 Variations as a Cause of Leber Congenital Amaurosis. Invest. Ophthalmol. Vis. Sci. *56*, 7784–7793. Berson, E.L., Sandberg, M.A., Rosner, B., and Sullivan, P.L. (1983). Color Plates to Help Identify Patients with Blue Cone Monochromatism. Am. J. Ophthalmol. *95*, 741–747.

Berson, E.L. (1993) Retinitis pigmentosa. The Friedenwald Lecture. Invest. Ophthalmol. Vis. Sci. *34*, 1659-1676.

Bhagavatula, G., Rich, M.S., Young, D.L., Marin, M., and Fields, S. (2017). A Massively Parallel Fluorescence Assay to Characterize the Effects of Synonymous Mutations on TP53 Expression. Mol. Cancer Res. *15*, 1301–1307.

Birot, A.-M., Bouton, O., Froissart, R., Maire, I., and Bozon, D. (1999). IDS gene-pseudogene exchange responsible for an intragenic deletion in a hunter patient. Hum. Mutat. *8*, 44–50.

Boldt, K., van Reeuwijk, J., Gloeckner, C.J., Ueffing, M., and Roepman, R. (2009). Chapter 9 - Tandem Affinity Purification of Ciliopathy-Associated Protein Complexes. In Methods in Cell Biology, S.M. King, and G.J. Pazour, eds. (Academic Press), pp. 143–160.

Boughman, J.A., Conneally, P.M., and Nance, W.E. (1980). Population genetic studies of retinitis pigmentosa. Am. J. Hum. Genet. *32*, 223–235.

Bowes, C., Li, T., Danciger, M., Baxter, L.C., Applebury, M.L., and Farber, D.B. (1990). Retinal degeneration in the rd mouse is caused by a defect in the β subunit of rod cGMP-phosphodiesterase. Nature *347*, 677–680.

Braun, T.A., Mullins, R.F., Wagner, A.H., Andorf, J.L., Johnston, R.M., Bakall, B.B., Deluca, A.P., Fishman, G.A., Lam, B.L., Weleber, R.G., et al. (2013). Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. Hum. Mol. Genet. *22*, 5136–5145.

Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic Integration of Splicing within Gene Regulatory Pathways. Cell *152*, 1252–1269.

Bravo-Gil, N., Pozo, M.G., Martín-Sánchez, M., Méndez-Vidal, C., Rúa, E.R. la, Borrego, S., and Antiñolo, G. (2017). Unravelling the genetic basis of simplex Retinitis Pigmentosa cases. Sci. Rep. *7*, 41937.

Brázda, V., Kolomazník, J., Lýsek, J., Bartas, M., Fojta, M., Šťastný, J., and Mergny, J.-L. (2019). G4Hunter web application: a web server for G-quadruplex prediction. Bioinformatics *35*, 3493–3495.

Brody, E., and Abelson, J. (1985). The "spliceosome": yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science *228*, 963–967.

Buena-Atienza, E., Nasser, F., Kohl, S., and Wissinger, B. (2018). A 73,128 bp de novo deletion encompassing the *OPN1LW/OPN1MW* gene cluster in sporadic Blue Cone Monochromacy: a case report. BMC Med. Genet. *19*, 107.

Buena-Atienza, E., Rüther, K., Baumann, B., Bergholz, R., Birch, D., De Baere, E., Dollfus, H., Greally, M.T., Gustavsson, P., Hamel, C.P., et al. (2016). *De novo* intrachromosomal gene conversion from *OPN1MW* to *OPN1LW* in the male germline results in Blue Cone Monochromacy. Sci. Rep. *6*, 28253.

Bujakowska, K.M., Fernandez-Godino, R., Place, E., Consugar, M., Navarro-Gomez, D., White, J., Bedoukian, E.C., Zhu, X., Xie, H.M., Gai, X., et al. (2017). Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations. Genet. Med. *19*, 643–651.

Bundey, S., and Crews, S.J. (1984). A study of retinitis pigmentosa in the City of Birmingham. I Prevalence. J. Med. Genet. *21*, 417–420.

Bunge, S., Rathmann, M., Steglich, C., Bondeson, M.-L., Tylki-Szymanska, A., Popowska, E., and Gal, A. (1998). Homologous nonallelic recombinations between the iduronate-sulfatase gene and pseudogene cause various intragenic deletions and inversions in patients with mucopolysaccharidosis type II. Eur. J. Hum. Genet. *6*, 492–500.

Bunyan, D.J., and Robinson, D.O. (2008). Multiple de novo mutations in the MECP2 gene. Genet. Test. *12*, 373–375.

Buratti, E., and Baralle, F.E. (2004). Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. Mol. Cell. Biol. *24*, 10505–10514.

Burke, C.M., and Darling, A.E. (2016). A method for high precision sequencing of near full-length 16S rRNA genes on an Illumina MiSeq. PeerJ *4*, e2492.

Burn, T.C., Connors, T.D., Klinger, K.W., and Landes, G.M. (1995). Increased exon-trapping efficiency through modifications to the pSPL3 splicing vector. Gene *161*, 183–187.

Bystrykh, L.V., Haan, G. de, and Verovskaya, E. (2014). Barcoded Vector Libraries and Retroviral or Lentiviral Barcoding of Hematopoietic Stem Cells. In Hematopoietic Stem Cell Protocols, (Humana Press, New York, NY), pp. 345–360.

Caceres, J.F., Stamm, S., Helfman, D.M., and Krainer, A.R. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science *265*, 1706–1709.

Cagliani, R., Riva, S., Pozzoli, U., Fumagalli, M., Comi, G.P., Bresolin, N., Clerici, M., and Sironi, M. (2011). Balancing selection is common in the extended MHC region but most alleles with opposite risk profile for autoimmune diseases are neutrally evolving. BMC Evol. Biol. *11*, 171.

Caminsky, N.G., Mucaki, E.J., and Rogan, P.K. (2014). Interpretation of mRNA splicing mutations in genetic disease: review of the literature and guidelines for information-theoretical analysis. F1000Research *3*.

Caputi, M., and Zahler, A.M. (2001). Determination of the RNA Binding Specificity of the Heterogeneous Nuclear Ribonucleoprotein (hnRNP) H/H′/F/2H9 Family. J. Biol. Chem. *276*, 43850–43859.

Cardoso, A.R., Oliveira, M., Amorim, A., and Azevedo, L. (2016). Major influence of repetitive elements on disease-associated copy number variants (CNVs). Hum. Genomics *10*.

Carroll, J., Neitz, M., Hofer, H., Neitz, J., and Williams, D.R. (2004). Functional photoreceptor loss revealed with adaptive optics: An alternate cause of color blindness. Proc. Natl. Acad. Sci. U.S.A. *101*, 8461–8466.

Carroll, J., Baraas, R.C., Wagner-Schuman, M., Rha, J., Siebe, C.A., Sloan, C., Tait, D.M., Thompson, S., Morgan, J.I.W., Neitz, J., et al. (2009). Cone photoreceptor mosaic disruption associated with Cys203Arg mutation in the M-cone opsin. Proc. Natl. Acad. Sci. U.S.A. *106*, 20948–20953.

Carroll, J., Dubra, A., Gardner, J.C., Mizrahi-Meissonnier, L., Cooper, R.F., Dubis, A.M., Nordgren, R., Genead, M., Connor, T.B., Stepien, K.E., et al. (2012). The Effect of Cone Opsin Mutations on Retinal Structure and the Integrity of the Photoreceptor Mosaic. Invest. Ophthalmol. Vis. Sci. *53*, 8006–8015.

Cartegni, L., and Krainer, A.R. (2002). Disruption of an SF2/ASF-dependent exonic splicing enhancer in *SMN2* causes spinal muscular atrophy in the absence of *SMN1*. Nat. Genet. *30*, 377–384.

Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat. Rev. Genet. *3*, 285–298.

Carvalho, C.M.B., and Lupski, J.R. (2016). Mechanisms underlying structural variant formation in genomic disorders. Nat. Rev. Genet. *17*, 224–238.

Casbon, J.A., Slatter, A.F., Musgrave-Brown, E., Osborne, R.J., Lichtenstein, C.P., and Brenner, S. (2013). Reflex: intramolecular barcoding of long-range PCR products for sequencing multiple pooled DNAs. Nucleic Acids Res. *41*, e112–e112.

Casola, C., Zekonyte, U., Phillips, A.D., Cooper, D.N., and Hahn, M.W. (2012). Interlocus gene conversion events introduce deleterious mutations into at least 1% of human genes associated with inherited disease. Genome Res. *22*, 429–435.

Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., Gardner, E.J., Rodriguez, O.L., Guo, L., Collins, R.L., et al. (2019). Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat. Commun. *10*, 1784.

Chang, B., Hawes, N.L., Pardue, M.T., German, A.M., Hurd, R.E., Davisson, M.T., Nusinowitz, S., Rengarajan, K., Boyd, A.P., Sidney, S.S., et al. (2007). Two mouse retinal degenerations caused by missense mutations in the β-subunit of rod cGMP phosphodiesterase gene. Vision Res. *47*, 624–633.

Chang, B., Grau, T., Dangel, S., Hurd, R., Jurklies, B., Sener, E.C., Andreasson, S., Dollfus, H., Baumann, B., Bolz, S., et al. (2009). A homologous genetic basis of the murine cpfl1 mutant and human achromatopsia linked to mutations in the PDE6C gene. Proc. Natl. Acad. Sci. U. S. A. *106*, 19581–19586.

Chen, E.Y., Zollo, M., Mazzarella, R., Ciccodicola, A., Chen, C., Zuo, L., Heiner, C., Repetto, M., Burough, F., Schlessinger, D., et al. (1996). Long-Range Sequence Analysis in Xq28: Thirteen Known and Six Candidate Genes in 219.4 kb of High GC DNA Between the RCP/GCP and G6PD Loci. Hum. Mol. Genet. *5*, 659–668.

Chen, J.-M., Cooper, D.N., Chuzhanova, N., Férec, C., and Patrinos, G.P. (2007). Gene conversion: mechanisms, evolution and human disease. Nature Reviews Genetics *8*, 762–775.

Chen, L., Zhou, W., Zhang, L., and Zhang, F. (2014). Genome Architecture and Its Roles in Human Copy Number Variation. Genomics Inform. *12*, 136–144.

Cheung, R., Insigne, K.D., Yao, D., Burghard, C.P., Jones, E.M., Goodman, D.B., and Kosuri, S. (2018). Many rare genetic variants have unrecognized large-effect disruptions to exon recognition. BioRxiv 199927.

Cheung, R., Insigne, K.D., Yao, D., Burghard, C.P., Wang, J., Hsiao, Y.-H.E., Jones, E.M., Goodman, D.B., Xiao, X., and Kosuri, S. (2019). A Multiplexed Assay for Exon Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions. Mol. Cell *73*, 183-194.e8.

Chiang, J. (pei-wen), and Trzupek, K. (2015). The current status of molecular diagnosis of inherited retinal dystrophies. Curr. Opin. Ophthalmol. *26*, 346–351.

Chuang, K., Fields, M.A., and Del Priore, L.V. (2017). Potential of Gene Editing and Induced Pluripotent Stem Cells (iPSCs) in Treatment of Retinal Diseases. Yale J. Biol. Med. *90*, 635–642.

Chuzhanova, N., Chen, J.-M., Bacolla, A., Patrinos, G.P., Férec, C., Wells, R.D., and Cooper, D.N. (2009). Gene conversion causing human inherited disease: Evidence for involvement of non-B-DNA-forming sequences and recombination-promoting motifs in DNA breakage and repair. Hum. Mutat. *30*, 1189–1198.

Cideciyan, A.V., Hufnagel, R.B., Carroll, J., Sumaroka, A., Luo, X., Schwartz, S.B., Dubra, A., Land, M., Michaelides, M., Gardner, J.C., et al. (2013). Human Cone Visual Pigment Deletions Spare Sufficient Photoreceptors to Warrant Gene Therapy. Hum. Gene Ther. *24*, 993–1006.

Cook, G.W., Konkel, M.K., Walker, J.A., Bourgeois, M.G., Fullerton, M.L., Fussell, J.T., Herbold, H.D., and Batzer, M.A. (2013). A Comparison of 100 Human Genes Using an Alu Element-Based Instability Model. PLOS ONE *8*, e65188.

Cooper, T.A. (2005). Use of minigene systems to dissect alternative splicing elements. Methods *37*, 331–340.

Cornelis, S.S., Bax, N.M., Zernant, J., Allikmets, R., Fritsche, L.G., den Dunnen, J.T., Ajmal, M., Hoyng, C.B., and Cremers, F.P.M. (2017). In Silico Functional Meta-Analysis of 5,962 ABCA4 Variants in 3,928 Retinal Dystrophy Cases. Hum. Mutat. *38*, 400–408.

Corton, M., Blanco, M.J., Torres, M., Sanchez-Salorio, M., Carracedo, A., and Brion, M. (2010). Identification of a novel mutation in the human PDE6A gene in autosomal recessive retinitis pigmentosa: homology with the nmf28/nmf28 mice model. Clin. Genet. *78*, 495–498.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Cremers, F.P.M., Boon, C.J.F., Bujakowska, K., and Zeitz, C. (2018). Special Issue Introduction: Inherited Retinal Disease: Novel Candidate Genes, Genotype–Phenotype Correlations, and Inheritance Models. Genes *9*, 215.

Crognale, M.A., Fry, M., Highsmith, J., Haegerstrom-Portnoy, G., Neitz, M., Neitz, J., and Webster, M.A. (2004). Characterization of a novel form of X-linked incomplete achromatopsia. Vis. Neurosci. *21*, 197–203.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Res. *14*, 1188-1190.

Cummings, B.B., Marshall, J.L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A.R., Bolduc, V., Waddell, L.B., Sandaradura, S.A., O'Grady, G.L., et al. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci. Transl. Med. *9*, 1–25.

Daguenet, E., Dujardin, G., and Valcárcel, J. (2015). The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. EMBO Rep. *16*, 1640–1655.

Daiger, S.P., Bowne, S.J., and Sullivan, L.S. (2007). Perspective on Genes and Mutations Causing Retinitis Pigmentosa. Arch. Ophthalmol. *125*, 151–158.

Dal Mas, A., Rogalska, M.E., Bussani, E., and Pagani, F. (2015). Improvement of SMN2 Pre-mRNA Processing Mediated by Exon-Specific U1 Small Nuclear RNA. Am. J. Hum. Genet. *96*, 93–103.

D'Amore, R., Johnson, J., Haldenby, S., Hall, N., Hughes, M., Joynson, R., Kenny, J.G., Patron, N., Hertz-Fowler, C., and Hall, A. (2017). SMRT Gate: A method for validation of synthetic constructs on Pacific Biosciences sequencing platforms. BioTechniques *63*, 13–20.

De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip. Rev. RNA *4*, 49–60.

Deeb, S.S., Lindsey, D.T., Hibiya, Y., Sanocki, E., Winderickx, J., Teller, D.Y., and Motulsky, A.G. (1992). Genotype-phenotype relationships in human red/green color-vision defects: molecular and psychophysical studies. Am. J. Hum. Genet. *51*, 687–700.

Deininger, P. (2011). Alu elements: know the SINEs. Genome Biol. *12*, 236.

Dennis, M.Y., Harshman, L., Nelson, B.J., Penn, O., Cantsilieris, S., Huddleston, J., Antonacci, F., Penewit, K., Denman, L., Raja, A., et al. (2017). The evolution and population diversity of human-specific segmental duplications. Nat. Ecol. Evol. *1*, 69

Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., and Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. *37*, e67.

Desviat, L.R., Pérez, B., and Ugarte, M. (2012). Minigenes to Confirm Exon Skipping Mutations. Methods Mol Biol. *867*, 37–47.

Doktor, T.K., Schrøder, L.D., Andersen, H.S., Brøner, S., Kitewska, A., Sørensen, C.B., and Andresen, B.S. (2014). Absence of an intron splicing silencer in porcine Smn1 intron 7 confers immunity to the exon skipping mutation in human SMN2. PloS One *9*, e98841.

Dominguez, C., and Allain, F.H.-T. (2006). NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. Nucleic Acids Res. *34*, 3634–3645.

Dominguez, C., Fisette, J.-F., Chabot, B., and Allain, F.H.-T. (2010). Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. Nat. Struct. Mol. Biol. *17*, 853.

Donadon, I., McVey, J.H., Garagiola, I., Branchini, A., Mortarino, M., Peyvandi, F., Bernardi, F., and Pinotti, M. (2018). Clustered F8 missense mutations cause hemophilia A by combined alteration of splicing and protein biosynthesis and activity. Haematologica *103*, 344–350.

Dougherty, M.L., Underwood, J.G., Nelson, B.J., Tseng, E., Munson, K.M., Penn, O., Nowakowski, T.J., Pollen, A.A., and Eichler, E.E. (2018). Transcriptional fates of human-specific segmental duplications in brain. Genome Res. *28*, 1566–1576

Dryja, T.P., Finn, J.T., Peng, Y.W., McGee, T.L., Berson, E.L., and Yau, K.W. (1995). Mutations in the gene encoding the alpha subunit of the rod cGMP-gated channel in autosomal recessive retinitis pigmentosa. Proc. Natl. Acad. Sci. U. S. A. *92*, 10177–10181.

Dryja, T.P., Rucinski, D.E., Chen, S.H., and Berson, E.L. (1999). Frequency of Mutations in the Gene Encoding the α Subunit of Rod cGMP-Phosphodiesterase in Autosomal Recessive Retinitis Pigmentosa. Invest. Ophthalmol. Vis. Sci. *40*, 1859–1865.

Dumont, B.L. (2015). Interlocus gene conversion explains at least 2.7 % of single nucleotide variants in human segmental duplications. BMC Genomics *16*, 456.

Dumont, B.L., and Eichler, E.E. (2013). Signals of Historical Interlocus Gene Conversion in Human Segmental Duplications. PLOS ONE *8*, e75949.

Duyzend, M.H., Nuttle, X., Coe, B.P., Baker, C., Nickerson, D.A., Bernier, R., and Eichler, E.E. (2016). Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. Am. J. Hum. Genet. *98*, 45–57.

Dvir, L., Srour, G., Abu-Ras, R., Miller, B., Shalev, S.A., and Ben-Yosef, T. (2010). Autosomal-Recessive Early-Onset Retinitis Pigmentosa Caused by a Mutation in PDE6G, the Gene Encoding the Gamma Subunit of Rod cGMP Phosphodiesterase. Am. J. Hum. Genet. *87*, 258–264.

Eisenberger, T., Neuhaus, C., Khan, A.O., Decker, C., Preising, M.N., Friedburg, C., Bieg, A., Gliem, M., Charbel Issa, P., Holz, F.G., et al. (2013). Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. PloS One *8*, e78496.

Elkon, R., and Agami, R. (2017). Characterization of noncoding regulatory DNA in the human genome. Nat. Biotechnol. *35*, 732–746.

Ellingford, J.M., Horn, B., Campbell, C., Arno, G., Barton, S., Tate, C., Bhaskar, S., Sergouniotis, P.I., Taylor, R.L., Carss, K.J., et al. (2017). Assessment of the incorporation of CNV surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. J. Med. Genet. *55*, 114–121.

Erkelenz, S., Theiss, S., Otte, M., Widera, M., Peter, J.O., and Schaal, H. (2014). Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic Acids Res *42*, 10681–10697.

Eslami Rasekh, M., Chiatante, G., Miroballo, M., Tang, J., Ventura, M., Amemiya, C.T., Eichler, E.E., Antonacci, F., and Alkan, C. (2017). Discovery of large genomic inversions using long range information. BMC Genomics *18*, 65.

Falanga, A., Stojanović, O., Kiffer-Moreira, T., Pinto, S., Millán, J.L., Vlahoviček, K., and Baralle, M. (2014). Exonic splicing signals impose constraints upon the evolution of enzymatic activity. Nucleic Acids Res. *42*, 5790–5798.

Farber, D.B., and Lolley, R.N. (1974). Cyclic guanosine monophosphate: elevation in degenerating photoreceptor cells of the C3H mouse retina. Science *186*, 449–451.

Farkas, M.H., Grant, G.R., White, J.A., Sousa, M.E., Consugar, M.B., and Pierce, E.A. (2013). Transcriptome analyses 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 *14*, 486.

Fawcett, J.A., and Innan, H. (2011). Neutral and Non-Neutral Evolution of Duplicated Genes with Gene Conversion. Genes *2*, 191–209.

Fei, T., Chen, Y., Xiao, T., Li, W., Cato, L., Zhang, P., Cotter, M.B., Bowden, M., Lis, R.T., Zhao, S.G., et al. (2017). Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. Proc. Natl. Acad. Sci. U.S.A. *114*, E5207–E5215.

Findlay, G.M., Boyle, E.A., Hause, R.J., Klein, J.C., and Shendure, J. (2014). Saturation editing of genomic regions by multiplex homology-directed repair. Nature *513*, 120–123.

Fishman, G.A. (1978). Retinitis pigmentosa. Genetic percentages. Arch. Ophthalmol. Chic. Ill 1960 *96*, 822–826.

Flynn, E.K., Kamat, A., Lach, F.P., Donovan, F.X., Kimble, D.C., Narisu, N., Sanborn, E., Boulad, F., Davies, S.M., Gillio, A.P., et al. (2014). Comprehensive Analysis of Pathogenic Deletion Variants in Fanconi Anemia Genes. Hum. Mutat. *35*, 1342–1353.

Fraile-Bethencourt, E., Díez-Gómez, B., Velásquez-Zapata, V., Acedo, A., Sanz, D.J., and Velasco, E.A. (2017). Functional classification of DNA variants by hybrid minigenes: Identification of 30 spliceogenic variants of BRCA2 exons 17 and 18. PLOS Genet. *13*, e1006691.

Fraile-Bethencourt, E., Valenzuela-Palomo, A., Díez-Gómez, B., Acedo, A., and Velasco, E.A. (2018). Identification of Eight Spliceogenic Variants in BRCA2 Exon 16 by Minigene Assays. Front. Genet. *9*, 188.

Franke, G., Bausch, B., Hoffmann, M.M., Cybulla, M., Wilhelm, C., Kohlhase, J., Scherer, G., and Neumann, H.P.H. (2009). Alu-Alu recombination underlies the vast majority of large VHL germline deletions: Molecular characterization and genotype-phenotype correlations in VHL patients. Hum. Mutat. *30*, 776–786.

Frederiksen, A.L., Duno, M., and Welinder, L.G. (2013). Blue cone monochromatism in a female due to skewed X-inactivation. Ophthalmic Genet. *34*, 101–104.

Fu, Y. (1995). Phototransduction in Rods and Cones. In Webvision: The Organization of the Retina and Visual System, H. Kolb, E. Fernandez, and R. Nelson, eds. (Salt Lake City (UT): University of Utah Health Sciences Center), https://www.ncbi.nlm.nih.gov/pubmed/21413389

Gaildrat, P., Killian, A., Martins, A., Tournier, I., Frébourg, T., and Tosi, M. (2010). Use of Splicing Reporter Minigene Assay to Evaluate the Effect on Splicing of Unclassified Genetic Variants. In Cancer Susceptibility: Methods and Protocols, M. Webb, ed. (Totowa, NJ: Humana Press), pp. 249–257.

Gallego-Paez, L.M., Bordone, M.C., Leote, A.C., Saraiva-Agostinho, N., Ascensão-Ferreira, M., and Barbosa-Morais, N.L. (2017). Alternative splicing: the pledge, the turn, and the prestige. Hum. Genet. *136*, 1015–1042.

Gallinaro, H., Lazar, E., Jacob, M., Krol, A., and Branlant, C. (1981). Small RNAs in HnRNP fibrils and their possible function in splicing. Mol. Biol. Rep. *7*, 31–39.

Gamsiz, E.D., Ouyang, Q., Schmidt, M., Nagpal, S., and Morrow, E.M. (2012). Genome-wide transcriptome analysis in murine neural retina using high-throughput RNA sequencing. Genomics *99*, 44–51.

Garanto, A., Riera, M., Pomares, E., Permanyer, J., Castro-Miró, M. de, Sava, F., Abril, J.F., Marfany, G., and Gonzàlez-Duarte, R. (2011). High Transcriptional Complexity of the Retinitis Pigmentosa CERKL Gene in Human and Mouse. Invest. Ophthalmol. Vis. Sci. *52*, 5202–5214.

Garanto, A., Beersum, S.E.C. van, Peters, T.A., Roepman, R., Cremers, F.P.M., and Collin, R.W.J. (2013). Unexpected CEP290 mRNA Splicing in a Humanized Knock-In Mouse Model for Leber Congenital Amaurosis. PLOS ONE *8*, e79369.

Garanto, A., Duijkers, L., and Collin, R.W.J. (2015). Species-Dependent Splice Recognition of a Cryptic Exon Resulting from a Recurrent Intronic CEP290 Mutation that Causes Congenital Blindness. Int. J. Mol. Sci. *16*, 5285–5298.

Gardner, J.C., Michaelides, M., Holder, G.E., Kanuga, N., Webb, T.R., Mollon, J.D., Moore, A.T., and Hardcastle, A.J. (2009). Blue cone monochromacy: causative mutations and associated phenotypes. Mol. Vis. *15*, 876–884.

Gardner, J.C., Webb, T.R., Kanuga, N., Robson, A.G., Holder, G.E., Stockman, A., Ripamonti, C., Ebenezer, N.D., Ogun, O., Devery, S., et al. (2010). X-Linked Cone Dystrophy Caused by Mutation of the Red and Green Cone Opsins. The American Journal of Human Genetics *87*, 26–39.

Gardner, J.C., Liew, G., Quan, Y.-H., Ermetal, B., Ueyama, H., Davidson, A.E., Schwarz, N., Kanuga, N., Chana, R., Maher, E.R., et al. (2014). Three Different Cone Opsin Gene Array Mutational Mechanisms with Genotype–Phenotype Correlation and Functional Investigation of Cone Opsin Variants. Hum. Mutat. *35*, 1354–1362.

Gardner, J.C., Michaelides, M., and Hardcastle, A.J. (2016). Cone opsins, colour blindness and cone dystrophy: Genotype-phenotype correlations. South Afr. Med. J. Suid-Afr. Tydskr. Vir Geneeskd. *106*, S75-78.

Garneau, D., Revil, T., Fisette, J.-F., and Chabot, B. (2005). Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. J. Biol. Chem. *280*, 22641–22650.

Gasperini, M., Starita, L., and Shendure, J. (2016). The power of multiplexed functional analysis of genetic variants. Nat. Protoc. *11*, 1782–1787.

Gasperini, M., Findlay, G.M., McKenna, A., Milbank, J.H., Lee, C., Zhang, M.D., Cusanovich, D.A., and Shendure, J. (2017). CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. Am. J. Hum. Genet. *101*, 192–205.

Georgomanolis, T., Sofiadis, K., and Papantonis, A. (2016). Cutting a Long Intron Short: Recursive Splicing and Its Implications. Front. Physiol. *7*, 598.

Geuens, T., Bouhy, D., and Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. Hum. Genet. *135*, 851–867.

Gilbert, W. (1978). Why genes in pieces? Nature *271*, 501.

Girirajan, S. (2017). Missing heritability and where to find it. Genome Biol. *18*, 89.

Giudice, G., Sánchez-Cabo, F., Torroja, C., and Lara-Pezzi, E. (2016). ATtRACT—a database of RNA-binding proteins and associated motifs. Database J. Biol. Databases Curation 2016:baw035.

Gladman, J.T., Bebee, T.W., Edwards, C., Wang, X., Sahenk, Z., Rich, M.M., and Chandler, D.S. (2010). A humanized Smn gene containing the SMN2 nucleotide alteration in exon 7 mimics SMN2 splicing and the SMA disease phenotype. Hum. Mol. Genet. *19*, 4239–4252.

Glöckle, N., Kohl, S., Mohr, J., Scheurenbrand, T., Sprecher, A., Weisschuh, N., Bernd, A., Rudolph, G., Schubach, M., Poloschek, C., et al. (2014). Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. Eur. J. Hum. Genet. EJHG *22*, 99–104.

Goina, E., Skoko, N., and Pagani, F. (2008). Binding of DAZAP1 and hnRNPA1/A2 to an Exonic Splicing Silencer in a Natural BRCA1 Exon 18 Mutant. Mol. Cell. Biol. *28*, 3850–3860.

Gould, G.M., Grauman, P.V., Theilmann, M.R., Spurka, L., Wang, I.E., Melroy, L.M., Chin, R.G., Hite, D.H., Chu, C.S., Maguire, J.R., et al. (2018). Detecting clinically actionable variants in the 3′ exons of PMS2 via a reflex workflow based on equivalent hybrid capture of the gene and its pseudogene. BMC Med. Genet. *19*, 176.

Grau, T., Artemyev, N.O., Rosenberg, T., Dollfus, H., Haugen, O.H., Cumhur Sener, E., Jurklies, B., Andreasson, S., Kernstock, C., Larsen, M., et al. (2011). Decreased catalytic activity and altered activation properties of PDE6C mutants associated with autosomal recessive achromatopsia. Hum. Mol. Genet. *20*, 719–730.

Gray, V.E., Hause, R.J., Luebeck, J., Shendure, J., and Fowler, D.M. (2018). Quantitative Missense Variant Effect Prediction Using Large-Scale Mutagenesis Data. Cell Syst. *6*, 116-124.e3.

Green, D.G. (1972). Visual acuity in the blue cone monochromat. J. Physiol. *222*, 419–426.

Greenwald, S.H., Kuchenbecker, J.A., Rowlan, J.S., Neitz, J., and Neitz, M. (2017). Role of a Dual Splicing and Amino Acid Code in Myopia, Cone Dysfunction and Cone Dystrophy Associated with L/M Opsin Interchange Mutations. Transl. Vis. Sci. Technol. *6*.

Grimm, T., Meng, G., Liechti-Gallati, S., Bettecken, T., Müller, C.R., and Müller, B. (1994). On the origin of deletions and point mutations in Duchenne muscular dystrophy: most deletions arise in oogenesis and most point mutations result from events in spermatogenesis. J. Med. Genet. *31*, 183–186.

Grover, D., Mukerji, M., Bhatnagar, P., Kannan, K., and Brahmachari, S.K. (2004). Alu repeat analysis in the complete human genome: trends and variations with respect to genomic composition. Bioinformatics *20*, 813–817.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA Websuite. Nucleic Acids Res. *36*, W70–W74.

Gu, W., Zhang, F., and Lupski, J.R. (2008). Mechanisms for human genomic rearrangements. Pathogenetics *1*, 4.

Guiblet, W.M., Cremona, M.A., Cechova, M., Harris, R.S., Kejnovská, I., Kejnovsky, E., Eckert, K., Chiaromonte, F., and Makova, K.D. (2018). Long-read sequencing technology indicates genome-wide effects of non-B DNA on polymerization speed and error rate. Genome Res. *28*, 1767–1778.

Gunther, K.L., Neitz, J., and Neitz, M. (2008). Nucleotide Polymorphisms Upstream of the X-chromosome Opsin Gene Array Tune L:M Cone Ratio. Vis. Neurosci. *25*, 265–271.

Hai, Y., Cao, W., Liu, G., Hong, S.-P., Elela, S.A., Klinck, R., Chu, J., and Xie, J. (2008). A G-tract element in apoptotic agents-induced alternative splicing. Nucleic Acids Res. *36*, 3320–3331.

Haber, J.E., Ira, G., Malkova, A., and Sugawara, N. (2004). Repairing a double-strand chromosome break by homologous recombination: revisiting Robin Holliday's model. Philos Trans R Soc Lond B Biol Sci *359*, 79–86.

Haim, M., Fledelius, H.C., Skarsholm. (1988) X-linked myopia in Danish family. Acta Ophthalmol. (Copenh). *66*, 450-456. Haim, M. (2002). Epidemiology of retinitis pigmentosa in Denmark. Acta Ophthalmol. Scand. Suppl. 1–34.

Hallast, P., Nagirnaja, L., Margus, T., and Laan, M. (2005). Segmental duplications and gene conversion: Human luteinizing hormone/chorionic gonadotropin beta gene cluster. Genome Res. *15*, 1535–1546.

Halldorsson, B.V., Hardarson, M.T., Kehr, B., Styrkarsdottir, U., Gylfason, A., Thorleifsson, G., Zink, F., Jonasdottir, A., Jonasdottir, A., Sulem, P., et al. (2016). The rate of meiotic gene conversion varies by sex and age. Nat. Genet. *48*, 1377–1384.

Haque, A., Buratti, E., and Baralle, F.E. (2010). Functional properties and evolutionary splicing constraints on a composite exonic regulatory element of splicing in CFTR exon 12. Nucleic Acids Res. *38*, 647–659.

Hartasánchez, D.A., Vallès-Codina, O., Brasó-Vives, M., and Navarro, A. (2014). Interplay of Interlocus Gene Conversion and Crossover in Segmental Duplications Under a Neutral Scenario. G3 GenesGenomesGenetics *4*, 1479–1489.

Hastings, P. J., Ira, G., and Lupski, J. R. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genetics, *5*, e1000327. He, F., Seryshev, A.B., Cowan, C.W., and Wensel, T.G. (2000). Multiple Zinc Binding Sites in Retinal Rod cGMP Phosphodiesterase, PDE6αβ. J. Biol. Chem. *275*, 20572–20577.

Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nat. Rev. Mol. Cell Biol. *18*, 637–650.

Hiller, M., Zhang, Z., Backofen, R., and Stamm, S. (2007). Pre-mRNA Secondary Structures Influence Exon Recognition. PLoS Genet. *3*.

Hiwatashi, T., Mikami, A., Katsumura, T., Suryobroto, B., Perwitasari-Farajallah, D., Malaivijitnond, S., Siriaroonrat, B., Oota, H., Goto, S., and Kawamura, S. (2011). Gene conversion and purifying selection shape nucleotide variation in gibbon L/M opsin genes. BMC Evol. Biol. *11*, 312.

Hofer, H., Carroll, J., Neitz, J., Neitz, M., and Williams, D.R. (2005). Organization of the Human Trichromatic Cone Mosaic. J. Neurosci. *25*, 9669–9679.

den Hollander, A.I., Koenekoop, R.K., Yzer, S., Lopez, I., Arends, M.L., Voesenek, K.E.J., Zonneveld, M.N., Strom, T.M., Meitinger, T., Brunner, H.G., et al. (2006). Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. Am. J. Hum. Genet. *79*, 556–561.

Hong, D.-H., and Li, T. (2002). Complex Expression Pattern of RPGR Reveals a Role for Purine-Rich Exonic Splicing Enhancers. Invest. Ophthalmol. Vis. Sci. *43*, 3373–3382.

Hoon, M., Okawa, H., Santina, L.D., and Wong, R.O.L. (2014). Functional Architecture of the Retina: Development and Disease. Prog. Retin. Eye Res. *42*, 44–84.

Hu, D.N. (1982). Genetic aspects of retinitis pigmentosa in China. Am. J. Med. Genet. *12*, 51–56.

Huang, S.H., Pittler, S.J., Huang, X., Oliveira, L., Berson, E.L., and Dryja, T.P. (1995). Autosomal recessive retinitis pigmentosa caused by mutations in the α subunit of rod cGMP phosphodiesterase. Nat. Genet. *11*, 468–471.

Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., et al. (2012). Integrative Genome-wide Analysis Reveals Cooperative Regulation of Alternative Splicing by hnRNP Proteins. Cell Rep. *1*, 167–178.

Hunt, D.M., Dulai, K.S., Cowing, J.A., Julliot, C., Mollon, J.D., Bowmaker, J.K., Li, W.-H., and Hewett-Emmett, D. (1998). Molecular evolution of trichromacy in primates. Vision Res. *38*, 3299–3306.

Hurles, M. (2005). How homologous recombination generates a mutable genome. Hum. Genomics *2*, 179.

Ingram, N.T., Sampath, A.P., and Fain, G.L. (2016). Why are rods more sensitive than cones? J. Physiol. *594*, 5415–5426.

Inoue, F., Kircher, M., Martin, B., Cooper, G.M., Witten, D.M., McManus, M.T., Ahituv, N., and Shendure, J. (2017). A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. Genome Res. *27*, 38–52.

Ira, G., Satory, D., and Haber, J.E. (2006). Conservative Inheritance of Newly Synthesized DNA in Double-Strand Break-Induced Gene Conversion. Mol Cell Biol *26*, 9424–9429.

Isalan, M. (2006). Construction of semi-randomized gene libraries with weighted oligonucleotide synthesis and PCR. Nat. Protoc. *1*, 468–475.

Ito, K., Patel, P.N., Gorham, J.M., McDonough, B., DePalma, S.R., Adler, E.E., Lam, L., MacRae, C.A., Mohiuddin, S.M., Fatkin, D., et al. (2017). Identification of pathogenic gene mutations in LMNA and MYBPC3 that alter RNA splicing. Proc. Natl. Acad. Sci. U. S. A. *114*, 7689–7694.

Itsara, A., Wu, H., Smith, J.D., Nickerson, D.A., Romieu, I., London, S.J., and Eichler, E.E. (2010). De novo rates and selection of large copy number variation. Genome Res. *20*, 1469–1481.

Jablonski, J.A., Buratti, E., Stuani, C., and Caputi, M. (2008). The Secondary Structure of the Human Immunodeficiency Virus Type 1 Transcript Modulates Viral Splicing and Infectivity. J. Virol. *82*, 8038–8050.

Jagla, W.M., Jägle, H., Hayashi, T., Sharpe, L.T., and Deeb, S.S. (2002). The molecular basis of dichromatic color vision in males with multiple red and green visual pigment genes. Hum. Mol. Genet. *11*, 23–32.

Jain, M., Koren, S., Miga, K.H., Quick, J., Rand, A.C., Sasani, T.A., Tyson, J.R., Beggs, A.D., Dilthey, A.T., Fiddes, I.T., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat. Biotechnol. *36*, 338–345.

Jay, M. (1982). On the heredity of retinitis pigmentosa. Br. J. Ophthalmol. *66*, 405–416.

Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K., and Krzyzosiak, W.J. (2016). Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases. Nucleic Acids Res. *44*, 9050–9070.

Jeffreys, A.J., and May, C.A. (2004). Intense and highly localized gene conversion activity in human meiotic crossover hot spots. Nat. Genet. *36*, 151–156.

Jobbins, A.M., Reichenbach, L.F., Lucas, C.M., Hudson, A.J., Burley, G.A., and Eperon, I.C. (2018). The mechanisms of a mammalian splicing enhancer. Nucleic Acids Res. *46*, 2145–2158.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 167.

Julien, P., Miñana, B., Baeza-Centurion, P., Valcárcel, J., and Lehner, B. (2016). The complete local genotype–phenotype landscape for the alternative splicing of a human exon. Nat. Commun. *7*, 11558.

Jurka, J., Kohany, O., Pavlicek, A., Kapitonov, V.V., and Jurka, M.V. (2004). Duplication, coclustering, and selection of human Alu retrotransposons. Proc. Natl. Acad. Sci. U. S. A. *101*, 1268–1272.

Kajiwara, K., Berson, E.L., and Dryja, T.P. (1994). Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. Science *264*, 1604–1608.

Karpen, J.W., Zimmerman, A.L., Stryer, L., Baylor, D.A.. (1988). Gating kinetics of the cyclic-GMP-activated channel of retinal rods: flash photolysis and voltage-jump studies. Proc. Natl. Acad. Sci. U. S. A. *85*, 1287-1291.

Kashima, T., Rao, N., David, C.J., and Manley, J.L. (2007). hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. Hum. Mol. Genet. *16*, 3149–3159.

Katagiri, S., Iwasa, M., Hayashi, T., Hosono, K., Yamashita, T., Kuniyoshi, K., Ueno, S., Kondo, M., Ueyama, H., Ogita, H., et al. (2018). Genotype determination of the OPN1LW / OPN1MW genes: novel disease-causing mechanisms in Japanese patients with blue cone monochromacy. Sci. Rep. *8*, 11507.

Kawamura, S., Hiramatsu, C., Melin, A.D., Schaffner, C.M., Aureli, F., and Fedigan, L.M. (2012). Polymorphic Color Vision in Primates: Evolutionary Considerations. In Post-Genome Biology of Primates, H. Hirai, H. Imai, and Y. Go, eds. (Tokyo: Springer Tokyo), pp. 93–120.

Ke, S., Shang, S., Kalachikov, S.M., Morozova, I., Yu, L., Russo, J.J., Ju, J., and Chasin, L.A. (2011). Quantitative evaluation of all hexamers as exonic splicing elements. Genome Res. *21*, 1360–1374.

Ke, S., Anquetil, V., Zamalloa, J.R., Maity, A., Yang, A., Arias, M.A., Kalachikov, S., Russo, J.J., Ju, J., and Chasin, L.A. (2018). Saturation mutagenesis reveals manifold determinants of exon definition. Genome Res. *28*, 11–24.

Keeler, C.E. (1924). The Inheritance of a Retinal Abnormality in White Mice. Proc. Natl. Acad. Sci. U. S. A. *10*, 329–333.

Kellner, U., and Foerster, M.H. (1992). Color electroretinography. A method for separation of dysfunctions of cones. Doc. Ophthalmol. Adv. Ophthalmol. *80*, 13–23.

Kellner, U., Wissinger, B., Tippmann, S., Kohl, S., Kraus, H., and Foerster, M.H. (2004). Blue cone monochromatism: clinical findings in patients with mutations in the red/green opsin gene cluster. Graefes Arch. Clin. Exp. Ophthalmol. Albrecht Von Graefes Arch. Klin. Exp. Ophthalmol. *242*, 729–735.

Kikin, O., D'Antonio, L., and Bagga, P.S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res. *34*, W676–W682.

Kim, S., Cho, C.-S., Han, K., and Lee, J. (2016). Structural Variation of Alu Element and Human Disease. Genomics Inform. *14*, 70–77.

Klein, J.C., Chen, W., Gasperini, M., and Shendure, J. (2018). Identifying Novel Enhancer Elements with CRISPR-Based Screens. ACS Chem. Biol. *13*, 326–332.

van der Klift, H., Wijnen, J., Wagner, A., Verkuilen, P., Tops, C., Otway, R., Kohonen-Corish, M., Vasen, H., Oliani, C., Barana, D., et al. (2005). Molecular characterization of the spectrum of genomic deletions in the mismatch repair genes MSH2, MLH1, MSH6, and PMS2 responsible for hereditary nonpolyposis colorectal cancer (HNPCC). Genes. Chromosomes Cancer *44*, 123–138.

Kolb, H. (1995a). Simple Anatomy of the Retina. In Webvision: The Organization of the Retina and Visual System, H. Kolb, E. Fernandez, and R. Nelson, eds. (Salt Lake City (UT): University of Utah Health Sciences Center), https://www.ncbi.nlm.nih.gov/pubmed/21413389

Kolb, H. (1995b). Facts and Figures Concerning the Human Retina. In Webvision: The Organization of the Retina and Visual System, H. Kolb, E. Fernandez, and R. Nelson, eds. (Salt Lake City (UT): University of Utah Health Sciences Center), https://www.ncbi.nlm.nih.gov/pubmed/21413389

Kolb, H. (1995c). Photoreceptors. In Webvision: The Organization of the Retina and Visual System, H. Kolb, E. Fernandez, and R. Nelson, eds. (Salt Lake City (UT): University of Utah Health Sciences Center), https://www.ncbi.nlm.nih.gov/pubmed/21413389

Kornblihtt, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M.J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat. Rev. Mol. Cell Biol. *14*, 153–165.

Krainer, A.R., and Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for Pre-mRNA splicing in vitro. Cell *42*, 725–736.

Kremer, L.S., Bader, D.M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., Haack, T.B., Graf, E., Schwarzmayr, T., Terrile, C., et al. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat. Commun. *8*, 15824.

Kuma, K., Hayashida, H., and Miyata, T. (1988). Recent gene conversion between genes encoding human red and green visual pigments. Idengaku Zasshi *63*, 367–371.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. *4*, 1073–1081.

Ladekjær-Mikkelsen, A.-S., Rosenberg, T., and Jørgensen, A.L. (1996). A new mechanism in blue cone monochromatism. Hum. Genet. *98*, 403–408.

Lai, M.C., Bechy, A.-L., Denk, F., Collins, E., Gavriliouk, M., Zaugg, J.B., Ryan, B.J., Wade-Martins, R., and Caffrey, T.M. (2017). Haplotype-specific MAPT exon 3 expression regulated by common intronic polymorphisms associated with Parkinsonian disorders. Mol. Neurodegener. *12*, 79.

Lee, M., Roos, P., Sharma, N., Atalar, M., Evans, T.A., Pellicore, M.J., Davis, E., Lam, A.-T.N., Stanley, S.E., Khalil, S.E., et al. (2017). Systematic Computational Identification of Variants That Activate Exonic and Intronic Cryptic Splice Sites. Am. J. Hum. Genet. *100*, 751–765.

Li, P., Bellows, A.B., and Thompson, J.N. (1999). Molecular basis of iduronate-2-sulphatase gene mutations in patients with mucopolysaccharidosis type II (Hunter syndrome). J. Med. Genet. *36*, 21–27.

Liu, G.E., Alkan, C., Jiang, L., Zhao, S., and Eichler, E.E. (2009). Comparative analysis of Alu repeats in primate genomes. Genome Res. *19*, 876.

Liu, H.-X., Zhang, M., and Krainer, A.R. (1998). Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev. *12*, 1998–2012.

López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? FEBS Lett. *579*, 1900–1903.

Lorson, C.L., Hahnen, E., Androphy, E.J., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc. Natl. Acad. Sci. *96*, 6307–6311.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nat. Rev. Mol. Cell Biol. *8*, 479–490.

Majithia, A.R., Tsuda, B., Agostini, M., Gnanapradeepan, K., Rice, R., Peloso, G., Patel, K.A., Zhang, X., Broekema, M.F., Patterson, N., et al. (2016). Prospective functional classification of all possible missense variants in PPARG. Nat. Genet. *48*, 1570–1575.

Mansai, S.P., Kado, T., and Innan, H. (2011). The Rate and Tract Length of Gene Conversion between Duplicated Genes. Genes (Basel) *2*, 313–331.

Marín-Béjar, O., and Huarte, M. (2015). RNA Pulldown Protocol for In Vitro Detection and Identification of RNA-Associated Proteins. In Regulatory Non-Coding RNAs, (Humana Press, New York, NY), pp. 87–95.

Martínez-Gimeno, M., Gamundi, M.J., Hernan, I., Maseras, M., Millá, E., Ayuso, C., García-Sandoval, B., Beneyto, M., Vilela, C., Baiget, M., et al. (2003). Mutations in the pre-mRNA splicing-factor genes PRPF3, PRPF8, and PRPF31 in Spanish families with autosomal dominant retinitis pigmentosa. Invest. Ophthalmol. Vis. Sci. *44*, 2171–2177.

Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. *15*, 108.

Matunis, M.J., Xing, J., and Dreyfuss, G. (1994). The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization. Nucleic Acids Res. *22*, 1059–1067.

Mauger, D.M., Lin, C., and Garcia-Blanco, M.A. (2008). hnRNP H and hnRNP F Complex with Fox2 To Silence Fibroblast Growth Factor Receptor 2 Exon IIIc. Mol. Cell. Biol. *28*, 5403–5419.

Mayer, A.K., Rohrschneider, K., Strom, T.M., Glöckle, N., Kohl, S., Wissinger, B., and Weisschuh, N. (2016). Homozygosity mapping and whole-genome sequencing reveals a deep intronic *PROM1* mutation causing cone–rod dystrophy by pseudoexon activation. Eur. J. Hum. Genet. *24*, 459.

Mayer, A.K., Van Cauwenbergh, C., Rother, C., Baumann, B., Reuter, P., De Baere, E., Wissinger, B., Kohl, S., and ACHM Study Group (2017). CNGB3 mutation spectrum including copy number variations in 552 achromatopsia patients. Hum. Mutat. *38*, 1579–1591.

Mazoyer, S., Puget, N., Perrin-Vidoz, L., Lynch, H.T., Serova-Sinilnikova, O.M., and Lenoir, G.M. (1998). A BRCA1 nonsense mutation causes exon skipping. Am. J. Hum. Genet. *62*, 713–715.

McClements, M., Davies, W.I.L., Michaelides, M., Young, T., Neitz, M., MacLaren, R.E., Moore, A.T., and Hunt, D.M. (2013). Variations in Opsin Coding Sequences Cause X-Linked Cone Dysfunction Syndrome with Myopia and Dichromacy. Invest. Ophthalmol. Vis. Sci. *54*, 1361–1369.

McCullough, A.J., and Berget, S.M. (1997). G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol. Cell. Biol. *17*, 4562–4571.

McCullough, A.J., and Berget, S.M. (2000). An Intronic Splicing Enhancer Binds U1 snRNPs To Enhance Splicing and Select 5′ Splice Sites. Mol. Cell. Biol. *20*, 9225–9235.
McLaughlin, M.E., Sandberg, M.A., Berson, E.L., and Dryja, T.P. (1993). Recessive mutations in the gene encoding the β–subunit of rod phosphodiesterase in patients with retinitis pigmentosa. Nat. Genet. *4*, 130–134.
McLaughlin, M.E., Ehrhart, T.L., Berson, E.L., and Dryja, T.P. (1995). Mutation spectrum of the gene encoding the beta subunit of rod phosphodiesterase among patients with autosomal recessive retinitis pigmentosa. Proc. Natl. Acad. Sci. U. S. A. *92*, 3249–3253.
McMahill, M.S., Sham, C.W., and Bishop, D.K. (2007). Synthesis-Dependent Strand Annealing in Meiosis. PLOS Biology *5*, e299.
McMahon, C., Neitz, J., and Neitz, M. (2004). Evaluating the human X-chromosome pigment gene promoter sequences as predictors of L:M cone ratio variation. J. Vis. *4*, 7–7.
Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Jr, C.G.C., Kinney, J.B., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. *30*, 271–277.
Merbs, S.L., and Nathans, J. (1992). Absorption spectra of the hybrid pigments responsible for anomalous color vision. Science *258*, 464–466.
Merker, J.D., Wenger, A.M., Sneddon, T., Grove, M., Zappala, Z., Fresard, L., Waggott, D., Utiramerur, S., Hou, Y., Smith, K.S., et al. (2018). Long-read genome sequencing identifies causal structural variation in a Mendelian disease. Genet. Med. *20*, 159.
Michaelides, M., Hunt, D.M., Moore, A.T. (2004). The cone dysfunction syndromes. Br. J. Ophthalmol. *88*, 291–7.
Michaelides, M., Johnson, S., Simunovic, M.P., Bradshaw, K., Holder, G., Mollon, J.D., Moore, A.T., and Hunt, D.M. (2005). Blue cone monochromatism: a phenotype and genotype assessment with evidence of progressive loss of cone function in older individuals. Eye *19*, 2–10.
Mighell, T.L., Evans-Dutson, S., and O'Roak, B.J. (2018). A Saturation Mutagenesis Approach to Understanding PTEN Lipid Phosphatase Activity and Genotype-Phenotype Relationships. Am. J. Hum. Genet. *102*, 943–955.
Min, H., Chan, R.C., and Black, D.L. (1995). The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. Genes Dev. *9*, 2659–2671.
Mizrahi-Meissonnier, L., Merin, S., Banin, E., and Sharon, D. (2010). Variable Retinal Phenotypes Caused by Mutations in the X-Linked Photopigment Gene Array. Invest. Ophthalmol. Vis. Sci. *51*, 3884–3892.
Mollon, J.D. (1989). "Tho'' she kneel'd in that place where they grew..." The uses and origins of primate colour vision." J. Exp. Biol. *146*, 21–38.
Monani, U.R., Lorson, C.L., Parsons, D.W., Prior, T.W., Androphy, E.J., Burghes, A.H.M., and McPherson, J.D. (1999). A Single Nucleotide Difference That Alters Splicing Patterns Distinguishes the SMA Gene SMN1 From the Copy Gene SMN2. Hum Mol Genet *8*, 1177–1183.
Morales, M.E., White, T.B., Streva, V.A., DeFreece, C.B., Hedges, D.J., and Deininger, P.L. (2015). The Contribution of Alu Elements to Mutagenic DNA Double-Strand Break Repair. PLoS Genet. *11*.
Mort, M., Sterne-Weiler, T., Li, B., Ball, E.V., Cooper, D.N., Radivojac, P., Sanford, J.R., and Mooney, S.D. (2014). MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. Genome Biol. *15*, R19.
Moss, W.N. (2017). RNA2DMut: A web tool for the design and analysis of RNA structure mutations. RNA rna.063933.117.
Mueller, W.F., Larsen, L.S.Z., Garibaldi, A., Hatfield, G.W., and Hertel, K.J. (2015). The Silent Sway of Splicing by Synonymous Substitutions. J. Biol. Chem. *290*, 27700–27711.
Muradov, H., Boyd, K.K., Kerov, V., and Artemyev, N.O. (2012). Atypical retinal degeneration 3 in mice is caused by defective PDE6B pre-mRNA splicing. Vision Res. *57*, 1–8.
Muradov, K.G., Boyd, K.K., Martinez, S.E., Beavo, J.A., and Artemyev, N.O. (2003). The GAFa Domains of Rod cGMP-phosphodiesterase 6 Determine the Selectivity of the Enzyme Dimerization. J. Biol. Chem. *278*, 10594–10601.
Murphy, D., Singh, R., Kolandaivelu, S., Ramamurthy, V., and Stoilov, P. (2015). Alternative Splicing Shapes the Phenotype of a Mutation in BBS8 To Cause Nonsyndromic Retinitis Pigmentosa. Mol. Cell. Biol. *35*, 1860–1870.
Murphy, D., Cieply, B., Carstens, R., Ramamurthy, V., and Stoilov, P. (2016). The Musashi 1 Controls the Splicing of Photoreceptor-Specific Exons in the Vertebrate Retina. PLOS Genet. *12*, e1006256.
Nagy, A.L., and Boynton, R.M. (1979). Large-field color naming of dichromats with rods bleached. J. Opt. Soc. Am. *69*, 1259–1265.
Nathans, J. (1987). Molecular biology of visual pigments. Annu. Rev. Neurosci. *10*, 163–194.
Nathans, J., Thomas, D., and Hogness, D.S. (1986). Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science *232*, 193–202.
Nathans, J., Davenport, C.M., Maumenee, I.H., Lewis, R.A., Hejtmancik, J.F., Litt, M., Lovrien, E., Weleber, R., Bachynski, B., Zwas, F., et al. (1989). Molecular genetics of human blue cone monochromacy. Science *245*, 831–838.
Nathans, J., Maumenee, I.H., Zrenner, E., Sadowski, B., Sharpe, L.T., Lewis, R.A., Hansen, E., Rosenberg, T., Schwartz, M., Heckenlively, J.R., et al. (1993). Genetic heterogeneity among blue-cone monochromats. Am. J. Hum. Genet. *53*, 987–1000.
Nazim, M., Masuda, A., Rahman, M.A., Nasrin, F., Takeda, J., Ohe, K., Ohkawara, B., Ito, M., and Ohno, K. (2017). Competitive regulation of alternative splicing and alternative polyadenylation by hnRNP H and CstF64 determines acetylcholinesterase isoforms. Nucleic Acids Res. *45*, 1455–1468.
Neidhardt, J., Glaus, E., Barthelmes, D., Zeitz, C., Fleischhauer, J., and Berger, W. (2007). Identification and characterization of a novel RPGR isoform in human retina. Hum. Mutat. *28*, 797–807.
Nielsen, K.B., Sørensen, S., Cartegni, L., Corydon, T.J., Doktor, T.K., Schroeder, L.D., Reinert, L.S., Elpeleg, O., Krainer, A.R., Gregersen, N., et al. (2007). Seemingly Neutral Polymorphic Variants May Confer Immunity to Splicing-Inactivating Mutations: A Synonymous SNP in Exon 5 of MCAD Protects from Deleterious Mutations in a Flanking Exonic Splicing Enhancer. Am. J. Hum. Genet. *80*, 416–432.
Ning, G., Cheng, X., Luo, P., Liang, F., Wang, Z., Yu, G., Li, X., Wang, D., and Bao, M. (2017). Hybrid sequencing and map finding (HySeMaFi): optional strategies for extensively deciphering gene splicing and expression in organisms without reference genome. Sci. Rep. *7*, 43793.
Nuttle, X., Huddleston, J., O'Roak, B.J., Antonacci, F., Fichera, M., Romano, C., Shendure, J., and Eichler, E.E. (2013). Rapid and accurate large-scale genotyping of duplicated genes and discovery of interlocus gene conversions. Nat. Methods *10*, 903–909.
Nykamp, K., Anderson, M., Powers, M., Garcia, J., Herrera, B., Ho, Y.-Y., Kobayashi, Y., Patil, N., Thusberg, J., Westbrook, M., et al. (2017). Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. Genet. Med. *19*, 1105–1117.

Odenthal-Hesse, L., Berg, I.L., Veselis, A., Jeffreys, A.J., and May, C.A. (2014). Transmission Distortion Affecting Human Noncrossover but Not Crossover Recombination: A Hidden Source of Meiotic Drive. PLoS Genet. *10*.

Ohno, K., Takeda, J., and Masuda, A. (2018). Rules and tools to predict the splicing effects of exonic and intronic mutations. Wiley Interdiscip. Rev. RNA *9*, e1451 10.1002/wrna.1451.

Orosz, O., Rajta, I., Vajas, A., Takács, L., Csutak, A., Fodor, M., Kolozsvári, B., Resch, M., Sényi, K., Lesch, B., et al. (2017). Myopia and Late-Onset Progressive Cone Dystrophy Associate to LVAVA/MVAVA Exon 3 Interchange Haplotypes of Opsin Genes on Chromosome XRare Opsin Haplotypes Cause Late-Onset Cone Dystrophy. Invest. Ophthalmol. Vis. Sci. *58*, 1834–1842.

Osterberg, G (1937). Topography of the Layer of Rods and Cones in the Human Retina. J. Am. Med. Assoc. *108*, 232–232.

Ottaviani, D., LeCain, M., and Sheer, D. (2014). The role of microhomology in genomic structural variation. Trends Genet. *30*, 85–94.

Pagani, F., Stuani, C., Tzetis, M., Kanavakis, E., Efthymiadou, A., Doudounakis, S., Casals, T., and Baralle, F.E. (2003a). New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum. Mol. Genet. *12*, 1111–1120.

Pagani, F., Buratti, E., Stuani, C., and Baralle, F.E. (2003b). Missense, Nonsense, and Neutral Mutations Define Juxtaposed Regulatory Elements of Splicing in Cystic Fibrosis Transmembrane Regulator Exon 9. J. Biol. Chem. *278*, 26580–26588.

Pagani, F., Raponi, M., and Baralle, F.E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc. Natl. Acad. Sci. U. S. A. *102*, 6368–6372.

Papasaikas, P., and Valcárcel, J. (2016). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. Trends Biochem. Sci. *41*, 33–45.

Parfitt, D.A., Lane, A., Ramsden, C.M., Carr, A.-J.F., Munro, P.M., Jovanovic, K., Schwarz, N., Kanuga, N., Muthiah, M.N., Hull, S., et al. (2016). Identification and Correction of Mechanisms Underlying Inherited Blindness in Human iPSC-Derived Optic Cups. Cell Stem Cell *18*, 769–781.

Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The Expanding Landscape of Alternative Splicing Variation in Human Populations. Am. J. Hum. Genet. *102*, 11–26.

Parmley, J.L., and Hurst, L.D. (2007). Exonic Splicing Regulatory Elements Skew Synonymous Codon Usage near Intron-exon Boundaries in Mammals. Mol. Biol. Evol. *24*, 1600–1603.

Patterson, E.J., Wilk, M., Langlo, C.S., Kasilian, M., Ring, M., Hufnagel, R.B., Dubis, A.M., Tee, J.J., Kalitzeos, A., Gardner, J.C., et al. (2016). Cone Photoreceptor Structure in Patients With X-Linked Cone Dysfunction and Red-Green Color Vision Deficiency. Invest. Ophthalmol. Vis. Sci. *57*, 3853–3863.

Patterson, E.J., Kalitzeos, A., Kasilian, M., Gardner, J.C., Neitz, J., Hardcastle, A.J., Neitz, M., Carroll, J., and Michaelides, M. (2018). Residual Cone Structure in Patients With X-Linked Cone Opsin Mutations. Invest. Ophthalmol. Vis. Sci. *59*, 4238–4248.

Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.-I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. *30*, 265–270.

Peng, G.-H., and Chen, S. (2011). Active opsin loci adopt intrachromosomal loops that depend on the photoreceptor transcription factor network. Proc. Natl. Acad. Sci. *108*, 17821–17826.

Pezeshkpoor, B., Rost, S., Oldenburg, J., and El-Maarri, O. (2012). Identification of a third rearrangement at Xq28 that causes severe hemophilia A as a result of homologous recombination between inverted repeats. J. Thromb. Haemost. *10*, 1600–1608.

Piñol-Roma, S., and Dreyfuss, G. (1993). hnRNP proteins:Localization and transport between the nucleus and the cytoplasm. Trends Cell Biol. *3*, 151–155.

Piva, F., Giulietti, M., Burini, A.B., and Principato, G. (2012). SpliceAid 2: A database of human splicing factors expression data and RNA target motifs. Human Mutation *33*, 81–85.

Qu, H.-Q., Lu, Y., Marchand, L., Bacot, F., Fréchette, R., Tessier, M.-C., Montpetit, A., and Polychronakos, C. (2007). Genetic control of alternative splicing in the TAP2 gene: possible implication in the genetics of type 1 diabetes. Diabetes *56*, 270–275.

Rahbari, R., Wuster, A., Lindsay, S.J., Hardwick, R.J., Alexandrov, L.B., Turki, S.A., Dominiczak, A., Morris, A., Porteous, D., Smith, B., et al. (2016). Timing, rates and spectra of human germline mutation. Nat. Genet. *48*, 126.

Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., Consortium, U.B.E., Hardy, J., Ryten, M., et al. (2014). Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat. Neurosci. *17*, 1418–1428.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177.

Redin, D., Borgström, E., He, M., Aghelpasand, H., Käller, M., and Ahmadian, A. (2017). Droplet Barcode Sequencing for targeted linked-read haplotyping of single DNA molecules. Nucleic Acids Res. *45*, e125–e125.

Reed, R., and Maniatis, T. (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. Cell *41*, 95–105.

Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. J. Comput. Biol. J. Comput. Mol. Cell Biol. *4*, 311–323.

Reiter, L.T., Hastings, P.J., Nelis, E., Jonghe, P.D., Broeckhoven, C.V., and Lupski, J.R. (1998). Human Meiotic Recombination Products Revealed by Sequencing a Hotspot for Homologous Strand Exchange in Multiple HNPP Deletion Patients. Am. J. Hum. Genet. *62*, 1023–1033.

Remenyi, J., Bajan, S., Fuller-Pace, F.V., Arthur, J.S.C., and Hutvagner, G. (2016). The loop structure and the RNA helicase p72/DDX17 influence the processing efficiency of the mice miR-132. Sci. Rep. *6*.

Reyniers, E., Vanthienen, M.-N., Meire, F., Deboulle, K., Devries, K., Kestelijn, P., and Willems, P.J. (1995). Gene Conversion between Red and Defective Green Opsin Gene in Blue Cone Monochromacy. Genomics *29*, 323–328.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). 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.

Rimoldi, V., Soldà, G., Asselta, R., Spena, S., Stuani, C., Buratti, E., and Duga, S. (2013). Dual Role of G-runs and hnRNP F in the Regulation of a Mutation-Activated Pseudoexon in the Fibrinogen Gamma-Chain Transcript. PLOS ONE *8*, e59333.

Rivera-Barahona, A., Sánchez-Alcudia, R., Viecelli, H.M., Rüfenacht, V., Pérez, B., Ugarte, M., Häberle, J., Thöny, B., and Desviat, L.R. (2015). Functional Characterization of the spf/ash Splicing Variation in OTC Deficiency of Mice and Man. PLOS ONE *10*, e0122966.

Rivolta, C., Berson, E.L., and Dryja, T.P. (2002). Paternal Uniparental Heterodisomy With Partial Isodisomy of Chromosome 1 in a Patient With Retinitis Pigmentosa Without Hearing Loss and a Missense Mutation in the Usher Syndrome Type II Gene USH2A. Arch. Ophthalmol. *120*, 1566–1571.

Robinson, D.O., Bunyan, D.J., Gabb, H.A., Temple, I.K., and Yau, S.C. (1997). A small intraexonic deletion within the dystrophin gene suggests a possible mechanism of mutagenesis. Hum. Genet. *99*, 658–662.

Roca, X., Krainer, A.R., and Eperon, I.C. (2013). Pick one, but be quick: 5' splice sites and the problems of too many choices. Genes Dev. *27*, 129–144.

Rogalska, M.E., Tajnik, M., Licastro, D., Bussani, E., Camparini, L., Mattioli, C., and Pagani, F. (2016). Therapeutic activity of modified U1 core spliceosomal particles. Nat. Commun. *7*, 11168.

Rose, A.M., Mukhopadhyay, R., Webster, A.R., Bhattacharya, S.S., and Waseem, N.H. (2011). A 112 kb Deletion in Chromosome 19q13.42 Leads to Retinitis Pigmentosa. Invest. Ophthalmol. Vis. Sci. *52*, 6597–6603.

Rosenberg, A.B., Patwardhan, R.P., Shendure, J., and Seelig, G. (2015). Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences. Cell *163*, 698–711.

Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P., et al. (2005). The DNA sequence of the human X chromosome. Nature *434*, 325–337.

Rossetti, L.C., Goodeve, A., Larripa, I.B., and De Brasi, C.D. (2004). Homeologous recombination between AluSx-sequences as a cause of hemophilia. Hum. Mutat. *24*, 440.

Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell *38*, 317–331.

Ruskin, B., Zamore, P.D., and Green, M.R. (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell *52*, 207–219.

Russo, A., Siciliano, G., Catillo, M., Giangrande, C., Amoresano, A., Pucci, P., Pietropaolo, C., and Russo, G. (2010). hnRNP H1 and intronic G runs in the splicing control of the human rpL3 gene. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1799*, 419–428.

Sabesan, R., Schmidt, B.P., Tuten, W.S., and Roorda, A. (2016). The elementary representation of spatial and color vision in the human retina. Sci. Adv. *2*, e1600797.

Sakamoto, K., McCluskey, M., Wensel, T.G., Naggert, J.K., and Nishina, P.M. (2009). New mouse models for recessive retinitis pigmentosa caused by mutations in the Pde6a gene. Hum. Mol. Genet. *18*, 178–192.

Sangermano, R., Bax, N.M., Bauwens, M., van den Born, L.I., De Baere, E., Garanto, A., Collin, R.W.J., Goercharn-Ramlal, A.S.A., den Engelsman-van Dijk, A.H.A., Rohrschneider, K., et al. (2016). Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T→C Mutation in Stargardt Disease. Ophthalmology *123*, 1375–1385.

Sangermano, R., Garanto, A., Khan, M., Runhart, E.H., Bauwens, M., Bax, N.M., Born, L.I. van den, Khan, M.I., Cornelis, S.S., Verheij, J.B.G.M., et al. (2019). Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. Genet. Med. 1.

Sanz, D.J., Acedo, A., Infante, M., Durán, M., Pérez-Cabornero, L., Esteban-Cardeñosa, E., Lastra, E., Pagani, F., Miner, C., and Velasco, E.A. (2010). A High Proportion of DNA Variants of BRCA1 and BRCA2 Is Associated with Aberrant Splicing in Breast/Ovarian Cancer Patients. Clin. Cancer Res. *16*, 1957–1967.

Savisaar, R., and Hurst, L.D. (2017). Estimating the prevalence of functional exonic splice regulatory information. Hum. Genet. 1–20.

Savisaar, R., and Hurst, L.D. (2018). Exonic splice regulation imposes strong selection at synonymous sites. Genome Res. *28*, 1442–1454.

Schaub, M.C., Lopez, S.R., and Caputi, M. (2007). Members of the Heterogeneous Nuclear Ribonucleoprotein H Family Activate Splicing of an HIV-1 Splicing Substrate by Promoting Formation of ATP-dependent Spliceosomal Complexes. J. Biol. Chem. *282*, 13617–13626.

Schultze, M. (1866). Zur Anatomie und Physiologie der Retina. Arch. Für Mikrosk. Anat. *2*, 175–286.

Schulz, H.L., Grassmann, F., Kellner, U., Spital, G., Rüther, K., Jägle, H., Hufendiek, K., Rating, P., Huchzermeyer, C., Baier, M.J., et al. (2017). Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients From a Multicenter German Cohort—Impact of Selected Deep Intronic Variants and Common SNPsMutation Spectrum of ABCA4 in STGD1 Patients From Germany. Invest. Ophthalmol. Vis. Sci. *58*, 394–403.

Schwartz, S., Hall, E., and Ast, G. (2009). SROOGLE: webserver for integrative, user-friendly visualization of splicing signals. Nucleic Acids Res. *37*, W189-192.

Schwarz, J.M., Rödelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nat. Methods *7*, 575–576.

Scotti, M.M., Swanson, M.S. (2015). RNA mis-splicing in disease. Nat. Rev. Genet. *17*, 19-32.

Seyedahmadi, B.J., Rivolta, C., Keene, J.A., Berson, E.L., and Dryja, T.P. (2004). Comprehensive screening of the USH2A gene in Usher syndrome type II and non-syndromic recessive retinitis pigmentosa. Exp. Eye Res. *79*, 167–173.

Shamoo, Y., Abdul-Manan, N., and Williams, K.R. (1995). Multiple RNA binding domains (RBDs) just don't add up. Nucleic Acids Res. *23*, 725–728.

Sharma, N., Sosnay, P.R., Ramalho, A.S., Douville, C., Franca, A., Gottschalk, L.B., Park, J., Lee, M., Vecchio-Pagan, B., Raraigh, K.S., et al. (2014). Experimental Assessment of Splicing Variants Using Expression Minigenes and Comparison with In Silico Predictions. Hum. Mutat. *35*, 1249.

Sharpe, L.T., Stockman, A., Jaegle, H., and Nathans, J. (1999). 1 Opsin genes , cone photopigments , color vision , and color blindness.

Shen, S.Q., Myers, C.A., Hughes, A.E.O., Byrne, L.C., Flannery, J.G., and Corbo, J.C. (2016). Massively parallel cis-regulatory analysis in the mammalian central nervous system. Genome Res. *26*, 238–255.

Shi, W., Massaia, A., Louzada, S., Banerjee, R., Hallast, P., Chen, Y., Bergström, A., Gu, Y., Leonard, S., Quail, M.A., et al. (2018). Copy number variation arising from gene conversion on the human Y chromosome. Hum. Genet. *137*, 73–83.

Shyue, S.K., Li, L., Chang, B.H., and Li, W.H. (1994). Intronic gene conversion in the evolution of human X-linked color vision genes. Mol. Biol. Evol. *11*, 548–551.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Singh, R.K., and Cooper, T.A. (2012). Pre-mRNA splicing in disease and therapeutics. Trends Mol. Med. *18*, 472–482.

Singh, N.N., Lee, B.M., and Singh, R.N. (2015). Splicing regulation in spinal muscular atrophy by an RNA structure formed by long-distance interactions. Ann. N. Y. Acad. Sci. *1341*, 176–187.

Sleight, S.C., Bartley, B.A., Lieviant, J.A., and Sauro, H.M. (2010). In-Fusion BioBrick assembly and re-engineering. Nucleic Acids Res. *38*, 2624–2636.
Smallwood, P.M., Wang, Y., and Nathans, J. (2002). Role of a locus control region in the mutually exclusive expression of human red and green cone pigment genes. Proc. Natl. Acad. Sci. *99*, 1008–1011.
Smith, V.C., Pokorny, J., Delleman, J.W., Cozijnsen, M., Houtman, W.A., and Went, L.N. (1983). X-linked incomplete achromatopsia with more than one class of functional cones. Invest. Ophthalmol. Vis. Sci. *24*, 451–457.
Soemedi, R., Cygan, K.J., Rhine, C.L., Wang, J., Bulacan, C., Yang, J., Bayrak-Toydemir, P., McDonald, J., and Fairbrother, W.G. (2017). Pathogenic variants that alter protein code often disrupt splicing. Nat. Genet. *49*, 848–855.
Soens, Z.T., Branch, J., Wu, S., Yuan, Z., Li, Y., Li, H., Wang, K., Xu, M., Rajan, L., Motta, F.L., et al. (2017). Leveraging splice-affecting variant predictors and a minigene validation system to identify Mendelian disease-causing variants among exon-captured variants of uncertain significance. Hum. Mutat. *38*, 1521–1533.
Sohail, M., Cao, W., Mahmood, N., Myschyshyn, M., Hong, S.P., and Xie, J. (2014). Evolutionarily emerged G tracts between the polypyrimidine tract and 3′ AG are splicing silencers enriched in genes involved in cancer. BMC Genomics *15*, 1143.
Sothilingam, V., Garrido, M.G., Jiao, K., Buena-Atienza, E., Sahaboglu, A., Trifunović, D., Balendran, S., Koepfli, T., Mühlfriedel, R., Schön, C., et al. (2015). Retinitis pigmentosa: impact of different Pde6a point mutations on the disease phenotype. Hum. Mol. Genet. *24*, 5486–5499.
Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frébourg, T., Tosi, M., and Martins, A. (2016). Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. PLOS Genet. *12*, e1005756.
Starita, L.M., and Fields, S. (2015). Deep Mutational Scanning: A Highly Parallel Method to Measure the Effects of Mutation on Protein Function. Cold Spring Harb. Protoc. *2015*, 711–714.
Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J., and Fields, S. (2015). Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. Genetics *200*, 413–422.
Sterne-Weiler, T., Howard, J., Mort, M., Cooper, D.N., and Sanford, J.R. (2011). Loss of exon identity is a common mechanism of human inherited disease. Genome Res. *21*, 1563–1571.
Stratton, M.R., and Rahman, N. (2008). The emerging landscape of breast cancer susceptibility. Nat. Genet. *40*, 17–22.
Strein, C., Alleaume, A.-M., Rothbauer, U., Hentze, M.W., and Castello, A. (2014). A versatile assay for RNA-binding proteins in living cells. RNA *20*, 721.
Stryer, L. (1986). Cyclic GMP cascade of vision. Annu. Rev. Neurosci. *9*, 87-119.
Tai, D.J.C., Ragavendran, A., Manavalan, P., Stortchevoi, A., Seabra, C.M., Erdin, S., Collins, R.L., Blumenthal, I., Chen, X., Shen, Y., et al. (2016). Engineering microdeletions and microduplications by targeting segmental duplications with CRISPR. Nat. Neurosci. *19*, 517.
Taliaferro, J.M., Lambert, N.J., Sudmant, P.H., Dominguez, D., Merkin, J.J., Alexis, M.S., Bazile, C., and Burge, C.B. (2016). RNA sequence context effects measured in vitro predict in vivo protein binding and regulation. Mol. Cell *64*, 294–306.
Talukdar, I., Sen, S., Urbano, R., Thompson, J., Iii, J.R.Y., and Webster, N.J.G. (2011). hnRNP A1 and hnRNP F Modulate the Alternative Splicing of Exon 11 of the Insulin Receptor Gene. PLOS ONE *6*, e27869.
Tammaro, C., Raponi, M., Wilson, D.I., and Baralle, D. (2014). BRCA1 Exon 11, a CERES (Composite Regulatory Element of Splicing) Element Involved in Splice Regulation. Int. J. Mol. Sci. *15*, 13045–13059.
Tayebi, N., Stubblefield, B.K., Park, J.K., Orvisky, E., Walker, J.M., LaMarca, M.E., and Sidransky, E. (2003). Reciprocal and Nonreciprocal Recombination at the Glucocerebrosidase Gene Region: Implications for Complexity in Gaucher Disease. Am. J. Hum. Genet. *72*, 519–534.
Terao, K., Mikami, A., Saito, A., Itoh, S., Ogawa, H., Takenaka, O., Sakai, T., Onishi, A., Teramoto, M., Udono, T., et al. (2005). Identification of a protanomalous chimpanzee by molecular genetic and electroretinogram analyses. Vision Res. *45*, 1225–1235.
Tessereau, C., Léoné, M., Buisson, M., Duret, L., Sinilnikova, O.M., and Mazoyer, S. (2015). Occurrence of a non deleterious gene conversion event in the BRCA1 gene. Genes. Chromosomes Cancer *54*, 646–652.
Tewhey, R., Kotliar, D., Park, D.S., Liu, B., Winnicki, S., Reilly, S.K., Andersen, K.G., Mikkelsen, T.S., Lander, E.S., Schaffner, S.F., et al. (2016). Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. Cell *165*, 1519–1529.
The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.
Tilgner, H., Grubert, F., Sharon, D., and Snyder, M.P. (2014). Defining a personal, allele-specific, and single-molecule long-read transcriptome. Proc. Natl. Acad. Sci. *111*, 9869–9874.
Todorov, T., Todorova, A., Motoescu, C., Dimova, P., Iancu, D., Craiu, D., Stoian, D., Barbarii, L., Bojinova, V., and Mitev, V. (2012). Spontaneous recurrent mutations and a complex rearrangement in the MECP2 gene in the light of current models of mutagenesis. Mutat. Res. Mol. Mech. Mutagen. *734*, 69–72.
Travers, K.J., Chin, C.-S., Rank, D.R., Eid, J.S., and Turner, S.W. (2010). A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. *38*, e159.
Trombetta, B., Fantini, G., D'Atanasio, E., Sellitto, D., and Cruciani, F. (2016). Evidence of extensive non-allelic gene conversion among LTR elements in the human genome. Sci Rep *6*, 28710.
Truty, R., Paul, J., Kennemer, M., Lincoln, S.E., Olivares, E., Nussbaum, R.L., and Aradhya, S. (2019). Prevalence and properties of intragenic copy-number variation in Mendelian disease genes. Genet. Med. *21*, 114.
Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science *249*, 505–510.
Turunen, J.J., Niemelä, E.H., Verma, B., and Frilander, M.J. (2013). The significant other: splicing by the minor spliceosome. Wiley Interdiscip. Rev. RNA *4*, 61–76.
Tyanova, S., and Cox, J. (2018). Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. In Cancer Systems Biology: Methods and Protocols, L. von Stechow, ed. (New York, NY: Springer New York), pp. 133–148.
Tyanova, S., Temu, T., and Cox, J. (2016a). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319.
Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016b). The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods *13*, 731–740.
Ueyama, H., Muraki-Oda, S., Yamade, S., Tanabe, S., Yamashita, T., Shichida, Y., and Ogita, H. (2012). Unique haplotype in exon 3 of cone opsin mRNA affects splicing of its precursor, leading to congenital color vision defect. Biochem. Biophys. Res. Commun. *424*, 152–157.
Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212–1215.

Van Cauwenbergh, C., Van Schil, K., Cannoodt, R., Bauwens, M., Van Laethem, T., De Jaegere, S., Steyaert, W., Sante, T., Menten, B., Leroy, B.P., et al. (2017). arrEYE: a customized platform for high-resolution copy number analysis of coding and noncoding regions of known and candidate retinal dystrophy genes and retinal noncoding RNAs. Genet. Med. Off. J. Am. Coll. Med. Genet. *19*, 457–466.

Van Schil, K., Naessens, S., Van de Sompele, S., Carron, M., Aslanidis, A., Van Cauwenbergh, C., Mayer, A.K., Van Heetvelde, M., Bauwens, M., Verdin, H., et al. (2018). Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations. Genet. Med. *20*, 202–213.

Veltman, J.A., and Brunner, H.G. (2012). *De novo* mutations in human genetic disease. Nat. Rev. Genet. *13*, 565–575.

Verbakel, S.K., van Huet, R.A.C., Boon, C.J.F., den Hollander, A.I., Collin, R.W.J., Klaver, C.C.W., Hoyng, C.B., Roepman, R., and Klevering, B.J. (2018). Non-syndromic retinitis pigmentosa. Prog. Retin. Eye Res. *66*, 157–186.

Verrelli, B.C., and Tishkoff, S.A. (2004). Signatures of Selection and Gene Conversion Associated with Human Color Vision Variation. Am. J. Hum. Genet. *75*, 363–375.

Verrelli, B.C., Lewis, C.M., Stone, A.C., and Perry, G.H. (2008). Different Selective Pressures Shape the Molecular Evolution of Color Vision in Chimpanzee and Human Populations. Mol. Biol. Evol. *25*, 2735–2743.

Vervoort, R., Lennon, A., Bird, A.C., Tulloch, B., Axton, R., Miano, M.G., Meindl, A., Meitinger, T., Ciccodicola, A., and Wright, A.F. (2000). Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. Nat. Genet. *25*, 462–466.

Vidal, F., Farssac, E., Tusell, J., Puig, L., and Gallardo, D. (2002). First molecular characterization of an unequal homologous alu-mediated recombination event responsible for hemophilia. Thromb. Haemost. *88*, 12–16.

Vissers, L., Bhatt, S., M Janssen, I., Xia, Z., R Lalani, S., Pfundt, R., Derwinska, K., B A de Vries, B., Gilissen, C., Hoischen, A., et al. (2009). Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. Hum. Mol. Genet. *18*, 3579–3593.

Vollger, M.R., Dishuck, P.C., Sorensen, M., Welch, A.E., Dang, V., Dougherty, M.L., Graves-Lindsay, T.A., Wilson, R.K., Chaisson, M.J.P., and Eichler, E.E. (2019). Long-read sequence and assembly of segmental duplications. Nat. Methods *16*, 88.

Vollrath, D., Nathans, J., and Davis, R.W. (1988). Tandem array of human visual pigment genes at Xq28. Science *240*, 1669–1672.

Wagner-Schuman, M., Neitz, J., Rha, J., Williams, D.R., Neitz, M., and Carroll, J. (2010). Color-deficient cone mosaics associated with Xq28 opsin mutations: A stop codon versus gene deletions. Vision Res. *50*, 2396–2402.

Wahl, M.C., and Lührmann, R. (2015a). SnapShot: Spliceosome Dynamics I. Cell *161*, 1474-1474.e1.

Wahl, M.C., and Lührmann, R. (2015b). SnapShot: Spliceosome Dynamics II. Cell *162*, 456-456.e1.

Wahl, M.C., and Lührmann, R. (2015c). SnapShot: Spliceosome Dynamics III. Cell *162*, 456-456.e1.

Wald, G. (1968). The molecular basis of visual excitation. Nature. *219*, 800-807.Walley, A.J., and Harris, A. (1993). A novel point mutation (D380A) and a rare deletion (1255delI55) in the glucocerebrosidase gene causing Gaucher's disease. Hum. Mol. Genet. *2*, 1737–1738.

Wang, E., and Cambi, F. (2009). Heterogeneous Nuclear Ribonucleoproteins H and F Regulate the Proteolipid Protein/DM20 Ratio by Recruiting U1 Small Nuclear Ribonucleoprotein through a Complex Array of G Runs. J. Biol. Chem. *284*, 11194–11204.

Wang, G.-S., and Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. Nat. Rev. Genet. *8*, 749.

Wang, L., and Yan, F. (2017). Molecular insights into the specific recognition between the RNA binding domain qRRM2 of hnRNP F and G-tract RNA: A molecular dynamics study. Biochem. Biophys. Res. Commun. *494*, 95–100.

Wang, Y., and Wang, Z. (2014). Systematical identification of splicing regulatory cis-elements and cognate trans-factors. Methods *65*, 350–358.

Wang, C., Hosono, K., Kachi, S., Suto, K., Nakamura, M., Terasaki, H., Miyake, Y., Hotta, Y., and Minoshima, S. (2016). Novel OPN1LW/OPN1MW deletion mutations in 2 Japanese families with blue cone monochromacy. Hum. Genome Var. *3*, 16011.

Wang, E., Dimova, N., and Cambi, F. (2007). PLP/DM20 ratio is regulated by hnRNPH and F and a novel G-rich enhancer in oligodendrocytes. Nucleic Acids Res. *35*, 4164–4178.

Wang, Y., Macke, J.P., Merbs, S.L., Zack, D.J., Klaunberg, B., Bennett, J., Gearhart, J., and Nathans, J. (1992). A locus control region adjacent to the human red and green visual pigment genes. Neuron *9*, 429–440.

Wang, Y., Friedl, W., Lamberti, C., Jungck, M., Mathiak, M., Pagenstecher, C., Propping, P., and Mangold, E. (2003). Hereditary nonpolyposis colorectal cancer: Frequent occurrence of large genomic deletions in MSH2 and MLH1 genes. Int. J. Cancer *103*, 636–641.

Warnecke, T., Weber, C.C., and Hurst, L.D. (2009). Why there is more to protein evolution than protein function: splicing, nucleosomes and dual-coding sequence. Biochem. Soc. Trans. *37*, 756–761.

Waseem, N.H., Vaclavik, V., Webster, A., Jenkins, S.A., Bird, A.C., and Bhattacharya, S.S. (2007). Mutations in the Gene Coding for the Pre-mRNA Splicing Factor, PRPF31, in Patients with Autosomal Dominant Retinitis Pigmentosa. Invest. Ophthalmol. Vis. Sci. *48*, 1330–1334.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Weidenhammer, E.M., Ruiz-Noriega, M., and Woolford, J.L. (1997). Prp31p promotes the association of the U4/U6 x U5 tri-snRNP with prespliceosomes to form spliceosomes in Saccharomyces cerevisiae. Mol. Cell. Biol. *17*, 3580–3588.

Weile, J., and Roth, F.P. (2018). Multiplexed assays of variant effects contribute to a growing genotype–phenotype atlas. Hum. Genet.

Weisschuh, N., Mayer, A.K., Strom, T.M., Kohl, S., Glöckle, N., Schubach, M., Andreasson, S., Bernd, A., Birch, D.G., Hamel, C.P., et al. (2016). Mutation Detection in Patients with Retinal Dystrophies Using Targeted Next Generation Sequencing. PLOS ONE *11*, e0145951.

Weisschuh, N., Buena-Atienza, E., Wissinger, B (2020). Splicing mutations in inherited retinal diseases [published online ahead of print, 2020 Jun 15]. Prog Retin Eye Res. 100874.

Welby, E., Lakowski, J., Foggia, V.D., Budinger, D., Gonzalez-Cordero, A., Lun, A.T.L., Epstein, M., Patel, A., Cuevas, E., Kruczek, K., et al. (2017). Isolation and Comparative Transcriptome Analysis of Human Fetal and iPSC-Derived Cone Photoreceptor Cells. Stem Cell Rep. *9*, 1898.

White, R., Gonsior, C., Bauer, N.M., Krämer-Albers, E.-M., Luhmann, H.J., and Trotter, J. (2012). Heterogeneous Nuclear Ribonucleoprotein (hnRNP) F Is a Novel Component of Oligodendroglial RNA Transport Granules Contributing to Regulation of Myelin Basic Protein (MBP) Synthesis. J. Biol. Chem. *287*, 1742–1754.

White, T.B., Morales, M.E., and Deininger, P.L. (2015). Alu elements and DNA double-strand break repair. Mob. Genet. Elem. *5*, 81–85.

Wiehe, T., Mountain, J., Parham, P., and Slatkin, M. (2000). Distinguishing recombination and intragenic gene conversion by linkage disequilibrium patterns. Genet. Res. *75*, 61–73.
Winderickx, J., Battisti, L., Motulsky, A.G., and Deeb, S.S. (1992). Selective expression of human X chromosome-linked green opsin genes. Proc. Natl. Acad. Sci. U. S. A. *89*, 9710.
Winderickx, J., Battisti, L., Hibiya, Y., Motulsky, A.G., and Deeb, S.S. (1993). Haplotype diversity in the human red and green opsin genes: evidence for frequent sequence exchange in exon 3. Hum. Mol. Genet. *2*, 1413–1421.
Wong, K.H.Y., Levy-Sakin, M., and Kwok, P.-Y. (2018). De novo human genome assemblies reveal spectrum of alternative haplotypes in diverse populations. Nat. Commun. *9*, 3040.
Workman, R.E., Tang, A., Tang, P.S., Jain, M., Tyson, J.R., Zuzarte, P.C., Gilpatrick, T., Razaghi, R., Quick, J., Sadowski, N., et al. (2018). Nanopore native RNA sequencing of a human poly(A) transcriptome. BioRxiv 459529.
Wray, G.A. (2007). The evolutionary significance of *cis*-regulatory mutations. Nat. Rev. Genet. *8*, 206–216.
Wu, L., Wen, C., Qin, Y., Yin, H., Tu, Q., Nostrand, J.D.V., Yuan, T., Yuan, M., Deng, Y., and Zhou, J. (2015). Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. BMC Microbiol. *15*, 125.
Xiao, X., Wang, Z., Jang, M., Nutiu, R., Wang, E.T., and Burge, C.B. (2009). Splice site strength–dependent activity and genetic buffering by poly-G runs. Nat. Struct. Mol. Biol. *16*, 1094.
Yang, H.-J., Ratnapriya, R., Cogliati, T., Kim, J.-W., and Swaroop, A. (2015). Vision from next generation sequencing: Multi-dimensional genome-wide analysis for producing gene regulatory networks underlying retinal development, aging and disease. Progress in Retinal and Eye Research *46*, 1.
Yatsenko, S.A., Bakos, H.A., Vitullo, K., Kedrov, M., Kishore, A., Jennings, B.J., Surti, U., Wood-Trageser, M.A., Cercone, S., Yatsenko, A.N., et al. (2016). High-resolution microarray analysis unravels complex Xq28 aberrations in patients and carriers affected by X-linked blue cone monochromacy. Clin. Genet. *89*, 82–87.
Yoshida, T., Kokura, K., Makino, Y., Ossipow, V., and Tamura, T. (1999). Heterogeneous nuclear RNA-ribonucleoprotein F binds to DNA via an oligo(dG)-motif and is associated with RNA polymerase II. Genes Cells *4*, 707–719.
Young, T.L., Deeb, S.S., Ronan, S.M., Dewan, A.T., Alvear, A.B., Scavello, G.S., Paluru, P.C., Brott, M.S., Hayashi, T., Holleschau, A.M., et al. (2004). X-linked high myopia associated with cone dysfunction. Arch. Ophthalmol. *122*, 897–908.
Zangenberg, G., Huang, M.-M., Arnheim, N., and Erlich, H. (1995). New HLA–DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. Nat. Genet. *10*, 407–414.
Zearfoss, N.R., Johnson, E.S., and Ryder, S.P. (2013). hnRNP A1 and secondary structure coordinate alternative splicing of Mag. RNA *19*, 948–957.
Zeitz, C., Michiels, C., Neuillé, M., Friedburg, C., Condroyer, C., Boyard, F., Antonio, A., Bouzidi, N., Milicevic, D., Veaux, R., et al. (2019). Where are the missing gene defects in inherited retinal disorders? Intronic and synonymous variants contribute at least to 4% of CACNA1F-mediated inherited retinal disorders. Hum. Mutat. *40*, 765–787.
Zeng, C., and Berget, S.M. (2000). Participation of the C-Terminal Domain of RNA Polymerase II in Exon Definition during Pre-mRNA Splicing. Mol. Cell. Biol. *20*, 8290–8301.
Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J., and Shi, Y. (2017). An Atomic Structure of the Human Spliceosome. Cell *169*, 918-929.e14.
Zhang, Z., Zhou, L., Wang, P., Liu, Y., Chen, X., Hu, L., and Kong, X. (2009). Divergence of exonic splicing elements after gene duplication and the impact on gene structures. Genome Biol. *10*, R120.
Zhao, Z., Hewett-Emmett, D., and Li, W.H. (1998). Frequent gene conversion between human red and green opsin genes. J. Mol. Evol. *46*, 494–496.
Zheng, G.X.Y., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D.A., Merrill, L., Terry, J.M., et al. (2016). Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat. Biotechnol. *34*, 303.
Zhou, Y.H., and Li, W.H. (1996). Gene conversion and natural selection in the evolution of X-linked color vision genes in higher primates. Mol. Biol. Evol. *13*, 780–783.

11 Abbreviations

Abbreviation	Description
ACMG	American College of Medical Genetics and Genomics
adRP	Autosomal dominant RP
AMP	Association for Molecular Pathology
arRP	Autosomal recessive RP
BCM	Blue Cone Monochromacy
BED	Bornholm Eye Disease
bp	base pair
CCS	Circular Consensus Sequencing
CERES	Composite Exonic Regulatory Elements of Splicing
CF	Cystic Fibrosis
cfu	colony forming units
cGMP	cyclic guanosine monophosphate
Cluster PF	Cluster Passing Filter
CNG channel	cyclic nucleotide-gated channel
CNV	Copy Number Variation
CRD	Cone-rod dystrophy
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DSBs	Double Strand Breaks
dsDNA	double-stranded DNA
ED	Ensemble diversity
EMSA	Electrophoresis Mobility Shift Assays
eQTL	expression Quantitative Trait Locus
ERG	Electroretinography
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
ExAC	Exome Aggregation Consortium
FAM	Fluorescein Amidite
FDR	False-discovery Rate
FRAM	Fossil Right Arm Monomer
gnomAD	Genome Aggregation Database
HEK293	Human Embryonic Kidney 293 Cells
HEK293T	Human Embryonic Kidney 293 Cells that contain the SV40 T-antigen
HGMD	Human Gene Mutation Database
HI-NHEJ	homeology-induced Non-Homologous End-Joining
hnRNPF	heterogeneous nuclear ribonucleoprotein F
hnRNPH	heterogeneous nuclear ribonucleoprotein H (family)
hnRNPH1	heterogeneous nuclear ribonucleoprotein H1
hnRNPH2	heterogeneous nuclear ribonucleoprotein H2
hnRNPH3	heterogeneous nuclear ribonucleoprotein H3
HPLC	high performance liquid chromatography
HSF	Human Splicing Finder
IC	Interval Confidence
IFC	In-Fusion Cloning
iPSCs	Induced Pluripotent Stem Cells
IRDs	Inherited retinal dystrophies
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
kb	kilobase
LCR	Locus Control Region
LD	Linkage Disequilibrium
LD-PCR	Long Distance Polymerase Chain Reaction
LIAVA	c.(453A > G; 457A > C; 465C > G; 511G > A; 513G > T; 521C; 532A > G; 538T > G) and thus, enconding for a cone photopigment protein with the amino acid exchanges: p.[(=); M153L; (=); V171I; A174A; I178V; S180A]

Abbreviation	Description
LOVD	Leiden Open Variation Database
LPSP	Library Preparation and Sequencing Protocol
LWS	long-wavelength sensitive
MAF	Minor Allele Frequency
MaPSy	Massively Parallel Splicing Assay
MAVE	Multiplexed assays for variant effect
Mb	Megabase
MCS	Multiple Cloning Site
mESC	mouse embryonic stem cells
MFASS	Multiplexed Functional Assay of Splicing using Sort-seq
MFE	Minimum Free Energy
MPRA	Massively parallel reporter assay
MWS	middle-wavelength sensitive
MYA	Million Years Ago
NAHR	Non-Allelic Homologous Recombination
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End-Joining
NMD	Nonsense-mediated Decay
OPN1LW	Human Opsin 1, long-wavelength sensitive gene
OPN1MW	Human Opsin 1, middle-wavelength sensitive gene
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PDE6A	Human Rod cGMP Phosphodiesterase 6A gene
PE	paired-end
PPT	Polypyrimidine Tract
PTC	Premature Termination Codon
QTL	Quantitative Trait Locus
RBPs	RNA Binding Proteins
RetNet	Retinal Information Network database
RFLP	Restriction Fragment Length Polymorphism
ROS	rod photoreceptor outer segments
RP	Retinitis Pigmentosa
RRMs	RNA Recognition Motifs
RT-PCR	Reverse Transcriptase PCR
SELEX	Systematic Evolution of Ligands by Exponential enrichment
SF1	Splicing Factor 1
SILAC	Stable Isotope Labeling with Amino Acids in Cell Culture
siRNA	small interference RNA
SMA	Spinal Muscular Atrophy
SMRT	Single-Molecule Real-Time
SNPs	Single Nucleotide Polymorphisms
snRNPs	small nuclear ribonucleoproteins
SNVs	Single Nucleotide Variants
sQTL	splicing Quantitative Trait Locus
SR	Serine/ariginine-rich
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
UTR	untranslated region
VUS	Variant of Uncertain Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
XLCD	X-linked Cone Dysfunction
ΔG	Predicted MFE change in the Gibb's folding energy
3'ss	3' splice site
5'ss	5' splice site

12 Personal Contribution

Publication I (Section 6.1; Appendix, 13.1.1): Buena-Atienza E, Nasser F, Kohl S, Wissinger B (2018) A 73,128 bp *de novo* deletion encompassing the *OPN1LW/OPN1MW* gene cluster in sporadic Blue Cone Monochromacy: a case report. *BMC Medical Genetics* 19: 107.

Elena Buena-Atienza: wrote the manuscript, collected genetic and clinical data, contacted clinicians and discussed interpretation of results. I performed the experiments of microsatellite marker analysis, the breakpoint PCR segregation analysis, and the bioinformatic exploration of repetitive elements. I performed genotyping of family members.

The other authors: wrote and revised the manuscript, clinical data collection, establishment of the breakpoint PCR

Publication II (Section 6.3; Appendix, 13.1.2): Buena-Atienza E, Rüther K, Baumann B, Bergholz R, Birch D, De Baere E, Dollfus H, Greally MT, Gustavsson P, Hamel CP, Heckenlively JR, Leroy BP, Plomp AS, Pott JWR, Rose K, Rosenberg T, Stark Z, Verheij JBGM, Weleber R, Zobor D, Weisschuh N, Kohl S, Wissinger B (2016) *De novo* intrachromosomal gene conversion from *OPN1MW* to *OPN1LW* in the male germline results in Blue Cone Monochromacy. *Scientific Reports* (2016) 6, 28253

Elena Buena-Atienza: completed genotyping of patients by copy number assays and sequencing when needed. I performed all cloning experiments to generate minigenes for haplotypes found in the BCM/XLCD patients, including verification of clones. Splicing assays for all minigenes including relative quantification of splice products and the downstream characterization of the transcripts including cloning of RT-PCR products and screening of several hundreds of colonies. I performed the microsatellite marker analysis and developed strategies to identify the gene converted tract. Contributed to summarize the clinical data and correlate it with genotypes and functional haplotype data available. Conceived, wrote and revised the manuscript.

The other authors: genotyping of patients and collection of clinical data had been started, shared expertise in minigene splicing assays, contacting clinicians, supervision of this project. The manuscript was corrected by all authors.

Publication III (Section 6.4; Appendix, 13.1.3): Sothilingam V, Garrido MG, Jiao K, Buena-Atienza E, Sahaboglu A, Trifunović D, Balendran S, Koepfli T, Mühlfriedel R, Schön C, Biel M, Heckmann A, Beck SC, Michalakis S, Wissinger B, Seeliger MW, Paquet-Durand F (2015) Retinitis pigmentosa: impact of different Pde6a point mutations on the disease phenotype. *Human Molecular Genetics* (2015) 24, 5486–5499

Elena Buena-Atienza: contributed to writing and revision of the manuscript, generation of figures, qualitatively characterized the mice retinal transcripts and performed cloning experiments and screening of hundreds of colonies. Established all minigene constructs by cloning and mutagenesis and performed the splicing assays in two cell lines.

The other authors and service: wrote and revise the manuscript, generation of *Pde6a* mutant mice model, protein assay, pyrosequencing, histology, immunohistochemistry, immunofluorescence, calpain activity and TUNEL assay, microscopy, *in vivo* imaging and ERG.

Unpublished Results embedded in Section 6.4
Elena Buena-Atienza: performed experiments with *rd10* mice, human minigenes and *in silico* searches.

The other authors: provided retinal RNA from *rd10* mice

Publication IV (Section 6.6, Appendix, 13.1.4): Buena-Atienza E, Codrea MC, Czemmel S, Baumann, Kohl S, Poths S, Nahnsen S, Wissinger B. A comprehensive reference for exon 3 splicing defects of the human *OPN1LW* and *OPN1MW* opsin genes by means of a parallelized minigene splicing assay. *Manuscript in preparation*

Elena Buena-Atienza: I designed the project, interpreted the results and performed rational troubleshooting. I prepared all amplicon-based libraries derived from the RT-PCR Output Library upon splicing assay. I conceived the detailed strategy and methodology, I performed all the experiments autonomously, designed all the libraries, I prepared all the libraries for sequencing except for the SMRT library of the Input Library. I did troubleshooting and proposed alternatives and improvements when necessary. I performed downstream analysis, mined the data, statistical analysis, wrote the manuscript and prepared all figures with the exception of Figure 8.

The other authors and outsourced services: MiSeq sequencing, long-red sequencing of the Input and Output Library, CCS analysis in SMRTLink (v.5.0.0), scripts for alignments, assembly, barcode deconvolution, genotyping of disease and control groups, generation of Figure 8. The co-authors read the manuscript and contributed to its revision with comments and suggestions.

Publication V (Appendix, 13.1.5). Nicole Weisschuh, Elena Buena-Atienza, Bernd Wissinger. Splicing Mutations in Inherited Retinal Diseases. *Manuscript accepted in Progress in Retinal and Eye Research*

This is a Review article which compiles work from other authors in the field but also work from the group of Prof. Bernd Wissinger including parts of the work presented in this thesis in Sections 6.4 and 6.6.

Elena Buena-Atienza: I contributed to the writing of the manuscript, generation of Figures 7 and 8 and revision of the whole manuscript.

The other authors: wrote the manuscript, generated figures and revised the whole manuscript.

Unpublished Results I (Section 6.2, Appendix, 13.2.1): “A 142 bp intra-exonic deletion in both *OPN1LW* and *OPN1MW* underlying Blue Cone Monochromacy”

Elena Buena Atienza: performed detailed investigations of the BCM262 family, segregation analysis, sequencing of the breakpoint PCR and bioinformatic analysis and interpretation.

Mutation screening was performed by Britta Baumann and Prof. Bernd Wissinger.

Unpublished Results II (Section 6.5): Development of a multiplexed splicing assay

Elena Buena Atienza: conceived the strategies, designed and performed the experiments, established the primary analysis and performed the secondary analysis

Sequencing was provided by the c.ATG Core Facility for NGS Tübingen. The original idea was conceived by Prof. Bernd Wissinger for which he received a joint grant from the European Union’s Seventh Framework Programme for research, technological development and demonstration [grant agreement no 317472].

Unpublished Results III (Section 6.7, Appendix, 13.2.3): **“**Putative splicing factors binding *cis*-regulatory elements within exon 3 of *OPN1LW/MW* haplotypes**”**

Elena Buena Atienza: presented the idea, designed the RNA baits and prepared the RNA. Performed the RNA-pulldown assay and prepared samples for mass spectrometry. I contributed considerably to interpretation and design. I performed the siRNA knockdown experiments combined with minigene assays and subsequent quantification at protein and transcript level.

Dr. Tina Beyer assisted with the RNA-pulldown assay and preparation of the samples for mass spectrometry and performed MaxQuant and Perseus analysis and, together with Dr. Karsten Boldt assisted in design and interpretation. Running of the mass spectrometer instrument was performed by trained technicians of the Medical Proteomic Center Tübingen.

13 Appendix

13.1 Publications

13.1.1 Publication I

Elena Buena-Atienza, Fadi Nasser, Susanne Kohl and **Bernd Wissinger**. A 73,128 bp *de novo* deletion encompassing the *OPN1LW/OPN1MW* gene cluster in sporadic Blue Cone Monochromacy: a case report. *BMC Medical Genetics* (2018) 19

Online version available: https://doi.org/10.1186/s12881-018-0623-8

Buena-Atienza *et al. BMC Medical Genetics* (2018) 19:107
https://doi.org/10.1186/s12881-018-0623-8

BMC Medical Genetics

CASE REPORT Open Access

A 73,128 bp *de novo* deletion encompassing the *OPN1LW/OPN1MW* gene cluster in sporadic Blue Cone Monochromacy: a case report

Elena Buena-Atienza, Fadi Nasser, Susanne Kohl and Bernd Wissinger*

Abstract

Background: Blue Cone Monochromacy (BCM) is a rare congenital cone dysfunction disorder with X-linked recessive mode of inheritance. BCM is caused by mutations at the *OPN1LW/MW* cone opsin gene cluster including deletions of the locus control region (LCR) and/or parts of the gene cluster. We aimed at investigating the clinical presentation, genetic cause and inheritance underlying a sporadic case of BCM.

Case presentation: We report a 24-year-old male presenting with congenital photophobia, nystagmus and colour vision abnormalities. There was no history of retinal dystrophy in the family. Clinical diagnosis of BCM was supported by genetic investigations of the patient and his family members. Molecular genetic analysis of the *OPN1LW/OPN1MW* gene cluster revealed a novel deletion of about 73 kb in the patient encompassing the LCR. The deletion was absent in the X-chromosomes of both the mother and transmitting grandfather.

Conclusions: The present report provides the clinical findings and the genetic basis underlying a sporadic BCM case which is caused by a *de novo* deletion within the *OPN1LW/MW* gene cluster originating from the mother's germline due to Alu-repeat mediated recombination. This is the first report of a *de novo* deletion resulting in BCM, highlighting the importance to consider BCM and perform genetic testing for this condition in male patients with cone dysfunction also in the absence of a positive family history.

Keywords: *De novo* mutations, Sporadic cases, Blue Cone Monochromacy, Colour vision deficiency, Alu-mediated recombination, Retinal dystrophy

Background

Blue Cone Monochromacy (BCM; OMIM#303700) is characterized by reduced visual acuity, photophobia, colour vision deficiency and it is regularly accompanied by congenital nystagmus and myopia [1]. The clinical findings in BCM patients largely overlap with those of achromatopsia patients. An X-linked recessive pattern of inheritance is a practical distinctive trait of BCM. Yet, lack of family history in simplex cases hinders prioritization of genetic tests. In addition, genetic testing for BCM is not regularly available at most diagnostic genetic testing laboratories due to the complexity of the opsin gene cluster.

The genes for the long wavelength sensitive (*OPN1LW*; LW opsin gene) and middle wavelength sensitive (*OPN1MW*; MW opsin gene) pigments in LW and MW cone photoreceptors are located on the X-chromosome and arranged as a head-to-tail tandem array in the so-called *OPN1LW/OPN1MW* gene cluster. The first position in the cluster is generally occupied by an *OPN1LW* gene copy followed by one or multiple *OPN11MW* gene copies. Each gene copy possesses a direct upstream promoter, while the expression of the genes in the gene cluster is regulated by an upstream locus control region (LCR). The LCR is a 0.6 kb *cis*-regulatory sequence essential to drive expression in a distance dependent manner so

* Correspondence: wissinger@uni-tuebingen.de
Institute for Ophthalmic Research, Centre for Ophthalmology, University of Tuebingen, Elfriede-Aulhorn 7, D-72076 Tuebingen, Germany

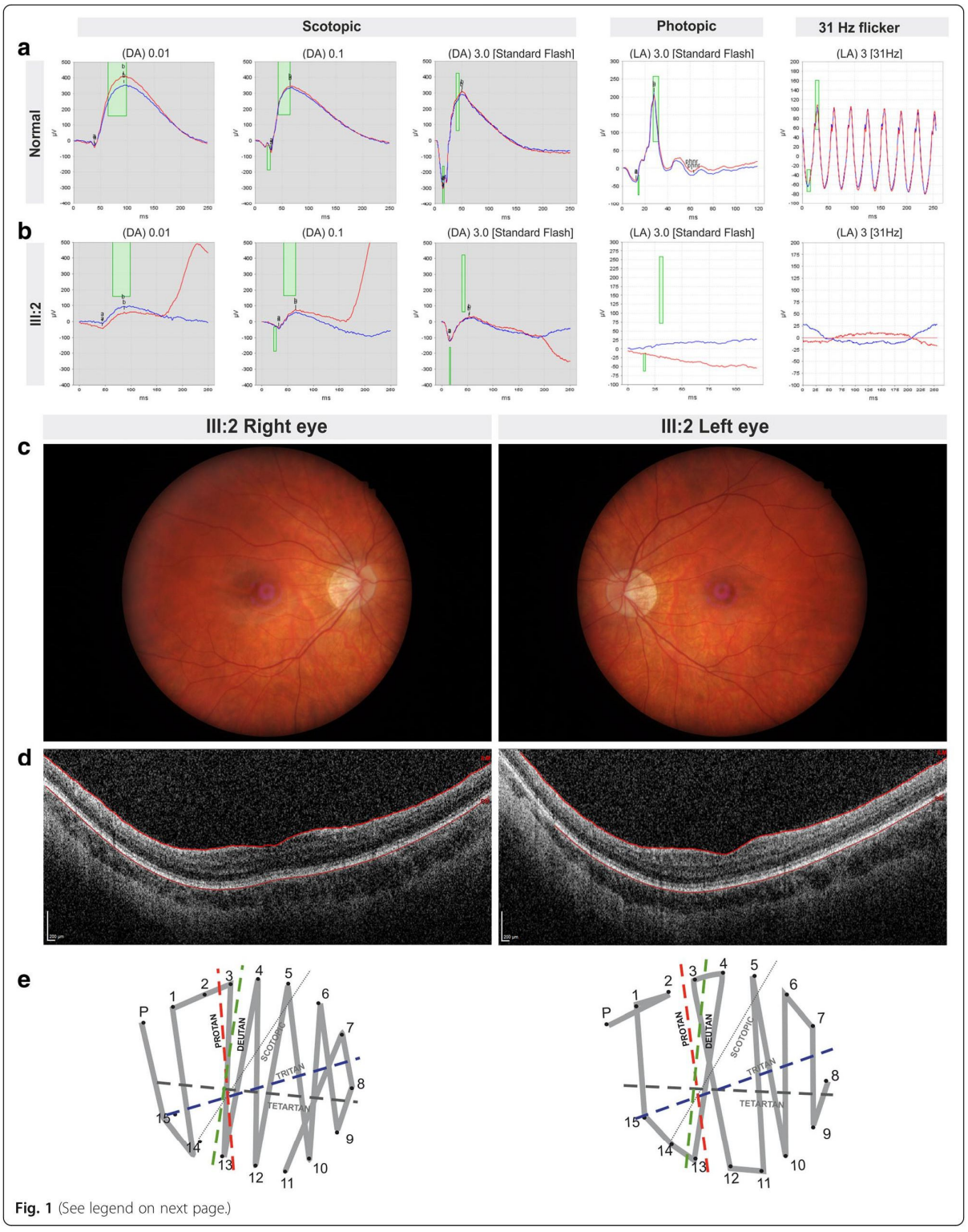


Fig. 1 (See legend on next page.)

(See figure on previous page.)
Fig. 1 Ophthalmological findings in BCM patient III:2 **a** Full-field ERG recordings according to ISCEV Standards of a normal subject **b** ERG of the index patient (III:2) at last clinical examination (24 years old) showing mild reduction under dark-adapted (DA) scotopic conditions and undetectable responses under light-adapted (LA) photopic responses in both eyes. Strength of the light stimulus is given in photopic cd*s/m^2, if not stated otherwise. Grey sections indicate scotopic recording conditions. Red colour represents right eye and blue the left eye. Green rectangles indicate 5–95% confidence intervals for amplitudes and implicit times of a normal population. **c** Right and left eye fundus image of the patient showing tilted optic disc. **d** Right and left eye OCT shows normal retinal architecture in the BCM patient with thinned photoreceptor layers. **e** Farnsworth Panel D-15 saturated shows confusion errors along the protan-deutan axes for both eyes

that essentially only the first two gene copies within the gene cluster are expressed [2].

Deletions affecting the *OPN1LW/OPN1MW* gene cluster on Xq28 have been estimated to account for 30–40% of all mutations found in BCM patients [3, 4]. Whereas deletions may be restricted to the LCR region, a number of deletions have been reported to extend towards the human cone opsin gene cluster, i.e. deleting partially or completely the *OPN1LW* and/or *OPN1MW* gene copies [1, 3–7].

In this report, we present a male with a clinical phenotype of BCM carrying a novel deletion originated by a de novo mutation event in the *OPN1LW/OPN1MW* gene cluster. Herein, we describe the clinical findings of the patient, present the results from molecular genetic investigations of the patient and his family members, and discuss the putative underlying mechanism leading to this de novo mutation.

Case presentation

A 24-year-old male presented with photophobia since birth. No family history for colour vision defects or retinal dystrophies was reported. Myopia with an refractive error of – 5.50 D (right eye) and – 6.50 D (left eye) and astigmatism were found in the patient (III:2) at the age of 8 months along with nystagmus but devoid of strabismus. Glasses were given at the age of 1 year. Difficulties distinguishing colours were noticed by his parents at the age of 3 years. Achromatopsia was the first suspected diagnosis. At the age of 4 years occlusion therapy alternating in both eyes for 2 months was attempted to treat amblyopia but was unsuccessfull. A best-corrected visual acuity of 20/200 was measured with Snellen charts at the age of 6 years. No brain injuries were detected by magnetic resonance imaging. Visual evoked potential flash and B-scan ultrasonography performed normal for both eyes. At the age of 11 years a visual acuity of 20/250 was measured. At the latest exam at the age of 24 years, full-field light- and dark-adapted electroretinogram (ERG) recordings were performed according to the International Society for Clinical Electrophysiology of Vision (ISCEV) standard protocol [8]. Subnormal amplitudes under scotopic conditions and extinct responses under photopic conditions were observed in both eyes of the patient (III:2) in comparison to normal controls (Fig. 1a-b). A visual acuity of 20/400 was measured for both eyes with a myopic correction of – 12.00 D (right eye) and – 11.50 D (left eye). Anterior segment, pupillary reflexes and intraocular pressure revealed no abnormalities. Eye fundus examination revealed normal retinal vessels, optic nerve heads showing tilted optic discs with myopic conus and the maculae had elapsed reflex without waxy reflex (Fig. 1c). Spectral domain optical coherence tomography (SD-OCT) performed with Spectralis OCT (Heidelberg Engineering) showed normal retinal architecture with thinned photoreceptor layer (Fig. 1d). Colour vision test evaluated with the Farnsworth D-15 Colour Test revealed protan-deutan confusion errors (Fig. 1e).

Mutation screening of *OPN1LW/OPN1MW* gene cluster in the patient (III:2) was performed as previously described [9]. Genotyping PCRs with genomic DNA from the patient (III:2, see Fig. 2a for a pedigree of the genetically investigated family members) revealed absence of the LCR and promoter regions of both *OPN1LW* and *OPN1MW* genes – indicative for a large deletion – but presence of the 3′ parts of the *OPN1MW* gene, namely exons 4, 5 and 6. Upon fine sequence tag site content mapping, the deletion was finally bridged with a PCR amplicon performed with primers BCM#27_F (5′- TCGACCCAGAATTAACCTC TCT -3′) and BCM#27BPR (5'-TCTAAAAATGGACA AGGATTAACCA -3′) which was sequenced with Sanger to determine the exact breakpoints in patient III:2 (Fig. 2b and Fig. 2c). The deletion, NC_000023.11:g.154,118,184_154,191,311del, encompasses 73,128 bp with the centromeric breakpoint located in the intergenic region between *MECP2* and *OPN1LW* and the telomeric breakpoint within intron 3 of *OPN1MW* (Fig. 2d).

A sequence alignment of the breakpoint junction sequence in the patient (III:2) with the corresponding non-mutant sequence sections from his grandfather (I:1) revealed an overlapping stretch of 13 bp between the centromeric and the telomeric breakpoint sequences (Fig. 2b) shared by two Alu elements. The sequence remnants embedded within the deletion breakpoints resembled the junction of two Alu elements; (1) a fossil right Alu momomer (FRAM) element at the centromeric breakpoint of the deletion, and (2) an *AluJo* element in intron 3 of *OPN1MW* at the telomeric breakpoint of the deletion (Fig. 2d). Microsatellite marker analysis revealed

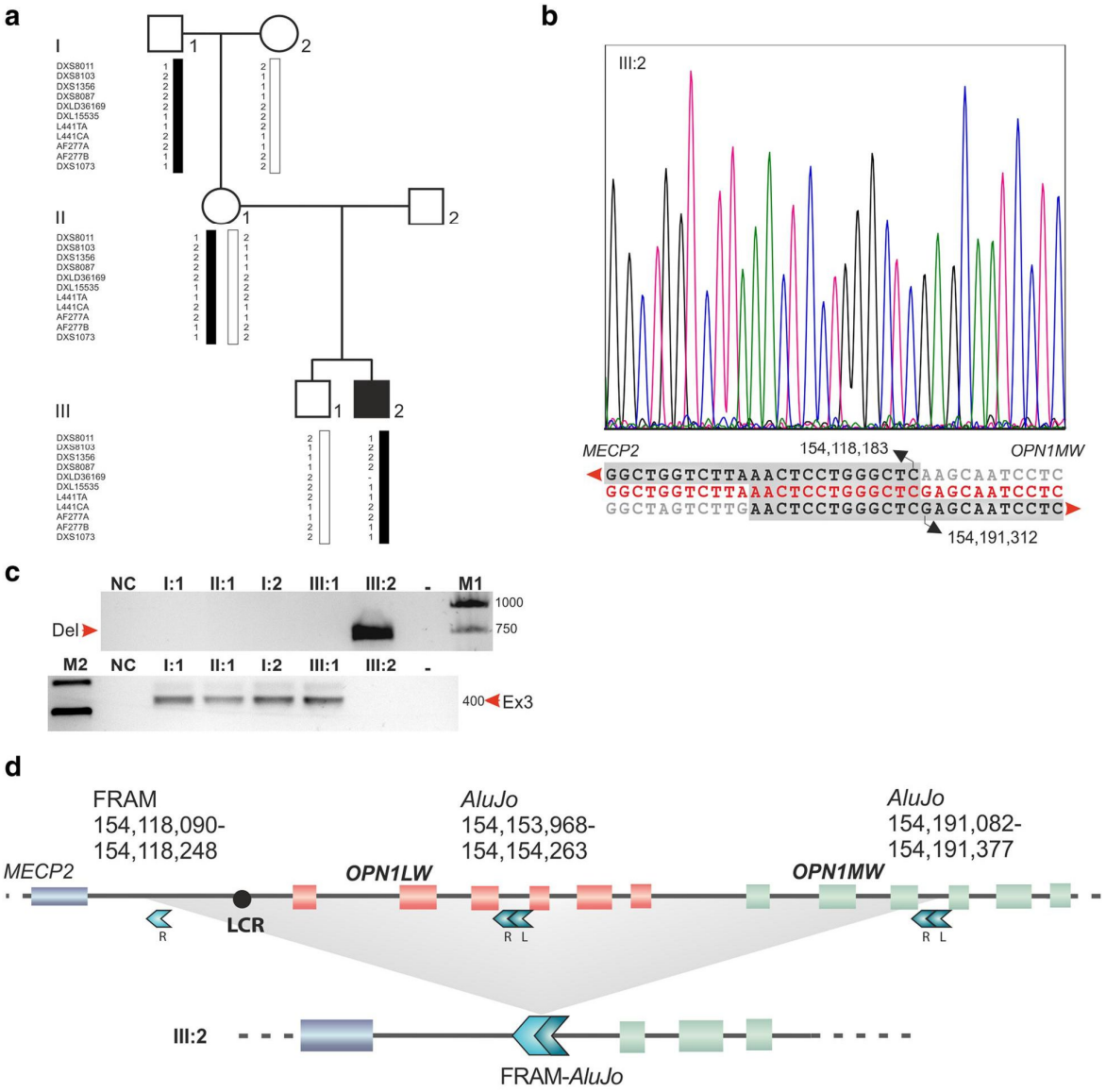


Fig. 2 Identification of a *de novo* 73,128 bp deletion at the cone opsin gene cluster **a** Pedigree of the family with a single affected male (III:2). Reconstructed haplotypes based on microsatellite markers flanking the *OPN1LW/MW* gene cluster within Xq28 revealed transmission of the X-chromosome from the maternal grandfather (I:1) to the index patient (III:2). **b** Sanger sequencing traces of the deletion breakpoints observed in the index patient (III:2) and its sequence depicted in red letters in the alignment underneath. Wildtype sequence at the centromeric and the telomeric breakpoint are depicted in black letters at the top and the bottom of the alignment, respectively. The breakpoints define a 73,128 bp deletion (NC_000023.11:g.154,118,184_154,191,311del) including the LCR, the entire *OPN1LW* gene and the first three exons of the *OPN1MW* gene. **c** Segregation analysis demonstrated the *de novo* occurence of the deletion. A breakpoint PCR amplicon yielded only a product for patient III:2 ("Del", upper agarose gel panel), whereas a PCR amplicon for exon 3 of *OPN1LW* and *OPN1MW* ("Ex3", lower agarose gel panel) yielded products for the mother (II:1) and the grandfatehr (I:1) but not for the patient (III:2). M1 & M2: DNA size standards; NC: no DNA template negative control; –: empty well. **d** Schematic representation of the 73,128 bp deletion encompassing the LCR, the entire *OPN1LW* and 5′ parts of *OPN1MW* gene in the BCM patient (III:2). Red and green boxes depict individual exons of the *OPN1LW* and *OPN1MW* genes, respectively. The localization and composition of Alu elements most likely involved in the mutation event are shown as arrowheads. Physical coordinates refer to Genbank entry NC_000023.11

that the X-chromosome present in the patient (III:2) had been transmitted from his maternal grandfather (I:1, Fig. 2a). Segregation analysis performed by means of breakpoint PCR amplification showed that neither the patient's grandfather (I:1) nor the mother (II:1) carry the deletion (Fig. 2c). The CARE guidelines were followed in reporting this case.

Discussion and conclusions

In this report we describe a deletion spanning 73,128 bp within the human *OPN1LW/OPN1MW* cone opsin gene cluster in a male with clinical features fully compatible with BCM but lacking family history for this condition. The molecular findings were critical to establish the correct diagnosis for the patient. A more thorough clinical investigation at 24 years of age showed results concordant with a BCM phenotype (Fig. 1a-e). While there is evidence for a link between emmetropization and the M to L cone ratio [10], the mechanism(s) underlying myopia in BCM are still unsolved. Presumably due to the high myopia and the lack of cone contribution to the dark-adapted responses, a mild reduction of scotopic signals was observed (Fig. 1a-b). Previously, rod responses to dim stimuli were reported to be in the lower normal or subnormal range in BCM patients [4, 6].

The NC_000023.11:g.154,118,184_154,191,311del mutation in this patient has not been previously reported. The *de novo* nature of the deletion was confirmed by haplotype reconstruction and segregation analysis of all family members available. While the X-chromosome present in the patient (III:2) had been transmitted from his maternal grandfather (I:1, Fig. 2a), the NC_000023.11:g.154,118,184_154,191,311del mutation was absent in the patient's grandfather (I:1) and mother (II:1, Fig. 2c). Therefore the mutation identified in the patient of this particular case represents a *de novo* event that occurred in the mother's germline (II:1, Fig. 2c).

The breakpoint sequences of the deletion are localized within Alu repetitive elements. Based on the sequence composition prior to deletion and the actual Alu sequence remnants found at the deletion breakpoints, it is tempting to propose a mechanism underlying this *de novo* mutation event. We hypothesize that the monomeric FRAM element underwent pairing with the left arm of the bipartite *AluJo* element and induced subsequent recombination within the 13 bp of microhomology. FRAM and *AluJo* most likely promoted the origin of the deletion during meiosis through misalignment of the homologous stretch [11]. The FRAM and *AluJo* elements may have acted in this very event most likely as nucleation spots allowing mispairing and making the sequence prone to homologous recombination [11]. The resulting deletion gave rise to a hybrid FRAM-*AluJo* element. Notably, more than 95% of the Alu-mediated deletion events identified in the human genome occurred as Alu-mediated unequal homologous recombination [12]. According to the aforementioned model proposed, the Alu element mediated illegitimate recombination may have occurred either by mispairing of sister or non-sister chromatids (interchromosomal), or between elements on the same chromosome (intrachromosomal).

Although several studies have previously reported deletions including the LCR to be associated with a BCM phenotype [1, 3–7], this is the first *de novo* deletion reported to underlie a BCM phenotype. We have previously described a *de novo* intrachromosomal gene conversion event in the male germline transferring a pathogenic haplotype from *OPN1MW* to *OPN1LW* [12]. Notwithstanding, the BCM patient's grandfather presented with a linked colour vision deficiency due to a pathogenic haplotype in *OPN1MW* and the patient's mother was a carrier of the converted *OPN1LW* gene [13]. Our case report herein describes a BCM patient bearing a unique mutation that stem from an independent *de novo* event in a family with no history of BCM or colour vision abnormalities.

In summary, we were able to elucidate the origin of a newly emerged mutation in an isolated case of BCM. These findings emphasize the need to test for X-linked BCM in cases with a clinical diagnosis of cone dysfunction, even in the absence of a family history in order to obtain a correct clinical diagnosis and a valid basis for proper genetic counselling. The novel 73,128 bp deletion within the *OPN1LW/MW* gene cluster occurred most likely as a result of an Alu recombination-mediated mechanism. This is the first report of a *de novo* deletion resulting in BCM.

Abbreviations

BCM: Blue Cone Monochromacy; ERG: Electroretinogram; FRAM: Fossil right Alu momomer; LCR: Locus control region; SD-OCT: Spectral domain optical coherence tomography

Acknowledgements

We thank Britta Baumann for excellent technical assistance. We further acknowledge support by the Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of the University of Tübingen.

Funding

This work was supported by the European Union's Seventh Framework Programme for research, technological development and demonstration [grant agreement no 317472] and by the BCM Families Foundation.

Availability of data and materials

The mutation has been deposited and stored publicly available in ClinVar under the link https://www.ncbi.nlm.nih.gov/clinvar/variation/448956/ with accession number SCV000612182.

Authors' contributions

FN collected clinical data. EB-A performed the genetic investigations and summarized molecular data. SK managed the family case. BW provided the study concept and design. All authors wrote, read and approved the final manuscript.

Ethics approval and consent to participate
All procedures performed involving human participants were in accordance with the tenets of the WMA Declaration of Helsinki and have been approved by the Ethics Committee of the Medical Faculty, Eberhard-Karls University Tuebingen.

Consent for publication
Written informed consent was obtained from the patient to publish the medical information supplied in this case report.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Received: 15 January 2018 Accepted: 12 June 2018
Published online: 26 June 2018

References

1. Nathans J, Davenport CM, Maumenee IH, Lewis RA, Hejtmancik JF, Litt M, et al. Molecular genetics of human blue cone monochromacy. Science. 1989; 245(4920):831–8.
2. Winderickx J, Battisti L, Motulsky AG, Deeb SS. Selective expression of human X chromosome-linked green opsin genes. Proc Natl Acad Sci U S A. 1992;89(20):9710.
3. Nathans J, Maumenee IH, Zrenner E, Sadowski B, Sharpe LT, Lewis RA, et al. Genetic heterogeneity among blue-cone monochromats. Am J Hum Genet. 1993;53(5):987–1000.
4. Ayyagari R, Kakuk LE, Bingham EL, Szczesny JJ, Kemp J, Toda Y, et al. Spectrum of color gene deletions and phenotype in patients with blue cone monochromacy. Hum Genet. 2000;107(1):75–82.
5. Ayyagari R, Kakuk LE, Coats CL, Bingham EL, Toda Y, Felius J, et al. Bilateral macular atrophy in blue cone monochromacy (BCM) with loss of the locus control region (LCR) and part of the red pigment gene. Mol Vis. 1999;5:13.
6. Kellner U, Wissinger B, Tippmann S, Kohl S, Kraus H, Foerster MH. Blue cone monochromatism: clinical findings in patients with mutations in the red/ green opsin gene cluster. Graefes Arch Clin Exp Ophthalmol. 2004;242(9): 729–35.
7. Gardner JC, Liew G, Quan Y-H, Ermetal B, Ueyama H, Davidson AE, et al. Three different cone Opsin gene Array mutational mechanisms with genotype–phenotype correlation and functional investigation of cone Opsin variants. Hum Mutat. 2014;35(11):1354.
8. McCulloch DL, Marmor MF, Brigell MG, Hamilton R, Holder GE, Tzekov R, et al. ISCEV standard for full-field clinical electroretinography (2015 update). Doc Ophthalmol. 2015;130(1):1–12.
9. Cideciyan AV, Hufnagel RB, Carroll J, Sumaroka A, Luo X, Schwartz SB, et al. Human cone visual pigment deletions spare sufficient photoreceptors to warrant gene therapy. Hum Gene Ther. 2013;24(12):993–1006.
10. Gisbert S, Schaeffel F. M to L cone ratios determine eye sizes and baseline refractions in chickens. Exp Eye Res. 2018;172:104–11.
11. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. Nat Rev Genet. 2002;3(5):nrg798.
12. Liu GE, Alkan C, Jiang L, Zhao S, Eichler EE. Comparative analysis of Alu repeats in primate genomes. Genome Res. 2009;19(5):876.
13. Buena-Atienza E, Rüther K, Baumann B, Bergholz R, Birch D, Baere ED, et al. De novo intrachromosomal gene conversion from OPN1MW to OPN1LW in the male germline results in blue cone Monochromacy. Sci Rep. 2016;6: 28253.

13.1.2 Publication II

Elena Buena-Atienza, Klaus Rüther, Britta Baumann, Richard Bergholz, David Birch, Elfride De Baere, Helene Dollfus, Marie T. Greally, Peter Gustavsson, Christian P. Hamel, John R. Heckenlively, Bart P. Leroy, Astrid S. Plomp, Jan Willem R. Pott, Katherine Rose, Thomas Rosenberg, Zornitza Stark, Joke B. G. M. Verheij, Richard Weleber, Ditta Zobor, Nicole Weisschuh, Susanne Kohl and **Bernd Wissinger**. *De novo* intrachromosomal gene conversion from *OPN1MW* to *OPN1LW* in the male germline results in Blue Cone Monochromacy. *Scientific Reports* (2016) 6, 28253

Online version available: https://www.nature.com/articles/srep28253

www.nature.com/scientificreports

OPEN

De novo intrachromosomal gene conversion from *OPN1MW* to *OPN1LW* in the male germline results in Blue Cone Monochromacy

Received: 03 February 2016
Accepted: 01 June 2016
Published: 24 June 2016

Elena Buena-Atienza[1], Klaus Rüther[2], Britta Baumann[1], Richard Bergholz[3], David Birch[4], Elfride De Baere[5], Helene Dollfus[6], Marie T. Greally[7], Peter Gustavsson[8], Christian P. Hamel[9], John R. Heckenlively[10], Bart P. Leroy[5], Astrid S. Plomp[11], Jan Willem R. Pott[12], Katherine Rose[13], Thomas Rosenberg[14], Zornitza Stark[15], Joke B. G. M. Verheij[16], Richard Weleber[17], Ditta Zobor[1], Nicole Weisschuh[1], Susanne Kohl[1] & Bernd Wissinger[1]

X-linked cone dysfunction disorders such as Blue Cone Monochromacy and X-linked Cone Dystrophy are characterized by complete loss (of) or reduced L- and M- cone function due to defects in the *OPN1LW/OPN1MW* gene cluster. Here we investigated 24 affected males from 16 families with either a structurally intact gene cluster or at least one intact single (hybrid) gene but harbouring rare combinations of common SNPs in exon 3 in single or multiple *OPN1LW* and *OPN1MW* gene copies. We assessed twelve different *OPN1LW/MW* exon 3 haplotypes by semi-quantitative minigene splicing assay. Nine haplotypes resulted in aberrant splicing of ≥20% of transcripts including the known pathogenic haplotypes (i.e. 'LIAVA', 'LVAVA') with absent or minute amounts of correctly spliced transcripts, respectively. *De novo* formation of the 'LIAVA' haplotype derived from an ancestral less deleterious 'LIAVS' haplotype was observed in one family with strikingly different phenotypes among affected family members. We could establish intrachromosomal gene conversion in the male germline as underlying mechanism. Gene conversion in the *OPN1LW/OPN1MW* genes has been postulated, however, we are first to demonstrate a *de novo* gene conversion within the lineage of a pedigree.

The apo-proteins of the human long-wavelength and middle-wavelength sensitive cone photoreceptor pigments are encoded by the *OPN1LW* (*LW*; OMIM 300822) and *OPN1MW* genes (*MW*; OMIM 300821), respectively. Arranged in a head-to-tail tandem array on the long arm of the X-chromosome[1], these duplicated genes share 98% sequence identity[2]. The prototypic gene array structure consists of a *LW* followed by a *MW* gene, however, considerable variability has been observed in gene copy number in the human *LW/MW* gene array[3,4]. Yet, studies on

[1]Institute for Ophthalmic Research, Centre for Ophthalmology, Tuebingen, Germany. [2]Sankt Gertrauden-Krankenhaus, Berlin, Germany. [3]Department of Ophthalmology, Charité – Universitätsmedizin Berlin, Campus Virchow-Klinikum, Berlin, Germany. [4]Retina Foundation of the Southwest, Tom and Dorothy Anderson Vision Research Center, Texas, USA. [5]Department of Ophthalmology & Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium. [6]Centre de Référence pour les Affections Rares en Génétique Ophtalmologique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France. [7]National Centre for Medical Genetics, Our Lady's Children's Hospital, Dublin, Ireland. [8]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden. [9]Genetic Sensory Diseases - Hopital Gui de Chauliac, Centre Hospitalier Universitaire, Montpellier, France. [10]Department of Ophthalmology and Visual Sciences, W. K. Kellogg Eye Center, University of Michigan, Ann Arbor, MI, USA. [11]Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands. [12]Department of Ophthalmology, University Medical Centre Groningen, University of Groningen, The Netherlands. [13]Genetic Health Services Victoria, Monash Medical Centre, Parkville, Australia. [14]National Eye Clinic for the Visually Impaired, Kennedy Center, Glostrup, Denmark. [15]Victorian Clinical Genetics Services, Murdoch Childrens Research Institute, Parkville, Victoria, Australia. [16]Department of Medical Genetics, University Medical Centre Groningen, University of Groningen, The Netherlands. [17]Department of Ophthalmic Genetics, Casey Eye Institute, Portland, OR, USA. Correspondence and requests for materials should be addressed to B.W. (email: wissinger@uni-tuebingen.de)

the human retina showed that only the two most proximal genes in the array are expressed at appreciable levels[5]. Duplicated genes are prone to unequal homologous recombination and gene conversion, the non-reciprocal transfer of genetic information in which a recipient sequence is replaced by a donor sequence that remains unaltered[6]. Concurrently, gene conversion and recombination themselves, enhance the interchange of DNA variants, thus reducing the divergence between duplicates while expediting a considerably high haplotype diversity[7]. In accordance with gene variability at the population level, both mechanisms are presumed to occur at the *LW/MW* gene cluster; nonetheless, they have never been directly observed.

Blue Cone Monochromacy (BCM, OMIM 303700) is an X-linked inherited cone dysfunction disorder. Affected subjects present with colour vision abnormalities and reduced visual acuity, but nystagmus and photophobia are also common[8]. In this condition S-cone function is retained, while the function of both L- and/or M-cones is absent. In the allelic condition, X-linked Cone Dystrophy (XLCD), cone function is strongly and progressively impaired. Such patients regularly present with myopia and astigmatism and dichromatic or monochromatic colour vision. The two most common genetic causes of BCM are large deletions encompassing the Locus Control Region (LCR) preventing the expression of the *LW/MW* genes[9,10] or the presence of either a single *LW/MW* hybrid opsin gene or multiple *LW/MW* opsin genes inactivated by the p.C203R missense mutation[9,11]. Very few additional disease-causing point mutations or small internal deletions in the *LW/MW* array have been reported[12–15].

A particular combination of common polymorphisms in *LW* exon 3 segregating in a single BCM family was first noted by Nathans and colleagues but not considered as disease-causing at that time[9]. This haplotype, referred to as 'LIAVA' in the literature, comprises the following SNPs: c.(453A > G; 457A > C; 465C > G; 511G > A; 513G > T; 521C; 532A > G; 538T > G) and a thereof deduced cone pigment variant with a certain combination of amino acid exchanges: p.[(=); M153L; (=); V171I; A174A; I178V; S180A]. Neither the spectral properties of this variant cone pigment nor its membrane trafficking is altered[16]. Still the 'LIAVA' haplotype has never been reported in individuals with normal colour vision but in patients with incomplete achromatopsia or XLCD[17,18] and is associated with widespread alterations of the retinal morphology and disorganized cone structure, different from findings in BCM patients carrying the p.C203R mutation[19]. In a recent study involving subjects with protan colour vision defects, the 'LIAVA' haplotype and other rare exon 3 haplotypes in *LW* were found to induce exon 3 skipping *in vitro*[20]. These findings have lately been corroborated and extended to patients with severe cone disorders including BCM and the identification of two further haplotypes, 'LVAVA' and 'MIAVA', that impair splicing[21].

Hitherto, the interplay of *cis*-regulatory elements and their contribution to exon 3 splicing is still unclear. To better ascertain to what extent different exon 3 haplotypes lead to splicing aberrations and thereby contribute to severe cone dysfunction disorders, we pursued a semi-quantitative assessment of transcripts from minigene assays performed on a comprehensive set of rare exon 3 haplotypes observed in a total of sixteen families diagnosed with BCM or XLCD. Gene conversion has been proposed as one mechanism underlying the formation of rare exon 3 haplotyopes. However, little is known about the specific features of such gene conversion in the human cone opsin genes. Taking advantage of a family with strikingly different ocular phenotype between the grandfather and grandson, we were able to identify an intrachromosomal *de novo* gene conversion event in the male germline which results in the replacement of a permissive haplotype by the strongly deleterious 'LIAVA' haplotype in the *LW* gene and explains the severe BCM phenotype in the grandson.

Materials and Methods

Patient recruitment and clinical evaluation. The study was performed in compliance with the tenets of the WMA Declaration of Helsinki. Study participants were recruited *ad hoc* at different centers specialized in inherited retinal diseases during routine clinical diagnostics. All participants gave written informed consent – approved by the respective local research and ethical review boards - for participation in the study for which blood or DNA samples were sent to Tuebingen for genetic analysis. Procedures of the genetic analysis were approved by the Ethics Committee of the Medical Faculty, Eberhard-Karls University Tuebingen. Patients underwent basic ophthalmologic examination according to the standards of the recruiting centers (for details see Supplementary information).

Genotyping of the *LW/MW* gene cluster. Genomic DNA was isolated from blood samples according to standard procedures. We analyzed the basic structure and integrity of the *LW/MW* gene cluster and the absence of the common p.C203R mutation by means of an established PCR and PCR/RFLP protocol. For those subjects having a structurally intact array, *LW* or *MW* specific long distance PCRs (LD-PCRs) were performed and LD-PCR products sequenced. For subjects with multiple gene copies, the total number of *LW/MW* opsin genes was determined by means of real-time quantitative PCR (for details see Supplementary information).

Minigene preparation. The prototype *LW* opsin minigene construct was kindly provided by Dr. Ueyama (Shiga University, Japan) and used to generate minigene constructs with *LW/MW* gene variants in exon 3, flanked by its native intronic sequences and the remaining human *LW* cDNA sequence (for details see Supplementary information).

Transfection and RNA extraction. HEK293 cells at 80–90% confluency were transfected with 4 μg DNA of the minigene construct using 20 μl Lipofectamine 2000 (ThermoFischer GmbH, Dreieich, Germany) per well. 24 h post-transfection, total RNA was extracted (for details see Supplementary information).

RT-PCR and relative quantification. First strand cDNA synthesis was performed using 2 μg of total RNA and random hexamer primers. Subsequent PCR was performed with a 5′ FAM (6-carboxyfluorescein) labeled forward primer and using the QIAGEN Multiplex PCR Kit (Qiagen, Hilden, Germany). FAM-labeled RT-PCR products were diluted 1:10 in water; mixed with 1 μl of GeneScan ROX500 size standard (Life Technologies, Darmstadt, Germany) and 8 μl of Hi-Di Formamide (Life Technologies) in a total volume of 10 μl. Mixes were

separated by capillary electrophoresis on an ABI 3130XL Genetic Analyzer instrument (Life Technologies). The area-under-the-curve (AUC) was calculated with GeneMapper 5 (Life Technologies) software. Ratios of splicing products were determined as the AUC for individual peaks divided by the sum of AUC of all differentially spliced products.

Microsatellite analysis. Microsatellite markers locating either centromeric (DXS8011, DXS8103, DXS1356, DXS8087) or telomeric (L441TA, L441CA, AF277A, AF277B and DXS1073) to the *LW/MW* cluster were genotyped in the three BCM72 family members. Markers for which the mother BCM72-II:1 was heterozygous were used to reconstruct haplotypes.

Mapping of the gene conversion event. LD-PCRs were performed for all three members of BCM72 for the proximal gene copy and for the distal copies. Upon cloning of digested LD-PCR products, we selected two independent clones of the following gene copies for further analysis: (i) *LW*-derived clones bearing the 'LIAVA' haplotype from subjects BCM72-II:1 and BCM72-III:1, (ii) *MW*-derived clones bearing the 'LIAVA' haplotype from subject BCM72-I:1, and (iii) *LW*-derived clones bearing the 'LIAVS' haplotype from subject BCM72-I:1. We sequenced the clones by primer walking (for details see Supplementary information).

Bioinformatic predictions and reference sequences. See Supplementary information.

Results

***LW/MW* genotypes of subjects with X-linked cone dysfunction disorders.** We genotyped the *LW/MW* gene cluster in twenty-four affected subjects from sixteen families diagnosed with BCM or XLCD not carrying mutations commonly reported in BCM. Eleven families had structurally intact arrays with a proximal *LW* gene followed by one (n = 6) or multiple copies of the *MW* gene (n = 5). The remaining five families harboured either a single *LW* (n = 3) or a single *LW/MW* hybrid gene (n = 2) (Fig. 1a).

Sequencing of *LW/MW* exons revealed a high proportion of different (n=12) rare exon 3 haplotypes (Fig. 1b). Table 1 depicts the LW/MW gene array composition and the exon 3 haplotypes for each subject. For consistency we designate haplotypes according to the amino acid residues they encode (i.e. 'LIAVA' for p.[L153-I171-A174-V178-A180]). Exon 3 haplotypes comprise two synonymous variants, c.453A > G and c.465C > G. While the c.453A allele was in strict linkage disequilibrium with c.457A/p.V153, we observed both alleles of the variant c.465C > G in *cis* with c.457A/p.V152. For ease, we distinguish these alleles hereafter with a superscript add-on where appropriate (e.g. MIAVA$^{c.465C}$ versus MIAVA$^{c.465G}$). Except for a missense mutation c.556C > T/p.P186S found in the *LW* gene of subject BCM142–21958, all other variants comprised in the haplotypes are *per se* common in the population of colour normal subjects (minor allele frequency; MAF > 0.05; 22, 23). A large fraction of these haplotypes is rare: seven out of the twelve haplotypes observed here (Table 2) were not present in the population sample investigated by Winderickx and colleagues[22]. In all our sixteen families the opsin gene harbouring such a rare haplotype occupies the most proximal position with respect to the LCR. In subjects ZD379–19194, BCM101–19818, and BCM126–20616 also the distal gene copies bear rare exon 3 haplotypes.

Minigene assay for rare exon 3 haplotypes. We analyzed exon 3 transcript processing for a total of twelve different exon 3 haplotypes observed in our patient cohort and the 'MVAIS' control haplotype by an established minigene assay[20]. RT-PCR showed three differently spliced products with relative quantities depending on the actual haplotype (Fig. 2a). The 450 bp product is derived from the correctly processed transcript, whereas the two smaller products are derived from aberrantly spliced transcripts either lacking exon 3 (281 bp) or lacking exon 3 and 72 bp of the 3′ end of exon 2 (214 bp), respectively (Fig. 2a). Both aberrant transcripts cause a frame shift leading to a premature termination codon (PTC) in exon 4. To quantify the splicing defect more precisely we performed RT-PCR with a FAM-labeled forward primer and separated the PCR products on a capillary electrophoresis instrument. From the AUC of fluorescence intensity for each fragment we calculated the relative abundance of individual fragments and correctly spliced RT-PCR products (Fig. 2b, Table 2). The control haplotype 'MVAIS' yielded no aberrantly spliced products while the 'LIAVA' construct resulted in the complete absence of correctly spliced products (Fig. 2b). Three haplotypes, 'LVAVA', 'MIAVA$^{c.465C}$' and 'MIAVA$^{c.465G}$' were found to yield only minor amounts (below ~10%) of correctly spliced products and were classified as strongly deleterious (+++, Table 2). An intermediate fraction (20–50%) of correctly spliced products was obtained with three further haplotypes, 'LIAIA', 'LIAVS' and 'MVAVA' and were considered intermediate (++) in terms of the magnitude of the splicing defect. The remaining five haplotypes, 'LVAIA', 'LVAISS', 'MVAIA', 'MVVVA$^{c.465G}$' and 'MVVVA$^{c.465C}$' yielded predominantly (>75%) correctly spliced products. (Fig. 2b, Table 2) and therefore were considered minor defective (+).

Correlation of *LW/MW* arrays and splicing defects with clinical presentation. *LW/MW* gene cluster genotypes were greatly diverse in our cohort of families with clinical diagnosis of BCM or XLCD. One family had an *LW/MW* gene array made up of a total of four *LW/MW* gene copies, four families had three copies, six families had two copies and five families carried a single *LW,* or *LW/MW* hybrid gene.

All families with single gene arrays carried strongly deleterious haplotypes: 'LIAVA' in BCM73 and BCM93, and 'LVAVA' in another three families (BCM66, BCM112, and BCM194). Clinical data were available for nine subjects from these families (Supplementary Table S1). Except for one very young subject (BCM194–25474), all these subjects had a best corrected visual acuity (BCVA) of ≤0.3, absent or strongly reduced photopic and flicker ERG responses, and impaired colour vision. Seven out of nine subjects had nystagmus and two experienced mild photophobia. Funduscopy revealed minor macular pigmentary irregularities but were dominated by myopia-related alterations such as optic disc crescents and peripapillary atrophy. All but two subjects in this

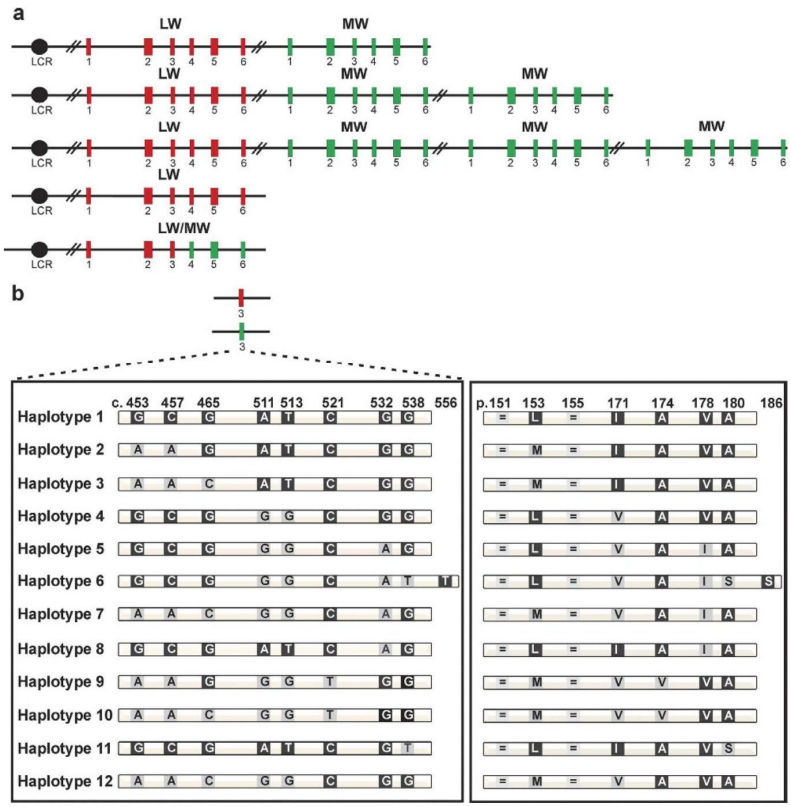


Figure 1. Structural diversity of the *LW/MW* gene array and exon 3 haplotypes in BCM and XLCD families analyzed in this study. (**a**) Overview of the various *LW/MW* gene arrays observed in the study cohort. The *LW/MW* gene cluster comprises the LCR, represented by a black circle, *LW* exons depicted as red and *MW* exons as green boxes, respectively. We observed single *LW* or *LW/MW* hybrid arrays and multiple gene arrays with one *LW* copy and one up to three *MW* copies. (**b**) The twelve different exon 3 haplotypes found in this cohort comprise eight common SNPs, depicted by the coding nucleotide position (left) and the corresponding variant amino acid residues (right) and a novel missense variant c.556C > T/p.P186S as part of haplotype 6.

group were highly myopic. Within this group we noted a tendency that subjects with the 'LIAVA' haplotype are more severely affected than those with the 'LVAVA' haplotype.

The families with multiple *LW/MW* opsin genes were more heterogeneous in terms of genotypes and clinical presentation. Clinical data were available for fifteen subjects from eleven families. Current genotyping technologies cannot determine the actual order of multiple distal *MW* gene copies. Therefore, for ordering of multiple *MW* gene copies we took into account impaired *MW* cone function in the patients along with the positional bias of the opsin gene expression[5]. Three families had identical *MW* gene copies with respect to exon 3 haplotypes whereas BCM142 and BCM72 had different *MW* gene copies (BCM142–21958, BCM72–17075 and BCM72–16874, Table 1). In the latter we assume that the *MW* gene copy harbouring the exon 3 haplotype with the most severe splicing defect occupies the second position in the array. Four families (ZD379, BCM101, BCM126 and BCM72–17075 [BCM72-III:1]) had highly deleterious (+++) haplotypes in all gene copies either being completely (BCM126) or partially isogenic (i.e. all copies with identical haplotypes or two of three gene copies sharing an identical haplotype, respectively). While all subjects in this group were clinically typical for BCM, myopia was rather mild and one was even slightly hyperopic.

The second sub-category of subjects with multiple copies are characterized by a *LW* gene comprising a strongly deleterious exon 3 haplotype and a single or multiple *MW* genes carrying exon 3 haplotypes causing intermediate or minor splicing defects. Although the clinical findings of the seven subjects from five families (BCM51, BCM133, BCM160, ZD314, and ZD547) were rather variable (namely visual acuity, photophobia and nystagmus), we consistently noted myopia, impaired colour vision and reduced but never absent cone or flicker ERG responses in these subjects. Consistent with some preserved central, cone-mediated vision there was at least for the better eye a BCVA of $\geq$0.5 in four of them.

The remaining subjects are less likely to be explained simply by the splicing defect due to exon 3 haplotypes: BCM98–19713 comprises a 'LIAIA' haplotype in the *LW* gene associated with a moderate splicing defect and a common haplotype with minimally compromised splicing in the *MW* gene. BCM142–21958 only had exon 3 haplotypes that result in minor amounts of aberrant transcripts, but carries a novel missense variant,

Index subjects	Exon 3					Clinical diagnosis
	LW		*MW*			
	Haplotype	p.	Haplotype	p.	n[d]	
ZD379-19194	1	LIAVA	2	MIAVA$^{c.465C}$	2	BCM
BCM101-19818	1	LIAVA	3	MIAVA$^{c.465G}$	2	BCM
BCM51-12359	1	LIAVA	9	MVVVA$^{c.465G}$	2	XLCD
BCM160-23130	1	LIAVA	9	MVVVA$^{c.465G}$	1	BCM
ZD314-18057	1	LIAVA	10	MVVVA$^{c.465C}$	1	XLCD
BCM126-20616	4	LVAVA	4	LVAVA	1	BCM
BCM133-20961	4	LVAVA	5	LVAIA	1	BCM
ZD547-4544	4	LVAVA	12	MVAVA	1	XLCD
BCM98-19713	8	LIAIA	7	MVAIA	1	XLCD
BCM142-21958	6	LVAIS[c]	5, 7	LVAIA, MVAIA	2	BCM
BCM72-17075[a]	1	LIAVA	1, 10	LIAVA, MVVVA$^{c.465C}$	3	BCM
BCM72-16874[a]	11	LIAVS	1, 10	LIAVA, MVVVA$^{c.465C}$	3	Deutan/Macular Dystrophy
BCM73-16953[b]	1	LIAVA	–	–	0	BCM/XLCD
BCM93-19164	1	LIAVA	–	–	0	XLCD
BCM66-16407	4	LVAVA	–	–	0	BCM
BCM112-23518[b]	4	LVAVA	–	–	0	CRD[e]
BCM194-25474	4	LVAVA	–	–	0	BCM

Table 1. *LW/MW* gene cluster composition and *LW* and *MW* exon 3 haplotypes in BCM and XLCD families analysed in this study. [a]Affected subjects from the same family with distinct genotypes. [b]Subjects harbouring single *LW/MW* hybrid genes. [c]This haplotype LVAIS includes an additional missense variant (c.556C > T/p.P186S; RefSeq: NM_020061.5) in exon 3 of the *LW* gene. [d]Number of *MW* gene copies deduced from qPCR. [e]Cone-Rod Dystrophy due to additionally impaired rod function (see Supplementary Table S1).

	Exon 3 (p.)[a]	Subjects[b]	% Correctly spliced[c]	SD[d]	Splicing defect[e]	Population Frequency (*LW*/*MW*)[f]
Control Haplotype[g]	MVAIS	–	100	–	–	0.027/0.008
Haplotype 1	LIAVA	9	n.d.	–	+++	0.0/0.0
Haplotype 2	MIAVA$^{c.465G}$	1	10.41	1.45	+++	0.0/0.0
Haplotype 3	MIAVA$^{c.465C}$	1	8.78	3.19	+++	0.0/0.0
Haplotype 4	LVAVA	6	6.71	0.27	+++	0.0/0.0
Haplotype 5	LVAIA	2	79.40	0.97	++	0.23/0.033
Haplotype 6	LVAIS[h]	1	98.73	1.10	+	0.0/0.0
Haplotype 7	MVAIA	1	97.62	0.19	+	0.094/0.6
Haplotype 8	LIAIA	1	40.75	0.23	++	0.0/0.0
Haplotype 9	MVVVA$^{c.465G}$	2	80.07	0.35	+	0.0/0.025
Haplotype 10	MVVVA$^{c.465C}$	3	75.59	0.65	+	0.027/0.21
Haplotype 11	LIAVS	1	20.3	0.0	++	0.0/0.0
Haplotype 12	MVAVA	1	53.0	0.0	++	0.013/0.016

Table 2. Relative quantification of the proportion of correctly spliced transcripts for different exon 3 haplotypes. [a]Exon 3 haplotypes referring to the encoded amino acid combination. [b]Number of subjects (index) carrying this haplotype. [c]Proportion of RT-PCR products from correctly spliced transcripts (450 bp). [d]Standard deviation (SD), calculated from technical triplicates. [e]Relevance of the splicing defect (−, no splice defect; +, more than 75% correctly spliced product; ++, 20–50% of correctly spliced product; +++, below ~10% of correctly spliced product). [f]Frequency of haplotypes in the population as reported by Winderickx *et al.*[22]. [g]Control haplotype as reported by Ueyama *et al.*[20]. [h]This haplotype LVAIS includes an additional missense variant (c.556 C > T/p.P186S; RefSeq: NM_020061.5) in exon 3 of the *LW* gene.

c.556C > T/p.P186S, in exon 3 of the *LW* opsin gene which does not compromise splicing, but may impair folding and/or function of the derived photopigment. Notably, an analogous proline to serine substitution at amino acid position 187 has been reported previously in the *LW* opsin gene of a deuteranope subject[24].

Gene conversion results in the deleterious haplotype 'LIAVA'. We observed a family, BCM72, with rather discordant phenotypes (Fig. 3) between the index subject (BCM72-III:1, Fig. 4a) diagnosed with BCM, and his grandfather (BCM72-I:1, Fig. 4a) who developed macular dystrophy at the age of 40 and presented with a deutan defect. A crucial difference is documented in the cone-derived ERG recordings: Responses in BCM72-I:1

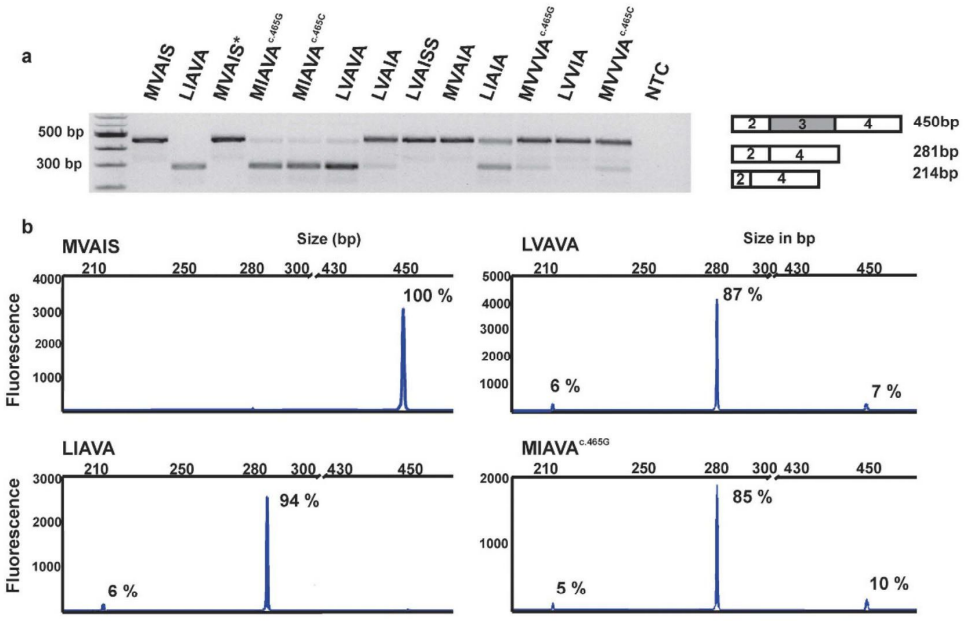


Figure 2. Qualitative and quantitative analysis of RT-PCR products from minigene splicing assays. (**a**) Agarose gel electrophoresis of RT-PCR products obtained with RNA from HEK293 cells transfected with minigene constructs bearing various exon 3 haplotypes. The tested haplotype is given above the corresponding gel lane. A 100 bp ladder size standard was loaded in the leftmost lane. Both lanes 'MVAIS' and 'MVAIS*' refer to minigenes carrying the control haplotype. 'MVAIS*' has a modified Multiple Cloning Site from the prototype construct 'MVAIS' (see Supplementary Materials and Methods). NTC: non-template negative control. A scheme on the composition of the RT-PCR products is given on the right. Full length gel picture is presented in Supplementary Fig. S3. (**b**) Examples of capillary electrophoresis and quantitative analysis of fluorescent labeled RT-PCR products of the minigene assay for four different haplotypes ('MVAIS', 'LVAVA', 'LIAVA' and 'MIAVA$^{c.465G}$'). The fragment size scale is given on the x-axis and fluorescence intensity (in arbitrary units) on the y-axis. Relative amounts of each fragment are given for the corresponding peak as determined by Gene Mapper. The three different sized products correspond to correctly spliced transcripts (450 bp), aberrantly spliced transcripts lacking exon 3 (281 bp), and a minor species of aberrantly spliced products lacking exon 3 and further 72 bp of exon 2 (214 bp).

were not impaired, whereas in BCM72-III:1 photopic single-flash and 30 Hz-flicker-responses were extinguished as shown by a flat light-adapted (LA) ERG (Fig. 3a). Rod-derived responses were normal in both cases (Fig. 3a). Subject BCM72-I:1 performed as a deuteranomalous in the Farnsworth Panel D-15 desaturated test while BCM72-III:1 had a majority of confusions along the protan axis (Fig. 3b). A protan-like arrangement of colour discs in the Panel D-15 is frequently observed in BCM subjects[8,25]. The index subject's mother (BCM72-II:2) had normal visual acuity and colour vision but borderline reduced photopic ERG responses. The macular dystrophy of subject BCM72-I:1 is evident in the fundus autofluorescence image of his left eye (Fig. 3c, upper image). OCT of the left eye of both subjects is shown in Fig. 3d. Retinal pigment epithelium and mostly the outer cone photoreceptor layers, namely the cone outer segment termination and the inner segment ellipsoid layers, were severely damaged in the macular area in subject BCM72-I:1, but the retinal thickness is also diminished in subject BCM72-III:1.

Genotyping of the grandson BCM72-III:1 revealed an intact *LW/MW* gene array with a proximal *LW* gene and three copies of the *MW* gene (Fig. 4b). Sequence analysis showed a 'LIAVA' exon 3 haplotype for the *LW* gene and 'LIAVA' and 'MVVVA$^{c.465C}$' for the *MW* gene copies. No further mutations were found in other exons (including exon 6, which is not covered by the LD-PCRs) of the opsin genes of the affected subjects. Considering the expression bias of the opsin gene cluster, and the BCM phenotype of the subject, we reasoned that the 'LIAVA'-bearing *MW* gene copy occupies the first position downstream of the *LW* gene. Surprisingly, the same downstream *MW* haplotypes but a distinct *LW* exon 3 haplotype 'LIAVS' were found in the grandfather BCM72-I:1. Applying the minigene splicing assay we found that the 'LIAVS' haplotype resulted in an intermediate splicing defect with still approximately 20% correctly spliced transcripts (Fig. 5b).

Segregation analysis with microsatellite markers flanking the *LW/MW* gene cluster supported the transmission of the Xq28 segment from grandfather to grandson without evidence for recombination (Fig. 4 and Supplementary Fig. S1), suggesting a *de novo* gene conversion in the *LW/MW* gene cluster in this family. We further explored whether the gene conversion event has occurred in *trans* between X-chromosomes (in the mother's germline), or in *cis* between *LW* and *MW* gene copies in either the grandfather's or the mother's germline.

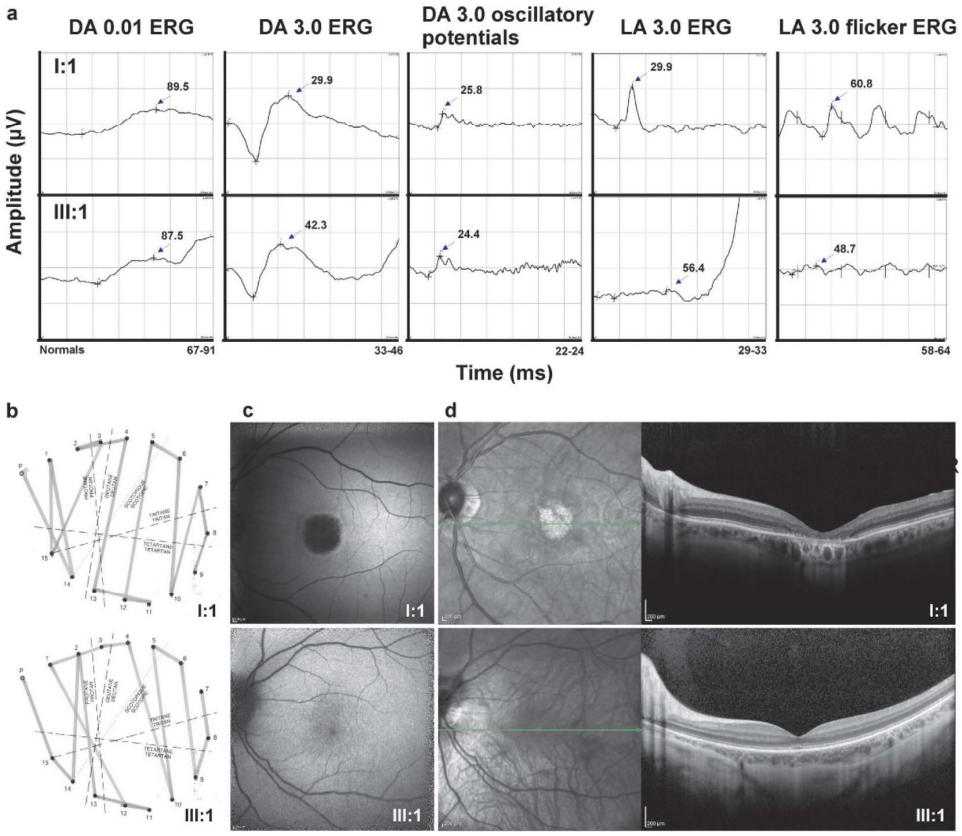


Figure 3. Clinical findings in family BCM72 with strikingly different phenotypes in the grandfather (I:1) and his grandson (III:1). (**a**) Fullfield-ERG with nearly normal responses in BCM72-I:I and not detectable photopic responses in BCM72-III:1. DA: dark-adapted, LA: light-adapted, stimulus strengths: 0.01 or 3.0 cd.s.m^{-2}. (**b**) Panel D-15 desaturated with protan defects in BCM72-I:I, and deutan defects in BCM72-III:I. (**c**) Autofluorescence and (**d**) infrared picture (left panel) and OCT (right panel) with macular dystrophy in BCM72-I:I and normal retinal architecture with thinned photoreceptor layer in BCM72-III:I.

Genotyping of the mother BCM72-II:2 revealed a common 'LIAIS' exon 3 haplotype in one of her proximal *LW* genes and the 'LIAVA'-bearing haplotype in the other *LW* gene copy (Fig. 4b). This finding strongly supports transmission of the 'LIAVA'-bearing *LW* gene from her father (BCM72-I:1) and from BCM72-II:2 to her offspring and thereby asserts the occurrence of the gene conversion in the grandfather's germline as a *de novo* mutation. To define the extent of this intrachromosomal event we sequenced exon 3 and the flanking introns of the putative gene conversion donor ('LIAVA'-bearing *MW* gene copy) and recipient gene copy ('LIAVS'-bearing *LW* gene copy) in BCM72-I:1 and its product ('LIAVA'-bearing *LW* gene copy) in both BCM72-II:2 and BCM72-III:1. By comparative sequencing we narrowed down the size of the maximal converted sequence in the recipient *LW* to a region of 1,297 bp (c.409 + 950_578 + 90conNM_000513.2:c.409 + 950_578 + 90), which is delimited by the discriminative SNP rs3788802 (c.409 + 949 G > A) in intron 2 and rs369018729 (c.578 + 91 G > A) in intron 3, and includes the discriminating variant c.538 T > G/p.S180A/rs949431 in exon 3 (Fig. 4a and Supplementary Fig. S2).

Discussion

Splicing defects caused by rare *LW/MW* exon 3 haplotypes. The presence of rare combinations (foremost 'LIAVA') of otherwise common coding variants in exon 3 of the *LW/MW* genes in subjects with X-linked colour vision or cone dysfunction disorders has been noted in several publications[9,18,24]. Ueyama and colleagues were the first who showed that such rare combinations of variants induce a splicing defect and hence, explain the protan defect in probands with otherwise normal *LW* gene sequence[20]. Lately – and during the course of our study – Gardner *et al.* reported *LW/MW* gene splicing defects as underlying disease mechanism in nine BCM/XLCD families carrying exon 3 "interchange haplotypes"[21]. These patients had only a single *LW/MW* gene copy carrying an exon 3 haplotype inducing a splicing defect or multiple opsin gene copies, in which at least the two most proximal carry such haplotypes. In this study we could corroborate their findings to a broader extent by including a large series of subjects with BCM or XLCD (24 affected subjects from 16 families). We observed an extensive diversity in terms of structure and composition of the *LW/MW* gene cluster and exon 3 haplotypes. Besides, we explored the functional consequences of twelve exon 3 haplotyes on RNA processing by means of

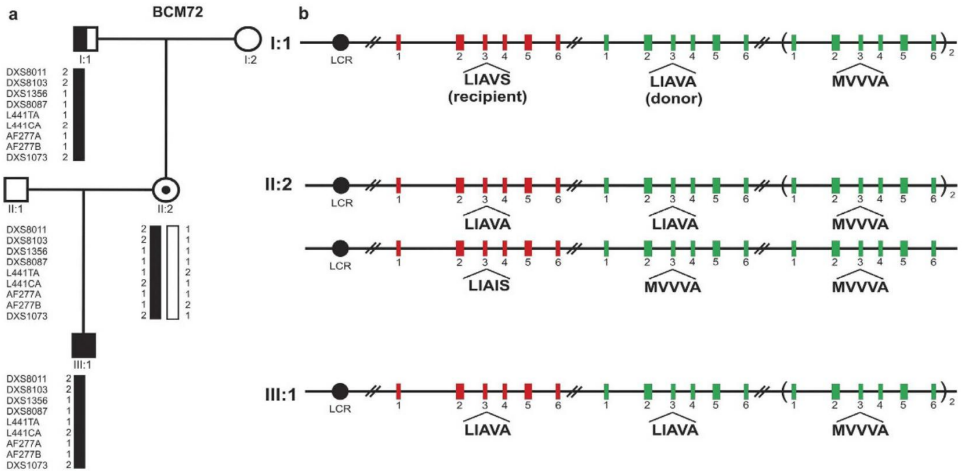


Figure 4. Gene conversion at the *LW/MW* opsin gene cluster in family BCM72. (**a**) Pedigree of family BCM72. Subject BCM72-I:1 presented with macular dystrophy and deuteranopia, subject BCM72-II:2 is an asymptomatic female carrier and subject BCM72-III:1 was diagnosed with BCM. Reconstructed haplotypes based on microsatellite markers encompassing the *LW/MW* gene cluster on Xq28 revealed inheritance of the X-chromosome from the grandfather to the grandson with no evidence for recombination. (**b**) Scheme of the structure of the *LW/MW* gene array and the *LW* and *MW* exon 3 haplotypes in crucial members of the BCM72 family. The *LW/MW* gene array on the transmitted chromosome comprises the locus control region (LCR), a single *LW* gene and three *MW* gene copies. Red and green coloured and numbered boxes represent *LW* exons and *MW* exons, respectively. Exon 3 haplotypes are indicated below the respective exon boxes. For multiple distinct *MW* copies, their most likely order with respect to exon 3 haplotypes is depicted. Note that female subject BCM72-II:2 is heterozygous for *LW* exon 3 haplotypes 'LIAVA' and 'LIAIS', the latter inherited from her mother. The findings support an intrachromosomal gene conversion event transforming the more ancestral haplotype 'LIAVS' into the 'LIAVA' haplotype in the germline of subject BCM72-I:1.

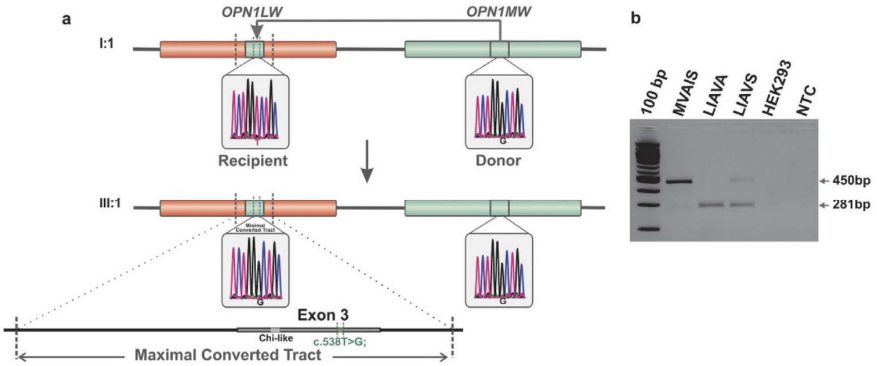


Figure 5. Gene conversion replaces a permissive 'LIAVS' haplotype by a strongly deleterious 'LIAVA' haplotype in the *LW* gene. (**a**) The gene conversion occurred in a 'telomeric to centromeric' direction in the X-chromosome of subject BCM72-I:1 with the 'LIAVA' bearing *MW* gene serving as donor and the ancestral 'LIAVS' bearing *LW* gene serving as recipient. Below, the X-chromosome of subject BCM72-III:1 presents the product of the gene conversion event with the 'LIAVA' haplotype present on both *LW* and the most proximal *MW* as demonstrated by corresponding sequences traces for the c.538T > G variant. Sequencing of cloned *LW* and *MW* gene fragments revealed a maximal converted region of 1,297 bp (c.409 + 950_c.578 + 90conNM_000513.2: c.409 + 950_c.578 + 90) covering exon 3 and flanking intron sequences (see also Supplementary Fig. S2). (**b**) Direct comparison of RT-PCR products from the minigene splicing assays shows a substantial amount of correctly spliced transcripts (450 bp) for the 'LIAVS' exon 3 haplotype, whereas such products are undetectable for the 'LIAVA' haplotype. Full length gel picture is presented in Supplementary Fig. S4.

heterologous minigene assays and took advantage of fluorescence-based capillary electrophoresis of RT-PCR products for accurate relative quantification of the proportions of differentially spliced transcripts. Most of the investigated haplotypes resulted in a substantial fraction of mis-spliced transcript with two main aberrant mRNA

species that result in a PTC in exon 4. Such transcripts presumably elicit nonsense-mediated mRNA decay[26]. Accurate quantification of the relative amounts of the different species disclosed clearly distinct ratios of correctly and aberrantly spliced transcripts depending on the haplotype. 'LIAVA' as the extreme led to complete exon skipping. The two related haplotypes 'MIAVA$^{c.465C}$' and 'MIAVA$^{c.465G}$', as well as 'LVAVA', yielded high levels of aberrantly spliced transcripts but still detectable amounts of correctly spliced transcripts. Notably, we observed the latter haplotype being associated with a less severe clinical phenotype in comparison with patients harbouring 'LIAVA'. Three further haplotypes, 'LIAVS', 'LIAIA' and 'MVAVA' have an intermediate effect in terms of aberrant splicing (20–50% of correctly spliced opsin), whereas the remaining six haplotypes confer only minor splicing defects. The (patho-)physiological consequences of such intermediate or minor defects on disease expression or colour in general still needs to be explored.

The consequences of genotypes that comprise highly deleterious exon 3 haplotypes in the relevant, most proximal gene copies may approximate the situation of a full *null* mutation like patients with larger deletions at the *LW/MW* gene cluster[9,12–14]. Less straightforward are cases with multiple gene copies combining copies with highly deleterious exon 3 haplotypes and copies with haplotypes causing intermediate splicing defects (e.g. *LW*: 'LVAVA'+*MW*: 'MVAVA' as in ZD547). The diversity of genotypes and quantitative difference in the magnitude of the splicing defect caused by different haplotypes is reflected by the variability in clinical presentations ranging from a BCM typical phenotype to cone disorders with considerably preserved cone function and rather good visual acuity. This clinical variability is in line with previous reports on subjects with BCM, "atypical BCM", or XLCD who carry rare *LW/MW* exon 3 haplotypes[9,16–18].

Since accurate quantification of the different spliced forms is lacking in prior studies, we could not directly compare their results at the quantitative level. However, a qualitative comparison points out some differences. While in this study as well as in the study of Gardner and colleagues[21] the three distinct transcript species were observed, the correctly spliced form, an exon 3-skipped transcript and an additional aberrant transcript, the nature of the latter one seems to differ. This third minor species is produced for instance from the 'LVAVA' and 'MIAVA' haplotypes (and from 'LIAVA' in our study) which in both studies is a result of internal exon 2 splicing but involves different acceptor sites in intron 3 (in the study of Gardner and colleagues) and at the correct intron 3/exon 4 junction (in this study), respectively. Microheterogeneity in terms of minor aberrant splicing products have not only been observed in many heterologous minigene assays but also in native tissue of subjects carrying splicing-inducing mutations[27] and may be influenced by technical parameters.

SROOGLE[28] predicted several Splicing Regulatory Sequences (SRSs) to be disrupted and/or created when exon 3 SNPs from deleterious haplotypes were interchanged. In fact, every other SNP within the haplotype disrupts or creates at least one SRS. Half of the overall putative Exonic Enhancers of Splicing (EESs) overlapping the eight SNPs were predicted to be disrupted by 'LIAVA'. For instance, the c.521C > T substitution is foreseen to create a novel EES, which is consistent with the increased correctly spliced transcripts abundance *in vitro* (Table 2, haplotype 9–10). Comparison of closely related haplotypes indicates that the c.532A/p.178Ile allele confers "protection": for instance 'LVAVA' yielded about 7% of correctly spliced transcripts whereas 'LVAIA' about 80% (Table 2, haplotype 4–5). These patterns may suggest the presence of motifs that exert combined enhancer and silencer effects to a certain extent as described in *SMN1* and *CFTR* genes[29].

Gene conversion as a mechanism leading to a pathogenic haplotype in the cone opsin array. We here report a *de novo* gene conversion event in the *LW/MW* gene cluster which we proved in the lineage of a single family (BCM72). This rare instance further enabled us (i) to discriminate between inter- and intrachromosomal event, (ii) to differentiate between its occurrence in either the male versus the female germline, (iii) to define the directionality of the event (from *LW* to *MW* or *vice versa*), (iv) to determine the actual sequence homology between the donor and recipient prior to the gene conversion event and finally (v) to refine the extent of the converted sequence. These findings are novel since no other *de novo* gene conversion events in a parent-offspring transmission have been reported for the *LW/MW* gene cluster so far.

Our study is in congruency with the population-based evidence for gene conversion in the opsin genes as described by Verrelli and Tishkoff[23]. Gene conversion together with selective forces are proposed to have an influence on the observed variation in *LW/MW* genes, as mainly advantageous polymorphisms have been spread[6,23]. However, beyond the evolutionary perspective, gene conversion in the *LW/MW* gene cluster may also lead to the occurrence of deleterious genotypes that are associated with colour vision deficiencies, XLCD or BCM.

Signatures of gene conversion in the *LW/MW* gene cluster at the level of individual chromosomes have so far been proposed specifically in the context of BCM subjects with rare point mutations present in all copies of the *LW/MW* gene cluster[13,30], yet the gene conversion event could not be directly pinpointed, hindering a full characterization. For instance, Gardner and colleagues reported a BCM family in which the c. 529T > C/p. W177R missense mutation in exon 3 was observed in both *LW* and *MW* genes[13]. Albeit a gene conversion event was deduced based on the observation that both genes carry this unique mutation and share a block of SNPs in exon 3, occurrence of the gene conversion event (i.e. a family member with the ancestral genotype prior to the event) could not be observed in this family.

Different from these previous reports, we herein were able to study for the first time a gene conversion event at the *LW/MW* gene cluster within the lineage of a single pedigree (i.e. the genotypes prior to and following the gene conversion event) and, importantly, to associate these two different genotypes to their cognate phenotypes. Furthermore, we could correlate these differences at the phenotypic level prior and post-gene conversion with a quantitative difference in the *in vitro* levels of correctly spliced *LW* gene transcripts that results from the distinct exon 3 haplotypes prior and post gene conversion. While 'LIAVA' has been previously assessed for splicing[20], the ancestral prior gene conversion 'LIAVS' haplotype has been reported in the literature[18] but has – to our knowledge – never been tested for splicing before.

With respect to directionality of the gene conversion, we could demonstrate in our family that the *MW* copy acted as donor and *LW* copy served as recipient gene copy. The same 'telomeric to centromeric' direction has been proposed for the "spreading" mutation cases[13,30]. Seemingly, the position of a recombination hotspot has an influence over the directionality of gene conversion[31]. The donor *MW* copy carries the haplotype 'LIAVA', which most likely occupies the first position downstream of the *LW* gene, explaining the deutan defect in BCM72-I:1. The *LW* gene of this subject carries the exon 3 haplotype 'LIAVS', which only differs from 'LIAVA' by a single SNP (c.538G > T/p.S180A). Compared to the latter haplotype which results in a fully penetrant splicing defect, the 'LIAVS' haplotype produces 20% of correctly spliced transcripts which is perfectly compatible with the much milder phenotype of the grandfather (BCM72-I:1). In contrast, the grandson (BCM72-III:1) presenting with BCM has the two first *LW/MW* copies carrying the fully deleterious haplotype 'LIAVA', consistent with his BCM phenotype. The ERG responses were the most informative clinical feature that documents the phenotypic differences between BCM72-I:1 and BCM72-III:1. The 'LIAVS' haplotype has been reported before in affected males of two families with X-linked cone-dominated phenotypes[18], who compared to the 'LIAVA'-carrying individuals from an independent family seemed to have a better visual performance. One of the 'LIAVS'-carrying individuals presented with evidence of maculopathy, similarly to BCM72-I:1 here, however, BCM72-I:1 developed this maculopathy later in life.

Given the sequence identity between *LW* and *MW*, we could only define the outermost borders of the gene conversion event in intron 2 and intron 3 (Fig. 5a) setting a maximal size of about 1,300 bp which is beyond the average size estimates for converted tracts in humans[32]. Significantly reduced linkage disequilibrium of variants in a region of about 400 bp covering *LW* exon 3 and flanking intronic sequences supported the presence of a hotspot for gene conversion at this region[23] centering around a *Chi*-like sequence element[22]. Although gene conversion at a certain site is expected to be a rare event, recent single cell analysis technologies may allow to actually determining the frequency and pattern of such events at the *LM/MW* locus in human sperm cells[33].

While formally a *de novo* point mutation could be an alternative explanation, we reason that a *de novo* gene conversion is the most likely event since (i) the very same c.538T > G variant is present in the downstream *MW* gene copy, (ii) the c.538T > G is a common variant frequently observed in *LW* genes[22], (iii) population studies have revealed strong evidence for frequent gene conversion between *LW* and *MW* genes[23,34], and (iv) it has been estimated that the *de novo* rate of gene conversion between paralogous loci in the human genome is approximately 100-fold higher than the *de novo* point mutation rate[35–38].

The current study, not only supports the predictions of population genetics studies and "spreading" mutation cases on gene conversion in the opsin genes, but also, adds direct evidence on how gene conversion events may arise spontaneously in the germline having an important impact on human disease. Moreover, this type of mutations might be overlooked by employing next-generation sequencing technologies which complicates its identification as well as its prediction.

In conclusion, our study confirms that *LW/MW* gene array genotypes bearing certain rare exon 3 haplotypes cause BCM and XLCD in a large patient cohort and explains the deleterious functional effect of these haplotypes. Reliable quantitative analysis of splice products has been implemented and revealed that the relative proportions of correctly and aberrantly spliced transcripts obtained for the twelve different haplotypes cluster in three main categories. Lastly, we traced back the origin of the strongly deleterious 'LIAVA' haplotype in a BCM subject. This haplotype was inherited from his mother and, in all likelihood arose as a result of a *de novo* intrachromosomal gene conversion event from an ancestral 'LIAVS' in the grandfather's germline and explains the phenotypic transformation from deuteranopia in the grandfather to BCM in his grandson.

References

1. Vollrath, D., Nathans, J. & Davis, R. W. Tandem array of human visual pigment genes at Xq28. *Science* **240,** 1669–1672 (1988).
2. Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B. & Hogness, D. S. Molecular genetics of inherited variation in human color vision. *Science* **232,** 203–210 (1986).
3. Macke, J. P. & Nathans, J. Individual variation in size of the human red and green visual pigment gene array. *Invest Ophthalmol Vis Sci* **38,** 1040–1043 (1997).
4. Wolf, S. *et al.* Direct visual resolution of gene copy number in the human photopigment gene array. *Invest Ophthalmol Vis Sci* **40,** 1585–1589 (1999).
5. Winderickx, J., Battisti, L., Motulsky, A. G. & Deeb, S. S. Selective expression of human X chromosome-linked green opsin genes. *Proc Natl Acad Sci USA* **89,** 9710–9714 (1992).
6. Chen, J. M., Cooper, D. N., Chuzhanova, N., Ferec, C. & Patrinos, G. P. Gene conversion: Mechanisms, evolution and human disease. *Nat Rev Genet* **8,** 762–775 (2007).
7. Innan, H. & Kondrashov, F. The evolution of gene duplications: Classifying and distinguishing between models. *Nat Rev Genet* **11,** 97–108 (2010).
8. Michaelides, M. *et al.* Blue cone monochromatism: A phenotype and genotype assessment with evidence of progressive loss of cone function in older individuals. *Eye (Lond.)* **19,** 2–10 (2005).
9. Nathans, J. *et al.* Molecular genetics of human blue cone monochromacy. *Science* **245,** 831–838 (1989).
10. Smallwood, P. M., Wang, Y. & Nathans, J. Role of a locus control region in the mutually exclusive expression of human red and green cone pigment genes. *Proc Natl Acad Sci USA* **99,** 1008–1011 (2002).
11. Kazmi, M. A., Sakmar, T. P. & Ostrer, H. Mutation of a conserved cysteine in the X-linked cone opsins causes color vision deficiencies by disrupting protein folding and stability. *Invest Ophthalmol Vis Sci* **38,** 1074–1081(1997).
12. Nathans, J. *et al.* Genetic heterogeneity among blue-cone monochromats. *Am J Hum Genet* **53,** 987–1000 (1993).
13. Gardner, J. C. *et al.* X-linked cone dystrophy caused by mutation of the red and green cone opsins. *Am J Hum Genet* **87,** 26–39 (2010).
14. Ladekjaer-Mikkelsen, A. S., Rosenberg, T. & Jorgensen, A. L. A new mechanism in blue cone monochromatism. *Hum Genet* **98,** 403–408 (1996).
15. Kellner, U. *et al.* Blue cone monochromatism: Clinical findings in patients with mutations in the red/green opsin gene cluster. *Graefes Arch Clin Exp Ophthalmol* **242,** 729–735 (2004).
16. McClements, M. *et al.* Variations in opsin coding sequences cause x-linked cone dysfunction syndrome with myopia and dichromacy. *Invest Ophthalmol Vis Sci* **54,** 1361–1369 (2013).

17. Crognale, M. A. *et al.* Characterization of a novel form of X-linked incomplete achromatopsia. *Vis Neurosci* **21,** 197–203 (2004).
18. Mizrahi-Meissonnier, L., Merin, S., Banin, E. & Sharon, D. Variable retinal phenotypes caused by mutations in the X-linked photopigment gene array. *Invest Ophthalmol Vis Sci* **51,** 3884–3892 (2010).
19. Carroll, J. *et al.* The effect of cone opsin mutations on retinal structure and the integrity of the photoreceptor mosaic. *Invest Ophthalmol Vis Sci* **53,** 8006–8015 (2012).
20. Ueyama, H. *et al.* Unique haplotype in exon 3 of cone opsin mRNA affects splicing of its precursor, leading to congenital color vision defect. *Biochem Biophys Res Commun* **424,** 152–157 (2012).
21. Gardner, J. C. *et al.* Three Different Cone Opsin Gene Array Mutational Mechanisms with Genotype–Phenotype Correlation and Functional Investigation of Cone Opsin Variants. *Hum Mutat* **35,** 1354–1362 (2014).
22. Winderickx, J., Battisti, L., Hibiya, Y., Motulsky, A. G. & Deeb, S. S. Haplotype diversity in the human red and green opsin genes: Evidence for frequent sequence exchange in exon 3. *Hum Mol Genet* **2,** 1413–1421 (1993).
23. Verrelli, B. C. & Tishkoff, S. A. Signatures of selection and gene conversion associated with human color vision variation. *Am J Hum Genet* **75,** 363–375 (2004).
24. Neitz, M. *et al.* Variety of genotypes in males diagnosed as dichromatic on a conventional clinical anomaloscope. *Vis Neurosci* **21,** 205–216 (2004).
25. Weiss, A. H. & Biersdorf W. R. Blue cone monochromatism. *J Pediatr Ophthalmol Strabismus* **26,** 218—223 (1989).
26. Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rew Mol Cell Biol* **5,** 89–99 (2004).
27. Steffensen, A. Y. *et al.* Functional characterization of BRCA1 gene variants by mini-gene splicing assay. *Eur J Hum Genet* **22,** 1362–1368 (2014).
28. Schwartz, S., Hall, E. & Ast, G. SROOGLE: Webserver for integrative, user-friendly visualization of splicing signals. *Nucleic Acids Res* **37,** 189–192 (2009).
29. Pagani, F. & Baralle, F. E. Genomic variants in exons and introns: Identifying the splicing spoilers. *Nat Rev Genet* **5,** 389–396 (2004).
30. Reyniers, E. *et al.* Gene conversion between red and defective green opsin gene in blue cone monochromacy. *Genomics* **29,** 323–328 (1995).
31. Hallast, P., Nagirnaja, L., Margus, T. & Laan, M. Segmental duplications and gene conversion: Human luteinizing hormone/chorionic gonadotropin β gene cluster. *Genome Res* **15,** 1535–1546 (2005).
32. Reiter, L. T. *et al.* Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. *Am J Hum Genet* **62,** 1023–1033 (1998).
33. Bhang, H. E. *et al.* Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med* **21,** 440–448 (2015).
34. Zhou, Y. H. & Li, W. H. Gene conversion and natural selection in the evolution of X-linked color vision genes in higher primates. *Mol Biol Evol* **13,** 780–783 (1996).
35. Kong, A. *et al.* Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488,** 471–475 (2012).
36. Campbell, D. *et al.* Estimating human mutation rate using autozygosity in a founder population. *Nat Genet* **44,** 1277–1281 (2012).
37. Dumont, B. L. & Eichler, E. E. Signals of historical interlocus gene conversion in human segmental duplications. *PLoS One* **8,** e75949 (2013).
38. Dumont, B. L. Interlocus gene conversion explains at least 2.7% of single nucleotide variants in human segmental duplications. *BMC Genomics* **16,** 456 (2015).

Acknowledgements

The authors like to thank Dr. Ueyama (Shiga University, Japan) for providing the minigene construct. This work was supported by the European Union's Seventh Framework Programme for research, technological development and demonstration [grant agreement no 317472], and by Dr. Renata Sarno (BCM Families Foundation). E.D.B. and B.P.L. are senior clinical investigators of the Research Foundation Flanders (FWO), and are further supported by FWO grant number OZP 3G0C6715. We further acknowledge support by the Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of the University of Tübingen.

Author Contributions

E.B.-A. and B.B. performed the experiments. K.R., R.B., D.B., E.D.B., H.D., M.T.G., P.G., C.P.H., J.R.H., B.P.L., A.S.P., J.W.R.P., K.R., T.R., Z.S., J.B.G.M.V., R.W. and D.Z. recruited the families and collected the clinical data. N.W., S.K. and B.W. supervised the study and reviewed the data. E.B.-A., S.K. and B.W. drafted the manuscript. All authors read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Buena-Atienza, E. *et al. De novo* intrachromosomal gene conversion from *OPN1MW* to *OPN1LW* in the male germline results in Blue Cone Monochromacy. *Sci. Rep.* **6**, 28253; doi: 10.1038/srep28253 (2016).

Supplementary Materials and Methods

Patient recruitment and clinical evaluation

Subjects diagnosed either with BCM, XLCD or similar disorders were recruited at different clinical centers. The study was performed in compliance with the tenets of the WMA Declaration of Helsinki. Study participants were recruited ad hoc at different centers specialized in inherited retinal diseases during routine clinical diagnostics. All participants gave written informed consent – approved by the respective local research and ethical review boards - for participation in the study for which blood or DNA samples were sent to Tuebingen for genetic analysis. Procedures of the genetic analysis were approved by the Ethics Committee of the Medical Faculty, Eberhard-Karls University Tuebingen. Venous blood samples were collected as well as family history. Patients underwent basic ophthalmologic examination which included evaluation of ocular motility and slit lamp examination of the anterior segment, visual acuity (VA), refraction, colour vision and visual field testing, assessment of retinal morphology applying imaging technologies such as funduscopy, optical coherence tomography (OCT), autofluorescence imaging, and ERG recordings under scotopic and photopic conditions.

Genotyping of the *MW/LW* gene cluster

Genomic DNA was isolated from blood samples according to standard procedures. The structure and integrity of the *LW*/*MW* gene cluster (e.g. single *versus* multiple opsin gene copies, presence of hybrid genes) on the X-chromosome was investigated by a two-step genotyping protocol. First, we used a protocol that screens for mutations most commonly found in BCM/XLCD patients and specifically tests for large genomic deletions at the *OPN1LW*/*OPN1MW* gene cluster and the upstream locus control region (LCR), the presence of single or multiple opsin gene copies, and the presence or absence of the common point mutation c.607T>C, p.C203R. Briefly, duplex PCRs including primers for the tested fragment and an autosomal control fragment were performed for: 1) a sequence defined as the common overlap of known deletions of the LCR (Nathans et al., 1989, 1993); 2) the *OPN1LW* promoter; and 3) the *OPN1MW* promoter to test for large genomic deletions and the presence of single or multiple opsin gene copies. In addition, two PCR fragments for exons 4 and 5, respectively, for both *OPN1LW* and *OPN1MW* were amplified and digested with B*stUI* and R*saI*, respectively. The restriction fragment length polymorphism (RFLP) in exon 4 via B*stUI* analysed for the c.607T>C p.C203R mutation (RFLP positive cases are further validated by Sanger sequencing), while the RFLP in exon 5 via R*saI* addressed the dimorphism at amino acid position 277 – p.Y277 being specific for the *OPN1LW* gene, and p.F277 for the *OPN1MW* gene. The second screening step analyzed for other deleterious point mutations or haplotypes for those subjects either single or multiple structurally intact gene copies by PCR and Sanger-sequencing. Specific long distance PCRs (LD-PCRs) were performed with primers LCR1F (5'-CACCCTTCTGCAAGAGTGTGGG-3') and RGCP2-5R (5'-GCGGACTAGTGATCTGCTGATGGTGTTGCTTA-3') for the proximal gene copy, and E1G-F (5'-GAGTACAGGTATTTGCCACTAAGC-3') and RGCP2-5R for distal gene copies. LD-PCR products were directly sequenced with internal primers or re-amplified with exon-specific primers prior to sequencing. PCR products were purified by ExoSAP-IT treatment (Affymetrix, Santa Clara, CA) and cycle-sequenced applying BigDye Terminator V1.1 chemistry (Life Technologies, Darmstadt, Germany). For the determination of exon 3 haplotypes we amplified a 400 bp fragment using primers BCM-Ex3-F (5'-TGGTGGAAAGAAAGATGTCG-3') and BCM-Ex3-R (5'-GCCCAGAGAAAGGAAGTGATT-3') that was subsequently sequenced directly or after an intermitting cloning step if multiple mixed base positions were observed. For subjects with multiple gene copies, the total number of *LW/MW* opsin genes was determined by means of real-time quantitative PCR (qPCR) with genomic DNA as template. We used two different TaqMan assays that target different parts of the *LW/MW* genes: the HS_01912094 assay (Life Technologies) targeting exon 6, and a custom-designed TaqMan assay (employing RGCP_TQF [5'-CCCAACAGAAAGCTGAAAGC-3'] as forward and RGCP_TQR [5'-GTGCAAAACTTTCGGATTGG-3'] as reverse primers, respectively, and RGCP_TQP [5'-CAGCCCGAGTCCTGCCATTGG-3'] with 5'-FAM and 3'-BHQ1 modifications as probe primer) targeting a common segment of intron 1. In parallel, we performed qPCR reactions for a human genome single copy reference sequence (RNaseP TaqMan Copy Number Reference Assay; Life Technologies). We used a series of male controls with defined *LW/MW* copy number [n = 1-6; (4)] to generate a copy number dependent ΔCt calibration curve. All three assays were performed in triplicate for each sample. Obtained ΔCt values were used to infer copy numbers from the calibration curves.

Minigene Constructs

A *Hind*III site in the polylinker of the prototype construct was eliminated for the present study by means of inverse PCR with primers *Not*I-Inv3-Rv (5'-AAAAAAGCGGCCGCTGACGGTTCACTAAACGAGCTCT-3') and *Not*I-Inv3-Fw (5'-AAAAAAGCGGCCGCCGGATCCGGTACCATGGCCCAGCA-3'), followed by digestion with *Not*I and re-circulation. Derived minigenes with novel exon 3 haplotypes were generated by replacing a 279 bp *Hind*III-*Afl*II fragment with a homologous fragment from defined subject obtained through PCR amplification with primers BCM-Ex3-r and BCM-Ex3-f and digestion with *Hind*III and *Afl*II. Alternatively, PCR products were first cloned in pCR2.1 (TA Cloning Kit, Invitrogen, Life technologies) before excision of the *Hind*III-*Afl*II fragment and replacement cloning into the revised prototype construct. In vitro mutagenesis with primers IVM-Ex3-174t-Fw (5'-TGGCCATCGTGGGCATTGTCTTCTCCTGGATCTG-3') and IVM-Ex3-174t-Rv (5'-CAGATCCAGGAGAAGACAATGCCCACGATGGCCA-3') was performed to obtain the haplotype 'LVVIA'. All constructs were verified by Sanger sequencing.

Transfection and RNA extraction

HEK293 cells were routinely maintained in Dulbecco´s modified Eagle´s medium, DMEM (Gibco, Life Technologies) supplemented with 10% fetal bovine serum (FBS; Gibco, Life Technologies), Penicillin-Streptomycin (Sigma-Aldrich Chemie GmbH) at 100 µg/ml, and Amphotericin B (Biochrom GmbH) at 2.5 µg/ml. Cells were seeded in 6-well plates in DMEM with 10% FBS and the following day, at 80-90% confluency, cells were transfected with 4 µg DNA of the minigene construct using 20 µl Lipofectamine 2000 per well and OptiMEM® supplemented with GlutaMAX™ (Life Technologies) as diluents and medium. After 6 h incubation, cells were harvested by trypsinization with 0.05% Trypsine-EDTA (Gibco, Life Technologies), centrifuged at 1500 rpm for 5 min and transferred to a 6 cm dish with DMEM supplemented with 10% FCS and antibiotics. 24 h post-transfection, cells were lysed and total RNA was extracted applying the peqGOLD Total RNA Kit (PEQLAB Biotechnologie GmbH).

RT-PCR and relative quantification

First strand cDNA synthesis was performed using 2 µg of total RNA and random hexamer primers, according to manufacturer's instructions (Transcriptor First Strand cDNA Synthesis Kit, Roche). Subsequent PCR was performed with a 5' FAM (6-carboxyfluorescein) labeled forward primer, FEO35: 5'-ACCATGAAGTTCAAGAAGCT-3', and O4-104-Rv:5'-AGCAGGTGACCATGAGGA-3' as reverse primer and using the QIAGEN Multiplex PCR Kit reagent chemistry (Qiagen) including 1/10 volume of Q-solution. Cycling conditions were 95°C for 15s, 40 cycles of 94°C for 30s, 60°C for 90s and 72°C for 45 to 60s, and a final extension at 72°C for 7 min. PCR products were separated by electrophoresis on a 2% agarose gel. FAM-labeled RT-PCR products were diluted 1:10 in water and mixed with 1 µl of GeneScan ROX500 size standard (Life Technologies) and 8 µl of Hi-Di Formamide (Life Technologies) in a total volume of 10 µl. Mixes were separated by capillary electrophoresis on an ABI 3130XL Genetic Analyzer instrument (Life Technologies). The area-under-the-curve (AUC) was calculated with GeneMapper 5 (Life Technologies) software. Ratios of splicing products were determined as the AUC for individual peaks divided by the sum of AUC of all differentially spliced products.

Microsatellite analysis

Centromeric (DXS8011, DXS8103, DXS1356, DXS8087) and telomeric (L441TA, L441CA, AF277A, AF277B and DXS1073) markers to the *LW/MW* cluster were used to genotype the three BCM72 family members (see Figure. 4A, Supplementary Table S2 and Supplementary Fig. S1). Primers were labeled either with TET, HEX or FAM. Standard PCR reactions were performed with 50-100 ng of genomic DNA. PCR products were diluted 2-10 fold with water, mixed with GeneScan ROX500 size standard and Hi-Di Formamide, and separated by capillary electrophoresis. GeneMapper 5 Software was used for allele calling. Markers for which the mother BCM72-II:1 was heterozygous were used to reconstruct haplotypes.

Mapping of the gene conversion event

LD-PCRs were performed for all three members of BCM72 using primers LCR1F and RGCP2-5R for the proximal gene copy, and E1G-F and RGCP2-5 for the distal copies. Upon *Avr*II-*Kpn*I digestion, the resulting 7494 bp and 6211 bp fragments from *LW* (intron 1 to intron 5) and *MW* (intron 1 to intron 4) respectively, were purified with CHROMA SPIN+TE-1000 columns (Takara Bio Europe) and ligated to *Xba*I-*Kpn*I double-digested pGEM3Zf(+). We selected two independent clones of the following gene copies for further analysis: (1) *LW*-derived clones bearing the 'LIAVA' haplotype from subjects BCM72-II:1 and BCM72-III:1, (2) *MW*-derived clones bearing the 'LIAVA' haplotype from subject BCM72-I:1, and (3) *LW*-derived clones bearing the 'LIAVS' haplotype from subject BCM72-I:1. We sequenced exon 3 and flanking introns in these clones by means of primer walking.

Bioinformatic predictions and reference sequences

In silico splicing predictions were performed with SROOGLE (http://sroogle.tau.ac.il/) (28). NG_009105.2, NG_011606.1, NM_020061.5 and NM_000513.2 as reference sequences of the human *OPN1LW* and *OPN1MW* genes (NG) and mRNA transcripts (NM), respectively (NCBI database). All four reference sequences carry the exon 3 haplotype 'MVAIA'.

Supplementary Tables and Figures

Supplementary Table S1. Compilation of patients' demographics, clinical findings and diagnosis, and genotypes.

Subject code	Age at (last) clinical examination	VA: OD OS	Refraction: OD OS	Fundus	Color vision	ERG	Photophobia	Nystagmus	Progression	Clinical diagnosis	Exon 3 Genotype(s)
Single *LW* or *LW/MW* Hybrid Gene Array											
BCM73-16953	51 years	0.15 0.15	-12 -12	Peripapillary atrophy, slight pigmentary changes in macula	Impaired, OD: blue-yellow axis OS: no specific axis	Photopic / 30 Hz flicker: no response Scotopic: normal	No	Yes	No	BCM	Single LW/MW hybrid; 'LIAVA'
BCM73-17481	14 years	0.05 0.05	-14 -14	Peripapillary atrophy, slight pigmentary changes in macula and blond appearance of fundus	ND	Photopic / 30 Hz flicker: no response Scotopic: normal	No	Yes	No	BCM	Single LW/MW hybrid; 'LIAVA'
BCM73-20537	3 years	0.2 0.2	-4.0(-1.75x30°) -4.0(-2.0x170°)	Peripapillary atrophy, blond appearance of fundus, slight pigmentary changes in macula	ND	Photopic: severely reduced Scotopic: normal	No	Yes	No	XLCD	Single LW/MW hybrid; 'LIAVA'
BCM73-20770	3 months	0.2 0.2	-8.0(-1.25x30°) -9.0(-1.50x155°)	Blond appearance of fundus	ND	ND	No	Yes	No	BCM	Single LW/MW hybrid; 'LIAVA'
BCM93-19164	14 years	0.3 0.3	-5 -3	Myopic optic nerve, peripapillar choroid atrophy, retinal atrophy at posterior pole	Impaired	Photopic: no response Scotopic: normal	No	Yes	No	XLCD	Single MW gene; 'LIAVA'
BCM66-16407	41 years	0.3 0.3	-13 -14	Peripapillary atrophy, myopic choroidosis	Extreme red-green defect; Berson test positive for BCM	Photopic: absent Scotopic: normal	Very mild	No	No	BCM	Single LW gene: 'LVAVA'
BCM112-23518	14 years	0.16 0.16	-24 -23	Peripapillar choroid atrophy, thinning of RPE at posterior pole	Protan defect, consistent with incomplete achromatopsia	Photopic: no response Scotopic: severely reduced	No	Yes	Slowly progressive	CRD	Single LW/MW hybrid: 'LVAVA'
BCM112-22852	6 years	0.3 0.3	-9 -9.5	Diminished foveal reflexes	Protan defect, consistent with incomplete achromatopsia	Photopic: severely reduced Scotopic: slightly reduced	No	Yes	Slowly progressive	CRD	Single LW/MW hybrid: 'LVAVA'
BCM194-25474	5 years	0.5 0.5	-16.00(-1.75x35°) -16.25(-1.00x35°)	Temporal optic nerve atrophy, no true fovea reflex, global hypo-pigmented fundus	Normal children's Ishihara test in both eyes.	Photopic: strongly reduced, 30Hz Flicker: reduced and delayed Scotopic: normal	Mild	No	Progressive myopia	BCM	Single LW gene: 'LVAVA'

Supplementary Table S1 (continued) - Compilation of patients' demographics, clinical data and diagnosis, and genotypes.

Subject code	Age at (last) clinical examination	VA: OD OS	Refraction: OD OS	Fundus	Color vision	ERG	Photo-phobia	Nystag-mus	Progre-ssion	Clinical diagnosis	Exon 3 Genotype(s)
Multigene LW/MW Gene Array											
ZD379-19194	12 years	0.16 0.16	ND	Mild optic disc pallor	ND	Photopic / 30 Hz flicker: strongly reduced Scotopic: normal	ND	Yes	No	BCM	Multigene (n = 3) LW(prox.): 'LIAVA' MW(distal): 'MIAVA$^{c.465C,}$
ZD379-19195	7 years	0.16 0.16	+1 +1	Peripapillary atrophy and myopic fundus; some macular RPE mottling	ND	Photopic / 30 Hz flicker: strongly reduced Scotopic: normal	ND	ND	ND	BCM	Multigene (n = 3) LW(prox.):' LIAVA' MW(distal): 'MIAVA$^{c.465C,}$
BCM101-19818	3 years	0.1 0.1	-6 -6	Normal	Impaired	Photopic: no response Scotopic: normal	Yes	Yes	No	BCM	Multigene (n = 3) LW(prox.): 'LIAVA'; MW(distal): 'MIAVA$^{c.465G,}$
BCM126-20616	41 years	0.25 0.20	ND	Temporal optic atrophy, no bone spicules or pigment deposits	Impaired; F-M: several confusions along deutan axis	Photopic / 30 Hz flicker: strongly reduced and delayed Scotopic: normal	Yes	ND	Yes	BCM	Multigene (n = 2) LW(prox.): 'LVAVA' MW(distal): 'LVAVA'
BCM72-17075 (BCM72-III:1)	12 years	0.1 0.1	+0.5(-1.5x8°) +0.5(-1.5x48°)	Slight pigmentary changes in macula	Protan defect	Photopic: no response Scotopic: normal	No	Yes (during first year of life)	Yes	BCM	Multigene (n = 4) LW(prox.): 'LIAVA' MW(distal): 'LIAVA'/ 'MVVVA$^{c.465C,}$
BCM133-20960	10 years	0.6 0.5	-5.00(-3.75x25°) -4.00(-3.25x160°)	Peripapillary atrophy related to myopia	Moderate red/green axis in both eyes	Photopic / 30 Hz flicker: reduced by 15% Scotopic: normal	Moderate photophobia	Fine, inconstant nystagmus	No apparent progression (in 3 years follow-up)	BCM	Multigene (n = 2) LW(prox.): 'LVAVA' MW(distal): 'LVAIA'
BCM133-20961	32 years	0.6 0.4	-20.75 -19.00	Atrophic posterior pole related to high myopia, slightly narrowed retinal vessels	ND	Photopic: strongly reduced Scotopic: moderately reduced	Severe photophobia	No	Very moderate progression, maybe due to high myopia	BCM	Multigene (n = 2) LW(prox.): 'LVAVA' MW(distal): 'LVAIA'

Supplementary Table S1 (continued) - Compilation of patients' demographics, clinical data and diagnosis, and genotypes.

Subject code	Age at (last) clinical examination	VA: OD OS	Refraction: OD OS	Fundus	Color vision	ERG	Photo-phobia	Nystag-mus	Progre-ssion	Clinical diagnosis	Exon 3 Genotype(s)
Multigene LW/MW Gene Array (continued)											
BCM133-23364	1 year	Strabism	-5 -5	ND	ND	Photopic: decreased Scotopic: normal	No	No	No follow up	BCM	Multigene (n = 2) LW(prox.): 'LVAVA' MW(distal): 'LVAIA'
ZD547-4544	46 years	0.8 0.8	-5.0(-3.75x15°) -6.0(-3.75x175°)	Myopic optic nerve	Impaired; D15/Nagel: several confusions along protan axis	Photopic / 30 Hz flicker strongly reduced and delayed Scotopic: normal	Yes	No	No	XLCD	Multigene (n = 2) LW(prox.): 'LVAVA' MW(distal): 'MVAVA'
BCM160-23130	6 years	0.3 0.5	-3.75(-5.50x8°) -5.0(-5.0x166°)	Papilla with mild *conus myopicus*, thin retina, no true fovea reflex	Impaired, red/green defect	Photopic / 30 Hz flicker reduced and delayed Scotopic: essentially normal	Subjective mild light sensitive	No	Progressive myopia	BCM	Multigene (n = 2) LW(prox.): 'LIAVA' MW(distal): 'MVVVA$^{c.465G}$,
BCM51-12359	12 years	0.2 0.2	-14.5(-2.00x50°) -14.5(-1.50x130°)	Myopic discs with peripapillary crescent, stretched and attenuated retinal vessels, thinned RPE in the posterior pole and macula, and fine RPE mottling with granularity and early lacunar or coin-shaped lesions in the mid-periphery, OU.	Protan defect consistent with incomplete achromatopsia	Photopic: DA cone responses to red stimulus, photopic single flash and 30Hz flicker responses strongly subnormal and delayed. Spectral LA cone ERG supports normal distribution of cones Scotopic: mildly subnormal b-waves amplitudes	ND	Small amplitude nystagmus	ND	XLCD	Multigene (n = 3) LW(prox.): 'LIAVA' MW(distal): 'MVVVA$^{c.465G}$,
ZD314-18057	10 years	0.2 0.2	-6.5(+1.25x100°) -6.0(+2.25x75°)	Pale disk, no foveal reflex, no pigmentary abnormalities	Impaired color vision, no specific axis	Photopic / 30Hz flicker strongly reduced and delayed Scotopic: normal	Yes	No	No	XLCD	Multigene (n = 2) LW(prox.) 'LIAVA' MW(distal): 'MVVVA$^{c.465C}$,
BCM98-19713	23 years	0.3 0.4	-2.75(-3,5x5°) -0.25(-2.75x180°)	Normal	D-15/saturated: normal; D-15/desaturated: color confusion w/o specific color confusion axis	Photopic / 30Hz flicker response reduced Scotopic: normal	No	No	No	XLCD	Multigene (n = 2) LW(prox.): 'LIAIA' MW(distal): 'MVAIA'
BCM142-21958	41 years	0.25 0.20	ND	Temporal optic atrophy, no bone spicules or pigment deposits	Impaired; F-M: several confusions along deutan axis	Photopic / 30Hz Flicker strongly reduced and delayed Scotopic: normal	Yes	ND	Yes	BCM	Multigene (n = 3) LW(prox.): 'LVAISS' MW(distal): 'MVAIA'/'LVAIA'

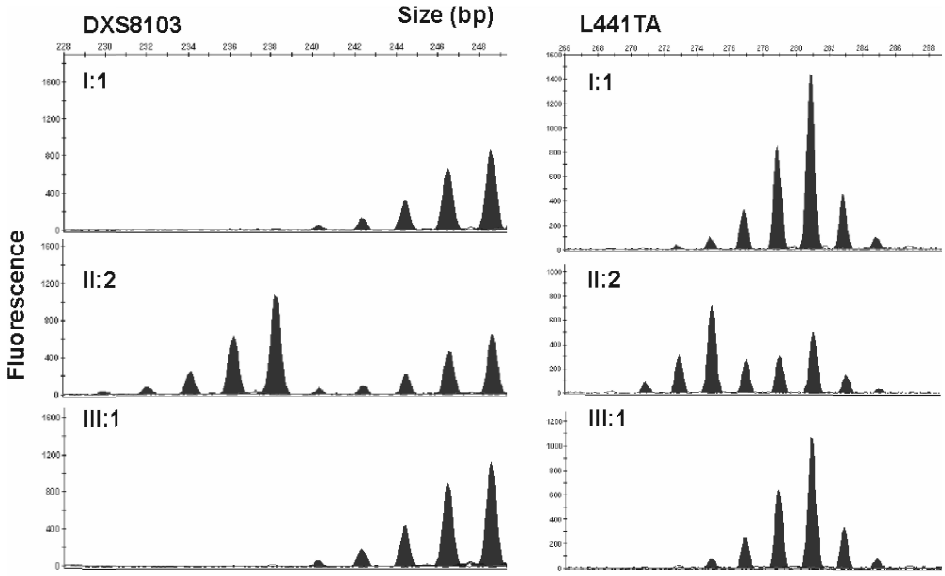
Supplementary Table S1 (continued) - Compilation of patients' demographics, clinical data and diagnosis, and genotypes.

Subject code	Age at (last) clinical examination	VA: OD OS	Refraction: OD OS	Fundus	Color vision	ERG	Photophobia	Nystagmus	Progression	Clinical diagnosis	Exon 3 Genotype(s)
BCM72 family											
BCM72-16874 (BCM72-I:1)	71 years	0.1 0.1	+1.75(-2.5x91°) +0.75(–0.5x95°)	Macular dystrophy	D-15/desat.: Deuteranomaly	normal	No	No	Yes	Deutan/ Macular dystrophy	Multigene (n = 4) LW(prox.): 'LIAVS' MW(distal):: 'LIAVA'/ 'MVVVA$^{c.465C,}$
BCM72-17075 (BCM72-III:1)	12 years	0.1 0.1	+0.5(-1.5x8°) +0.5(-1.5x48°)	Slight pigmentary changes in macula	D-15/desat.: Protan defect	Photopic: no response Scotopic: normal	No	Yes (during first year of life)	Yes	BCM	Multigene (n = 4) LW(prox.): 'LIAVA' MW(distal): 'LIAVA'/ 'MVVVA$^{c.465C,}$
BCM72-16876 (BCM72-II:1)	31 years	0.7 ambly op 1.0	-0.25 -0.75(-1.0x157°)	No data	D-15/desat.: Normal	Fullfield ERG: normal; multifocal ERG: slightly reduced	No	No	No	normal sighted	female carrier

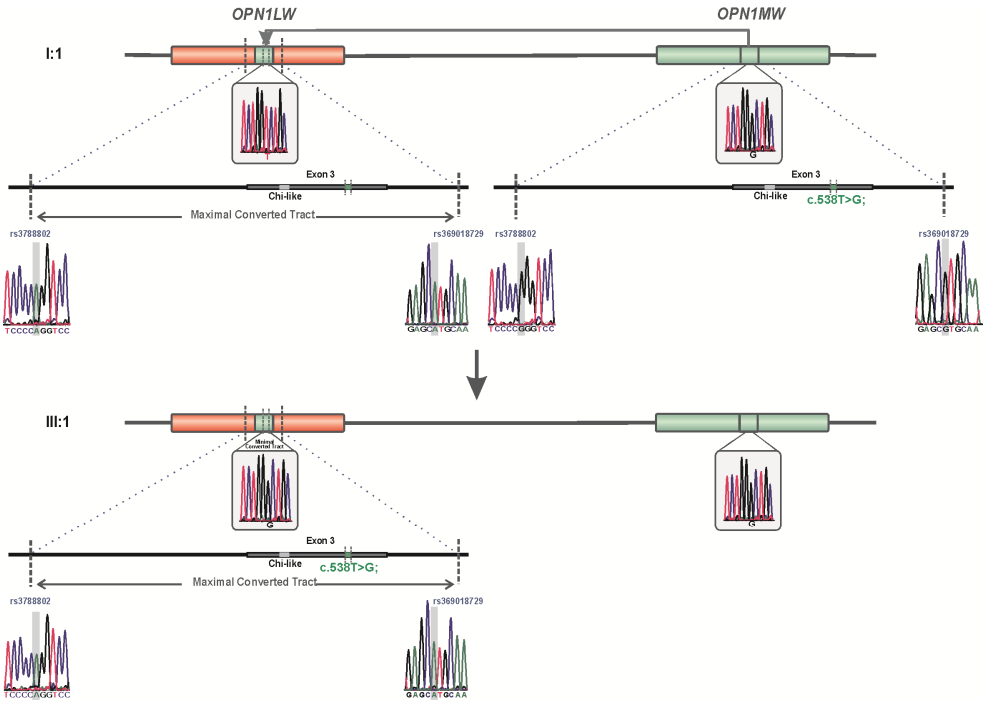
D-15 – Farnsworth Panel D-15, DA – dark-adapted, F-M – Farnsworth Munsell, LA – light-adapted, ND – no data, OD – *oculus dexter* (right eye), OS – *oculus sinister* (left eye), OU – *oculus uterque* (both eyes), prox. – proximal, RPE – retinal pigment epithelium, VA – Visual acuity, prox.: proximal.

Supplementary Table S2- Physical location and genetic map distances of the microsatellite markers used for segregation analysis.

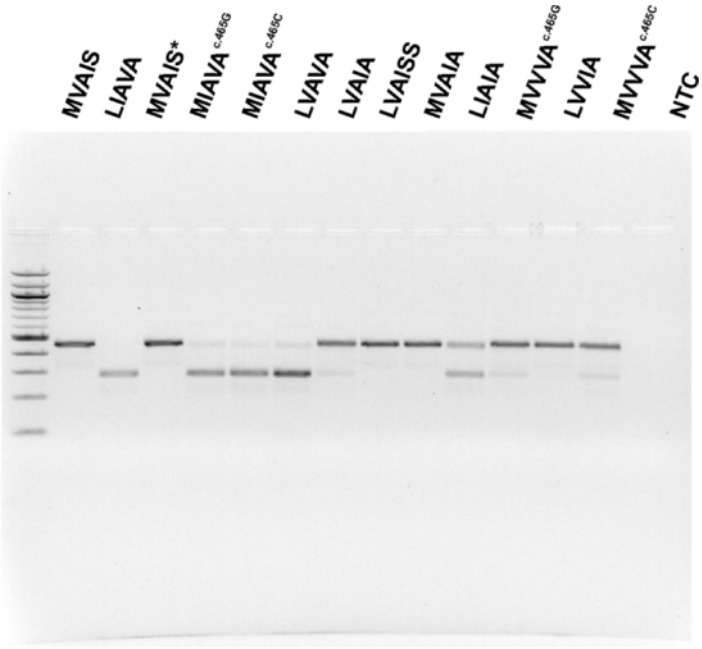
Microsatellite marker	Marshfield genetic map distance (cM)	Physical distance (bp) according to GRCh37/hg19 assembly
DXS8011	98.2	-
DXS8103	100.73	-
DXS1356	-	chrX:152,698,178-152,698,407
DXS8087	102.35	-
L441TA	-	chrX:153,747,967-153,748,233
L441CA	-	chrX:153,754,802-153,755,044
AF277A	-	chrX:153,778,324-153,778,552
AF277B	-	chrX:153,828,907-153829042
DXS1073	102.35	--



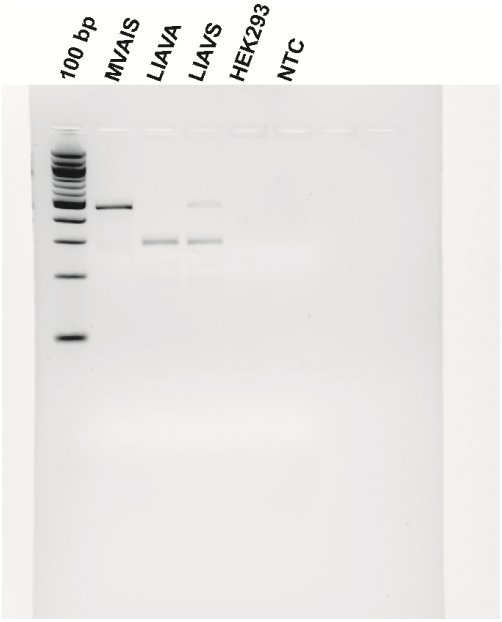
Supplementary Figure S1: Segregation of two informative microsatellite markers demonstrating transmission of the grandparental X-chromosome (BCM72-I:1) to the grandson (BCM72-III:1) in family BCM72. DXS8103 maps centromeric to the *LW/MW* gene cluster, whereas L441TA is the closest marker telomeric to the *LW/MW* gene cluster.



Supplementary Figure S2: Mapping of the outermost borders of the gene conversion event in family BCM72. The maximal converted tract in family BCM72 is delimited by SNPs rs3788802 (c.409+949G>A) in intron 2 and rs69018729 (c.578+91G>A) in intron 3, for which the 'LIAVA' bearing *MW* gene copy in BCM72-I:1 differs from the *LW* gene copies in subjects BCM72-I:1, BCM72-II:2 and BCM72-III:1. Note that the gene conversion event converts the 'LIAVS' bearing *LW* gene copy in BCM72-I:1 into a 'LIAVA' bearing *LW* copy that is transmitted to the mother and the grandson.



Supplementary Figure S3: Full size gel image of Figure 2a. Agarose gel electrophoresis of RT-PCR products obtained with RNA from HEK293 cells transfected with minigene constructs bearing various exon 3 haplotypes. The tested haplotype is given above the corresponding gel lane. A 100 bp ladder size standard was loaded in the leftmost lane. Both lanes 'MVAIS' and 'MVAIS*'refer to minigenes carrying the control haplotype. 'MVAIS*' has a modified Multiple Cloning Site from the prototype construct 'MVAIS' (see Supp. Materials and Methods). NTC: non-template negative control. A scheme on the composition of the RT-PCR products is given on the right.



Supplementary Figure S4: Full size gel image of Figure 5a Direct comparison of RT-PCR products from the minigene splicing assays shows a substantial amount of correctly spliced transcripts (450 bp) for the 'LIAVS' exon 3 haplotype, whereas such products are undetectable for the 'LIAVA' haplotype. No products were obtained in the RT-PCR with RNA from untransfected HEK293 cells (lane 'HEK293') and the no template control PCR (lane 'NTC').

13.1.3 Publication III

Vithiyanjali Sothilingam, Marina Garcia Garrido, Kangwei Jiao, **Elena Buena-Atienza**, Ayse Sahaboglu, Dragana Trifunović, Sukirthini Balendran, Tanja Koepfli, Regine Mühlfriedel, Christian Schön, Martin Biel, Angelique Heckmann, Susanne C. Beck, Stylianos Michalakis, **Bernd Wissinger**, Mathias W. Seeliger and François Paquet-Durand. Retinitis pigmentosa: impact of different Pde6a point mutations on the disease phenotype. *Human Molecular Genetics* (2015) 24, 5486–5499

Retinitis Pigmentosa: Impact of different *Pde6a* point mutations on the disease phenotype

Authors

Vithiyanjali Sothilingam [1§], Marina Garcia-Garrido [1§], Kangwei Jiao [2, 3§], Elena Buena-Atienza [4], Ayse Sahaboglu [2], Dragana Trifunović [2], Sukirthini Balendran [4†], Tanja Koepfli [4*], Regine Mühlfriedel [1], Christian Schön [5], Martin Biel [5], Angelique Heckmann [6], Susanne C. Beck [1], Stylianos Michalakis [5], Bernd Wissinger [4#], Mathias W. Seeliger [1#], François Paquet-Durand [2#]

Affiliations

[1] Division of Ocular Neurodegeneration, Institute for Ophthalmic Research, Centre for Ophthalmology, University of Tuebingen, Schleichstr.4/3, 72076 Tuebingen, Germany

[2] Cell death mechanisms group, Institute for Ophthalmic Research, Centre for Ophthalmology, University of Tuebingen, Roentgenweg 11, 72076 Tuebingen, Germany

[3] 2nd People's Hospital Of Yunnan Province & 4th Affiliated Hospital Of Kunming Medical University, 176 Qingnian Rd., Wuhua, 650021 Kunming, Yunnan, China

[4] Molecular Genetics Laboratory, Centre for Ophthalmology, University Clinics Tuebingen, Roentgenweg 11, 72076 Tuebingen, Germany; † current address: Department of Obstetrics and Gynaecology, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria; * current address: CeGat GmbH, Tübingen, Germany

[5] Center for Integrated Protein Science Munich (CIPSM) at the Department of Pharmacy – Center for Drug Research, Ludwig-Maximilians-Universität München, 81377 Munich, Munich, Germany

[6] *GenOway, 181 Avenue Jean Jaures, 69362 Lyon, France*

[§] *The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors'*

[#] *These authors should be considered as equivalent corresponding authors*

Abstract

Mutations in the *PDE6A* gene can cause rod photoreceptors degeneration and the blinding disease Retinitis Pigmentosa (RP). While a number of pathogenic *PDE6A* mutations have been described, little is known about their impact in compound heterozygous situations and potential interactions of different disease-causing alleles.

Here, we used a novel mouse model for the *Pde6a* R562W mutation in combination with an existing line carrying the V685M mutation to generate compound heterozygous *Pde6a* V685M/R562W animals, exactly homologous to a case of human RP. We compared the progression of photoreceptor degeneration in these compound heterozygous mice with the homozygous V685M and R562W mutants, and additionally with the D670G line that is known for a relatively mild phenotype. We investigated PDE6A expression, cGMP accumulation, calpain and caspase activity, *in vivo* retinal function and morphology, as well as photoreceptor cell death and survival.

This analysis confirms the severity of different *Pde6a* mutations and indicates that compound heterozygous mutants behave like intermediates of the respective homozygous situations. Specifically, the severity of the four different *Pde6a* situations may be categorized by the pace of photoreceptor degeneration: V685M (fastest) > V685M/R562W > R562W > D670G (slowest). While calpain activity was strongly increased in all four mutants, caspase activity was not. This points to the execution of non-apoptotic cell death and may lead to the identification of new targets for therapeutic interventions. For

individual RP patients, our study may help to predict time-courses for *Pde6a*-related retinal degeneration and thereby facilitate the definition of a window-of-opportunity for clinical interventions.

Key Words

Cyclic GMP; necrosis; electroretinography; retina; phototransduction

Introduction

Retinitis Pigmentosa (RP) is a hereditary neurodegenerative disease of the retina which affects photoreceptors and is a major cause of early-onset blindness in the industrialized world (1). The genetic mutations triggering RP usually lead to a disturbance of the phototransduction cascade, often associated with an elevation of cyclic guanosine mono-phosphate (cGMP) (2). In the dark, high guanylate cyclase activity increases cGMP levels (3). Phototransduction starts with a light-induced, conformational change of opsin molecules, causing sequential activation of transducin and phospho-diesterase-6 (PDE6), the latter of which subsequently reduces the intracellular cGMP concentration. High cGMP levels maintain cyclic-nucleotide-gated (CNG) cation channels in the open state, allowing Ca^{2+} influx (4). PDE6 activation reduces cGMP levels, thus closing CNG channels and causing hyperpolarization and signal transmission to 2nd order neurons.

Genetic mutations affecting PDE6 function lead to an excessive accumulation of cGMP and subsequent rod photoreceptor death (5,6), followed by a mutation independent, secondary death of cone photoreceptors (7). In rod photoreceptors, PDE6 is composed of two catalytic subunits – alpha (A) and beta (B) – and in the inactive state associates with two inhibitory gamma subunits. When all mutations affecting any of the three PDE6 alpha, beta, or gamma subunits are considered together, these are responsible for up to 4-8% of human RP patients (8-10). Previously, PDE6 dysfunction was investigated primarily in animal models affected by mutations in the *Pde6b* gene, such as the *rd1* (11) or the *rd10* mouse (12). A variety of *Pde6a* mutations are also known to cause RP (13) but, so far these have been relatively little studied.

RP animal models are normally studied in the homozygous state (*e.g.* as *Pde6b*$^{rd1/rd1}$) even though, in human RP patients, in outbred populations, homozygosity is relatively rare and, in fact, compound heterozygosity, where two different disease-causing alleles come together in one individual, is more frequent (*e.g. PDE6A*$^{V685M/R562W}$). To account for this, we have used various *Pde6a* mutant mice and studied the progression of retinal degeneration both in the homozygous and compound heterozygous situations. To facilitate a comparison with the human situation, we focussed on *Pde6a* mouse mutants for which homologous RP patients have previously been identified. This relates to point mutations resulting in an amino acid exchange in various positions of the PDE6A protein, *i.e.* R562W, D670G, and V685M. For the sake of clarity, in the following these animal models will be referred to by the position of their respective mutations.

The previously reported *Pde6a* D670G and V685M mouse mutants have been generated by N-ethyl-N-nitrosourea (ENU) mutagenesis and were identified through a screen for altered fundus pigmentation (13). In this study, we generated and studied an additional *Pde6a* R562W knock-in mutant. The generation of this mutant followed the identification of a human RP patient compound heterozygous for the c.1684C>T/p.Arg562Trp and c.2053G>A/p.Val685Met mutations in *PDE6A*. Therefore the generation of the R562W knock-in mutant gave the opportunity to study the exact homologous genotype of the patient through crossbreeding of the R562W knock-in with the V685M mutant. We then used both homozygous and compound heterozygous *Pde6a* mutant animals to assess the relative impact of different genetic insults on the progression and the severity of retinal degeneration. The data generated here may serve as a reference for further pre-clinical studies in RP animal models and – extrapolated to the human situation – may guide future clinical trials for RP therapy development.

Material and Methods

Animals

All *Pde6a* mutants used were generated and maintained on the C57BL6/J strain background, were housed under standard white cyclic lighting, had free access to food and water, and were used irrespective of gender. The V685M and D670G animals were obtained from the Jackson Labs (Bar Harbor, MA, USA). The R562W knock-in mutant was generated by GenOway (Lyon, France) using standard procedures of homologous recombination in murine embryonic stem (ES) cells. Briefly, left arm and right arm homology fragments of 4,066 bp (covering exon 10 to exon 12 and parts of the flanking introns of *Pde6a*) and of 3,830 bp (covering exon 13 and parts of the flanking introns of *Pde6a*), respectively, were PCR amplified from a C57BL/6-derived BAC clone carrying *Pde6a* sequences. To generate the knock-in mutation the 'CGG' arginine codon was replaced by a 'TGG' tryptophan codon at the respective position in exon 13 in the right arm homology fragment by *in vitro* mutagenesis. The cloned fragments were verified by Sanger sequencing and then assembled for the final targeting construct that comprised a neomycin positive selection cassette flanked by loxP sites and inserted in intron 12 between the left and right arm homology fragments, and a diphtheria toxin expression cassette for negative selection. The targeting construct was electroporated into C57BL/6-derived ES cells according to GenOway's electroporation procedures (*i.e.* 108 ES cells in presence of 100 µg of linearized plasmid, 260 Volt, 500 µF). Positive selection was started 48 hours after electroporation, by addition of 200 µg/ml of G418 (150 µg/ml of active component, Life Technologies, Inc.). Resistant clones were isolated and amplified in 96-well plates. Duplicates of 96-well plates were made. The set of plates containing ES cell clones amplified on gelatine were genotyped by both PCR and Southern blot analysis and the presence of the mutation was verified by sequencing. One fully characterized ES clone was used for injection into blastocysts of C57BL/6J-Tyr$^{c\text{-}2J}$/J mice and obtained two highly chimeric male mice were obtained. These mice were mated with C57BL/6 cre deleter female mice to excise the neomycin selection cassette. The F1 progeny was tested for proper excision of the neomycin cassette by means of PCR and Southern blotting. Since the used C57BL/6-derived ES cells carry the *rd8* allele (14), we mated the F1 with C57BL/6J animals and then crossbred F2 animals to obtain homozygous R562W mice devoid of *rd8*. For all animals used in this study day of birth was considered as postnatal day (P) 0. All procedures carried out on animals were reviewed and approved by the competent authority (Regierungspräsidium Tuebingen). All efforts were made to minimize the number of animals used and their suffering.

Functional analysis of the p.R562W mutant gene

We used an established system for the functional expression of PDE6 protein based on a PDE6C/PDE5 fusion construct and the use of the baculovirus/Sf9 insect cell system for recombinant protein expression (15,16). The mutated construct bearing a tryptophane codon instead of an arginine codon at the site homologous to p.R562 of *Pde6a* was generated by *in vitro* mutagenesis (Quik Change *in vitro* Mutagenesis Kit from Agilent, Waldbronn, Germany). Generation of recombinant bacmids, transfection of Sf9 insect cells and viral amplifications were carried out according to the manufacturers´ recommendations (Life Technologies/Invitrogen, Darmstadt, Germany). Lysis of cells and purification of recombinant protein was carried out as previously described (15) and PDE activity was measured using [^{3}H]-cGMP (GE Healthcare) as substrate (17). Briefly, 5µg purified wild type and mutant protein, respectively, were incubated in a total volume of 40µl of 20mM Tris-HCl pH8.0, 50mM NaCl, 15mM $MgSO_4$, 2mM beta-mercaptoethanol, 0.1µM shrimp alkaline phosphatase and 5µM [^{3}H]cGMP (100.000cpm) for 10 min at room temperature. The reaction was stopped by the addition of 0.5ml AG1-X2 anion exchange resin (*Bio-Rad*) in a 20% bed volume suspension. Samples were incubated with the resin for 10 min with occasional vortexing and then centrifuged at 9000xg for 2 min. 250 µl aliquots of the clear supernatant were removed and measured in a scintillation counter (Beckman LS 6000 Liquid Scintillation Counter, Beckman Coulter GmbH, Krefeld, Germany).

RT-PCR and allelic quantification

Homozygous and heterozygous *Pde6a*:R562W-knockin mice were sacrificed at the age of P14 and P80, respectively. In addition, we used P13 and adult C57BL6 wild-type mice as control. Retinas were dissected from enucleated eyes and used to prepare total RNA. The tissue was lysed mechanically in a Precellys homogenizer (Peqlab Biotechnologie GmbH, Erlangen, Germany) and total RNA isolated through affinity chromatography on silica membrane (peqGold Total RNA Kit; Peqlab Biotechnologie GmbH). Single-stranded cDNA was synthesized from 1-2 µg of total RNA by reverse transcription applying random hexamers for priming (Transcriptor First Strand cDNA Synthesis Kit, Roche, Mannheim, Germany). For qualitative RT-PCR we amplified 1/10 volume of the first-strand cDNA with primers MmPde6a_cDNA_ex12Fnw (5-ACGCGGAGTCATACGAAATC-3) and MmPde6a_cDNA_ex15Rnw (5-ATGATGCCTTTCCAAGATGG-3) or Pde6a_cDNA_Ex11 (5'-AGAGGTGTACGGCAAAGAGC-3') and Pde6a_cDNA_Ex14 (5'-GTTGTTCGTGCCTCTGTGGT-3') applying standard PCR conditions RT-PCR products were cloned into pCR2.1 using the TA Cloning Kit (Invitrogen/Life Technologies) and plasmid DNA purified from single bacterial clones was sequenced applying cycle sequencing and BigDye Terminator V1.1 chemistry (Applied Biosystems/Life Technologies). Sequencing products were separated on an ABI 3130XL capillary sequencer (Applied Biosystems/Life Technologies). Raw sequences were processed using Sequencing Analysis Software V5.2 (Applied Biosystems/Life technologies) and assembled into contigs using SeqMan (Lasergene, Madison, WI, USA). Relative quantification between the R562W knock-in allele and the wild-type allele was done by pyrosequencing. For relative quantification of allelic *Pde6a* transcripts we amplified 1/10 volume of first strand retinal cDNA of heterozygous mutant mice with primers Pde6a_Mm_Ex12-13_F (5'-TTCCACATCCCGCAAGAG-3') and a 5'-biotinylated reverse primer Pde6a_Mm_Ex12-13_R (5-Bio-CCAGCAAGGAGAACATGGTC-3') applying standard PCR conditions. For comparison we amplified an exon 13 internal fragment of the Pde6A gene from genomic DNA of heterozygous

R562W knock-in mice (extracted from ear punches) with primers Pde6a_Mm_Ex13-DNA_F (5'-GTTCACAGGCCCTGGTGC-3') and 5' biotinylated Pde6a_Mm_Ex12-13_R (see above). PCR products were purified by immobilization on streptavidin coated sepharose beads (GE Healthcare, Munich, Germany) and vacuum filtration (Vacuum Manifold, Qiagen, Hilden, Germany). Single-stranded PCR products were sequenced with primer Pde6a_Mm_Ex12-13_S (5'-GAATCACTTACCACAACTG -3') on a Pyromark Q96 instrument according to the manufacturer's recommendations (Qiagen).

Minigene splicing assays

Minigene constructs were generated by cloning both allelic products of a PCR with primers EcoRI-PDE6A-MMf (5'-aaGAATTCTGTATCAGTACGACCCAAGAC-3'; EcoRI recognition site underlined) and BamHI-PDE6A-MMr (5'-aaGGATCCGAGTTGTATACTTCTCTATTCTTGGAA-3'; BamHI recognition site underlined) with genomic DNA of a heterozygous R562W-knockin mutant into the EcoRI and BamHI sites of the pSPL3 vector. The amplified fragment from the wild-type allele encompasses exon 13 of *Pde6a* and 424 bp of upstream intron 12 sequence and 308 bp of downstream intron 13 sequence. The fragment from the knock-in allele covers an additional 96 bp insert (comprising a single loxP site and flanking multiple cloning site sequences derived from the targeting vector) inserted 247 bp upstream of exon 13. Further constructs were obtained by introducing the c.1684C>T/p.R562W mutation into the wild-type allele construct and by reverting the mutation in the knock-in allele construct through *in vitro* mutagenesis according to the manufacturer's protocol (Quik Change Mutagenesis Kit; Agilent). The inserts of the minigene constructs were verified by Sanger sequencing as outlined above. Purified plasmid DNA of the constructs were used to transfect HEK293 and 661W as follows: Cells were seeded in 6-well plates in DMEM (Gibco/Life Technologies) with 10% FBS (Biochrom GmbH, Berlin, Germany) and the following day, at 80-90% confluency, cells were transfected with 4-10 µg plasmid DNA of the minigene construct using 20 µl Lipofectamine 2000 per well and OptiMEM® supplemented with GlutaMAX™ (Gibo/Life Technologies) as diluents and medium. After 6 hours incubation, cells were harvested by trypsinization with 0.05% Trypsine-EDTA (Gibco/Life Technologies), centrifuged at 1500 rpm for 5 min and transferred to a 6 cm dish with DMEM supplemented with 10% FCS and antibiotics (10 ml/l of Penicillin-Streptomycin solution [P4333; Sigma-Aldrich Chemie GmbH, Munich, Germany] and 10 ml/L Amphotericin B [250µg/ml in water; Biochrom AG]). 24 hours post-transfection, cells were lysed and total RNA was extracted applying the peqGOLD Total RNA Kit (Peqlab Biotechnologie GmbH). Single stranded cDNA was synthesized as outlined above and cDNA was amplified with primers SA2 (5'-ATCTCAGTGGTATTTGTGAGC-3') and SD6 (5'-TCTGAGTCACCTGGACAACC-3') applying standard PCR conditions. RT-PCR products obtained from transfected HEK293 cells were cloned into pCR2.1 using the TA Cloning Kit (Invitrogen/Life Technologies) and plasmid DNA isolated from single bacterial clones sequenced as outlined above.

Histology, immunohistochemistry, and immunofluorescence

Animals were sacrificed in the morning (10-11 am), their eyes enucleated and fixed in 4 % paraformaldehyde (PFA) in 0.1 M phosphate buffer (pH 7.4) for 45 min at 4°C. PFA fixation was followed by cryoprotection in graded sucrose solutions (10, 20, 30 %). Unfixed eyecups were directly embedded in cryomatrix (Tissue-Tek, Leica, Bensheim, Germany). Sagittal 12 µm sections were obtained and stored at -20°C. For immunofluorescence, sections were incubated overnight at 4°C with primary rabbit antibody against PDE6A (Novus Biologicals, Cambridge, UK; # NBP1-87312), rabbit antibody against caspase-3 (Cell Signaling, Danvers, MA; #9664), or sheep antibody against cGMP (1:400; kindly provided by Harry Steinbusch, Maastricht, The Netherlands), then washed in PBS and incubated for 1h with Alexa Fluor 488-conjugated, matching secondary antibodies (Molecular Probes, Inc. Eugene, USA). Negative controls were carried out by omitting the primary antibody. Sections were mounted with Vectashield (Vectorlabs, Burlingame, CA, USA) for microscopy.

Calpain activity and TUNEL Assay

Calpain activity was investigated using an enzymatic in situ assay (18). Unfixed cryosections were incubated for 15 min in calpain reaction buffer (CRB; 25 mM HEPES, 65 mM KCl, 2 mM MgCl2, 1,5 mM CaCl2, 2 mM DTT) and then incubated at 35°C for 1h in the dark in CRB with 2 mM fluorescent calpain substrate 7-amino-4-chloromethylcoumarin, t-BOC-Leucyl-Lmethionine amide (Life technologies, Darmstadt, Germany; #A6520). Fluorescence was uncaged by calpain-dependent cleavage of t-Boc-Leu-Met-CMAC. Terminal deoxynucleotidyl transferase dUTP nick end labelling (TUNEL) assay was performed using an *in situ* cell death detection kit (Fluorescein or TMR; Roche Diagnostics GmbH, Mannheim, Germany). For controls terminal deoxynucleotidyl transferase enzyme was either omitted from the labelling solution (negative control), or sections were pre-treated for 30 min with DNAse I (Roche, 3U/ml) in 50mM Tris-HCl, pH 7.5, 1 mg/ml BSA to induce DNA strand breaks (positive control). While negative control gave no staining, positive control stained all nuclei in all layers of the retina (19).

Microscopy, cell counting, and statistical analysis

Light and fluorescence microscopy were usually performed at room temperature on an Axio Imager Z.1 ApoTome Microscope, equipped with a Zeiss Axiocam MRm digital camera. Images were captured using Zeiss Axiovision 4.8 software; representative pictures were taken from central areas of the retina using a 20x/0,8 Zeiss Plan-APOCHROMAT objective. Adobe Photoshop CS3 (Adobe Systems Incorporated, San Jose, CA) was used for primary image processing.For the quantifications of positively labelled cells (TUNEL, cGMP) pictures were captured on three entire sagittal sections for at least three different animals for each genotype and age using Mosaic mode of Axiovision 4.8 at 20x magnification. The average area occupied by a photoreceptor cell (*i.e.* cell size) for each genotype and age was determined by counting DAPI-stained nuclei in 9 different areas (50 x 50 µm) of the retina. The total number of photoreceptor cells was estimated by dividing the outer nuclear layer (ONL) area by this average cell size. The number of positively labelled cells in the ONL was counted manually. We considered cells as positively labelled only if

they showed a strong staining of either the photoreceptor nuclei (*e.g.* for TUNEL) or perinuclear areas (*e.g.* for cGMP, caspase-3). Values obtained are given as fraction of total cell number in ONL (*i.e.* as percentage) and expressed as mean ± standard error of the mean (SEM). For the quantification of PDE6A protein expression, fluorescent pictures were loaded into ImageJ (Vers. 1.44; Wayne Rasband, National Institute of Mental Health, Bethesda, MA) and the line plot option was used to obtain maximum intensity values for the outer nuclear layer (ONL) and outer segment (OS) areas. The (unstained) ONL values were taken for background subtraction, the average pixel intensity in wt OS was arbitrarily set to 1; the mutant values were expressed as a function of that For statistical comparisons the unpaired Student t-test as implemented in Prism 5 for Windows (GraphPad Software, La Jolla, CA) was employed.

Non-invasive *in vivo* imaging and functional testing

A baseline characterization with electroretinography (ERG), spectral domain optical coherence tomography (SD-OCT) and scanning-laser ophthalmoscopy (SLO) was performed at 4 weeks of age. The groups were composed of 4 animals of each genotype.

Electroretinography (ERG)

ERGs were recorded binocularly from different *Pde6a* mutants as described previously (20). Mice were dark-adapted overnight and anaesthetized using a combination of Ketamine (66.7 mg/kg body weight) and Xylazine (11.7 mg/kg body weight). Their pupils were dilated and single flash ERG responses were obtained under scotopic (dark-adapted) and photopic (light-adapted with a background illumination of 30 cd/m^2, starting 10 min before recording) conditions. Single white-flash stimuli ranged from -4 to 1.5 log cd*s/m^2 under scotopic and from -1.0 to 1.5 log cd*s/m^2 under photopic conditions. Ten responses were averaged with inter-stimulus intervals of 5 s (for -4 to -0.5 log cd*s/m^2) or 17 s (for 0 to 1.5 log cd*s/m^2).

Spectral-Domain Optical Coherence Tomography (SD-OCT)

Retinal structures of the still anesthetized animals were visualized via OCT imaging with a Spectralis^{TM} HRA+OCT (Heidelberg Engineering GmbH, Heidelberg, Germany). This device features a superluminescent diode at 870 nm as low coherence light source. Scans are acquired at a speed of 40.000 scans per second and each two-dimensional B-scan contains up to 1536 A-scans (21-23). The images were taken with the equipment set of 30° field of view and with the software Heidelberg Eye Explorer (HEYEX version 5.3.3.0, Heidelberg, Germany). Resulting images were exported as a 8 bit colour bitmap files and processed with CorelDraw X3 (Corel corporation, Ottawa, ON Canada).

Scanning-Laser Ophthalmoscopy (SLO)

Eyes were kept moisturized with Methocel (Omnivision, Puchheim, Germany) so that SLO imaging was performed in the same session as OCT. It was carried out with a HRA 1 system (Heidelberg Engineering, Heidelberg, Germany) according to previously described procedures (24). Briefly, the HRA 1 system feature lasers in the short (visible) wavelength range (488 nm in both and 514 nm), and also in the long (infrared) wavelength range (795/830 nm and 785/815 nm). The 488 and 795 nm lasers are used for fluorescein (FLA) and indocyanine green (ICG) angiography, respectively. GFP excitation was detected in the autofluorescence mode at 488 nm with a 500 nm barrier filter.

Results

Functional analysis of the p.R562W mutation in PDE6A

To verify the pathogenicity and to assess the functional consequence of a recently identified new missense variant (p.R562W) in the PDE6A gene, we expressed the mutant PDE6 protein in Sf9 insect cells by using an established chimeric PDE6C/PDE5 construct (15). Purified recombinant protein was analysed for cGMP hydrolysis activity. In comparison with wild-type (wt) protein, the catalytic activity of R562W mutant protein was reduced to about 10% (Fig.1).

Generation of a new mouse model for the R562W mutation

In order to study the retinal phenotype caused by the R562W mutation, we generated (under contract at GenOway) a *Pde6a*:R562W-knockin mouse mutant applying state-of-the art homologous recombination in mouse embryonic stem cells. The knock-in allele bore the actual c.1684C>T/p.R562W mutation in exon 13 and, in addition, upstream, in intron 12, remnant sequences of the targeting process including a single loxP site flanked by 62 bp of sequence of the multiple cloning site derived from the targeting vector. We used heterozygous Pde6a:R562-knockin animals to test for proper expression of the mutant allele in the retina. Upon allelic quantification of *Pde6a* RT-PCR products, we found an excess of wild-type allele derived transcripts in heterozygous animals (Fig. 2A).

Further RT-PCR experiments with primers located in exon 12 and 14 revealed two major products in homozygous R562W mutants. The larger RT-PCR product was identical in size (403 bp) to the RT-PCR obtained from control animals, and the other about 100 bp smaller (Fig. 2B). Sequencing of cloned products from homozygous R562W mutant animals revealed that the smaller cDNA fragment lacks

exon 13. Additionally, we performed RT-PCR experiments in heterozygous R562W animals. Sequencing of cloned RT-PCR products showed absence of exon 13 in the smaller cDNA fragment while the larger cDNA fragment represented a mixture of correctly spliced wild-type and knock-in allele derived transcripts, respectively (Supplementary Material, Fig.S1). These findings indicated an incompletely penetrant splicing defect caused by the knock-in allele. Bioinformatic analysis suggested that the C>T transition created an Exonic Splicing Silencer (ESS) site which most likely explained the impaired exon 13 splicing (data not shown). Skipping of exon 13 resulted in an in-frame deletion of the open reading frame and translated into a PDE6A protein shortened by 36 amino acid residues (p.541_576del), which lacks the amino-terminal portion of the catalytic domain including the first metal binding motif.

Since we could not rule out an effect of the resident modifications (loxP site and multiple cloning site sequences; loxP/MCS insert) in intron 12 in the knock-in allele on transcript splicing, we generated minigene constructs that include the murine exon 13 and 424 bp of upstream intron 12 and 309 bp of downstream intron 13 sequences inserted in to the pSPL3 exon trapping vector. Four different variants of the minigene constructs were cloned: (1) wt, (2) wild-type exon 13 and the 96 bp loxP/MCS insert in intron 12, (3) c.1684C>T/p.R562W mutation in exon 13 with wt intron 12 portion, and (4) the knock-in allele (c.1684C>T/p.R562W and loxP/MCS insert in intron 12) (Supplementary Material, Fig. S2). Constructs were transfected into HEK293 cells and total RNA was prepared 24 h post transfection. RT-PCR for the expression construct with primers located in flanking exons of the pSPL3 vector indicate that the splicing defect observed in the knock-in mouse mutant was caused by the c.1684C>T/p.R562W mutation and not primed or influenced by the modifications in intron 12 (Supplementary Material, Fig. S3). Given that a considerable proportion of exon 13 skipped transcripts were also expressed from the wt minigene construct in HEK293T cells, we re-performed minigene assays for the c.1684C>T/p.R562W mutant construct in the mouse cone photoreceptor-like cell line 661W. RT-PCR with RNA from transfected 661W cells showed a strongly reduced proportion of transcripts with exon 13 skipping from the wild-type minigene construct and roughly equal proportions of correctly spliced and exon 13 skipped transcripts from the mutant construct (Supplementary Material, Fig. S4). This corresponded to the findings in mutant mouse retina (Fig. 2B).

Decreased PDE6A expression causes cGMP accumulation and photoreceptor death

We first assessed the expression of the PDE6A protein in the retina of wt and *Pde6a* mutants, at post-natal day (PN) 11, *i.e.* a time-point just before the onset of photoreceptor degeneration and widespread destruction of the outer nuclear layer (ONL). In wt retina PDE6A is present almost exclusively in photoreceptor outer segments (OS). In contrast, in V685M, V685M*R562W, and R562W PDE6A is essentially undetectable, with a faint, dot-like staining pattern in R562W OS indicative of a very low PDE6A protein expression. In D670G retina, PDE6A protein is detectable in OS; however, the amount of protein detected is clearly lower than in wt (Fig. 3A-E).

The decreased detection of PDE6A in mutant retinas correlated with an increased accumulation of cGMP (Fig3F-J). Wt retina did not show signs of important cGMP accumulation (wt: 0.003% cGMP positive cells ± 0.003 SEM, 6), while in all four *Pde6a* mutants, the numbers of rod photoreceptors showing high levels of cGMP was elevated already at PN11 (V685M: 0.61% ± 0.04; V685M*R562W: 0.19% ± 0.06; R562W: 0.17% ± 0.08; D670G: 0.007% ± 0.002; n = 3 for all).

To further assess the progression of retinal degeneration in the various *Pde6a* mutants, we used the TUNEL assay to label dying photoreceptors and, conversely, quantified the number of rows of surviving photoreceptors (Fig. 4). In the V685M ONL the percentage of TUNEL positive cells peaks already at PN12, something that is also evidenced by the early and rapid loss of photoreceptor rows (Fig. 4B, G, L). In the compound heterozygous V685M/R562W and in the R562W cell death peaked at PN15 although the peak amplitude was slightly lower in the R562W mutant (Fig. 4C, D, H, I, M, N).. When compared to the V685M situation, this corresponded to a delay in the onset and progression of degeneration of approximately two days in V685M/R562W retina and four days in R562W retina. The

D670G mutant showed a relatively small peak of cell death at PN21 and also the slowest overall progression of degeneration among the four *Pde6a* mutant genotypes (Fig, 4E, J, O).

Retinal degeneration has frequently been associated with the execution of classical apoptosis and the activity of caspase-type proteases (25,26). However, more recent studies point to the activity of non-apoptotic and caspase-independent mechanisms of cell death in hereditary retinal degeneration (27) which may instead rely on the activity of Ca^{2+} activated calpain type proteases (2,28). To address this question for the *Pde6a* mutants used here, we performed an immunostaining for activated caspase-3, a key effector in classical apoptosis. Since the chances for a positive detection of caspase-3 activity are highest when cell death is high, we focused on the peaks of degeneration as assessed by the TUNEL assay. In the early post-natal wt, a low level of caspase-3 activity was present, likely relating to developmental cell death in rodent retina (29). Remarkably, none of the *Pde6a* mutants showed any significant increase of caspase-3 activity when compared to wt (Fig. 5A-E; quantification in K). An *in situ* assay for Ca^{2+} activated calpain-type proteases (28) revealed very low levels of activation in wt retina. In *Pde6a* mutants however, the numbers of photoreceptors cells showing calpain activation was dramatically increased (Fig. 5F-J). When we then assessed the progression of calpain activity over time, we found a strong temporal correlation between calpain activity and cell death (Fig. 5L).

Characterization of *Pde6a* mutants *in vivo*

The four *PDE6a* mutant lines were morphologically and functionally characterized *in vivo* by means of SLO and OCT imaging and ERG recording at the age of 4 weeks. In SLO imaging, overall fundus appearance was visualized with the green laser at 514 nm (Fig. 6 A, F, K, P); the retinal vasculature was studied with angiography applying indocyanine green (Fig. 6 B, G, L, Q) and fluorescein (Fig. 6 C, H, M, R) and the appropriate laser wavelengths (795 nm and 488 nm, respectively). Retinal layering was studied by means of OCT imaging (Fig. 6 D-E, I-J, N-O, S-T). D670G mice revealed a heavily spotted fundus appearance in the native fundus imaging (Fig. 6 A) as well as in the fluorescein angiography mode (Fig. 6 C) whereas in the OCT analysis, a very thin ONL was visualized depicting the presence of only few photoreceptor rows (Fig. 6 D-E). A spotty fundus was also found in the R562W mice (Fig. 6 F) together with a further decrease in the retinal thickness where no outer retina was detected (Fig. 6 I-J). Compound heterozygous V685M/R562W mice showed a heavily affected retinal fundus with large areas of degeneration (Fig. 6 K, M). In these mutants the retinal thickness was strongly decreased, which resulted in enhanced visibility of choroidal structures in each SLO imaging mode (Fig. 6 K-M). Accordingly, the OCT analysis revealed a highly degenerated retina (Fig. 6 N-O). V685M was the mouse line with the strongest degeneration, retinal and choroidal structures were difficult to distinguish due to the severe decrease in the retinal thickness (Fig. 6 S-T). Altogether, a different degree of retinal degeneration was detected and a gradient based on the severity (from less to more affection) of the disease was established: D670G> R562W>V685M/R562W>V685M.

The *in vivo* morphological findings correlated well with the functional data obtained with ERG. Full-field ERG measurements under both scotopic and photopic conditions allow the assessment of retinal function (Fig. 7A, B). Depending on the extent of morphological alterations, different ranges of ERG recordings could be observed. Retinal function was mostly preserved in the D670G variants (blue box and whisker plot (B&W)), greatly reduced in the R562W mice (green B&W) and completely missing in the V685M mutant (red B&W). The V685M/R562W variant, as an intermediate mutant line, is positioned between the R562W and V685M variants. Scotopic and photopic ERG traces at the highest stimulus intensity (Fig. 7C) were similar to those of a $Rho^{-/-}$ animal used as a functional control for cone-only responses (20,30). A comparison of the size of ERG amplitudes of different *Pde6a* mutants with age-matched C57/BL6 wt mice revealed that amplitudes of *Pde6a* mutants were considerably smaller and represent 1/3 to 1/4 of wt animals (20). Taken together, these *in vivo* morphological and functional observations were closely matching the results obtained in *ex vivo* quantification of photoreceptor cell death and survival (Fig. 4).

Discussion

In this report, we investigated the pathologic consequences of three different *Pde6a* point mutations for retinal photoreceptor degeneration. We analysed the effects of these mutations on PDE6 activity, photoreceptor cGMP accumulation, and progression and mechanisms of retinal degeneration. Previously, a variety of RP animal models have been generated and studied that carried human gene mutations homozygously (31-33). Here, we created an *in vivo* model carrying two different point mutations (*Pde6a* $^{R562W/V685M}$), exactly homologous to a human RP patient genotype, representing one of the first ever attempts to create a compound heterozygous, patient-matched basis for the development and assessment of individualized RP therapies.

Loss of PDE6 function, cGMP signalling, and cell death

PDE6 is the main photoreceptor cGMP hydrolysing enzyme and displays the highest activity levels among all PDE´s (34). Although excessive accumulation of cGMP has been found in various animal models for RP, it is particularly prominent in animals that carry mutations in PDE6 (2). However, it is still unclear to what extent different *PDE6* mutations cause cGMP accumulation and how this correlates with progression speed and severity of retinal degeneration. In all *Pde6a* mutants studied here the genetic defects led to a reduced PDE6 function as evidenced by an accumulation of cGMP. Compared to mutations where PDE6 function is abolished entirely, such as in the *rd1* mouse (35), the accumulation of cGMP in the *Pde6a* point-mutants examined here is less prominent.

High levels of cGMP are thought to cause an over-activation of CNG channels leading to excessive amounts of Ca^{2+}-influx (28,36). This is compatible with the observed increase in the numbers of photoreceptor cells showing a high degree of Ca^{2+}-activated calpain-type protease activity. Calpain activity is often associated with non-apoptotic forms of cell death (37) and in all *Pde6a* mutants it is strongly correlated in time with the progression of photoreceptor cell death. The relatively low amount of calpain activity positive photoreceptors, when compared to the *rd1* mouse (38), likely reflects the lower extent of cGMP accumulation and thus lower Ca^{2+}-influx. This may be particularly evident in the V685M mutant where the numbers of dying, TUNEL positive cells exceeds the numbers of calpain activity positive cells by approximately 2:1. In the more slowly degenerating mutants the relative extent of calpain activity is more important, to reach a ratio of approximately 1:1 in the D670G mutant. In either case these results are suggestive of a causal involvement of calpain activity in the degenerative process.

Remarkably, in all four mutant genotypes the numbers of photoreceptors displaying high caspase-3 activity was very low, at wt levels, and not correlated to the progression of the mutation-induced degeneration. Since caspase-3 is key to the execution of numerous terminal proteolytic events in classical apoptotic cell death (27), this result indicates that apoptosis does not play any major role in *Pde6a* mutation induced retinal degeneration. This is in agreement with a recent study on the prevalence of non-apoptotic cell death in ten different rodent retinal degeneration models (2). Similar observations were also made in monkey retinal cell cultures subjected to hypoxia/reoxygenation where calpain proteolytic activity was in fact found to inactivate caspase-type proteases (39). Overall, our results may provide important insights for potential therapy developments, suggesting the targeting of non-apoptotic processes, such as excessive cGMP-signalling or calpain activity as a feasible treatment approach.

Variable speed of progression, functional and degeneration phenotypes

The four different *Pde6a* mutants used here were found to have a variable pace of degeneration, corresponding to the effect of the genetic mutations on cGMP accumulation and – as a deduced from this – PDE6 function.

Remarkably, there is a clear discrepancy between the extent of cGMP accumulation observed in *Pde6b* mutant *rd1* photoreceptors and *Pde6a* V685M photoreceptors. Even though the speed of photoreceptor loss is comparable in both RP models, and both PDE6 protein isoforms have been suggested to be enzymatically equivalent (40), in *rd1* retina 6-15% of photoreceptors show dramatic cGMP accumulation

(6) compared to only 0.6% in V685M retina. The extent of *rd1* cGMP accumulation might be explained by a complete loss of PDE6 expression caused by the *rd1* mutation. In the V685M mutant a minor amount of residual PDE6 activity could prevent such a dramatic accumulation of cGMP; however, even a more moderate elevation of cGMP (undetectable by immunohistochemistry) would still result in rapid photoreceptor degeneration.

The phenotype of the novel R562W mutant is intermediate between that of the V685M and the D670G mutant. Unexpectedly, we found that the pathogenic effect of this underlying c.1684C>T substitution is twofold, at two distinct levels: pre-mRNA splicing and protein enzymatic activity. The C>T transition (+64 of exon 13) is predicted to create an Exonic Splicing Silencer (ESS) site (-TGGTGG-) which results in an in-frame skipping of exon 13. The splicing defect induced by the mutation is only partially penetrant and therefore correctly spliced transcripts derived from the mutant allele, which accounts for about 50-60% of transcripts (Fig.2B), will be translated into a full-length protein carrying the R562W substitution. This R562W mutant protein exhibits a reduction of its catalytic activity to about 10% compared to the WT protein. The shortened p.541_576del protein derived from *Pde6a* transcripts lacking exon 13, lacks a strongly conserved segment of the protein's catalytic domain, including one of the two metal ion binding motifs which are important for functional activity (41). Therefore, we reason that the shortened protein may essentially be non-functional or with very strongly reduced activity. Moreover, protein stability of such shortened proteins is often reduced as also documented for mouse models for retinal dystrophies (42). The findings with the R562W mutant complements prior reports of splicing defects in retinal PDE6 genes caused by single nucleotide exonic mutations such as the c.1814G>A/p.N605S in *Pde6b* in the *atrd3* mouse model (43) or the c.2368G>A/p.E790K in the human *PDE6C* in a family with achromatopsia (44). It also emphasizes the need for a thorough transcript analysis of the mutant gene for a proper assessment of its consequences.

Typical for RP, the loss of the primarily affected rod photoreceptors was followed by a secondary loss of cone photoreceptors via a well-known but still etiologically unresolved mechanism (45). Differences in rod loss among the mutants thus corresponded well to the functional data in Ganzfeld ERG at PN30, although these essentially represent cone function. In particular, ERG amplitudes in the D670G mutants were at level with $Rho^{-/-}$ animals (30), while no more responses were left in V685M mutants. The results in the other lines were in between these extreme values. This correlation allows to provide an estimate of the minimal number of rod photoreceptors that still need to be physically present (*i.e.* surviving) in order to maintain residual cone photoreceptor function at a recordable level. In our setting, at least one intact row of rod photoreceptors was required in order to still preserve cone ERG at P30.

R562W/V685M: A new patient-matched, compound heterozygous animal model for RP

Current research on the mechanisms underlying recessive RP usually employs homozygous animal models, such as the $Pde6b^{rd1/rd1}$ mouse (11,46). In RP patients in outbred populations homozygosity is, however, less common than compound heterozygous mutations with two different disease-causing alleles (47,48). This creates incongruence between the currently used animal models for RP and the actual patient situation, in particular because the pathological consequences of compound heterozygous versus homozygous mutations are to date only poorly understood.

According to H.J. Muller the action of recessive mutations may be explained be either constituting an amorph (classical *null* mutation) or a hypomorphic mutation, the latter associated with lower amounts of the gene product or reduced function or activity (49). If the amounts of the gene product or its activity is strictly additive and proportional to the phenotypic outcome, then one would expect for compound heterozygous mutations an intermediate disease phenotype, which lies between the two respective homozygous disease phenotypes. However this must not be necessarily true, for instance in traits that develop in a threshold dependent manner. Moreover one needs to consider situation of intra-allelic complementation in which different mutations in the same gene may "neutralize" each other, a phenomenon most likely to occur in proteins that act as oligomers (50). Thus, comparative studies between homozygotes and compound heterozygotes bear important insights into the mode of action of

individual mutations, potential interactions between different mutant proteins, and the correlation of the disease phenotype with the amount and functional activity of the mutant gene product.

Our results with the V685M/R562W retinal degeneration suggest that for the *Pde6a* gene a classic additive model for the interaction of the two mutations is valid, resulting in a proportional relationship between the amount of or the functional activity of the mutant gene product and the phenotypic outcome is valid. If generalized, this model enables to predict the outcome of an unknown genotype in terms of severity and disease progression if the phenotype of two adjoined genotypes in *Pde6a* is known. Our findings also demonstrate that a genotype-phenotype correlation with high predictability exists for *Pde6a*-associated retinal degeneration at least in murine mutants raised on the same genetic background. Therefore, the lack of consistent genotype-phenotype correlation sometimes observed in human RP patients is most likely due to secondary genetic factors and/or differences in lifestyle (51,52).

Nevertheless, our study provides a rational basis for predictions on human RP phenotypes and disease progression in compound heterozygous situations, provided the pathological consequences of at least one of the disease-causing alleles are known. This knowledge could be very valuable for patient counselling and also – if suitable therapeutic approaches become available in the future – for the determination of the best possible time-frame for therapeutic interventions.

The large genetic diversity in RP causing mutations - even if affecting the same gene - may cause very different degeneration phenotypes. Our detailed characterization of different homozygous and compound heterozygous mouse models for RP will provide a basis for further investigations, in particular on the development of future individualized therapies. The wide availability of a large variety of animal models will likely facilitate the generation of ‘personalized’ compound heterozygous animal models that match the human disease condition very closely. Such models could greatly speed up the development of a more personalized medicine in the field of retinal degenerations. For the individual RP patient, the prediction of time-courses for *Pde6a*-related retinal degeneration is currently of particular value. Upon availability of causative or symptomatic treatment options, the definition of a window-of-opportunity for clinical interventions will gain further importance.

Acknowledgements

We thank K. Masarini, N. Rieger, and Britta Baumann for skilful technical assistance. We thank Dr. Muayyad Al-Ubaidi (University of Oklahoma, Oklahoma City, USA) for kindly providing the 661W cell line.

Funding

This work was supported by the Kerstan Foundation (RD-CURE), the 2nd People’s Hospital Of Yunnan Province & 4th Affiliated Hospital Of Kunming Medical University/China, Deutsche Forschungsgemeinschaft [DFG PA1751/4-1], Alcon Research Institute, European Commission [DRUGSFORD: HEALTH-F2-2012-304963; EyeTN: [FP7-People-2012-ITN-317472].

Abbreviations

CNG, cyclic nucleotide gated; ERG, electroretinography; GCL, ganglion cell layer; INL, inner nuclear layer; IPL, inner plexiform layer; OCT, optic coherence tomography; ONL, Outer nuclear layer; OS, outer segment; PDE, phosphodiesterase; PN, post-natal; RP, Retinitis Pigmentosa; RPE, retinal pigment epithelium; SLO, scanning laser ophthalmoscopy; TUNEL, terminal UDP nick-end labelling.

References

1. Chizzolini,M., Galan,A., Milan,E., Sebastiani,A., Costagliola,C., Parmeggiani,F. (2011) Good epidemiologic practice in retinitis pigmentosa: from phenotyping to biobanking. Curr. Genomics, 12, 260-266.
2. Arango-Gonzalez,B., Trifunovic,D., Sahaboglu A, Kranz K, Michalakis S, Farinelli P, Koch S, Koch F, Cottet S, Janssen Bienhold U, et al. (2014) Identification of a common non-apoptotic cell death mechanism in hereditary retinal degeneration. PLoS One., 9, e112142.
3. Olshevskaya,E.V., Ermilov,A.N., Dizhoor,A.M. (2002) Factors that affect regulation of cGMP synthesis in vertebrate photoreceptors and their genetic link to human retinal degeneration. Mol. Cell Biochem., 230, 139-147.
4. Michalakis,S., Geiger,H., Haverkamp,S., Hofmann,F., Gerstner,A., Biel,M. (2005) Impaired opsin targeting and cone photoreceptor migration in the retina of mice lacking the cyclic nucleotide-gated channel CNGA3. Invest Ophthalmol. Vis. Sci., 46, 15161524.
5. Farber,D.B., Lolley,R.N. (1974) Cyclic guanosine monophosphate: elevation in degenerating photoreceptor cells of the C3H mouse retina. Science, 186, 449-451.
6. Sahaboglu,A., Paquet-Durand,O., Dietter,J., Dengler,K., Bernhard-Kurz,S., Ekström,P., Hitzmann,B., Ueffing,M., Paquet-Durand,F. (2013) Retinitis Pigmentosa: Rapid neurodegeneration is governed by slow cell death mechanisms. Cell Death & Dis., 4, e488.

7. Hamel,C. (2006) Retinitis pigmentosa. Orphanet. J Rare. Dis., 1, 40.
8. Bayes,M., Giordano,M., Balcells,S., Grinberg,D., Vilageliu,L., Martinez,I., Ayuso,C., Benitez,J., Ramos-Arroyo,M.A., Chivelet,P., . (1995) Homozygous tandem duplication within the gene encoding the beta-subunit of rod phosphodiesterase as a cause for autosomal recessive retinitis pigmentosa. Hum. Mutat., 5, 228-234.
9. Dryja,T.P., Rucinski,D.E., Chen,S.H., Berson,E.L. (1999) Frequency of mutations in the gene encoding the alpha subunit of rod cGMP-phosphodiesterase in autosomal recessive retinitis pigmentosa. Invest Ophthalmol. Vis. Sci., 40, 1859-1865.
10. Dvir,L., Srour,G., Abu-Ras,R., Miller,B., Shalev,S.A., Ben-Yosef,T. (2010) Autosomalrecessive early-onset retinitis pigmentosa caused by a mutation in PDE6G, the gene encoding the gamma subunit of rod cGMP phosphodiesterase. Am J Hum Genet., 87, 258-264.
11. Keeler,C.E. (1924) The Inheritance of a Retinal Abnormality in White Mice. Proc. Natl. Acad. Sci. U. S. A., 10, 329-333.
12. Chang,B., Hawes,N.L., Hurd,R.E., Davisson,M.T., Nusinowitz,S., Heckenlively,J.R. (2002) Retinal degeneration mutants in the mouse. Vision Research, 42, 517-525.
13. Sakamoto,K., McCluskey,M., Wensel,T.G., Naggert,J.K., Nishina,P.M. (2009) New mouse models for recessive retinitis pigmentosa caused by mutations in the Pde6a gene. Hum. Mol. Genet., 18, 178-192.
14. Mattapallil,M.J., Wawrousek,E.F., Chan,C.C., Zhao,H., Roychoudhury,J., Ferguson,T.A., Caspi,R.R. (2012) The Rd8 mutation of the Crb1 gene is present in vendor lines of C57BL/6N mice and embryonic stem cells, and confounds ocular induced mutant phenotypes. Invest Ophthalmol. Vis. Sci., 53, 2921-2927.
15. Grau,T., Artemyev,N.O., Rosenberg,T., Dollfus,H., Haugen,O.H., Cumhur,S.E., Jurklies,B., Andreasson,S., Kernstock,C., Larsen,M., et al. (2011) Decreased catalytic activity and altered activation properties of PDE6C mutants associated with autosomal recessive achromatopsia. Hum Mol. Genet., 20, 719-730.
16. Muradov,K.G., Boyd,K.K., Martinez,S.E., Beavo,J.A., Artemyev,N.O. (2003) The GAFa domains of rod cGMP-phosphodiesterase 6 determine the selectivity of the enzyme dimerization. J Biol. Chem., 278, 10594-10601.
17. Natochin,M., Artemyev,N.O. (2000) Mutational analysis of functional interfaces of transducin. Methods Enzymol., 315, 539-554.
18. Ekstrom,P.A., Ueffing,M., Zrenner,E., Paquet-Durand,F. (2014) Novel in situ activity assays for the quantitative molecular analysis of neurodegenerative processes in the retina. Curr. Med. Chem., 21, 3478-3493.
19. Paquet-Durand,F., Silva,J., Talukdar,T., Johnson,L.E., Azadi,S., van Veen,T., Ueffing,M., Hauck,S.M., Ekstrom,P.A. (2007) Excessive activation of poly(ADP-ribose) polymerase contributes to inherited photoreceptor degeneration in the retinal degeneration 1 mouse. J. Neurosci., 27, 10311-10319.
20. Tanimoto,N., Muehlfriedel,R.L., Fischer,M.D., Fahl,E., Humphries,P., Biel,M., Seeliger,M.W. (2009) Vision tests in the mouse: Functional phenotyping with electroretinography. Front Biosci., 14, 2730-2737.
21. Fischer,M.D., Huber,G., Beck,S.C., Tanimoto,N., Muehlfriedel,R., Fahl,E., Grimm,C., Wenzel,A., Reme,C.E., van de Pavert,S.A., et al. (2009) Noninvasive, in vivo assessment of mouse retinal structure using optical coherence tomography. PLoS. ONE., 4, e7507.
22. Huber,G., Beck,S.C., Grimm,C., Sahaboglu-Tekgoz,A., Paquet-Durand,F., Wenzel,A., Humphries,P., Redmond,T.M., Seeliger,M.W., Fischer,M.D. (2009) Spectral domain optical coherence tomography in mouse models of retinal degeneration. Invest Ophthalmol. Vis. Sci., 50, 5888-5895.
23. Garcia,G.M., Beck,S.C., Muhlfriedel,R., Julien,S., Schraermeyer,U., Seeliger,M.W. (2014) Towards a quantitative OCT image analysis. PLoS One., 9, e100080.
24. Seeliger,M.W., Beck,S.C., Pereyra-Munoz,N., Dangel,S., Tsai,J.Y., Luhmann,U.F., van de Pavert,S.A., Wijnholds,J., Samardzija,M., Wenzel,A., et al. (2005) In vivo confocal imaging of the retina in animal models using scanning laser ophthalmoscopy. Vision Res., 45, 3512-3519.
25. Chang,G.Q., Hao,Y., Wong,F. (1993) Apoptosis: final common pathway of photoreceptor death in rd, rds, and rhodopsin mutant mice. Neuron, 11, 595-605.
26. Sanges,D., Comitato,A., Tammaro,R., Marigo,V. (2006) Apoptosis in retinal degeneration involves cross-talk between apoptosis-inducing factor (AIF) and caspase12 and is blocked by calpain inhibitors. Proc. Natl. Acad. Sci. U. S. A., 103, 1736617371.
27. Zeiss,C.J., Neal,J., Johnson,E.A. (2004) Caspase-3 in postnatal retinal development and degeneration. Invest Ophthalmol. Vis. Sci., 45, 964-970.
28. Paquet-Durand,F., Beck,S., Michalakis,S., Goldmann,T., Huber,G., Muhlfriedel,R., Trifunovic,D., Fischer,M.D., Fahl,E., Duetsch,G., et al. (2011) A key role for cyclic nucleotide gated (CNG) channels in cGMP-related retinitis pigmentosa. Hum. Mol. Genet., 20, 941-947.
29. Young,R.W. (1984) Cell death during differentiation of the retina in the mouse. J. Comp Neurol., 229, 362-373.
30. Jaissle,G.B., May,C.A., Reinhard,J., Kohler,K., Fauser,S., Lutjen-Drecoll,E., Zrenner,E., Seeliger,M.W. (2001) Evaluation of the rhodopsin knockout mouse as a model of pure cone function. Invest Ophthalmol. Vis. Sci., 42, 506-513.
31. Samardzija,M., Tanimoto,N., Kostic,C., Beck,S., Oberhauser,V., Joly,S., Thiersch,M., Fahl,E., Arsenijevic,Y., von,L.J., et al. (2009) In conditions of limited chromophore supply rods entrap 11-cis-retinal leading to loss of cone function and cell death. Hum Mol. Genet., 18, 1266-1275.
32. Price,B.A., Sandoval,I.M., Chan,F., Simons,D.L., Wu,S.M., Wensel,T.G., Wilson,J.H. (2011) Mislocalization and degradation of human P23H-rhodopsin-GFP in a knockin mouse model of retinitis pigmentosa. Invest Ophthalmol. Vis. Sci., 52, 9728-9736.
33. Stuck,M.W., Conley,S.M., Naash,M.I. (2014) The Y141C knockin mutation in RDS leads to complex phenotypes in the mouse. Hum Mol. Genet., 23, 6260-6274.
34. Rybalkin,S.D., Hinds,T.R., Beavo,J.A. (2013) Enzyme assays for cGMP hydrolyzing phosphodiesterases. Methods Mol. Biol., 1020, 51-62.
35. Bowes,C., Li,T., Danciger,M., Baxter,L.C., Applebury,M.L., Farber,D.B. (1990) Retinal degeneration in the rd mouse is caused by a defect in the beta subunit of rod cGMPphosphodiesterase. Nature, 347, 677-680.
36. Fox,D.A., Poblenz,A.T., He,L. (1999) Calcium overload triggers rod photoreceptor apoptotic cell death in chemical-induced and inherited retinal degenerations. Ann. N. Y. Acad. Sci., 893, 282-285.
37. Yamashima,T. (2004) Ca2+-dependent proteases in ischemic neuronal death: a conserved 'calpain-cathepsin cascade' from nematodes to primates. Cell Calcium, 36, 285-293.
38. Paquet-Durand,F., Azadi,S., Hauck,S.M., Ueffing,M., van Veen,T., Ekstrom,P. (2006) Calpain is activated in degenerating photoreceptors in the rd1 mouse. J. Neurochem., 96, 802-814.
39. Nakajima,E., Hammond,K.B., Rosales,J.L., Shearer,T.R., Azuma,M. (2011) Calpain, not caspase, is the causative protease for hypoxic damage in cultured monkey retinal cells. Invest Ophthalmol. Vis. Sci., 52, 7059-7067.
40. Muradov,H., Boyd,K.K., Artemyev,N.O. (2010) Rod phosphodiesterase-6 PDE6A and PDE6B subunits are enzymatically equivalent. J Biol. Chem., 285, 39828-39834.

41. He,F., Seryshev,A.B., Cowan,C.W., Wensel,T.G. (2000) Multiple zinc binding sites in retinal rod cGMP phosphodiesterase, PDE6alpha beta. J Biol. Chem., 275, 2057220577.
42. Chang,B., Khanna,H., Hawes,N., Jimeno,D., He,S., Lillo,C., Parapuram,S.K., Cheng,H., Scott,A., Hurd,R.E., et al. (2006) In-frame deletion in a novel centrosomal/ciliary protein CEP290/NPHP6 perturbs its interaction with RPGR and results in early-onset retinal degeneration in the rd16 mouse. Hum Mol. Genet., 15, 1847-1857.
43. Muradov,H., Boyd,K.K., Kerov,V., Artemyev,N.O. (2012) Atypical retinal degeneration 3 in mice is caused by defective PDE6B pre-mRNA splicing. Vision Res., 57, 1-8.
44. Chang,B., Grau,T., Dangel,S., Hurd,R., Jurklies,B., Sener,E.C., Andreasson,S., Dollfus,H., Baumann,B., Bolz,S., et al. (2009) A homologous genetic basis of the murine cpfl1 mutant and human achromatopsia linked to mutations in the PDE6C gene. Proc. Natl. Acad. Sci. U. S. A, 106, 19581-19586.
45. Kranz,K., Paquet-Durand,F., Weiler,R., Janssen-Bienhold,U., Dedek,K. (2013) Testing for a gap junction-mediated bystander effect in retinitis pigmentosa: secondary cone death is not altered by deletion of connexin36 from cones. PLoS One., 8, e57163.
46. Dalke,C., Graw,J. (2005) Mouse mutants as models for congenital retinal disorders. Exp. Eye Res., 81, 503-512.

Figures

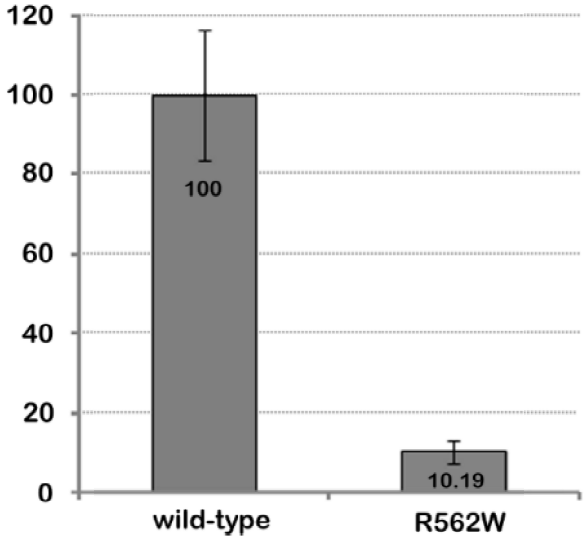


Figure 1: The R562W mutant PDE6 protein shows reduced catalytic activity. cGMP hydrolysis activity of wild-type and R562W mutant chimeric PDE6C/PDE5 protein was determined *in vitro* on purified recombinant protein. The bars were normalized to wild-type enzymatic activity (100%).

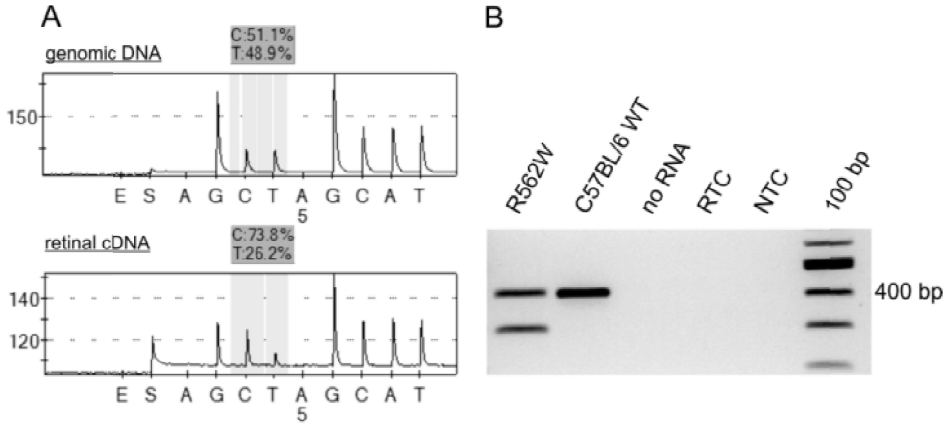


Figure 2: *Pde6a* cDNA analysis in the R562W mouse mutant. (**A**) Pyrosequencing-based relative quantification of wild-type and knock-in *Pde6a* alleles at the genomic DNA level (top) and in reverse transcribed retinal cDNA of a heterozygous *Pde6a*:R562W-knockin mutant. Note that the assayed variant nucleotide position is the actual c.1684C>T mutation and that allelic quantification of cDNA was done on RT-PCR products with primers in exon 12 and 13. (**B**) Qualitative analysis of the integrity of retinal *Pde6a* transcripts in the *Pde6a*:R562W-knockin mouse mutant. RT-PCR products with primers in exons 12 and 15 of *Pde6a* were amplified from retinal RNA of a homozygous *Pde6a*:R562W-knockin mutant and a C57BL/6 wild-type control. The larger sized product of 403 bp is derived from correctly spliced transcripts and the smaller product represent mis-spliced transcripts lacking exon 13. Controls “no RNA”: w/o RNA in the DN*as*e I digestion and subsequent RT and PCR reactions, “RTC”: w/o RNA in the cDNA synthesis and subsequent PCR reaction, “NTC”: No template control for the PCR reaction. 100 bp ladder size standard is shown in the rightmost lane.

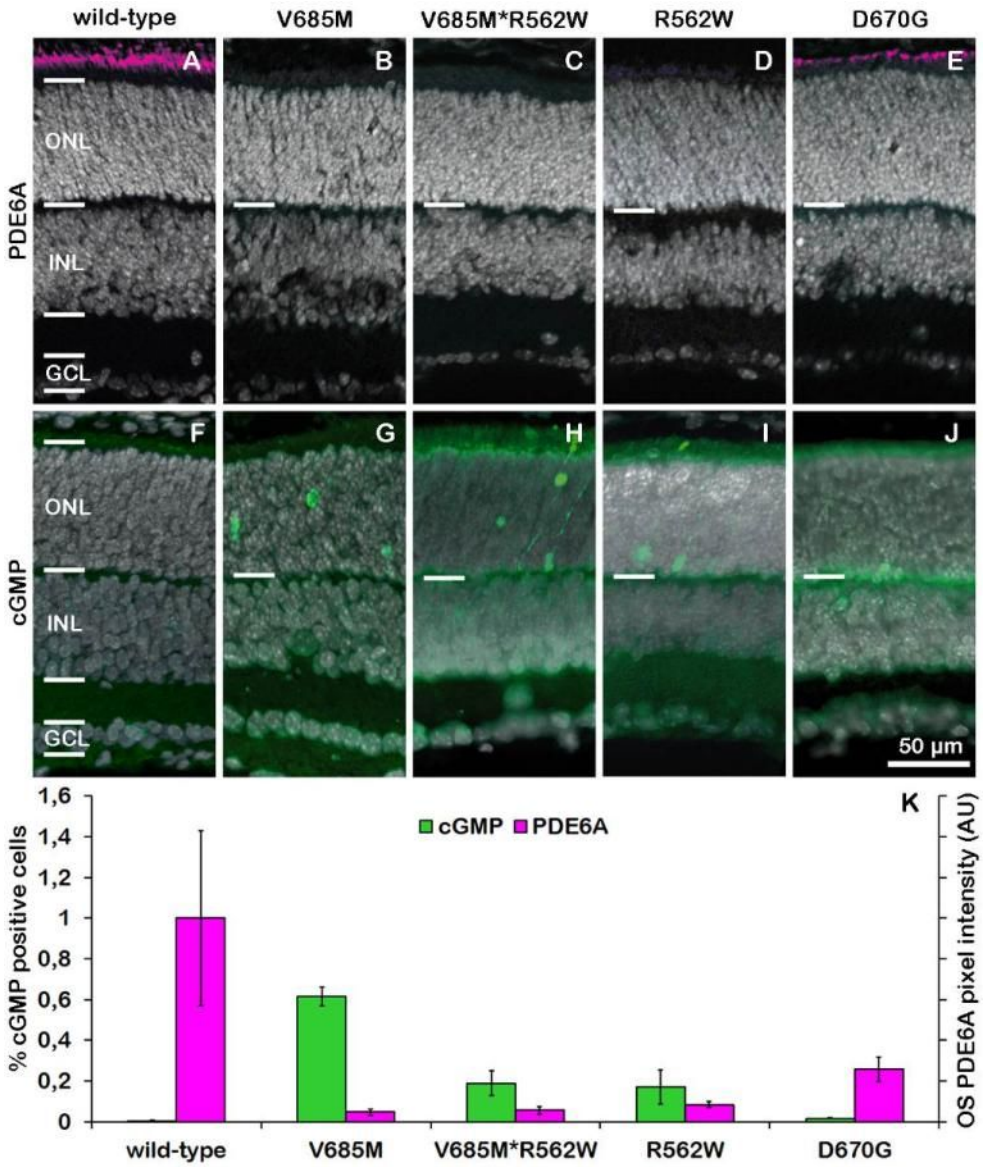


Figure 3: Loss of PDE6A expression causes cGMP accumulation. In the PN11 wild-type (wt) retina (**A**), immunostaining for PDE6A shows strong protein expression in photoreceptor outer segments. In contrast, at the same age, in V685M retina (**B**), the PDE6A protein is undetectable. Similarly, in the compound heterozygous V685M*R562W retina (**C**), the protein is essentially absent. In R562W retina (**D**) small immunoreactive dots at the level of the outer segments (OS) may indicate a minimal protein expression. In the D670G mutant (**E**), however, there is a clear PDE6A protein expression; albeit at reduced levels when compared to wt. At PN11, wt retina is essentially negative for cGMP immunoreactivity (**F**). All *Pde6a* mutants, however, display individual rod photoreceptor cells that have accumulated large amounts of cGMP (**G**-**J**). The quantification of cGMP positive cells in the outer nuclear layer (ONL) and the PDE6A pixel intensity in the OS (arbitrary units; AU) shows an inverse correlation (**K**). Images are representative for immunostaining performed on retinal sections from at least three independent animals for each genotype.

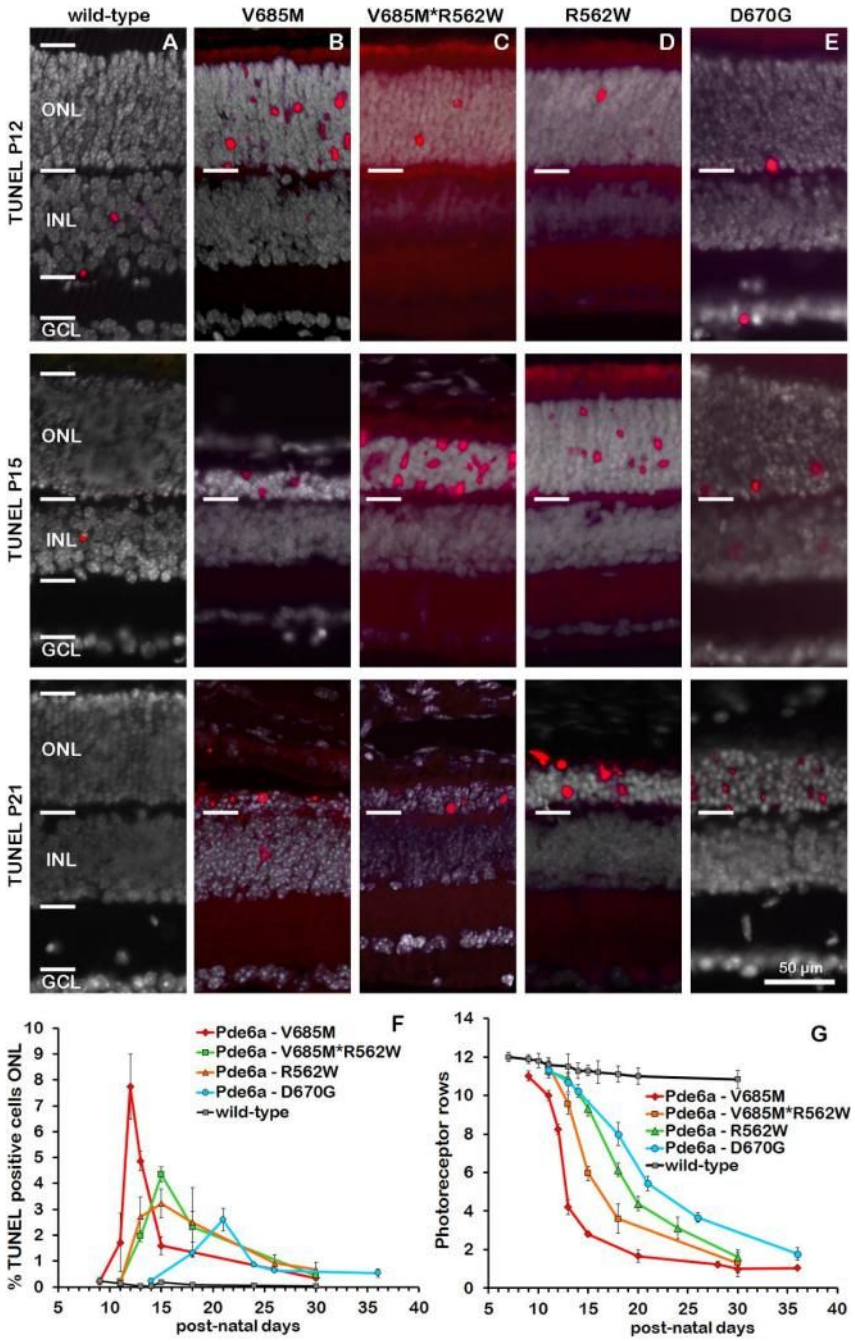


Figure 4: Photoreceptor cell death and survival. The TUNEL assay in the PN11 wild-type (wt) retina (**A**), occasionally labels cells dying due to developmental processes. In contrast, in all *Pde6a* mutants (**B-E**), photoreceptor cell death is dramatically increased. The images show the situation at the peak of cell death for the respective models. The line graph at the bottom left (**F**) illustrates the progression of photoreceptor cell death as evidenced by the TUNEL assay in the different *Pde6a* mutants. The peak times as well as the peak amplitudes correspond to the speed of retinal degeneration, which is illustrated by the loss of photoreceptors (**G**). Images are representative for TUNEL assays performed on retinal sections from at least three independent animals, quantifications in F, G include data from 3-7 animals per genotype and time-point.

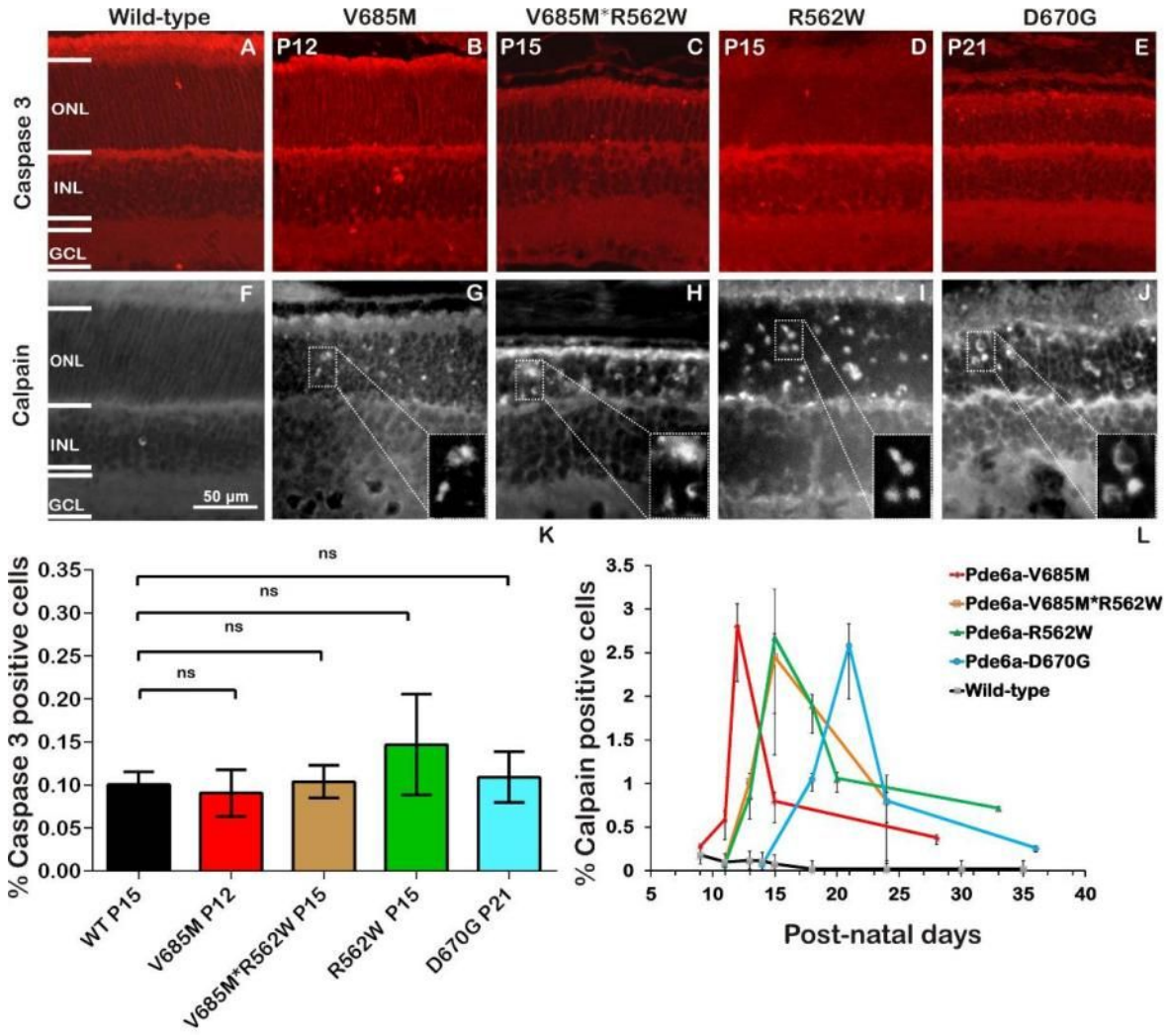


Figure 5: Photoreceptor degeneration in *Pde6a* mutants correlates with calpain not caspase activity. Immunofluorescence for activated caspase-3 – a key marker for classical apoptosis – reveals no major differences between wild-type and mutant retinae (**A-E**). This is also illustrated by the quantification of caspase-3 positive cells, which shows only extremely low numbers of cells at the respective peaks of degeneration, with no significant (n.s.) differences to wild-type (WT) (**K**). In contrast, an *in situ* calpain activity assay reveals strong differences between wild-type (**F**) and all four *Pde6a* mutant genotypes (**G-J**). The progression of calpain activity was analysed over time (**K**) and showed a strong correlation to the extent of cell death and the progression of retinal degeneration (*cf.* Figure 4). Images shown are representative for at least three different stainings obtained at the respective mutant's peak of degeneration. Note that the differences in the progression of *Pde6a* mutant degenerations are also illustrated here by the differences in retinal ONL sizes. Data shown in K and L was obtained from observations made on 4-6 independent specimens for each genotype and time-point.

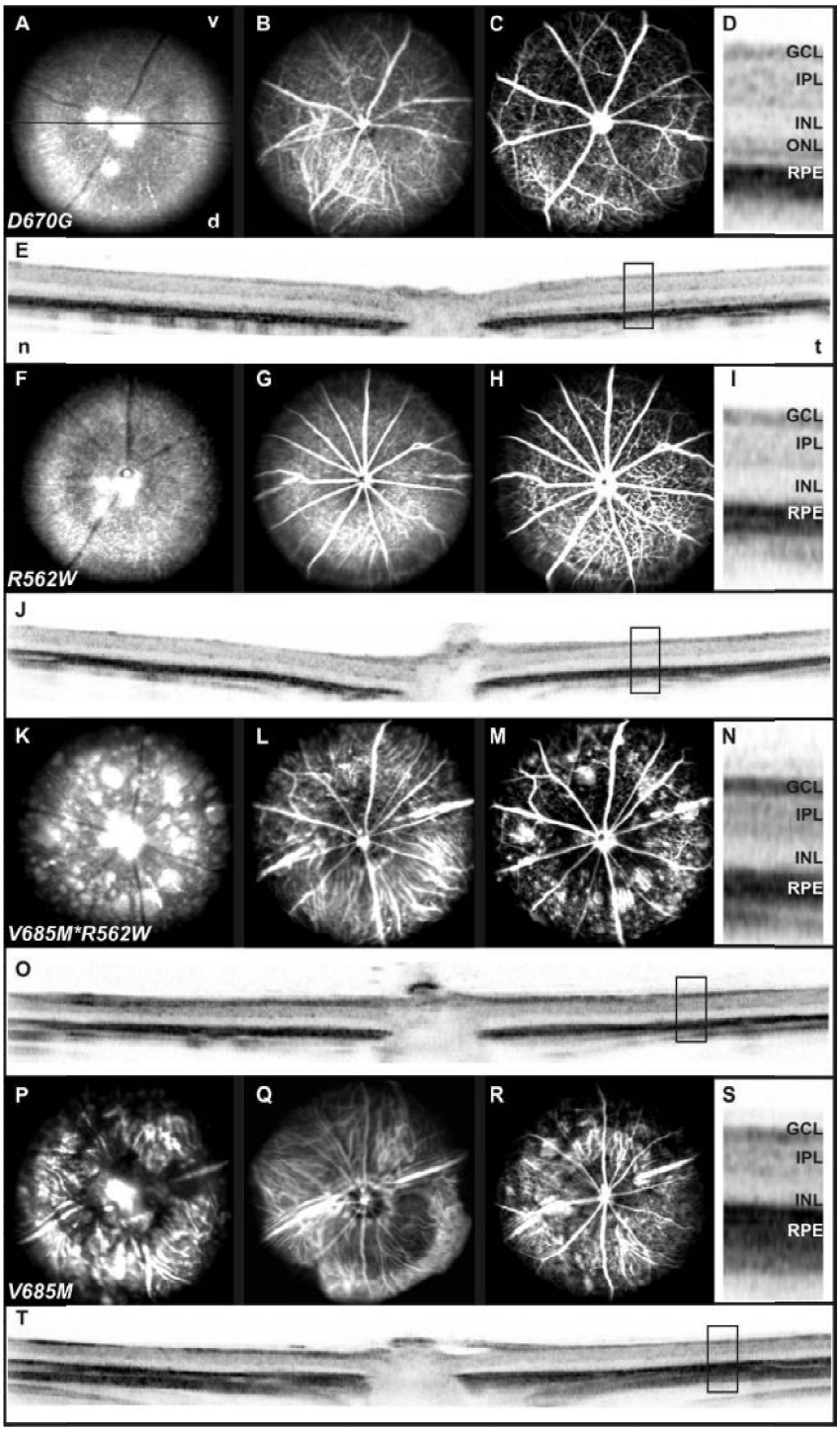


Figure 6: Retinal morphology in *Pde6a* mutants *in vivo*. *D670G* (**A**-**E**)*, R562W* (**F**-**J**)*, V685M*R562W* (**K**-**O**) *and V685M* (**P**-**T**) mutant mice were examined with SLO and SD-OCT at PN30. Examinations included fundus native imaging (**A**, **F**, **K**, **P**), indocyanine-green angiography (**B**, **G**, **L**, **Q**) and fluorescein angiography (**C**, **H**, **M**, **R**). For the OCT analyses horizontal scans through the optic disc head are shown (**E**, **J**, **O**, **T**; magnifications; **D**, **I**, **N**, **S**). All of the mutants analysed revealed retinal degeneration. D670G mice showed several spots spread over the fundus (**A**), which were also visible in fluorescein angiography examination (**C**). Retinal thickness was also affected and a very thin ONL was detected (**D**, **E**). A similar fundus appearance (**F**-**H**) but a more severe decrease in the retinal thickness was found in the R562W mutants (**I**, **J**). V685M*R562W compound heterozygous mice showed large areas of retinal degeneration in native fundus imaging (**K**). The loss of retinal layers demonstrated in OCT imaging (**O**, **N**), allowed the enhanced visualization of choroidal structures (**L**, **M**). V685M exhibited a very severe retinal degeneration in both SLO and OCT imaging (**P**-**T**). Abbreviations: GCL, ganglion cell layer, IPL, inner plexiform layer; INL, inner nuclear layer; ONL, outer nuclear layer; RPE, retinal pigment epithelium; d, dorsal; n, nasal; t, temporal; v, ventral.

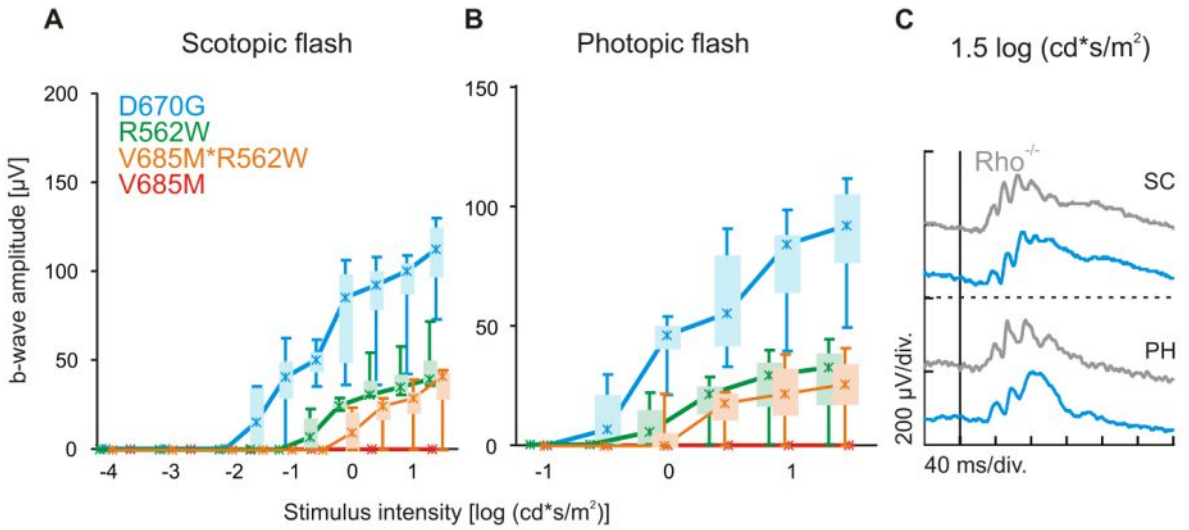
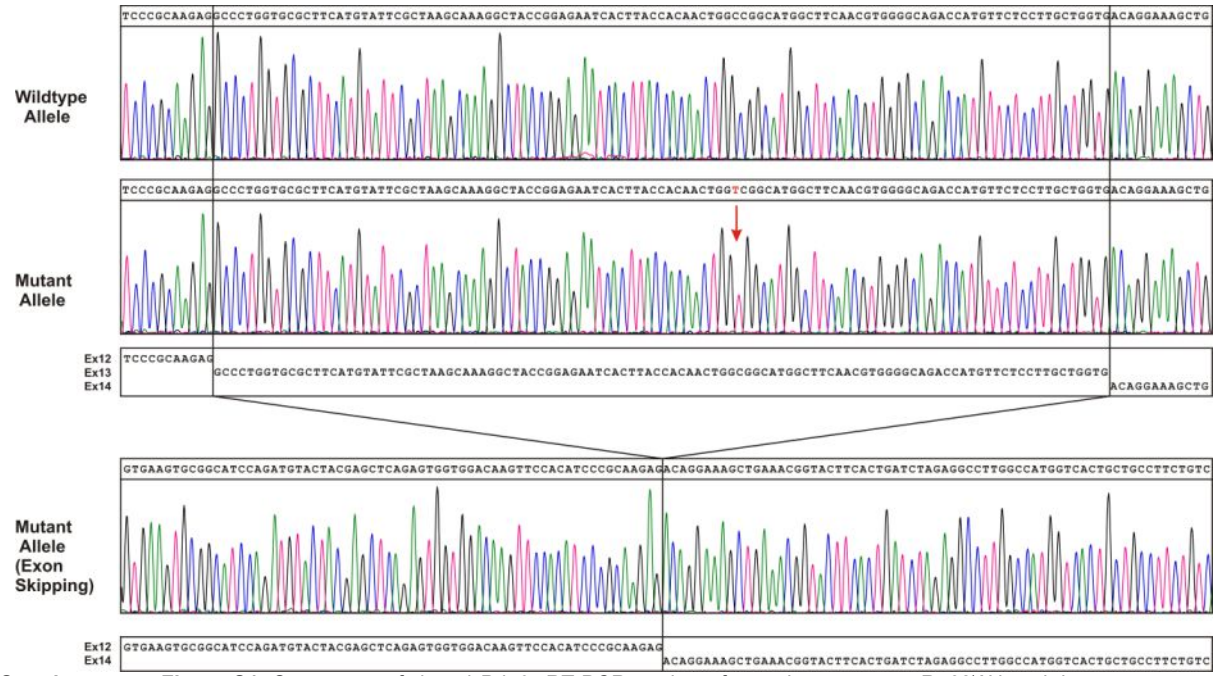
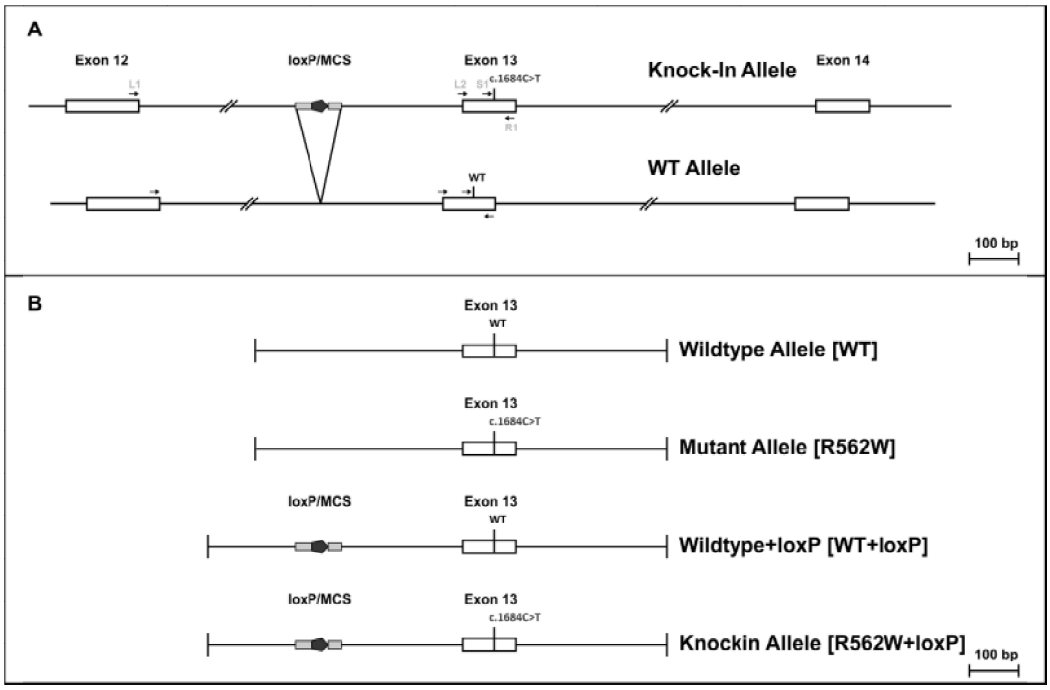


Figure 7: Functional assessment of different *Pde6a* mutants based on electroretinography. Quantitative evaluation of the scotopic (dark-adapted (**A**)) and photopic (light-adapted (**B**)) b-wave amplitude data, acquired at PN30 and shown as box-and whisker-plot (B&W), for the different *Pde6a* variants. In B&W, boxes indicate the 25– 75% quantile range, whiskers the 5% and 95% quantiles, and asterisks the medians of the data. Scotopic (top) and photopic (bottom) single flash ERG at 1.5 log (cd*s/m^2) maximal light intensity (**C**). ERGs from D670G mutant (blue) are similar to ERGs from a Rho$^{-/-}$ mouse, a valid model for pure cone function. D670G (blue), R562W (green), V685M*R562W (orange), and V685M mice (red).

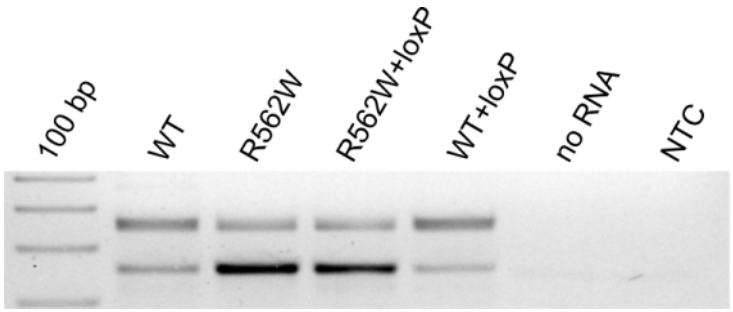
Supplementary Material



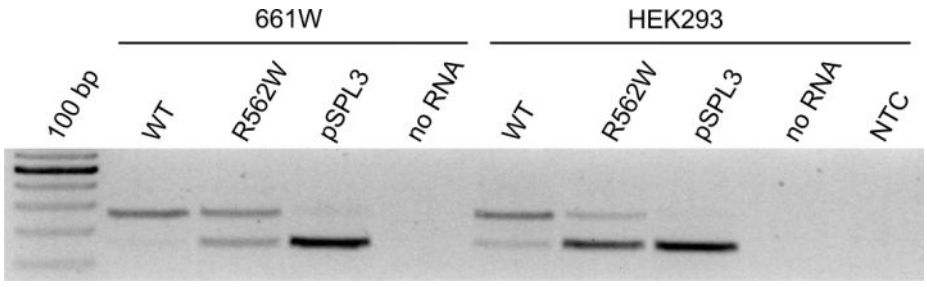
Supplementary Figure S1: Sequences of cloned *Pde6a* RT-PCR products from a heterozygous R562W knock-in mouse mutant. Shown are cDNA sequences derived from properly spliced wild-type (c.1684C marked by a red arrow; top panel) and mutant transcripts (c.1684T; middle panel) and a cDNA from a transcript lacking exon 13 (bottom panel).



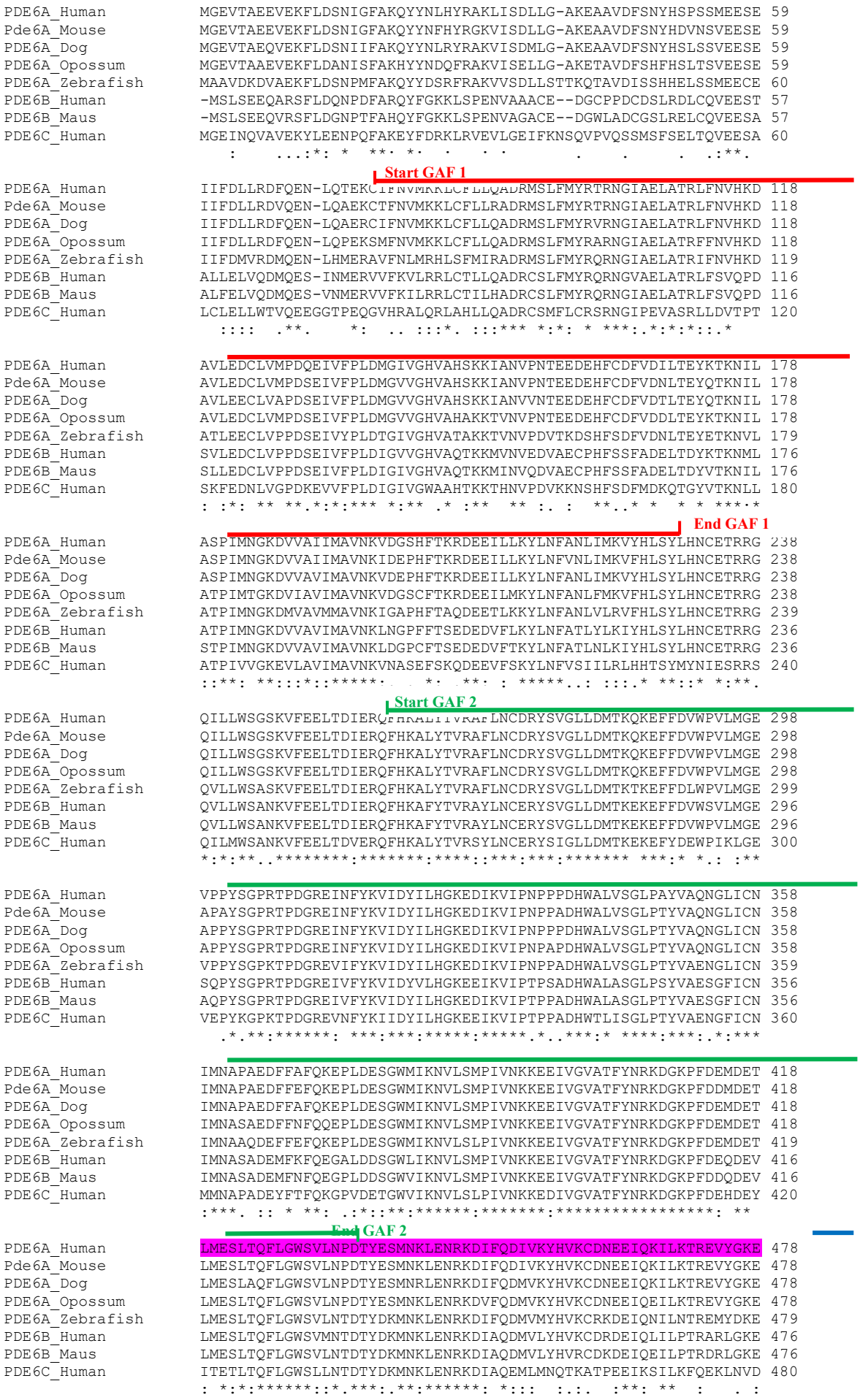
Supplementary Figure S2: Schematic overview of the genomic structure of the central proportion (exon 12 – exon 14) of the R562W knock-in allele and the wild-type allele at the *Pde6a* locus, respectively, (A) and the cloned segment and composition of the minigene constructs used to assess splicing competence (B).



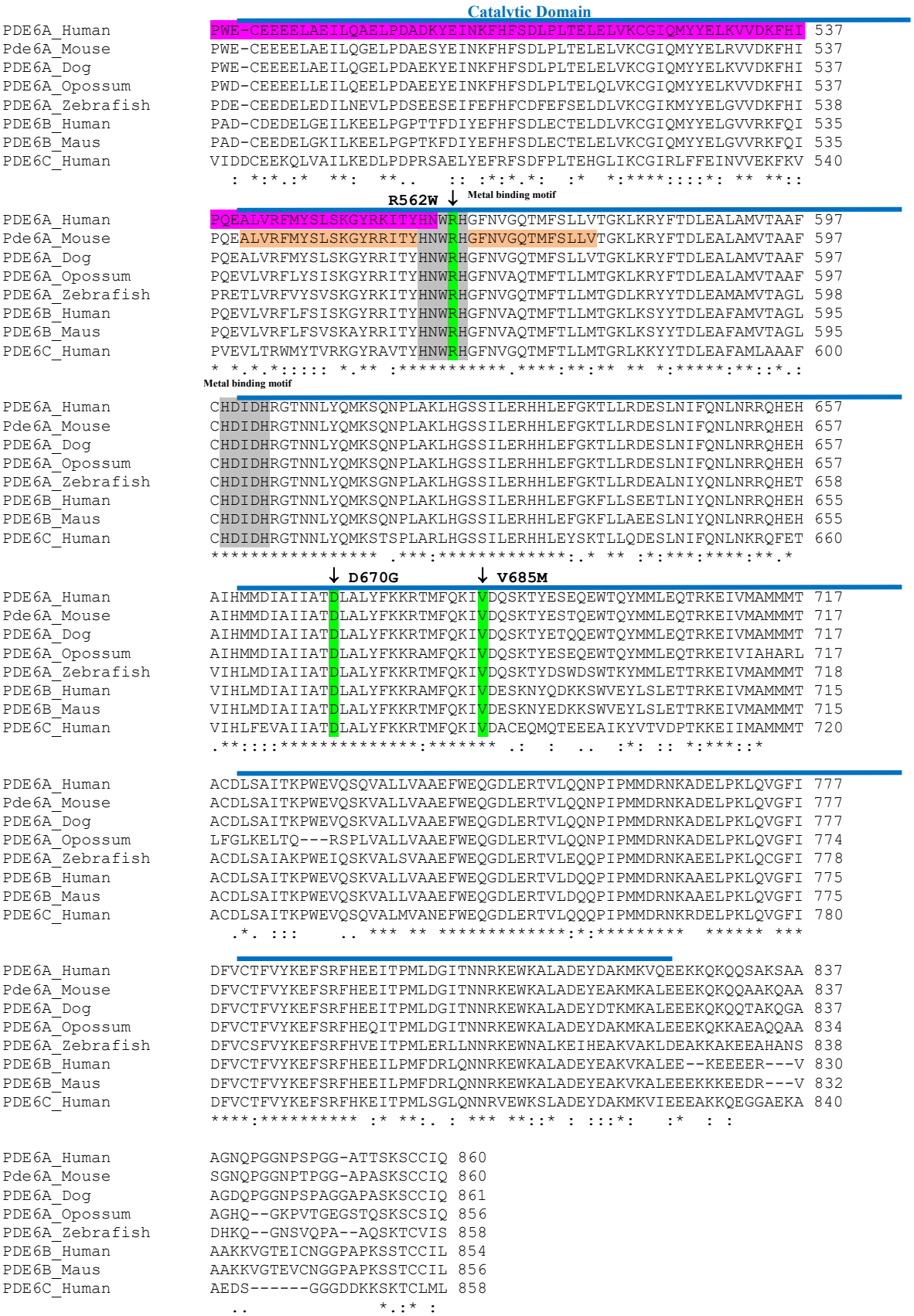
Supplementary Figure S3: RT-PCR from minigene constructs (WT, R562W, R562W and loxP/MCS insert, WT and loxP/MCS insert) transiently expressed in HEK293 cells. Primers located in flanking exons of the pSPL3 vector were used for amplification. Further lanes: '100 bp' - 100 bp ladder size standard, 'no RNA' - no RNA added to cDNA synthesis reaction, and 'NTC' - non-template control for the PCR reaction.



Supplementary Figure S4: RT-PCR from minigene constructs (WT and R562W) transfected either in the mouse photoreceptor-like cell line 661W (left) and HEK293 (right). Further lanes: '100 bp' - 100 bp ladder size standard, 'no RNA' - no RNA added to cDNA synthesis reaction, and 'NTC' - non-template control for the PCR reaction.




```
PDE6A_Human          MGEVTAEEVEKFLDSNIGFAKQYYNLHYRAKLISDLLG-AKEAAVDFSNYHSPSSMEESE 59
Pde6A_Mouse          MGEVTAEEVEKFLDSNIGFAKQYYNFHYRGKVISDLLG-AKEAAVDFSNYHDVNSVEESE 59
PDE6A_Dog            MGEVTAEQVEKFLDSNIIFAKQYYNLRYRAKVISDMLG-AKEAAVDFSNYHSLSSVEESE 59
PDE6A_Opossum        MGEVTAAEVEKFLDANISFAKHYYNDQFRAKVISELLG-AKETAVDFSHFHSLTSVEESE 59
PDE6A_Zebrafish      MAAVDKDVAEKFLDSNPMFAKQYYDSRFRAKVVSDLLSTTKQTAVDISSHHELSSMEECE 60
PDE6B_Human          -MSLSEEQARSFLDQNPDFARQYFGKKLSPENVAAACE--DGCPPDCDSLRDLCQVEEST 57
PDE6B_Maus           -MSLSEEQVRSFLDGNPTFAHQYFGKKLSPENVAGACE--DGWLADCGSLRELCQVEESA 57
PDE6C_Human          MGEINQVAVEKYLEENPQFAKEYFDRKLRVEVLGEIFKNSQVPVQSSMSFSELTQVEESA 60
                        :    ...:*: *  **· *·  ·   · ·       .    .     .  .:**.

PDE6A_Human          IIFDLLRDFQEN-LQTEKCTFNVMKKLCFLLQADRMSLFMYRTRNGIAELATRLFNVHKD 118
Pde6A_Mouse          IIFDLLRDVQEN-LQAEKCTFNVMKKLCFLLRADRMSLFMYRTRNGIAELATRLFNVHKD 118
PDE6A_Dog            IIFDLLRDFQEN-LQAERCIFNVMKKLCFLLQADRMSLFMYRVRNGIAELATRLFNVHKD 118
PDE6A_Opossum        IIFDLLRDFQEN-LQPEKSMFNVMKKLCFLLQADRMSLFMYRARNGIAELATRFFNVHKD 118
PDE6A_Zebrafish      IIFDMVRDMQEN-LHMERAVFNLMRHLSFMIRADRMSLFMYRQRNGIAELATRIFNVHKD 119
PDE6B_Human          ALLELVQDMQES-INMERVVFKVLRRLCTLLQADRCSLFMYRQRNGVAELATRLFSVQPD 116
PDE6B_Maus           ALFELVQDMQES-VNMERVVFKILRRLCTILHADRCSLFMYRQRNGIAELATRLFSVQPD 116
PDE6C_Human          LCLELLWTVQEEGGTPEQGVHRALQRLAHLLQADRCSMFLCRSRNGIPEVASRLLDVTPT 120
                       ::::  .**.    *:  .. :::*. :::*** *:*: * ***:.*:*:*::.*

PDE6A_Human          AVLEDCLVMPDQEIVFPLDMGIVGHVAHSKKIANVPNTEEDEHFCDFVDILTEYKTKNIL 178
Pde6A_Mouse          AVLEDCLVMPDSEIVFPLDMGVVGHVAHSKKIANVPNTEEDEHFCDFVDNLTEYQTKNIL 178
PDE6A_Dog            AVLEECLVAPDSEIVFPLDMGVVGHVAHSKKIANVVNTEEDEHFCDFVDTLTEYQTKNIL 178
PDE6A_Opossum        AVLEDCLVMPDSEIVFPLDMGVVGHVAHAKKTVNVPNTEEDEHFCDFVDDLTEYKTKNIL 178
PDE6A_Zebrafish      ATLEECLVPPDSEIVYPLDTGIVGHVATAKKTVNVPDVTKDSHFSDFVDNLTEYETKNVL 179
PDE6B_Human          SVLEDCLVPPDSEIVFPLDIGVVGHVAQTKKMVNVEDVAECPHFSSFADELTDYKTKNML 176
PDE6B_Maus           SLLEDCLVPPDSEIVFPLDIGIVGHVAQTKKMINVQDVAECPHFSSFADELTDYVTKNIL 176
PDE6C_Human          SKFEDNLVGPDKEVVFPLDIGIVGWAAHTKKTHNVPDVKKNSHFSDFMDKQTGYVTKNLL 180
                     : :*: ** **.*:*:*** *:** .* :**  ** :. :  **..* *  * * ***·*

PDE6A_Human          ASPIMNGKDVVAIIMAVNKVDGSHFTKRDEEILLKYLNFANLIMKVYHLSYLHNCETRRG 238
Pde6A_Mouse          ASPIMNGKDVVAIIMAVNKIDEPHFTKRDEEILLKYLNFVNLIMKVFHLSYLHNCETRRG 238
PDE6A_Dog            ASPIMNGKDVVAVIMAVNKVDEPHFTKRDEEILLKYLNFANLIMKVYHLSYLHNCETRRG 238
PDE6A_Opossum        ATPIMTGKDVIAVIMAVNKVDGSCFTKRDEEILMKYLNFANLFMKVFHLSYLHNCETRRG 238
PDE6A_Zebrafish      ATPIMNGKDMVAVMMAVNKIGAPHFTAQDEETLKKYLNFANLVLRVFHLSYLHNCETRRG 239
PDE6B_Human          ATPIMNGKDVVAVIMAVNKLNGPFFTSEDEDVFLKYLNFATLYLKIYHLSYLHNCETRRG 236
PDE6B_Maus           STPIMNGKDVVAVIMAVNKLDGPCFTSEDEDVFTKYLNFATLNLKIYHLSYLHNCETRRG 236
PDE6C_Human          ATPIVVGKEVLAVIMAVNKVNASEFSKQDEEVFSKYLNFVSIILRLHHTSYMYNIESRRS 240
                     ::**: **:::*::*****·  ·  *·  **: : *****..: :::.* **::* *:**.

PDE6A_Human          QILLWSGSKVFEELTDIERQFHKALFTVRAFLNCDRYSVGLLDMTKQKEFFDVWPVLMGE 298
Pde6A_Mouse          QILLWSGSKVFEELTDIERQFHKALYTVRAFLNCDRYSVGLLDMTKQKEFFDVWPVLMGE 298
PDE6A_Dog            QILLWSGSKVFEELTDIERQFHKALYTVRAFLNCDRYSVGLLDMTKQKEFFDVWPVLMGE 298
PDE6A_Opossum        QILLWSGSKVFEELTDIERQFHKALYTVRAFLNCDRYSVGLLDMTKQKEFFDVWPVLMGE 298
PDE6A_Zebrafish      QVLLWSASKVFEELTDIERQFHKALYTVRAFLNCDRYSVGLLDMTKTKEFFDLWPVLMGE 299
PDE6B_Human          QVLLWSANKVFEELTDIERQFHKAFYTVRAYLNCERYSVGLLDMTKEKEFFDVWSVLMGE 296
PDE6B_Maus           QVLLWSANKVFEELTDIERQFHKAFYTVRAYLNCERYSVGLLDMTKEKEFFDVWPVLMGE 296
PDE6C_Human          QILMWSANKVFEELTDVERQFHKALYTVRSYLNCERYSIGLLDMTKEKEFYDEWPIKLGE 300
                     *:*:**..********:*******:****::***:***:******* ***:* *.: :**

PDE6A_Human          VPPYSGPRTPDGREINFYKVIDYILHGKEDIKVIPNPPPDHWALVSGLPAYVAQNGLICN 358
Pde6A_Mouse          APAYSGPRTPDGREINFYKVIDYILHGKEDIKVIPNPPADHWALVSGLPTYVAQNGLICN 358
PDE6A_Dog            APPYSGPRTPDGREINFYKVIDYILHGKEDIKVIPNPPPDHWALVSGLPTYVAQNGLICN 358
PDE6A_Opossum        APPYSGPRTPDGREINFYKVIDYILHGKEDIKVIPNPAPDHWALVSGLPTYVAQNGLICN 358
PDE6A_Zebrafish      VPPYSGPKTPDGREVIFYKVIDYILHGKEDIKVIPNPPADHWALVSGLPTYVAENGLICN 359
PDE6B_Human          SQPYSGPRTPDGREIVFYKVIDYVLHGKEEIKVIPTPSADHWALASGLPSYVAESGFICN 356
PDE6B_Maus           AQPYSGPRTPDGREIVFYKVIDYILHGKEDIKVIPTPPADHWALASGLPTYVAESGFICN 356
PDE6C_Human          VEPYKGPKTPDGREVNFYKIIDYILHGKEEIKVIPTPPADHWTLISGLPTYVAENGFICN 360
                       .*.**:******: ***:***:*****:*****.*..***:* ****:***:.*:***

PDE6A_Human          IMNAPAEDFFAFQKEPLDESGWMIKNVLSMPIVNKKEEIVGVATFYNRKDGKPFDEMDET 418
Pde6A_Mouse          IMNAPAEDFFEFQKEPLDESGWMIKNVLSMPIVNKKEEIVGVATFYNRKDGKPFDDMDET 418
PDE6A_Dog            IMNAPAEDFFAFQKEPLDESGWMIKNVLSMPIVNKKEEIVGVATFYNRKDGKPFDEMDET 418
PDE6A_Opossum        IMNASAEDFFNFQQEPLDESGWMIKNVLSMPIVNKKEEIVGVATFYNRKDGKPFDEMDET 418
PDE6A_Zebrafish      IMNAAQDEFFEFQKEPLDESGWMIKNVLSLPIVNKKEEIVGVATFYNRKDGKPFDEMDET 419
PDE6B_Human          IMNASADEMFKFQEGALDDSGWLIKNVLSMPIVNKKEEIVGVATFYNRKDGKPFDEQDEV 416
PDE6B_Maus           IMNASADEMFNFQEGPLDDSGWVIKNVLSMPIVNKKEEIVGVATFYNRKDGKPFDDQDEV 416
PDE6C_Human          MMNAPADEYFTFQKGPVDETGWVIKNVLSLPIVNKKEDIVGVATFYNRKDGKPFDEHDEY 420
                     :***. :: * **: .:*::**:******:*******:*****************: **

PDE6A_Human          LMESLTQFLGWSVLNPDTYESMNKLENRKDIFQDIVKYHVKCDNEEIQKILKTREVYGKE 478
Pde6A_Mouse          LMESLTQFLGWSVLNPDTYESMNKLENRKDIFQDIVKYHVKCDNEEIQKILKTREVYGKE 478
PDE6A_Dog            LMESLAQFLGWSVLNPDTYESMNRLENRKDIFQDMVKYHVKCDNEEIQKILKTREVYGKE 478
PDE6A_Opossum        LMESLTQFLGWSVLNPDTYESMNKLENRKDVFQDMVKYHVKCDNEEIQEILKTREVYGKE 478
PDE6A_Zebrafish      LMESLTQFLGWSVLNTDTYDKMNKLENRKDIFQDMVMYHVKCRKDEIQNILNTREMYDKE 479
PDE6B_Human          LMESLTQFLGWSVMNTDTYDKMNKLENRKDIAQDMVLYHVKCDRDEIQLILPTRARLGKE 476
PDE6B_Maus           LMESLTQFLGWSVLNTDTYDKMNKLENRKDIAQDMVLYHVRCDKDEIQEILPTRDRLGKE 476
PDE6C_Human          ITETLTQFLGWSLLNTDTYDKMNKLENRKDIAQEMLMNQTKATPEEIKSILKFQEKLNVD 480
                     : *:*:******::*.***:.**:******: *:::  :.:.  :**: **  :   . :
```



Supplementary Figure S5: Amino acid sequence alignment of retinal PDE proteins.
A – Variant Sites in the studied Mutants

A – Immunogen used to raise the anti-PDE6A antibody

A – Metal binding motif

V – Sequence deleted in the p.541_576del mutant (caused by exon 13 skipping)

13.1.4 Publication IV

Elena Buena-Atienza, Marius Cosmin Codrea, Stefan Czemmel, Britta Baumann, Susanne Kohl, Sven Poths, Sven Nahnsen, **Bernd Wissinger**. A comprehensive reference for exon 3 splicing defects of the human *OPN1LW* and *OPN1MW* opsin genes by means of a parallelized minigene splicing assay

Manuscript in preparation

A comprehensive reference for exon 3 splicing defects of the human *OPN1LW* and *OPN1MW* opsin genes by means of a parallelized minigene splicing assay

Authors:

Elena Buena-Atienza[1,#], Marius Cosmin Codrea[2], Stefan Czemmel[2], Britta Baumann[1], Susanne Kohl[1], Sven Poths[3], Sven Nahnsen[2], Bernd Wissinger[1]*

Affiliations:

[1] Institute for Ophthalmic Research, Centre for Ophthalmology, University of Tübingen, Tübingen, Germany

[2] Quantitative Biology Center (QBiC), University of Tübingen, Tübingen, Germany

[3] Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany

[#] Current address: Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany

*Corresponding author: wissinger@uni-tuebingen.de

Abstract

Background: Interpretation of the consequences of DNA variants observed in genetic testing remains challenging. The impact of exonic variants that are located distant from the canonical splice sites on transcript processing is often neglected, while the putative pathogenic effect of the amino acid substitution remains the main focus. Based on previous reports that certain combinations of common variants within exon 3 of the human X-linked *OPN1LW* and *OPN1MW* cone opsin genes result in aberrant splicing and cause disease, we systematically assessed the impact of exon 3 haplotypes on splicing applying a parallelized minigene approach.

Results: We determined the fraction of correctly spliced transcripts for a total of 232 *OPN1LW/MW* exon 3 haplotypes assayed as pools in parallel transfection experiments. More than half of the haplotypes resulted in a fraction of correctly spliced transcripts below 50%. Splicing is significantly impaired for haplotypes harbouring the c.532G or the c.538G alleles which caused an overall reduction of ~38% and ~26% in the exon 3 retention levels, respectively. An additive effect was observed for haplotypes harbouring both c.532G and c.538G. Review of 1,175 exon 3 haplotypes compiled from controls and patients with X-linked cone dysfunction and related disorders showed that haplotypes enriched in the patient group induce high levels of exon skipping *in vitro*.

Conclusions: Our systematic analysis showed that accumulation of exon 3 variants increased the fraction of mis-spliced transcripts. The c.532G and c.538G variants had the strongest allelic impact and together explain more than 50% of the haplotype-associated splicing disruption effect. Beyond this additive effect, our *in vitro* data underscores extensive and complex interactions among *cis*-regulatory elements involved in splicing of exon 3 of *OPN1LW/OPN1MW.*

Keywords: Splicing, minigene, NGS, long-read sequencing, VUS, MAVEs, *OPN1LW*, *OPN1MW*

Introduction

The advent of massive parallel sequencing or next-generation sequencing (NGS) technologies has marked a milestone for diagnostic and research of Mendelian and complex diseases. While high throughput sequencing enables to generate data in unprecedented scale and speed, the interpretation of biological and clinical significance of NGS-discovered variants is still a major bottleneck. Community-driven efforts have recognized the worth of standardized criteria for variant interpretation and classification (Richards et al., 2015; Karbassi et al., 2016; Nykamp et al., 2017). Yet, only a minimal fraction (~2%) of all missense variants reported in the Genome Aggregation Database has a clinical interpretation available in ClinVar (Starita et al., 2017). Even more, the so-called variants of uncertain significance (VUS) account for >50% of all missense variants with interpretation in ClinVar (Starita et al., 2017). VUS and variants without any interpretation represent a major burden for geneticists, clinicians and patients.

Functional evidence using well-established assays is considered a strong criterion to further assess VUS regarding potential pathogenicity (Woods et al., 2016) although it should be accompanied by further evidence (Sobrido, 2016); including for instance, clinical history, segregation analysis, frequency of the variant in the general population and evolutionary conservation. Valid bioassays are yet lacking for many disease-associated genes or gene products. If available, these are often laborious and usually designed to assess individual variants one-by-one. Typically such bioassays are not standardized but done by experts in a particular field of research and at a pace much slower and in lower numbers than the identification of novel variants in the very gene. Currently, assessment and prioritization of exonic missense variants mainly relies on filtering attending to frequency, conservation, biophysical amino acid properties and protein features by means of *in silico* tools such as SIFT (Kumar et al., 2009) and PolyPhen-2 (Adzhubei et al., 2010). On the other hand, it is well-known that exonic variants – regardless of being non-synonymous or synonymous (D'Souza et al., 1999; Monani et al., 1999; Mueller et al., 2015) – and haplotypes (Caffrey et al., 2008; Park et al., 2016) may interfere with pre-mRNA splicing, putatively resulting in mispliced transcripts that might turn into altered or absence of a given protein and thus cause disease. Bioinformatic algorithms are available to evaluate the putative effect of single nucleotide variants on splicing alteration (listed in Table 2; Anna and Monika, 2018). However, *in silico* tools provide only predictive results which in some instances are conflicting due to limited sensitivity and specificity.

Splicing entails recognition of the canonical 3' and 5' splice sites and the branch site by the spliceosome (Burge et al., 1999). The spliceosome requires additional guidance to find the *bona-fide* splice sites and differentiate these from decoy splice sites to establish the exon-intron definition (Cartegni et al., 2002). Providing an additional layer of regulation, *cis*-regulatory elements of splicing that are located in exons can be classified, attending to the regulatory function exerted, into Exonic Splicing Enhancers (ESEs) and Exonic Splicing Silencers (ESSs). ESEs and ESSs can modulate splicing by promoting or inhibiting the binding of *trans*-acting RNA-binding proteins with enhancing or silencing activity, respectively. This in turn controls the assembly and conformation of the core components of the splicing machinery (Cartegni et al., 2002; Staffa and Cochrane, 1995). The current understanding of splicing regulation is still very limited. The discrete regulatory splicing landscape modeled by ESE or ESS hexamers lacks integration of sequence context-dependent effects and interactions between *cis*-elements and thus, may turn out overly simplified (Haque et al., 2010; Julien et al., 2016; Tammaro et al., 2014; Zhu et al., 2001). The most reliable assessment of putative splicing

mutations is direct cDNA transcript analysis with RNA derived from native tissue. However, such material is frequently unavailable or inaccessible; for instance, if the gene-of-interest is expressed only in certain organs or tissues. Alternatively, *in vitro* minigene splicing assays have been proven reliable for functional characterization of variants (Steffensen et al., 2014; Weisschuh et al., 2012).

The high number of VUS underscores the shortcomings of current variant assessment strategies and *post hoc* one-by-one experimental validation of variants. Multiplexed assays of variant effect (MAVEs) provide novel means to test and assess a multitude of variants simultaneously, either known variants or the complete landscape of potential single-base variants in a gene or exon. The latter may result in a comprehensive 'fitness' landscape which can serve as a reference of functionality once a novel variant is identified. Using NGS as qualitative and/or quantitative proxy, MAVEs can accommodate a large number of variants to be assayed at different molecular levels: gene expression regulation (Melnikov et al., 2012; Patwardhan et al., 2012), protein function such as enzymatic activity and interactions through phage display and yeast-based assays, respectively (Starita and Fields, 2015; Starita et al., 2015), and impact on splicing by means of minigene reporter assays (Julien et al., 2016; Rosenberg et al., 2015).

In this study we developed and implemented a parallelized minigene splicing assay to experimentally analyze the effect of common variants in exon 3 of the human *OPN1LW* and *OPN1MW* genes and combinations of such variants on splicing. *OPN1LW* and *OPN1MW* are present as a tandem repeat gene cluster on the X chromosome and encode for the apoprotein of the long-wavelength sensitive (LWS) and middle-wavelength-sensitive (MWS) cone photopigments, respectively (Nathans et al., 1989). These genes share >98% nucleotide sequence identity including intronic and intergenic regions and are subject to frequent unequal homologous recombination and gene conversion events which potentially result in novel combinations of variants (haplotypes) of uncertain relevance. Notably, Ueyama and colleagues showed that certain rare haplotypes in exon 3 induces exon skipping and thereby, cause colour vision defects (Ueyama et al., 2012). We and others have confirmed and extended this finding to severe forms of X-linked cone dysfunction (Buena-Atienza et al., 2016; Gardner et al., 2014; Greenwald et al., 2017. In our previous study, we individually assessed 12 different rare exon 3 haplotypes – defined by eight common single nucleotide polymorphisms (SNPs) – by means of an *in vitro* minigene splicing assay. We observed a rather continuous range of exon 3 skipping levels for the studied haplotypes most likely reflecting a complex regulatory network of *cis*-acting elements to maintain exon definition. In contrast to the previous one-by-one minigene testing approach, we present in this report a strategy that leverages NGS for parallel functional assessment of splicing efficiencies. We sought to develop an assay to enable unbiased *a priori* functional testing of all possible 256 exon 3 haplotypes of *OPN1LW/OPN1MW* deduced from the eight bi-allelic variant sites commonly observed in the general population.

This study focused on three aims: i) to develop a new MAVE assay applied to *OPN1LW/OPN1MW* exon 3 haplotypes which could serve as a model for other gene systems requiring connectivity over large sequence distances between *cis*-elements, ii) to predict the putative pathogenic effect of novel exonic variants and/or haplotypes on splicing efficiency in the context of X-linked colour vision deficiencies and severe cone dysfunction disorders and, iii) to gain understanding on the splicing regulatory landscape of the *OPN1LW/OPN1MW* genes.

Results

Generation of a traceable *OPN1LW/OPN1MW* minigene reporter library

We aimed at generating a minigene library to comprise all possible combinations (haplotypes) of the eight common single-nucleotide variants in exon 3 of *OPN1LW* and *OPN1MW*: c.453A>G, c.457A>C, c.465C>G, c.511G>A, c.513G>T, c. 521C>T, c.532A>G, and c.538T>G; Winderickx et al., 1993). The haplotype c.(453A; 457A; 465C; 511G; 513G; 521C; 532A; 538T), which does not induce detectable levels of aberrantly spliced transcripts in a uniplex minigene splicing assay in HEK293 cells (Buena-Atienza et al., 2016) and the haplotype c.(453G; 457C; 465G; 511A; 513T; 521T; 532G; 538G), which

induces fully aberrant splicing served as control 1 (“Reference”) and control 2, respectively. Note that in many prior studies exon 3 haplotypes have been denoted according to the corresponding combination of encoded amino acid residues. According to this scheme, controls 1 and 2 correspond to ‘MVAIS’ and ‘LIAVA’, respectively (see Supplementary Material; Supplementary. Fig. S1). For the generation of the minigene library, we performed targeted mutagenesis of *OPN1LW* exon 3 using the control 1 minigene as template and added an 8-bp barcode tag (Fig. 1A) to generate a large series of minigene plasmids carrying different haplotypes tagged by unique barcodes (Input Library; Fig.1B). We adopted a two-step site-directed mutagenesis strategy with overlapping degenerate Ultramers and their assembly by overlap extension PCR to account for the size of exon 3 (169 bp) and the 85 bp distance between the most proximal and most distal variant site (Supp. Figure S2). The resulting exon 3 mutagenesis PCR product included homology arms that allowed retro-fitting into the linearized Barcode Library by In-Fusion Cloning (Supp. Fig. S2). This ensured that the targeted exon 3 haplotypes were tagged with unique 8-bp barcodes. A total of 90 single clones were Sanger sequenced to validate and quantify the presence, completeness and diversity of barcodes and exon 3 haplotypes during the library construction. To favour tagging of haplotypes by unique barcodes we generated a highly complex Barcode Library in far excess to the number of clones in the final Input library; see Materials and Methods.

The parallelized minigene splicing assay was carried out by *en masse* transfection (Fig. 1C) of HEK293T cells with the Input Library followed by the qualitative and quantitative analysis of minigene-derived transcripts in the Output Library. The most commonly observed splicing defect induced by (disease-associated) exon 3 haplotypes is skipping of the entire exon 3, and thus the loss of the sequences which induce this defect in the transcripts. Therefore, the barcode which is maintained in the transcripts serves as *key* or *tag* to trace back the original exon 3 haplotype. Readout (and data analysis) of the parallelized minigene assay thus is based on i) qualitative NGS sequencing of the mutagenized barcoded-minigene plasmid pool (Input Library; Fig. 1D), and ii) qualitative and quantitative NGS sequencing of the pool of minigene-derived transcripts isolated from the transfected HEK293T cells (Output Library; Fig. 1E). Long-read PCR-free sequencing of the Input Library (Fig. 1D) was used to phase exon 3 haplotype with its unique cognate barcode tag for each single plasmid molecule, accounting for the large distance between exon 3 and the barcode site in our 'native-like' minigene construct (which harbours the complete sequence of the exon 3-flanking introns and framed by cDNA fragments of exons 1-2 and exons 4-6 of *OPN1LW*; Ueyama et al. 2011). Output Library sequencing was performed by long-read and short-read paired-end sequencing of RT-PCR amplicons (Fig. 1E). In this context long reads were essential to capture correct phasing of haplotype-barcode and to generate a *de novo* reference assembly serving as a scaffold to enable mapping of the more abundant short reads using available Bioinformatics tools (Fig. 1F). Two main measures were obtained from the Output Library datasets: the first querying whether a particular read bears exon 3 sequences (exon 2-to-3 junction = exon retention) or exon 3 is skipped (by for instance the observation of exon 2-to-4 junction), and, secondly, to determine the relative ratio of the two differentially spliced transcripts (retained/skipped) expressed as percentage of reads with retained exon 3 (Supp. Fig. S3). While long reads faithfully recapitulate the completeness of the sequence encompassing the relevant exon-exon boundaries from differentially spliced transcripts (isoform structure) including the unique identifier, short-read sequencing achieved a considerable higher read count required for reliable relative quantification levels, as previously described (Fig. 2; Park et al., 2018).

Performance and reproducibility of the parallelized minigene splicing assay

We observed 248 haplotypes uniquely tagged in the Input Library. Targeted haplotypes represented 89.7% of circular consensus sequencing (CCS; see also Materials and Methods) read counts among uniquely tagged haplotypes (Supp. Fig. S4A). The latter represented 27.8% of the Input Library; corresponding to 8,003 out of a total 28,968 CCS reads obtained from sequencing of the Input Library. Unique barcodes tagging targeted haplotypes represented 77% of the total (n = 1952) unique barcodes sequenced (Supp. Fig. S4B). On average, each unique targeted haplotype was tagged by six different

unique barcodes and observed in 29 CCS reads on average. The control 1 haplotype with no variants identified, which was used as template for the *in vitro* mutagenesis was observed in 0.6% (n = 44) of total CCS reads (Supp. Fig. S4C) and tagged by <2% (n = 31) of the total unique barcodes (Supp. Fig. S4D).

Long-read sequencing of the Output Library yielded a total of 10,384 CCS reads containing unique barcodes. Targeted haplotypes in the Output Library were tagged by an average of ~5.5 different unique barcodes and ~27% of the CCS reads were retained upon filtering by quality barcode uniqueness. These figures are comparable to those of the Input Library sequencing indicating the absence of major constraints or bias during propagation (between Input and Output library). This also reflects that the diversity and uniqueness of barcodes and haplotypes is inherent to the constructed library (Input Library).

The exon 3 retention fraction determined for each unique-barcoded haplotype in the Output Library (Supp. Table S3) with short-read sequencing (mapping assisted by long-reads) highly correlated across replicates (Fig. 2A; three replicates shown) with correlation coefficients ranging between 0.89 and 0.97 for all pairwise comparisons of the four replicates (p-value < 2.2e-16; Supp. Fig. S5A). When compared with results from individual splicing assays for 20 different haplotypes harboured in barcoded constructs which were available from the quality control of the Input Library construction (see Materials and Methods), the outcome from the parallelized minigene splicing assay was largely concordant (Fig. 2B). A high positive correlation was found between long-read sequence data and individually quantified assays for the corresponding haplotype (r = 0.87, n = 20, p-value = 8.02e-07; Supp. Fig. S5B; Supp. Table S4 and Supp. Table S5). A high correlation was maintained between the parallelized and individually assayed datasets when both long- and short-read data were combined for hybrid mapping (r = 0.83, n = 20, p-value = 5.3e-06; Supp. Fig. S5B; Supp. Table S4 and Supp. Table S5). As expected, the mean exon 3 retention fraction calculated by the combined analysis *versus* the average distance scores alone correlated positively for the 20 haplotypes (r = 0.95, n = 20, p-value = 3.06e-10; Supp. Fig. S5B; Supp. Table S4 and Supp. Table S5) and for all haplotypes tested in the parallelized assay (r = 0.86, n = 232, p-value < 2.2e-16; Supp. Fig. S5 and Supp. Table S5).

Based on the results of 20 haplotypes assayed individually and quantified by fluorescence-labelled PCR fragment analysis, three categories of predominant exon skipping (“S”), intermediate levels (“I”) and minor (“M”) exon skipping were defined (Fig. 2B; Supp. Table S4) in agreement with our previous study (Buena-Atienza et al., 2016). When compared with the data from the parallelized experiment a sensitivity of 83.3% and a specificity of 57% were calculated for the parallelized splicing assay taking into account true positive, false positive, true negative and false negative values (Supp. Table S4). The false negative value stems from one haplotype sequenced by a low number of reads (Supp. Table S4) and the false positive values come from haplotypes with an exacerbated number of reads in comparison to other identifiers which could reflect biases towards smaller amplicons (=RT-PCR with skipping of exon 3) being preferably sequenced on the Illumina flow cell and thus, lower ratios for the mean % of exon retention.

Impact of individual exonic variants and influence of the haplotype context on *OPN1LW/OPN1MW* exon 3 retention

We compared the percentage of correctly spliced *OPN1LW/OPN1MW* exon 3 transcripts of all haplotypes carrying each of the eight variants with those harbouring the reference allele at a given position (regardless of the additional variants entailed in the haplotype). The four variant sites that significantly reduced the fraction of correctly spliced transcripts across all haplotypes were c.457C, c.511A, c.532G and c.538G (Fig. 3A). Notably, the c.532G variant induces the overall highest (~38%; Supp. Table S6) reduction of correctly spliced transcripts in comparison to haplotypes harbouring the c.532A reference (Wilcoxon test, p = <2e-16; Fig. 3A), followed by the overall ~27% reduction exerted by the variant c.538G (Wilcoxon test, p = 6.7e-10; Fig. 3A; Supp. Table S6). We further asked whether these two variants with the strongest individual effect may interact with each other and evaluated the

effect strength of the four possible variant combinations at these two sites. Haplotypes were sorted in four groups according to the variant combinations at c.532 and c.538; i.e. GG, GT, AG, and AT (Fig. 3B). In fact, we found that haplotypes harbouring c.532G and c.538G simultaneously resulted on average in the lowest level of correctly spliced transcripts whereas haplotypes harbouring c.532A and c.538T resulted in the highest level of correctly spliced transcripts. Heterotypic variant combinations at these sites (i.e. 'GT' or 'AG') performed intermediate. The differences between the two homotypic variant combinations and between homotypic and heterotypic variant combinations were highly significant (Wilcoxon test, p = 1.0e-11; Fig. 3B).

Since the c.511G>A variant showed the third highest difference for exon 3 retention levels (Fig. 3A), we performed pairwise comparisons between the sub-groups of haplotype comprising all different variant combinations at c.511, c.532 and c.538 variants applying the Wilcoxon test and p-values adjusted with Bonferroni correction (Fig. 4). Twenty-one of the 28 pairwise comparisons reached significance levels of p-value <0.01 (Fig. 4). Average levels of 'exon 3 retention' of sub-groups of haplotypes increased rather linearly ranging from the “AGG” (c.511A, c.532G and c.538G) subgroup with <5% of correctly exon 3-spliced transcripts up to ~80% for the “GAT” (c.511G, c.532A, c.538T) subgroup of haplotypes (Fig. 4). Notably, these two extremes comprise opposite variants at all three positions. Moreover, the order of effect strengths on exon 3 retention, c.532A>G, c.538T>G and c.511G>A (from strongest to least) is maintained in the clustered analysis of triple variants (Fig. 4). Haplotypes carrying the c.511A in phase with c.532G and c.538G result in an overall significant reduction of exon 3 retention compared to those with c.511G (AGG *vs* GGG).

Next, we pursued to analyze for each individual site the effect driven by the accumulation of additional variants. We divided the haplotypes based on the alternate alleles at a given site (e.g. c.453A or c.453G) and counted the number of additional variants with respect to the Reference haplotype (Fig. 5A-H). For all sites, we found that an increased number of variants correlated with a decrease in exon 3 retention. For c.532 as well as c.538 the presence of the variant allele significantly outweighed the cumulative effect of other additional variants (in conjunction with the reference allele) for most categories of numerically accumulated variants (Fig. 5G-H). A significant increase in exon skipping is observed for all haplotypes entailing two, three, four and five variants including the c.532G variant (Fig. 5G). Similarly, haplotypes with the c.538G variant within a total of three, four and five accumulated variants showed a significant increase in mispliced transcripts (Fig. 5H). An opposite counteracting effect was observed for the variant allele at c.521, (and to a lesser extent) at c.513 and c.465, but only at one (e.g. c.465 in the category with two variants; Fig. 5C) or two numerical categories (e.g. c.513 in the categories with four and five variants; Fig. 5E, or c.521 with three and four variants; Fig. 5F). Overall, the four variants located towards the 3’ end of *OPN1LW/MW* exon 3 (Fig. 5E-H) have a greater effect on splicing regulation than those towards the 5’ end (Fig. 5A-D).

Combinatorial analysis and interactions between double and triple variants

The results presented in Fig. 3-5 suggest effects beyond that of single variants, indicating possible interactions between variants in the modulation of splicing. To further explore potential interactions between these variants, we tested whether the level of exon skipping is best explained by an additive model. A total of 28 combinatorial haplotypes with two variants sites different from the “Reference haplotype” are expected from our combinatorial design and were present in the Input Library (see Fig. S4C-D). First, we determined the ratios between the experimentally observed level of exon skipping for all 26 assessed haplotypes with two variant sites available (“double variants”) and the sum of values from the corresponding individual variants (expected values under a purely additive model). We calculated epistasis scores as the residuals of observations from expected values over the four replicates and performed a paired t-test; 95% confidence interval (Supp. Table S7). The observed exon skipping levels of only six of the 26 “double variants” could be explained by an additive model (Fig. 6; bottom arcs), while the others show positive and negative epistasis to variable extent (Fig. 6; upper arcs). Notably, in four out of the six “double variants” fitting an additive model, c.457A>C was one of the

two variants. Half of the “double variants” revealed a strong epistatic interaction which accounted for a >20% increase or decrease in exon skipping levels compared with the expected values from the corresponding single mutants. The analysis also showed only a low positive epistasis for c.532A>G and c.538T>G (p-value = 0.04072): the two variants with the strongest individual impact on exon 3 splicing. The variant c.453A>G is involved in positive and negative epistatic interactions. The latter deviate stronger from the expected values under the additive model, and are observed with downstream variants c.511G>A, c.532A>G and c.538T>G (p-values: 0.001, 0.001 and 0.001, respectively). Similarly, positive and negative epistatic interactions were noted for the variant c.521C>T; positive epistasis with c.453A>G, c.511G>A, c.513G>T and c.532A>G (p-values: 0.007, 0.003, 0.011 and 0.003, respectively) and negative epistasis with c.538T>G (p-value = 0.024) (Fig. 6, Supp. Table S7). We further explored additive or epistatic interactions for the 49 experimentally assessed haplotypes with triple variants. Exon 3 retention values for 14 of these do not significantly differ from the sum of the values for the corresponding individual variants (Supp. Table S8). The remaining 35 haplotypes with triple variants showed evidence for positive epistasis (Supp. Table S8).

Finally, all tested haplotypes can be visualized by applying hierarchical clustering (Fig. 7A). A pattern is observed in the heatmap where c.532G and c.538G cluster in a ‘red’ block phased with other variants (Fig.7A). Haplotypes lacking the latter two alleles (‘dark blue’ for c.532G and c.538G) are found in clusters that do not induce high levels of exon skipping (blue or yellow; Fig. 7A). In a second alignment-based approach according to the sequence identity among haplotypes, tree-based clustering was performed and three major clusters with high exon skipping rates (red) were identified (Fig. 7B).

Reference tables for haplotype classification based on splicing efficiency

We searched for the extremes of the distribution for either exon skipping or exon retention. A selection of 30 haplotypes from the whole dataset (Supp. Table S3) showing the highest and lowest rates of exon retention are summarized in Table 1. The c.511A, c.532G and c.538G alleles are enriched in haplotypes with highest scores for exon 3 skipping and c.532G is completely absent in haplotypes with lowest exon 3 misplicing rates. Haplotypes ranked in the top 15 for exon skipping seemed to be more abundant and to be tagged by more barcodes which could indicate biases derived from the Input Library construction. However, this deviation likely is derived from two overrepresented haplotypes with four exonic variants with respect to the reference haplotype; as shown by the exacerbated number of reads or high frequency in the Input Library (Supp. Fig. S4D) and thus high frequency (Table 1), that turned out to result in high levels of exon 3 skipping in the Output Library.

Diversity of exon 3 haplotypes in human subjects: enrichment of deleterious exon 3 haplotypes in cone photoreceptor disease

We determined *OPN1LW/MW* exon 3 haplotypes from 163 male subjects diagnosed with BCM, cone dystrophy, cone-rod dystrophy and one female BCM carrier. After filtering for unambiguous haplotypes and removal of haplotypes linked to other point mutations in *OPN1LW* or *OPN1MW* we obtained 198 haplotypes (n = 151 from the proximal gene copy and 47 from distal gene copies) which were further sub-classified into a 'disease group' (i.e. potentially pathogenic haplotypes; n = 168) and a 'control group' (n = 30). We further compiled *OPN1LW/MW* exon 3 haplotypes from the available literature including population studies as well as studies on patients with BCM, XLCD/BED, and congenital colour vision defects (Buena-Atienza et al., 2016; Carroll et al., 2012; Carroll et al., 2010; Crognale et al., 2004; Deeb et al., 1995; Gardner et al., 2010; Gardner et al., 2014; Greenwald et al., 2017; Hayashi et al., 2001; Jordan et al., 2010; Katagiri et al., 2018; Li et al., 2015; McClements et al., 2013; Mizrahi-Meissonnier et al., 2010; Mountford et al., 2019; Neitz et al., 1999; Orosz et al., 2017; Patterson et al., 2016; Ueyama et al., 2004). By that we obtained n = 1664 exon 3 haplotypes, 684 of which were incomplete, i.e. not all common exon 3 variants were reported. We further excluded three haplotypes linked to other point mutations in the *OPN1LW/MW* genes. The final study sample included 1175 fully resolved exon 3 haplotypes including 812 'control group' haplotypes and 243 haplotypes in the 'disease group'. In total, 33 different exon 3 haplotypes were observed, 25 of which were rare, i.e. < 1% in the

'control group' (Table 2). Seven of these haplotypes were exclusively observed in the 'disease group' including four haplotypes observed at least three times (in the 'disease group'). Three further haplotypes were strongly enriched, i.e. ≥8 fold increased frequency, in the 'disease group'. In fact, the profile of exon 3 haplotypes was evidently different between the 'disease group' and the 'control group' with eight of the haplotypes being significantly different in frequency ($p<0.05$ adjusted for multiple testing, Table 2). The lowest p-values for haplotypes enriched in the 'disease group' were obtained for 'LIAVA' ('G-C-G-A-T-C-G-G', p=2.47E-20) and 'LVAVA' ('G-C-G-G-G-C-G-G', p= 1.40E-15), two well-documented pathogenic haplotypes. The difference in haplotype frequencies was still significant when we compared proximal (mostly *OPN1LW*) and distal (mostly *OPN1MW* gene copies) separately (data not shown). For twenty-eight haplotypes observed in human subjects we obtained reliable data with the parallelized minigene splicing assay. We plotted the cumulated percentage of observed haplotypes in the 'disease group' and the 'control group' with respect to the splicing performance of such haplotypes (inverse Kaplan-Meier plot, Fig. 8). The graph indicates an enrichment of haplotypes with low splicing performance in the 'disease group'. Enriched haplotypes in the 'disease group' such as 'LIAVA', 'LVAVA' and 'MIAVA' (Table 2) are also found within the lowest exon inclusion scores in Table 1 (id_85, id_70 and id_235, respectively). Likewise, haplotypes enriched in the 'control group' such as 'MVAIS' (Control 1) and 'LVAIS' (id_68; Table 1) show high levels of exon retention.

Discussion

We have developed a cost-effective method to parallelize minigene splicing assays for hundreds of variant combinations through the use of barcoded-minigene constructs which can be simultaneously assayed in a single experiment (Fig. 1). This strategy enabled parallel functional characterization of *OPN1LW/MW* exon 3 haplotypes at the transcript level. For this purpose, we generated a targeted mutagenized minigene library which was subjected to assay the effect of almost all possible combinations of eight common natural variants on splicing efficiency. In this pilot, experimental reproducibility of measured exon 3 retention across replicates performed with high correlation (Fig. 2A; Supp. Fig.S4A). In comparison with 20 selected haplotypes assayed in individual splicing assays, the results from the parallelized assay tended to yield a higher level of exon skipping, in particular for haplotypes predominantly inducing exon skipping (Fig. 2B, Supp. Fig. S5B). The data from the individual assays and the parallelized assay do show a non-linear correlation. While the individual assay and the parallelized assay differ in multiple technical steps and may each have its limitations, we reason that some elements of the parallelized assay such as preference of shorter molecules to be sequenced on NGS platforms and the hybrid read mapping strategy may have introduced additional bias towards an increased proportion of exon skipping in the parallelized approach. The sensitivity and specificity of the parallelized minigene assay was calculated taking the individual minigene assay as the gold-standard. Nonetheless, it remains challenging to evaluate reliability of each method due to the lack of a reference from splicing in the native tissue and cell type: the human red (long wavelength sensitive) and green (middle wavelength sensitive) cone photoreceptors.

While the output of this study needs further supporting evidence for interpretation, our results most likely reflect useful predictors, especially for the extremes of the distribution. We would like to emphasize certain improvements and unique technical features of our parallelized minigene splicing assay compared with other studies. Firstly, splicing of *OPN1LW/MW* exon 3 was assessed in its native sequence context including the entire exon 3, full-length non-modified flanking introns 2 and 3, and upstream and downstream exon 1-2 and 4-6 sequences. Secondly, we applied PacBio long-read PCR-free sequencing of the Input Library to unambiguously link exon 3 haplotypes with unique barcodes thereby avoiding PCR errors (Laver et al., 2016) and falsely attributed barcodes which may occur in other cloning workflows (e.g. subassembly cloning; Kitzman et al., 2015) or short-read mate-pair NGS sequencing. Thirdly, we used a hybrid NGS sequencing (PacBio single molecule real time sequencing and paired-end short-read sequencing) strategy for the RT-PCR amplicons of the minigene transcript pool of the Output Libraries. Long-read sequences ensured unequivocal correct assignment of splice junctions retaining or skipping exon 3 for each barcode-tagged molecule, and were used to generate a

de novo reference sequence scaffold for the exon 3-retained/skipped haplotype isoforms. Additional short-read paired-end sequencing-by-synthesis of amplicons covering the splice site of interest on one end and the barcode on the other was done to obtain deeper sequencing for relative quantification of differentially spliced transcripts. Short-read mapping required non-standard methods for read alignment and also taking advantage of the *de novo* reference scaffold to increase the fraction of mapped reads and to allow the search for haplotypes. A combination of the advantages from both long- and short-read sequencing technologies was pursued, similarly to hybrid approaches (Au et al., 2012, 2013; Deshpande et al., 2013; Choudhury et al., 2018).

The obtained results emphasize a complex relationship between common exonic variants and exon 3 splicing of *OPN1LW* and *OPN1MW* genes. Nonetheless, we were able to rank variants according to their impact on splicing with the strongest negative effect exerted by the variants closest to the downstream splice donor site (Fig. 3A). The single variant effect analysis (Fig. 3) clearly demonstrated that the c.532A>G has a strong negative impact on exon 3 splicing. Moreover, all TOP15 haplotypes with the highest levels of exon retention had c.532A and all but one of the TOP15 haplotypes with the highest levels of exon skipping had c.532G (Table 1). These results obtained from our comprehensive approach corroborate previous observations from a set of individual minigene assays with supporting genotype-phenotype associations (Ueyama et al., 2012; Gardner et al., 2014; Buena-Atienza et al., 2016). We observed a similarly strong negative effect of the c.538T>G variant on exon 3 retention (Fig. 3A), and co-occurence of variants c.532G and c.538G on the same haplotype exacerbated the effect in a close to additive manner (Fig. 3B, Supp. Table S7). Despite this negative impact on exon 3 retention, both variants are rather common in the human population, in particular in *OPN1MW* (frequencies: 0.138 and 0.916 for c.532G and c.538G, respectively, in our control dataset, (Supp. Table S9). The c.538G is a coding variant which encodes an alanine residue at amino acid position 180. Such a photopigment has a spectral shift of about 4 nm to lower wavelength compared to a photopigment with serine at this very position and thus, results in an increase of the spectral difference between LWS and MWS photopigments (Merbs and Nathans, 1992). It remains to be determined whether this is an ecological advantage that may outweigh the 'lower fitness' of *OPN1MW* transcripts bearing the c.538G variants in terms of correct splicing. We further explored a potential functional interaction of the c.532G>A and c.538G>T variants with c.511G>A, the third most effective single variant. In this triple variant analysis we noted a significant negative effect of the c.511A variant on exon retention for haplotypes with both c.532G and c.538G alleles (Fig. 4). These findings further corroborate the cumulative effect of exon 3 variants on the relative quantity of exon skipping.

Instead, we investigated to which further extent the splicing defect can be attributed to the accumulation of additional variants deviant from the reference haplotype. We observed a general tendency towards exon skipping as more variants accumulate (Fig. 5A-H). A pathogenic threshold is seemingly surpassed when i) a certain number of variants are accumulated and ii) these occur at specific nucleotide positions affected by these variants, suggesting redundancy in exonic splicing motifs and thus robustness of exon 3 splicing to individual variants. Among all variants the c.532G had the strongest impact of all single changes (number of variants, n = 1) to the reference haplotype accounting for an almost 50% reduction in transcripts retaining exon 3. The haplotype that accumulated all other variants with the exception of the c.532A>G variant, also accounts for about 50% of exon 3-skipped transcripts (Fig. 5G). The inherent difference in the number of haplotypes per category (with a defined number of variants compared to the reference) in this test decreases statistical power for the categories with low or large number of accumulated variants. For instance, there are only 8 different haplotypes in the ‘7 variants’ category (i.e. a total of seven variants compared to the reference), whereas the number of possible haplotypes for the ‘3 variants’ and the ‘4 variants’ category is much higher with n = 56 and n = 70 haplotypes, respectively. This bias in statistical power likely explains that there was no significant difference in exon 3 retention between the c.532G and c.532A bearing haplotypes in the ‘7 variants‘ category but in the ‘4 variant’ category of haplotypes, the difference in exon retention between haplotypes carrying either the c.532A or the c.532G (n = 35 haplotypes each) is highly significant (Fig.

5G). In comparison, no significant differences in exon 3 retention were observed in the ‘4 variant’ category if stratified for c.453, c.457, c.465 or c.511. Although approximately half of the assayed haplotypes result in exon retention levels of <50%, our data also indicate that recognition of exon 3 of *OPN1LW/MW* during splicing is still highly tolerant given that accumulation of several variants seems to be needed in order to achieve levels that strongly alter splicing (e.g. ≤10% exon retention). As a general rule, at least four variants – one being either the c.532G or the c.538G – are agglomerated to exert a negative effect by reducing the splicing efficiency in about or more than 50%. Does the discrete creation or abolishment of ESS or ESE suffice to recruit, block and/or modulate splicing factors? A less simplistic model envisages synergy among *cis*-regulatory elements to regulate the probability of a splicing event to occur or shift probabilities between alternative splicing outcomes. Bioinformatic tools such as ESEfinder (Cartegni et al., 2003) or SROOGLE (Schwartz et al., 2009) predict several regulatory elements with distinct functions as overlapping hexamers across exons and introns. Hexamer-based algorithms such as HExplorer (Erkelenz et al., 2014) do not fully recapitulate evidence obtained by individual and parallelized *OPN1LW/MW* minigene assays (Supp. Fig. S6), the latter corroborating previous reports (Soukarieh et al., 2016).

Rosenberg and colleagues found that the additive model explained >90% of the combinatorial effect sizes of splice site selection induced by non-overlapping 4-mers at either the 5’ or 3’ end of the exon (Rosenberg et al., 2015) and could explain >60% of the effects driven by known variants causing exon skipping in *SMN1/2*, *CFTR*, and *BRCA2*. Conversely, Julien and colleagues found that ~50% of the total effect size induced by the double mutants could be explained by a linear additive model without interactions, while extensive epistasis between variants was identified across the whole exon (Julien et al., 2016). We found combinatorial haplotypes with double and triple variants to be highly context-dependent, albeit 50% and 30% of the effects observed by double and triple variants, respectively, could be explained by an additive model (Fig. 6; Supp .Table S7; Supp. Table S8). Double variants that showed additive behaviour, such as for instance the haplotype with c.457A>C and c.532A>G, showed no epistasis in triple combinations harbouring in addition c.511G>A, c.513G>T or c.538T>G variants (Supp. Table S7; Supp. Table S8). Possibly the additive effects for certain pairs of variants preclude epistasis with a third variant in some circumstances. This might be dependent on the recruitment and interactions among *trans*-acting splicing factors and/or secondary RNA structures.

Through clustering of haplotypes based on their splicing competence exerted and by sorting exonic variants within the haplotype according to the exon skipping levels, we observed a pattern of high susceptibility for aberrant splicing including the c.532G or c.538G variants which in many instances are linked with c.511A (Fig. 7A), as for instance in the highly deleterious ‘LIAVA’ haplotype. These findings further support the single variant assessments (Fig. 3). A secondary cluster of haplotypes inducing high levels of exon skipping comprises the c.511G and c.513G variants as in the ‘LVAVA’ haplotype (Fig. 7B). In the tree-based clustering however, the haplotypes 'LIAVA' and 'LVAVA' – enriched in the 'disease group' (Table 2) – cluster together for sequence similarity and high levels of aberrant splicing with other additional seven related haplotypes. The two variant combinations encoding the 'MIAVA' haplotype (with either c.465C or c.465G) were found in a second major hotspot for high exon skipping clustering with four related haplotypes (Fig. 7B). A third cluster is highlighted with three haplotypes derivative of the 'LIAVA' haplotype which cause high levels of aberrant splicing but have not yet been observed in the population. Besides, it is apparent that the rules underlying the regulation of *OPN1LW/MW* exon 3 splicing are not simple but rather based on the dual and joint interactions within variants (Fig. 4-6).

Purifying selection may act against haplotypes that have a stronger pathogenic impact (Castel et al., 2018). Notably, the three variants with highest overall effect on splicing competence are non-synonymous variants. Only the c.538G>A/p.Ala180Ser variant has a significant effect on spectral tuning that may explain evolutionary constraints. However, there may be other constraints such as protein

folding, stability, chromophore accessibility, half-life of excitation states among others which may shape and balance the landscape of coding variants in the X-linked cone pigment genes.

Other parallelized splicing assays have tested up to several thousands of variants (Julien et al., 2016; Soemedi et al., 2017; Adamson et al., 2018; Cheung et al., 2019). Yet, the targeted exons in these studies were ≤ 100 bp long, shorter than the average human exon length of 134 bp (Savisaar and Hurst, 2017) and the 169-bp exon 3 of *OPN1LW/MW* targeted in our study. The assay strategy applied herein enable testing of haplotypes scattered through multiple large exons, including complete flanking endogenous introns and may well be adapted for a larger spectrum of variants through the implementation of an increase in barcode library complexity and higher sequencing depth of the Output Library.

In our study we focused on the assessment of all haplotypes derived from the eight common exon 3 variants observed in the human population. In our survey of exon 3 haplotypes in control (n = 812) and disease (n = 243) individuals, we observed 33 different haplotypes, 16 of which were rare (<1%) in the entire sample (Table 2). The frequencies of about a quarter of all haplotypes were significantly different between the disease and control group with five haplotypes being significantly more prevalent in the disease group. For ~85% of all haplotypes observed in human subjects we could retrieve the corresponding splicing competence *tag* from our dataset generated with the parallelized minigene splicing assay, which revealed high exon 3 skipping levels among those haplotypes enriched in the disease group (Fig. 8). The main difference in the population frequency *versus* exon 3 retention level plot (Fig. 8) is due to exclusivity or elevated frequencies of haplotypes with retention levels below 10% in the disease group, likely demarking the threshold of severe phenotypic consequence. A more complete distinction between ‘disease-causing’ and ‘control’ haplotypes is limited by the sampling (e.g. rarely clear evidence that the disease in the ‘disease group’ is caused by mutations in the *OPN1LW/MW* gene cluster and diversity of included clinical phenotypes) as well as the multiplicity of *OPN1LW/MW* gene copies, some of which are phenotypically irrelevant. While our study provides a comprehensive picture on the splicing competence of naturally occurring exon 3 haplotypes, it lacks single-nucleotide resolution to define potential cooperative or epistatic interactions between *cis*-acting elements in this exon. Future studies may therefore target exon 3 in its entirety or should look into whether binding sites for splicing factors are created or abolished and/or different mechanisms are involved, such as alteration of RNA secondary structures. We simplified the splicing assay by assessing exon 3 variants in the context of a 'partially spliced' *OPN1LW* construct (Ueyama et al. 2012). Future studies may contemplate the effect of long-distance interactions in the full endogenous genomic context maintaining the structural arrangement of the opsin genes (e.g. inserting the library by CRISPR/Cas9).

In summary, we have scrutinized a confined gene region by a middle-throughput functional pilot assay that enabled testing the individual effect of >200 *OPN1LW/MW* haplotypes on exon 3 splicing, the assessment of the overall influence of each variant within the haplotype and searching for patterns in the splicing regulatory landscape by addressing an unprecedented combinatorial approach. Although further validation including deeper high-throughput sequencing and standardized pipelines as well as further research is required, the results derived from haplotypes hitherto not characterized may assist interpretation of uncertain genotypes observed in XLCD patients. This method can be adapted and implemented for putative intronic and exonic splice-disrupting variants of any disease-causing gene. The development of parallel functional splicing assays for unbiased classification of variants will enable prospective functional testing in the future.

Materials and Methods

Barcode Pool

A single-stranded 91-mer oligonucleotide Ultra_BC.2: 5' AAG GAC GAC GAT GAC AAG TAG NNN NNN NNC CCT TTA GTG AGG GTT AAT CTA TAG TGA GTC GAC TTA GAG CTA CTG GGT GGC ATC CCT GTG A 3', including an 8-mer random barcode tag and a *SalI* restriction site, and a 21 bp oligonucleotide Seed_BC.2: 5' TCA CAG GGA TGC CAC CCA GTA 3' were custom synthesized by IDT (Leuven, Belgium). Seed_BC.2 and Ultra_BC.2 oligos were hybridized to extend the single-stranded 91-mer into double-stranded molecules with CloneAmp HiFi PCR Premix (Clontech, Saint-Germain-en-Laye, France) at a final

concentration of 18 µM of Seed_BC.2 and 1.7 µM of Ultra_BC.2. Incubation conditions were 98°C for 3 min; 60°C for 10 s and extension at 72°C for 60 min.

Exon 3 mutagenesis

Site-directed mutagenesis was performed to generate the two alternate alleles at the eight common variant sites comprised within the haplotype previously defined in exon 3 of *OPN1LW/OPN1MW*: c.(453A>G; 457A>C; 465C>G; 511G>A; 513G>T; 521C>T; 532A>G; 538T>G; Ueyama et al., 2012; Buena-Atienza et al., 2016; see also Supp. Fig. S1). Ultramers were purchased from IDT. Ultra-Ex3-01-Fw: 5' ATG TGA GAT TTG ATG CCA AGC TGG CCA TCR TKG GCA TTG YCT TCT CCT GGR TCT GGK CTG CTG TGT GGA CAG CCC 3' and Ultra-Ex3-01-Rv: 5' CAG CTT GGC ATC AAA TCT CAC ATT GCC AAA GGG CTT GCA SAC CAC CAK CCA YCT CTC CCA GGA AAT GAT GGC C 3' were synthesized with equimolar mixtures of phosphoramidite monomers at the respective positions and share a 26 bp non-degenerate complementary region. Two PCR products were amplified – corresponding to the 5' and 3' end of the exon 3 and flanking intronic sequence upstream and downstream of exon 3 (108 and 111 bp, respectively) – with PfuUltra II Fusion HS DNA Polymerase (Agilent, Hamburg, Germany), 5 ng of the reference minigene construct as template (Buena-Atienza et al., 2016) and the following two primer pairs, i) Ex3-01-Fw: 5' GCC CTG AAT TCT GTG TGC AGA CGT TTG GGG TCT AAG CAG GAC AGT GGG AAG CTT TGC T and Ultra-Ex3-01-Rv, and ii) Ex3-01-Rv: 5' ATA CAT TGA TAG ACA TTG CAC GCT CAG AGG GGC CCA GAG AAA GGA AGT GAT TTG CCT TAA GGT CAC and Ultra-Ex3-01-Fw, respectively. PCR cycling conditions for i) and ii) were an initial denaturation of 96°C for 2 min, 96°C for 20s, 65ºC (i) or 70ºC (ii) for 20s and 72°C for 15s for 30 cycles, and a final extension step at 72°C for 2 min. Upon *Dpn*I digestion at 37°C for 1 h, the two PCR products were column-purified (MinElute PCR Purification Kit, Qiagen, Hilden, Germany) and used as template for an overlap-extension PCR (OE-PCR) with PfuUltra II Fusion HS DNA Polymerase (Agilent) and the outermost primers Ex-01-Fw and Ex-01-Rv. PCR cycling conditions for the OE PCR were an initial denaturation of 96°C for 2 min, 96°C for 20s, 56ºC for 20s and 72°C for 30s for 30 cycles, and a final extension step at 72°C for 2 min. The final 388 bp OE-PCR product was column-purified (MinElute PCR Purification Kit, Qiagen).

Barcode library cloning

The vector backbone was generated by inverse PCR with the reference minigene construct as template (Buena-Atienza et al., 2016) and primers IFC-inv-BC-Rv: 5'- CTA CTT GTC ATC GTC GTC CTT GTA A -3' and IFC-inv-BC-Fw: 5'- TAC TGG GTG GCA TCC CTG TGA -3'. The PfuUltra II Fusion HS DNA Polymerase (Agilent) was used for the inverse PCR with following cycling conditions: 96°C for 2 min, 96°C for 20s, 65ºC for 20s and 72°C for 145s for 30 cycles, and a final extension step at 72°C for 5 min Upon *Dpn*I digestion at 37°C for 1 h, the inverse PCR product was gel-purified with the MinElute Gel Extraction Kit (Qiagen). The double-stranded 91-mer barcode pool was cloned into the inverse PCR product downstream of exon 6 at the 3' of the stop codon by In-Fusion Cloning (IFC) with 21 bp of homology overlaps at each side. The IFC reaction was optimized to a molar ratio of 1:20 (vector:insert). The IFC reaction was incubated for 15 min at 50°C, following the IFC kit instructions (Clontech), diluted 1:25 in TE buffer and transformed in Stellar competent cells (Clontech). Bacterial outgrowth in liquid culture was avoided. A total of 33,000 colonies were harvested from eight independent transformations. All Petri dishes containing colonies stemming from the same transformation were harvested together as a pool for one single plasmid DNA isolation. Plasmid DNA was extracted with a total of eight midiprep columns (GenElute™ Plasmid Miniprep Kit, Sigma-Aldrich, Hamburg, Germany) and pooled in equal amounts of plasmid DNA.

Exon 3 mutagenesis library cloning

The barcode library was digested with *Hind*III and *Afl*II and the 9.3 kb band vector backbone purified by gel extraction (MinElute Gel Extraction Kit, Qiagen). The exon 3 mutagenesis library (=388 bp purified PCR product) was cloned by IFC into the digested barcode library with homology overlaps of 50 bp and 60 bp at the *Hind*III and *Afl*II overhangs, respectively. The IFC reaction was optimized to a molar ratio of 1:12 (vector:insert). Incubation and transformation was performed as indicated above. A total of 3,500 colonies from six transformations were pooled. The resulting plasmid library DNA was extracted as indicated above and we will refer to it as 'Input library'.

Input library Long-Read sequencing

The Input library (20 µg) was digested with *Hind*III and *Sal*I at a final concentration of 0.6U/µl and 0.4U/µl, respectively, for 3 h at 37°C and thereafter purified on a MinElute Spin Column (Qiagen). The 2.3 kb insert band was gel-purified – avoiding ethidium bromide staining and UV light exposure by only staining gel lanes with a small amount of sample at each side – using the MinElute Gel Extraction Kit (Qiagen) and eluted in 50 µl of 10 mM of Tris, pH 8. The total yield of the size-selected Input library 2.3 kb fragment was 1.5 µg as determined with the Qubit dsDNA HS Assay Kit on a Qubit 4.0 fluorometer (Thermo Firsher Scientific, Darmstadt, Germany) and used to generate the SMRT bell library according to the manufacturer's instructions (Pacific Biosciences, Menlo Park, California). The SMRTbell library with the Input Library as insert was sequenced on a Pacbio RS II instrument (Pacific Biosciences, Menlo Park, California) at the DNA sequencing facility of the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden and yielded 28,968 reads of insert with a mean of 21 passes. The mean read quality of the insert was 0.9977.

Data analysis Input Library sequencing

A total of 28,968 circular consensus sequence (CCS) reads were obtained from the alignment of subreads taken from a single zero-mode waveguide. Sequences extracted from the CCS fasta files were aligned to the sequence of the reference minigene construct *Hin*dIII-*Sal*I fragment (Buena-Atienza et al., 2016) using Clustal Omega (Goujon et al., 2010; Sievers et al., 2011).

Reads were reverse-complemented when necessary. A custom R script called the SNPs in exon 3 and filtered out sequences that did not meet the requirements, e.g., unaligned sequences, sequences lacking either barcode or exon 3 or partially truncated. Sequences that passed the filters were grouped by unique barcodes and subsequently, by unique haplotypes. In total, the library comprised 649 uniquely barcoded haplotypes tagged by a total of 1,952 unique barcodes. These uniquely barcoded haplotypes could be divided into three groups: i) haplotypes comprising targeted variants ("targeted haplotypes"), ii) haplotypes comprising targeted and non-targeted variants, and iii) haplotypes comprising only non-targeted variants. The first group consists of 248 different haplotypes out of the 256 (96.9%) theoretical maximum of the site-directed mutagenesis strategy. The eight haplotypes not represented in this group (3.1%) were either not observed or not uniquely tagged. On average, each unique targeted haplotype was tagged by six different unique barcodes and covered by 29 CCS (as a sum of all reads with different barcodes per unique haplotype).

Parallelized minigene splicing assay

The splicing assay was performed by transfecting HEK293T cells with the Input Library in four replicate experiments. HEK293T cells were routinely maintained in Dulbecco´s modified Eagle´s medium, DMEM (Gibco, Life Technologies) supplemented with 10% fetal bovine serum (FBS; Gibco, Life Technologies), Penicillin-Streptomycin (Sigma-Aldrich Chemie GmbH) at 100 µg/ml, and Amphotericin B (Biochrom GmbH, Merck, Darmstadt, Germany) at 2.5 µg/ml. Cells were seeded in 6-well plates in DMEM with 10% FBS and the following day, at 80-90% confluency, four biological replicates were transfected with 4 µg DNA of the Input Library pool and 5 ng of an equimolar mixture of two minigene control plasmids: "Reference" (Control 1)- and 'LIAVA'(Control 2)-minigenes tagged with the unique barcodes "TAATGGGG" and "CATTGGCC", respectively. 20 µl Lipofectamine 2000 (ThermoFischer GmbH) per well and OptiMEM® (Life Technologies) were used as diluents and medium. Following 6 h incubation, cells were detached with 0.05% Trypsine-EDTA (Gibco, Life Technologies), centrifuged at 1500 rpm for 5 min and transferred to a 6 cm dish with DMEM supplemented with 10% FBS, Penicillin-Streptomycin and Amphotericin B at the aforementioned concentrations. 24 h post-transfection, cells were lysed and total RNA was extracted using the peqGOLD Total RNA Kit (PEQLAB Biotechnologie GmbH, Erlangen, Germany). Individual minigene splicing assays and semi-quantification of correctly spliced transcripts were performed as previously described (Buena-Atienza et al., 2016) for a total of 20 barcoded minigene constructs collected during quality control of Input Library clones by colony PCR and Sanger sequencing.

Output Library Long-Read sequencing

First strand cDNA synthesis was performed with Transcriptor High Fidelity cDNA Synthesis Kit (Roche, Mannheim, Germany) with 2 µg of total RNA in a total volume of 20 µl. A minigene-specific primer binding downstream of the barcode, IFC-downstream-Rv: 5'- ATT AGG ACA AGG CTG GTG GGC -3', was used for priming and reverse transcription was incubated at 50°C for 60 min. For each of the four replicates, a total of 10µl cDNA was amplified with KAPA Lib Amp Kit (Roche) and with one pair of HPLC-purified primers (Supplementary Table S1). Reverse transcriptase (RT-)PCR cycling conditions were 98°C for 45s, 98°C for 15s and 72°C for 10s for 24 cycles, and a final extension step at 72°C for 1 min. Amplicons were immediately purified with AMPure PB beads (Pacific Biosciences) at 0.78X and pooled in an equimolar ratio and ~1µg of DNA was used for SMRTbell sequencing library preparation according to manufacturer's instructions (SMRTbell™ Template Prep Kit, Pacific Biosciences). Adaptors for the blunt end ligation were used in a 42X molar excess. All purifications within the SMRTbell library preparation were performed with a 0.78X ratio of AMPure PB beads. The SMRTbell sequencing library was loaded at 2 pM by diffusion on one SMRT cell for sequencing on a Sequel instrument (Pacific Biosciences).

Output Library Short-Read sequencing

First strand cDNA synthesis was done as described above. A two-step PCR approach was applied to introduce offset spacers, random N bases and sequences for Illumina sequencing primers in the first PCR round while indices and grafting P5 and P7 adaptors were added during the second round of amplification. Ultramer oligonucleotides (Supp. Table S2) for amplification introduced 16 bases (spacers and random bases to cope with the low diversity for sequencing; see also Supplementary Figure S3B) for each primer combination and contained a phosphorothioate bond (*) between the last and penultimate base. For each of the four replicates (L1, L2, L3 and L4), a total of 10µl cDNA was used for eight parallel PCR reactions and amplified with KAPA Lib Amp Kit (Roche). Cycling conditions for the first PCR were 98°C for 45s, 98°C for 15s, 60°C for 5s and 72°C for 10s for 24 cycles, and a final extension step at 72°C for 1 min. PCRs were cleaned-up with QIAquick PCR Purification Kit QIAGEN columns and quantified with dsDNA HS Assay kit and Qubit 4. For each replicate L1-4, two primer combinations were used in the first amplification and equal amounts of PCR products pooled upon quantification of purified amplicons. For the second PCR, eight reactions for each L1-4 with 1 ng of the first PCR amplicons as template were used for amplification with KAPA Lib Amp Kit (Roche) to introduce indices (Supp.Table S2; Supp. Fig. S3B) that enable multiplexing the four replicates for sequencing. PCR cycling conditions were 98°C for 45s, 98°C for 15s, 60°C for 30s and 72°C for 15s for 10 cycles, and a final extension step at 72°C for 1 min. Products of the second PCR were size-selected with 1.8X AMPure XP beads (Beckman Coulter, Krefeld, Germany). Quality control of L1-4 libraries was performed with a DNA7500 assay on the 2100 Bioanalyzer (Agilent Technologies), quantitation with Qubit dsDNA HS Assay Kit and quantitative PCR with the KAPA Library Quantification Kit (Roche). A MiSeq paired-end Nano flow cell (v2, Illumina) was loaded with 2.7 pM of an equimolar pool of L1-4 Output Libraries and sequencing was performed with 115, 8 and 202 cycles for Read 1, Index Read and Read 2, respectively. The "Chemistry.file" was modified to adapt the number of amplification cycles and the duration of First Extension. We obtained a total of 664,730 paired-end reads with an averaged ≥Q30 score of 89.7% (96.8M).

Data processing and integration of Output and Input Library data

Generation of the custom 'reference assembly' using the Output Library from PacBio long-read sequencing and its combination with Illumina data by mapping short paired-end reads against this 'reference assembly' is described in Supplementary Material, Supplementary Material and Methods. Minigene-tagging 8-bp barcodes were identified for all reads and the exon 3 haplotype in phase was assigned. CCS reads were grouped by unique barcodes. A unique identifier was given for each haplotype of the Input Library with at least one unique barcode. For each CCS read an alignment to the transcript reference sequence including exon 3 and relevant exonic junctions with exon 2 and exon 4 was calculated. A distance score threshold of 0.2 was used to discern the exon 3 retention reads from the exon 3 skipped reads based on the reads distribution (concordant with the expected size distribution). Short-read counts for retained or skipped exon 3 were obtained by mapping of short Illumina reads on the PacBio based 'reference assembly' as well as an average distance score for long CCS reads for each haplotype. Relative percentages of retention levels using short-read data were calculated for each replicate sample.

***OPN1LW* and *OPN1MW* exon 3 haplotypes from control and patient population**

Exon 3 sequences of the proximal (*OPN1LW* or *OPN1LW/MW* hybrids) and the distal (*OPN1MW* or *OPN1MW/LW* hybrid) gene copies of the cone opsin gene cluster were determined as described previously (Buena-Atienza et al. 2016) from a sample of 243 male subjects clinically diagnosed with Blue Cone Monochromacy (BCM), Achromatopsia or Cone Dystrophy. Ambiguous haplotypes (e.g. multiple polymorphic variants in exon 3 sequences) or haplotypes from *OPN1LW/MW* gene copies with other pathogenic variants (e.g. the common p.[C203R] mutation) were discarded. Haplotypes obtained from subjects with no other gene mutation explaining the disease were combined in the 'disease group'. Haplotypes obtained from subjects with a mutation in a gene other than in *OPN1LW* and/or *OPN1MW* explaining the disease (e.g. a homozygous or compound heterozygous mutation in a known Achromatopsia gene) served as population controls ('control group'). We further compiled *OPN1LW/MW* exon 3 haplotypes from the available literature (Buena-Atienza et al., 2016; Carroll et al., 2012; Carroll et al., 2010; Crognale et al., 2004; Deeb et al., 1995; Gardner et al., 2010; Gardner et al., 2014; Greenwald et al., 2017; Hayashi et al., 2001; Jordan et al., 2010; Katagiri et al., 2018; Li et al., 2015; McClements et al., 2013; Mizrahi-Meissonnier et al., 2010; Mountford et al., 2019; Neitz et al., 1999; Orosz et al., 2017; Patterson et al., 2016; Ueyama et al., 2004) and manually attributed these haplotypes to the 'disease group' (subjects with BCM, XLCD, or colour vision defects not explained by other genetic defects) or the 'control group' (healthy controls, colour vision defects explained by single gene copies or hybrid gene copies).

Acknowledgements

We would like to thank members of the c.ATG Core Facility at the Institute of Medical Genetics and Applied Genomics in Tübingen and Christa Lanz from the Max Planck Institute for Developmental Biology in Tübingen for excellent technical assistance. This work was supported by the European Union's Seventh Framework Programme for research, technological development and demonstration [grant agreement no 317472] and by the BCM Families Foundation.

Authors' contributions

EB-A performed and designed all experiments, secondary data analysis and wrote the manuscript. MCC performed computation of long-read data from Input and Output Library sequencing and generated the reference scaffold. MCC and SC worked on the hybrid mapping of short-read data on the reference scaffold and generated primary data analysis. SK and BB screened haplotypes from our patient and control groups. SP supported sequencing of the Output Library. SN assisted with the design of bioinformatic analysis. BW conceived the study, assisted with experimental design, compiled haplotypes from disease and control groups, wrote the manuscript and supervised the whole study. All authors read and revised the manuscript.

Disclosure Declaration

The authors declare that they have no competing interests.

References

Adamson, S.I., Zhan, L., and Graveley, B.R. (2018). Vex-seq: high-throughput identification of the impact of genetic variation on pre-mRNA splicing efficiency. Genome Biol. *19*, 71.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.

Anna, A., and Monika, G. (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. J. Appl. Genet. 1–16.

Au, K.F., Underwood, J.G., Lee, L., and Wong, W.H. (2012). Improving PacBio Long Read Accuracy by Short Read Alignment. PLOS ONE *7*, e46679.

Au, K.F., Sebastiano, V., Afshar, P.T., Durruthy, J.D., Lee, L., Williams, B.A., Bakel, H. van, Schadt, E.E., Reijo-Pera, R.A., Underwood, J.G., et al. (2013). Characterization of the human ESC transcriptome by hybrid sequencing. Proc. Natl. Acad. Sci. *110*, E4821–E4830.

Buena-Atienza, E., Rüther, K., Baumann, B., Bergholz, R., Birch, D., Baere, E.D., Dollfus, H., Greally, M.T., Gustavsson, P., Hamel, C.P., et al. (2016). De novo intrachromosomal gene conversion from OPN1MW to OPN1LW in the male germline results in Blue Cone Monochromacy. Sci. Rep. *6*, 28253.

Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). Splicing of Precursors to mRNAs by the Spliceosomes. In The RNA world Second edition (ed. Gesteland RF et al.), 525–560 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Caffrey, T.M., Joachim, C., and Wade-Martins, R. (2008). Haplotype-specific expression of the N-terminal exons 2 and 3 at the human MAPT locus. Neurobiol. Aging *29*, 1923–1929.

Carroll, J., Dubra, A., Gardner, J.C., Mizrahi-Meissonnier, L., Cooper, R.F., Dubis, A.M., Nordgren, R., Genead, M., Connor, T.B. Jr., Stepien, K.E., Sharon, D., Hunt, D.M., Banin, E., Hardcastle, A.J., Moore, A.T., Williams, D.R., Fishman, G., Neitz, J, Neitz, M., Michaelides, M. (2012). The effect of cone opsin mutations on retinal structure and the integrity of the photoreceptor mosaic. Invest. Ophthalmol. Vis. Sci. *53,*8006–80015. doi: 10.1167/iovs.12-11087.

Carroll, J., Dubra, A., Wagner-Schuman, M., Neitz, J., McClements, M., Hunt, D.M., Mollon, J.D., Moore, A.T., Neitz, M., Michaelides, M. (2010). Genetic mutations within the L/M array account for both cone dysfunction and the color vision phenotype in Bornholm eye disease. XIVth Symposium on Retinal Degeneration, Mont-Treblant, Canada (July, 2010).

Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat. Rev. Genet. *3*, 285–298.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. (2003). ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res. *31*, 3568–3571.

Castel, S.E., Cervera, A., Mohammadi, P., Aguet, F., Reverter, F., Wolman, A., Guigo, R., Iossifov, I., Vasileva, A., and Lappalainen, T. (2018). Modified penetrance of coding variants by cis-regulatory variation shapes human traits. BioRxiv 190397.

Cheung, R., Insigne, K.D., Yao, D., Burghard, C.P., Wang, J., Hsiao, Y.-H.E., Jones, E.M., Goodman, D.B., Xiao, X., and Kosuri, S. (2019). A Multiplexed Assay for Exon Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions. Mol. Cell *73*, 183–194.

Choudhury, O., Chakrabarty, A., and Emrich, S.J. (2018). HECIL: A Hybrid Error Correction Algorithm for Long Reads with Iterative Learning. Sci. Rep. *8*, 9936.

Crognale, M.A., Fry, M., Highsmith, J., Haegerstrom-Portnoy, G., Neitz, M., Neitz, J., Webster, M.A. (2004). Characterization of a novel form of X-linked incomplete achromatopsia. Vis. Neurosci. *21*, 197–203.

Deeb, S.S., Alvarez, A., Malkki, M., Motulsky, A.G. (1995). Molecular patterns and sequence polymorphisms in the red and green visual pigment genes of Japanese men. Hum. Genet. *95*, 501–506.

Deshpande, V., Fung, E.D.K., Pham, S., and Bafna, V. (2013). Cerulean: A Hybrid Assembly Using High Throughput Short and Long Reads. Springer Berlin Heidelberg, 349–363.

D'Souza, I., Poorkaj, P., Hong, M., Nochlin, D., Lee, V.M.-Y., Bird, T.D., and Schellenberg, G.D. (1999). Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. Proc. Natl. Acad. Sci. *96*, 5598–5603.

Erkelenz, S., Theiss, S., Otte, M., Widera, M., Peter, J.O., and Schaal, H. (2014). Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic Acids Res. *42*, 10681–10697.

Gardner, J.C., Webb, T.R., Kanuga, N., Robson, A.G., Holder, G.E., Stockman, A., Ripamonti, C., Ebenezer, N.D., Ogun, O., Devery, S., Wright, G.A., Maher, E.R., Cheetham, M.E., Moore, A.T., Michaelides, M., Hardcastle, A.J. (2010). X-linked cone dystrophy caused by mutation of the red and green cone opsins. Am. J. Hum. Genet. *87,* 26-39.

Gardner, J.C., Liew, G., Quan, Y.-H., Ermetal, B., Ueyama, H., Davidson, A.E., Schwarz, N., Kanuga, N., Chana, R., Maher, E.R., et al. (2014). Three Different Cone Opsin Gene Array Mutational Mechanisms with Genotype–Phenotype Correlation and Functional Investigation of Cone Opsin Variants. Hum. Mutat. *35*, 1354–1362.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL–EBI. Nucleic Acids Res. *38*, W695–W699.

Greenwald, S.H., Kuchenbecker, J.A., Rowlan, J.S., Neitz, J., Neitz, M. (2017). Role of a Dual Splicing and Amino Acid Code in Myopia, Cone Dysfunction and Cone Dystrophy Associated with L/M Opsin Interchange Mutations. Transl. Vis. Sci Technol. *6*, 2. doi: 10.1167/tvst.6.3.2.

Haque, A., Buratti, E., and Baralle, F.E. (2010). Functional properties and evolutionary splicing constraints on a composite exonic regulatory element of splicing in CFTR exon 12. Nucleic Acids Res. *38*, 647–659.

Hayashi, S., Ueyama, H., Tanabe, S., Yamade, S., Kani, K. (2001). Number and variations of the red and green visual pigment genes in Japanese men with normal color vision. Jpn. J. Ophthalmol. *45*, 60–67.

Julien, P., Miñana, B., Baeza-Centurion, P., Valcárcel, J., and Lehner, B. (2016). The complete local genotype–phenotype landscape for the alternative splicing of a human exon. Nat. Commun. *7*, 11558.

Jordan, G., Deeb, S.S., Bosten, J.M., Mollon, J.D. (2010). The dimensionality of color vision in carriers of anomalous trichromacy. J. Vis. *10,*12.

Karbassi, I., Maston, G.A., Love, A., DiVincenzo, C., Braastad, C.D., Elzinga, C.D., Bright, A.R., Previte, D., Zhang, K., Rowland, C.M., et al. (2016). A Standardized DNA Variant Scoring System for Pathogenicity Assessments in Mendelian Disorders. Hum. Mutat. *37*, 127.

Katagiri, S., Iwasa, M., Hayashi, T., Hosono, K., Yamashita, T., Kuniyoshi, K., Ueno, S., Kondo, M., Ueyama, H., Ogita, H., Shichida, Y., Inagaki, H., Kurahashi, H., Kondo, H., Ohji, M., Hotta, Y., Nakano, T. (2018). Genotype determination of the OPN1LW/OPN1MW genes: novel disease-causing mechanisms in Japanese patients with blue cone monochromacy. Sci. Rep. *8,* 11507.

Kitzman, J.O., Starita, L.M., Lo, R.S., Fields, S., and Shendure, J. (2015). Massively Parallel Single Amino Acid Mutagenesis. Nat. Methods *12*, 203–206.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. *4*, 1073–1081.

Laver, T.W., Caswell, R.C., Moore, K.A., Poschmann, J., Johnson, M.B., Owens, M.M., Ellard, S., Paszkiewicz, K.H., and Weedon, M.N. (2016). Pitfalls of haplotype phasing from amplicon-based long-read sequencing. Sci. Rep. *6*, 21746.

Li, J., Gao, B., Guan, L., Xiao, X., Zhang, J., Li, S., Jiang, H., Jia, X., Yang, J., Guo, X., Yin, Y., Wang, J., Zhang, Q. (2015). Unique Variants in *OPN1LW* Cause Both Syndromic and Nonsyndromic X-Linked High Myopia Mapped to MYP1. Invest. Ophthalmol. Vis. Sci. *56,* 4150–4155.

McClements, M., Davies, W.I., Michaelides, M., Young, T., Neitz, M., MacLaren, R.E., Moore, A.T., Hunt, D.M. (2013). Variations in opsin coding sequences cause x-linked cone dysfunction syndrome with myopia and dichromacy. Invest. Ophthalmol. Vis. Sci. *54*, 1361–1369.

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Jr, C.G.C., Kinney, J.B., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. *30*, 271–277.

Mizrahi-Meissonnier, L., Merin, S., Banin, E., Sharon, D. (2010). Variable retinal phenotypes caused by mutations in the X-linked photopigment gene array. Invest. Ophthalmol. Vis. Sci. *51*, 3884–3892.

Merbs S.L., Nathans J. (1992). Absorption spectra of the hybrid pigments responsible for anomalous color vision. Science *16*, 464-466.

Monani, U.R., Lorson, C.L., Parsons, D.W., Prior, T.W., Androphy, E.J., Burghes, A.H.M., and McPherson, J.D. (1999). A Single Nucleotide Difference That Alters Splicing Patterns Distinguishes the SMA Gene SMN1 From the Copy Gene SMN2. Hum. Mol. Genet. *8*, 1177–1183.

Mountford, J.K., Davies, W.I.L., Griffiths, L.R., Yazar, S., Mackey, D.A., Hunt, D.M. (2019). Differential stability of variant *OPN1LW* gene transcripts in myopic patients. Mol. Vis. *25*, 183–193.

Mueller, W.F., Larsen, L.S.Z., Garibaldi, A., Hatfield, G.W., and Hertel, K.J. (2015). The Silent Sway of Splicing by Synonymous Substitutions. J. Biol. Chem. *290*, 27700–27711.

Nathans, J., Davenport, C.M., Maumenee, I.H., Lewis, R.A., Hejtmancik, J.F., Litt, M., Lovrien, E., Weleber, R., Bachynski, B., Zwas, F., et al. (1989). Molecular genetics of human blue cone monochromacy. Science *245*, 831–838.

Neitz, J., Neitz, M., He, J.C., Shevell, S.K. (1999). Trichromatic color vision with only two spectrally distinct photopigments. Nat Neurosci. *2,* 884–888.

Nykamp, K., Anderson, M., Powers, M., Garcia, J., Herrera, B., Ho, Y.-Y., Kobayashi, Y., Patil, N., Thusberg, J., Westbrook, M., et al. (2017). Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. Genet. Med. *19*, 1105–1117.

Orosz, O., Rajta, I., Vajas, A., Takács, L., Csutak, A., Fodor, M., Kolozsvári, B., Resch, M., Sényi, K., Lesch, B., Szabó, V., Berta, A., Balogh, I., Losonczy, G. (2017). Myopia and Late-Onset Progressive Cone Dystrophy Associate to LVAVA/MVAVA Exon 3 Interchange Haplotypes of Opsin Genes on Chromosome X. Invest Ophthalmol Vis Sci. *58,* 1834–1842.

Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The Expanding Landscape of Alternative Splicing Variation in Human Populations. Am. J. Hum. Genet. *102*, 11–26.

Park, S.A., Ahn, S.I., and Gallo, J.-M. (2016). Tau mis-splicing in the pathogenesis of neurodegenerative disorders. BMB Rep. *49*, 405.

Patterson, E.J., Wilk, M., Langlo, C.S., Kasilian, M., Ring, M., Hufnagel, R.B., Dubis, A.M., Tee, J.J., Kalitzeos, A., Gardner, J.C., Ahmed, Z.M., Sisk, R.A., Larsen, M., Sjoberg, S., Connor, T.B., Dubra, A., Neitz, J., Hardcastle, A.J,, Neitz, M., Michaelides, M., Carroll, J. (2016). Cone Photoreceptor Structure in Patients With X-Linked Cone Dysfunction and Red-Green Color Vision Deficiency. Invest Ophthalmol Vis Sci. *57*,3853–3863.

Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.-I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. *30*, 265–270.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). 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.

Rosenberg, A.B., Patwardhan, R.P., Shendure, J., and Seelig, G. (2015). Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences. Cell *163*, 698–711.

Savisaar, R., and Hurst, L.D. (2017). Estimating the prevalence of functional exonic splice regulatory information. Hum. Genet. *136*, 1059–1078.

Schwartz, S., Hall, E., and Ast, G. (2009). SROOGLE: webserver for integrative, user-friendly visualization of splicing signals. Nucleic Acids Res. *37*, W189-W192.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high• quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Sobrido, M.-J. (2016). Are Functional Assays for Pathogenicity Assessment of Genetic Variants Overrated? Hum. Mutat. *38*, 5–5.

Soemedi, R., Cygan, K.J., Rhine, C.L., Wang, J., Bulacan, C., Yang, J., Bayrak-Toydemir, P., McDonald, J., and Fairbrother, W.G. (2017). Pathogenic variants that alter protein code often disrupt splicing. Nat. Genet. *49*, 848–855.

Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frébourg, T., Tosi, M., and Martins, A. (2016). Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. PLOS Genet. *12*, e1005756.

Staffa, A., and Cochrane, A. (1995). Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of human immunodeficiency virus type 1. Mol. Cell. Biol. *15*, 4597–4605.

Starita, L.M., and Fields, S. (2015). Deep Mutational Scanning: A Highly Parallel Method to Measure the Effects of Mutation on Protein Function. Cold Spring Harb. Protoc. *2015*, 711–714.

Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J., and Fields, S. (2015). Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. Genetics *200*, 413–422.

Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J., and Fowler, D.M. (2017). Variant Interpretation: Functional Assays to the Rescue. Am. J. Hum. Genet. *101*, 315–325.

Steffensen, A.Y., Dandanell, M., Jønson, L., Ejlertsen, B., Gerdes, A.-M., Nielsen, F.C., and Hansen, T. vO (2014). Functional characterization of BRCA1 gene variants by mini-gene splicing assay. Eur. J. Hum. Genet. *22*, 1362–1368.

Tammaro, C., Raponi, M., Wilson, D.I., and Baralle, D. (2014). BRCA1 Exon 11, a CERES (Composite Regulatory Element of Splicing) Element Involved in Splice Regulation. Int. J. Mol. Sci. *15*, 13045–13059.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. *28*, 2731–2739.

Ueyama, H., Kuwayama, S., Imai, H., Oda, S., Nishida, Y., Tanabe, S., Shichida, Y., Yamade, S. (2004). Analysis of L-cone/M-cone visual pigment gene arrays in Japanese males with protan color-vision deficiency. Vision Res. *44*, 2241–2252.Ueyama, H., Muraki-Oda, S., Yamade, S., Tanabe, S., Yamashita, T., Shichida, Y., and Ogita, H. (2012). Unique haplotype in exon 3 of cone opsin mRNA affects splicing of its precursor, leading to congenital color vision defect. Biochem. Biophys. Res. Commun. *424*, 152–157.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Weisschuh, N., Wissinger, B., Gramer, E. (2012). A splice site mutation in the PAX6 gene which induces exon skipping causes autosomal dominant inherited aniridia. Mol. Vis. *18*, 751–757

Winderickx, J., Battisti, L., Hibiya, Y., Motulsky, A.G., Deeb, S.S. (1993). Haplotype diversity in the human red and green opsin genes: evidence for frequent sequence exchange in exon 3. Hum. Mol. Genet. *2*, 1413–1421.

Woods, N.T., Baskin, R., Golubeva, V., Jhuraney, A., De-Gregoriis, G., Vaclova, T., Goldgar, D.E., Couch, F.J., Carvalho, M.A., Iversen, E.S., et al. (2016). Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. Npj Genomic Med. *1*, 16001.

Zhu, J., Mayeda, A., and Krainer, A.R. (2001). Exon Identity Established through Differential Antagonism between Exonic Splicing Silencer-Bou.nd hnRNP A1 and Enhancer-Bound SR Proteins. Mol. Cell *8*, 1351–1361.

List of Abbreviations

Abbreviation	Description
BCM	Blue Cone Monochromacy
CCS	Circular Consensus Sequencing
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
IFC	In-Fusion Cloning;
LWS	long-wavelength sensitive
MAVEs	multiplex assay of variant effect
MWS	middle-wavelength-sensitive
NGS	next-generation sequencing
OE-PCR	overlap-extension PCR
OPN1LW	human long wavelength-sensitive cone opsin gene
OPN1MW	human middle wavelength-sensitive cone opsin gene
RT-PCR	Reverse transcriptase PCR
SNV	single nucleotide polymorphism
VUS	variant of uncertain significance
XLCD	X-linked cone dysfunction
BCM	Blue Cone Monochromacy

Table 1. TOP15 Haplotypes inducing exon retention or skipping

Nr.	ID*	c.453 A>G	c.457 A>C	c.465 C>G	c.511 G>A	c.513 G>T	c.521 C>T	c.532 A>G	c.538 T>G	% Exon Retention †	SD‡	Distance Score§	Total CCS¶	Read Count#	Mutation Rate**	Frequency Input Library††	Nr. of Barcodes‡‡
TOP15 Haplotypes Exon Retention																	
1	id_7	-	C	G	-	-	T	-	-	99.5	0.3	0.14	31	5,147	0.38	0.0028	4
2	id_215	-	-	G	-	T	T	-	-	98.7	0.7	0.18	34	28,322	0.38	0.0017	3
3	id_190	-	-	G	-	-	-	-	-	97.7	1.0	0.15	30	25,362	0.13	0.0028	2
4	id_123	G	C	-	-	T	T	-	-	97.6	1.2	0.21	33	30,108	0.50	0.0053	4
5	id_220	-	-	G	-	-	-	-	G	95.1	6.5	0.16	6	311	0.25	0.0006	2
6	id_55	-	C	-	-	T	-	-	-	93.4	1.3	0.15	58	15,727	0.25	0.0068	6
7	id_246	-	-	-	-	T	T	-	-	93.2	1.7	0.17	23	5,646	0.25	0.0026	4
8	id_212	-	-	G	-	T	-	-	-	91.7	2.0	0.18	65	10,874	0.25	0.0060	6
9	id_222	-	-	-	-	-	T	-	-	91.4	0.3	0.12	76	11,057	0.13	0.0067	9
10	id_154	G	-	G	-	T	-	-	-	91.3	1.3	0.16	21	2,926	0.38	0.0021	2
11	id_132	G	-	G	-	-	-	-	-	91.2	0.5	0.13	10	794	0.25	0.0019	3
12	id_49	-	C	-	A	T	T	-	-	86.8	6.3	0.23	21	1,659	0.50	0.0029	3
13	id_1	-	C	-	-	-	-	-	-	86.4	3.8	0.17	78	9,508	0.13	0.0085	10
14	id_197	-	-	G	A	-	-	-	-	86.1	3.8	0.14	10	667	0.25	0.0015	4
15	id_68	G	C	G	-	-	-	-	-	85.0	4.7	0.17	40	3,057	0.38	0.0050	5
TOP15 Haplotypes Exon Skipping																	
1	id_235	.	-	-	A	T	-	G	G	2.6	0.2	0.48	88	4,800	0.50	0.0106	11
2	id_70	G	C	G	-	C	-	G	G	2.5	0.9	0.49	45	1,245	0.63	0.0051	9
3	id_28	-	C	G	-	T	-	G	G	1.9	1.2	0.48	57	2,682	0.63	0.0044	10
4	id_177	G	-	-	A	T	T	G	G	1.8	1.1	0.49	60	5,707	0.75	0.0053	8
5	id_85	G	C	G	A	T	-	G	G	1.6	1.0	0.49	57	2,291	0.88	0.0046	5
6	id_19	-	C	G	A	T	-	G	G	1.5	1.4	0.51	71	2,566	0.75	0.0074	8
7	id_41	-	C	-	A	-	-	G	G	1.0	0.6	0.49	201	37,388	0.50	0.0132	15
8	id_127	G	C	-	-	T	-	-	G	1.0	0.5	0.51	277	34,155	0.50	0.0050	8
9	id_80	G	C	G	A	-	T	G	G	1.0	0.6	0.49	37	3,813	0.88	0.0022	7
10	id_48	-	C	-	A	T	-	G	G	0.8	1.1	0.49	60	3,808	0.63	0.0050	8
11	id_109	G	C	-	A	-	T	G	G	0.8	0.2	0.50	97	30,768	0.75	0.0078	14
12	id_44	-	C	-	A	-	T	G	G	0.5	0.7	0.50	135	8,536	0.63	0.0099	16
13	id_84	G	C	G	A	T	-	G	-	0.3	0.2	0.49	44	29,696	0.75	0.0053	6
14	id_168	G	-	-	A	-	-	G	G	0.3	0.4	0.49	74	10,735	0.50	0.0060	11
15	id_116	G	C	-	A	T	T	G	G	0.2	0.1	0.49	65	27,308	0.88	0.0053	8

*Identification code for each exon 3 haplotype uniquely tagged found in the Output Library; †Mean % of exon 3 retention rates of four replicates; ‡Standard Deviation of exon retention % over the four replicates; §Mean distance score indicator for exon retention (0.0-0.2) or skipping (0.2-0.5) calculated as the mean for each ID from CCS reads aligned to exon 3. ¶Total count of CCS reads (PacBio) for each haplotype ID; #Total count of Short reads (Illumina) mapped to the reference assembly for all four replicates; **Rate of variant changes present in the haplotype (0-1); ††Frequency of each haplotype in the Input Library among all uniquely tagged exon 3 haplotypes calculated from the read count of Input Library sequencing dataset; ‡‡Total number of unique barcodes detected in both Input and Output Library for each haplotype.

Table 2. Exon 3 haplotypes in human subjects

Haplotype	'trivial' AA haplotype*	No. 'disease group'†	No. 'control group'‡	p-value (<0.05) §
A-A-C-A-T-C-A-T	MIAIS	1	1	
A-A-C-A-T-C-G-G	MIAVA	5	0	↑ 0.0207
A-A-C-G-G-C-A-G	MVAIA	28	244	↓ 3.82E-08
A-A-C-G-G-C-A-T	MVAIS	4	80	↓ 0.00016
A-A-C-G-G-C-G-G	MVAVA	6	3	
A-A-C-G-G-T-A-G	MVVIA	1	3	
A-A-C-G-G-T-A-T	MVVIS	2	6	
A-A-C-G-G-T-G-G	MVVVA	27	10	↑ 2.25E-09
A-A-C-G-G-T-G-T	MVVVS	1	0	
A-A-G-A-T-C-A-G	MIAIA	0	1	
A-A-G-A-T-C-G-G	MIAVA	3	0	
A-A-G-G-G-C-A-G	MVAIA	0	12	
A-A-G-G-G-C-A-T	MVAIS	1	2	
A-A-G-G-G-C-G-G	MVAVA	2	1	
A-A-G-G-G-T-A-T	MVVIS	1	1	
A-A-G-G-G-T-G-G	MVVVA	5	21	
G-C-C-G-G-C-A-G	LVAIA	0	1	
G-C-C-G-G-C-A-T	LVAIS	0	4	
G-C-G-A-G-C-A-T	LIAIS	0	1	
G-C-G-A-G-C-G-T	LIAVS	0	1	
G-C-G-A-T-C-A-G	LIAIA	1	1	
G-C-G-A-T-C-A-T	LIAIS	5	26	
G-C-G-A-T-C-G-G	LIAVA	32	0	↑ 2.47E-20
G-C-G-A-T-C-G-T	LIAVS	3	0	
G-C-G-A-T-T-G-G	LIVVA	2	0	
G-C-G-G-G-C-A-G	LVAIA	28	94	
G-C-G-G-G-C-A-T	LVAIS	39	284	↓ 1.84E-07
G-C-G-G-G-C-G-G	LVAVA	27	1	↑ 1.40E-15
G-C-G-G-G-C-G-T	LVAVS	1	2	
G-C-G-G-G-T-A-G	LVVIA	1	2	
G-C-G-G-G-T-A-T	LVVIA	2	7	
G-C-G-G-G-T-G-G	LVVVA	14	3	↑ 9.92E-06
G-C-G-G-G-T-G-T	LVVVS	1	0	
	Σ	**243**	**812**	

*Trivial haplotype nomenclature based on the amino acid changes encoded by the haplotype; †Number of haplotypes observed in the *OPN1LW* or *OPN1MW* gene copies in the defined 'disease group'; ‡Number of haplotypes observed in the *OPN1LW* or *OPN1MW* gene copies in the defined 'control group'; §Fishers exact test (test haplotype *versus* the sum of all other haplotypes, Bonferroni corrected). Note that the 'disease group' represents a heterogeneous group which might include subjects with other disease-causing mutations not linked to pathogenic variants in the *OPN1LW/MW* gene cluster.

Figures

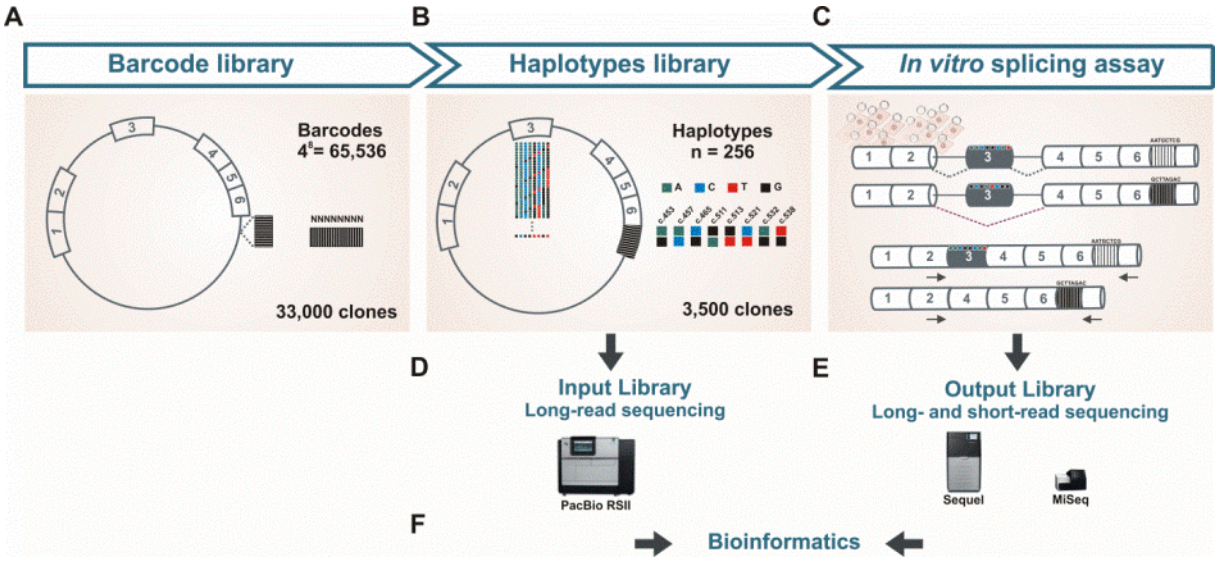


Figure 1. Workflow of the parallelized minigene splicing assay. A) A barcode library is generated by the insertion of a random octamer sequence in the backbone of the *OPN1LW* minigene vector. **B)** Site-directed mutagenesis is used to generated exon 3 haplotypes and retro-fitted into the barcode library vector backbone yielding the Input Library (Supp. Fig. S2) consisting of 256 haplotypes in exon 3 of the *OPN1LW* barcoded minigene. **C**) Long-read sequencing (PacBio RSII) of the Input Library allows matching the haplotype with cognate barcode for each molecule. **D**) *In vitro* splicing assay by *en masse* transfection of the Input Library pool of 256 *OPN1LW/MW* barcoded minigenes **E**) Short- and Long-read sequencing (MiSeq and Sequel, respectively) are used to sequence RT-PCR amplicons of the *OPN1LW/MW* minigene transcript pool including 'exon 3 retained' and 'exon 3 skipped' transcripts (Output Library; Supp. Fig. S3). **F**) Bioinformatic analysis. The proportion of 'exon 3 included' and 'exon 3 skipped' transcripts for each haplotype is calculated using barcodes as proxies.

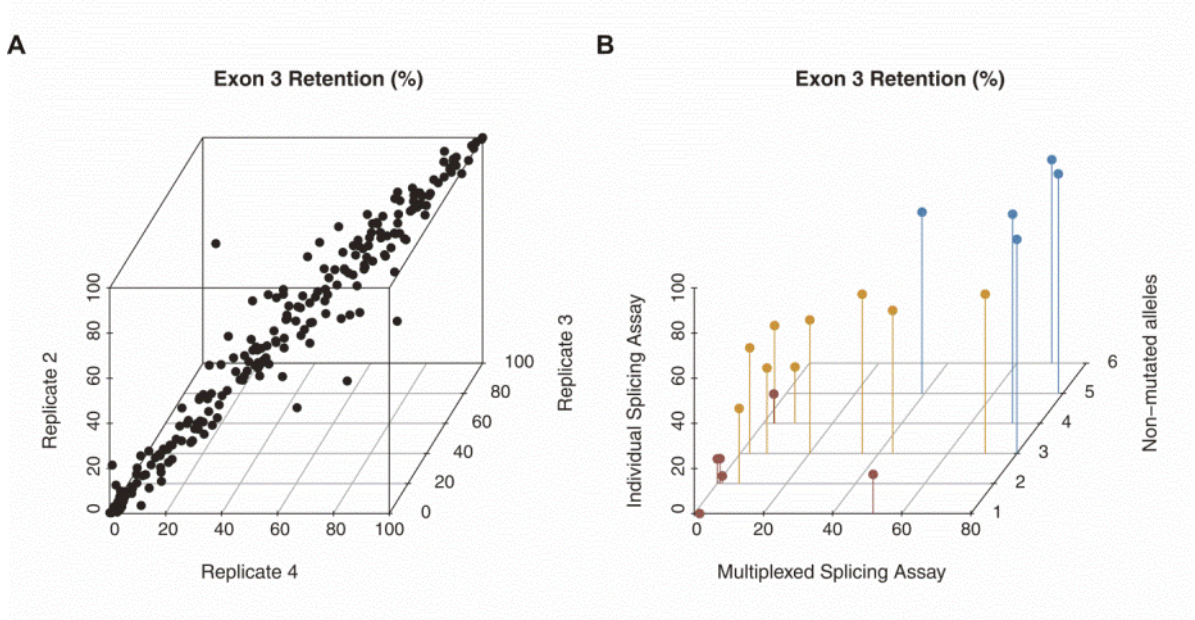


Figure 2. Correlation analysis of exon 3 retention levels measured (figure legend continues in next page). A) Mean exon 3 retention (%) of three biological replicates of the Output Library for all uniquely tagged targeted haplotypes. For all four biological replicates the Spearman's rank correlation coefficient rho ranges from 0.89 to 0.97; p-value < 2.2e-16 (Supp. Figure S5A). **B)** Correlation between the mean % of exon 3 retention measured individually and by the parallelized splicing assay for 20 control barcoded-haplotypes entailing 2-7 variants. Haplotype classification based on exon 3 retention (%) measured by the individual splicing assay (Buena-Atienza et al., 2016) into rather deleterious effect (<20%; red), intermediate (20-70%; orange) and minor effect (>70%; blue) for “S”, “I” and “M” (Supp. Table S4; Supp. Fig. S5B), respectively.

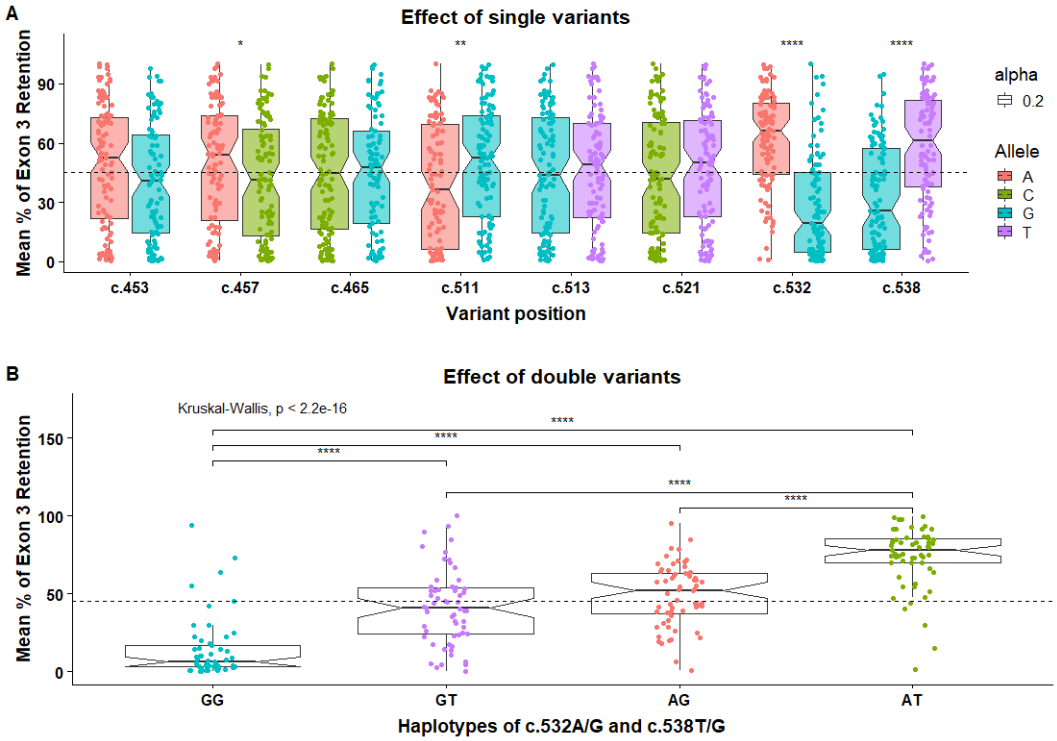


Figure 3. Effect of exonic variants and haplotypes on *OPN1LW/MW* exon 3 retention. The mean % of exon 3 retention of all uniquely tagged haplotypes was compared **A)** for each of the 8 common exonic variants within the haplotype (c.453, c.457, c.466, c.511, c.513, c.521, c.532 and c.538) according to the allele carried (i.e. 128 haplotypes are expected to carry c.453A and 128 haplotypes c.453G). The variants c.532G and c.538G individually resulted in a highly significant change of the mean % exon 3 retention in comparison to the counterpart alleles c.532A and c.538T with the overall haplotype backgrounds; Wilcoxon test, p = <2e-16 and p = 6.7e-10, respectively (Supp. Table S6), **B)** for all haplotypes grouped by the four possible combinations attending to the combination of phased alleles in c.532A and c.538G sites of exon 3 referred to as "AG" (c.532A and c.538G), "AT" (c.532A and c.538T), "GG" (c.532G and c.538G) and "GT" (c.532G and c.538T), being the mean % of exon 3 retention significantly reduced for haplotypes harbouring both the c.532G and c.538G alleles ("GG"); Wilcoxon test, p = 1.0e-11 and significantly increased for haplotypes entailing the combination of c.532A and c.538T alleles ("AT"); Wilcoxon test, p = 6.6e-11, in comparison to the mean of all haplotypes. Kruskall Wallis test and the post-hoc Dunn test for pairwise comparison of haplotype group means was performed with Bonferrini correction (not significant is not shown; *, p-value <0.05; **, p-value <0.01; ***, p-value <0.001: ****, p-value <0.0001), resulting in significantly different means for all comparisons except of the "GT" (c.532G and c.538T) and "AT" (c.532A and c.538T), having only one of the c.532G or c.538G variants leads to the same mean of exon 3 retained levels.

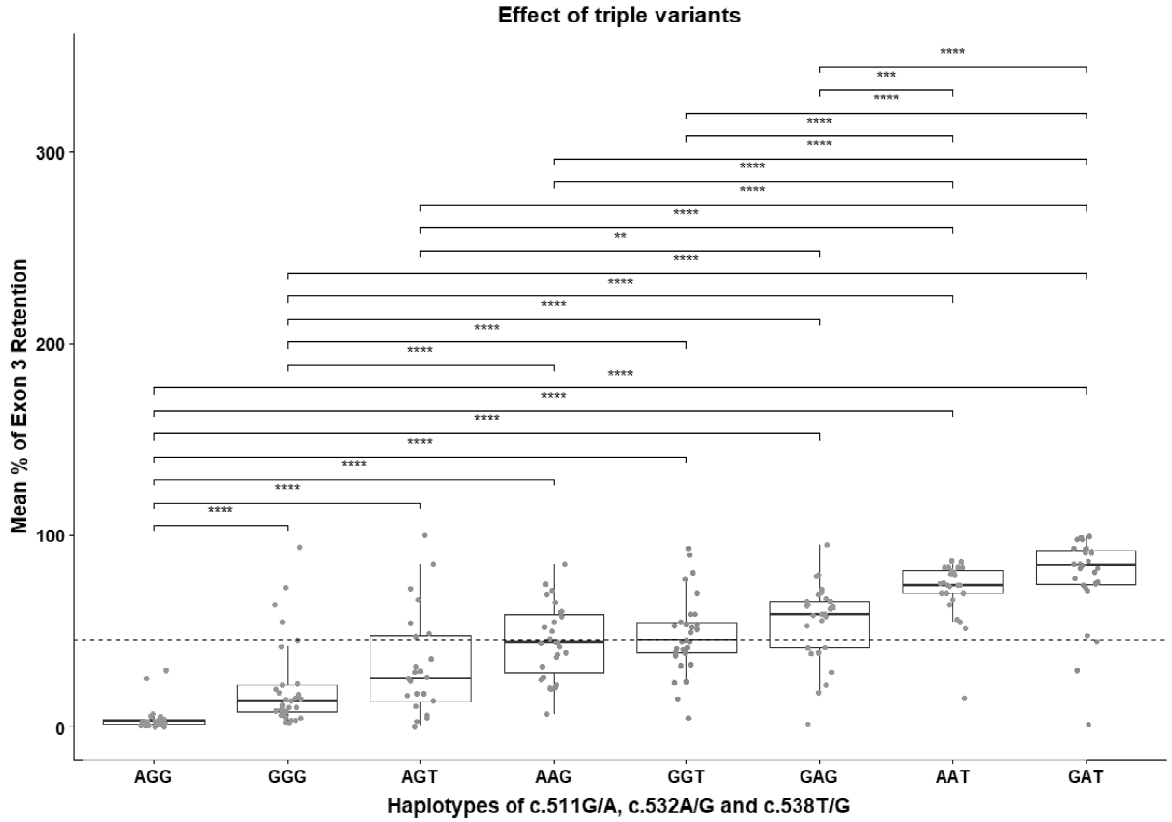


Figure 4. Effect of triple combinations of exonic variants on *OPN1LW/MW* exon 3 retention. Haplotypes were grouped by three variants, c.511G/A, c.532A/G and c.538T/G, as c. 511G>A showed the third highest difference for exon 3 retention in Fig.3. All 28 pairwise comparisons of the mean of the 8 haplotype groups (based on the c.511G/A, c.532A/G and c.538T/G variants) were performed with Wilcoxon test and p-values adjusted with Bonferroni correction (not significant is not shown; *, p-value <0.05; **, p-value <0.01; ***, p-value <0.001: ****, p-value <0.0001). Haplotypes carrying the c.511A in phase with the c.532G and c.538G result in an overall significant reduction of exon 3 retention compared to those with c.511G (AGG *versus* GGG) whereas c.511G>A in phase with any of the other three combinations of c.532A>G and c.538T>G does not result in a significant reduction (AAT *versus* GAT, AAG *versus* GAG and AGT *versus* GGT). The variant c.511G>A has a significant effect when the other two G mutant alleles are present in the haplotype.

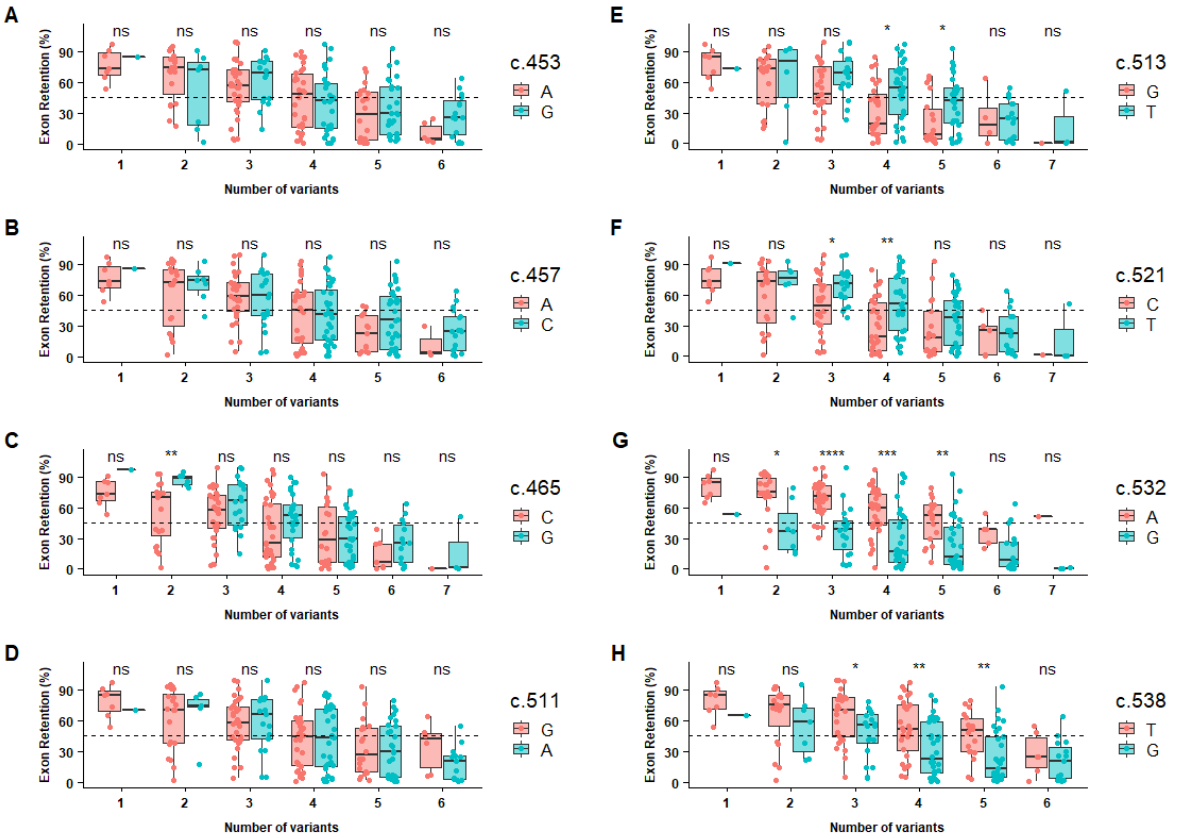


Figure 5. Effect of the accumulation of exonic variants in *OPN1LW/MW* exon 3 retention. The mean % of exon 3 retention levels is plotted for haplotype groups entailing from 1 up to 7 variants with respect to the "Reference" haplotype ("Number of additional variants"). For each of the 8 variants at c.453, c.457, c.466, c.511, c.513, c.521, c.532 and c.538 (**A-H**), values for each allele are represented side-by-side; variant allele (turquoise) or ancestral allele (tomato red). Overall mean (%) exon 3 retention is shown as a stripped line. Note that for some of the tests significance levels were not significant mainly due to the low number of samples. Wilcoxon test was performed to compare the distribution of the means (ns, not significant; *, p-value <0.05; **, p-value <0.01; ***, p-value <0.001: ****, p-value <0.0001).

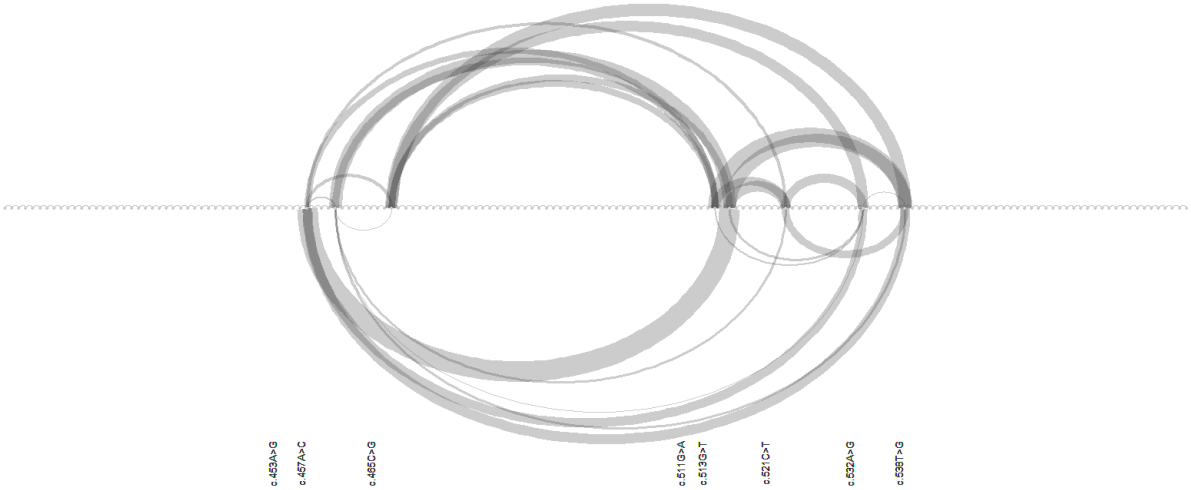


Figure 6. Epistatic and additive interactions from the double variant analysis. Each nucleotide of exon 3 is represented by a node (total of 169 bp). Only interactions within the 8 nucleotides of the haplotype were assessed. Positive and negative epistasis are shown by arc edges above and below the exon 3 plane, respectively. Additive interactions are depicted below. The width of the arc line depicts proportionally the absolute value of the median difference between observed and expected exon skipping values over the four biological replicates (Supp. Table S7).

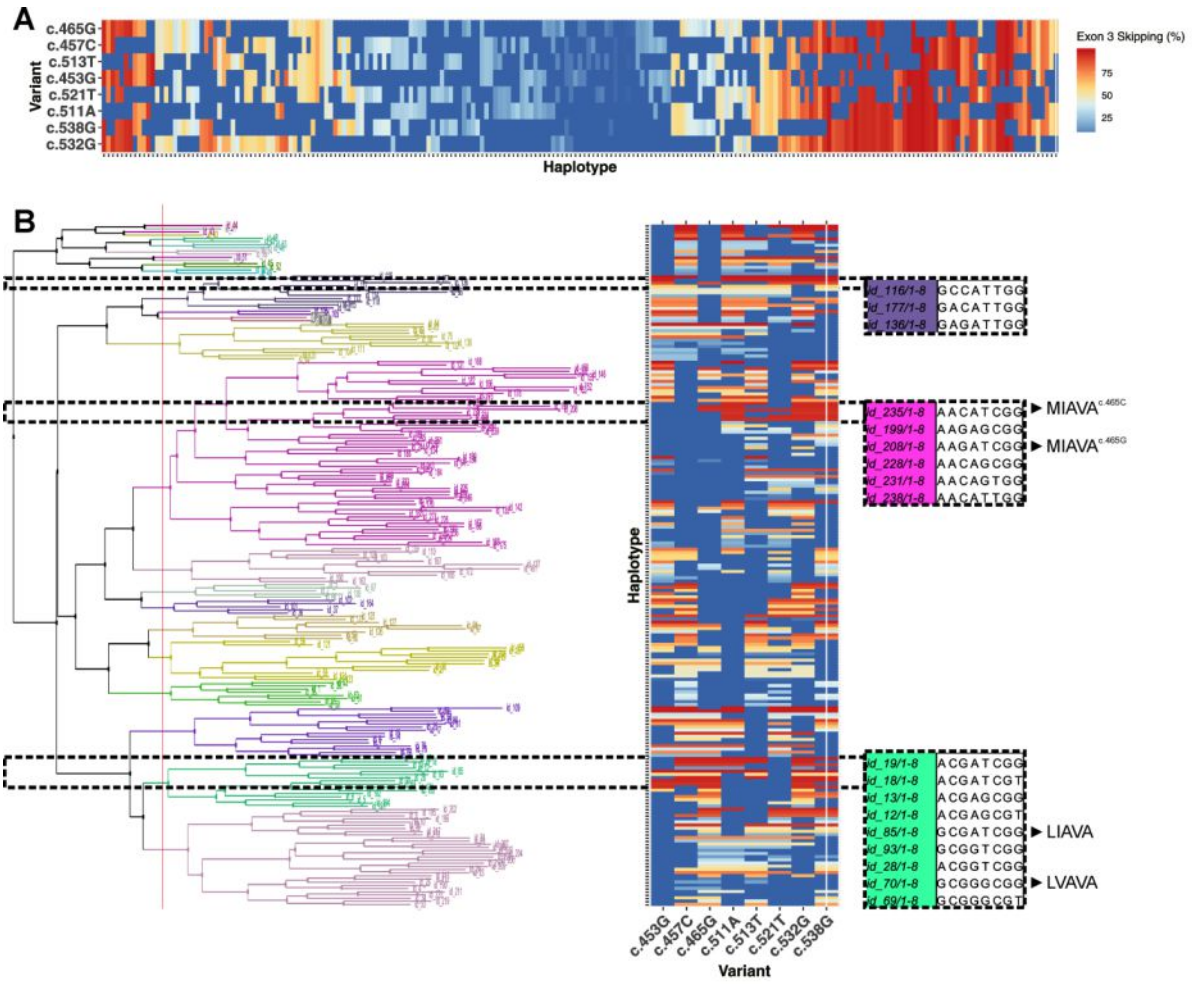


Figure 7. Clustering of haplotypes according to variants and splicing outcome. **A)** Heatmap showing the degree of exon 3 skipping (mean % colour-coded in the legend; 'red' indicates high exon skipping levels and 'blue' low exon skipping levels) induced by all haplotypes analysed. Haplotypes are clustered for the 8 variants entailed and the % of induced exon 3 skipping. For those variant positions where the haplotype harbours the allele of the Reference haplotype, the position is depicted in dark blue. Clustering was performed by ordering the haplotypes (x-axis, columns), the variant profile of exonic variants comprised in the haplotype of exon 3 (y-axis, rows) according to the level of exon 3 skipping (colour code) from the parallelized minigene splicing assay. **B)** Alignment-based tree clustering of haplotypes. A multiple sequence alignment of the 8-bp sequences (variants in the haplotype) was performed with MEGA5 (Tamura et al., 2011) and used as input to generate a tree with Jalview (Waterhouse et al., 2009). Phyloclustering was performed by rooting the tree and clusters were thereafter coloured. The haplotypes (y-axis, rows) in the heatmap of Fig.7B were ordered accordingly to the alignment after phyloclustering. Clusters with high sequence identity and inducing high exon skipping rates are squared and sequence is shown. The known haplotypes 'LIAVA' and 'LVAVA' clustered together for high levels of aberrant splicing. The 'MIAVA' haplotype with c.465C or c.465G were found in the same cluster. Hierarchical clustering for the heatmaps in A) and B) was performed with hclust() and ggplot() in RStudio.

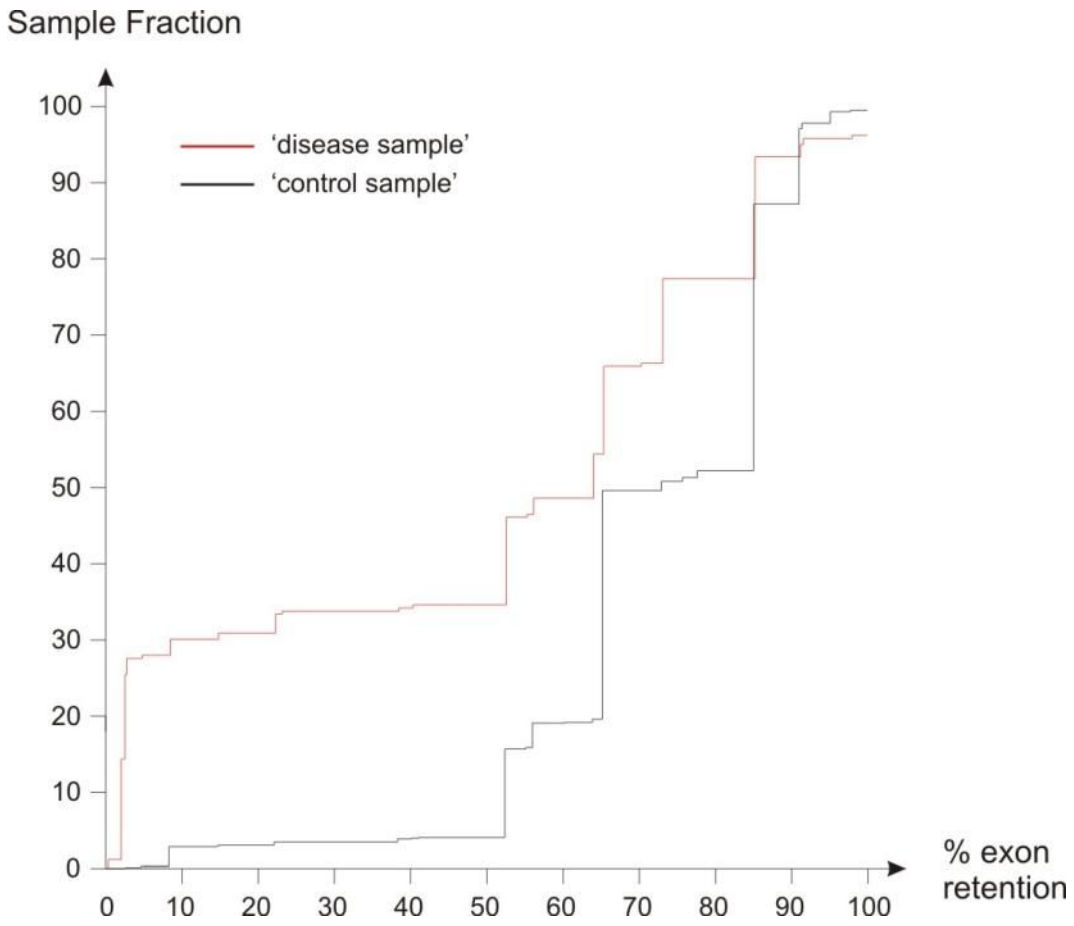


Figure 8. Kaplan-Meier plot of the haplotype fraction and exon retention in the human 'disease group' and 'control group'. The cumulative percentage of observed haplotypes in the 'disease group' and the 'control group' is plotted against the level of exon retention for such haplotypes.

Supplementary Materials and Methods

Computation and analysis of data derived from short- and long-read sequencing of Input and Output libraries.

Quality of CCS reads was assessed with SMRT Link (version 5.0.1.9585). Raw bam2bam_barcode files barcoding all input subreads were converted into CCS reads by calling the consensus sequences from subreads. Then standard fasta/fastq files were generated from bam2fasta and bam2fastq. Quality from short-reads was assessed by FastQC (version 0.11.4), Cutadatapt (version 1.8.3) removed unwanted adapter sequences and performed quality based filtering and Tophat (version 2.2.3.0) was used for mapping. All reads not meeting a quality threshold of Q=30 were removed and up to 7 mismatches were allowed for mapping. Tophat Alignment bam files, the PacBio assembly reference fasta and a custom annotation file (in the gtf file format) were used in the R language (version 3.4.0) to count raw read counts using the function featureCounts() from the Rsubread package (version 1.26.1) on each of the barcode that is linked to an haplotype, allowing for multi-mapped and multi-assigned reads. For each CCS read an alignment to the transcript reference sequence including exon 3 and relevant exonic junctions with exon 2 and exon 4 was calculated. The distance score, an Euclidean distance calculated as the usual distance between the two vectors, $\mathrm{sqrt}(\mathrm{sum}((x_i - y_i)^2))$, derived from the alignment can be used as an indicator of whether the sequence entailed an exon 3-retained or -skipped transcript. The threshold of 0.2 to discern between retention or skipping of exon 3 was included in the alignment file of short reads derived from Output Library mapped against the reference sequence built from the Output Library PacBio data. In this sense we could separate Illumina reads according to whether they mapped rather to CCS reads with a distance score indicating exon retention or pointing to aberrant splicing.

All single 8-bp barcodes found in CCS reads of the Output Library were extracted. For each barcode, the following information was listed: size length of the amplified amplicon, the sequence, whether the sequence belongs to a haplotype group in exon 3 based on the definition of the Input Library, the number of variants with respect to exon 3 reference and the distance score. We found 38,221 barcodes listed, from which 1863 were associated with a haplotype different or equal to the Reference haplotypes of the Input Library, of which 1819 were haplotypes different from the Reference haplotype. The rest do not represent only non-targeted haplotypes and mixed haplotypes, but also sequences bearing exon 3-skipped amplicons which do bear the same

unique barcode found in the Input Library. Thus, barcodes were grouped by means of an identifier ("id_xy") which contained all reads with the same unique barcode also sequenced in the Input Library even if no grouping for a haplotype sequence in exon 3 was available for the Output Library. We found a total of 248 id's being 1 the Reference haplotype and 2 the Control haplotypes from minigene constructs transfected mixed with the Input Library. All sequences for each id were classified as either "exon 3 retention" or "exon 3 skipping" based on the distance score (threshold=0.2) and the mapped short reads to the built reference. When counts for "exon 3 retention" and " exon 3 skipping" were combined for each id, a total of 232 id's (uniquely barcoded-haplotypes) were observed having both types of amplicons from differentially spliced transcripts. A total of 14 id's (uniquely barcoded haplotypes) were found associated with amplicons exclusively coming from one spliced outcome. A joint table with relative distance scores and relative ratios for exon 3 retention were calculated for each haplotype. A mean ratio and standard deviation was calculated from counts for the four replicates. Control barcoded-haplotypes (Reference and 'LIAVA') transfected mixed with the Input Library were manually added to the analysis and relative distance scores were at the extreme sides of the distribution. Upon confirmation of expected results from control haplotypes, these were not pooled with the rest of the haplotypes for downstream analysis.

Supplementary Tables

Supplementary Table S1. Primer sequences for long-read sequencing of the Output Library

Name	Sequence (5' to 3')
Exon2_0012_Forward	GCGCACGCACTACAGAAGCATTGTGAACCAGGTCTCTGGCTA
Exon2_0016_Forward	TAGATGCGAGAGTAGAAGCATTGTGAACCAGGTCTCTGGCTA
Exon2_0005_Forward	ACACGCATGACACACTAGCATTGTGAACCAGGTCTCTGGCTA
Exon2_0032_Forward	TATCTCTGTAGAGTCTAGCATTGTGAACCAGGTCTCTGGCTA
Barc_mini_0012_Reverse	TCTGTAGTGCGTGCGCATGCCACCCAGTAGCTCTAAGTCGA
Barc_mini_0016_Reverse	TCTACTCTCGCATCTAATGCCACCCAGTAGCTCTAAGTCGA
Barc_mini_0005_Reverse	AGTGTGTCATGCGTGTATGCCACCCAGTAGCTCTAAGTCGA
Barc_mini_0032_Reverse	AGACTCTACAGAGATAATGCCACCCAGTAGCTCTAAGTCGA

Supplementary Table S2. Primers for short-read sequencing of the Output Library

PCR	Name	Sequence (5' to 3')
First	Rd1_4S_Ex2_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGCATTGTGAACCAGGTCTCTGGCT*A
	Rd1_7S_Ex2_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCTAGCATTGTGAACCAGGTCTCTGGCT*A
	Rd1_10S_Ex2_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTACAGAGCATTGTGAACCAGGTCTCTGGCT*A
	Rd1_12S_Ex2_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTAGTCATAAGCATTGTGAACCAGGTCTCTGGCT*A
	Rd1_4S_BC_R	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTAGCTCTAAGTCGACTCACTATAGATTAA*C
	Rd1_6S_BC_R	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTTAGCTCTAAGTCGACTCACTATAGATTAA*C
	Rd1_9S_BC_R	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGCTCTAGCTCTAAGTCGACTCACTATAGATTAA*C
	Rd1_12S_BC_R	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGCAGAAGTTAGCTCTAAGTCGACTCACTATAGATTAA*C
	Rd2_12S_BC_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNGCAGAAGTTAGCTCTAAGTCGACTCACTATAGATTAA*C
	Rd2_9S_BC_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNAGCTCTAGCTCTAAGTCGACTCACTATAGATTAA*C
	Rd2_6S_BC_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNCTTAGCTCTAAGTCGACTCACTATAGATTAA*C
	Rd2_4S_BC_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNTAGCTCTAAGTCGACTCACTATAGATTAA*C
	Rd2_12S_Ex2_F	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNTAGTCATAAGCATTGTGAACCAGGTCTCTGGCT*A
	Rd2_10S_Ex2_F	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNCTACAGAGCATTGTGAACCAGGTCTCTGGCT*A
	Rd2_7S_Ex2_F	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNGCTAGCATTGTGAACCAGGTCTCTGGCT*A
	Rd2_4S_Ex2_F	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNAGCATTGTGAACCAGGTCTCTGGCT*A
Second	P5-PE-Rd1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
	P7-701-Rd2	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
	P7-702-Rd2	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
	P7-703-Rd2	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T
	P7-704-Rd2	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T

Supplementary Table S3. Mean ratio and average distance scores for each unique haplotype

Identifier	Exon 3 Haplotype	Counts Replicate 1	Counts Replicate 2	Counts Replicate 3	Counts Replicate 4	Mean (%) exon 3 retention	Standard deviation	Total CCS Reads	Relative distance score
id_127	c.453A>G; c.457A>C; c.513G>T; c.538T>G	5987	6632	9968	11568	1.01	0.53	277	0.51
id_19	c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	574	391	757	844	1.46	1.38	71	0.51
Control 2	LIAVA	562	1041	645	1659	2.48	1.17	57	0.50
id_109	c.453A>G; c.457A>C; c.511G>A; c.521C>T; c.532A>G; c.538T>G	7925	5180	10042	7621	0.80	0.22	97	0.50
id_44	c.457A>C; c.511G>A; c.521C>T; c.532A>G; c.538T>G	1192	1886	1518	3940	0.51	0.74	135	0.50
id_134	c.453A>G; c.465C>G; c.532A>G; c.538T>G	415	368	103	354	3.48	1.81	40	0.49
id_85	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	403	597	493	798	1.58	0.96	57	0.49
id_80	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.521C>T; c.532A>G; c.538T>G	371	1132	1026	1284	0.96	0.62	37	0.49
id_48	c.457A>C; c.511G>A; c.513G>T; c.532A>G; c.538T>G	714	949	917	1228	0.85	1.13	60	0.49
id_168	c.453A>G; c.511G>A; c.532A>G; c.538T>G	1855	2568	2496	3816	0.33	0.40	74	0.49
id_41	c.457A>C; c.511G>A; c.532A>G; c.538T>G	5343	5654	12987	13404	1.02	0.59	201	0.49
id_13	c.457A>C; c.465C>G; c.511G>A; c.532A>G; c.538T>G	201	218	173	332	5.85	9.27	46	0.49
id_70	c.453A>G; c.457A>C; c.465C>G; c.532A>G; c.538T>G	219	287	206	533	2.54	0.91	45	0.49
id_180	c.453A>G; c.513G>T	5744	9518	7832	13657	1.28	0.46	422	0.49
id_171	c.453A>G; c.511G>A; c.521C>T; c.532A>G; c.538T>G	351	654	392	748	3.82	1.26	62	0.49
id_177	c.453A>G; c.511G>A; c.513G>T; c.521C>T; c.532A>G; c.538T>G	1002	1377	902	2426	1.81	1.11	60	0.49
id_202	c.465C>G; c.511G>A; c.521C>T; c.532A>G; c.538T>G	616	587	253	495	3.10	0.91	54	0.49
id_84	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G	7240	4180	9693	8583	0.34	0.19	44	0.49
id_116	c.453A>G; c.457A>C; c.511G>A; c.513G>T; c.521C>T; c.532A>G; c.538T>G	6851	4536	9162	6759	0.21	0.10	65	0.49
id_28	c.457A>C; c.465C>G; c.513G>T; c.532A>G; c.538T>G	350	583	619	1130	1.89	1.17	57	0.48
id_199	c.465C>G; c.511G>A; c.532A>G; c.538T>G	130	12	94	71	3.01	3.67	26	0.48
id_3	c.457A>C; c.532A>G; c.538T>G	2589	2462	2867	3483	3.64	0.13	195	0.48
id_185	c.453A>G; c.513G>T; c.521C>T; c.532A>G; c.538T>G;	152	197	121	127	8.50	3.28	24	0.48
id_125	c.453A>G; c.457A>C; c.513G>T; c.521C>T; c.532A>G; c.538T>G	831	1558	821	1644	6.33	0.51	86	0.48
Id_235	c.511G>A; c.513G>T; c.532A>G; c.538T>G	901	1290	901	1708	2.58	0.23	88	0.48
id_231	c.511G>A; c.521C>T; c.532A>G; c.538T>G	408	1090	1123	1198	3.24	1.48	73	0.47
id_208	c.465C>G; c.511G>A; c.513G>T; c.521C>T; c.532A>G; c.538T>G	805	1469	998	1765	3.97	1.28	62	0.47
id_173	c.453A>G; c.511G>A; c.513G>T; c.532A>G	168	321	249	380	6.04	2.45	26	0.47
id_182	c.453A>G; c.513G>T; c.532A>G; c.538T>G	538	655	654	1063	15.06	1.80	88	0.47
id_67	c.453A>G; c.457A>C; c.532A>G; c.538T>G	568	1031	877	1411	9.14	3.32	84	0.47
id_12	c.457A>C; c.465C>G; c.511G>A; c.532A>G	1782	2164	1766	4367	84.74	2.21	52	0.47
id_91	c.453A>G; c.457A>C; c.465C>G; c.513G>T	512	687	596	919	43.91	12.31	28	0.47
id_106	c.453A>G; c.457A>C; c.511G>A; c.532A>G; c.538T>G	60	89	37	128	6.63	4.88	12	0.47
id_156	c.453A>G; c.465C>G; c.513G>T; c.532A>G; c.538T>G	197	295	366	1229	4.59	1.93	33	0.47
id_228	c.511G>A; c.532A>G; c.538T>G	243	334	258	437	5.12	3.31	54	0.47
id_238	c.511G>A; c.513G>T; c.521C>T; c.532A>G; c.538T>G	957	1352	877	1957	4.21	1.46	81	0.46
id_60	c.457A>C; c.513G>T; c.521C>T; c.532A>G; c.538T>G	567	829	675	850	22.62	6.17	73	0.46
id_170	c.453A>G; c.511G>A; c.521C>T; c.532A>G	70	111	82	268	17.27	2.61	9	0.46
id_51	c.457A>C; c.511G>A; c.513G>T; c.521C>T; c.532A>G; c.538T>G	1277	1581	1174	1794	3.01	0.85	96	0.46
id_214	c.465C>G; c.513G>T; c.532A>G; c.538T>G	678	297	220	541	29.57	35.09	38	0.46
id_31	c.457A>C; c.465C>G; c.513G>T; c.521C>T; c.532A>G; c.538T>G	514	1108	1191	1759	5.32	2.33	42	0.46
id_57	c.457A>C; c.513G>T; c.532A>G; c.538T>G	225	351	261	549	11.43	1.85	59	0.46

Identifier	Exon 3 Haplotype	Counts Replicate 1	Counts Replicate 2	Counts Replicate 3	Counts Replicate 4	Mean (%) exon 3 retention	Standard deviation	Total CCS Reads	Relative distance score
id_102	c.453A>G; c.457A>C; c.521C>T; c.532A>G; c.538T>G	1052	1408	896	1650	7.71	2.32	95	0.46
id_92	c.453A>G; c.457A>C; c.465C>G; c.513G>T; c.532A>G	2136	2317	2496	3450	76.90	4.29	77	0.46
id_152	c.453A>G; c.465C>G; c.511G>A; c.513G>T; c.538T>G	201	274	197	281	22.09	3.84	15	0.45
id_98	c.453A>G; c.457A>C; c.465C>G; c.513G>T; c.538T>G	29	4	71	98	17.79	14.09	10	0.45
id_37	c.457A>C; c.521C>T; c.532A>G; c.538T>G	653	849	710	1142	9.87	1.40	105	0.45
id_192	c.465C>G; c.532A>G; c.538T>G	156	258	112	232	14.59	4.31	38	0.45
id_90	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.538T>G	133	56	95	211	6.53	5.01	21	0.45
id_113	c.453A>G; c.457A>C; c.511G>A; c.513G>T; c.532A>G; c.538T>G	276	390	433	722	25.08	9.65	61	0.45
id_9	c.457A>C; c.465C>G; c.521C>T; c.532A>G; c.538T>G	461	479	595	921	13.30	3.74	68	0.45
id_130	c.453A>G; c.532A>G	945	1775	1045	2054	14.59	2.47	167	0.45
id_131	c.453A>G; c.532A>G; c.538T>G	428	527	653	692	14.16	3.02	57	0.45
id_136	c.453A>G; c.465C>G; c.521C>T; c.532A>G; c.538T>G	497	1096	957	1188	10.08	4.43	48	0.45
id_15	c.457A>C; c.465C>G; c.511G>A; c.521C>T; c.532A>G	511	984	843	1320	4.59	2.42	38	0.45
id_73	c.453A>G; c.457A>C; c.465C>G; c.521C>T; c.532A>G; c.538T>G	4847	10556	8123	16618	63.84	5.51	102	0.45
id_72	c.453A>G; c.457A>C; c.465C>G; c.521C>T; c.532A>G	101	71	92	99	22.97	1.49	11	0.45
id_6	c.457A>C; c.465C>G; c.532A>G; c.538T>G	2248	2964	2250	4482	19.56	3.51	113	0.44
id_79	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.521C>T; c.532A>G	113	145	67	303	10.69	2.82	17	0.44
id_40	c.457A>C; c.511G>A; c.532A>G	720	3432	2241	4096	4.83	3.06	74	0.44
id_105	c.453A>G; c.457A>C; c.511G>A; c.532A>G	90	151	273	478	13.59	2.56	28	0.44
id_227	c.511G>A; c.532A>G	25	151	49	33	17.27	13.44	10	0.43
id_248	c.513G>T; c.521C>T; c.532A>G; c.538T>G	283	163	167	316	16.73	2.48	30	0.43
id_164	c.453A>G; c.521C>T; c.532A>G; c.538T>G	272	292	219	500	18.15	2.64	50	0.43
id_18	c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G	652	1465	819	2141	2.72	0.96	29	0.43
id_24	c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.538T>G	60	38	47	44	44.05	7.98	11	0.43
id_119	c.453A>G; c.457A>C; c.511G>A; c.538T>G	271	589	388	982	19.47	2.21	57	0.43
id_118	c.453A>G; c.457A>C; c.511G>A; c.513G>T; c.538T>G	171	154	260	453	20.09	4.38	35	0.43
id_108	c.453A>G; c.457A>C; c.511G>A; c.521C>T; c.532A>G	770	1409	1057	2132	66.60	8.08	78	0.43
id_21	c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.521C>T; c.532A>G	66	152	113	144	24.98	8.15	17	0.43
id_97	c.453A>G; c.457A>C; c.465C>G; c.513G>T; c.521C>T; c.538T>G	233	244	163	300	38.61	4.38	18	0.42
id_47	c.457A>C; c.511G>A; c.513G>T; c.532A>G	1879	1842	2259	2882	71.88	8.52	33	0.42
id_94	c.453A>G; c.457A>C; c.465C>G; c.513G>T; c.521C>T	508	830	579	1717	29.47	7.33	51	0.42
id_115	c.453A>G; c.457A>C; c.511G>A; c.513G>T; c.521C>T; c.532A>G	237	433	270	731	24.31	2.38	41	0.42
id_189	c.532A>G; c.538T>G	2087	2550	1761	3091	22.10	2.37	165	0.42
id_43	c.457A>C; c.511G>A; c.521C>T; c.532A>G	378	541	464	827	16.22	4.04	67	0.42
id_167	c.453A>G; c.511G>A; c.532A>G	162	242	172	378	31.08	4.60	28	0.42
id_33	c.457A>C; c.465C>G; c.513G>T; c.538T>G	207	153	80	205	28.32	6.37	11	0.41
id_198	c.465C>G; c.511G>A; c.532A>G	3	54	6	41	52.26	42.43	7	0.41
id_179	c.453A>G; c.511G>A; c.538T>G	46	61	3	96	42.67	38.94	9	0.41
id_195	c.465C>G; c.521C>T; c.532A>G; c.538T>G	506	909	377	1503	8.29	2.77	36	0.41
id_25	c.457A>C; c.465C>G; c.511G>A; c.538T>G	38	38	23	99	37.68	4.83	6	0.41
id_122	c.453A>G; c.457A>C; c.513G>T; c.532A>G; c.538T>G	7120	7927	11341	11033	93.87	1.84	89	0.40
id_153	c.453A>G; c.465C>G; c.511G>A; c.538T>G	164	122	244	265	43.60	4.86	27	0.40
id_45	c.457A>C; c.511G>A; c.521C>T; c.538T>G	300	584	245	980	24.86	1.02	21	0.40
id_54	c.457A>C; c.511G>A; c.538T>G	103	75	107	25	31.08	8.69	11	0.40
id_69	c.453A>G; c.457A>C; c.465C>G; c.532A>G	830	3801	1930	4044	4.61	2.07	33	0.40
id_245	c.513G>T; c.532A>G; c.538T>G	2678	2855	2663	3803	54.92	9.22	81	0.39

Identifier	Exon 3 Haplotype	Counts Replicate 1	Counts Replicate 2	Counts Replicate 3	Counts Replicate 4	Mean (%) exon 3 retention	Standard deviation	Total CCS Reads	Relative distance score
id_147	c.453A>G; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	240	297	374	635	29.80	3.47	31	0.39
id_142	c.453A>G; c.465C>G; c.511G>A; c.521C>T; c.532A>G	54	78	104	61	28.31	8.89	10	0.39
id_146	c.453A>G; c.465C>G; c.511G>A; c.513G>T; c.532A>G	133	62	82	237	29.35	3.52	10	0.39
id_17	c.457A>C; c.465C>G; c.511G>A; c.521C>T; c.538T>G	188	570	401	712	59.68	9.24	29	0.39
id_56	c.457A>C; c.513G>T; c.532A>G	948	920	855	1663	23.48	2.30	64	0.39
id_81	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.521C>T; c.538T>G	217	171	319	417	25.88	3.85	31	0.39
id_133	c.453A>G; c.465C>G; c.532A>G	318	273	306	773	39.22	5.91	41	0.38
id_50	c.457A>C; c.511G>A; c.513G>T; c.521C>T; c.532A>G	184	264	303	268	35.24	7.97	31	0.38
id_217	c.465C>G; c.513G>T; c.521C>T; c.532A>G; c.538T>G	181	364	251	513	42.01	4.88	42	0.38
id_224	c.521C>T; c.532A>G; c.538T>G	3214	3798	3462	5593	72.90	4.23	133	0.38
id_121	c.453A>G; c.457A>C; c.513G>T; c.532A>G	330	359	292	523	31.90	3.67	37	0.38
id_110	c.453A>G; c.457A>C; c.511G>A; c.521C>T; c.538T>G	81	56	150	158	36.18	11.09	12	0.38
id_237	c.511G>A; c.513G>T; c.521C>T; c.532A>G	95	264	286	427	25.96	4.57	19	0.38
id_5	c.457A>C; c.465C>G; c.532A>G	97	29	72	222	41.05	4.46	14	0.37
id_213	c.465C>G; c.513G>T; c.532A>G	7	43	15	34	32.48	14.74	5	0.37
id_181	c.453A>G; c.513G>T; c.532A>G	30	2	28	27	22.37	24.20	9	0.37
id_135	c.453A>G; c.465C>G; c.521C>T; c.532A>G	23	55	28	86	58.61	8.40	6	0.37
id_75	c.453A>G; c.457A>C; c.465C>G; c.511G>A	611	701	559	907	14.90	4.05	20	0.37
id_2	c.457A>C; c.532A>G	417	506	441	909	38.81	3.36	46	0.37
id_176	c.453A>G; c.511G>A; c.513G>T; c.521C>T; c.532A>G	297	167	210	492	51.98	20.14	33	0.37
id_59	c.457A>C; c.513G>T; c.521C>T; c.532A>G	709	683	547	1187	50.92	3.49	49	0.36
id_95	c.453A>G; c.457A>C; c.465C>G; c.513G>T; c.521C>T; c.532A>G	226	418	238	401	48.92	8.11	30	0.36
id_66	c.453A>G; c.457A>C; c.532A>G	175	236	212	326	44.76	8.11	29	0.36
id_88	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.521C>T; c.538T>G	77	90	70	55	51.75	9.47	9	0.36
id_52	c.457A>C; c.511G>A; c.513G>T; c.521C>T; c.538T>G	1013	1443	1177	1808	70.83	4.49	49	0.36
id_117	c.453A>G; c.457A>C; c.511G>A; c.513G>T; c.521C>T; c.538T>G	181	270	299	347	38.74	5.96	29	0.36
id_103	c.453A>G; c.457A>C; c.521C>T; c.538T>G	252	431	384	805	41.18	1.64	32	0.35
id_23	c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.521C>T; c.538T>G	109	253	434	759	20.55	8.93	14	0.35
id_209	c.465C>G; c.511G>A; c.513G>T; c.521C>T; c.538T>G	354	380	564	1418	49.92	5.60	25	0.35
id_128	c.453A>G; c.457A>C; c.538T>G	2996	3826	3983	7380	41.08	13.75	95	0.35
id_101	c.453A>G; c.457A>C; c.521C>T; c.532A>G	1025	1195	1101	1481	41.60	2.22	80	0.35
id_158	c.453A>G; c.465C>G; c.513G>T; c.521C>T; c.532A>G	74	103	97	341	40.42	5.76	11	0.35
id_207	c.465C>G; c.511G>A; c.513G>T; c.521C>T; c.532A>G	229	292	343	660	48.83	5.59	33	0.35
id_112	c.453A>G; c.457A>C; c.511G>A; c.513G>T; c.532A>G	212	279	305	580	54.39	9.80	30	0.34
id_124	c.453A>G; c.457A>C; c.513G>T; c.521C>T; c.532A>G	1230	1178	918	1965	53.51	7.82	93	0.34
id_99	c.453A>G; c.457A>C; c.465C>G; c.538T>G	1046	782	1597	1125	52.37	29.22	45	0.34
id_8	c.457A>C; c.465C>G; c.521C>T; c.532A>G	481	764	729	1313	45.37	5.11	70	0.34
id_178	c.453A>G; c.511G>A; c.513G>T; c.538T>G	85	74	85	110	45.72	12.80	12	0.33
id_234	c.511G>A; c.513G>T; c.532A>G	425	566	460	761	100.00	0.00	7	0.33
id_62	c.457A>C; c.513G>T; c.538T>G	396	335	499	663	57.92	8.87	23	0.33
id_172	c.453A>G; c.511G>A; c.521C>T; c.538T>G	343	361	420	856	44.17	2.57	43	0.33
id_74	c.453A>G; c.457A>C; c.465C>G; c.521C>T; c.538T>G	193	131	343	334	55.11	8.50	32	0.33
id_86	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.521C>T	149	215	185	522	54.56	4.46	13	0.32
id_155	c.453A>G; c.465C>G; c.513G>T; c.532A>G	47	170	142	185	54.63	4.23	18	0.32
id_83	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T	258	321	335	395	55.98	7.96	23	0.32
id_163	c.453A>G; c.521C>T; c.532A>G	671	364	553	1212	44.64	2.05	38	0.32

Identifier	Exon 3 Haplotype	Counts Replicate 1	Counts Replicate 2	Counts Replicate 3	Counts Replicate 4	Mean (%) exon 3 retention	Standard deviation	Total CCS Reads	Relative distance score
id_230	c.511G>A; c.521C>T; c.532A>G	375	943	456	979	46.84	0.42	50	0.32
id_232	c.511G>A; c.521C>T; c.538T>G	1041	1312	1207	2119	57.37	3.54	89	0.31
id_30	c.457A>C; c.465C>G; c.513G>T; c.521C>T; c.532A>G	143	118	132	98	51.71	5.68	13	0.31
id_149	c.453A>G; c.465C>G; c.511G>A; c.513G>T; c.521C>T; c.532A>G	92	152	29	117	43.68	22.37	9	0.31
id_32	c.457A>C; c.465C>G; c.513G>T; c.521C>T; c.538T>G	159	195	168	302	52.50	8.20	24	0.31
id_210	c.465C>G; c.511G>A; c.513G>T; c.538T>G	168	270	285	590	60.24	2.55	27	0.31
id_144	c.453A>G; c.465C>G; c.511G>A; c.521C>T; c.538T>G	770	767	749	1363	28.39	15.21	28	0.31
id_27	c.457A>C; c.465C>G; c.513G>T; c.532A>G	245	231	333	358	52.81	2.78	18	0.30
id_157	c.453A>G; c.465C>G; c.513G>T; c.521C>T	249	97	281	390	47.69	11.14	13	0.30
id_36	c.457A>C; c.521C>T; c.532A>G	1044	1872	1592	3404	38.52	3.92	57	0.30
id_239	c.511G>A; c.513G>T; c.521C>T; c.538T>G	398	516	761	885	64.56	5.57	46	0.29
id_240	c.511G>A; c.513G>T; c.538T>G	92	155	67	229	41.94	3.19	13	0.29
id_211	c.465C>G; c.511G>A; c.538T>G	183	256	393	569	54.63	4.37	20	0.29
id_34	c.457A>C; c.465C>G; c.538T>G	331	358	519	573	41.61	7.54	37	0.29
id_241	c.511G>A; c.538T>G	604	1115	1041	1733	74.48	2.15	44	0.29
id_165	c.453A>G; c.521C>T; c.538T>G	150	527	155	229	71.71	13.47	16	0.29
id_61	c.457A>C; c.513G>T; c.521C>T; c.538T>G	683	809	760	1562	63.52	3.91	56	0.28
id_38	c.457A>C; c.521C>T; c.538T>G	462	736	958	1464	61.47	5.89	60	0.28
id_29	c.457A>C; c.465C>G; c.513G>T; c.521C>T	68	61	73	61	46.85	18.82	5	0.28
id_111	c.453A>G; c.457A>C; c.511G>A; c.513G>T	1104	1031	1143	2602	73.93	1.76	48	0.28
id_151	c.453A>G; c.465C>G; c.511G>A; c.513G>T; c.521C>T; c.538T>G	24	106	62	145	32.64	17.31	7	0.28
id_46	c.457A>C; c.511G>A; c.513G>T	377	687	591	1317	69.78	4.51	28	0.28
id_14	c.457A>C; c.465C>G; c.511G>A; c.521C>T	820	1030	964	1370	51.44	3.05	48	0.28
id_137	c.453A>G; c.465C>G; c.521C>T; c.538T>G	69	174	126	302	63.10	3.85	15	0.28
id_104	c.453A>G; c.457A>C; c.511G>A	131	221	290	437	81.02	5.19	18	0.28
id_201	c.465C>G; c.511G>A; c.521C>T; c.532A>G	43	158	67	216	72.09	24.67	14	0.27
id_194	c.465C>G; c.521C>T; c.532A>G	1496	1420	1238	2710	48.94	5.44	88	0.27
id_78	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.521C>T	319	517	336	744	63.61	1.28	22	0.27
id_247	c.513G>T; c.521C>T; c.532A>G	1454	1210	1186	2504	58.86	1.77	71	0.27
id_244	c.513G>T; c.532A>G	238	993	512	1291	36.76	4.22	41	0.27
id_184	c.453A>G; c.513G>T; c.521C>T; c.532A>G	760	550	2022	2352	93.15	6.00	12	0.27
id_126	c.453A>G; c.457A>C; c.513G>T; c.521C>T; c.538T>G	1130	1686	1234	1935	62.57	4.45	79	0.27
id_216	c.465C>G; c.513G>T; c.521C>T; c.532A>G	1446	2811	2504	5845	89.88	0.58	48	0.27
id_148	c.453A>G; c.465C>G; c.511G>A; c.513G>T; c.521C>T	23	31	18	120	54.41	22.85	3	0.26
id_188	c.532A>G	416	540	459	851	53.49	3.65	40	0.26
id_10	c.457A>C; c.465C>G; c.521C>T; c.538T>G	64	138	65	120	61.78	9.14	12	0.26
id_203	c.465C>G; c.511G>A; c.521C>T; c.538T>G	831	998	990	1933	84.93	1.76	34	0.26
id_223	c.521C>T; c.532A>G	2732	3360	2891	7547	70.09	4.73	97	0.26
id_53	c.457A>C; c.511G>A; c.513G>T; c.538T>G	17	183	59	104	69.25	7.97	6	0.25
id_93	c.453A>G; c.457A>C; c.465C>G; c.513G>T; c.532A>G; c.538T>G	117	180	216	396	45.12	14.93	9	0.24
id_114	c.453A>G; c.457A>C; c.511G>A; c.513G>T; c.521C>T	662	1094	939	1811	79.97	8.51	39	0.24
id_42	c.457A>C; c.511G>A; c.521C>T	113	108	40	143	81.64	3.68	12	0.24
id_186	c.453A>G; c.513G>T; c.538T>G	15	75	29	108	65.48	10.45	3	0.24
id_20	c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.521C>T	143	341	296	559	73.50	2.82	22	0.24
id_120	c.453A>G; c.457A>C; c.513G>T	73	153	79	144	74.68	9.84	9	0.24
id_233	c.511G>A; c.513G>T	43	16	21	88	40.13	15.70	6	0.23

Identifier	Exon 3 Haplotype	Counts Replicate 1	Counts Replicate 2	Counts Replicate 3	Counts Replicate 4	Mean (%) exon 3 retention	Standard deviation	Total CCS Reads	Relative distance score
id_107	c.453A>G; c.457A>C; c.511G>A; c.521C>T	429	400	747	1135	75.35	7.00	40	0.23
id_225	c.521C>T; c.538T>G	704	1265	768	1650	38.28	6.69	36	0.23
id_138	c.453A>G; c.465C>G; c.511G>A	60	93	187	241	83.01	2.60	7	0.23
id_219	c.465C>G; c.513G>T; c.538T>G	94	95	142	261	57.31	2.06	10	0.23
id_63	c.457A>C; c.538T>G	172	143	248	303	58.64	4.65	20	0.23
id_249	c.513G>T; c.521C>T; c.538T>G	198	254	448	661	69.54	4.17	26	0.23
id_200	c.465C>G; c.511G>A; c.521C>T	273	347	180	679	66.34	3.96	15	0.23
id_187	c.453A>G; c.538T>G	9459	6390	11087	9316	21.67	4.17	23	0.23
id_49	c.457A>C; c.511G>A; c.513G>T; c.521C>T	339	344	392	584	86.81	6.34	21	0.23
id_206	c.465C>G; c.511G>A; c.513G>T; c.521C>T	1210	1433	1290	2296	74.00	3.12	51	0.23
id_251	c.538T>G	528	909	570	1299	65.15	2.66	50	0.23
id_166	c.453A>G; c.511G>A	289	426	464	315	72.92	7.21	29	0.23
id_218	c.465C>G; c.513G>T; c.521C>T; c.538T>G	296	352	437	495	79.39	2.36	20	0.22
id_58	c.457A>C; c.513G>T; c.521C>T	924	1142	790	2121	72.87	2.84	56	0.22
id_141	c.453A>G; c.465C>G; c.511G>A; c.521C>T	6	32	30	30	60.57	19.60	5	0.22
id_71	c.453A>G; c.457A>C; c.465C>G; c.521C>T	121	151	159	329	77.60	11.66	12	0.22
id_191	c.465C>G; c.532A>G	764	858	704	1232	80.22	3.76	40	0.22
id_250	c.513G>T; c.538T>G	568	592	615	1109	70.27	3.73	49	0.22
id_123	c.453A>G; c.457A>C; c.513G>T; c.521C>T	7879	4825	9808	7596	97.56	1.18	33	0.21
id_169	c.453A>G; c.511G>A; c.521C>T	332	204	374	710	79.47	2.19	20	0.21
id_204	c.465C>G; c.511G>A; c.513G>T	32	107	32	53	70.13	4.17	5	0.21
id_196	c.465C>G; c.521C>T; c.538T>G	449	205	379	413	78.63	8.45	23	0.21
id_236	c.511G>A; c.513G>T; c.521C>T	373	186	276	275	82.98	6.56	17	0.21
id_160	c.453A>G; c.465C>G; c.513G>T; c.538T>G	16	19	15	37	58.70	14.36	5	0.20
id_65	c.453A>G; c.457A>C	1868	1929	1471	3713	75.66	3.16	67	0.20
id_161	c.453A>G; c.465C>G; c.538T>G	168	247	275	438	67.30	3.42	15	0.20
id_26	c.457A>C; c.465C>G; c.513G>T	367	595	408	835	82.95	2.41	20	0.20
id_175	c.453A>G; c.511G>A; c.513G>T; c.521C>T	81	76	39	282	83.19	1.38	7	0.20
id_226	c.511G>A	185	233	381	717	70.11	3.40	15	0.19
id_215	c.465C>G; c.513G>T; c.521C>T	6679	4967	9350	7326	98.68	0.74	34	0.18
id_100	c.453A>G; c.457A>C; c.521C>T	843	704	1225	1652	80.35	2.17	54	0.18
id_243	c.513G>T	131	198	281	439	73.81	8.57	11	0.18
id_212	c.465C>G; c.513G>T	2135	2195	2166	4378	91.75	1.98	65	0.18
id_39	c.457A>C; c.511G>A	83	117	192	359	75.02	6.05	14	0.17
id_68	c.453A>G; c.457A>C; c.465C>G	694	603	522	1238	85.00	4.72	40	0.17
id_1	c.457A>C	1098	1911	2983	3516	86.44	3.80	78	0.17
id_246	c.513G>T; c.521C>T	785	1430	1437	1994	93.16	1.66	23	0.17
Control 1	Reference	3824	4957	3419	12605	90.96	3.79	52	0.17
id_154	c.453A>G; c.465C>G; c.513G>T	347	798	489	1292	91.29	1.32	21	0.16
id_4	c.457A>C; c.465C>G	200	177	196	503	82.47	2.50	18	0.16
id_220	c.465C>G; c.538T>G	29	68	82	132	95.07	6.47	6	0.16
id_162	c.453A>G; c.521C>T	1366	822	1294	2015	84.44	1.73	56	0.15
id_190	c.465C>G	5395	2821	9024	8122	97.73	1.05	30	0.15
id_55	c.457A>C; c.513G>T	3610	3477	3546	5094	93.42	1.28	58	0.15
id_35	c.457A>C; c.521C>T	1488	2165	3556	5290	71.16	8.93	71	0.15
id_64	c.453A>G	1736	1259	1375	2382	84.74	1.22	87	0.14

Identifier	Exon 3 Haplotype	Counts Replicate 1	Counts Replicate 2	Counts Replicate 3	Counts Replicate 4	Mean (%) exon 3 retention	Standard deviation	Total CCS Reads	Relative distance score
id_197	c.465C>G; c.511G>A	133	183	137	214	86.13	3.81	10	0.14
id_7	c.457A>C; c.465C>G; c.521C>T	800	1419	736	2192	99.50	0.28	31	0.14
id_229	c.511G>A; c.521C>T	562	843	641	1214	83.07	2.89	28	0.13
id_132	c.453A>G; c.465C>G	116	196	173	309	91.21	0.54	10	0.13
id_222	c.521C>T	2787	2632	1882	3756	91.35	0.26	76	0.12

Supplementary Table S4. Mean (%) of exon 3 retention and distance scores for haplotypes measured by individual and parallelized minigene splicing assays

		Individual assay				Parallelized assay					
Identifier	**Exon 3 Haplotype**	**Non-Variant sites***	**Mean (%) exon 3 retention**	**SD**	**Splicing outcome category**	**Mean (%) exon 3 retention**	**SD**	**Splicing outcome category**	**Prediction Accuracy**	**Total Counts**	**Average Distance score**
id_85	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	1	0.00		S	1.58	0.96	S	TP	2291	0.494
id_19	c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G; c.538T>G	2	3.32	0.71	S	1.46	1.38	S	TP	2566	0.508
id_77	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.532A>G; c.538T>G	2	10.89	2.79	S	0	NA	S	TP	3556	0.497
id_109	c.453A>G; c.457A>C; c.511G>A; c.521C>T; c.532A>G; c.538T>G	2	11.02	0.97	S	0.80	0.22	S	TP	30768	0.498
id_199	c.465C>G; c.511G>A; c.532A>G; c.538T>G	4	12.98	1.74	S	3.01	3.67	S	TP	307	0.479
id_88	c.453A>G; c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.521C>T; c.538T>G	1	17.36	1.56	S	51.75	9.47	I	FN	292	0.359
id_67	c.453A>G; c.457A>C; c.532A>G; c.538T>G	4	24.98	0.51	I	9.14	3.32	S	FP	3887	0.469
id_125	c.453A>G; c.457A>C; c.513G>T; c.521C>T; c.532A>G; c.538T>G	2	33.27	0.71	I	6.33	0.51	S	FP	4854	0.476
id_102	c.453A>G; c.457A>C; c.521C>T; c.532A>G; c.538T>G	3	37.87	2.33	I	7.71	2.32	S	FP	5006	0.459
id_231	c.511G>A; c.521C>T; c.532A>G; c.538T>G	4	43.31	0.69	I	3.24	1.48	S	FP	3819	0.474
id_18	c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.532A>G	3	46.80	3.44	I	2.72	0.96	S	FP	5077	0.429
id_118	c.453A>G; c.457A>C; c.511G>A; c.513G>T; c.538T>G	3	59.15	3.62	I	20.09	4.38	I	TN	1038	0.427
id_24	c.457A>C; c.465C>G; c.511G>A; c.513G>T; c.538T>G	3	63.38	6.18	I	44.05	7.98	I	TN	189	0.428
id_50	c.457A>C; c.511G>A; c.513G>T; c.521C>T; c.532A>G	3	70.55	8.28	I	35.24	7.97	I	TN	1019	0.381
id_52	c.457A>C; c.511G>A; c.513G>T; c.521C>T; c.538T>G	3	70.58		I	70.83	4.49	I	TN	5441	0.356
id_133	c.453A>G; c.465C>G; c.532A>G	5	80.31	6.57	M	39.22	5.91	I	FP	1670	0.383
id_223	c.521C>T; c.532A>G	6	90.19	4.63	M	70.09	4.73	M	TN	16530	0.257
id_201	c.465C>G; c.511G>A; c.521C>T; c.532A>G	4	92.69	2.91	M	72.09	24.67	M	TN	484	0.273
id_114	c.453A>G; c.457A>C; c.511G>A; c.513G>T; c.521C>T	3	94.87	1.38	M	79.97	8.51	M	TN	4506	0.445
id_196	c.465C>G; c.521C>T; c.538T>G	5	97.25	0.55	M	78.63	8.45	M	TN	1446	0.207

*Number of non-variant sites compared to the reference haplotype; SD: Standard Deviation; Splicing outcome categories, S: predominant exon skipping, I: intermediate levels of exon skipping, M: minor exon skipping levels, NA: Not Available, TP: True Positive; FN: False Negative, FP: False Positive, TN: True Negative

Supplementary Table S5. Correlation between individual and parallelized minigene splicing assays

Data for Spearman correlation test	P-value	rho	n
Individually assayed (FAM-quantified) *versus* Average distance scores	8.02E-07	0.87	20.00
Individually assayed (FAM-quantified) *versus* Parallelized mean (%) Short-Long Hybrid	5.30E-06	0.83	20.00
Average distance scores *versus* Parallelized mean (%) Short-Long Hybrid	3.06E-10	0.95	20.00
Average distance scores *versus* Parallelized mean (%) Short-Long Hybrid	2.20E-16	0.86	232.00

p-value (95% Interval Confidence); rho: Spearman rank correlation coefficient; n: number of identifiers

Supplementary Table S6. Effect of single variants on the overall mean exon 3 retention

Variant*	W[†]	P-value	Significance	95% IC[‡]	Difference[§]
c.453	6693	0.05133	ns	(-0.02; 17.15)	7.879512
c.457	6279	0.04281	*	(0.25; 16.88)	8.013108
c.465	5779	0.9878	ns	(-7.86;8.89)	0.05931689
c.511	4524	0.005633	**	(-19.48; -3.20)	-11.05062
c.513	5682	0.7652	ns	(-10.15; 6.42)	-0.9586008
c.521	5327	0.2724	ns	(-12.96; 3.10)	-4.226906
c.532	9653	< 2.2e-16	****	(30.50; 45.00)	37.93396
c.538	2996	6.66E-10	****	(-34.31; -18.58)	-26.57216

*Variant location in the *LW/MW* exon 3 haplotypes; [†]Mann-Whitney U test or Wilcoxon rank sum test statistic calculated between the haplotypes carrying either of the two alleles for each SNP shown in Fig. 3A; [‡]95 percent confidence interval; [§]Difference in location.

Supplementary Table S7. Additive model analysis for double variants

Double variants AB	ObsAB_1	ExpAB_1	Diff_Obs_ExpAB_1	ObsAB_2	ExpAB_2	Diff_Obs_ExpAB_2	ObsAB_3	ExpAB_3	Diff_Obs_ExpAB_3	ObsAB_4	ExpAB_4	Diff_Obs_ExpAB_4	P-value	Epistasis
c.453A>G; c.532A>G	85.40	67.15	18.25	87.77	63.13	24.64	82.01	56.88	25.13	86.47	59.91	26.56	0.001023	Negative
c.453A>G; c.513G>T	98.59	48.25	50.34	99.16	50.63	48.53	98.14	37.26	60.88	99.00	29.67	69.33	0.001309	Negative
c.453A>G; c.538T>G	82.29	48.38	33.91	74.84	49.95	24.89	81.57	53.26	28.32	74.61	48.83	25.78	0.0008044	Negative
c.453A>G; c.511G>A	30.80	43.75	-12.96	17.14	42.89	-25.76	26.72	46.33	-19.61	33.65	47.62	-13.97	0.00874	Positive
c.453A>G; c.457A>C	29.07	34.31	-5.24	22.97	30.78	-7.81	22.64	23.85	-1.22	22.68	26.33	-3.65	0.04805	Positive
c.453A>G; c.521C>T	14.64	24.98	-10.34	15.33	24.60	-9.27	14.22	23.39	-9.17	18.06	22.65	-4.58	0.007345	Positive
c.453A>G; c.465C>G	8.62	17.78	-9.16	9.18	19.83	-10.64	9.25	16.10	-6.85	8.09	16.42	-8.33	0.001592	Positive
c.511G>A; c.532A>G	100.00	78.53	21.47	86.75	73.46	13.29	71.43	73.54	-2.12	72.73	80.07	-7.34	0.4137	NA
c.532A>G; c.538T>G	79.20	83.16	-3.95	78.55	80.52	-1.97	74.39	80.47	-6.08	79.46	81.28	-1.83	0.04072	Positive
c.513G>T; c.532A>G	59.66	83.02	-23.36	69.28	81.20	-11.91	61.33	64.47	-3.14	62.66	62.13	0.54	0.1729	NA
c.457A>C; c.532A>G	64.75	69.09	-4.34	63.24	61.35	1.89	57.60	51.07	6.53	59.19	58.78	0.41	0.6504	NA
c.521C>T; c.532A>G	32.32	59.75	-27.43	34.17	55.17	-21.01	29.82	50.60	-20.79	23.35	55.10	-31.75	0.002484	Positive
c.465C>G; c.532A>G	24.74	52.56	-27.82	20.63	50.40	-29.77	17.19	43.31	-26.12	16.56	48.88	-32.32	0.0002117	Positive
c.521C>T; c.538T>G	60.23	40.99	19.24	67.11	41.98	25.13	52.86	46.98	5.89	66.67	44.02	22.64	0.02383	Negative
c.457A>C; c.538T>G	36.63	50.32	-13.69	38.46	48.16	-9.70	46.77	47.44	-0.66	43.56	47.70	-4.14	0.09275	NA
c.513G>T; c.538T>G	26.58	64.26	-37.67	32.09	68.01	-35.91	26.50	60.84	-34.34	33.72	51.05	-17.33	0.006944	Positive
c.457A>C; c.521C>T	21.30	26.91	-5.61	41.76	22.82	18.94	26.94	17.57	9.37	25.37	21.52	3.85	0.2869	NA
c.511G>A; c.538T>G	23.51	59.76	-36.25	27.35	60.27	-32.92	23.82	69.92	-46.09	27.41	69.00	-41.59	0.0008803	Positive
c.457A>C; c.511G>A	24.10	45.69	-21.60	31.62	41.10	-9.48	17.19	40.51	-23.33	27.02	46.49	-19.47	0.009442	Positive
c.457A>C; c.465C>G	20.00	19.72	0.28	16.95	18.04	-1.09	14.29	10.28	4.00	18.89	15.30	3.59	0.2671	NA
c.511G>A; c.521C>T	20.82	36.36	-15.54	15.42	34.93	-19.51	14.20	40.05	-25.85	17.30	42.81	-25.51	0.003228	Positive
c.465C>G; c.511G>A	19.55	29.16	-9.61	12.57	30.15	-17.59	11.68	32.76	-21.08	11.68	36.59	-24.91	0.01116	Positive
c.465C>G; c.513G>T	11.15	33.66	-22.51	7.88	37.89	-30.01	7.11	23.68	-16.57	6.88	18.64	-11.77	0.01429	Positive
c.513G>T; c.521C>T	6.75	40.85	-34.10	4.83	42.66	-37.84	6.89	30.97	-24.09	8.88	24.86	-15.99	0.01092	Positive
c.457A>C; c.513G>T	6.34	50.18	-43.84	6.10	48.84	-42.74	5.47	31.44	-25.97	8.42	28.54	-20.12	0.0115	Positive
c.465C>G; c.538T>G	0.00	33.79	-33.79	0.00	37.21	-37.21	6.10	39.68	-33.59	13.64	37.80	-24.16	0.001414	Positive

A:Variant A; B: Variant B for each couple of double variants; A_1-4 and B_1-4 from each 4 replicates; ObsAB: observed ratio for combination of A and B variants; ExpAB: expected ratio following an additive model; Diff_ObsAB_ExpAB: difference between observed and expected ratio for each given pair of variants

Supplementary Table S8. Additive model analysis for triple variants

Triple mutants	O_ABC_1	E_ABC_1	OE_ABC_1	O_ABC_2	E_ABC_2	OE_ABC_2	O_ABC_3	E_ABC_3	OE_ABC_3	O_ABC_4	E_ABC_4	OE_ABC_4	P-value	Epistasis
c.453A>G; c.532A>G; c.538T>G	89.72	99.35	-9.63	85.96	96.80	-10.84	85.30	95.31	-10.01	82.37	95.01	-12.64	0.00053	Positive
c.453A>G; c.511G>A; c.532A>G	72.22	94.72	-22.49	73.14	89.74	-16.60	63.37	88.38	-25.01	66.93	93.80	-26.87	0.00202	Positive
c.453A>G; c.465C>G; c.532A>G	52.20	68.74	-16.54	61.54	66.68	-5.14	64.71	58.15	6.56	64.68	62.61	2.08	0.56370	NA
c.453A>G; c.521C>T; c.532A>G	57.23	75.94	-18.71	55.49	71.46	-15.96	56.24	65.44	-9.20	52.48	68.83	-16.35	0.00517	Positive
c.453A>G; c.457A>C; c.532A>G	44.57	85.27	-40.70	61.44	77.63	-16.19	53.30	65.90	-12.60	61.66	72.51	-10.85	0.06325	NA
c.453A>G; c.457A>C; c.538T>G	41.52	66.51	-24.99	69.55	64.44	5.11	54.28	62.28	-7.99	70.33	61.43	8.89	0.57930	NA
c.453A>G; c.513G>T; c.538T>G	33.33	80.44	-47.11	38.67	84.29	-45.62	20.69	75.68	-54.99	45.37	64.78	-19.41	0.01244	Positive
c.453A>G; c.465C>G; c.538T>G	33.33	49.98	-16.64	27.94	53.49	-25.56	33.45	54.52	-21.07	36.07	51.53	-15.46	0.00336	Positive
c.453A>G; c.521C>T; c.538T>G	39.33	57.17	-17.84	9.30	58.27	-48.97	36.13	61.81	-25.68	28.38	57.75	-29.37	0.01929	Positive
c.453A>G; c.457A>C; c.513G>T	26.03	66.37	-40.34	33.99	65.12	-31.13	11.39	46.27	-34.88	29.86	42.27	-12.41	0.01629	Positive
c.453A>G; c.511G>A; c.521C>T	18.07	52.55	-34.47	20.10	51.21	-31.11	20.59	54.89	-34.30	23.38	56.54	-33.16	0.00003	Positive
c.453A>G; c.457A>C; c.521C>T	21.59	43.10	-21.51	21.45	39.10	-17.65	17.63	32.41	-14.78	17.92	35.25	-17.33	0.00102	Positive
c.453A>G; c.457A>C; c.511G>A	21.37	61.88	-40.50	17.65	57.39	-39.74	12.41	55.35	-42.94	24.49	60.22	-35.73	0.00012	Positive
c.453A>G; c.465C>G; c.511G>A	16.67	45.35	-28.68	13.98	46.44	-32.46	20.32	47.60	-27.27	17.01	50.32	-33.30	0.00024	Positive
c.453A>G;c.457A>C; c.465C>G;	10.09	35.90	-25.82	20.23	34.32	-14.09	12.07	25.12	-13.05	17.61	29.02	-11.41	0.01632	Positive
c.453A>G; c.465C>G; c.513G>T	9.80	49.84	-40.04	9.65	54.17	-44.52	6.95	38.52	-31.57	8.44	32.37	-23.93	0.00461	Positive
c.457A>C; c.532A>G; c.538T>G	96.29	101.28	-4.99	96.43	95.01	1.42	96.51	89.49	7.03	96.21	93.88	2.33	0.60010	NA
c.457A>C; c.511G>A; c.532A>G	90.69	96.65	-5.96	97.44	87.96	9.48	95.76	82.56	13.20	96.78	92.67	4.11	0.29980	NA
c.511G>A; c.532A>G; c.538T>G	96.30	110.73	-14.43	90.12	107.12	-17.00	97.67	111.97	-14.29	95.42	115.18	-19.75	0.00105	Positive
c.465C>G; c.532A>G; c.538T>G	83.33	84.75	-1.42	82.56	84.06	-1.50	83.93	81.73	2.20	91.81	83.98	7.83	0.47750	NA
c.457A>C; c.513G>T; c.532A>G	75.53	101.15	-25.62	76.85	95.69	-18.84	74.15	73.49	0.67	79.56	74.73	4.83	0.27910	NA
c.465C>G; c.513G>T; c.532A>G	71.43	84.62	-13.19	81.40	84.74	-3.34	46.67	65.73	-19.06	70.59	64.82	5.77	0.26620	NA
c.457A>C; c.521C>T; c.532A>G	64.18	77.88	-13.70	58.97	69.67	-10.69	57.35	59.62	-2.27	65.42	67.70	-2.28	0.09008	NA
c.457A>C; c.511G>A; c.532A>G	55.67	70.68	-15.01	65.52	64.89	0.63	56.94	52.33	4.62	57.66	61.48	-3.82	0.48130	NA
c.511G>A; c.521C>T; c.532A>G	52.80	87.32	-34.52	53.76	81.78	-28.02	53.07	82.10	-29.03	53.01	88.99	-35.98	0.00052	Positive
c.465C>G;c.521C>T; c.532A>G	43.98	61.35	-17.36	53.17	58.72	-5.55	50.24	51.87	-1.62	56.86	57.80	-0.93	0.19270	NA
c.513G>T; c.532A>G; c.538T>G	36.45	115.22	-78.77	54.89	114.86	-59.97	38.04	102.89	-64.85	50.96	97.23	-46.27	0.00262	Positive
c.513G>T; c.521C>T; c.532A>G	41.13	91.81	-50.69	42.81	89.52	-46.71	38.70	73.02	-34.32	41.93	71.05	-29.11	0.00421	Positive
c.521C>T; c.532A>G; c.538T>G	26.54	91.95	-65.41	31.04	88.84	-57.79	21.40	89.02	-67.62	29.39	90.20	-60.81	0.00010	Positive
c.511G>A; c.513G>T; c.532A>G	0.00	110.59	-110.59	0.00	107.80	-107.80	0.00	95.96	-95.96	0.00	96.02	-96.02	0.01632	Positive
c.457A>C; c.511G>A; c.538T>G	58.25	77.89	-19.64	76.00	74.77	1.23	65.42	78.93	-13.51	76.00	81.59	-5.59	0.13180	NA
c.457A>C; c.465C>G; c.538T>G	69.49	51.91	17.57	52.79	51.70	1.09	54.91	48.70	6.21	56.37	50.40	5.97	0.11430	NA
c.511G>A; c.513G>T; c.538T>G	57.61	91.83	-34.22	60.65	94.62	-33.97	53.73	92.34	-38.61	60.26	84.94	-24.68	0.00152	Positive
c.465C>G; c.511G>A; c.538T>G	46.99	61.36	-14.36	46.88	63.82	-16.94	38.93	71.18	-32.25	48.68	71.69	-23.01	0.01215	Positive
c.465C>G; c.513G>T; c.538T>G	45.74	65.85	-20.11	42.11	71.55	-29.45	41.55	62.10	-20.55	41.38	53.75	-12.37	0.00970	Positive
c.511G>A; c.521C>T; c.538T>G	43.32	68.56	-25.23	45.96	68.59	-22.63	37.61	78.47	-40.86	43.61	77.91	-34.31	0.00524	Positive
c.457A>C; c.513G>T; c.538T>G	53.54	82.38	-28.85	37.31	82.50	-45.19	33.27	69.86	-36.59	44.19	63.65	-19.46	0.00960	Positive
c.457A>C; c.521C>T; c.538T>G	37.01	59.11	-22.10	45.79	56.48	-10.69	31.63	55.99	-24.37	39.69	56.62	-16.94	0.00887	Positive
c.465C>G; c.511G>A; c.521C>T	35.90	37.95	-2.06	38.04	38.47	-0.43	31.11	41.31	-10.20	29.60	45.51	-15.90	0.14260	NA
c.513G>T; c.521C>T; c.538T>G	34.85	73.05	-38.20	30.71	76.33	-45.62	31.47	69.40	-37.92	24.81	59.97	-35.16	0.00041	Positive
c.457A>C; c.511G>A; c.513G>T	36.07	77.75	-41.68	28.53	75.45	-46.92	25.38	62.93	-37.55	30.90	62.44	-31.53	0.00121	Positive
c.465C>G; c.511G>A; c.513G>T	25.00	61.22	-36.22	28.04	64.50	-36.46	34.38	55.18	-20.80	32.08	52.53	-20.46	0.00815	Positive

A:Variant A; B: Variant B and C: Variant C for each combination of triple variants; O_ABC: observed ratio for combination of A, B and C variants; E_ABC: expected ratio for the combination of A, B and C variants following an additive model, and OE_ABC: difference between observed and expected for each given trio of variants

Supplementary Table S8. Additive model analysis for triple variants (continued)

Triple mutants	O_ABC_1	E_ABC_1	OE_ABC_1	O_ABC_2	E_ABC_2	OE_ABC_2	O_ABC_3	E_ABC_3	OE_ABC_3	O_ABC_4	E_ABC_4	OE_ABC_4	P-value	Epistasis
c.457A>C; c.513G>T; c.521C>T	26.95	58.98	-32.03	27.50	57.16	-29.66	30.51	39.99	-9.49	23.57	37.46	-13.89	0.03241	Positive
c.465C>G; c.521C>T; c.538T>G	10.02	42.58	-32.56	26.34	45.53	-19.19	20.05	48.24	-28.19	29.06	46.72	-17.66	0.00644	Positive
c.457A>C; c.511G>A; c.521C>T	15.93	54.48	-38.55	19.44	49.43	-29.98	15.00	49.07	-34.07	23.08	55.41	-32.33	0.00034	Positive
c.457A>C; c.465C>G; c.513G>T	15.26	51.78	-36.52	14.79	52.38	-37.59	18.38	32.70	-14.32	19.76	31.24	-11.48	0.03761	Positive
c.511G>A; c.513G>T; c.521C>T	9.65	68.42	-58.77	17.74	69.27	-51.53	15.22	62.47	-47.25	25.45	58.76	-33.30	0.00299	Positive
c.465C>G; c.513G>T; c.521C>T	0.63	42.45	-41.82	2.17	46.21	-44.03	0.78	32.24	-31.46	1.68	27.56	-25.88	0.00362	Positive
c.457A>C; c.465C>G; c.521C>T	0.88	28.51	-27.63	0.35	26.36	-26.01	0.54	18.84	-18.29	0.23	24.21	-23.99	0.00132	Positive

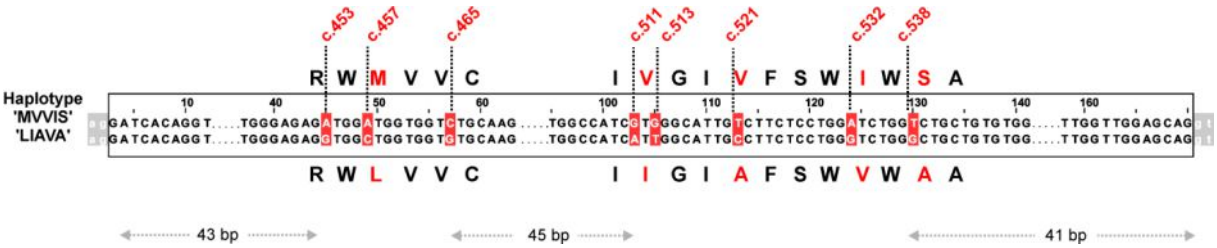
A:Variant A; B: Variant B and C: Variant C for each combination of triple variants; O_ABC: observed ratio for combination of A, B and C variants; E_ABC: expected ratio for the combination of A, B and C variants following an additive model, and OE_ABC: difference between observed and expected for each given trio of variants

Supplementary Table S9. Allele Frequencies in Controls

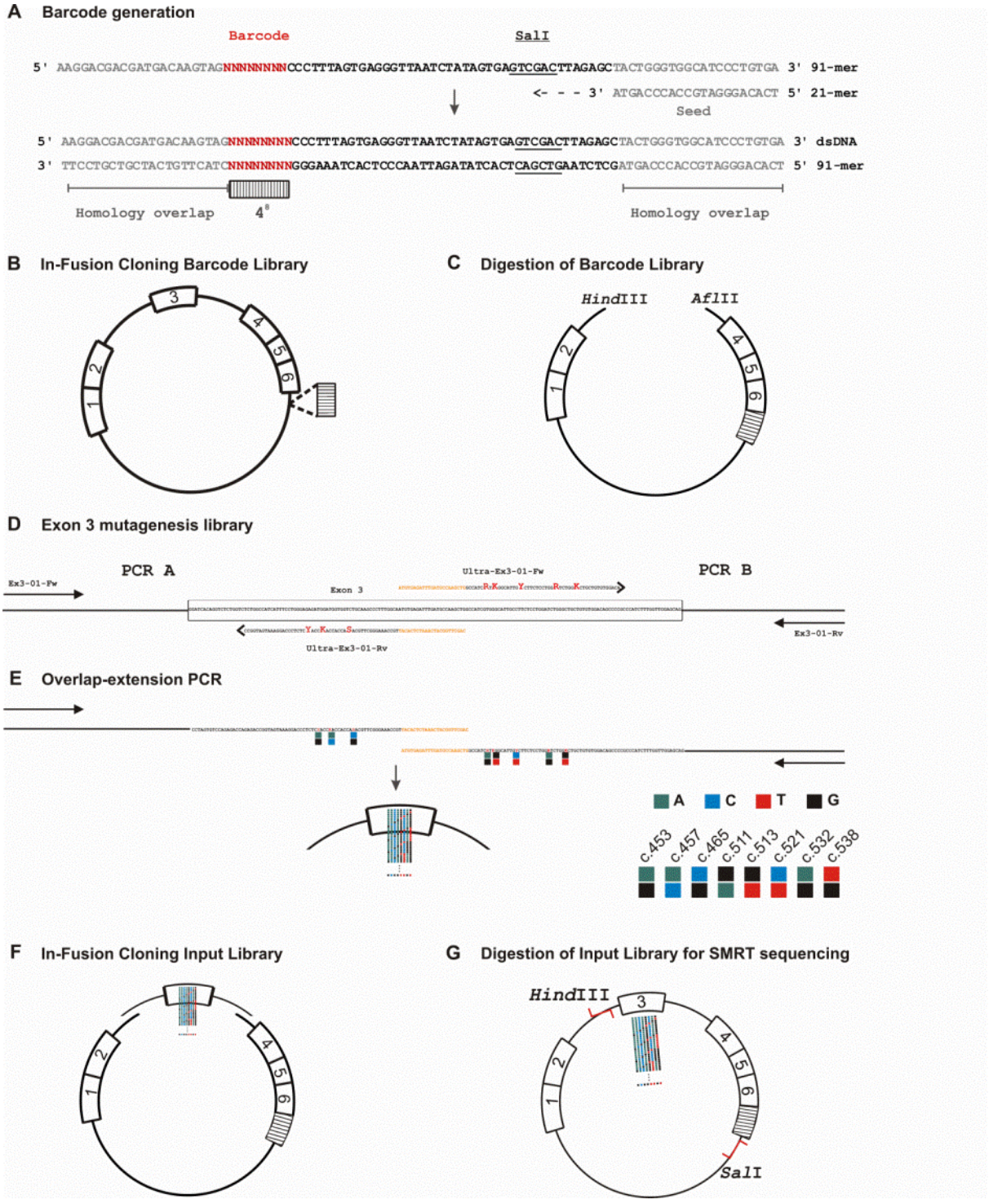
Variant	*OPN1LW* (proximal)	*OPN1MW* (distal)
c.453A*	0.283	0.904
c.457A	0.334	0.885
c.465C*	0.283	0.772
c.511A	0.058	0.005
c.513T	0.057	0.005
c.521T	0.086	0.122
c.532G	0.051	0.138
c.538G	0.434	0.916

Allele frequencies were calculated with data from n = 1,418 exon 3 haplotypes of control except for variants (*) for which allele frequencies were derive from n = 812 exon 3 haplotypes of controls.

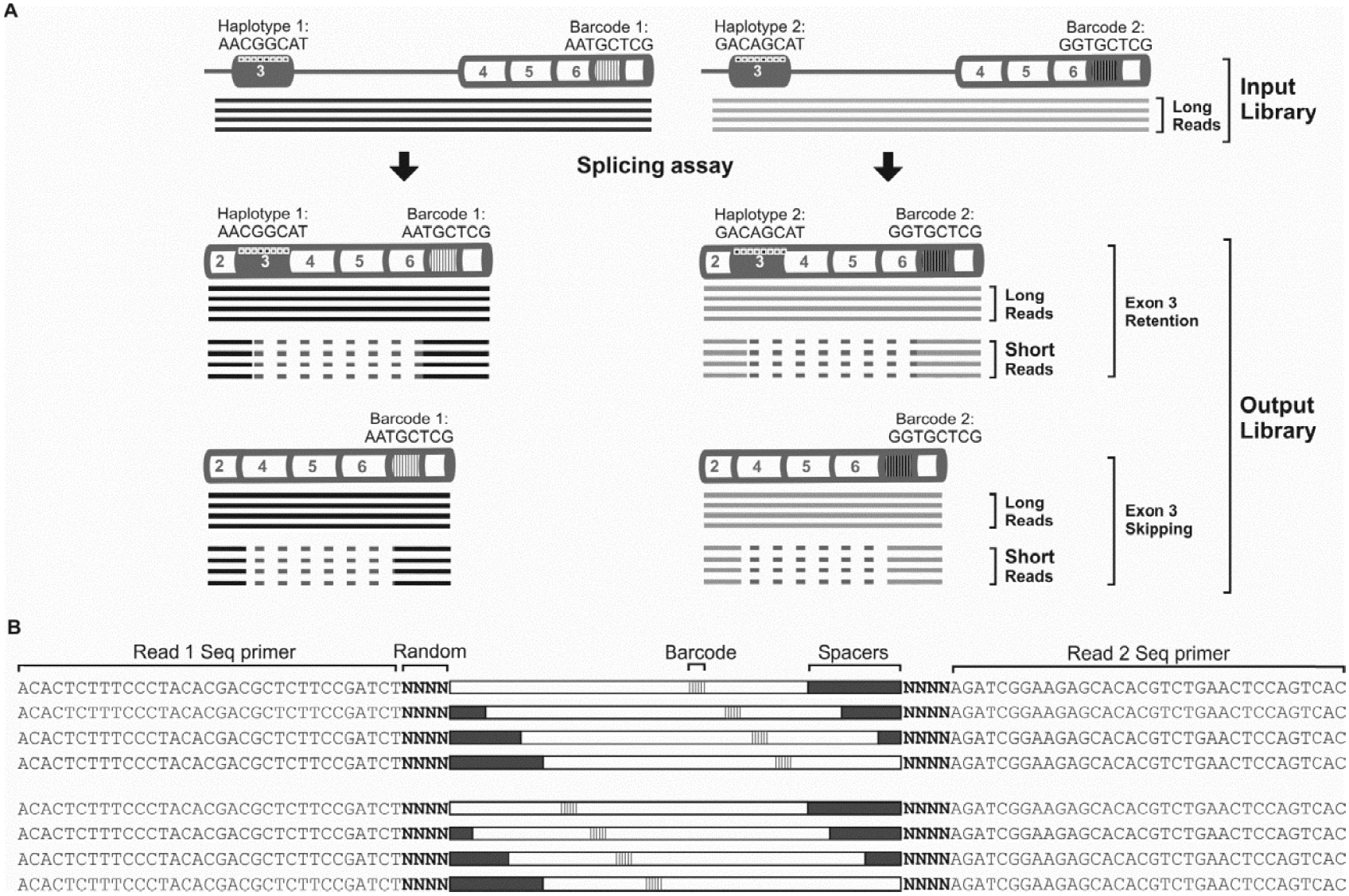
Supplementary Figures



Supplementary Figure S1. Haplotype definition. The haplotype in exon 3 of *OPN1LW/MW* comprises the following SNPs: c.453A>G, c.457A>C, c.465C>G, c.511G>A, c.513G>T, c. 521C>T, c.532A>G, and c.538T>G (Winderickx et al., 1993). The deduced cone pigment variant encodes for the following amino acid exchanges: p.[(=); M153L; (=); V171I; V174A; I178V; S180A]. The variants and amino acid exchanges are highlighted in red. The distance from the variants at the extremes of the haplotype to the donor (“gt”) and acceptor (“ag”) splice sites is indicated in base pairs (41 and 43 bp, respectively). The distance between the 5’ block and the 3’ block of the haplotype is also indicated (45 bp).

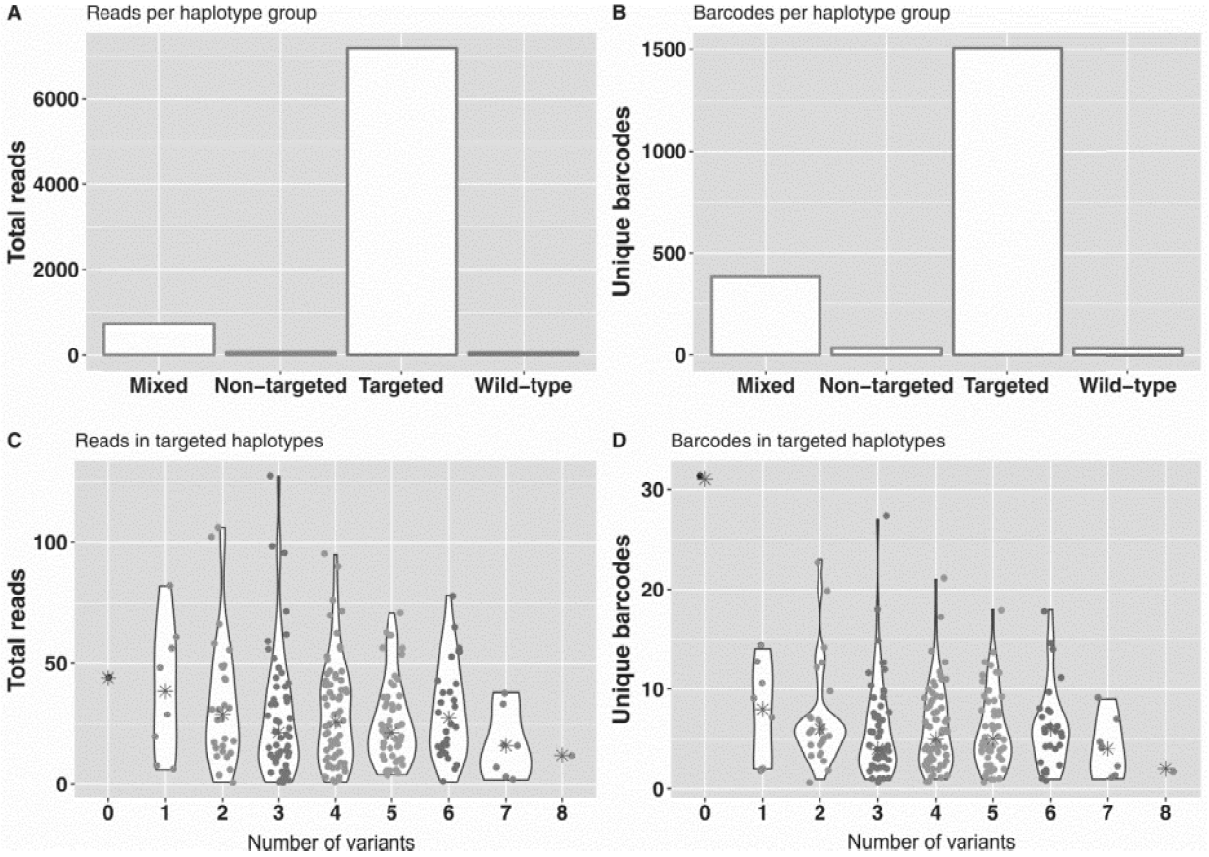


Supplementary Figure S2. Input Library Construction. **A)** Generation of a dsDNA 91-mer with an 8-bp degenerate tag and 21-bp homology overlaps for In-Fusion Cloning. **B)** Generation of the barcode library pool from 33,000 scrapped bacterial colonies. **C)** Digestion of the barcode library. **D)** Two separate PCRs with degenerate primers introduced targeted mutations in exon 3. **E)** Overlap-extension PCR generates the exon 3 mutagenesis library. **F)** In-Fusion Cloning of the exon 3 mutagenesis library into the digested barcode library produces the Input Library. **G)** Digestion of the Input Library for sequencing with PacBio.

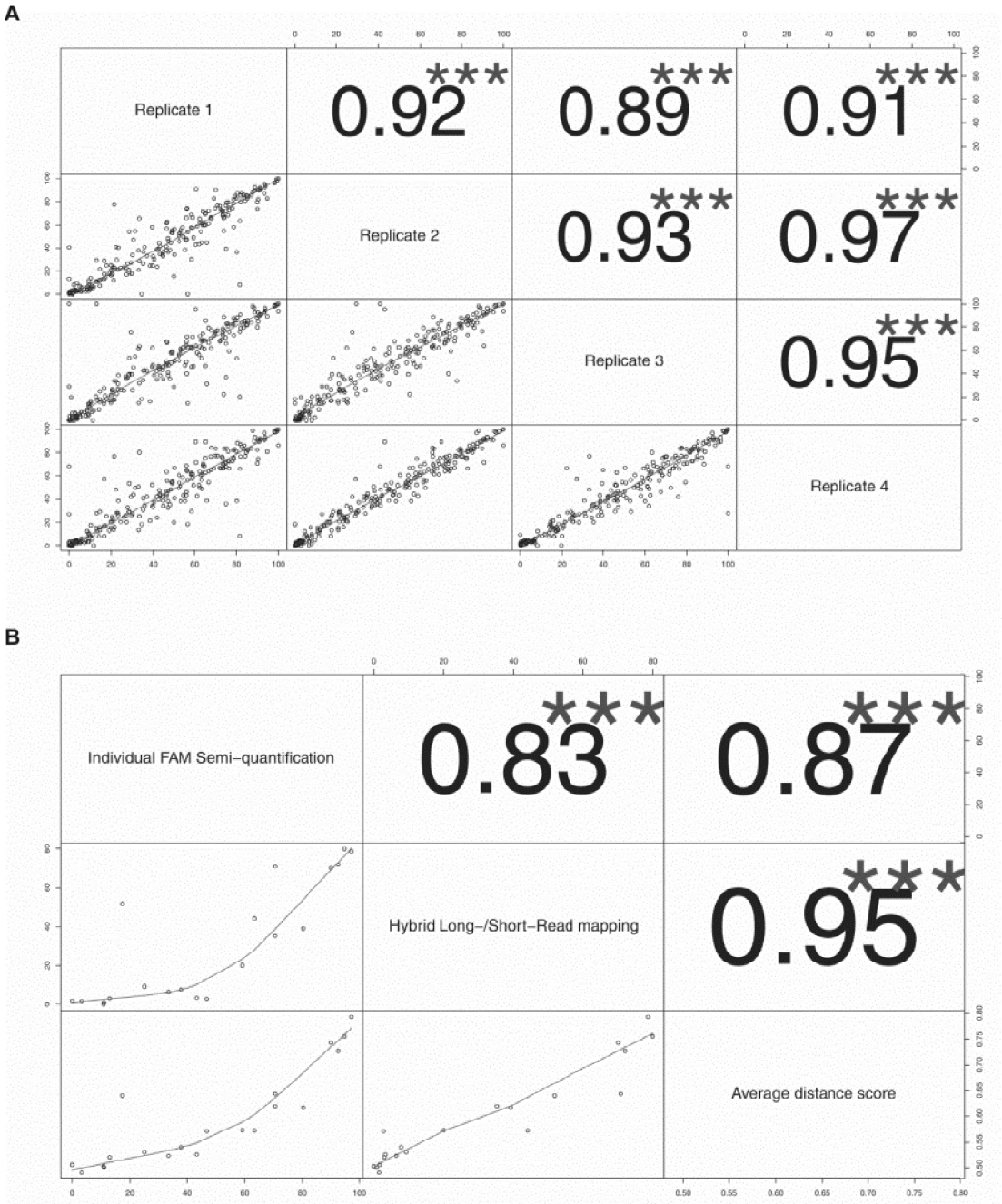


Supplementary Figure S3. Output Library construction (figure legend continues in next page)

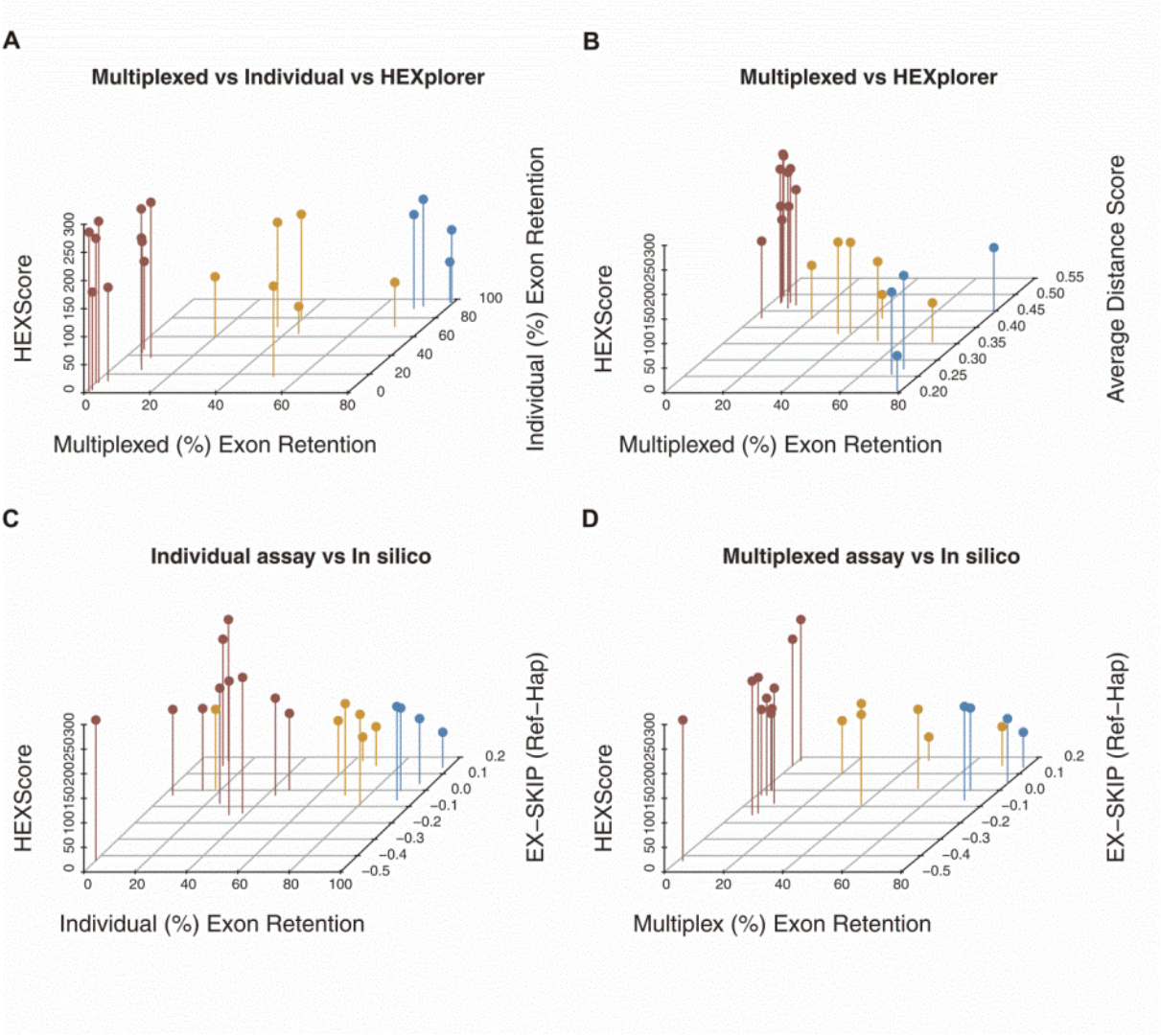
Supplementary Figure S3. Output Library construction. **A)** Simplified scheme for the preparation of the Output Library for NGS. Scheme exemplifies two *OPN1LW/MW* minigenes: Haplotype 1 (c.453A; c.457A; c.465C; c.511G; c.513G; c.521C; c.532A; c.538T) and Haplotype 2 (c.453G; c.457A; c.465C; c.511A; c.513G; c.521C; c.532A; c.538T), each tagged by unique barcodes: Barcode 1 ('AATGCTCG') and Barcode 2 ('GGTGCTCG'). Long-reads (depicted as continuous lines) are generated for the digested Input Library constructs and for the RT-PCR amplicons (Output Library) while paired-end short-reads (depicted as continuous lines with dotted lines in between representing lacking sequence) were generated only for the RT-PCR amplicons derived from transcripts with retained or skipped exon 3. **B)** Strategy for short-read amplicon sequencing of the Output Library. The scheme represents the amplicons obtained from the first round PCR with primers (Supp. Table S2) to introduce Read 1 and Read 2 sequences, 8 random bases and 8 bp of offset spacers per amplicon. The position of the 'Barcode' within the dual-direction strategy (to increase the nucleotide diversity) is depicted by a striped stretch and refers to the internal 8-bp barcode tagging exon 3 haplotypes. The second round PCR with primers (Supp. Table S2) introduces the adaptor P5 upstream of Read 1 and the adaptor P7 upstream of Read 2 with following indices in between for the Index Read at the P7 site (single indexing): 701 (5' -TCGCCTTA- 3'), 702 (5' -CTAGTACG- 3'), 703 (5' -TTCTGCCT- 3') and 704 (5' -GCTCAGGA- 3') for multiplexing the four replicates.



Supplementary Figure S4. Characteristics of the Input Library. **A)** Bar graph showing the total number of CCS reads per uniquely tagged haplotype and **B)** the total number of unique barcodes for each haplotype category: targeted, non-targeted, mixed (non-targeted and targeted) and wild-type (Reference haplotype). **C)** In the violin plot each dot represents a unique haplotype and the number of CCS read observations and **D)** unique barcodes is depicted only for targeted haplotypes with 0-8 variants with respect to the Reference haplotype. The median number of CCS reads (C) or unique barcodes (D) is represented by a star.



Supplementary Figure S5. Reproducibility and correlation of the Output Library with individually assayed minigenes. **A)** The exon 3 retention % levels determined by paired-end reads mapped on the custom reference genome highly correlated across the four replicates; Spearman correlation coefficient rho= 0.89 - 0.97; p-value < 2.2e-16 (95% Interval Confidence). **B)**. Correlation of exon 3 retention levels between parallelized and individual minigene splicing assays. A Spearman correlation coefficient was calculated for the exon 3 retention % levels measured by "Individual FAM Semi-quantification"; capillary electrophoresis of FAM-labeled RT-PCR amplicons derived from 20 different haplotypes generated by mutagenesis and tagged by a barcode assayed one-by-one through *in vitro* minigene assays and by the parallelized minigene assay in two formats; "Average distance score" refers to scores derived from only long-read dataset transformed as (-1)*(1-n), and the "Hybrid Long-/Short-Read mapping", a hybrid combined mapping of short-reads against the long-read assembly.



Supplementary Figure S6. Comparison of the splicing outcome derived from individual and parallelized *OPN1LW/MW* minigene assays with *in silico* prediction scores from two independent algorithms (HEXplorer and EX-SKIP). HEXScores represent the difference between HEXplorer scores of the Reference and a given haplotype (n = 20). HEXplorer scores are calculated as the average Z-score of all six overlapping hexamers for each nucleotide (Erkelenz et al., 2014). HEXScores (y-axis) are plotted against A) the individual (z-axis) and parallelized (x-axis) mean (%) exon 3 retention levels, B) the average distance score (z-axis) and parallelized (x-axis) mean (%) exon 3 retention levels, C) the difference between EX-SKIP scores for the Reference and a given haplotype (z-axis) and individual (x-axis) mean (%) exon 3 retention levels and D) the difference between EX-SKIP scores for the Reference and a given haplotype (z-axis) and parallelized (x-axis) mean (%) exon 3 retention levels. Colours represent categories according to splicing competence (Fig.2; Supp. Table S4).

13.1.5 Publication V

Nicole Weisschuh, **Elena Buena-Atienza**, **Bernd Wissinger**. Splicing Mutations in Inherited Retinal Diseases

This is a pre-copyedited, author-produced version of an article accepted for publication in *Progress in Retinal and Eye Research* following peer review. The version of record “Weisschuh N, Buena-Atienza E, Wissinger B. Splicing mutations in inherited retinal diseases [published online ahead of print, 2020 Jun 15]. Prog Retin Eye Res. 2020;100874.” is available online at: https://doi:10.1016/j.preteyeres.2020.100874

Abstract

Mutations which induce aberrant transcript splicing represent a distinct class of disease-causing genetic variants in retinal disease genes. Such mutations may either weaken or erase regular splice sites or create novel splice sites which alter exon recognition. While mutations affecting the canonical GU-AG dinucleotides at the splice donor and splice acceptor site are highly predictive to cause a splicing defect, other variants in the vicinity of the canonical splice sites or those affecting additional *cis*-acting regulatory sequences within exons or introns are much more difficult to assess or even to recognize and require additional experimental validation. Splicing mutations are unique in that the actual outcome for the transcript (e.g. exon skipping, pseudoexon inclusion, intron retention) and the encoded protein can be quite different depending on the individual mutation. In this article, we present an overview on the current knowledge about and impact of splicing mutations in inherited retinal diseases. We introduce the most common sub-classes of splicing mutations including examples from our own work and others and discuss current strategies for the identification and validation of splicing mutations, as well as therapeutic approaches, open questions, and future perspectives in this field of research.

Keywords

Splicing Mutations, Inherited Retinal Disease, Deep-intronic Mutations, Cryptic Splice Sites, Splicing Correction Therapies

Contents

1. Introduction 218

2. Splicing process and machinery 219

2.1 Splice sites and the spliceosome 219

2.2 *Cis*-acting regulatory elements and *trans*-acting splicing factors 220

3. Disease-causing splice variants in inherited retinal disease 220

3.1 Mutations leading to exon skipping 221

3.2 Mutations leading to intron retention 222

3.3 Mutations leading to the activation of pseudoexons 222

3.4 Exonic splicing mutations 223

3.5 *OPN1LW* and *OPN1MW* exon 3 haplotypes impair splicing 225

3.6 Mutations affecting *trans*-acting splicing factors 226

4. Validation of splicing mutations 227

4.1 *In silico* prediction tools 227

4.2 Direct transcript analysis 229

4.3 Heterologous splice assays 230

5. Therapeutic approaches for splicing defects 232

5.1 Sequence-adapted U1 and U6 snRNA 232

5.2 Small molecules with splicing rescue activity 233

5.3 Antisense Oligonucleotides 233

5.4 Genome Editing 235

5.5 *Trans*-Splicing 236

6. Future directions 236

1. Introduction

Technical improvements in genetic analysis and in particular the pace and amount of sequence data generated through the introduction of next generation sequencing technologies have not only boosted our knowledge on the number and diversity of genes linked to retinal disease but also yielded an unprecedented diversity of variants linked to inherited disease (Chaitankar et al., 2016; Broadgate et al., 2017). A considerable proportion of them cause a defect in transcript splicing (Bacchi et al., 2014).

Most genes in the human genome are discontinuous, i.e. they contain intervening sequences (introns) which are transcribed and part of the primary transcript (precursor messenger ribonucleic acid, pre-mRNA) but need to be removed during splicing to yield the mature functional mRNA. In fact, only about 5% of the protein-coding genes in the human genome are single-exon genes, i.e. the transcript does not need to be spliced, and the average gene comprises 10-11 exons (range 1-363). The size of introns is on average about 5,000 bp (range from as low as 43 bp to 1,049,910 bp) and thus exceeds the size of exons (mean: 365 bp, median: 141 bp, range: 2 bp to 24,927 bp) by far (Hubé & Francastel, 2015).

Not all splice sites are constitutively recognized and used during transcript processing. Alternative splicing (AS), i.e. the differential inclusion of subsets of exons into the mRNA, is rather the rule in multiexon mammalian genes and provides the opportunity to diversify gene and protein function with a limited number of genes. In fact, AS is considered a driving force during evolution and the extent and diversity of AS correlates with organismal complexity (Kim et al., 2008; Chen et al., 2014). There are five major types of AS: i) exon skipping, ii & iii)) the use of alternative splice donor or splice acceptor sites, iv) intron retention and v) inclusion of mutually exclusive exons. The about 20,000 protein coding genes listed in the latest GENCODE release of the human genome (V31; https://www.gencodegenes.org/) give rise to more than 83,000 distinct transcripts. This fourfold increase in diversity is to a large extent due to AS. The increasing depth of transcriptome analysis and the expanding coverage of cell and tissue diversity is yielding even higher numbers of alternatively spliced transcripts. However, some AS events are only observed in a minute proportion of transcripts and their biological relevance therefore remains questionable (Nellore et al., 2016). Additional variability in AS at the population level due to inherited variations adds to this complexity (Park et al. 2018).

The mammalian retina is a highly specialized tissue evolved for light detection, conversion of the absorbed light into an electrical signal and its neuronal processing along with supporting structural and biochemical components in retinal cells. Although the retina is composed of a few major cell types (photoreceptors, horizontal cells, bipolar cells, amacrine cells, retinal ganglion cells, glial cells, and retinal pigment epithelium cells) there is a considerable diversity of these cell types at the morphological and functional level (Baden et al., 2016; Franke et al., 2017; Grünert & Martin, 2020; Hoon et al., 2014) which is also reflected by gene expression diversity in single cell transcriptomics (Shekhar et al., 2016; Rheaume et al., 2018; Peng et al., 2019; Voigt et al., 2019; Lukowski et al., 2019). Recent RNAseq studies of the human retina have shown that a large fraction of the entire gene complement of the genome is expressed in this tiny piece of tissue (i.e. about 2/3 of all protein coding genes) (Farkas et al., 2013; Pinelli et al., 2016 and for review: Zelinger & Swaroop, 2018). Notably, these studies have shown an enormous magnitude of AS with about 50% of expressed genes displaying altered exon composition compared to the reference sequence including several thousands of novel exons. Mature photoreceptors in particular adhere to an exceptional developmentally regulated AS program that is characterized by a switch-like pattern with high levels of inclusion of certain exons in photoreceptors and their almost complete exclusion in extra-retinal tissues which is driven by a lack of prototypical neuronal splicing factors and upregulation of the Musashi 1 factor at least in the murine retina (Murphy et al., 2016).

Moreover, dynamic changes in AS have been observed during retinal development with differential splicing events occurring more frequently during early development (Wan et al., 2011; Kim et al., 2016). A large fraction of genes undergoing changes in AS during retinal development are genes involved in ciliogenesis and ciliary transport as well as genes underlying retinal dystrophies (Mellough et al., 2019).

An exquisite example is *RPGR* encoding the retinitis pigmentosa GTPase regulator which localizes to the connecting cilium and interacts with a variety of ciliary proteins. A variety of alternatively spliced transcripts have been identified in the distal portion of this gene. The $RPGR^{ORF15}$ transcript is an alternatively spliced isoform in which exon 15 extends into intron 15 due to skipping of the splice donor site for exon 15 and is the most abundant isoform in photoreceptor. Notably, ORF15 is a hotspot for mutations in *RGPR* linked to X-linked Retinitis pigmentosa (Vervoort et al., 2000).

Soon after the identification of the exon-intron structure of most eukaryotic genes, the first disease-causing mutations causing aberrant splicing, were identified (Busslinger et al., 1981; Orkin et al., 1981) and are now classified as a distinct category of mutations. Splicing is a complex process which requires a multitude of coordinated *trans*-acting splicing factors recognizing certain sequence motifs within the transcript and facilitating the transesterification reaction to join cognate exons and the excision of the intronic sequence in between. At the level of the transcript sequence there are some obligate and highly conserved features such as splice donor and splice acceptor sites which are necessary but not sufficient and require additional less conserved *cis*-acting sequences within the exon or the flanking intron sequence to direct the splicing process and regulate AS. This complex interplay comes at the expense of the multiplicity of sequence variants which induces perturbations in transcript splicing and eventually disease.

In this review, we provide an overview on the current knowledge about the impact of splicing mutations in inherited retinal diseases (IRDs) for both the expert and the clinical audience with emphasis on our own experiences and examples from recently published and unpublished data from our group.

2. Splicing process and machinery

The central dogma of molecular biology describes the directional flow of information from DNA to RNA to protein. In a first step the DNA sequence is transcribed into an initial, unprocessed RNA molecule termed precursor messenger RNA (pre-mRNA). For most transcripts, this pre-mRNA needs to be processed before the translation into a polypeptide sequence can commence. More precisely, the coding elements (exons) need to be recognized and joined together while the non-coding elements (introns) are removed from the immature transcript (Gilbert, 1978). This process is called splicing. Introns are also present in non-protein-coding genes, yet to a lesser extent in number and frequency compared to protein-coding genes (Piovesan et al., 2016).

2.1 Splice sites and the spliceosome

The splicing reaction generally takes place in the nucleus through the action of a ribonucleoprotein complex termed the ‘spliceosome‘ which contains several hundreds of proteins associated with five small nuclear RNAs (snRNAs). Since the spliceosome is not a preformed catalytic core but assembles on the pre-mRNA, its components are disassembled and recycled for reuse after each splicing reaction. Most eukaryotic introns (~99.5%) are spliced by the major spliceosome, which comprises U1, U2, U4, U5 and U6 snRNAs, whereas a small fraction of introns are recognized by the minor spliceosome that is linked to U11, U12, U4atac, U5 and U6atac snRNAs (Turunen et al., 2013).

Specificity of the splicing reaction mainly relies on the recognition of conserved sequences at the exon-intron boundaries to which the snRNAs bind. The splice site located at the upstream exon-intron boundary (splice donor site or 5´ splice site) comprises an almost invariable GU dinucleotide while the splice site located at the downstream intron-exon boundary (splice acceptor site or 3´ splice site) contains a similarly conserved AG dinucleotide (Fig. 1). Approximately 99% of mammalian introns obey to this GU-AG rule (Burset et al., 2000). However, in some instances non-canonical GC-AG and AU-AC dinucleotides represent genuine outermost intron boundaries as well (Shapiro & Senapathy, 1987). Besides these canonical dinucleotides there is some preference for nucleotide occupancy at neighboring sites in the intron as well as the exon sequence, although its conservation is less strict and functional relevance of variants at these sites is less predictable. In addition to the splice sites, a so called branch point (BP) site and a polypyrimidine tract (PPT) are crucial for the splicing process. The

BP is a short conserved sequence containing an adenosine residue between 10 and 50 nucleotides upstream of the splice acceptor site (Gao et al., 2008) while the PPT comprises 15 to 20 pyrimidine nucleotides, and is located between 5 and 40 nucleotides upstream of the splice acceptor site (Reed & Maniatis, 1985).

Two sequential transesterification reactions are required to join exons and release of the interjacent intron. In a first step, the BP adenosine residue serves as the nucleophile for the first transesterification reaction with the intron's 5' terminal nucleotide resulting in a lariat structure. Subsequently, the free 3' OH of the upstream exon acts as a nucleophile to induce a transesterification with the 5' phosphate group of the first nucleotide of the downstream exon which joins the exons and releases the intronic lariat. The splicing reaction is initiated by U1 snRNP binding to the splice donor site, followed by U2 snRNP interactions at the BP sequence and finally U4, U5 and U6 snRNP interactions near both splice sites. A series of conformational changes enables the joining of adjacent exons and the removal of the intron lariat (Ruskin et al., 1984). Details of the splicing process alongside intelligible illustrations are delineated in numerous excellent reviews (e.g. Matera & Wang, 2014; Scotti & Swanson, 2015).

2.2 *Cis*-acting regulatory elements and *trans*-acting splicing factors

The splicing process yields mature mRNA which serves as template for protein synthesis or alternatively, in the case of non-protein-coding genes, processed transcripts which exert their function as RNA molecules. snRNAs and proteins (‘*trans*‘ factors) assemble into the functional spliceosome in the context of additional *cis*-acting sequence elements located within the pre-mRNA itself (Black, 2003; Matlin et al., 2005). These *cis*-acting sequences are short motifs much more degenerate than the conserved splice sites. Their classification is based on their activity – either stimulating or repressing splicing efficiency, and on their location – either in exonic or intronic sequences. Exonic splicing enhancers (ESEs) are thought to recruit serine-arginine rich (SR) proteins which interact with the basal spliceosome and promote the use of nearby splice sites, thereby facilitating the inclusion of exons in which they reside (Blencowe, 2000). In contrast, exonic splicing silencers (ESSs) inhibit the use of adjacent splice sites by interacting for instance with members of the heterogeneous nuclear ribonucleoproteins (hnRNP) family (Wang et al., 2006). Equivalent splicing signals are found in the intronic regions. These are coined as intronic splicing enhancer (ISE) and intronic splicing silencer (ISS) elements. However, this categorical classification is somewhat oversimplified and the true underlying mechanistic rules go beyond these two features. For instance, some splicing factors do not have discrete functional roles yet recognize *cis*-regulatory elements that are able to induce antagonistic or ambivalent effects depending on their location (Zhou et al., 2014), sequence context (Rimoldi et al., 2013) or interaction with additional factors (Cáceres et al., 1994; Huelga et al., 2012).

Splicing enhancers and silencers also play an important role in tissue-specific alternative splicing. This mechanism has been intensively studied in muscle or brain. Moreover, predominant expression of the *RPGR*^{ORF15} transcript isoform in photoreceptors is mediated by multiple repeat units of ESEs present in the purine-rich region of this exon (Hong & Li, 2002). For a review on regulation of alternative splicing during tissue development and cell differentiation refer to Baralle & Giudice, 2017.

3. Disease-causing splice variants in inherited retinal disease

Splicing defects have a profound impact on protein sequence, structure and function and have been reported for almost all known hereditary diseases (Wang & Cooper, 2007). As of May 2020, the Human Gene Mutation Database (HGMD®) lists 282,895 unique mutations reported in 11,031 human genes. Missense and nonsense mutations are the most frequent variant categories and together account for 58% of mutations, followed by small deletions (14.5%). Splicing mutations represent the third largest category with 8.8%. Gross deletions and small insertions account for 7.2% and 6.1%, respectively. Other mutation types such as complex rearrangements, regulatory, or repeat variations are comparably rare. The figures are similar for the group of IRDs (Berger et al., 2010; Roosing et al., 2014). Of note, the spectrum of disease-linked mutations of some genes is dominated by splice mutations. A well-known example is the *VCAN* gene that is associated with Wagner syndrome and erosive

vitreoretinopathy (Mukhopadhyay et al., 2006). To date, the vast majority of pathogenic variants identified in *VCAN* affect the splice acceptor or splice donor sites of introns 7 and 8 (Kloeckener-Gruissem & Amstutz, 2016).

Mutations affecting splicing can occur at both acceptor and donor splice sites but seem to be more prevalent at donor sites. Table 1 shows a meta-analysis of splicing mutations in 20 frequently mutated genes in non-syndromic retinal disease sourced from HGMD®, indicating that donor splice site mutations are more prevalent than acceptor splice site variants (ratio 1.3:1). This ratio is similar to that observed by Krawczak and colleagues (2007), who investigated 478 splice mutations in 38 different human genes and found a ratio of donor to acceptor splice site mutations of 1.5:1. The reason for this imbalance is unknown.

Splicing defects are not only caused by mutations at the canonical GU-AG splice sites but also by genetic variations in exonic and intronic splicing enhancer and silencer motifs. In addition, mutations can activate or enhance cryptic splice sites that are located far away from the annotated exons (‘deep intron mutations‘). Apart from genetic variants in *cis*-acting regulatory elements, mutations can affect the functionality of *trans*-acting factors. Mutations affecting the BP and PPT are very rare. While no example related to retinal disease is known so far, another ocular phenotype has been linked to this kind of mutations: an intronic mutation in the lariat branchpoint sequence of the *LCAT* gene was shown to cause fish-eye disease which is characterized by corneal opacification (Kuivenhoven et al., 1996).

Figure 2 shows different types of splicing mutations which, in the following, are discussed in more detail attending to the type of splicing defect elicited by the mutation and/or the location of the mutation with respect to exon-intron definition, alongside examples illustrating each category based on our own work and that of others.

3.1 Mutations leading to exon skipping

Exon skipping is the most frequent consequence of splice site mutations (Nakai & Sakamoto, 1994). Of note, the same mutation can result in different aberrantly spliced transcripts lacking either one or multiple exons. This has been shown for instance for a splice site mutation in the *CDHR1* gene associated with cone dystrophy (Stingl et al., 2017). This gene encodes a member of the cadherin superfamily of calcium-dependent cell adhesion molecules and plays a role in outer segment disc morphogenesis (Rattner et al., 2001; Rattner et al., 2004). We performed an *in vitro* splicing assay for the c.2040+5G>T variant (affecting the splice donor site of exon 15) in *CDHR1* and found two aberrant transcripts in cells transfected with a minigene construct expressing the mutant allele. Subsequent sequencing of reverse transcription PCR (RT-PCR) products showed that the minor aberrant transcript lacked exon 16 whereas the major aberrant transcript lacked both exon 15 and 16 (Fig. 3). Of note, cells transfected with the mutant allele showed also a small amount of correctly spliced transcripts. It is not uncommon that splice site mutations produce small amounts of normal spliced mRNA in addition to aberrant transcripts. Such mutations are described as being not fully penetrant, as hypomorphic, or as “leaky”. They can be unmasked as being deleterious if occurring in compound heterozygous state with a second pathogenic allele.

Another example of differential exon skipping is found in the *ABCA4* gene, which belongs to the ATP-binding cassette (ABC) transporters. ABCA4 functions as a flippase in photoreceptor outer segments that translocates retinaldehyde conjugated to phosphatidylethanolamine across outer segment disc membranes (Weng et al., 1999). Transcript analyses of photoreceptor precursor cells and fibroblasts derived from patients harboring the recurrent c.5461-10T>C variant in *ABCA4* revealed the presence of transcripts that either skipped exon 39 alone or exon 39 and exon 40 together (Sangermano et al., 2016; Aukrust et al., 2017). The molecular mechanism of differential exon skipping (i.e. skipping of one or more exons) is not quite understood but most probably can be attributed to non-sequential intron removal that is influenced by disrupted splice sites (Pulyakhina et al., 2015; Takahara et al., 2002). The consequence of exon skipping on the protein depends on whether the reading frame is disturbed or not.

If the reading frame is continuous, the polypeptide will lack amino acids that might or might not be crucial for protein function. It has been postulated that skipping of exons encoding >60% of a protein domain is very likely to eliminate the functionality of the domain with which these exons overlap (Magen & Ast, 2005). If the reading frame is altered, there is a high probability of the formation of a premature termination codon (PTC), thereby leading to the degradation of the transcript by the nonsense mediated decay (NMD) process. NMD is a cellular surveillance mechanism that recognizes PTCs in a transcript and, instead of translating such transcripts, targets them for degradation (Frischmeyer & Dietz, 1999).

3.2 Mutations leading to intron retention

Intron sequences constitute approximately 25% of the human genome (Sakharkar et al., 2004). This vast amount of sequence by chance comprises motifs that are very similar or even identical to the consensus motifs of splice donor or splice acceptor sites. Such motifs are called cryptic splice sites. Activation of an intronic cryptic splice site in the proximity of a canonical exon can lead to the inclusion of intronic sequences into the mRNA (Krawzak et al., 2007). This process is often referred to as intron retention or exonification. Factors that favour intron retention are weak splice sites, short intron length and particular densities of regulatory elements (Sakabe & de Souza, 2007). The consequence of intron retention at the protein level depends on whether the reading frame is disturbed or not. If the reading frame is continuous, the intronic sequence might be translated, thereby becoming a part of the polypeptide, but the incorporation of additional amino acids might impact the function of the protein. However, most transcripts retaining introns contain one or more PTCs, which are recognized to initiate degradation by NMD (Wong et al., 2016).

We found a patient diagnosed with achromatopsia to be homozygous for a splice site mutation c.82+5G>T in the *ATF6* gene. ATF6 plays a role in the unfolded protein response pathway and was identified as the first gene causing achromatopsia which do not encode a component of the cone phototransduction cascade (Kohl et al., 2015). The c.82+5G>T variant is predicted to weaken or abolish the canonical donor site of exon 1 according to three different *in silico* tools (see chapter 4.1 for details), and a cryptic donor site is predicted 88 bp downstream of the canonical donor site of exon 1 (Fig. 4A). Since *ATF6* is expressed in blood lymphocytes, we could directly assess the impact of the c.82+5G>T variant on splicing by isolating RNA from a patient's blood sample and synthesizing complementary DNA (cDNA). Subsequent sequencing of RT-PCR products showed that the c.82+5G>T variant leads to exonification of 88 bp of intron 1 which confirmed the usage of the predicted cryptic donor site (Fig. 4C). Exonification of these 88 bp results in 38 novel amino acids followed by a PTC (Fig. 4D). Detection of such transcripts in blood lymphocytes indicates that a decent amount of mutant transcripts escapes NMD in these cells. As can be seen in Fig. 4B, the c.82+5G>T variant is leaky (i.e. produces residual amounts of correct transcript), suggesting that the patient may also produce small amounts of wild-type ATF6A polypeptide at least in blood lymphocytes. Whether the c.82+5G>T variant is fully penetrant in photoreceptors is not known.

Another well-studied example of intron retention is found in the *RPE65* gene which encodes a retinoid isomerohydrolase (Redmond et al., 1998) and was identified as a disease gene in patients with Leber congenital amaurosis (LCA; Gu et al., 1997). The c.246-11A>G variant in *RPE65* has been shown to result in the exonification of ten nucleotides of intron 3, thereby inducing a translational frame shift and creating a PTC nine amino acids downstream (Tucker et al., 2015). The authors could show that the majority of the c.246-11A>G transcript is lost due to NMD. If a cryptic splice site is located within an exon and this site is activated by a mutation that reduces or abolishes recognition of the natural splice site, the consequence is exon truncation. Again, the consequence at the protein level is dependent on the continuity of the reading frame as described above.

3.3 Mutations leading to the activation of pseudoexons

If cryptic splice sites are arranged in a way that places potential acceptor and donor sites spaced apart by the length of a typical exon (i.e. <200 bp; Sakharkar et al., 2004), these can define non-annotated, or ‘cryptic exons’ (Kapustin et al., 2011). Cryptic exons are also often referred to as ‘pseudoexons’ (PE).

Normally, cryptic splice sites are not recognized by the spliceosome, even if they exhibit high consensus values (Krawczak et al., 1992). Exons are only recognized as such if they are neighbored by a plethora of supplementary *cis*-acting elements that recruit and bind *trans*-acting factors. This redundancy of the exon definition mechanism contributes to the high fidelity of the splicing process. However, if a mutation creates or strengthens a deep intronic cryptic splice site, a potential complementary splice site in the proximity can be activated, resulting in pathological inclusion of a PE in the mRNA. Cryptic splice site activation events are not restricted to deep intronic regions but can also occur close to the ordinary splice sites. Alternatively, PE inclusion may result from the creation of an ESE or the disruption of an ESS as a consequence of an intronic nucleotide change. If the inclusion of the PE in the transcript retains the reading frame, it becomes part of the protein. In contrast, if a PE introduces a PTC, the resulting transcript will most probably be targeted to NMD. One of the most common and well-known deep intronic changes related to retinal disease is the c.2991+1655A>G variant one of the most frequent mutations in the *CEP290* gene responsible for Leber congenital amaurosis (den Hollander et al., 2006). The c.2991+1655A>G variant is located within intron 26 and creates a strong splice donor site that leads to the insertion of a cryptic exon of 128 bp encoding a PTC. With respect to retinal disease, deep intronic variants that lead to aberrant splicing by means of PE inclusion have also been described for *ABCA4* (for review, see Cremers et al., 2020), *CHM* (van den Hurk et al., 2003; Carss et al., 2017), *CNGB3* (Weisschuh et al., 2020; Bronstein et al., 2020), *OFD1* (Webb et al., 2012), *PRPF31* (Rio Frio et al., 2009) *RPGRIP1* (Jamshidi et al., 2019), *USH2A* (Vaché et al., 2012; Liquori et al., 2016), *USH3A* (Khan et al., 2017), and *PROM1* (Mayer et al., 2016).. The gene product of the latter is primarily found in photoreceptor outer segment disc membranes and is thought to play a structural role (Yang et al., 2008). We identified a deep intronic variant c.2077-521A>G in *PROM1* in homozygous state in two siblings with cone-rod dystrophy originating from a consanguineous background (Mayer et al., 2016; Fig. 5A). *In silico* assessment of the c.2077-521A>G variant using three splicing algorithms predicted the creation of a novel donor site (Fig. 5B). Functional analyses to test whether the c.2077-521A>G variant has an effect on splicing were performed by using a heterologous splicing assay in human embryonic kidney (HEK) 293T cells and in murine 661W cells, the latter displaying characteristics of photoreceptors. We could demonstrate that the c.2077-521A>G variant leads to the insertion of a 155-bp PE into the transcript (Fig. 5C+D). If translated, this would lead to an insertion of 22 novel amino acids followed by a PTC (Fig. 5E).

ABCA4 is by far the gene with the greatest number of described deep-intronic variants and such variants may constitute of up to 5% of all *ABCA4* disease alleles (Cremers et al., 2020). This is not surprising since intronic sequences cover 94.5% of the 128-kb *ABCA4* sized gene and one would expect a considerable number of motifs that resemble splice sites in this vast amount of non-coding sequence. Until recently, approximately 25% of Stargardt cases remained genetically unsolved because only a single heterozygous mutation was identified in the *ABCA4* gene (Zernant et al., 2014; Zernant et al., 2017). A relatively high percentage of monoallelic patients after routine screening is a strong indicator of the presence of pathogenic variants outside the coding regions. Indeed about 10% of mono-allelic Stargardt patients were demonstrated to carry deep-intronic variants (Braun et al., 2013; Zernant et al., 2014; Bax et al., 2015; Zaneveld et al., 2015; Schulz et al., 2017; Sangermano et al., 2019; Bauwens et al., 2019).

Of note, the majority of deep-intronic variants identified in retinal disease have been shown to create or strengthen cryptic splice donor sites. This imbalance of cryptic donor and acceptor sites has been described before (Dhir & Buratti, 2010) and is in line with the observation that canonical donor splice site mutations are more prevalent than canonical acceptor splice site variants (Krawczak et al., 2007).

3.4 Exonic splicing mutations

The potential impact of exonic variants on splicing is usually underestimated since it is common practice to only judge or prioritize variants based on their influence on the amino acid sequence. However, according to computational analysis, 25% of all variants within exons may affect splicing (Lim et al.,

2011). Hence, there is an enormous number of variants awaiting characterization at the transcript level. In particular, mutations that affect the first and the last three exonic positions have to be taken into consideration since they are an integral part of the acceptor and donor splice site consensus sequences. Surprisingly, few reports of such variants with an attributed and validated splicing defect have been published for genes associated with inherited retinal disease (IRD), most likely because these variants are more difficult to be properly scored as candidate splicing variants by the common bioinformatic tools. A recent report from our group illustrates this category of splicing mutation: a patient diagnosed with retinitis pigmentosa carrying an in-frame deletion and a silent substitution (c.783G>A/p.P261P) in the *CDHR1* gene in compound heterozygous state (Stingl et al., 2017). This synonymous substitution changes the last nucleotide of exon 8 without changing the encoded amino acid residue. The last exonic nucleotide of an exon is often guanine, and mutations at this position have been shown to result in splicing defects (Cartegni et al., 2002). Applying a heterologous splicing assay, we could demonstrate that the c.783G>A variant in *CDHR1* causes skipping of exon 8 (Fig. 6; Stingl et al., 2017). The reading frame of the transcript is not predicted to be changed by the skipping of exon 8. Therefore, the translated protein lacks 48 amino acid residues which is predicted to interfere with the formation of the second and third cadherin repeat and likely impairs protein function. A further example from our group is represented by the c.2368G>A variant in *PDE6C* associated with autosomal recessive achromatopsia. The c.2368G>A variant affects the first base of exon 21 and induces a partial exon skipping event in approximately 40% of transcripts. Skipping of exon 21 leads to a frameshift and formation of a PTC while the remaining correctly spliced transcript translates into a mutant PDE6C harbouring the E790K substitution (Chang et al., 2009). Another well-known example of an exonic variant leading to partial missplicing is the frequent c.2588G>C variant in *ABCA4.* This variant at the first nucleotide of exon 17 yields two splicing isoforms at the exon 16-17 junction with about equal abundance, one that is correctly spliced and deduced to encode a G863A substitution and a another which uses a splice acceptor site three nucleotides downstream and causes a single amino acid deletion (Maugeri et al., 1999). A more systematic analysis has been reported by Aparisi and colleagues (Aparisi et al., 2013) who investigated mutations at the outermost nucleotides in exons of several Usher syndrome type I genes by means of minigene assays and transcript analysis of nasal epithelial cells from patients. More recently, Soens and colleagues investigated a larger series of such splicing mutations due to mutations at the exon extremities of genes linked to IRD applying minigene assays (Soens et al., 2017). In most instances, exon skipping was observed and less frequently the activation of a novel splice acceptor or donor site close by and thus, the inclusion of an extension or shortening of the very exon in the mRNA.

Apart from mutations affecting the splice site consensus sequences, mutations may also affect exonic splicing regulatory elements. An excellent example for this category is the *BEST1* gene. Mutations in *BEST1* are associated with the phenotypic spectrum of ‘bestrophinopathies’, including Best disease, autosomal dominant vitreoretinochoroidopathy (ADVIRC) and autosomal recessive bestrophinopathy (ARB) (Johnson et al., 2017). A number of variants located at exon internal sites such as the synonymous variant c.102C>T in exon 2 which introduces a cryptic donor site about 50 bp upstream of the exon 2 splice donor site and the c.256G>A variant at nucleotide position 9 in exon 4 observed in patients with ARB and ADVIRC, respectively, have been shown to induce splicing defects through skipping of the corresponding exons (Yardley et al., 2004, Davidson et al., 2010). Sporadically exon internal mutations which induce splicing defects have been reported for some other IRD genes. This includes the c.1430G>A variant in the *RPE65* gene which alters the recognition of the genuine weak acceptor site of exon 13. The variant leads to a frameshift and PTC as has been demonstrated by minigenes in both human and mouse cells as well as by direct transcript analysis *in vivo* (Li et al., 2019). Further examples include a 1 bp deletion, c.2299del, in exon 13 of *USH2A* which induces skipping of one or two exons (Lenassi et al., 2014), the synonymous c.1359C>T variant in *CHM* (da Palma et al., 2020) and the c.5237G>A variant in *CDH23* in Usher syndrome type 1. The latter creates

a novel acceptor site and results in a 51 bp deletion in exon 40, thereby shortening the protein by 17 amino acids (Becirovic et al., 2008).

Exon internal mutations may also exert dual functional effects. We have recently described a novel missense mutation, the c.1684C>T/p.R562W in *PDE6A* which was found in *trans* with the recurrently reported c.2053G>A/p.V685M mutation in a patient with retinal dystrophy (Sothilingam et al., 2015). PDE6A encodes the alpha-subunit of the catalytic heterodimeric of the rod phosphodiesterase and mutations in this gene are a common cause of autosomal recessive retinitis pigmentosa. Using a minigene assay we could demonstrate that the c.1684C>T variant located 45 bases apart from the closest exon-intron boundary induces a partially penetrant splicing defect with skipping of exon 13. However, we additionally observed a considerable fraction of correctly spliced transcripts. Functional analysis of the recombinantly expressed R562W mutant protein revealed a drastic reduction of the enzymatic activity to about 10% of the wildtype protein level (Sothilingam et al., 2015). Thus, we reason that there is a combined defect owing to reduced levels of full-length transcripts due to the splicing defect and a reduced enzymatic activity of the R562W mutant protein. Notably, the partial penetrant splicing defect holds true when native expression of *Pde6a* transcripts in the retina of a R562W-knockin mouse mutant is directly assessed suggesting conserved splicing determinants for this site in mouse and human (Sothilingam et al., 2015). Interestingly, the c.1684C>T/p.R562W variant of the murine *Pde6a* is homologous to the well-known murine *rd10* mutation, c.1678C>T/p.R560C, in the paralogous *Pde6b* gene in a region with strong sequence conservation between the paralogs (Fig. 7A). However, upon experimental analysis of *Pde6b* transcripts in *rd10* mutant mice, we did not found any evidence for a splicing defect affecting exon 13 (Fig. 7B; unpublished results). Bioinformatic sequence analysis using SpliceAid2 (Piva et al., 2012) predicts an ESS motif created by the c.1684C>T/p.R562W variant in *Pde6a*, whereas the opposite effect, the creation of an ESE, is predicted for the c.1678C>T/p.R560C (*rd10*) mutation in *Pde6b*. In *Pde6a* exon 13 there is a guanine three nucleotides downstream of the mutation which participates in the formation of the novel ESS motif (-TGGTGG-). Instead, the homologous position in *Pde6b* is occupied by a cytosine (c.1680C) which prevents the formation of the aforementioned ESS motif. The predicted ESE motif in the *Pde6b rd10* mutant sequence rather enhances exon inclusion, the opposite regulatory activity as that exerted by the motif created by the homologous mutation in *Pde6a*. The ESE motif is not created by the c.1678C>T/p.R560C alone. The combination of c.1678C>T/p.R560C together with two neighbouring synonymous (paralog-specific) changes, c.1680C and c.1683C in *Pde6b*, create this additional ESE and thereby, an enhancer-rich region (Fig. 7C). This example well illustrates the importance of the sequence context, including silent changes, in determining whether a variant putatively impairs exon recognition and inclusion or even reinforces genuine splice sites recognition.

3.5 *OPN1LW* and *OPN1MW* exon 3 haplotypes impair splicing

Another intriguing example of the importance of the sequence context of exonic variants has recently been elucidated for exon 3 variants in the *OPN1LW* and *OPN1MW* genes encoding the apoprotein of the human long-wavelength sensitive (LWS) and middle-wavelength-sensitive (MWS) cone photopigments. *OPN1LW* and *OPN1MW* are located on the X-chromosome and form a tandem repeat gene cluster with a single *OPN1LW* and one or more copies of the *OPN1MW* gene. Their expression is regulated by an upstream enhancer element (locus control region; LCR) which determines the mutually exclusive expression of LWS and MWS cone pigments in individual cone photoreceptors. The high level of sequence identity between *OPN1LW* and *OPN1MW* which extends to the introns and non-coding intergenic spacer between the gene copies favours unequal homologous recombination and, as a consequence, results in the gain or loss of opsin gene copies on the recombinant chromosomes and the formation of *OPN1LW-OPN1MW* hybrid genes which frequently underlie the common forms of red-green colourblindness (Nathans et al., 1986). Moreover, the high level of sequence conservation between *OPN1LW* and *OPN1MW* also fosters gene conversion, i.e. the non-reciprocal exchange of usually short sequences, between *OPN1LW* and *OPN1MW* copies during meiosis. This process results in an interchange of common polymorphisms between *OPN1LW* and *OPN1MW* and a unique pattern of

low linkage disequilibrium between these polymorphisms. Gene conversion between *OPN1LW* and *OPN1MW* has long been postulated at the level of populations and recently we could confirm the conversion of an exon 3 haplotype from *OPN1MW* to *OPN1LW* by comparing the sequence signatures within the lineage of a single pedigree (Buena-Atienza et al., 2016).

Remarkably, certain rare combinations of common variants in exon 3 of *OPN1LW* and *OPN1MW* have been repeatedly found in subjects with X-linked cone dysfunction disorders, Bornholm Eye Disease, X-linked cone dystrophy and X-linked incomplete achromatopsia (McClement et al., 2013; Mizrahi-Meissonier et al., 2010; Crognale et al., 2004) and a pathogenic effect was long thought to result from the combined interaction of the encoded amino acid exchanges. However, Ueyama and colleagues demonstrated by *in vitro* expression that such combinations of common variants (i.e. haplotypes) induce skipping of exon 3 during transcript processing (Ueyama et al., 2012). In the following, we and others could confirm and further extend the spectrum of deleterious rare exon 3 haplotypes in patients with X-linked cone dystrophy and blue cone monochromacy (Gardner et al., 2014; Buena-Atienza et al., 2016; Greenwald et al., 2017). The so called 'LIAVA' haplotype (a term historically referring to the amino acid residues at the non-synonymous variant sites; see Fig. 8A) comprises the following exonic variants: c.(453A>G; 457A>C; 465C>G; 511G>A; 513G>T; 521C; 532A>G; 538T>G) and a thereof deduced cone pigment variant with a certain combination of amino acid exchanges: p.[(=); M153L; (=); V171I; A174A; I178V; S180A], has been repeatedly shown to induce fully or close to fully penetrant exon skipping in minigene assays (Fig. 8B-C). Two further haplotypes, namely 'LVAVA' and 'MIAVA', are strongly deleterious yielding less than 10% of properly spliced transcripts. The skipping event shifts the reading frame and results in a PTC, presumably targeting the transcript for NMD.

This case is rather unique in IRD since a combinatorial effect underlies misregulation of splicing and determines the extent to which exon inclusion is disturbed. Depending on the actual haplotype the percentage of correctly spliced transcripts (Fig. 8C) varies over the entire range of mixtures. Although we estimated that approximately 20% of correctly spliced transcripts might suffice to retain proper cone function based on phenotypic evidence (Buena-Atienza et al., 2016), additional work is required to establish the threshold below which the splicing defect has a phenotypic or physiological effect.

We recently extended the experimental analysis of *OPN1LW* and *OPN1MW* exon 3 haplotypes through targeted mutagenesis and the implementation of a multiplexed parallelized splicing assay to investigate all theoretically possible haplotypes from the eight common variants in exon 3 with two possible alleles per variant site (2^8=256) (Fig. 8D; Buena-Atienza et al., manuscript in preparation). We observed a considerable number of novel haplotypes resulting in low (<10%) levels of correct splicing. The 'output tags' generated by this investigation may serve as functional predictors once such haplotypes are identified in subjects. By assessing all tested haplotypes based on the genotype at individual sites, we found that variants c.532A>G and c.538T>G have the highest individual impact on promoting exon 3 skipping (Fig. 8E; Buena-Atienza et al., manuscript in preparation).

The relevance of local interactions between exonic variants in the context of splicing regulation has been emphasized by the existence of elements coined as composite exonic regulatory elements of splicing (CERES). Interestingly, it has been noticed that some elements overlapping a particular variant or group of variants may be able to exert opposite activities concurrently, i.e. enhancing or inhibiting). Evidence has been provided by investigations on *CFTR* exon 12 (Pagani et al., 2003) and *BRCA1* exon 11 (Tammaro et al., 2014). Certain interactions within localized regulatory signals in the three sub-divisions of the defined *OPN1LW* and *OPN1MW* exon 3 haplotype might resemble that of CERES. Yet, the interactions between non-overlapping *cis*-acting elements at distant regions might represent a new form of interplay between splicing determinants.

3.6 Mutations affecting *trans*-acting splicing factors

Trans splicing factors play an important role in the pathogenesis of retinal degeneration and have been reported to be the second most common cause of autosomal dominant retinitis pigmentosa (ADRP;

Sullivan et al., 2006). Among the 29 ADRP genes identified to date (source: RetNet, https://sph.uth.edu/retnet/), six code for ubiquitous core snRNP proteins (*PRPF3*, *PRPF4*, *PRPF6*, *PRPF8*, *PRPF31*, and *SNRNP200*) and one for a non-snRNP splicing factor (*RP9*), respectively. Mutations in the spliceosomal proteins that are associated with retinal disease have been reviewed elsewhere (Růžičková & Staněk, 2017; Krausová & Staněk, 2018). In short, pre-mRNA splicing factors, mainly PRPF8, function as scaffold proteins that hold the pre-mRNA in the correct orientation. Mutations in these components affect the stoichiometry and kinetics of spliceosome assembly (Tanackovic et al., 2011). RNA helicases such as SNRNP200 are the driving forces in the dynamic remodeling of the spliceosome, and mutations compromise the conformational rearrangements that are vital for the assembly (Zhao et al., 2009).

Probably the most interesting aspect of the involvement of splicing factors in the pathogenesis of ADRP is that mutations in the abovementioned genes, despite their ubiquitous expression, only affect retinal cells. The most likely explanation is that retinal cells, due to the high turnover-rate of photoreceptor outer segments (Bok & Young, 1972), have a high demand for splicing and are more sensitive to spliceosomal component deficiency than other tissues. This is reflected by the fact that the retina contains unusually high amounts of spliceosome components (Tanackovic et al., 2011).

The pathomechanism of *PRPF31*-linked retinal degeneration has been recently studied in much detail in patient-derived fibroblasts and induced pluripotent stem cells (iPSCs), and therefrom differentiated retinal organoids and retinal pigmented epithelium (RPE) cells. This study revealed that the alteration of the retinal-specific splicing program through mutated PRPF31 leads to extensive intron retention, and is related with the disruption of RPE and the cilium of photoreceptors (Buskin et al., 2018).

4. Validation of splicing mutations

Genetic testing is almost exclusively done at the genomic DNA level. Variant interpretation in terms of their functional consequences therefore mostly relies on general rules (i.e. universal genetic code, canonical splice sites, gene models) and 'experienced' interference (e.g. kind of mutation, kind of amino acid exchange, evolutionary conservation, etc.). However, the 'language' of splicing is equivocal and remains not fully understood which makes interpretation of variants putatively affecting splicing or the recognition of such variants still challenging. Variants affecting the highly conserved GU and AG splice acceptor and splice donor dinucleotides nearly invariably induce a splicing defect and therefore are generally accepted as deleterious mutations without further validation. However, the consequence of the individual mutation, whether it induces exon skipping or activates an alternative cryptic splice site in neighbouring exon or intron sequences thereby remains uncertain.

Variants in the vicinity but outside the canonical dinucleotides are less straightforward to interpret since these sites exhibit considerable sequence variation. Such variants have to be considered as putative splice variants and their effect on splicing needs to be validated. Direct analysis of native transcripts is the most reliable approach but is limited to those genes that are expressed in accessible tissues or where such tissue or cell types can be differentiated *in vitro* from iPSCs. Validating variants by *in vitro* assays is an alternative if source material is not available, but is laborious and therefore difficult to implement in routine workflows of diagnostic laboratories. While *in silico* prediction tools have become essential parts in workflows for analyzing putative splice variants, they can lead to misinterpretation and false predictions. In the following, several methodologies for the detection and analysis of splice mutations are introduced.

4.1 *In silico* prediction tools

Bioinformatic approaches to predict the effect of a variant on splicing use algorithms based on sequence conservation and thermodynamics. It is important to note that both parameters are descriptive of spliceosome recognition but not deterministic to assess splice potential (Lee et al., 2017). Application aspects, advantages and disadvantages of different *in silico* prediction tools have been reviewed elsewhere (Jian et al., 2014) and a comprehensive list of prediction algorithms can be found in

a recent review on splicing defects in human genetic disease (Abramowicz & Gos, 2018). In general, all algorithms compare the variant sequence with the reference (or consensus) sequence and provide a prediction score which reflects the strength of the splicing signal. Most algorithms are based on position weight matrices, which are calculated on the basis of a collection of known splice site sequences (Shapiro & Senapathy, 1987). Considering its frequency, these matrices assign a weight to each nucleotide within the analysed sequence. While the position weight matrix-based method assumes statistical independence between positions**,** maximum entropy models take into account the dependencies between positions within the sequences being analysed (Yeo & Burge, 2004). Other tools make use of neural networks based on the frequencies of dinucleotides within a sequence and use a dataset of known splice sites to train programs to identify splice sites (Reese et al., 1997). The different tools should always complement each other since they differ from each other with respect to the consensus sequences that are used for the comparison with the input sequences, statistical models and the training datasets that were used for machine learning.

As outlined in 3.3, we have identified a deep intronic variant in *PROM1* that causes PE inclusion into the transcript (Mayer et al., 2016). Figure 5B shows splice site prediction scores from three algorithms for the deep intronic variant. All algorithms, namely Human Splicing Finder (HSF) matrices (Desmet et al., 2009), MaxEntScan (Yeo & Burge, 2004) and NNSplice (Reese et al., 1997), predict the creation of a novel cryptic donor site with relatively high scores. Moreover, a cryptic acceptor site 155 nucleotides upstream of the new donor was predicted by all three algorithms. Making use of a heterologous splicing assay we could indeed verify the inclusion of a PE of 155 nucleotides into the transcript (Fig. 5C+D), in concordance with the bioinformatic predictions.

The major drawback of *in silico* tools is that the final interpretation of the results remains with the user, meaning that he or she needs to define an arbitrary cutoff value as to whether the variant of interest is (likely) causing a splicing defect or not. Since there are no consensus guidelines on the selection of cutoff values, the set threshold may turn out to be too restrictive (i.e. leading to false negative results) or too lax (i.e. providing a false positive result). To demonstrate how difficult it can be to define a cutoff value, we have analysed all canonical splice sites of the *CNGB3* gene using three different splicing algorithms. *CNGB3* encodes the beta subunit of the cyclic nucleotide-gated ion channel in cone photoreceptors and is the major achromatopsia gene in Europe and the US. Table 2 shows the prediction scores for all 18 exons of *CNGB3* obtained with HSF matrices (Desmet et al., 2009), MaxEntScan (Yeo&Burge, 2004) and NNSplice (Reese et al., 1997). Two important observations can be made. First, both the acceptor and donor splice site of exon 13 are comparably weak. One of the three algorithms does not even recognize the exon as such which makes it impossible to obtain prediction scores for any nucleotide substitution at the splice sites. Nevertheless, exon 13 is recognized as an exon and spliced into the mRNA by the splicing machinery of cone photoreceptors, demonstrating that predictions scores can be misleading. Second, most prediction scores are lower for acceptor sites than for donor sites. It has to be emphasized that assessing acceptor site variants is more difficult than donor site variants since they display a greater diversity of nucleotide combinations. This is due to a longer length of the acceptor consensus sequence (-3 to -14 compared to +3 to +6 for the donor consensus) and a certain flexibility of the PPT (Lee et al., 2017). It has to be noted, though, that substitutions of uracil and cytosine nucleotides within the PPT may have an impact on splicing none the less. The third most frequent *ABCA4* variant, c.5461-10T>C, has only a subtle effect on splicing according to *in silico* predictions. However, as outlined in 3.1, it exerts a distinct splicing pattern when assessed in photoreceptor progenitor cells of patients harboring this variant in homozygous or heterozygous state (Sangermano et al., 2016). Likewise, a strong effect of a thymine to cytosine change in the PPT has been demonstrated for the c.4004-5T>C variant in the *VCAN* gene, which has been found to result in exon 8 skipping in patients with Wagner disease (Mukhopadhyay et al., 2006).

Beyond canonical splice sites and nearby consensus sequences *cis*-regulatory elements of splicing are much less conserved. Yet, virtually any gene variant located elsewhere could potentially disturb a

splicing regulatory motif and thereby cause aberrant splicing. Algorithms embracing this type of putative mutations were first based on the search for enhancers exploiting experimental evidence derived from functional systematic evolution of ligands by exponential enrichment such as ESE-finder (Cartegni et al., 2003). Many algorithms rely on hexamer motif libraries/datasets such as RESCUE-ESE or ESRSeq scores (Fairbrother et al., 2004; Ke et al., 2011) or decamers for the search of silencers (FAS-ESS; Wang et al., 2004). Whereas most of the latter algorithms search for discrete categories of motifs, EX-SKIP retrieves relative ratios taking into account densities of ESEs and ESSs. Several of these datasets are compiled in user-friendly tools such as HSF (Desmet et al., 2009) and SROOGLE (Schwartz et al., 2009). The abovementioned tools lack context dependence and newly developed algorithms such as HEXplorer have introduced some degree of sequence context by including six overlapping hexamers per assessed site (Erkelenz et al., 2014). EX-SKIP, Hexplorer and ESRSeq scores perform comparably well in terms of specificity and sensitivity for mutations for which splicing was assessed in patient tissue (Grodecká et al., 2017). HSF provides some hits for RNA-binding proteins, namely splicing factors, whose binding could be compromised to the predicted change at the *cis*-sequence level. However, the SpliceAid2 tool (Piva et al., 2012) is more informative and only includes sites recognized by RNA-binding proteins based on experimental evidence. Nevertheless, additional work considering specificity and redundancy for binding sites is needed. This tool has an option to include some context, for example being able to select among different cell lines, normal and cancer tissues. The recently published SPANR algorithm used data from 16 tissues (not including the retina) for training purposes (Xiong et al., 2015).

To summarize, the reliability of bioinformatic analyses to assess splicing variants is limited. The recognition of exon-intron boundaries is not only dependent on the consensus sequences of the splice sites but on a plethora of auxiliary *cis*-acting elements which display high nucleotide diversity that are difficult to assess. Bearing in mind the limitations of bioinformatic analyses, they should be used primarily to prioritize variants for further validation using *in vitro* assays and not as a standalone evaluation tool.

4.2 Direct transcript analysis

A straightforward way to assess the functional impact of splice variants is to analyze the RNA extracted from accessible patient tissues (i.e. lymphocytes, keratinocytes and fibroblasts), provided that the gene of interest is expressed in this tissue. The synthesis of DNA from an RNA template, *via* reverse transcription, produces cDNA which can be used as a template for amplification and sequencing. An example of direct cDNA analysis is given in chapter 3.2, where we describe the analysis of a splice variant in the *ATF6* gene using RNA isolated from patient´s lymphocytes (Fig. 4). One drawback of this approach is that accessible cells (e.g. blood lymphocytes or fibroblasts) do not necessarily reflect the situation in the affected cells (e.g. photoreceptors; Albert et al., 2018). This has to be considered not only when obtaining negative results. Variants that showed no effect on splicing for instance in blood cells could still have an impact in retinal cells considering that blood cells lack retina-specific splicing proteins (i.e. *trans* splicing factors). Likewise, variants might show a splicing defect in blood cells but not in retinal cells. Even murine models may not always recapitulate the situation in the human retina (Garanto et al., 2013). RNA analysis is not restricted to study selected variants, but can be applied to identify splicing variants on a genome-wide scale and thus may complement whole exome and whole genome sequencing as a diagnostic tool (Gonorazky et al. 2019). This is achieved by sequencing the entire transcriptome (i.e. the complete set of transcripts in a cell) and scanning for misspliced transcripts. RNA sequencing (RNAseq) is therefore an ideal tool to detect splice variants as it allows for the detection of both coding and non-coding variants and provides transcript level information for interpreting splice changes (Chaitankar et al., 2016; Gonorazky et al., 2019). In RNAseq, a population of RNA (total or enriched subpopulations) is converted to a library of cDNA fragments with adaptors attached to one or both ends. The adaptors provide target sequences for subsequent amplification and sequencing in a high-throughput manner. Indications of abnormal mRNA splicing are sequence reads that demonstrate exon skipping, the use of cryptic splice sites, and high levels of intron inclusion. A

variety of bioinformatics tools has been developed for RNAseq read mapping, and the detection and quantification of variant splice forms (for an overview see Marco-Puche et al., 2019; Mehmood et al., 2019). Reads that represent correct splicing map within an investigated exon and span the junctions between this specific exon and the two upstream and downstream flanking exons. Pseudoexons that are integrated into the transcripts can be seen in the same way. An exon skipping event is demonstrated by reads spanning the junction between two flanking exons. Intron retention can be identified by reads that extend into the intron. All these missplicing events can be visualized in RNAseq datasets with so called Sashimi plots (Katz et al., 2013) which are also implemented in the popular web application Integrative Genomics Viewer (Thorvaldsdóttir et al., 2013). An example is shown in Figure 6C. As outlined in 3.4, we could demonstrate that the putative splice variant c.783G>A/p.P261P in *CDHR1* causes skipping of exon 8 (Stingl et al., 2017). The variant most likely represents a hypomorphic allele and is found with an allele frequency of 0.3% in the normal population (https://gnomad.broadinstitute.org/variant/10-85962879-G-A). Since *CDHR1* is not expressed in accessible tissues, we analyzed the splicing defect caused by the c.783G>A variant in an *in vitro* splicing assay. In addition, we performed RNAseq with RNA from the retina of a healthy eye donor who was heterozygous for the c.783G>A variant and compared it with two retina samples that were homozygous for the wildtype G allele. As shown in Figure 6C, the Sashimi plots for both wildtype samples (WT1 in blue and WT2 in green) show only reads that span the junction between exons 7 and 8, and between exons 8 and 9, respectively. In contrast, the retina sample heterozygous for the c.783 G>A variant (HET in red) shows a substantial number of reads (n = 149) linking exon 7 to exon 9, demonstrating the skipping of exon 8 in this sample. More recently, Bronstein and colleagues demonstrated the feasibility of an entire transcriptome approach on iPSC-derived retinal organoids to identify a deep intronic mutation in *CNGB3* inducing PE inclusion (Bronstein et al., 2020). A potential limitation of transcript analysis from native tissue or organoids is downregulation of the mutant transcript(s) by NMD. Since many splicing mutations result in aberrant transcripts harbouring PTCs, these transcripts undergo NMD and may become strongly underrepresented in heterozygous state compared to the transcripts derived from a non-NMD allele (Bonifert et al., 2014). Therefore, any evidence for a bias in allelic expression, for instance in the allelic read frequencies of a coding single nucleotide polymorphism (cSNP), may provide a hint for the action of NMD and the presence of an allele with low expression or lowered levels of steady state transcripts. If possible, RNA isolated from an intermitting cell culture of primary cells (e.g. fibroblasts or lymphocytes) under conditions which blocks NMD such as the addition of general translation inhibitors like cycloheximide, puromycin or anisomycin should be preferred.

Another issue that needs to take into account is the effect of tissue or cell-specific splicing factors and differences in the pattern of alternative splicing. Beyond the well documented examples of retina-specific splice variants such as the *RPGR*^{ORF15} transcript, recent reports described retina-specific splicing defects such as the closely located deep-intronic mutations c.4539+2001G>A and c.4539+2028C>T in *ABCA4* which do not result in aberrant splicing in fibroblasts but only in photoreceptor precursor cells derived from iPSCs (Albert et al. 2018). Even more convincing are the results observed for the c.2991+1665A>G in *CEP290*, where the fraction of misspliced transcripts is fourfold higher than normal spliced transcripts in eye cup organoids compared to fibroblasts or undifferentiated iPSCs were these transcript isoforms are about equal abundant (Parfitt et al., 2016).

To summarize, direct transcript analysis is a straightforward approach to assess splicing changes but is hampered by tissue- or cell-specific expression of the gene of interest. The majority of genes associated with retinal disorders are not or only poorly expressed in non-ocular tissues (Khan et al., 2019). The inaccessibility of the human retina has forced researchers to take a detour by reprogramming patient skin cells into iPSCs and differentiating them into retinal cells expressing the gene of interest (Parfitt et al., 2016; Sangermano et al., 2016; Lukovic et al., 2015; Yoshida et al., 2014; Tucker et al., 2015). While this approach circumvents the problem of RNA degradation in primary tissues caused by NMD, it remains a relatively costly method that requires specific skills, thereby

placing it far away from representing a high-throughput technique to assess the impact of splicing variants.

4.3 Heterologous splice assays

Heterologous splice assays are a viable alternative to analyze the impact of a putative splice variant if direct transcript analysis of patient RNA is not feasible. This is the case if the gene of interest is not expressed in readily accessible tissues, or if the resulting transcript is degraded by NMD as a consequence of the presence of a PTC.

Detailed workflows to conduct heterologous splice assays have already been published (Desviat et al., 2012). Briefly, a genomic segment encompassing the variant of interest along with flanking sequences is PCR-amplified from patient genomic DNA and is cloned into a so called minigene plasmid vector. Following cloning, the resulting constructs in their wild-type and mutant versions are used to transfect eukaryotic cell lines, which are then analyzed with respect to vector splicing. Ideally, transfection with the wild-type minigene results in correct splicing (i.e. inclusion of the cloned exon(s) between the vector-derived exons), while transfection with the mutant construct results in aberrant transcripts.

Minigene vectors are splice reporter vectors (also known as exon-trapping vectors) that code for two exons with functional splice sites. The exons are separated by a segment which comprises a multiple cloning site in which the sequence of interest can be cloned. Initially, exon trapping vectors have been developed to rapidly identify and clone coding regions (Buckler et al., 1991). For the analysis of splice variants, the cloned segment should comprise not only the variant with flanking intronic sequence but at least one downstream and upstream exon and their corresponding introns. This allows for testing the variants in their genomic context. Using DNA of a subject heterozygous for the splice variant of interest enables co-amplification of the mutant and the normal allele and subsequent simultaneous cloning of the mutant and wildtype minigene construct. In case only patients are available which are homozygous for the mutant allele, the wildtype allele can be generated by site directed mutagenesis.

Both wildtype and mutant minigenes are used to transfect eukaryotic cell lines, making use of their splicing machinery. After RNA isolation, cDNA synthesis and RT-PCR with vector-specific primers (to ensure selective amplification of vector-derived transcripts), a first analysis of transcripts can be performed by agarose gel electrophoresis. If the inserted fragment has functional splice acceptor and splice donor sites, splicing should occur between the vector and insert sequences, thereby including the exonic region in the processed transcript. Any aberrant splicing (e.g. exon skipping or intron retention) can be assessed by comparing the size of RT-PCR products derived from the wildtype and mutant minigenes. Subsequent sequencing enables the exact delineation of the splicing defect.

A widely used vector for splicing assays is the pSPL3 vector, which contains a small artificial gene composed of an SV40 promoter that provides replication and transcription in eukaryotic cells, an exon-intron-exon sequence with functional splice donor and acceptor sites, and a late polyadenylation signal. The vector also contains an origin of replication and an ampicillin resistance gene. The cell line for minigene experiments is selected depending on the tissue where the gene is expressed, e.g., for photoreceptor-specific genes, 661W cells can be used which display characteristics of photoreceptor cells (Tan et al., 2004). However, if such cell lines are not available, any established cell lines such as HEK293T or HeLa can be used. As outlined in the previous chapter, we have successfully validated several putative splice variants using minigene assays based on the pSPL3 vector. Note that heterologous splice assays relying on minigenes transfected into cells in culture are sometimes referred to as *in vitro* splicing assays or *ex vivo* splice assays. The latter term is typically used by studies that include *in vitro* splice assays with *in vitro* transcribed RNA and without cells.

As mentioned before, putative splice variants should always be cloned in their genomic context, that is, with at least one upstream and one downstream exon. However, this is not always possible. While eukaryotic exons are usually small (<200 bp; Sakharkar et al., 2004), the intervening introns often encompass several kilobases (kb) of sequence, which precludes cloning into conventional exon

trapping vectors. Minigenes lacking the genomic context of a variant, i.e. flanking exons and ***cis***-acting elements, might not correlate with splice defects observed in patient derived cells (Rivera et al., 2000; Sangermano et al., 2016). Hence, a major drawback of minigene vectors is the limited size of DNA segments that they can accommodate. The pSPL3 vector itself is 6 kb and while we successfully inserted segments of 8 kb (unpublished results), this likely represents the maximum insert size. Bacterial artificial chromosome clones facilitate the generation of large, multi-exon wild-type splice vectors, and have been developed for the analysis of splice variants in the *ABCA4* gene (Sangermano et al., 2018). However, the maximum insert size of segments that have been cloned into this so called midigene vector is 11.7 kb, representing so far a modest advantage compared with the pSPL3 vector.

Minigenes other than exon-trapping vectors exclusively relying on endogenous splice sites of flanking exons (Fig. 8B), have been used for the splicing assessment of exon 3 in *OPN1LW* and *OPN1MW* (Ueyama et al., 2012; Gardner et al., 2014; Buena-Atienza et al., 2016; Greenwald et al., 2017).

To summarize, heterologous splice assays represent a viable choice to assess the effect of putative splice variants in the absence of patient cells, or if the gene of interest is not expressed in available somatic cells. Besides allowing validation of putative splice variants, minigenes can support monitoring of therapeutic strategies aimed at modulating splicing or correction of splicing defects (Matos et al., 2014). Size constraints of minigene constructs represent a major limitation to fully assess the effect of splice variants. In addition, cell lines used for transfection might lack expression of specific *trans* splicing factors. Hence, some heterologous splice assays might not recapitulate accurately the *in vivo* situation. In case of negative results (i.e. no splicing defect is observed), analysis of a larger segment and/or switching to a tissue-specific cell line should be considered.

5. Therapeutic approaches for splicing defects

Efforts in gene augmentation clinical trials for retinitis pigmentosa (Dias et al,. 2018) and other IRDs by subretinal injections as well as the potential of intravitreal injections have been reviewed recently (Garafalo et al., 2019). Splicing mutations which behave recessive or act through haploinsufficiency may be targeted in a similar manner to other kinds of hypomorphic or nullimorphic mutations by supplementation therapies such as recombinant enzyme therapy, metabolite supplements (e.g. oral 9-*cis*-retinyl acetate in LRAT und RPE65 deficiency), or adeno-associated-viral (AAV) vector-based delivery of a functional gene copy. In addition, certain approaches specifically targeting splicing defects have been developed. Although these approaches are generally limited to small groups or even individual mutations, they have the advantage of circumventing current limitations in transgene size for AAV-based supplementation therapy which hampers its application for many large IRD genes (*ABCA4*, *USH2A*, *EYS*, *CEP290*, and others). In the following we will briefly review the major therapeutic approaches for splicing defects and their application for IRD genes.

5.1 Sequence-adapted U1 and U6 snRNA

During initiation of the splicing process, U1 and U6 snRNP participate in the recognition of the splice donor site which involves base pairing between the U1 and to a lesser extent the U6 RNA moiety and the primary transcript. Harnessing this property, engineered (i.e. sequence adapted) U1 RNA which improves base-pairing with splice donor site mutations has been developed as a strategy to restore proper mRNA splicing. The principle concept of this correction approach goes back to early *in vitro* experiments on adenovirus E1A splicing (Zhuang & Weiner, 1986) and since then has been applied successfully in a variety of gene defects and *in vitro* disease models. Proof-of-concept for its feasibility *in vivo* has been demonstrated in mouse models for aromatic L-amino-acid decarboxylase deficiency (A>U at splice donor position +4), factor 7 coagulation deficiency (G>A at splice donor position +5) and familial dysautonomia (U>C at splice donor position +6) (Lee et al., 2016; Balestra et al., 2014; Donadon et al., 2018) with sequence adapted U1 snRNA gene sequences delivered by AAV. Targeting non-conserved splice donor sequences reduces potentially deleterious off-target alternative splicing events. Rogalska and co-workers studied off-target activity of a transgenic mouse expressing an U1 snRNA designed to correct *SMN2* exon 7 inclusion and observed in the spinal cord significant changes

in gene expression for only 12 genes and just one alternative splicing off-target (Rogalska et al., 2016). Nevertheless, off-target effects are a concern of this approach and need to be carefully assessed in every instance for the targeted tissue and cell type.

Hitherto, there are only a few studies which applied adapted U1 snRNA - alone or in conjunction with adapted U6 snRNA - to target splice mutations in IRD genes (namely *RHO*, *BBS1*, and *RPGR*) and demonstrated rescue of splicing defects in minigene assays as well as in patient-derived fibroblasts (Tanner et al., 2009; Glaus et al., 2011; Schmid et al., 2013). Moreover, Schmid and colleagues explored more systematically the potential of this strategy using a series of substitutions from position -2 (last but one nucleotide of the exon) to position +5 (fifth position of the intron) at the splice donor sequence of *BBS1* exon 5 which induced a splicing defect in a minigene assay in COS7 cells. While single-base adapted U1 snRNA specifically re-introducing a base-pairing between U1 snRNA and the transcript at the site of the mutation yielded partial rescue at some sites, a greater rescue efficacy or even complete rescue was observed with fully adapted U1 snRNA. Notably, mutations at the canonical -GU- splice donor dinucleotide were not rescued by this approach. Rescue of the +5 position was further improved by simultaneous application of both sequence-adapted U1 and U6 snRNA (Schmid et al., 2013).

5.2 Small molecules with splicing rescue activity

Several small molecules which act as general splicing modulators have been usually identified through compound library screening in the context of alternative splicing. Some of these approaches have been used to specifically screen for compounds with rescue effects on certain specific splicing defects due to mutations at non-canonical splice site positions or binding sites for auxiliary splicing factors. This led to the identification of kinetin, a plant cytokinin, a synthetic chloroaminopurine derivate named 'rectifier of aberrant splicing' (RECTAS) which suppresses exon skipping in c.2204+6T>C mutant transcripts of *IKBKAP* in familial dysautonomia (Slaugenhaupt et al., 2004; Yoshida et al., 2015) and several synthetic compounds for the treatment of spinal muscular atrophy (SMA) (Naryshkin et al. 2014; Palacino et al. 2015). SMA is principally caused by inactivating mutations or deletions in *SMN1*. However, disease severity is modulated by expression levels (i.e. copy number) of *SMN2*, a partially inactivated paralog of *SMN1*. *SMN2* carries a C-to-T substitution at the 6th position of exon 7 which generates an ESS and results in a large fraction of *SMN2* transcripts lacking exon 7 which fail to complement loss of *SMN1*. Many therapies for SMA thus aim to increase the fraction of full-length *SMN2* transcripts (Lorson et al., 2010). Two small molecule compounds derived from compound screens, risdiplam and branaplam, both pyridazine derivates formulated for oral intake, have been shown to increase *SMN2* exon 7 recognition and inclusion (Ratni et al., 2019). Both compounds are currently being tested in clinical trials with risdiplam showing tolerability and efficacy in phase II and phase III trials (Baranello et al., 2019). Some limited data from a first short-term study with orally administered kinetin in eight familial dysautonomia patients have been published which showed safety and tolerability and demonstrated a significant increase in exon inclusion in blood leukocytes (Axelrod et al., 2011). However, results on clinically relevant efficacy and long-term tolerability are still missing.

To the best of our knowledge none of these compounds have been tested in retinal dystrophies nor have compound screens been performed for splicing defects in IRD genes.

5.3 Antisense Oligonucleotides

Antisense oligonucleotide (AON) technology offers a flexible instrument for tailored interference with splicing. Antisense oligonucleotides (AONs) are synthetic single stranded nucleic acid molecules of 16 to 20 nucleotides in length which bind to target transcript sequences through base-pairing. Certain chemical modifications on the phosphate backbone such as phosphorodiamidate morpholino chemistry or of the sugar moiety (e.g. 2' OH ribose substitutions such as 2‘-Methoxyethyl) have led to an increase in base-pairing stability, reduced toxicity and improved pharmacokinetic properties of AONs. To interfere with mRNA splicing, AONs are designed to hybridize to crucial *cis*-acting sequence elements (e.g. splice donor or acceptor sites, or exonic/intronic splicing regulatory elements) thereby blocking the

binding of *trans*-acting splice factors and spliceosomal assembly. Two main strategies are applied to exploit AON technology for the treatment of splicing defects: (1) targeted elimination of a genuine exon through induced exon skipping, and (2) blockage of cryptic splice donor or splice acceptor sites.

The first strategy applies to the elimination of an in-frame exon which carries a deleterious mutation from the transcript thereby maintaining the reading frame at the expense of missing an internal sequence within the thereof translated polypeptide. The validity of this strategy is well appreciated by genetic mutations in the *DMD* gene where nonsense or frameshift mutations usually result in Duchenne Muscular Dystrophy while intragenic deletions which maintain the reading frame result in the much milder disease expression of Becker muscular dystrophy. The design and chemistry of AONs and modalities for AON-based exon skipping in *DMD* has been explored intensively and showed impressive efficacy in mouse and dog disease models (Alter et al., 2006; McClorey et al., 2006). Exon 51 skipping AONs (eteplirsen/EXONDYS51 and drisapersen/PRO051) which potentially target about 13% of the human Duchenne patient population - mostly through correction of the reading frame in patients with exon 50 deletions - have been tested via systemic delivery in clinical trials and showed tolerability and clinical benefit in prolonged mobility (Cirak et al., 2011; Goemans et al., 2011; Mendell et al., 2016; Goemans et al., 2016). Eteplirsen (tradename EXONDYS51®) has obtained marketing authorization for the United States by the US Food and Drug Administration (FDA) while the European Medicinal Agency did not approve the drug due to still limited evidence of efficacy.

Transferring this strategy to IRD, Barny and colleagues showed AON-based exon skipping of exon 36 in patient-derived dermal fibroblasts of subjects with the recurrent c.4723A>T/p.Lys1575X nonsense mutation in *CEP290* exon 36 and achieved more than 60% of exon skipping. While the abundance of CEP290 protein - mainly representing the internally truncated CEP290$^{\Delta 36}$ isoform - significantly increased, the authors noted alterations in the dynamics of centriolar satellite formation, and the proportion of ciliated cells tended to diminish as did axonemal length (Barny et al., 2019). In contrast, correct localization of the CEP290$^{\Delta 41}$ protein to the ciliary transition zone, and restoration of normal cilia length was demonstrated upon AON-based exon 41 skipping in kidney cells derived from a patient with Joubert Syndrome carrying the c.5668G>T; p.G1890*nonsense mutation (Ramsbotton et al., 2018).

The application of the AON-based exon skipping strategy relies not only on suitable in-frame exons but also on the functional competence of the shortened protein product. This is most likely the case at the extremities, i.e. the amino- and carboxy-terminal of proteins or in proteins with modular, repeated domains such those encoded by many Usher syndrome genes, *CRB1*, *EYS* and *CEP290*. However, this functional tolerance of internal truncations needs to be carefully tested in each individual case. Recently Zhang and co-workers demonstrated that a shortened human CEP290 ('miniCEP290') with a deletion of amino acid residues 580-1180 is able to complement and ameliorate the retinal degeneration in the Cep290^{rd16} model upon subretinal AAV delivery (Zhang et al., 2018). Notably, the Foundation Fighting Blindness recently announced a 7.5 million US-Dollar support for an AON-based therapy program for *USH2A* which also included exon skipping modules for exon 13 and exon 50 as developed at Radboud University, Nijmegen (patent WO 2016/005514 A1) and currently in early clinical testing (NCT03780257 on exon 13 skipping). It should be noted that concerns on the impact of the exon 13 skipping strategy has been raised based on *USH2A* transcript analysis from native nasal epithelium tissue of Usher syndrome patients with the common c.2299delG mutation showing that a fraction of mutant transcripts lack exon 13 or both exons 12 and 13 (Lenassi et al. 2014).

AON technology can be alternatively exploited for the blockage of alternate splice donor and acceptor sites. Rescue or suppression of 'true' splice mutations which impair splicing through the formation of novel splice donor or splice acceptors sites has the potential to restore expression of the full-length protein. For retinal dystrophies this approach has been most exclusively used for deep-intronic mutations which results in the inclusion of a PE into the mRNA. Feasibility of this approach with profound rescue rates has been demonstrated for deep-intronic mutations in *CHM* (Garanto et al., 2018), *USH2A* (Slijkerman et al., 2016), several *ABCA4* mutations (Albert et al., 2018; Bauwens et al.,

2019; Sangermano et al., 2019) and the common c.2991+1655A>G mutation in *CEP290* (Gerard et al., 2012; Garanto et al., 2016). Whereas most of these rescue experiments were carried out *in vitro* using minigene constructs or on patient-derived fibroblasts, improved ciliogenesis and cilia length has been demonstrated upon AON-induced PE suppression of the *CEP290*:c.2991+1655A>G mutation in an iPSC-derived retinal organoid model (Dulla et al., 2018) and reduction of misspliced *CEP290* transcript has been reported *in vivo* in a humanized CEP290 mouse mutant (Garanto et al., 2016). Currently, the only example of an AON-approach targeting an exonic mutation has been performed for *USH1C* by means of *in vitro* and *in vivo* evaluation in a mouse model (Lentz et al., 2013). AON treatment of the mutation c.216G>A within exon 3, which generates a novel splice donor site, successfully ameliorated the hearing and vestibular dysfunction upon systemic delivery in neonatal mutant mice (Lentz et al., 2013). Unfortunately, *USH1C* expression in the retina was not explored in this study. We have recently applied AON-based splice switching oligonucleotides to rescue a deep-intronic mutation in *OPA1* in dermal fibroblasts of patients with dominant optic atrophy and could demonstrate dose-dependent restoration of properly spliced transcripts from the mutant allele and a concomitant increase in OPA1 protein. Actually, this represents the first example of AON-based treatment for a mitochondrial disease and up to now, the only example for AON-based treatment of a dominant allele in ocular disease (Bonifert et al., 2016).

Delivery of AONs to the retina needs to take into account the blood-retinal barrier which strongly limits the access following systemic application. Therefore, AONs need to be delivered directly to the eye, for instance via injection into the vitreous humor. Once present in the vitreous, AONs have proven to be able to freely diffuse into all retinal cells to interfere with splicing (Gerard et al., 2015).

Clinically most promising advances in AON-based therapy have been made for the treatment of the *CEP290* mutation c.2991+1655A>G in patients with Leber congenital amaurosis. The first data on eight patients from a phase I/II study with 3-6 month follow-up upon intravitreal injection of the AON (QR-110, Sepofarsen) have been published and showed good tolerability and some improvement in visual function on average. Particularly, one of the patients responded very strongly and quickly with an improvement of visual acuity from light perception at baseline to 20/400 over a period of 6 months (Cideciyan et al., 2019). A fast-track phase 2/3 clinical trial is now underway.

A drawback and potential advantage at the same time is the limited half-life of AON-based drugs. Although backbone modifications have led to an increase in stability and pharmacodynamics of AONs, the decay over time necessitates regular re-administrations (e.g. every 3-6 month). However, this obstacle can be overcome by adapting dosing over time and allows tapering of the drug in case of adverse effects.

5.4 Genome Editing

Recent developments in the application of gene editing technologies for the treatment of ocular diseases in general have been discussed in other reviews in this journal (Lee et al., 2019; Burnight et al., 2018). As an alternative to splice-switching oligonucleotides, genome editing strategies are currently being explored for stable correction of splicing mutations at the DNA level in somatic cells. Whereas homology directed repair-based genome editing is still very inefficient in particular in non-dividing cells, several groups propose to harness the non-homologous end joining-based repair pathway to eliminate mutations in non-coding sequences such as deep intronic mutations where imprecise repair is tolerable.

Recent research in the field of retinal dystrophies has pioneered this approach. Targeting the common c.2991+1655A>G mutation in *CEP290*, Maeder and colleagues designed two guide RNAs to direct Cas9 induced double-strand breaks up- and down-stream of the pseudoexon (Maeder et al., 2019). Delivering the *S.aureus* Cas9 and the two guide RNA sequences through packaging into a single AAV vector, the authors could demonstrate successful genome editing and a subsequent increase in *CEP290* expression in human cells, human explant cultures and a humanized mouse model. Moreover, the feasibility of the approach was demonstrated *in vivo* in a non-human primate with a rate of up to

28% of productive editing (Maeder et al., 2019). Pioneering the application of CRISPR/Cas9 technology to correct a somatic gene defect in human patients, this approach is now entering clinical translation. The FDA recently approved an Investigational New Drug application to start a first clinical trial in human LCA patients using subretinal AAV-based delivery of the editing components (NCT03872479). However, in the pre-clinical study productive editing rate (i.e. inversion or deletion of the PE expected to rescue normal splicing) was only about 40% of all observed editing events in this approach which may not suffice for other genetic defects.

With the pace of improvements and novel tools that become available in the field of genome editing, it can be expected that there is still room for greater improvements. Notwithstanding, issues such as efficient delivery and off-target sites need to be thoroughly monitored, for example by whole genome sequencing.

5.5 *Trans*-Splicing

Finally, a therapeutic approach which does not specifically target splicing mutations, but makes use of the splicing machinery shall be mentioned. *Trans*-splicing between independent precursor mRNA molecules is a naturally occurring process in many lower eukaryotes and a few examples have also been reported in mammals (Lei et al., 2016). *Trans*-splicing in mammalian cells can also be induced through the introduction of base-pairing between the 5' and 3' intronic portions of the two precursor transcripts and has been exploited to rescue gene mutations in a technology called spliceosome-mediated RNA *trans*-splicing (SMaRT) in *in vitro* and *in vivo* models of various disease such as cystic fibrosis, hemophilia A, SMA, and others (Liu et al., 2002; Chao et al., 2003; Berger et al., 2016). Efficacies range from as low as 5.4% to up to 40% (Berger et al., 2016). Berger and colleagues investigated *trans*-splicing for human *RHO* transcripts employing a 3' pre-mRNA *trans*-splicing module (PTM) that induces *trans*-splicing in intron 1. *Trans*-splicing efficacy of up to 40% was obtained in HEK293 cells and enabled rescue of mutations in stable cell lines. *Trans*-splicing was also observed *in vivo* upon subretinal AAV-delivery with *trans*-splicing of up to 20% in the transduced area but no therapeutic effect on retinal degeneration in a $Rho^{+/-}RHO^{P347S/-}$ mouse model (Berger et al., 2015). Most recently, Dooley and co-workers demonstrated SMaRT-induced *trans*-splicing in intron 26 of *CEP290* with a 5' PTM in HEK293T cells as well as in a transgenic *CEP290* minigene mouse model upon AAV7m8-based subretinal delivery albeit with rather modest efficacy (Dooley et al., 2018).

6. Future directions

Although pre-mRNA splicing has been studied intensively over more than four decades, our understanding of the splicing *language* and the underlying *grammar* rules is still incomplete. Especially intricate are *cis*-acting sequence elements outside the canonical sequence which modulate or critically impair exon recognition. Current bioinformatic tools are still insufficient in terms of sensitivity and specificity to call splicing mutations reliably. Moreover, prediction algorithms do not necessarily provide a conclusive support for the actual consequence of a splicing mutation on transcript processing. It is likely that many exonic variants which are currently classified as alterations in the coding sequence in fact cause splicing defects.

Progress in the development and implementation of deep learning algorithms and artificial intelligence may further improve sensitivity and specificity to call splicing mutations. Such approaches have lately being applied to score variants from entire variant databases (Xiong et al., 2015) or to generate lookup-tables of all potential nucleotide changes in all human genes (Jaganathan et al., 2019). According to our own experience, these deep-learning algorithms have higher sensitivity and specificity compared with individual splice prediction tools but are still not applicable for diagnostic applications without experimental validation, in particular for deep-intronic mutations. However, the combination of established bioinformatics tools with deep learning algorithms is expected to further enhancing the predictive power of *in silico* analysis.

Minigene assays have proved a valid method to assay potential splicing variants to complement *in silico* predictions. However, construct cloning is usually performed for each variant individually and thus is a laborious and time-consuming procedure. Recently, highly parallelized functional assay formats have been developed which use barcoding and next generation sequencing to deconvolute the data sets into individual bins. Read counts are the primary outcome, a variable cognate to the splicing perturbation elicited by each of the thousands of intronic and exonic variants introduced in the system by synthetic mutagenesis (Rosenberg et al., 2015; Soemedi et al., 2017; Adamson et al., 2018). Such high throughput assays and their outputs can also serve as valuable sources for the development of bioinformatic tools and deep-learning algorithms with improved sensitivity and specificity (see paragraph above). The need for improved variant interpretation including potential splice variants has led to the formation of the CAGI (Critical Assessment of Genome Interpretation) research community (https://genomeinterpretation.org/). CAGI promotes public competitions to develop and benchmark computational methods to predict the molecular impact of genomic variation. In a recent call CAGI5 solicited tools to identify the impact of genetic variants on exon splicing using the Vex-seq dataset from 2059 human genetic variants spanning 110 alternative exons (Adamson et al., 2018). Yet, such parallelized assays still present shortcomings in terms of the short length of the assayed exons, the synthetic nature of constructs (e.g. hybrid up- and down-stream introns), and the use of cell culture models. Some of these limitations of parallel minigene reporters have been obliterated in our recent approach to study the impact of *OPN1LW/MW* exon 3 haplotypes on splicing where near-native constructs with full-length genuine upstream and downstream intronic sequences were used (Buena-Atienza et al., manuscript in preparation; Fig. 8C). The implementation of long-read sequencing in this context enables large minigenes to be assayed in parallel, for example minigene constructs accommodating genomic sequence segments of several kb in length including multiple exons and their flanking introns. Such midigene constructs closer reflect the native genomic context and integrate interdependency of splice site and exon recognition as well as long-distance interactions of *cis*-acting sequences (Sangermano et al., 2018).

Direct transcript analysis aiming at obtaining the continuous full-length sequence of transcripts is amenable by long-read sequencing technologies which provide exon connectivity and will allow detection of multiple aberrant splicing events or so far undetected misspliced products in transcripts of the length of *ABCA4*. Moreover, the readout could provide insight into the quantification of transcripts (e.g. relative expression of isoforms) carrying heterozygous variants and the phasing of those variants. The outcome of alternative splicing events as well as aberrant splicing caused by splicing defects is also influenced by cell and tissue specific factors or splicing programs and thus, can vary qualitatively as well as quantitatively. One recommendation to circumvent the lack of native retinal context is that future minigene assays and in particular parallelized assays testing a large variety of putative splicing mutations take advantage of testing systems approximating the *in vivo* context; for instance, iPSC-derived photoreceptor precursors, RPE cells or organoids, human donor retinal explants or *in vivo* expression in suitable mammalian systems. An intriguing example are the findings of Becirovic and colleagues showing that strong cell type-specific differences in the splicing pattern of *PRPH2* and mutations in this gene as assayed with an AAV delivered minigene vector (Becirovic et al., 2016).

High penetrant splicing mutations causing inherited (retinal) diseases are probably only the 'tip of the iceberg'. Genetically determined variations in the level of gene expression is a phenomenon that has recently attracted attention in particular in the context of common diseases and the unsolved problem of variability of disease expression and reduced penetrance in mendelian disorders (GTEx consortium, 2017). A decent fraction of *cis*-acting expression quantitative trait loci (eQTLs) are due to variants affecting the pattern and quantity of alternative splicing of a given gene (Zhang et al., 2015; Park et al., 2018). Notably, a large number of such splicing QTLs (sQTLs) co-localize with loci for complex disease mapped by genome wide association studies (Park et al., 2018; Raj et al., 2018). Currently, data on eQTLs are very limited for the human retina (Llavona et al., 2017; Ratnapriya et al., 2019), and no single study currently exists which specifically addresses sQTLs in the retinal transcriptome. Given the

large fraction of alternatively spliced transcript in the retina and retina specific transcript isoforms, there is quite some potential for such studies. Future research may also address potential perturbations in the overall pattern of AS in disease and during retinal degeneration splicing using the many IRD animal models which are now available (Slijkerman et al., 2015).

There is also a need to further explore the full breadth of the retinal transcriptome and to characterize changes during development, maturation and ageing which should make use of long read sequencing for unambiguous determination of transcript isoforms. This will also enable to investigate the propensity of the retina for AS of microexons as seen in the brain (Irmia et al., 2014). While prior transcriptomic studies mostly dealt with the retina in its entirety, technological advances in single cell and single nuclei analytics now enters a stage which allows to address and to resolve the transcriptome and potentially also the pattern of AS at the level of individual cells and retinal cell populations (Shekhar et al., 2016; Rheaume et al., 2018; Liang et al., 2019).

More than 40 years after the discovery of split genes, we only lately begin to appreciate the full complexity of splicing and its importance in cellular differentiation, the maintenance of cellular homeostasis, and to drive physiological changes in response to external stimuli or stress. The ingenuity of this system in the first place goes at the expense of a decent vulnerability for perturbations such as splicing mutations which are responsible for a considerable fraction of cases suffering from IRD.

Tables

Table 1. Types of splicing mutations in 20 selected retinal disease genes. (Source: HGMD®)

Gene	All splicing variants	Variants of the canonical donor site (+1, +2)	Variants of the canonical acceptor site (-1, -2)	Exonic variants acting on splicing	Non-canonical splice site variants[a,b]	Deep intronic splice site variants[b]
ABCA4	190	67	46	10	49	18
USH2A	120	44	37	9	24	6
EYS	28	14	3	3	8	0
CRB1	18	6	6	3	3	0
CHM	49	16	23	1	7	2
RS1	23	12	5	1	5	0
BEST1	13	7	2	2	2	0
GUCY2D	21	9	8	1	3	0
RPGR	47	19	11	2	15	0
RHO	6	1	3	1	1	0
RPE65	27	7	9	2	9	0
CEP290	44	16	16	2	9	1
RP1	1	1	0	0	0	0
PRPF31	27	12	9	0	4	2
CACNA1F	24	12	6	0	6	0
PRPH2	5	2	1	0	2	0
RPGRIP1	23	9	4	2	5	3
PDE6B	18	7	7	0	4	0
CNGA3	4	2	1	0	1	0
CNGB3	22	10	8	0	4	0
Sum	710	273	205	39	161	32

[a]positions -20 to +20; [b]Pathogenicity not experimentally validated for all variants.

Table 2. Prediction scores for canonical splice sites of *CNGB3* (NM_019098.4)

	HSF Matrices [0 - 100]		MaxEntScan [0-16]		NNSplice [0 - 1]	
Exon	**donor**	**acceptor**	**donor**	**acceptor**	**donor**	**acceptor**
1[a]	91.2	-	9.8	-	1.0	-
2	91.69	87.13	6.8	7.96	0.92	0.95
3	96.67	85.21	9.6	7.37	0.99	0.98
4	93.08	81.44	8.69	8.8	1.0	0.83
5	76.54	79.33	7.26	7.22	0.6	0.41
6	83.72	90.85	8.65	9.53	0.97	0.99
7	91.66	87.59	10.28	7.75	0.98	0.28
8	96.71	89.86	10.57	9.05	1.0	0.45
9	85.9	82.86	9.65	7.6	0.92	0.94
10	84.57	87.79	10.15	7.66	0.99	0.86
11	97.83	92.54	9.88	10.19	0.99	0.99
12	95.84	79.64	8.89	7.38	1.0	0.81
13	72.21	78.42	3.24	4.9	-	-
14	97.89	85.46	10.47	9.78	1.0	0.99
15	79.18	79.28	7.33	6.07	0.68	0.14
16	76.7	90.14	7.09	11.54	0.95	0.99
17	85.35	90.37	8.55	12.09	0.94	0.87
18[b]	-	91.08	-	6.36	-	0.92

Default thresholds were used for all tools. Score ranges for each tool are given in brackets. [a]Being the first exon, exon 1 lacks the splice acceptor site. [b]Being the last exon, exon 18 lacks the splice donor site.

Figures

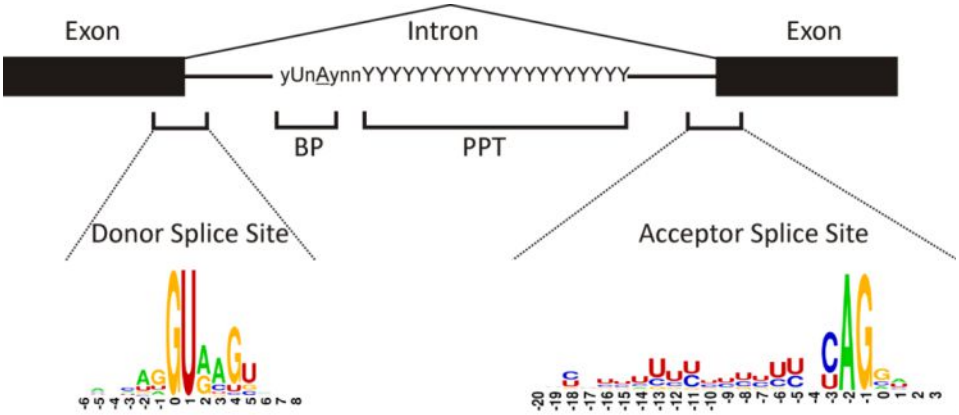


Figure 1. Prototypic consensus splice sites of human genes. Most introns start with GT nucleotides, and end with AG nucleotides. Consensus sequences were created using Weblogo (logo@compbio.berkeley.edu). The height of each letter is proportional to the frequency of the corresponding nucleotide at the given position. BP, branch point; PPT, polypyrimidine tract.

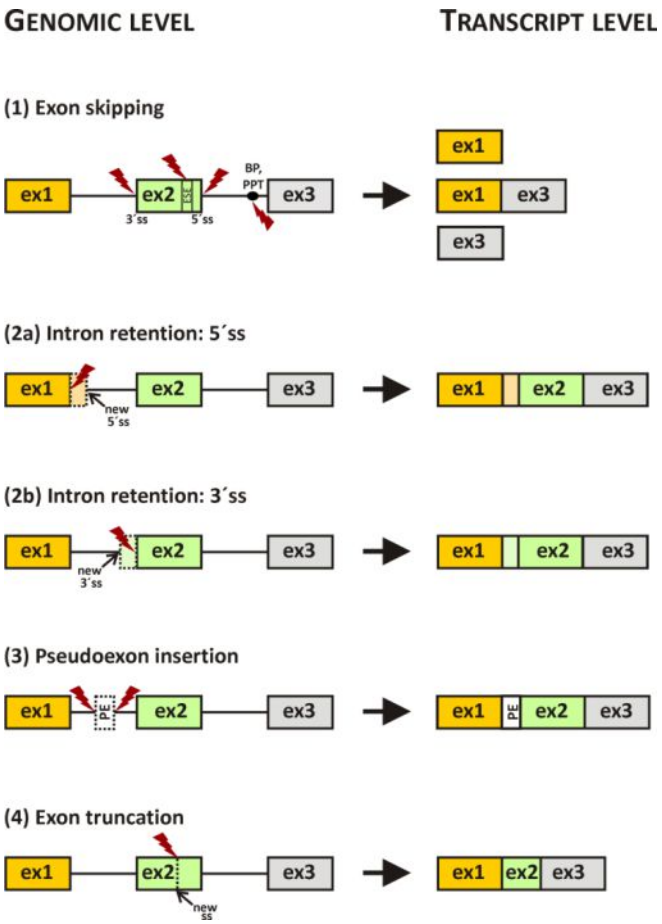


Figure 2. Overview of different splice type mutations. (1) Exon skipping is the most frequent consequence of splice site mutations. Both mutations of the splice acceptor (3´ss) and the splice donor (5´ss) can cause exon skipping. Mutations of branch point (BP) and polypyrimidine tract (PPT) are comparatively rare. Exonic mutations disrupting an exonic splicing enhancer (ESE) motif (as depicted) or creating an exonic splicing silencer (ESS) motif can lead to exon skipping. The same mutation can result in different transcripts lacking one or more exons. (2) Activation of an intronic cryptic splice site in the proximity of a constitutive exon can lead to the inclusion of intronic sequences into the mRNA downstream of the canonical 5'ss donor site (2a) or upstream of the 3'ss acceptor site (2b). (3) Creation or strengthening of a cryptic donor or acceptor site can activate a 'decoy' complementary splice site in the proximity, leading to pseudoexon (PE) insertion into the transcript. (4) Exonic mutations can create a new splice site leading to exon truncation.

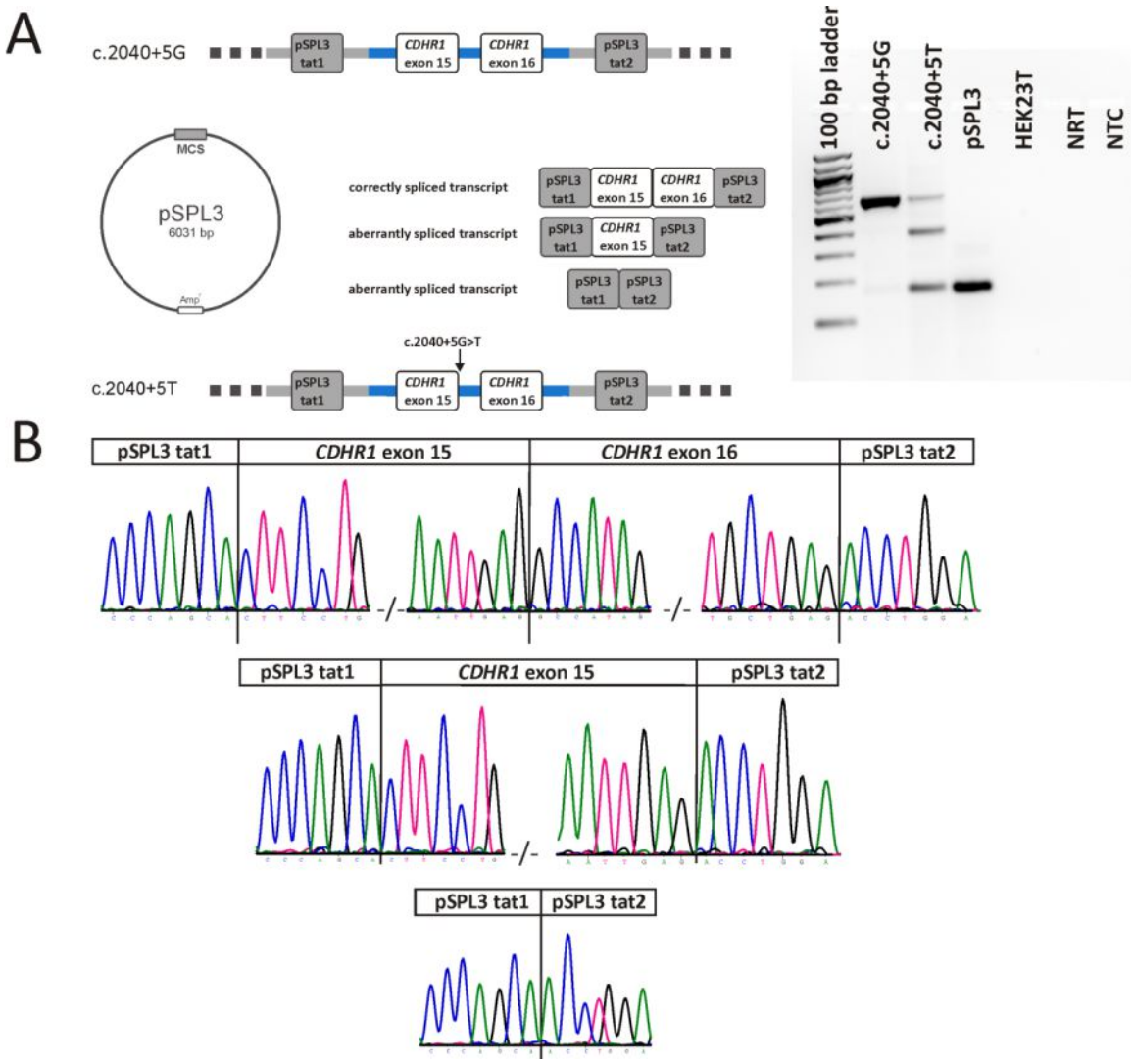


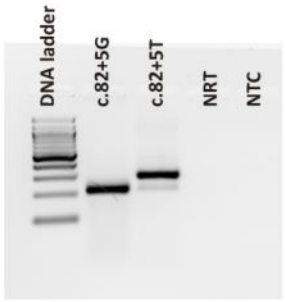
Figure 3. Analysis of the c.2040+5G>T variant in CDHR1. (A) Constructs comprising a fragment from intron 14 to intron 16 of *CDHR1* were generated to contain either the c.2040+5T mutant or the c.2040+5G control, respectively, and cloned into the exon trapping vector pSPL3. RT-PCR revealed one product in HEK293T cells transfected with the wildtype construct (lane 2) and three products for the mutant construct (lane 3). Transfection with empty pSPL3 vector (lane 4) and non-transfected cells (lane 5) served as controls. NRT, no reverse transcriptase control (lane 6), NTC, no template control (lane 7). A 100 bp ladder size standard was loaded in the leftmost lane. Schemes of the amplified products are presented on the left of the agarose gel. Grey boxes represent pSPL3 exons and white boxes *CDHR1* exons. (B) Sequence analysis showing that one aberrant RT-PCR product from the mutant minigene construct results from skipping of exon 16 and the other aberrant RT-PCR product results from skipping of both exon 15 and 16.

A

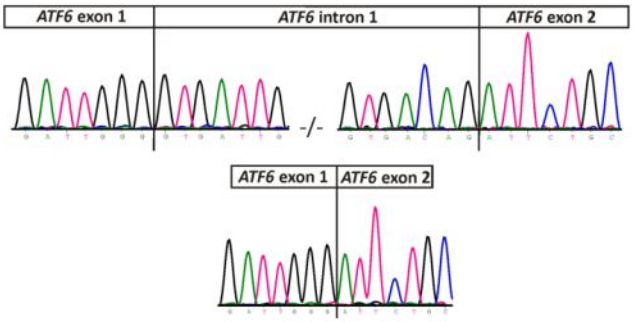
AGATATTAATCACGGAGTTCCAGGGAG
AAGGAACTTGTGAAATGGGGGAGCCGG
CTGGGGTTGCCGGCACCATGGAGTCAC
CTTTTAGCCCGGGACTCTTTCACAGGC
TGGATGAAGATTGGGgtgaktgggatc
tgagaatgtaccagggtggctcgggtt
cgcgtctaaaggtttcttcctccctac
ttccgcccactcgtggtgacaggtgtg
gaccaaccctgcct

Splice site	HSF matrices	MaxEnt	NNSplice
canonical donor +5G	92.22	7.93	0.97
canonical donor +5T	79.91	-	-
cryptic donor	88.1	6.34	0.8

B



C



D

```
CACAGGCTGGATGAAGATTGGGgtgattgggatctgagaatgtaccagggtggctcgggttcgcgtctaaaggtttcttcc
 H  R  L  D  E  D  W  G  D  W  D  L  R  M  Y  Q  G  G  S  G  S  R  L  K  V  S  S

tccctacttccgcccactcgtggtgacaggtgtggaccaaccctgcctggggagtgaggcccctcgtc.............
 S  L  L  P  P  T  R  G  D  R  C  G  P  T  L  P  G  E  *
```

Figure 4. *In silico* and direct transcript analysis of the c.82+5G>T variant in ATF6. (A) The sequence shows *ATF6* exon 1 in capital letters followed by intronic sequence. The invariable GT dinucleotides of the canonical and the cryptic donor site are shown in blue. The c.82+5 position (K in IUPAC code) is shown in red. The 88 bp that are predicted to be included in the aberrant transcript are underlined. The table provides the scores from three different *in silico* splice prediction tools. (B) RNA was extracted from PaxGene-isolated blood samples from a patient being homozygous for the c.82+5G>T variant as well as from a healthy control sample, reverse transcribed and PCR amplified. The healthy control sample (lane 2) shows a single transcript while the patient shows two transcripts (lane 3), the major one being clearly larger than the correct transcript. A 100 bp ladder size standard was loaded in the leftmost lane. NRT (lane 4), no reverse transcriptase control; NTC (lane 5), no template control. (C) Sequencing of subcloned RT-PCR products demonstrates that the larger transcript shows exonification of the first 88 bp of intron 1 while the smaller transcript shows correct splicing of exon 1 and exon 2. (D) Consequence of the c.82+5G>T variant at the amino acid level. The first 88 bp of intron 1 are translated into 38 novel amino acids followed by a PTC. Protein translation is shown below the nucleotide sequence in one letter amino acid code.

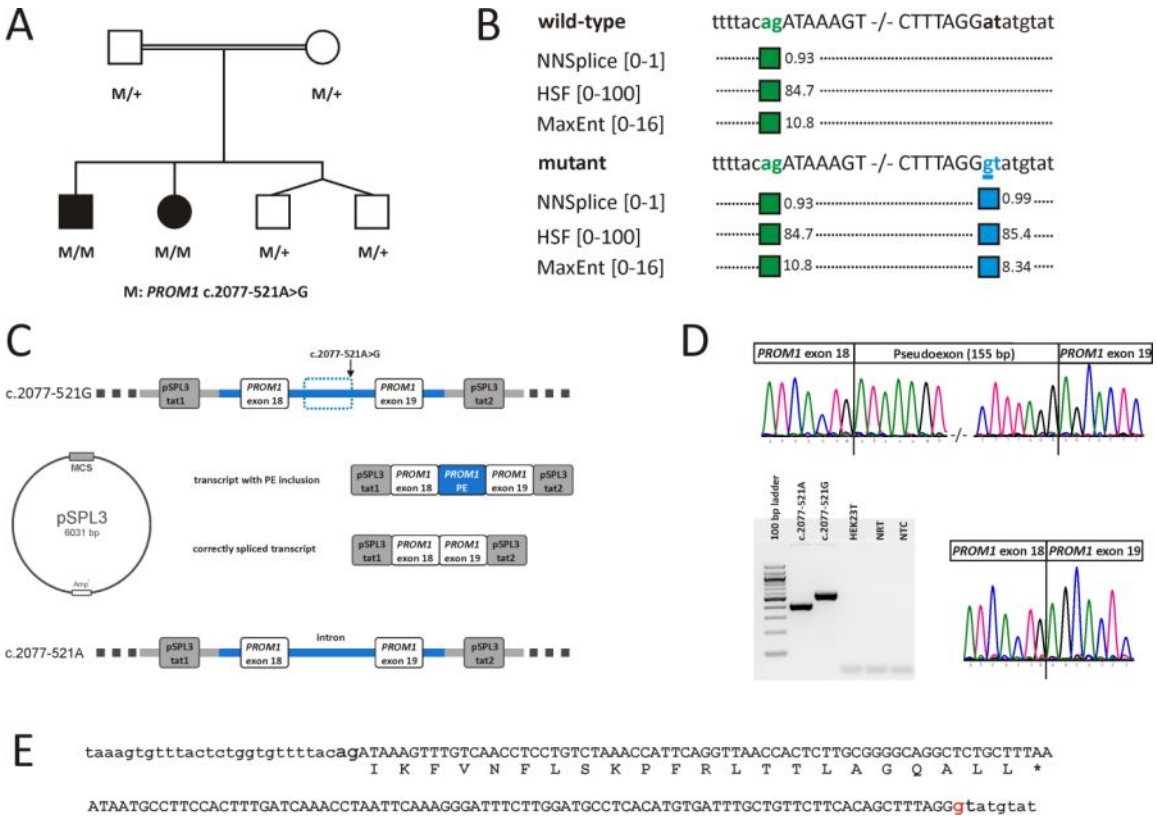


Figure 5. Analysis of the deep intronic variant c.2077-521A>G in PROM1. (A) Pedigree of the family segregating the deep intronic variant c.2077-521A>G in *PROM1*. Both affected siblings are homozygous for the variant (M/M), while the parents and the unaffected twin siblings are heterozygous (M/+). First-cousin marriage is indicated by double horizontal lines. (B) The sequence electropherograms show the cryptic donor site (blue) that is created by the c.2077-521A>G variant (underlined). Scores from three algorithms are given next to the blue box (score ranges are given in brackets). A complementary acceptor site (green) 155 nucleotides upstream is predicted by all three algorithms (scores next to the green boxes). (C) Wildtype and mutant minigene constructs were cloned into the exon-trapping vector pSPL3 which contains two exons tat1 and tat2 (grey). The cloned fragments comprised *PROM1* exon 18 with 239 bp of upstream sequence, intron 18 and exon 19 with 367 bp of downstream sequence. (D) RT-PCR and subsequent sequence analysis showed that the aberrant larger transcript is due to the insertion of a PE of 155 nucleotides between exons 18 and 19. (E) Consequence of the c.2077-521A>G variant (underlined). The PE (in capital letters) translates into 22 novel amino acids followed by a PTC (protein translation is shown below DNA sequence in one letter amino-acid code).

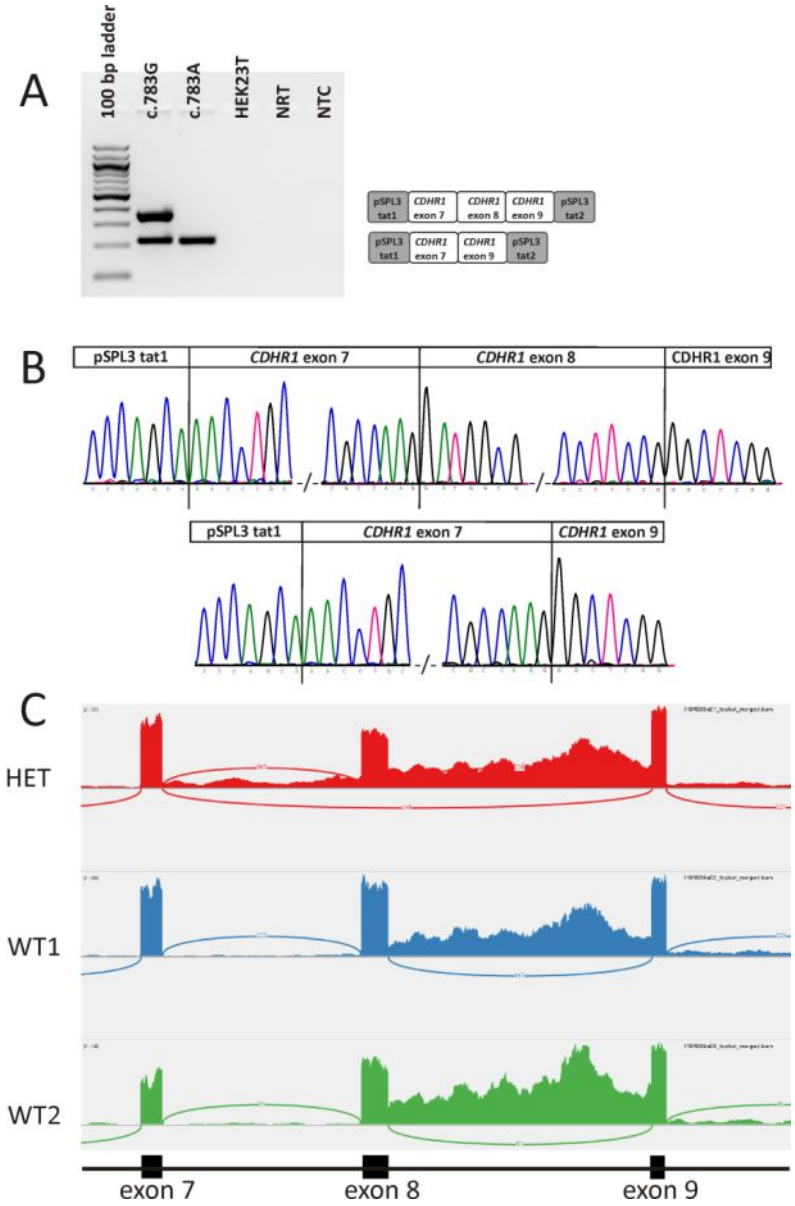


Figure 6. Analysis of the c.783G>A variant in CDHR1. (A) RT-PCR revealed two products in HEK293T cells transfected with the wildtype minigene (lane 2) and a single RT-PCR product for the mutant minigene (lane 3). Non-transfected cells served as control (lane 4). NRT, no reverse transcriptase control (lane 5), NTC, no template control (lane 6). A 100 bp ladder size standard was loaded in the leftmost lane. Schemes of the amplified products are presented on the right of the agarose gel. Grey boxes represent pSPL3 exons and white boxes *CDHR1* exons. (B) Sequence analysis showing that the aberrant product results from skipping of exon 8. The junction of *CDHR1* exon 9 and pSPL3 tat2 could not be shown since a reverse PCR primer was used that binds within exon 9. (C) Sashimi plots were obtained from RNASeq data of one retina sample being heterozygous for the c.783G>A variant in *CDHR1* (HET) and from two samples being homozygous for the wildtype G allele at position c.783 (WT1 and WT2). The arcs indicate splice junction reads, with the thickness of the arc correlating with the number of junction reads spanning the two exons being connected by the arc. A significant number of reads (n = 149) between exon 7 and 9, indicating skipping of exon 8, are only seen in sample HET.

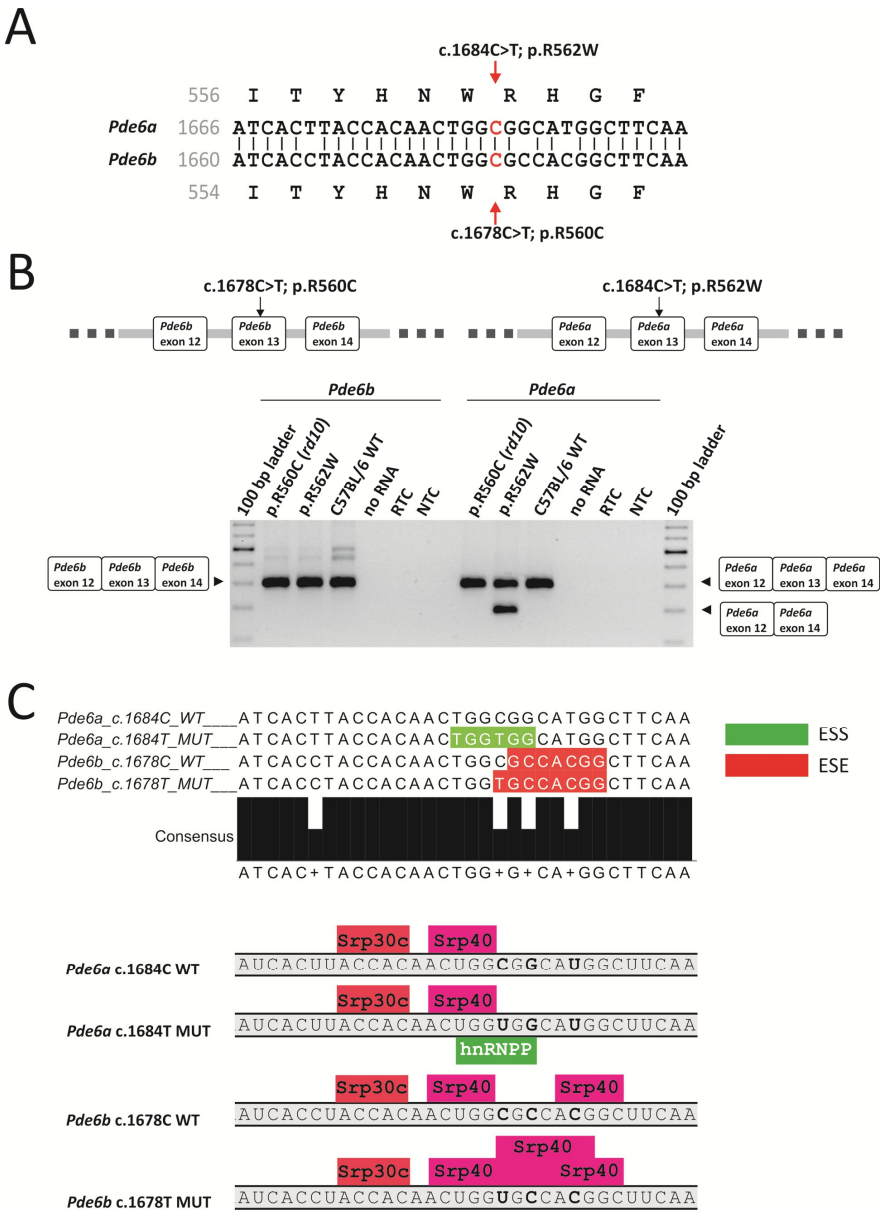


Figure 7. The impact of differential sequence context on splicing regulation. (A) Sequence at the vicinity of the homologous mutations c.1684C>T/p.R562W and c.1678C>T/p.R560C (*rd10*) of the murine *Pde6a* and *Pde6b* paralogous genes, respectively. Full conservation is present at the amino acid sequence and high conservation at the nucleotide level; 32 bp shown. (B) RT-PCR of *Pde6a* or *Pde6b* transcripts from murine retinae. Exon 13 skipping was detected only in the c.1684C>T/p.R562W knockin mouse mutant. No RNA, negative control without RNA template; RTC, negative control without reverse transcriptase; NTC, no template control in the PCR reaction (C) Caption of the multiple sequence alignment of exon 13 carrying either the wildtype (C-allele) or the mutant (T-allele) for each of the *Pde6a* and *Pde6b* genes. The predicted exonic splicing silencer (ESS) and exonic splicing enhancer (ESE) motifs are depicted by green and red filled boxes, respectively. Selected RNA-binding proteins or splicing factors predicted to be recruited by the *cis*-regulatory elements of splicing differentially present in the pre-mRNA are shown (SpliceAid2). The mutations and synonymous variants are depicted in bold. Serine-arginine rich (SR) proteins and a member of the heterogeneous nuclear ribonucleoproteins (hnRNP) family are shown by green and red filled boxes below or above the predicted ESS or ESE motif they putatively recognize.

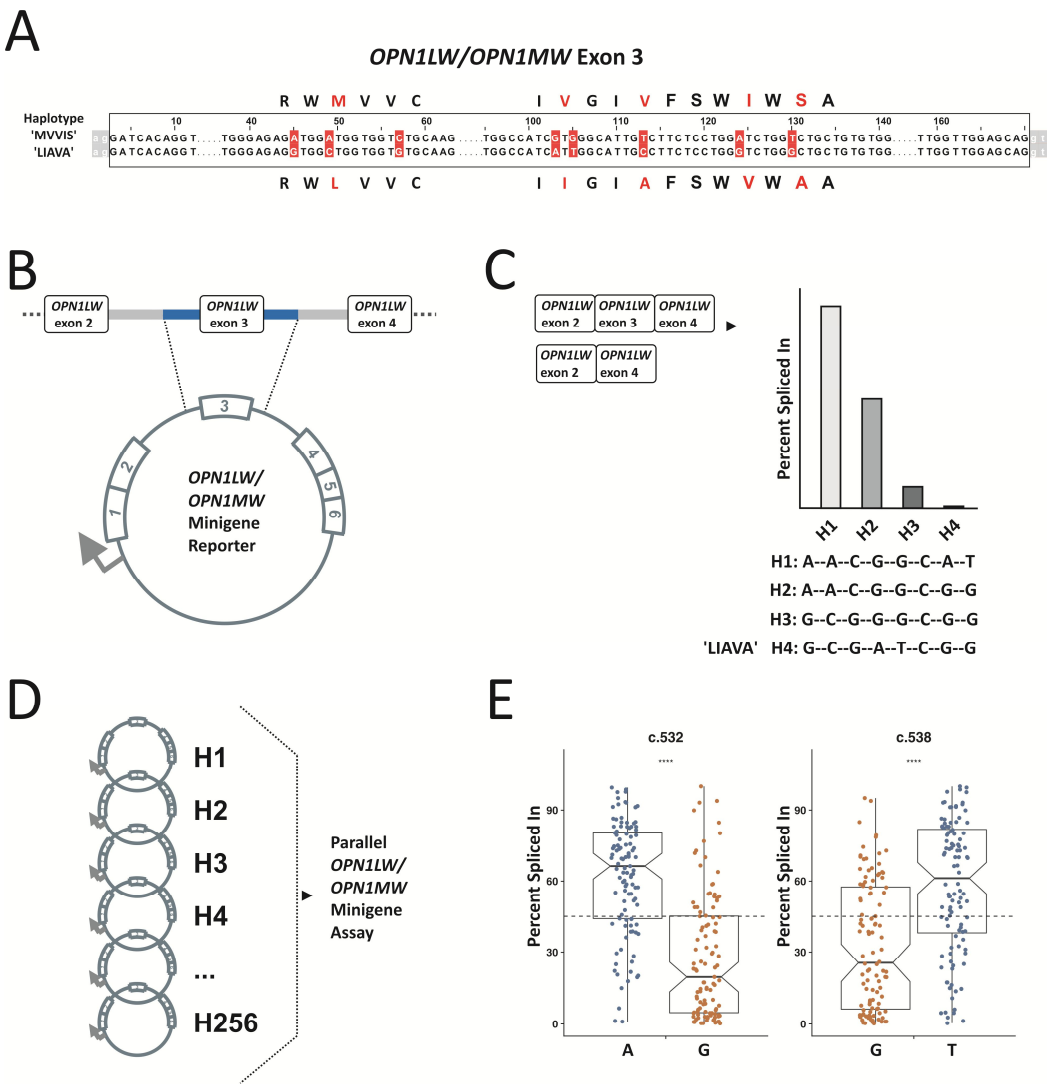


Figure 8. Analysis of exon 3 haplotypes in *OPN1LW/OPN1MW*. (A) Exon 3 of *OPN1LW* and *OPN1MW* is depicted by a white box with highlighted polymorphic single nucleotide sites and the amino acid changes encoded in red. Two aligned sequences harbouring distinct haplotypes with deviation for each of the eight possible polymorphic sites in exon 3 are shown. Note the position of the exonic variants within the internal part of exon 3 with respect to donor and acceptor splice sites (in grey boxes). Numbers represent coordinates of the alignment ("1" being the last but one nucleotide at the end of intron 2). (B) The minigene splicing reporter which has been used to assess *OPN1LW* and *OPN1MW* exon 3 haplotypes is a plasmid construct that harbours a strong promoter, all six endogenous exons and the full-length intronic sequence flanking the 169 bp long exon 3 with genuine splice sites and a polyA site (Ueyama et al., 2012). (C) Relative quantification of correctly and aberrantly spliced transcripts can be performed individually for each haplotype ("H") by fluorescence labelling of differentially sized RT-PCR products. (D) Parallel assessment of hundreds of targeted haplotypes ("H1-H256") in exon 3 of *OPN1LW* and *OPN1MW*. (E) Quantification of the proportion of correctly spliced transcripts for each assayed haplotype. The individual effect of c.532G and c.538G (the two most outermost variants on the haplotype closest to the donor splice site) on exon inclusion is shown. Haplotypes with the latter exonic variants (c.532G and c.538G) overall yield a significantly higher rate of exon skipping than those haplotypes with the corresponding alternate variants. Percent 'Spliced In' for each haplotype (dot) is shown in the graph as the mean % of exon 3 retention.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to thank the German Research Council (KFO 134, Wi1189/12-1), the Federal Ministry for Research and Education (01 GM 0850, 01GM1108A) the European Commission (ERA-NET 'ERMION' and 7th framework programme: 317472), the Tistou und Charlotte Kerstan - Stiftung, and the Blue Cone Monochromacy Family Foundation for support of our research of splicing defects in retinal disease genes.

References

Abramowicz, A., Gos, M., 2018. Splicing mutations in human genetic disorders: examples, detection, and confirmation. J Appl Genet 59, 253-268.

Adamson, S.I., Zhan, L., Graveley, B.R., 2018. Vex-seq: high-throughput identification of the impact of genetic variation on pre-mRNA splicing efficiency. Genome Biol 19, 71.

Albert, S., Garanto, A., Sangermano, R., Khan, M., Bax, N.M., Hoyng, C.B., et al., 2018. Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease. Am J Hum Genet 102, 517-527.

Alter, J., Lou, F., Rabinowitz, A., Yin, H., Rosenfeld, J., Wilton, S.D., Partridge, T.A., Lu, Q.L., 2006. Systemic delivery of morpholino oligonucleotide restores dystrophin expression bodywide and improves dystrophic pathology. Nat Med.12, 175-177.

Aparisi, M.J., García-García, G., Aller, E., Sequedo, M.D., Martínez-Fernández de la Cámara, C., Rodrigo, R., et al., 2013. Study of USH1 splicing variants through minigenes and transcript analysis from nasal epithelial cells. PLoS One 8, e57506.

Aukrust, I., Jansson, R.W., Bredrup, C., Rusaas, H.E., Berland, S., Jørgensen, A., et al., 2017. The intronic ABCA4 c.5461-10T>C variant, frequently seen in patients with Stargardt disease, causes splice defects and reduced ABCA4 protein level. Acta Ophthalmol 95, 240-246.

Bacchi, N., Casarosa, S., Denti, M.A., 2014. Splicing-correcting therapeutic approaches for retinal dystrophies: where endogenous gene regulation and specificity matter. Invest Ophthalmol Vis Sci 55, 3285-3294.

Baden, T., Berens, P., Franke, K., Román Rosón, M., Bethge, M., Euler, T., 2016. The functional diversity of retinal ganglion cells in the mouse. Nature 529, 345-350.

Balestra, D., Faella, A., Margaritis, P., Cavallari, N., Pagani, F., Bernardi, F., et al., 2014. An engineered U1 small nuclear RNA rescues splicing defective coagulation F7 gene expression in mice. J Thromb Haemost 12, 177-185.

Baralle, F.E., & Giudice, J., 2017. Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol 18, 437–451.

Baranello, G., Servais, L., Day, J., Deconinck, N., Mercuri, E., Klein, A., et al., 2019 FIREFISH Part 1: 1-Year Results on Motor Function in Babies with Type 1 SMA. Annual Meeting of the American Academy of Neurology, May 04-10, 2019, Philadelphia

Barny, I., Perrault, I., Michel, C., Goudin, N., Defoort-Dhellemmes, S., Ghazi, I., et al., 2019. AON-Mediated Exon Skipping to Bypass Protein Truncation in Retinal Dystrophies Due to the Recurrent CEP290 c.4723A > T Mutation. Fact or Fiction?. Genes (Basel) 10(5), pii: E368.

Bauwens, M., Garanto, A., Sangermano, R., Naessens, S., Weisschuh, N., De Zaeytijd, J., et al., 2019. 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 2019 Jan 23

Bax, N.M., Sangermano, R., Roosing, S., Thiadens, A.A., Hoefsloot, L.H., van den Born, L.I., et al., 2015. Heterozygous deep-intronic variants and deletions in ABCA4 in persons with retinal dystrophies and one exonic ABCA4 variant. Hum Mutat 36, 43-47.

Becirovic, E., Ebermann, I., Nagy, D., Zrenner, E., Seeliger, M.W., Bolz, H.J., 2008. Usher syndrome type 1 due to missense mutations on both CDH23 alleles: investigation of mRNA splicing. Hum Mutat 29, 452.

Becirovic, E., Böhm, S., Nguyen, O.N., Riedmayr, L.M., Koch, M.A., Schulze, E., et al., 2016. In Vivo Analysis of Disease-Associated Point Mutations Unveils Profound Differences in mRNA Splicing of Peripherin-2 in Rod and Cone Photoreceptors. PLoS Genet 12, e1005811.

Berger, A., Lorain, S., Joséphine, C., Desrosiers, M., Peccate, C., Voit, T., et al., 2015. Repair of rhodopsin mRNA by spliceosome-mediated RNA trans-splicing: a new approach for autosomal dominant retinitis pigmentosa. Mol Ther 23, 918-930.

Berger, A., Maire, S., Gaillard, M.C., Sahel, J.A., Hantraye, P., Bemelmans, A.P., 2016. mRNA trans-splicing in gene therapy for genetic diseases. Wiley Interdiscip Rev RNA 7, 487-498.

Berger, W., Kloeckener-Gruissem, B., Neidhardt, J., 2010. The molecular basis of human retinal and vitreoretinal diseases. Prog Retin Eye Res 29, 335-375.

Black, D.L., 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72, 291-336.

Blencowe, B.J., 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci 25, 106-110.

Bok, D., Young, R.W., 1972. The renewal of diffusely distributed protein in the outer segments of rods and cones. Vision Res 12, 161-168.

Bonifert, T., Karle, K.N., Tonagel, F., Batra, M., Wilhelm, C., Theurer, Y., et al., 2014. Pure and syndromic optic atrophy explained by deep intronic OPA1 mutations and an intralocus modifier.

Brain 137, 2164-2177.

Bonifert, T., Gonzalez Menendez, I., Battke, F., Theurer, Y., Synofzik, M., Schöls, L., Wissinger, B., 2016. Antisense Oligonucleotide Mediated Splice Correction of a Deep Intronic Mutation in OPA1. Mol Ther Nucleic Acids 5, e390.

Braun, T.A., Mullins, R.F., Wagner, A.H., Andorf, J.L., Johnston, R.M., Bakall, B.B., et al., 2013. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. Hum Mol Genet 22, 5136-5145.

Broadgate, S., Yu, J., Downes, S.M., Halford, S., 2017. Unravelling the genetics of inherited retinal dystrophies: Past, present and future. Prog Retin Eye Res 59, 53-96.

Bronstein, R., Capowski, E.E., Mehrotra, S., Jansen, A.D., Navarro-Gomez, D., Maher, M., et al., 2020. A combined RNA-seq and whole genome sequencing approach for identification of non-coding pathogenic variants in single families. Hum Mol Genet 29(6), 967-979.

Buckler, A.J., Chang, D.D., Craw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A., et al., 1991. Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc Natl Acad Sci USA 88, 4005-4009.
Buena-Atienza, E., Rüther, K., Baumann, B., Bergholz, R., Birch, D., De Baere, E., et al., 2016. De novo intrachromosomal gene conversion from OPN1MW to OPN1LW in the male germline results in Blue Cone Monochromacy. Sci Rep 6, 28253.
Burnight, E.R., Giacalone, J.C., Cooke, J.A., et al., 2018. CRISPR-Cas9 genome engineering: Treating inherited retinal degeneration. Prog Retin Eye Res 65,28-49.
Burset, M., Seledtsov, I.A., Solovyev, V.V., 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Res 28, 4364-4375.
Buskin, A., Zhu, L., Chichagova, V., Basu, B., Mozaffari-Jovin, S., Dolan, D., 2018. Disrupted alternative splicing for genes implicated in splicing and ciliogenesis causes PRPF31 retinitis pigmentosa. Nat Commun 9, 4234.
Busslinger, M., Moschonas, N., Flavell, R.A., 1981 Beta + thalassemia: aberrant splicing results from a single point mutation in an intron. Cell 27, 289-298.
Cáceres, J.F., Stamm, S., Helfman, D.M., Krainer, A.R., 1994. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science 265, 1706-1709.
Carss, K.J., Arno, G., Erwood, M., Stephens, J., Sanchis-Juan, A., Hull, S., et al., 2017. Comprehensive Rare Variant Analysis via Whole-Genome Sequencing to Determine the Molecular Pathology of Inherited Retinal Disease. Am J Hum Genet 100(1), 75-90.
Cartegni, L., Chew, S.L., Krainer, A.R., 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3, 285-298.
Chaitankar, V., Karakülah, G., Ratnapriya, R., Giuste, F. O., Brooks, M. J., Swaroop, A., 2016. Next generation sequencing technology and genomewide data analysis: Perspectives for retinal research. Prog Retin Eye Res 55:1-31.
Chang, B., Grau, T., Dangel, S., Hurd, R., Jurklies, B., Sener, E.C., et al., 2009. A homologous genetic basis of the murine cpfl1 mutant and human achromatopsia linked to mutations in the PDE6C gene. Proc Natl Acad Sci U S A 106, 19581–-19586.
Chao, H., Mansfield, S.G., Bartel, R.C., Hiriyanna, S., Mitchell, L.G., Garcia-Blanco, M.A., Walsh, C.E., 2003. Phenotype correction of hemophilia A mice by spliceosome-mediated RNA trans-splicing. Nat Med 9, 1015-1019.
Chen, L., Bush, S.J., Tovar-Corona, J.M., Castillo-Morales, A., Urrutia, A.O., 2014. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. Mol Biol Evol 31, 1402-1413.
Cideciyan, A.V., Jacobson, S.G., Drack, A.V., Ho, A.C., Charng, J., Garafalo, A.V., et al., 2019. Effect of an intravitreal antisense oligonucleotide on vision in Leber congenital amaurosis due to a photoreceptor cilium defect. Nat Med 25, 225-228.
Cirak, S., Arechavala-Gomeza, V., Guglieri, M., Feng, L., Torelli, S., Anthony, K., et al., 2011. Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study. Lancet 378, 595-605.
Cremers, F.P.M., Lee, W., Collin, R.W.J., Allikmets. R., 2020. Clinical spectrum, genetic complexity and therapeutic approaches for retinal disease caused by ABCA4 mutations. Prog Retin Eye Res, Apr 9:100861. doi: 10.1016/j.preteyeres.2020.100861.
Crognale, M.A., Fry, M., Highsmith, J., Haegerstrom-Portnoy, G., Neitz, M., Neitz, J., Webster, M.A., 2004. Characterization of a novel form of X-linked incomplete achromatopsia. Vis Neurosci 21, 197-203.
da Palma, M.M., Motta, F.L., Gomes, C.P., Salles, M.V., Pesquero, J.B., Sallum, J.M.F., 2020. Synonymous Variant in the CHM Gene Causes Aberrant Splicing in Choroideremia. Invest Ophthalmol Vis Sci 61(2), 38.
Davidson, A.E., Sergouniotis, P.I., Burgess-Mullan, R., Hart-Holden, N., Low, S., Foster, P.J., et al., 2010. A synonymous codon variant in two patients with autosomal recessive bestrophinopathy alters in vitro splicing of BEST1. Mol Vis 16, 2916-2922.
den Hollander, A.I., Koenekoop, R.K., Yzer, S., Lopez, I., Arends, M.L., Voesenek, K.E., et al., 2006. Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis. Am J Hum Genet 79, 556-561.
Dias, M.F., Joo, K., Kemp, J.A., Fialho, S.L., da Silva Cunha, A., Woo, S.J., Kwon, Y.J., 2018. Molecular genetics and emerging therapies for retinitis pigmentosa: Basic research and clinical perspectives. Prog Retin Eye Res 63, 107-131.
Desmet, F.O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., Béroud, C., 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Research 37, e67.
Desviat, L.R., Pérez, B., Ugarte, M., 2012. Minigenes to Confirm Exon Skipping Mutations. In: Aartsma-Rus A. (eds) Exon Skipping. Methods in Molecular Biology (Methods and Protocols), vol 867. Humana Press
Dhir, A., Buratti, E., 2010. Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies. FEBS J 277, 841-855.
Dooley, S.J., McDougald, D.S., Fisher, K.J., Bennicelli, J.L., Mitchell, L.G., Bennett, J., 2018. Spliceosome-Mediated Pre-mRNA trans-Splicing Can Repair CEP290 mRNA. Mol Ther Nucleic Acids 12, 294-308.
Donadon, I., Pinotti, M., Rajkowska, K., Pianigiani, G., Barbon, E., Morini, E., et al. , 2018. Exon-specific U1 snRNAs improve ELP1 exon 20 definition and rescue ELP1 protein expression in a familial dysautonomia mouse model. Hum Mol Genet 27, 2466-2476.
Dulla, K., Aguila, M., Lane, A., Jovanovic, K., Parfitt, D.A., Schulkens, I., et al., 2018. Splice-Modulating Oligonucleotide QR-110 Restores CEP290 mRNA and Function in Human c.2991+1655A>G LCA10 Models. Mol Ther Nucleic Acids 12, 730-740.
Erkelenz, S., Theiss, S., Otte, M., Widera, M., Peter, J.O., Schaal, H., 2014. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. Nucleic Acids Res 42, 10681–10697
Fairbrother, W.G., Yeo, G.W., Yeh, R., Goldstein, P., Mawson, M., Sharp, P.A., Burge, C.B., 2004. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. Nucleic Acids Res 32, 187–190.
Farkas, M.H., Grant, G.R., White, J.A., 2013. Transcriptome analyses 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 14, 486.
Franke, K., Berens, P., Schubert, T., Bethge, M., Euler, T., Baden, T., 2017. Inhibition decorrelates visual feature representations in the inner retina. Nature 542, 439-444.
Frischmeyer, P.A., Dietz, H.C., 1999. Nonsense-mediated mRNA decay in health and disease. Hum Mol Genet 8, 1893-1900.
Gao, K., Masuda, A., Matsuura, T., Ohno, K., 2008. Human branch point consensus sequence is yUnAy. Nucleic Acids Res 36, 2257-2267.
Garafalo, A.V., Cideciyan, A.V., Héon, E., Sheplock, R., Pearson, A, WeiYang Yu, C., et al., 2019. Progress in treating inherited retinal diseases: Early subretinal gene therapy clinical trials and candidates for future initiatives Prog Retin Eye Res, 100827. doi:10.1016/j.preteyeres.2019.100827

Garanto, A., van Beersum, S.E.C., Peters, T.A., Roepman, R., Cremers, F.P.M., Collin, R.W.J., 2013. Unexpected CEP290 mRNA splicing in a humanized knock-in mouse model for leber congenital amaurosis. PLoS ONE 8, e79369.
Garanto, A., Chung, D.C., Duijkers, L., Corral-Serrano, J.C., Messchaert, M., Xiao, R., et al.,
2016. In vitro and in vivo rescue of aberrant splicing in CEP290-associated LCA by antisense oligonucleotide delivery. Hum Mol Genet 25, 2552-2563.
Garanto, A., van der Velde-Visser, S.D., Cremers, F.P.M., Collin, R.W.J., 2018. Antisense Oligonucleotide-Based Splice Correction of a Deep-Intronic Mutation in CHM Underlying Choroideremia. Adv Exp Med Biol 1074, 83-89.
Gardner, J.C., Liew, G., Quan, Y.H., Ermetal, B., Ueyama, H., Davidson, A.E., et al., 2014. Three different cone opsin gene array mutational mechanisms with genotype-phenotype correlation and functional investigation of cone opsin variants. Hum Mutat 35, 1354-1362.
Gerard, X., Perrault, I., Hanein, S., Silva, E., Bigot, K., Defoort-Delhemmes, S., et al., 2012. AON-mediated Exon Skipping Restores Ciliation in Fibroblasts Harboring the Common Leber Congenital Amaurosis CEP290 Mutation. Mol Ther Nucleic Acids 1, e29.
Gérard, X., Perrault, I., Munnich, A., Kaplan, J., Rozet, J.M., 2015. Intravitreal Injection of Splice-switching Oligonucleotides to Manipulate Splicing in Retinal Cells. Mol Ther Nucleic Acids 4, e250.
Gilbert, W.,1978. Why genes in pieces? Nature 271, 501.
Glaus, E., Schmid, F., Da Costa, R., Berger, W., Neidhardt, J., 2011. Gene therapeutic approach using mutation-adapted U1 snRNA to correct a RPGR splice defect in patient-derived cells. Mol Ther 19, 936-941.
Goemans, N.M., Tulinius, M., van den Akker, J.T., Burm, B.E., Ekhart, P.F., Heuvelmans, N., et al.,
2011. Systemic administration of PRO051 in Duchenne's muscular dystrophy. N Engl J Med 364, 1513-1522.
Goemans, N.M., Tulinius, M., van den Hauwe, M., Kroksmark, A.K., Buyse, G., Wilson, R.J., et al., 2016. Long-Term Efficacy, Safety, and Pharmacokinetics of Drisapersen in Duchenne Muscular Dystrophy: Results from an Open-Label Extension Study. PLoS One 11, e0161955.
Gonorazky, H.D., Naumenko, S., Ramani, A.K., Nelakuditi, V., Mashouri, P., Wang, P., et al., 2019. Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. Am J Hum Genet 104, 1007.
Greenwald, S.H., Kuchenbecker, J.A., Rowlan, J.S., Neitz, J., Neitz, M., 2017. Role of a Dual Splicing and Amino Acid Code in Myopia, Cone Dysfunction and Cone Dystrophy Associated with L/M Opsin Interchange Mutations. Transl Vis Sci Technol 6, 2.
Grodecká, L., Buratti, E., Freiberger, T., 2017. Mutations of pre-mRNA splicing regulatory elements: Are predictions moving forward to clinical diagnostics? Int J Mol Sci 18, 1668.
GTEx Consortium, 2017. Genetic effects on gene expression across human tissues. Nature 550, 204-213.
Grünert, U., Martin, P.R., 2020. Cell types and cell circuits in human and non-human primate retina. Prog Retin Eye Res Feb 5:100844. doi: 10.1016/j.preteyeres.2020.100844.
Gu, S.M., Thompson, D.A., Srikumari, C.R., Lorenz, B., Finckh, U., Nicoletti, A., et al., 1997. Mutations in RPE65 cause autosomal recessive childhood-onset severe retinal dystrophy. Nat Genet 17, 194-197.
Hong, D.H., Li, T., 2002. Complex expression pattern of RPGR reveals a role for purine-rich exonic splicing enhancers. Invest Ophthalmol Vis Sci 43, 3373-3382.
Hoon, M., Okawa, H., Della Santina, L., Wong, R.O., 2014. Functional architecture of the retina: development and disease. Prog Retin Eye Res 42, 44-84.
Hubé, F., Francastel, C., 2015. Mammalian introns: when the junk generates molecular diversity. Int J Mol Sci 16, 4429-4452.
Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., et al., 2012. Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep 1, 167–178.
Irimia, M., Weatheritt, R.J., Ellis, J.D., Parikshak, N.N., Gonatopoulos-Pournatzis, T., Babor, M., et al., 2014. A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell 159(7), 1511-1523.
Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., et al., 2019. Predicting Splicing from Primary Sequence with Deep Learning. Cell 176, 535-548.
Jamshidi, F., Place, E.M., Mehrotra, S., Navarro-Gomez, D., Maher, M., Branham, K.E., et al., 2019. Contribution of noncoding pathogenic variants to RPGRIP1-mediated inherited retinal degeneration. Genet Med 21(3), 694-704.
Jian, X., Boerwinkle, E., Liu, X., 2014. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. Genet Med 16, 497-503.
Johnson, A.A., Guziewicz, K.E., Lee, C.J., Kalathur, R.C., Pulido, J.S., Marmorstein, L.Y., Marmorstein, A.D., 2017. Bestrophin 1 and retinal disease. Prog Retin Eye Res 58, 45-69.
Kapustin, Y., Chan, E., Sarkar, R., Wong, F., Vorechovsky, I., Winston, R.M., et al., 2011. Cryptic splice sites and split genes. Nucleic Acids Res 39, 5837–5844.
Katz, Y., Wang, E.T., Silterra, J., Schwartz, S., Wong, B., Mesirov J.P., et al., 2013. Sashimi plots: Quantitative visualization of RNA sequencing read alignments. arXiv:1306.3466 [q-bio.GN]
Ke, S., Shang, S., Kalachikov, S.M., Morozova, I., Yu, L., Russo, J.J., et al., 2011. Quantitative evaluation of all hexamers as exonic splicing elements. Genome Res 21, 1360–1374.
Khan, A.O., Becirovic, E., Betz, C., Neuhaus, C., Altmüller, J., Riedmayr, L.M., et al., 2017. A deep intronic CLRN1 (USH3A) founder mutation generates an aberrant exon and underlies severe Usher syndrome on the Arabian Peninsula. Sci Rep 7(1), 1411.
Khan, M., Fadaie, Z., Cornelis, S.S., Cremers, F.P.M., Roosing, S., 2019. Identification and Analysis of Genes Associated with Inherited Retinal Diseases. Methods Mol Biol 1834, 3-27.
Kim, E., Goren, A., Ast, G., 2008. Alternative splicing: current perspectives. Bioessays 30, 38-47.
Kim, J.W., Yang, H.J., Brooks, M.J., Zelinger, L., Karakülah, G., Gotoh, N., et al., 2016. NRL-Regulated Transcriptome Dynamics of Developing Rod Photoreceptors. Cell Rep 17, 2460-2473.
Kloeckener-Gruissem, B., Amstutz, C. VCAN-Related Vitreoretinopathy. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, Amemiya A, editors. Source GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2019.
Kohl, S., Zobor, D., Chiang, W.C., Weisschuh, N., Staller, J., Gonzalez Menendez, I., et al., 2015. Mutations in the unfolded protein response regulator ATF6 cause the cone dysfunction disorder achromatopsia. Nat Genet 47, 757-765.
Krausová, M., Staněk, D., 2018. snRNP proteins in health and disease. Semin Cell Dev Biol 79, 92-102.
Krawczak, M., Reiss, J., Cooper, D.N., 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 90, 41–54.
Krawczak, M., Thomas, N.S., Hundrieser, B., Mort, M., Wittig, M., Hampe, J., et al., 2007. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. Hum Mutat 28, 150–158.

Kuivenhoven, J.A., Weibusch, H., Pritchard, P.H., Funke, H., Benne, R., Assmann, G., et al., 1996. An intronic mutation in a lariat branchpoint sequence is a direct cause of an inherited human disorder (fish-eye disease). J Clin Invest 98, 358-364.
Lee, N.C., Lee, Y.M., Chen, P.W., Byrne, B.J., Hwu, W.L., 2016. Mutation-adapted U1 snRNA corrects a splicing error of the dopa decarboxylase gene. Hum Mol Genet 25, 5142-5147.
Lee, M., Roos, P., Sharma, N., Atalar, M., Evans, T.A., Pellicore, M.J., et al., 2017. Systematic Computational Identification of Variants That Activate Exonic and Intronic Cryptic Splice Sites. Am J Hum Genet 100, 751-765.
Lee, J.H., Wang, J.H., Chen, J., Li, F., Edwards, T.L., Hewitt, A.W., Liu, G.S. 2019. Gene therapy for visual loss: Opportunities and concerns. Prog Retin Eye Res 68, 31-53.
Lei, Q., Li, C., Zuo, Z., Huang, C., Cheng, H., Zhou, R., 2016. Evolutionary Insights into RNA trans-Splicing in Vertebrates. Genome Biol Evol 8, 562-577.
Lenassi, E., Saihan, Z., Bitner-Glindzicz, M., Webster, A.R., 2014. The effect of the common c.2299delG mutation in USH2A on RNA splicing. Exp Eye Res 122, 9-12.
Lentz, J.J., Jodelka, F.M., Hinrich, A.J., McCaffrey, K.E., Farris, H.E., Spalitta, M.J., et al., 2013. Rescue of hearing and vestibular function by antisense oligonucleotides in a mouse model of human deafness. Nat Med 19, 345-350.
Li, Y., Furhang, R., Ray, A., Duncan, T., Soucy, J., Mahdi, R., et al., 2019. Aberrant RNA splicing is the major pathogenic effect in a knock-in mouse model of the dominantly inherited c.1430A>G human RPE65 mutation. Hum Mutat 40, 426-443.
Liang, Q., Dharmat, R., Owen, L., Shakoor, A., Li, Y., Kim, S., et al., 2019. Single-nuclei RNA-seq on human retinal tissue provides improved transcriptome profiling. Nat Commun 10, 5743.
Lim, K.H., Ferraris, L., Filloux, M.E., Raphael, B.J., Fairbrother, W.G., 2011. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. Proc Natl Acad Sci U S A 108, 11093–11098.
Liquori, A., Vaché, C., Baux, D., Blanchet, C., Hamel, C., Malcolm, S., et al. 2016 Whole USH2A Gene Sequencing Identifies Several New Deep Intronic Mutations. Hum Mutat 37(2), 184-193.
Liu, X., Jiang, Q., Mansfield, S.G., Puttaraju, M., Zhang, Y., Zhou, W., et al., 2002. Partial correction of endogenous DeltaF508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA trans-splicing. Nat Biotechnol 20, 47-52.
Llavona, P., Pinelli, M., Mutarelli, M., Marwah, V.S., Schimpf-Linzenbold, S., Thaler, S., et al., 2017. Allelic Expression Imbalance in the Human Retinal Transcriptome and Potential Impact on Inherited Retinal Diseases. Genes (Basel) 8, pii: E283.
Lorson, C.L., Rindt, H., Shababi. M., 2010. Spinal muscular atrophy: mechanisms and therapeutic strategies. Hum Mol Genet 19(R1), R111-118.
Lukovic, D., Artero Castro, A., Delgado, A.B., Bernal Mde, L., Luna Pelaez, N., Diez Lloret, A., et al., 2015. Human iPSC derived disease model of MERTK-associated retinitis pigmentosa. Sci Rep 5, 12910.
Lukowski, S.W., Lo, C.Y., Sharov, A.A., Nguyen, Q., Fang, L., Hung, S.S., et al., 2019. A single-cell transcriptome atlas of the adult human retina. EMBO J 38, e100811.
Maeder, M.L., Stefanidakis, M., Wilson, C.J., Baral, R., Barrera, L.A., Bounoutas, G.S., et al., 2019. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. Nat Med 25, 229-233.
Magen, A., Ast, G., 2005. The importance of being divisible by three in alternative splicing. Nucleic Acids Res 33, 5574-5582.
Marco-Puche, G., Lois, S., Benítez, J., Trivino, J.C., 2019. RNA-Seq Perspectives to Improve Clinical Diagnosis. Front Genet 10, 1152.
Matlin, A.J., Clark, F., Smith, C.W., 2005. Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol 6, 386-398.
Matera, A.G., Wang, Z., 2014. A day in the life of the spliceosome. Nat Rev Mol Cell Biol 15, 108–121.
Matos, L., Canals, I., Dridi, L., Choi, Y., Prata, M.J., Jordan, P., et al., 2014. Therapeutic strategies based on modified U1 snRNAs and chaperones for Sanfilippo C splicing mutations. Orphanet J Rare Dis 9, 180.
Maugeri, A., van Driel, M.A., van de Pol, D.J., Klevering, B.J., van Haren, F.J., Tijmes, N., et al., 1999. The 2588G-->C mutation in the ABCR gene is a mild frequent founder mutation in the Western European population and allows the classification of ABCR mutations in patients with Stargardt disease. Am J Hum Genet 64, 1024-1035.
Mayer, A.K., Rohrschneider, K., Strom, T.M., Glöckle, N., Kohl, S., Wissinger, B., et al., 2016. Homozygosity mapping and whole-genome sequencing reveals a deep intronic PROM1 mutation causing cone-rod dystrophy by pseudoexon activation. Eur J Hum Genet 24, 459-462.
McClements, M., Davies, W.I., Michaelides, M., Young, T., Neitz, M., MacLaren, R.E., et al., 2013. Variations in opsin coding sequences cause x-linked cone dysfunction syndrome with myopia and dichromacy. Invest Ophthalmol Vis Sci 54, 1361-1369.
McClorey, G., Moulton, H.M., Iversen, P.L., Fletcher, S., Wilton, S.D., 2006. Antisense oligonucleotide-induced exon skipping restores dystrophin expression in vitro in a canine model of DMD. Gene Ther 3, 1373-1381.
Mehmood, A., Laiho, A., Venäläinen, M.S., McGlinchey, A.J., Wang, N., Elo, L.L., 2019. Systematic evaluation of differential splicing tools for RNA-seq studies. Brief Bioinform Dec 5. pii: bbz126. doi: 10.1093/bib/bbz126.
Mellough, C.B., Bauer, R., Collin, J., Dorgau, B., Zerti, D., Dolan, D.W.P., et al., 2019. An integrated transcriptional analysis of the developing human retina. Development 146, pii: dev169474.
Mendell, J.R., Goemans, N., Lowes, L.P., Alfano, L.N., Berry, K., Shao, J., et al. 2016. Longitudinal effect of eteplirsen versus historical control on ambulation in Duchenne muscular dystrophy. Ann Neurol 79, 257-271.
Mizrahi-Meissonnier, L., Merin, S., Banin, E., Sharon, D., 2010. Variable retinal phenotypes caused by mutations in the X-linked photopigment gene array. Invest Ophthalmol Vis Sci 51, 3884-3892.
Mukhopadhyay, A., Nikopoulos, K., Maugeri, A., de Brouwer, A.P., van Nouhuys, C.E., Boon, C.J., et al., 2006. Erosive vitreoretinopathy and wagner disease are caused by intronic mutations in CSPG2/Versican that result in an imbalance of splice variants. Invest Ophthalmol Vis Sci 47, 3565-3572.
Murphy, D., Cieply, B., Carstens, R., Ramamurthy, V., Stoilov, P., (2016) The Musashi 1 Controls the Splicing of Photoreceptor-Specific Exons in the Vertebrate Retina. PLoS Genet 12, e1006256.
Nakai, K., Sakamoto, H., 1994. Construction of a novel database containing aberrant splicing mutations of mammalian genes. Gene 141, 171-177.
Nathans, J., Thomas, D., Hogness, D.S., 1986. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. Science 232, 193-202.
Naryshkin, N.A., Weetall, M., Dakka, A., Narasimhan, J., Zhao, X., Feng, Z., et al., 2014. Motor neuron disease. SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. Science 345, 688-693.
Nellore, A., Jaffe, A.E., Fortin, J.P., Alquicira-Hernández, .J, Collado-Torres, L., Wang, S., et al. 2016. Human splicing diversity and the extent of unannotated splice junctions across human RNA-seq samples on the Sequence Read Archive. Genome Biol 17, 266.
Orkin, S.H., Goff, S.C., Hechtman, R.L., 1981. Mutation in an intervening sequence splice junction in man. Proc Natl Acad Sci U S A. 78, 5041-5045

Palacino, J., Swalley, S.E., Song, C., Cheung, A.K., Shu, L., Zhang, X., et al., 2015. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. Nat Chem Biol 11, 511-517.
Pagani, F., Stuani, C., Tzetis, M., Kanavakis, E., Efthymiadou, A., Doudounakis, S., et al., 2003. New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum Mol Genet 12(10), 1111-1120.
Parfitt, D.A., Lane, A., Ramsden, C., Jovanovic, K., Coffey, P.J., Hardcastle, A.J., et al., 2016. Using induced pluripotent stem cells to understand retinal ciliopathy disease mechanisms and develop therapies. Biochem Soc Trans 44, 1245–1251.
Park, E., Pan, Z., Zhang, Z., Lin, L., Xing, Y., 2018. The Expanding Landscape of Alternative Splicing Variation in Human Populations. Am J Hum Genet 102, 11-26.Pinelli, M., Carissimo, A., Cutillo, L., Lai, C.H., Mutarelli, M., Moretti, M.N., etal., 2016. An atlas of gene expression and gene co-regulation in the human retina. Nucleic Acids Res 44, 5773-5784.
Peng, Y.R., Shekhar, K., Yan, W., Herrmann, D., Sappington, A., Bryman, G.S., et al., 2019. Molecular classification and comparative taxonomics of foveal and peripheral cells in primate retina. Cell 176(5), 1222–1237.
Piovesan, A., Caracausi, M., Antonaros, F., Pelleri, M.C., Vitale, L., 2016. GeneBase 1.1: a tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. Database (2016) Vol. 2016: article ID baw153; doi:10.1093/database/baw153.
Piva, F., Giulietti, M., Burini, A.B., Principato, G., 2012. SpliceAid 2: A database of human splicing factors expression data and RNA target motifs. Hum Mutat 33, 81–85.
Pulyakhina, I., Gazzoli, I., 't Hoen, P.A., Verwey, N., den Dunnen, J.T., Aartsma-Rus, A., et al., 2015. SplicePie: a novel analytical approach for the detection of alternative, non-sequential and recursive splicing. Nucleic Acids Res 43, 11068.
Raj, T., Li, Y.I., Wong, G., Humphrey, J., Wang, M., Ramdhani, S., et al., 2018. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. Nat Genet 50, 1584-1592.
Ramsbottom, S.A., Molinari, E., Srivastava, S., Silberman, F., Henry, C., Alkanderi, S., et al., 2018. Targeted exon skipping of a CEP290 mutation rescues Joubert syndrome phenotypes in vitro and in a murine model. Proc Natl Acad Sci U S A 115, 12489-12494.
Ratnapriya, R., Sosina, O.A., Starostik, M.R., Kwicklis, M., Kapphahn, R.J., Fritsche, L.G., et al., 2019. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. Nat Genet 51, 606-610.
Ratni, H., Mueller, L., Ebeling, M., 2019. Rewriting the (tran)script: Application to spinal muscular atrophy. Prog Med Chem 58, 119-156.
Rattner, A., Smallwood, P.M., Williams, J., Cooke, C., Savchenko, A., Lyubarsky, A., et al., 2001. A photoreceptor-specific cadherin is essential for the structural integrity of the outer segment and for photoreceptor survival. Neuron 32, 775-786.
Rattner, A., Chen, J., Nathans, J., 2004. Proteolytic shedding of the extracellular domain of photoreceptor cadherin. Implications for outer segment assembly. J Biol Chem 279, 42202–42210.
Redmond, T.M., Yu, S., Lee, E., Bok, D., Hamasaki, D., Chen, N., et al., 1998. Rpe65 is necessary for production of 11-cis-vitamin A in the retinal visual cycle. Nat Genet 20, 344-351.
Reed, R., Maniatis, T., 1985. Intron sequences involved in lariat formation during pre-mRNA splicing. Cell 41, 95-105.
Reese, M.G., Eeckman, F.H., Kulp, D., Haussler, D., 1997. Improved splice site detection in Genie. J Comput Biol 4, 311-323.
Rheaume, B.A., Jereen, A., Bolisetty, M., Sajid, M.S., Yang, Y., Renna, K., et al., 2018. Single cell transcriptome profiling of retinal ganglion cells identifies cellular subtypes. Nat Commun 9, 2759.
Rimoldi, V., Soldà, G., Asselta, R., Spena, S., Stuani, C., Buratti, E., Duga, S., 2013. Dual Role of G-runs and hnRNP F in the Regulation of a Mutation-Activated Pseudoexon in the Fibrinogen Gamma-Chain Transcript. PLoS One 8, e59333.
Rio Frio, T., McGee, T.L., Wade, N.M., Iseli, C., Beckmann, J.S., Berson, E.L., Rivolta, C., 2009. A single-base substitution within an intronic repetitive element causes dominant retinitis pigmentosa with reduced penetrance. Hum Mutat 30(9), 1340-1347
Rivera, A., White, K., Stöhr, H., Steiner, K., Hemmrich, N., Grimm, T., et al., 2000. A comprehensive survey of sequence variation in the ABCA4 (ABCR) gene in Stargardt disease and age-related macular degeneration. Am J Hum Genet 67, 800-813.
Rogalska, M.E., Tajnik, M., Licastro, D., Bussani, E., Camparini, L., Mattioli, C., Pagani, F., 2016. Therapeutic activity of modified U1 core spliceosomal particles. Nat Commun 7, 11168.
Roosing, S., Thiadens, A.A., Hoyng, C.B., Klaver, C.C., den Hollander, A.I., Cremers, F.P., 2014. Causes and consequences of inherited cone disorders. Prog Retin Eye Res 42, 1-26.
Rosenberg, A.B., Patwardhan, R.P., Shendure, J., Seelig, G., 2015. Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences. Cell 163, 698–711.
Ruskin, B., Krainer, A.R., Maniatis, T., Green, M.R., 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell 38, 317-331.
Růžičková, Š., Staněk, D., 2017. Mutations in spliceosomal proteins and retina degeneration. RNA Biol 14, 544-552.
Sakabe, N.J., de Souza, S.J., 2007. Sequence features responsible for intron retention in human. BMC Genomics 8, 59.
Sakharkar, M.K., Chow, V.T., Kangueane, P., 2004. Distributions of exons and introns in the human genome. In Silico Biol 4, 387-393.
Sangermano, R., Bax, N.M., Bauwens, M., van den Born, L.I., De Baere, E., Garanto, A., et al., 2016. Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T→C Mutation in Stargardt Disease. Ophthalmology 123, 1375-1385.
Sangermano, R., Khan, M., Cornelis, S.S., Richelle, V., Albert, S., Garanto, A., et al., 2018. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. Genome Res 28, 100–110.
Sangermano, R., Garanto, A., Khan, M., Runhart, E.H., Bauwens, M., Bax, N.M., et al., 2019. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. Genet Med 2019 Jan 15.
Schmid, F., Hiller, T., Korner, G., Glaus, E., Berger, W., Neidhardt, J., 2013. A gene therapeutic approach to correct splice defects with modified U1 and U6 snRNPs. Hum Gene Ther 24, 97-104.
Schulz, H.L., Grassmann, F., Kellner, U., Spital, G., Rüther, K., Jägle, H., et al., 2017. Mutation Spectrum of the ABCA4 Gene in 335 Stargardt Disease Patients from a Multicenter German Cohort-Impact of Selected Deep Intronic Variants and Common SNPs. Invest Ophthalmol Vis Sci 58, 394-403.
Scotti, M.M., Swanson, M.S., 2015. RNA mis-splicing in disease. Nat Rev Genet 17, 19-32.
Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 15, 7155–7174.
Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., 2016. Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell 166, 1308-1323.

Slaugenhaupt, S.A., Mull, J., Leyne, M., Cuajungco, M.P., Gill, S.P., Hims, M.M., et al., 2004. Rescue of a human mRNA splicing defect by the plant cytokinin kinetin. Hum Mol Genet 13, 429-436.

Slijkerman, R.W., Song, F., Astuti, G.D., Huynen, M.A., van Wijk, E., Stieger, K., Collin, R.W., 2015. The pros and cons of vertebrate animal models for functional and therapeutic research on inherited retinal dystrophies. Prog Retin Eye Res 48, 137-159.

Slijkerman, R.W., Vaché, C., Dona, M., García-García, G., Claustres, M., Hetterschijt, L., et al. 2016. Antisense Oligonucleotide-based Splice Correction for USH2A-associated Retinal Degeneration Caused by a Frequent Deep-intronic Mutation. Mol Ther Nucleic Acids 5, e381.

Soemedi, R., Cygan, K.J., Rhine, C.L., Wang, J., Bulacan, C., Yang, J., et al., 2017. Pathogenic variants that alter protein code often disrupt splicing. Nat Genet.49, 848–855.

Soens, Z.T., Branch, J., Wu, S., Yuan, Z., Li, Y., Li, H., et al., 2017. Leveraging splice-affecting variant predictors and a minigene validation system to identify Mendelian disease-causing variants among exon-captured variants of uncertain significance. Hum Mutat 38, 1521-1533.

Sothilingam, V., Garcia Garrido, M., Jiao, K., Buena-Atienza, E., Sahaboglu, A., Trifunović, D., et al., 2015. Retinitis pigmentosa: impact of different Pde6a point mutations on the disease phenotype.
Hum Mol Genet 24, 5486-5499.

Stingl, K., Mayer, A.K., Llavona, P., Mulahasanovic, L., Rudolph, G., Jacobson, S.G., et al., 2017. CDHR1 mutations in retinal dystrophies. Sci Rep 7, 6992.

Sullivan LS, Bowne SJ, Birch DG, Hughbanks-Wheaton D, Heckenlively JR, Lewis RA, et al., 2006. Prevalence of disease-causing mutations in families with autosomal dominant retinitis pigmentosa: a screen of known genes in 200 families. Invest Ophthalmol Vis Sci 47, 3052–3064.

Schwartz, S., Hall, E., Ast, G., 2009. SROOGLE: Webserver for integrative, user-friendly visualization of splicing signals. Nucleic Acids Res 37, 189–192.

Takahara, K., Schwarze, U., Imamura, Y., Hoffman, G.G., Toriello, H., Smith, L.T., et al., 2002. Order of intron removal influences multiple splice outcomes, including a two-exon skip, in a COL5A1 acceptorsite mutation that results in abnormal pro-alpha1(V) N-propeptides and Ehlers-Danlos syndrome type I. Am J Hum Genet 71, 451–465.

Tammaro, C., Raponi, M., Wilson, D. I., Baralle, D, 2014. BRCA1 EXON 11, a CERES (composite regulatory element of splicing) element involved in splice regulation. Int J Mol Sci 15, 13045-13059.

Tan, E., Ding, X.Q., Saadi, A., Agarwal, N., Naash, M.I., Al-Ubaidi, M.R., 2004. Expression of cone-photoreceptor-specific antigens in a cell line derived from retinal tumors in transgenic mice. Invest Ophthalmol Vis Sci 45, 764-768.

Tanackovic, G., Ransijn, A., Thibault, P., Abou Elela, S., Klinck, R., Berson, E.L., et al., 2011. PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. Hum Mol Genet 20, 2116-2130.

Tanner, G., Glaus, E., Barthelmes, D., Ader, M., Fleischhauer, J., Pagani, F., et al., 2009. Therapeutic strategy to rescue mutation-induced exon skipping in rhodopsin by adaptation of U1 snRNA. Hum Mutat 30, 255-263.

Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14, 178-192.

Tucker, B.A., Cranston, C.M., Anfinson, K.A., Shrestha, S., Streb, L.M., Leon, A., et al., 2015. Using patient-specific induced pluripotent stem cells to interrogate the pathogenicity of a novel retinal pigment epithelium-specific 65 kDa cryptic splice site mutation and confirm eligibility for enrollment into a clinical gene augmentation trial. Transl Res 166, 740-749.e1.

Turunen, J.J., Niemela, E.H., Verma, B., Frilander, M.J., 2013. The significant other: splicing by the minor spliceosome. Wiley Interdiscip Rev RNA 4, 61–76.

Ueyama, H., Muraki-Oda, S., Yamade, S., Tanabe, S., Yamashita, T., Shichida, Y., Ogita, H., 2012. Unique haplotype in exon 3 of cone opsin mRNA affects splicing of its precursor, leading to congenital color vision defect. Biochem Biophys Res Commun 424, 152-157.

Vaché, C., Besnard, T., le Berre, P., García-García, G., Baux, D., Larrieu, L., et al. 2012. Usher syndrome type 2 caused by activation of an USH2A pseudoexon: implications for diagnosis and therapy. Hum Mutat 33(1), 104-108.

van den Hurk, J.A., van de Pol, D.J., Wissinger, B., van Driel, M.A., Hoefsloot, L.H., de Wijs, I.J., et al., 2003. Novel types of mutation in the choroideremia (CHM) gene: a full-length L1 insertion and an intronic mutation activating a cryptic exon. Hum Genet 113(3), 268-275.

Vervoort, R., Lennon, A., Bird, A.C., Tulloch, B., Axton, R., Miano, M.G., et al., 2000. Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. Nat Genet 25, 462-466.

Voigt, A.P., Whitmore, S.S., Flamme-Wiese, M.J., Riker, M.J., Wiley, L.A., Tucker, B.A., et al., 2019. Molecular characterization of foveal versus peripheral human retina by single-cell RNA sequencing. Exp Eye Res 184, 234–242.

Wan, J., Masuda, T., Hackler, L., Torres, K.M., Merbs, S.L., Zack, D.J., et al., 2011. Dynamic usage of alternative splicing exons during mouse retina development. Nucleic Acids Res 39, 7920-7930.

Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., Burge, C.B., 2004. Systematic identification and analysis of exonic splicing silencers. Cell 119, 831–845.

Wang, Z., Xiao, X., Van Nostrand, E., Burge, C.B., 2006. General and specific functions of exonic splicing silencers in splicing control. Mol Cell 23, 61-70.

Wang, G.S., Cooper, T.A., 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet 8, 749-761.

Webb, T.R., Parfitt, D.A., Gardner, J.C., Martinez, A., Bevilacqua, D., Davidson, A.E., et al., 2012. Deep intronic mutation in OFD1, identified by targeted genomic next-generation sequencing, causes a severe form of X-linked retinitis pigmentosa (RP23). Hum Mol Genet 21(16), 3647-3654.

Weisschuh, N., Sturm, M., Baumann, B., Audo, I., Ayuso, C., Bocquet, B., et al., 2020. Deep-intronic variants in CNGB3 cause achromatopsia by pseudoexon activation. Hum Mutat 41(1), 255-264.

Weng, J., Mata, N.L., Azarian, S.M., Tzekov, R.T., Birch, D.G., Travis, G.H., 1999. Insights into the function of Rim protein in photoreceptors and etiology of Stargardt's disease from the phenotype in abcr knockout mice. Cell 98, 13-23.

Wong, J.J., Au, A.Y., Ritchie, W., Rasko, J.E., 2016. Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. Bioessays 38, 41-49.

Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K., et al. 2015. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science 347, 1254806.

Yang, Z., Chen, Y., Lillo, C., Chien, J., Yu, Z., Michaelides, M., et al., 2008. Mutant prominin 1 found in patients with macular degeneration disrupts photoreceptor disk morphogenesis in mice. J Clin Invest 118, 2908-2916.

Yardley, J., Leroy, B.P., Hart-Holden, N., Lafaut, B.A., Loeys, B., Messiaen, L.M., et al., 2004. Mutations of VMD2 splicing regulators cause nanophthalmos and autosomal dominant vitreoretinochoroidopathy (ADVIRC). Invest Ophthalmol Vis Sci 45:3683-3689.

Yeo, G., Burge, C.B., 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 11, 377–394.

Yoshida, T., Ozawa, Y., Suzuki, K., Yuki, K., Ohyama, M., Akamatsu, W., et al., 2014. The use of induced pluripotent stem cells to reveal pathogenic gene mutations and explore treatments for retinitis pigmentosa. Mol Brain 7, 45.

Yoshida, M., Kataoka, N., Miyauchi, K., Ohe, K., Iida, K., Yoshida, S., et al., 2015. Rectifier of aberrant mRNA splicing recovers tRNA modification in familial dysautonomia. Proc Natl Acad Sci U S A 112, 2764-2769.

Zaneveld, J., Siddiqui, S., Li, H., Wang, X., Wang, H., Wang, K., et al., 2015. Comprehensive analysis of patients with Stargardt macular dystrophy reveals new genotype-phenotype correlations and unexpected diagnostic revisions. Genet Med 17, 262-270.

Zelinger, L., Swaroop, A., 2018. RNA Biology in Retinal Development and Disease. Trends Genet 34, 341-351.

Zernant, J., Xie, Y.A., Ayuso, C., Riveiro-Alvarez, R., Lopez-Martinez, M.A., Simonelli, F., et al., 2014. Analysis of the ABCA4 genomic locus in Stargardt disease. Hum Mol Genet 23, 6797-6806.

Zernant, J., Lee, W., Collison, F.T., Fishman, G.A., Sergeev, Y.V., Schuerch, K. et al., 2017. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. J Med Genet 54, 404-412.

Zhang, W., Li, L., Su, Q., Gao, G., Khanna, H., 2018. Gene Therapy Using a miniCEP290 Fragment Delays Photoreceptor Degeneration in a Mouse Model of Leber Congenital Amaurosis. Hum Gene Ther 29, 42-50.

Zhang, X., Joehanes, R., Chen, B.H., Huan, T., Ying, S., Munson, P.J., 2015. Identification of common genetic variants controlling transcript isoform variation in human whole blood. Nat Genet 47, 345-352.

Zhao, C., Bellur, D.L., Lu, S., Zhao, F., Grassi, M.A., Bowne, S.J., et al., 2009. Autosomal-dominant retinitis pigmentosa caused by a mutation in SNRNP200, a gene required for unwinding of U4/U6 snRNAs. Am J Hum Genet 85, 617-627.

Zhou, X., Wu, W., Li, H., Cheng, Y., Wei, N., Zong, J., Feng, X., Xie, Z., Chen, D., Manley, J.L., 2014. Transcriptome analysis of alternative splicing events regulated by SRSF10 reveals position-dependent splicing modulation. Nucleic Acids Res 42, 4019–4030.

Zhuang, Y., Weiner, A.M., 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. Cell 46, 827-835.

List of Abbreviations

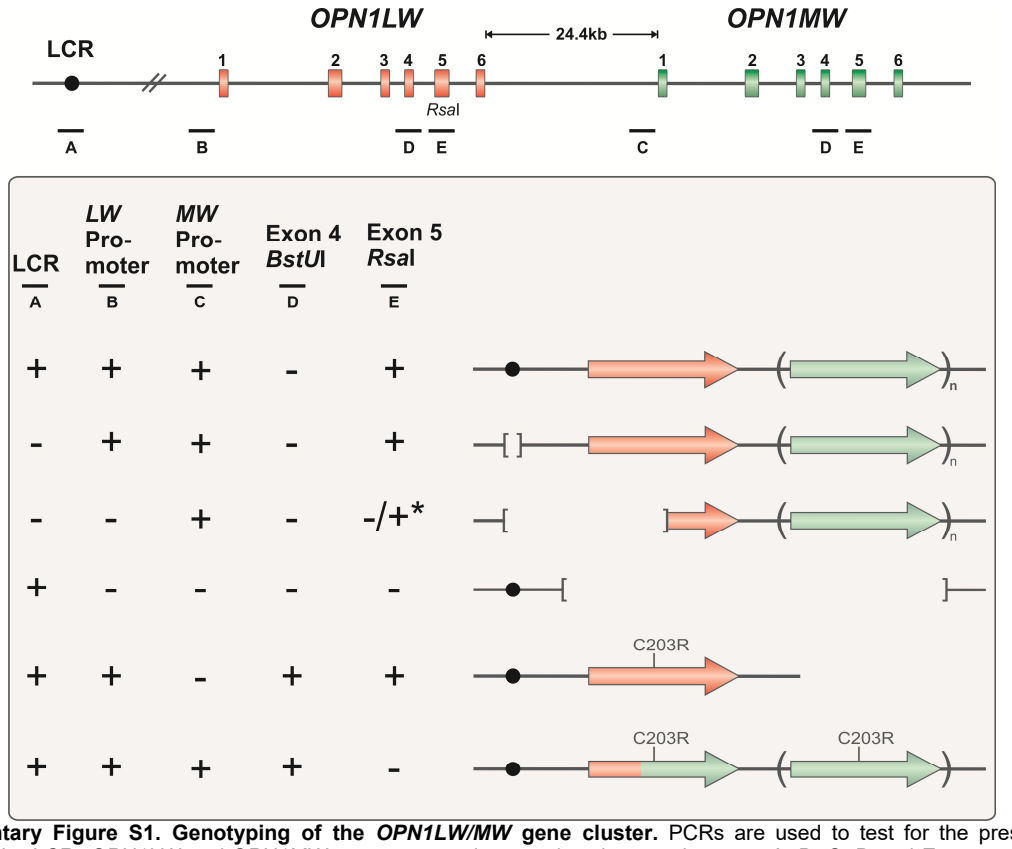
Abbreviation	Description
AAV	adeno associated virus
ABC	ATP-binding cassette
ADRP	autosomal dominant retinitis pigmentosa
ADVIRC	autosomal dominant vitreoretinochoroidopathy
AON	antisense oligonucleotide
ARP	autosomal recessive bestrophinopathy
AS	alternative splicing
bp	basepairs
BP	branch point
CAGI	critical assessment of genome interpretation
cDNA	complementary desoxyribonucleic acid
CERES	composite exonic regulatory elements of splicing
cSNP	coding single nucleotide polymorphism
DNA	desoxyribonucleic acid
eQTL	expression quantitative trait loci
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
FDA	Food and Drug Administration
HEK	human embryonic kidney
HGMD	Human Gene Mutation Database
hnRNP	heterogeneous nuclear ribonucleoprotein
HSF	Human Splicing Finder
iPSC	induced pluripotent stem cell
IRD	inherited retinal degeneration
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
IUPAC	International Union of Pure and Applied Chemistry
kb	kilobases (1000 bp)
LCA	Leber congenital amaurosis
LCR	locus control region
LWS	long-wavelength sensitive
mRNA	messenger ribonucleic acid
MWS	middle-wavelength sensitive
NMD	nonsense mediated decay
PCR	polymerase chain reaction
PE	pseudoexon
PPT	polypyrimidine tract
PTC	premature termination codon
PTM	pre-mRNA trans-splicing module
RECTAS	rectifier of aberrant splicing
RNA	ribonucleic acid
RNAseq	RNA sequencing
RPE	retinal pigmented epithelium
RT-PCR	reverse transcription-polymerase chain reaction
SMA	spinal muscular atrophy
SMaRT	spliceosome mediated RNA trans-splicing
SNP	single nucleotide polymorphism
snRNA	small nuclear ribonucleic acid
SR	serine-arginine rich

13.2 Supplementary Material

13.2.1 Supplementary material Introduction and Results

A gene-specific two-tiered genotyping assay for has been developed starting from most to least frequent mutations of the human *OPN1LW/MW* gene cluster. Targeted duplex- and simplex PCRs are used to test for the presence and integrity of the LCR, *OPN1LW* and *OPN1MW* promoters, and exons 4 and 5, respectively. The latter amplicons are to discern *OPN1LW* and *OPN1MW* genes and to detect the common c.607T>C/p.Cys203Arg mutation (Supp. Fig. S1).

For multiple-gene arrays, long-distance PCRs specifically amplifying each *OPN1LW* or *OPN1MW* by means of a differential primer binding site in the non-coding upstream region. Nested PCRs can be used to amplify each exon. Copy number assays based in quantitative PCR including male controls of known copy number enable determination of the total number of *OPN1LW/MW* opsin gene copies.



Supplementary Figure S1. Genotyping of the *OPN1LW/MW* gene cluster. PCRs are used to test for the presence and integrity of the LCR, *OPN1LW and OPN1MW* promoters, and exons 4 and 5 are shown as A, B, C, D and E, respectively. The RFLP based on *Rsa*I site in the exon 5 amplicon of *OPN1LW* only discriminates between p.277Tyr and p.277Phe. A novel *BstU*I site in the exon 4 amplicon is detected by RFLP when this carries the p.Cys203Arg (C203R) mutation. PCR products are obtained (+) or not (-). *- if exon 5 is deleted. Adapted from drawing of Prof. Bernd Wissinger.

13.2.2 Supplementary material Section 6.2

Screening of BCM262-III:2 and family members was performed using the standard BCM pipeline (Appendix 13.2.1) at the MGL in Tübingen. Amplification of exon 5 was performed with primers BCM-Ex5-F (5'-TCCAACCCCCGACTCAACTATC-3') and BCM-Ex5-R (5'-ACGGTATTTTGATGTGGATCTGCT-3'). RFLP with *Rsa*I was used for genotyping the dimorphism between *OPN1LW* and *OPN1MW* of BCM262-III:2 at amino acid position 27. To obtain increased amplification performance and longer sequences that include the surrounding the breakpoints of the deletion, the LD-PCRs as well as the re-amplification PCRs for exon 5, were optimized by using the following reverse primer RGCP2_intron5.2-Rv (5'- GCCGAGACTATAAAGGATTCC -3') which binds in intron 5 of both *OPN1LW* and *OPN1MW*; 208 bp further downstream as the RGCP2-5R primer (5'-GCGGACTAGTGATCTGCTGATGGTGTTGCTTA -3') used for the first amplification of exon 5 with the same amplification conditions and forward *OPN1LW/MW*-specific primers as in the standard protocol (Supp. Fig. S1C-E).

13.2.3 Supplementary material Section 6.4

RT-PCRs products from exon 12 to exon 15 of *Pde6b* and *Pde6a* were amplified from retinal transcripts of homozygous for c.1684C>T/p.Arg562Trp (p.R562W), c.1678C>T/p.Arg560Cys (*rd10*) and wild-type (C57BL/6 WT); Fig. 7. Primers used for amplifying transcripts for *Pde6a*: Mm_Pde6a_cDNA_ex12F_nw (5'- ACGCGGAGTCATACGAAATC -3') and Mm_Pde6a_cDNA_ex15R_nw (5'- ATGATGCCTTTCCAAGATGG -3') and *Pde6a*: Mm_Pde6b_cDNA_ex12F 5'- GCCGACCAAGTTTGACATCT -3') and Mm_Pde6b_cDNA_ex15R (5'- GTGGTGCCTTTCCAGAATTG -3') for *Pde6b*. Two human *PDE6A* exon 13 minigene constructs ("c.1684C" and "c.1684C>T/p.Arg562Trp") encompassing exon 13 and flanking intron sequences (336 bp of intron 12 and 590 bp of intron 13) were obtained analogously to the mouse minigene constructs (Publication III) by PCR amplification from a genomic DNA sample of the reference patient. Minigene constructs were generated by cloning both allelic products of a PCR with primers BamHI_PDE6A-H-r (5'- AAGGATCCAGCAAAGACTATAACCATGACCAA -3') and NotI_PDE6A-H-f (5'- AAGCGGCCGCAGCCCTTGTTAGCTTCTCCA -3)'.

13.2.4 Supplementary material Section 6.7

An overview of the methodology workflow is shown in Supp. Fig. S2.

RNA pulldown

Synthetic 5' biotinylated 109-nt ssRNA Ultramers (Scrambled, Reference haplotype and 'LIAVA' haplotype; Supp. Table S1) were chemically synthesized and purchased as RNAse-free and HPLC-purified from IDT. RNAs were gradually heated to 65ºC and slowly cooled down to 4ºC enabling RNA folding. Label-free whole lysates were prepared from HEK293T cells according to an established protocol (Agca et al., 2015). Briefly, cells were scraped from the wells with an RNA-binding solution and incubated at 4ºC for 15 min. RNA washing buffer contained 150mM NaCl, 50mM HEPES HCl pH 7.5, 0.1% NP40, 10mM MgCl2, and 10µl/ml of freshly added protein inhibitors Phosphatase Inhibitor Cocktail 2 (PI2) and Phosphatase Inhibitor Cocktail (PI3) (Sigma-Aldrich). The RNA-binding solution was further complemented with 20µl/ml of Proteinase Inhibitor Cocktail (Roche). After washing of the *Strep*-Tactin Superflow beads (IBA, Solutions for Life Sciences Göttingen, Germany), 2 µg of RNA bait were incubated on RNA-binding buffer with the washed beads for 30 min at 4ºC. Bead-bound RNA was incubated with 2 mg of whole lysate, yeast tRNA and 200 U of RNAse inhibitors for 1h at 4ºC. The biotin-bead-RNA-protein complexes were washed 3 times and bound proteins were eluted with Glycin buffer (200mM, pH 2.5) and subsequently neutralized by adding Tris buffer (1M Tris/HCl, pH 8).

Supplementary Material Table S1: Three RNA 109-bp Ultramers used as baits for the RNA pulldown assay.

Name Ultramer	GC (%)	MFE* (kcal/mol)	Oligonucleotide RNA Sequence (5'→3')
Scrambled	55.96	-33.70	GGUCGUGAGGCCUCAGUCUCGGCGGGUUCUUAAGGCUUGUCGUAGUCGUCAUGCC UGUCCGAUGCGCGUUAUUACAGACUCGGAUGAGAGUGCGUUAUGAAUCGGCGUG
Reference haplotype	55.05	-32.20	UUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAU GCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGG
'LIAVA' haplotype	56.88	-43.80	UUCCUGGGAGAGGUGGCUGGUGGUGUGCAAGCCCUUUGGCAAUGUGAGAUUUGAU GCCAAGCUGGCCAUCAUUGGCAUUGCCUUCUCCUGGGUCUGGGCUGCUGUGUGG

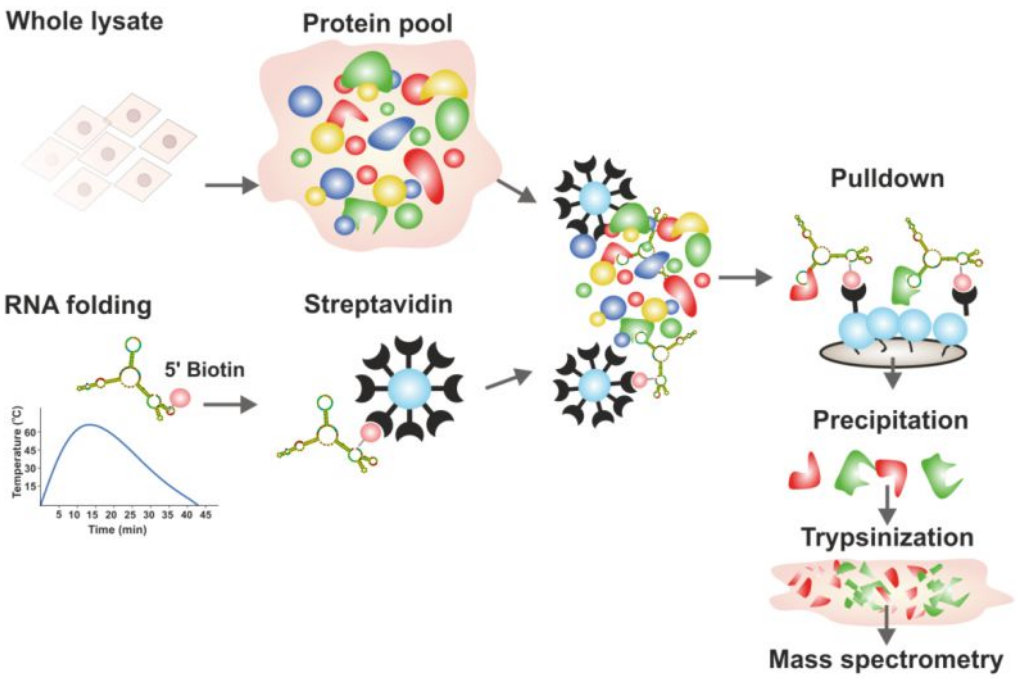
*Minimum Free Energy secondary structure (MFE) in kcal/mol was predicted by RNAfold from the Vienna RNA package (Gruber et al., 2008).

Mass spectrometry

Following affinity chromatography, eluted samples were precipitated with methanol and chloroform. In-solution trypsin cleavage was performed as previously described (Boldt et al., 2009). Upon dissolving precipitated samples in 30 µl, 50 mM ammonium bicarbonate (Sigma-Aldrich), supplemented with 2% RapiGest (Waters), samples were incubated with 1 µl 100 mM DTT (Merck) for 10 min at 60 °C. Following incubation with 1 µl 300 mM 2-iodacetamide at room temperature for 30 min in the dark, 1 µg of trypsin (Sigma-Aldrich, sequencing grade) was added to the reaction and was incubated overnight at 37 °C. Trifluoracetic acid was added to a final concentration of 1% to stop the reaction. Peptide samples were desalted and purified using stage tips before separation on a LC–MS/MS analysis which was performed on an Orbitrap Fusion™ system with standard parameters.

Proteomic data analysis

Raw files were processed with MaxQuant, Version 1.5.3.3,(Cox and Mann, 2008; Tyanova et al., 2016a) using the human SwissProt database 2014_11 for label-free quantification of proteins. Perseus (Version 1.5.5.3, doi:10.1038/nMeth.3901) was used to visualize and statistically validate the results. The peptide and protein false-discovery rates (FDR) were set to 1%. Quantification was performed with all, razor and unique peptides. An additional analysis was performed by using only unique peptides. Downstream analysis included significance A for identification of significant outliers with Benjamin-Hochberg FDR (threshold value 0.05) and a two-sample Student's T-test (Interval Confidence 95%, n = 6, being n the number of independent experiments) performed with Perseus software (Tyanova et al., 2016b; Tyanova and Cox, 2018)



Supplementary Figure S2. Workflow of the RNA-pulldown assay used to identify RNA-binding proteins as putative *trans*-acting splicing factors involved in splicing of *OPN1LW/MW* exon 3. HEK293T cell lysates were used as the pool of protein preys. Three RNA baits harbouring the sequences of interest in exon 3 (Reference or 'LIAVA' haplotypes) and a scrambled sequence were heated and slowly cooled down to allow RNA folding. The biotin molecule attached to the RNA baits enables binding with high affinity to Strep-Tactin which was leveraged to pull down proteins that bind the RNA baits. The interaction was broken by elution using Glycin buffer (pH2.5) and the eluted captured proteins were further precipitated, trypsinized and purified for mass spectrometry. Drawings and protocol adapted from (Agca et al., 2015; Marín-Béjar and Huarte, 2015; Wang and Wang, 2014).

RNA Interference

To knock down the human hnRNPF/H protein family, the following siRNA pools were purchased; si-HNRPF: ON-TARGETplus SMARTpool human HNRNPF (L-013449-01-0005), si-HNRPH1: ON-TARGETplus SMARTpool human HNRNPH1 (L-012107-00-0005), si-HNRPH2: ON-TARGETplus SMARTpool human HNRNPH2 (L-013245-02-0005) and ON-TARGETplus Non-targeting siRNA #1 (D-001810-01-05) from Dharmacon. A maximum of 100 nM of total siRNA was used per transfection. HEK293T cells at 40-60% confluency on 24-well plates were transfected with Lipofectamine RNAiMAX Reagent (Invitrogen, Life technologies) following manufacturer's instructions. Cells were trypsinized 24 hours post-transfection and reverse-transfected with 20 nM of siRNA and Lipofectamine RNAiMAX Reagent in 24-well plates. The following day, 500 ng of plasmid minigenes (Buena-Atienza et al., 2016) and 20 nM of siRNA were transfected with Lipofectamine 2000 (Invitrogen, Life technologies) and incubated for further 24 hours. Cells were harvested 72h after the first siRNA transfection and divided 1:1 for total RNA extraction with peqGOLD Total RNA Kit (PEQLAB Biotechnologie GmbH) and protein isolation with lysis buffer with Nonident P40 (NP40) (Roche) and protein inhibitors PI2, PI3 and PIC (Sigma-Aldrich). Silencing efficiency was validated by Western blotting. The effect of hnRNPF knockdown on exon 3 inclusion was determined by semi-quantitative fragment analysis of RT-PCR products.

Western blotting

Upon Bradford assay, 35 µg of protein samples derived from the siRNA assay were ran on pre-cast 10% polyacrylamide gels (Invitrogen, Thermo Scientific) and transferred on to polyvinylidene difluoride membranes (Millipore) using standard protocols. For eluted RNA-bound samples, the whole volume was loaded. Detection was achieved using the horseradish peroxidase chemiluminescent substrates Pierce™ enhanced chemiluminescence Western Blotting Substrate and Pierce™ enhanced chemiluminescence Plus Western Blotting Substrate (Thermo Scientific). The hnRNP F/H Antibody (1G11) sc-32310 (Santa Cruz Biotechnology) and GAPDH (Cell signalling) antibodies were used at 1:500 and 1:2000, respectively, and the secondary antibody at 1:7500 (Dianova).

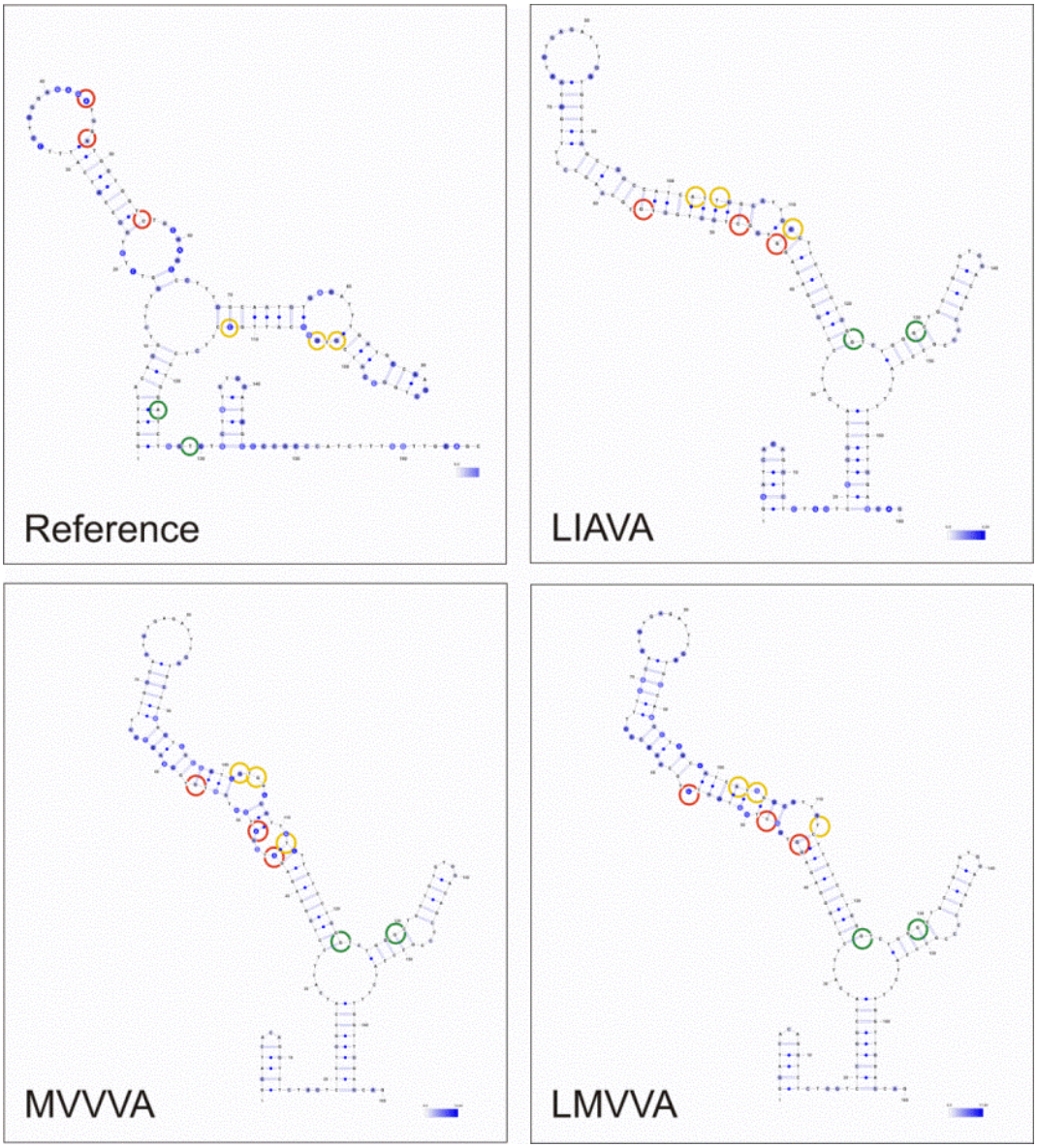
Semi-quantitative RT-PCR

Fluorescent RT-PCR was performed to quantify differentially spliced products as previously described (Buena-Atienza et al., 2016). Capillary electrophoresis and semi-quantitative fragment analysis of FAM-labeled RT-PCR products derived from minigene splicing assays and with Non-Targeting or hnRNPF/H siRNA treatment. GeneMapper (version 5) displays fluorescence (FAM-mediated) peaks that correspond to RT-PCR products derived from either the full-sized transcript with exon 3 retained or aberrantly spliced transcripts resulting in RT-PCR fragments of different length.

Bioinformatic predictions

In silico tools such as HSF (Desmet et al., 2009) and SpliceAid 2 (Piva et al., 2012) provide the predicted original and newly variant-induced motif(s) as well as splice factor(s) that putatively bind to a given motif. The MFE secondary structure of the RNA

baits used in kcal/mol was predicted by RNAfold from the Vienna RNA package (Gruber et al., 2008). The tool RNA2DMut (Moss, 2017) was used to predict RNA secondary structures of the RNA sequence comprising exon 3 either encoding for the Reference or different haplotypes (Fig.13). Sequences were submitted without constraints at the default temperature of 37°C.



Supplementary Figure S3. RNA secondary structures of pre-mRNA molecules bearing exon 3 haplotypes Predicted secondary structures by RNA2DMut (Moss, 2017) for the sequence (showed as DNA nucleotides) comprising the complete exon 3 of *OPN1LW/MW* for the Reference, 'LIAVA' (top), 'MVVVA' and , 'LMVVA' (bottom) haplotypes. The latter 'LMVVA' entails the c.453G; c.457C; c.465C; c.511A; c.513G; c.521T; c.532G; c.538G haplotype (siRNA knockdown experiment; Section 6.7; Table 3).

Supp. Table S2: Predicted MFE change in the Gibb's folding energy (ΔG) and ensemble diversity (ED) for the Reference exon 3 and for 22 mutants (bold and squared in grey) with highest (dark grey) and lowest scores (light grey) of ED (right side column).

Mut	c.	Sequence (w/o the last 24 nt)	ΔG	ED
0		*GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCGC*	*-58*	*50*
9		GG**U**UCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-55.6	82
398		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUG**G**UGUGUGGACAGCCCCG	-56.4	80
283		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCU**A**GCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-54.8	77
24		GGAUCAC**U**GGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-55.9	76
302		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAU**G**GUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-56.4	76
27		GGAUCACA**U**GUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-55.9	75
228		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUG**G**GAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-57.6	75
303		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAU**U**GUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-57.6	75
367	**c.532**	GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGG**G**UCUGGUCUGCUGUGUGGACAGCCCCG	-59.2	**75**
259		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAU**A**CCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCGC	-54.6	74
387	**c.538**	GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGG**G**CUGCUGUGUGGACAGCCCCG	-61.4	**74**
66		GGAUCACAGGUCUCUGGUCU**G**GGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-61.3	29
124		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAG**G**GAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCGC	-61.9	29
64		GGAUCACAGGUCUCUGGUCUC**A**GGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-62	29
327		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCA**G**UGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-59.7	28
187		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAG**A**CCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-62.4	28
130	**c.453**	GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAG**G**UGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-62	**27**
100		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUU**A**CUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-60.4	27
181		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCA**G**GCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-62.9	26
334		GGAUCACAGGUCUCUGGUCUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUG**A**CUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-58.5	26
54		GGAUCACAGGUCUCUGG**G**CUCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-64.5	26
57		GGAUCACAGGUCUCUGGU**U**UCUGGCCAUCAUUUCCUGGGAGAGAUGGAUGGUGGUCUGCAAGCCCUUUGGCAAUGUGAGAUUUGAUGCCAAGCUGGCCAUCGUGGGCAUUGCCUUCUCCUGGAUCUGGUCUGCUGUGUGGACAGCCCCG	-60.8	26

14 Acknowledgements

Firstly, I would like to thank my advisor **Prof. Bernd Wissinger** for the continuous support, for the opportunity to do research at his group, for the large knowledge and interest on my work, for the resources, the time and patience for discussions, for the motivation and critical thinking and for letting me be autonomous as well. I appreciate his guidance which was crucial for the research and writing of this thesis. I would like to thank my co-supervisor **Prof. Thomas Lahaye** as well for his time and his insightful comments to widen my research from various perspectives.

I greatly thank the support from **Dr. Renata Sarno** and **BCM Families Foundation**. Meeting the families and explaining them our research was one of the most rewarding moments during my PhD. This work received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration [grant agreement no 317472]. I greatly thank the PIs that conceived EyeTN, **Patricia** and the fellows: **Pablo**, **Julio**, **Laura**, **Riccardo**, **Carla**, **Evi**, **Simona**, **Raji**, **Prasun**, **Vicent** and **Giulia**. I will never forget the nights working for assignments and all the fun during trainings. I would like to thank Asper Biotech in Tartu, especially **Priit** and **Kristo** for treating me so well.

Big thanks to all the labmates, technicians and other members I had the pleasure to share time with back at *home* in Tübingen: **Eva**, **Britta**, **Jenny**, **Elke**, **Suki**, **Tobias**, **Joachim**, **Peggy**, **Susanne**, **Nicole**, **Julia**, **Anja**, **Irene**, **Simone**, **Barbara**, **Christian** and **Pablo**. Thanks **Tina** for your time as well. En especial a **Irene** por tu ayuda imprescindible para sobrevivir en los comienzos. **Susanne** thanks for always being supportive. Great thanks to **Nicole** for supervising; for shearing your expertise and rational thinking. **Peggy**, you were always much of a help for troubleshooting – Thanks. **Anja**, I could not have had a more comfortable company. **Julia** and **Anja**, it was great to share time for science with you but also for Zumba coreos! Thanks **Christian** for pre-warming us up with Delicatesse.

Thanks **Sven**, **Karin**, **Joshua**, **Antje**, **Marc**, **Jakob** and **Nicolas** for the immense flexibility and **Sabine** for listening to me. **Marius** thanks for letting me work with you on the Input Library code. **Stefan** thanks for your support and continuing the project.

Grazie **Enrico**, Teşekkür ederim **Ebru**, Ευχαριστώ **Melania** for your help, recipes and friendship!

Dr. Alfonso Lombana, dem besten spanischen Deutschlehrer, danke für die moralische Untersütztung, die Tipps, deine Weisheit und dein Optimismus! Vielen Lieben Dank **Lara**, mit dir konnte ich am besten üben. Danke für dein Verständnis bei den vielen *Neins*. Vielen herzlichen Dank **Isabel**, **Heidrun** und **Jörg** für Eure Geduld und Verpflegung. Danke **Bernhardt** für deine Großherzigkeit.

A mis maravillas: **María**, **Elisa**, **Silvia**, **Irene**, **Ángela** y **Clara** por nuestros reencuentros y los que vendrán, con nuestro pare **Fede** también**. Bea**, muchas gracias por compartir tus experiencias y por tus consejos. Por supuesto a las otras siete patas del octógono palentino **María**, **Eva**, **Marta**, **Marina**, **Andrea**, **Vero** y **Paola** gracias por venir, por estar siempre unidas y dispuestas para mis llegadas. Gracias **Carla** por sumar tanto.

Mil gracias a mis padres **María del Carmen** y **Fernando** y a mi hermano **Rodrigo** por respaldarme, por escucharme siempre y por aguantar la distancia. Gracias por inculcarme el valor del esfuerzo y el trabajo, a no rendirme jamás y a disfrutar del camino por arduo que sea.

Ich bedanke mich auch bei demjenige der jeden Tag die guten und nicht so guten Momente miterlebt hat. Liebsten Dank an **Sebastian**; schön, dass du mich begleitest und dass ich so viel von dir lerne.