Supplementary Materials for:

Specialization of the photoreceptor transcriptome by Srrm3-dependent microexons is required for outer segment maintenance and vision

Ludovica Ciampi 1, Federica Mantica 1,†, Laura Lopez-Blanch 1,†, Jon Permanyer 1,†, Cristina Rodríguez-Marín 1, Jingjing Zang 4, Damiano Cianferoni 1, Senda Jiménez-Delgado 1, Sophie Bonnal 1, Samuel Miravet-Verde 1, Verena Ruprecht 1,2, Stephan C.F. Neuhauss 4, Sandro Banfi 5,6, Sabrina Carrella 6, Luis Serrano 1,2,3,*, Sarah A. Head 1,* and Manuel Irimia 1,2,3,*

Luis Serrano
Centre for Genomic Regulation
Dr. Aiguader, 88, 08003 Barcelona, Spain
e-mail: luis.serrano@crg.eu
Phone: +34933160101 Fax: +34933160099

Sarah A. Head
Centre for Genomic Regulation
Dr. Aiguader, 88, 08003 Barcelona, Spain
e-mail: dibartolosa@gmail.com
Phone: +34933160101 Fax: +34933160099

Manuel Irimia
Centre for Genomic Regulation
Dr. Aiguader, 88, 08003 Barcelona, Spain
e-mail: mirimia@gmail.com
Phone: +34933160212 Fax: +34933160099
METHODS

Definition of tissue-enriched microexons

We downloaded information for all alternative splicing events of all types from VastDB¹ (https://vastdb.crg.eu/wiki/Downloads) including their inclusion levels (using the Percent Spliced In [PSI] metric as provided by vast-tools) across all VastDB samples and their genomic features (e.g. length, sequence, genomic coordinates, predicted protein impact) for all the species included in the manuscript (human: hg38, mouse: mm10, chicken: galGal4, zebrafish: danRer10). We considered as microexons all cassette exons with length ≤ 27 nucleotides. We then used the Get_Tissue_Specific_AS.pl² script (https://github.com/vastdb/pastdb/pastdb) to derive sets of tissue-enriched microexons from PSI tables. We defined as tissue-enriched in a tissue T all microexons presenting: (i) ΔPSI (delta PSI) between tissue T average (i.e. the PSI average among all samples from tissue T) and each of the other tissues’ averages ≥ 15, (ii) ΔPSI between tissue T average and the average of all the other tissues ≥ 25, (iii) coverage in at least N tissues (with N=10 for human and mouse; N=8 for chicken and zebrafish), and (iv) at least n samples per tissue (with n=2 for human and mouse; n=1 for chicken and zebrafish). In order to capture microexons with biased inclusion patterns in one or more tissues, for some tested tissues (e.g. Neural) we excluded tissue groups known to be characterized by partially overlapping microexon programs (e.g. Retina, Muscle, Heart) from the pool of compared tissues. We built a config file for each species, highlighting the correspondence between sample, relative tissue group and tissue groups excluded from the comparison, which we provided as input to Get_Tissue_Specific_AS.pl script. The config files used in the manuscript are listed in Dataset 15.

Definition of RetMICs, RetLONGs and RetGENEs

In order to identify microexons with retina-specific inclusion biases (i.e. enriched in no other tissue), we ran Get_Tissue_Specific_AS.pl as described in the previous section but without excluding any tissue groups from the pool of compared tissues (config files in Dataset 15) and printing out a table containing all the ΔPSI Retina-Others or Neural-Others (--test_tis Retina Neural option) for all the events with minimum coverage and number of replicates (see (iii) and (iv) in the previous section). We then devised a Retina Specificity Score (RSS) to help the identification of retina-specific exons (RetMICs and RetLONGs), and is defined as follows:

$$RSS = 2 \cdot (PSI_{retina} - (PSI_{neural} + PSI_{other}^{max})) - (2 \cdot SD_{retina} + SD_{neural} + SD_{other})$$
where:

PSI_[Retina|Neural]: exon average PSI among all [retina|neuronal] samples.

SD_[Retina|Neural|Other_tis]: standard deviation of exon PSI distributions across all [retina|neuronal|non-retina] samples.

Max_Other_tis: maximum average PSI among all non-retina tissues (in most cases, Max_Other_tis = PSI_Neural).

We identified RetMICs and RetLONGs for all species (human, mouse, chicken and zebrafish) requiring either (i) \( \Delta \)PSI between retina and the tissue with the highest PSI \( \geq 40 \), or (ii) average \( \Delta \)PSI between Retina and all other tissues \( \geq 10 \) and the RSS \( > 0 \). We then further divided RetMICs between microexons either exclusively included in retina (“retina-exclusive”; Max_Other_tis \( \leq 4 \ || \ (4 < \text{Max}_\text{Other}_\text{tis} < 15 \ && \text{PSI}_\text{Retina} \geq 90) \)) or presenting a substantially higher degree of retinal versus neural inclusion (“retina-differential”; all the other RetMICs).

Using a similar approach, we used the Get_Tissue_Specific_GE.pl script (https://github.com/vastdb-pastdb/pastdb) to derive the set of 505 retina-specific genes (specifically upregulated in retina compared to the other tissues). We defined as retina specific all genes presenting (1) minimum expression across all VastDB zebrafish samples (see above) \( \geq 5 \) cRPKMs \(--\text{min}_\text{expr} 5\), (2) minimum fold change of the median retina expression against the median expression in each of the other tissues \( \geq 3 \) \(--\text{min}_\text{fold} 3\), (3) minimum fold change of the median retina expression against the median of the median expression in all other tissues \( \geq 5 \) \(--\text{min}_\text{fold}_\text{glob} 5\), and (4) minimum absolute expression difference between the expression in retina and in each of the other tissues \( \geq 2 \) cRPKMs \(--\text{min}_\text{range} 2\). We only require one replicate with coverage for each tissue group to be considered \(--\text{min}_\text{rep} 1\) and we used the same config file \(--\text{groups}\) employed for the definition of RetMICs and RetLONGs (Dataset 15).

**Exon orthology calls**

We ran Broccoli (v1.0)\(^3\) with default parameters to infer gene orthogroups between human (hg38), mouse (mm10), chicken (galGal6) and zebrafish (danRer10), and we selected all gene orthogroups containing 20 genes or less (16,166 out of 16,313). We then ran ExOrthist main (v0.1.0)\(^4\) to derive genome-wide, multi-species exon orthogroups among all species. We
considered the following species pairwise evolutionary distance ranges [short: human-mouse (hg38-mm10), human-chicken (hg38-galGal6), mouse-chicken (mm10-galGal6); medium: human-zebrafish (hg38-danRer10), mouse-zebrafish (mm10-danRer10), chicken-zebrafish (galGal6-danRer10)] and set the evolutionary conservation cut-offs to the default values. In order to increase the sensitivity of the exon orthology calls, we also considered non-annotated exons identified by vast-tools\(^1\) (--extraexons option) and we provided pre-computed liftOver hits among each pair of species (computed with the ExOrthist companion script get_liftovers.pl) to be directly integrated in the exon orthogroups (--bonafide_pairs option). We identified a total of 206,644 exon orthogroups, recovering from a minimum of 77.89% of exons (chicken) to a maximum of 88.34% (mouse). Human RetMICs/RetLONGs exons were considered as genomically conserved in one of the other species when belonging to an exon orthogroup including at least one exon from that species (see Fig. 2B).

**GO enrichment analysis and hypergeometric tests**

All GO enrichment analyses were carried out with the gprofiler2 R package\(^5\), to which we provided custom annotations and backgrounds. We downloaded the human annotation (geneID-GO correspondence) from Ensembl (v102), and we combined it with the clueGO v2.5.5\(^6\) human annotation for biological processes (level 5) in order to include possibly unique annotations from the two sets. For the evolutionary comparison and to avoid biases due to different annotation qualities between species, we derived annotations for mouse, chicken and zebrafish starting from the assumption that orthologous genes likely share functional properties. We considered the gene orthogroups inferred between human, mouse, chicken and zebrafish (see “Exon orthology calls”) and assigned the functional annotation of each human gene to all the genes from the other species belonging to the same orthogroup. Then, for each species, we filtered for the GO categories with a number of genes included between 5 and 3,500. Background sets for each species consisted of genes containing events of any kind with the expression level requirements used to define tissue-enriched exons (see above). In particular, they were derived by running the Get_Tissue_Specific_AS.pl script with the same parameters used for the tissue-enriched microexon call and requiring sufficient read coverage in Retina, but for all event kinds (--event_type option set to EX, IR, Alt3, Alt5, in four different runs). The total number of background genes was 14,471 in human, 11,586 in mouse, 11,994 in chicken and 14,588 in zebrafish. False Discovery Rate (FDR) correction was applied to all performed tests. All the enriched GO categories are listed in Dataset 2.
For visualization purposes, GO enrichment analysis for genes downregulated (considering only the genes with log2FC(MUT/WT)< -1.5) upon *srm3* depletion in zebrafish eyes was performed directly using the software ClueGO\(^6\), selecting Biological Process as ontology. For this purpose, we converted the list of downregulated genes into human orthologs using DIOPT (DRSC Integrative Ortholog Prediction Tool)\(^7\) with a score ≥ 2 and “best score” required, as previously described\(^8\). All the enriched GO categories are listed in Dataset 9.

All hypergeometric (Fisher) tests mentioned in the manuscript were performed through the *phyper* function in baseR. The total number of genes used to compute the hypergeometric input values were derived from the background employed for the GO enrichment analysis (see previous section). Gene-disease information on inherited retinal disease comes from the database RetNet (https://sph.uth.edu/retnet/). The input values and resulting p-values for all the tests are provided in Dataset 3.

**Structural analysis of RetMICs**

A list of 75 human genes and their relative microexon events was used to retrieve the corresponding SwissProt (manually annotated) protein identifier (ID). The gene AC007246.3 was found to correspond to an antisense RNA and MAGI2-AS3 to a non-coding RNA. Protein Data Bank (PDB) structures were downloaded for those SwissProt IDs having any. For all 75 genes (excluding 7 causing ORF disruption), interaction models were searched within Interactome3D\(^9\) and ModBase\(^10\), respectively collecting protein-protein interactions (both experimental and modeled) and protein homology models. Redundant experimental interactions, already found in PDB structures, were excluded. Structures collected in ModBase came with a set of submodels with different quality levels and protein coverage. For each gene, the selected submodel was the one presenting the highest set of scores indicated by the authors\(^11\), with the condition of containing the microexon event and being based on a template structure with at least 40% of sequence similarity. Flanking exons were translated to their amino acid sequences and searched within their corresponding SwissProt entries. Four exon residues, flanking the microexon event, were used to match and determine the event position within collected structures. The structural characterization is available in Dataset 5. A graphical overview in Fig. S2 shows the events locations captured in 3D structures.

**Regulatory analysis of RetMICs and RetLONGs**
HEK293 lentiviral cell lines used in this study to ectopically express SRRM3, SRRM4 or MSI1 were generated using a previously described protocol\textsuperscript{12}. The gateway destination vector pCW57.1 containing SRRM4 was generated as described previously\textsuperscript{12}. SRRM3 and MSI1 cloned into the donor vector pDONR223 were obtained from the Biomolecular Screening & Protein Technologies Facility at CRG. The resulting entry clones were reacted with pCW57.1 using LR clonase II (11791020, Thermo Fisher Scientific). pCW57.1 lentiviral vector (empty vector (EV)) was used as a negative control. Prior to harvesting RNA, expression of the transgenes was induced using \(1 \mu\text{g/mL}\) of Doxycycline for 24 h. RNA was extracted with the RNeasy Mini kit (QIAGEN). Standard polyA-selected Illumina libraries were generated at the CRG Genomics unit and were sequenced in a Illumina HiSeq2500, producing an average of \(\sim80\) million paired-end 125 nt reads. Mapping statistics are provided in Dataset 17. Detection of microexons and longer exons upon induction was done by PCR using Q5® High-Fidelity DNA Polymerase kit (New England biolabs), according to the manufacturer's instructions. Primers were obtained directly from VastDB, which provides automatically designed primers annealing to the flanking upstream and downstream constitutive exons. PCR products were run on 3.5% agarose gel in TBE buffer to allow small isoform separation. All primer sequences are shown in Dataset 16.

To assess the enrichment of sequence motifs associated with SRRM3/4 (UGC) and MSI1 (AUG) regulation in the upstream and downstream RetMIC and RetLONG intronic sequences, we generated RNA maps running the \textit{rna\_maps} function from \textit{Matt}\textsuperscript{13} using sliding windows of 27 nucleotides and 75 nt upstream/downstream as region surrounding the exons. To evaluate the enrichment of NOVA (YCA\textit{Y}), RBFOX (GC\textit{T\textit{G}}) and ELAVL (\textit{T\textit{T\textit{T\textit{N\textit{T\textit{T\textit{T}})}}}) binding for RetMICs and neural microexons (Fig. S8) we employed a similar approach but considering 150 nt as the region surrounding the exon. For estimating FDR-corrected \(p\)-values in the RNA maps we used a permutation test using 1,000 permutations and a threshold of FDR < 0.05 as implemented in \textit{Matt}.

\textbf{Generation of \textit{srrm3} and \textit{srrm4} zebrafish mutant lines}

Fish procedures were approved by the Institutional Animal Care and Use Ethic Committee (PRBB–IACUEC). Zebrafish \textit{(Danio rerio)} were grown in the PRBB facility at 28°C 14 h light/10 h dark cycle. Individual and massive crosses were carried out in crossing tanks and the resulting eggs were screened for overall health from embryo collection to 5 dpf. Embryos were collected into Petri dishes with E3 medium with methylene blue and placed into the incubator.
Sibling WT and mutant larvae between 5 and 10 dpf were used for most of the procedures and WT adult fish of AB background for retina dissection. To create zebrafish mutant lines we used a CRISPR-Cas9 strategy adapting a previously described protocol\textsuperscript{14}. Single or double guide RNAs (gRNAs) were designed using CRISPRscan\textsuperscript{15} to target the eMIC domain of \textit{srrm3} (Tg(HuC:GFP; srrm3\_eMIC)) and \textit{srrm4} (Tg(HuC:GFP; srrm4\_eMIC)) and were injected at 1-2 cell embryos stage. Selected gRNA sequences are: GGGAAATAACTGCGTGAGCGGCGG (\textit{srrm3}) and GCTGTGCTTTCTGCTTTGCAGG and TGATTCTGCGGGGCTTCCAGGTGG (\textit{srrm4}). CRISPR-Cas9 injection mix (with a final volume of 5ul) consisted of Cas9 protein (300ng/μl) – PNABio Cas9 protein (CP01-50), the in-vitro transcribed gRNAs (50ng/μl). Additionally, we included 0.5 ul Phenol Red and sterile water to reach 5ul final volume. We injected 1 nl of CRISPR-Cas9 mix in each embryo. For each gene, we generated two independent founders. The screen for germline mutations of \textit{srrm4} at F0 revealed a founder containing a deletion of 1bp upstream the eMIC domain that produces a frameshift mutation causing a premature STOP codon breaking the eMIC domain (Srrm4\_F1). The screen for germline mutations for \textit{srrm3} revealed a founder for \textit{srrm3} eMIC MUT line containing a 5nt deletion that altered the reading frame upstream the eMIC, producing a non-functional domain and altering the C-terminal region of the protein (Srrm3\_F1). Additionally, we generated a second founder for \textit{srrm3} carrying a 19 bp deletion (Srrm3\_F2) and a second founder for \textit{srrm4} carrying 11 bp deletion, both confirming the gross phenotype (Srrm4\_F2). All the experiments performed in this manuscript have been done using the Srrm3\_F1 and Srrm4\_F1 founders but the second founders show similar phenotypes.

Founders were crossed with WT strain and F1 animals were genotyped. Heterozygous animals were incrossed to generate the single mutant lines. To obtain the DMUT line, \textit{srrm3} and \textit{srrm4} heterozygous mutants were crossed and \textit{srrm3}+/-,\textit{srrm4}-/- fish were maintained as viable and fertile. All experiments using DMUTs in this study were done through in-crosses of \textit{srrm3}+/-,\textit{srrm4}-/- fish, since \textit{srrm3} and \textit{srrm4} are linked in zebrafish making it impractical to use HET/HET incrosses (1/100 DMUT embryos). Therefore, \textit{srrm4} MUT fish were used as sibling controls for DMUT embryos. For fish genotyping, a piece of the caudal fin was cut and disaggregated using 100 uL of NaOH 50mM for 15 minutes at 96 degrees. To neutralize the reaction, 10uL of Tris-HCl pH=7.4 was added. 2 ul were used as templates for PCR with the \textit{srrm3} or \textit{srrm4} primers designed to amplify the genomic region of interest (listed in Dataset 16) and GoTaq® Flexi DNA Polymerase kit (Promega).

\textbf{In-vitro validation of srrm3 and srrm4 mutant isoforms}
We used previously generated pcDNA5 vectors containing full-length WT *srrm3* and *srrm4* sequences\(^{16}\). Inverse PCRs using these WT sequences as templates was used to reproduce the mutations generated in the different founders used in the study. Primer sequences are shown in Dataset 16. These constructs were transfected in HEK293 cells using Lipofectamine 2000 Transfection Reagent (Invitrogen #11668019). Cells transfected with pcDNA5 vector expressing GFP were used as a negative control. After 24 h, cells were harvested, RNA extracted using RNeasy minikit (Qiagen, 74104) and 1 ug of RNA was reverse-transcribed into cDNA using SuperScriptIII Reverse Transcriptase (Invitrogen #18080044). Detection of several representative microexons was done by RT-PCR as described above.

**Isolation of zebrafish larval tissues and adult retinae**

For tissue dissection procedures, animals were euthanized in 0.08% tricaine. For retina isolation, whole eyes were removed from adult WT animals. Retina isolation was performed as previously described\(^{17}\) and five retinae were pulled for RNA extraction. For larval tissue isolation, heads were dissected from WT embryos at 24 hpf while whole eyes were dissected from WT and *srrm3* MUT fish at all the other time-points described in the manuscript. To ensure sufficient RNA yield, from 20 to 60 eyes per genotype were pulled prior to RNA extraction. For that purpose, we used fresh tissue. Tissues were homogenized in RLT Buffer using a BeatBeater. RNA was then extracted using the RNeasy Micro kit (QIAGEN), according to manufacturer’s instructions. For tissue extraction, beta-mercaptoethanol was added to the RLT buffer, following manufacturer’s instructions. RNA for each sample was then used to generate standard polyA-selected Illumina libraries at the CRG Genomics unit that were sequenced in an Illumina HiSeq2500, producing an average of ~125 million paired-end 125 nt reads. Mapping statistics are provided in Dataset 17.

**Quantitative PCR (qPCR)**

qPCR reactions were carried out with the Roche Light Cycler 480 system, using Power SYBR-Green PCR Master Mix (4367659, Thermo Fisher Scientific) according to the manufacturer’s instructions. Expression quantification data from the different samples are expressed as cycle threshold (Ct). The *eef1a1* gene was used as endogenous reference control. Ct values were averaged for each in-plate technical duplicate and normalized as difference in Ct values (ΔCt) between the analyzed gene and the reference gene. Relative expression (\(2^{-\Delta Ct}\)) was calculated as fold change between MUT and WT (FC(MUT/WT)). Unpaired t-test was applied to measure significance between WT and MUT at each time-point. Detection of several
representative microexons was done by RT-PCR as described above. All primer sequences are shown in Dataset 16.

**Immunofluorescence and histological retinal analyses**

Zebrafish eyes were fixed in 4% paraformaldehyde (PFA) over-night, cryoprotected with 30% sucrose overnight, embedded in OCT and cryosectioned. Twenty-micrometer cryosections were collected on slides. For the PR marker staining, sections were blocked in 1% bovine serum albumin, 0.5% Triton X-100 in PBS for 1 h RT and incubated with primary antibodies (ZPR-1 (ZDB-ATB-081002-43) 1:400 and ZPR-3 (ZDB-ATB-081002-45) 1:200, Zebrafish International Resource Center (ZIRC), GS6 (Millipore) overnight in blocking solution. Sections were then incubated with the Alexa Fluor secondary antibodies (1:1000; Invitrogen) and counterstained with DAPI (Vector Laboratories). The Phalloidin staining was performed using a previously described protocol18. WT and MUT ZPR-3 fluorescence intensity was assessed by measuring the Integrated Density of the region of interest over the area using the *ImageJ* tool. Layer thickness along the apico-basal axis in WT and MUT eyes was manually measured using the *ImageJ* tool in superior, central and inferior retina, corresponding to the three values per fish listed in Dataset 11 at 3 and 5 dpf (1 retina section for each eye). Unpaired t-test was applied to measure significance between genotypes.

The Caspase3 assay was performed on retina sections following the protocol described above (anti-caspase3 (Fisher Scientific, 15889738) 1:500). Apoptotic cells were manually counted on the entire ONL in 1 section for each eye by the ImageJ tool. Unpaired t-test was applied to measure significance between groups. For all the experiments, the number of animals/genotypes is indicated in each figure legend.

**Electron Microscopy**

Larvae at 5 dpf have been euthanized in 0.08% tricaine, genotyped and fixed using a mixture of 2.5 % glutaraldehyde and 2 % paraformaldehyde prepared in 0.2 M HEPES buffer for 24 h at 4 °C. After fixation, the samples were washed 3 times with PBS 1X. The samples were post-fixed in 1% osmium tetroxide, dehydrated, and embedded in epoxy resin. The samples were cut on ultra-microtome Leica EM UC7 and collected on the single-slot oval grids and analyzed with FEI electron microscope. OS, mitochondrial and interphotoreceptor space area was determined using FEI software. The analysis was carried out on 4 eyes/genotype, for each genotype an N ≥ 13 fields was analyzed. One-way ANOVA test was applied for mitochondria Electron Microscopy analysis. As the statistical test is significant (p-value < 0.05), a post-hoc
analysis based on Tukey test was applied to determine pairwise comparisons between condition levels with correction for multiple testing (significance codes: $p < 0.0001$ ‘****’, $< 0.001$ ‘***’, $< 0.01$ ‘**’, $< 0.05$ ‘*’). To test the data are normally distributed we used Shapiro-Wilk test on the ANOVA residuals (i.e., if the $p$-value > 0.05, then the normality is not violated), while we applied Levene’s test to verify the variance across groups is homogeneous (i.e., if the $p$-value > 0.05, then we assume the homogeneity of variances in the different groups).

**Survivability tests**

For the survivability tests, 100 embryos per time point coming from a collective $srrm3$ HET cross were grown in two tanks. At each time point all the fish were individually euthanized in 0.08% tricaine and placed into 100 uL NaOH 50mM for further DNA extraction as previously described. Then genotyping was performed in order to check the ratios of survivability of each genotype. For the darkness survivability test, 100 embryos coming from a collective $srrm3$ HET cross were grown in 2 tanks covered with aluminum paper to mimic dark conditions (50 embryos per tank). At 13 dpf, the surviving fish were genotyped. The control experiment was performed growing embryos from the same $srrm3$ HET cross in standard conditions.

**Electroretinogram (ERG)**

For ERG experiments, $srrm3$ MUT, $srrm4$ MUT and DMUT larvae as well as different control siblings were recorded at 5 dpf as previously described. Briefly, larvae were dark adapted for at least 30 mins before recording. Then each of them was transferred to a piece of filter paper in a plastic recording chamber, which was filled with 1% agarose. The reference electrode was inserted into the agarose through a hole in the chamber. Afterwards, the eye ball of individual larvae was removed by a tungsten wire. The recording electrode, filled with E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl2, and 0.33 mM MgSO4) was placed against the center of the cornea. All these preparation steps were carried on under dim red light. The eye was then stimulated by 3 stimulation flashes. The duration of each flash was 100 ms and the interval between flashes was 12 seconds. The intensities were 0.6 lux, 6 lux and 60 lux (light source, HPX-2000, Ocean Optics). Electronic signals were amplified 1000 times by a pre-amplifier (P55 A.C. Preamplifier, Astro-Med. Inc, Grass Technology) with a band pass between 0.1 and 100 Hz and recorded via the self-developed National Instrument Labview program. The recordings were performed in 2 independent experiments for each genotype. Significance between genotypes was calculated using the One-Way ANOVA test.
Optokinetic responses (OKRs)

Spontaneous and compensatory eye movements, known as optokinetic responses, are induced by detection of motion cues in the visual environment. OKRs were recorded as described before\textsuperscript{20}. Briefly, 10 dpf fish were kept still in 3% prewarmed (28°C) methylcellulose in a small Petri dish. Contrast sensitivity was determined by applying different contrasts (5, 10, 20, 40, 70, and 100% of maximum), with a spatial frequency of 20 cycles/360°, and an angular velocity of 7.5°/s. SPSS (version26.0; Chicago, IL, USA) was used to analyze the data and generate the figures. Significance between genotypes was calculated using Kruskal-Wallis test.

Injection of eMIC domain mRNA in the \textit{srrm3} mutants

We used a previously generated vector\textsuperscript{16} containing the last two exons of the human SRRM3 gene (encoding the eMIC domain) cloned into a pcDNA3.1 plasmid (ThermoFisher #V79520). The vector was linearized by PsiI digestion and \textit{in vitro} transcribed using mMESSAGE mMACHINE™ T7 Transcription kit (ThermoFisher #AM1344). Fish embryos coming from a \textit{srrm3} HET cross were injected at 1 cell stage using 1 nl/cell of the mix containing RNA at different concentrations (0, 5 or 10 ng/ul) in RNase free water. Embryos injected with diluent only were used as a negative control. Upon injection, fish were grown until 5 dpf, sacrificed and genotyped as explained above. WT and \textit{srrm3} MUT fish were processed for histology as well as RNA extraction and validation of microexon inclusion, following the protocols described above.

Gene set enrichment analysis (GSEA)

All GSEA enrichments were performed through the R package fgsea (v1.10.1), using as input the log2 fold changes between the zebrafish \textit{srrm3} MUT and the relative WT at 48 hpf, 60 hpf, 72 hpf and 5 dpf. For each timepoint, we tested all GO categories defined for the zebrafish GO enrichment analysis (see above). Adjusted p-values were derived through FDR correction. All the enriched categories are listed in Dataset 14.
BIBLIOGRAPHY

1. Tapial, J. et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* **27**, 1759–1768 (2017).

2. Martín, G., Márquez, Y., Mantica, F., Duque, P. & Irimia, M. Alternative splicing landscapes in Arabidopsis thaliana across tissues and stress conditions highlight major functional differences with animals. *Genome Biol.* **22**, 35 (2021).

3. Derelle, R., Philippe, H. & Colbourne, J. K. Broccoli: Combining Phylogenetic and Network Analyses for Orthology Assignment. *Mol. Biol. Evol.* **37**, 3389–3396 (2020).

4. Márquez, Y. et al. ExOrthist: a tool to infer exon orthologies at any evolutionary distance. doi:10.1101/2021.02.22.432358.

5. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research* vol. 9 709 (2020).

6. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).

7. Hu, Y. et al. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12**, 357 (2011).

8. Torres-Méndez, A. et al. Parallel evolution of a splicing program controlling neuronal excitability in flies and mammals. doi:10.1101/2021.02.24.432780.

9. Mosca, R., Pons, T., Céol, A., Valencia, A. & Aloy, P. Towards a detailed atlas of protein–protein interactions. *Current Opinion in Structural Biology* vol. 23 929–940 (2013).

10. Pieper, U. MODBASE, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Research* vol. 32 217D–222 (2004).
11. Pieper, U. et al. ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **42**, D336–46 (2014).

12. Head, S. A. et al. Silencing of SRRM4 suppresses microexon inclusion and promotes tumor growth across cancers. *PLoS Biol.* **19**, e3001138 (2021).

13. Gohr, A. & Irimia, M. Matt: Unix tools for alternative splicing analysis. *Bioinformatics* **35**, 130–132 (2019).

14. Gagnon, J. A. et al. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* **9**, e98186 (2014).

15. Moreno-Mateos, M. A. et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR/Cas9 targeting in vivo. *Nat. Methods* **12**, 982 (2015).

16. Torres-Méndez, A. et al. A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. *Nature Ecology & Evolution* vol. 3 691–701 (2019).

17. Wiggenhauser, L. M., Kohl, K., Dietrich, N., Hammes, H. P. & Kroll, J. Studying Diabetes Through the Eyes of a Fish: Microdissection, Visualization, and Analysis of the Adult tg(fli:EGFP) Zebrafish Retinal Vasculature. *J. Vis. Exp.* (2017) doi:10.3791/56674.

18. Letelier, J. et al. Mutation of Vsx genes in zebrafish highlights the robustness of the retinal specification network. doi:10.1101/2022.01.20.477122.

19. Zang, J., Keim, J., Kastenhuber, E., Gesemann, M. & Neuhaus, S. C. F. Recoverin depletion accelerates cone photoresponse recovery. *Open Biol.* **5**, (2015).

20. Rinner, O., Rick, J. M. & Neuhaus, S. C. F. Contrast Sensitivity, Spatial and Temporal Tuning of the Larval Zebrafish Optokinetic Response. *Investigative Ophthalmology & Visual Science* vol. 46 137 (2005).

21. Krieger, E. & Vriend, G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* **30**, 2981–2982 (2014).
Fig. S1 - (A) Expression of PR differentiation marker genes (cRPKMs, left Y-axis) and average RetMIC inclusion levels (PSI, right Y-axis) along organoid development (in days [d], as described in Fig. 1C). (B) Heatmaps depicting PR marker gene expression (top) and individual RetMIC inclusion levels (bottom) along organoid development. Each row corresponds to a different gene/RetMIC. RetMIC PSIs correspond to the average of three biological replicates for each time point. Values with insufficient read coverage are shown in white. (C-D) Heatmap of inclusion levels for RetMICs (C) and neural-enriched microexons (NeuralMICs) (D) in hippocampal neurons, WT and Aipl1 KO retinas, corresponding to the values displayed in Fig. 1D,E. Each row corresponds to a different microexon with sufficient read coverage across all three samples.
**Fig. S2** - Protein 3D structures of RetMIC-containing genes for the exclusion isoform. Grey, blue and red boxes indicate experimental PDB structures, Interactome3D structures and ModBase models, respectively. The three residues around the RetMIC insertion sites are depicted in red. For each structure, the gene name and the number of RetMIC-encoded amino acids (aa) are indicated.
Fig. S3 - (A) Inclusion of a 3-nt microexon in CLTC (VastID: HsaEX0015816) across different tissues from VastDB. Retina is shown in blue, neural tissues in green. (B) Conservation of the microexon and neighboring intronic sequences in the genome across 100 vertebrate species, with seven representative species shown below. The flanking exons are shown as rectangles in the GENCODE V36 track, and the microexon is highlighted by a red box. Image generated from UCSC genome browser ([http://genome.ucsc.edu](http://genome.ucsc.edu)). (C) Amino acid sequence alignment of CLTC with and without the microexon insertion, indicated by the red arrow. Negatively/Positively charged amino acids are depicted in red/blue. The four beta strands comprising beta propeller blade #2 are indicated above the sequence. Amino acid numbering shown below corresponds to that of the CLTC gene without the microexon. (D) Ribbon structure of the CLTC terminal domain (PDB: 5M5R) with seven beta propeller blades numbered from N to C term. The AP2B1 clathrin-box peptide, which binds in the groove between blades 1 and 2, is represented by a stick model. The microexon-flanking amino acids of CLTC (K83 and A84) are highlighted in red, with the side-chain of A84 shown making hydrophobic contacts with the N-terminal lysine residue of the peptide clathrin box motif. Image generated by YASARA v19\(^2\). (E) Van der Waals surface of the CLTC terminal domain (PDB: 5M5R) showing the hydrophobic binding pocket of the AP2B1 clathrin-box peptide, with negatively charged side chains colored in red and positively charged side chains in blue. The hydrophobic side chain of A84 interacting with the N-terminal lysine residue of the peptide clathrin box motif of A2BP1 is circled. Insertion of the microexon results in a negatively charged Aspartic Acid residue at position 84, thereby disrupting this hydrophobic contact. Image generated by YASARA v19\(^2\).
**Fig. S4** - (A) Inclusion of a 9-nt microexon in MYO6 (VastID: HsaEX0041235) across different tissues from VastDB. Retina is shown in blue, neural tissues in green. **(B)** Conservation of the microexon sequence in the genome across 100 vertebrate species, with seven representative species shown below. The flanking exons are shown as rectangles in the GENCODE V36 track, and the microexon is highlighted by a red box. Image generated from UCSC genome browser ([http://genome.ucsc.edu](http://genome.ucsc.edu)). **(C)** Amino acid sequence alignment of MYO6 with and without the microexon insertion, indicated by the red arrow. Negatively/Positively charged amino acids are depicted in red/blue. The location of insert-1, which is unique to MYO6 among other myosin genes, is indicated above the sequence in orange, while the remaining residues of the loop that are conserved among myosins are indicated in magenta. Amino acid numbering shown below corresponds to that of the MYO6 gene without the microexon. **(D)** Ribbon structure of the MYO6 catalytic motor domain (PDB: 2BKI), with insert 1 and the conserved loop colored as in Fig. S4D, and the microexon-flanking amino acids (E299 and Y300) highlighted in red. The close proximity of this loop to the ATP binding site is indicated by the black arrows. Image generated by YASARA v19.21.
**Fig. S5 - (A-C)** Box plots showing RetMIC (A) and RetLONG (B) inclusion levels (PSI) and RetGENE expression levels (log10(cRPKM)) (C) across different cell and tissue types. Neurons correspond to FACS-sorted GFP-positive cells from 72 hpf embryos of the HuC:GFP transgenic line (data from VastDB) and Rods to FACS-sorted GFP-positive cells from adult retina of the xops:eGFP transgenic line (data from SRP109268). Telencephalon and Cerebellum data were retrieved from VastDB, and Retina data was generated in this study. No read coverage filter was applied to exons plotted in (A) and (B).
**Fig. S6** - (A) Heatmap showing the retina-inclusion bias compared to neural inclusion of genomically conserved human RetMICs (top row) and their respective orthologs (i.e. exons belonging to the same exon orthogroup) in mouse, chicken and zebrafish. Each column corresponds to a different RetMIC. The color represents the delta PSI between the average of the retina samples and the average of neural samples considered for the definition of RetMICs. Human RetMICs and their selected orthologs are plotted in the same order as in Fig. 2C, to facilitate the comparison. Blanks and ivory rectangles indicate missing orthologs and missing ΔPSI values, respectively. (B) Dotplot representing functional enrichments for genes harboring RetLONGs across species. The functional enrichment of genes containing RetLONGs was separately tested for each of the species, and significant categories in at least two species (FDR-adjusted p-value ≤ 0.05) were plotted. The color reflects the adjusted p-value, with yellow color depicting p ≥ 0.05. The size of the dots is proportional to the log2 of the observed vs. expected ratio (O/E), and black borders around the categories highlight log2 O/E ≥ 1. GEF: Guanyl-nucleotide Exchange Factor.
Fig. S7 – (A-C) cRPKMs for *MSI1* (A), *SRRM4* (B) and *SRRM3* (C) in each cell line as measured from RNA-seq data. The X-axes indicate whether the cells were transfected with the Empty Vector (EV) or overexpressing-constructs for *SRRM3, SRRM4* or *MSI1*.
Fig. S8 – (A-B) RNA maps of SRRM3/4 (A) and MSI1 (B) associated binding motifs in the regions surrounding retina-enriched exons by length group and 1,000 random exons used as a control set. Both upstream and downstream introns are shown. Regions with a significant difference (FDR<0.05) in motif coverage in the tested exon group with respect to the random exon set are marked by thicker lines. Sliding window = 27 bp. (C-E) RNA maps NOVA1 (C), RBFOX (D) and ELAVL (E) associated binding motifs in the regions surrounding RetMICs and neural-enriched microexons (Neural MICs) by length group and 1,000 random exons used as a control set. Regions with a significant difference (FDR<0.05) in the motif coverage with respect to the random exon set are marked as thick lines. (F) Expression levels (cRPKMs) for Elavl, Nova and Rbfox family members across mouse developing rods, cerebellum, neurons, embryonic brain and whole brain samples (data from VastDB).
Fig. S9  – (A) Schematic representation of the structure of *srrm4* and *srrm3* in the WT and MUT eMIC domain lines, highlighting their main protein domains; N-term: *srrm4* specific N-terminal domain; RS-domain: serine/arginine repetitive matrix domain; cwf21: PFAM domain binding to the spliceosome; eMIC: enhancer of microexons domain. (B) Sequences of *srrm3* and *srrm4* WTs and MUTs, highlighting the region of the mutation. For each gene, two founders have been generated. (C) RT-PCR assays showing the inclusion of four microexons in HEK293 cell lines upon ectopic expression of GFP, *srrm3*, founder 1 mutated-*srrm3* (srm3 F1), founder 2 mutated-*srrm3* (srm3 F2), *srrm4* and founder 1 mutated-*srrm4* (srm4 F1).
**Fig. S10** – OKR recordings for WT, *srrm3* HET and *srrm3* MUT siblings at 10 dpf, in response to different contrasts. N = 6 for WT, N = 13 for *srrm3* HET and N = 3 for *srrm3* MUT. P-values from Kruskal-Wallis tests.
**Fig. S11** - (A-G) ZPR-1 staining in WT (A), *srrm3* MUT (B) and *srrm4* MUT (C) at 5 dpf, and WT (D) and *srrm3* MUT (E) at 10 dpf. Arrows show partial *arr3a* mislocalization. Quantifications for ZPR-1 positive areas are provided in (F) and (G). N = 5 for WT and *srrm3* MUT 5 dpf fish, N = 4 for WT and *srrm3* MUT 10 dpf fish, and N = 3 for *srrm4* MUT. P-values correspond to one-way ANOVA (F) and unpaired t-test (G). (H-I) ZPR-3 staining for DMUT at 5 dpf; N = 2. (J) Glutamine Synthetase (GS) staining for *srrm3* MUT and WT siblings of retinae at 5 dpf; N = 3. (K) Quantification of mitochondrial number/field and total mitochondria area/field in WT, *srrm3* MUT and DMUT. One-way ANOVA test with Tukey post-hoc analysis was applied. At least 4 eyes/genotype and 13 fields/genotype were analyzed. Significance code: ****, \( P = 0 \); ***, \( 0 < P < 0.001 \); **, \( 0.001 \leq P < 0.01 \); *, \( 0.01 \leq P < 0.05 \).
**Fig. S12** - (A-C) Representative ERG tracks at maximum light intensity from individual WT, *srrm3* HET and *srrm3* MUT fish (A), DMUT and *srrm3* MUT (B), and WT and *srrm4* MUT at 5 dpf (C). (D) Average values for b-wave amplitudes from ERG recording from WT and *srrm4* MUT at 5 dpf, upon different light stimuli (1%, 10% and 100%). Recordings were done in two independent experiments. Combined N = 17 for WT and N = 23 for *srrm4* MUT. For statistical comparison, one-way ANOVA test was applied. Error bars correspond to SEM.
Fig. S13 – (A) *srrm3* and *srrm4* expression assessed by quantitative PCR (qPCR) assays in WT heads at 24 hpf and WT eyes at different time points (from 48 hpf to 20 dpf). For each time-point, two biological replicates were analyzed (N=2). Error bars correspond to SEM. (B) *srrm4* expression detected by qPCR in *srrm3* MUT eyes at different time points compared to WT siblings’ eyes. N = 3 for 5dpf, and N = 2 for all the other time-points. Error bars correspond to SEM. P-values from unpaired t-tests. (C) RT-PCR assays showing the inclusion of RetMICs (*mef2cb, cep290, cc2d2a* and *ifi88*) in WT heads at 24 hpf and WT and *srrm3* MUT eyes at different time points.
Fig. S14 - (A) Box plots showing the changes between WT and srrm3 MUT in inclusion levels (ΔPSI (MUT-WT), left Y-axis) for microexons and long exons mis-spliced at 5 dpf (ΔPSI(MUT-WT<-25), and in gene expression (log2FC(MUT/WT) for genes downregulated at 5 pdf (log2FC(MUT/WT<1.5), right Y-axis) at earlier time points. For each time point, only exons with sufficient read coverage are shown. (B) Selected GO categories (Y-axis) from a GSEA performed on the log2 expression ratio between the srrm3 MUT and the relative WT at different developmental time points (X-axis). The size of the dots is proportional the absolute value of the normalized enrichment score (NES), the color represents the adjusted p-value.
(Padj; where all adjusted p-values greater than 0.05 were set to 0.055), and the stroke further differentiate significant (Padj ≤ 0.05; thick border) and not significant (Padj > 0.05; thin border) GO categories. The background colors differentiate between the GO categories with positive (red) and negative (blue) NES, respectively associated with upregulated and downregulated genes in the MUT vs the WT. (C) Expression levels of PR-associated genes (opn1sw1, pde6ha, arr3a and rho) assessed by qPCR in WT and srrm3 MUT eyes at different time points. Two biological replicates were analyzed for all time points except for 5 dpf, where N = 3. Error bars correspond to SEM. P-values for 5 dpf from unpaired t-tests.
Fig. S15 – Representative images of ZPR-3 (A) and Phalloidin (B) staining on retina sections for srrm3 MUT and WT siblings at 60 hpf. Similar negative results were observed in two independent experiments.
Fig. S16—(A) RT-PCR assays showing the inclusion of microexons (in *impdh1*, *itsn1*, *arl6*, and *bbs9*) in the eyes of WT and *srrm3* MUT 5 dpf fish upon injection of different concentrations of mRNA encoding the human SRRM3 eMIC. Microexon inclusion levels quantified using *ImageJ* are shown. (B) Phalloidin staining for *srrm3* MUT and WT sibling's retinas at 5 dpf upon eMIC injection. The violin plots show the quantification of the thickness of the Outer Nuclear Layer (ONL) for each genotype and condition (N=4). One-way ANOVA was applied to measure significance of each injected MUT condition respect to the control WT.
Supplementary Table Legends

**Dataset 1 - Information about vertebrate RetMICs and RetLONGs.** List of RetMICs and RetLONGs for all the studied species: human (hg38), mouse (mm10), chicken (galGal6) and zebrafish (danRer10). From columns 1 to 4, general information for each exon: gene name, event VastID, event genomic coordinate and exon length. Columns 5 to 16 contain sample and inclusion information: number of Retina replicates, number of other tissue groups (Others), average PSI in Retina, average PSI in Others, ΔPSI(Retina-Others), Standard Deviation (SD) of PSI distributions in Retina, SD of PSI distributions in Others, minimum average PSI in Others, maximum average PSI in Others, number of Neural samples, average PSI in Neural samples, SD of PSI distribution in Neural samples and the Retina Specificity Score (RSS). The last two sheets contain the inclusion of RetMICs and RetLONGs using the PSI metric in WT and *srrm3* MUT fish 5dpf eyes.

**Dataset 2 - GO enrichment analysis of genes containing RetMICs and RetLONGs.** Results of all GO enrichment analysis mentioned in the manuscript (enrichment of RetMICs and RetLONGs in human (hg38), mouse (mm10), chicken (galGal6) and zebrafish (danRer10); enrichment of Neural UP genes in human), performed through gprofiler2 in R as described in Methods. Each sheet contains: Genome assembly (Species), GO category name, Significance (TRUE if Pvalue ≤ 0.05, FALSE otherwise), Pvalue (FDR corrected), Precision (Proportion of tested genes belonging to a given GO category), Observed (Number of tested genes belonging to a given GO category), Expected (Number of tested genes expected to belong to a given GO categories in a set of randomly chosen genes), Observed vs Expected (ratio Observed/Expected). GO categories are ordered by ascending Pvalue.

**Dataset 3 - Overlaps with disease-related gene sets.** Settings, input values and results for the enrichment tests of the human genes containing RetMICs among loci associated with different retinal diseases are listed in the sheet “Hypergeometric tests”. The sheet “Gene lists” includes the name of the genes associated with each category and disease.

**Dataset 4 - Domain annotations for RetMICs and RetLONGs.** VastDB automatic PFAM and PROSITE domain annotations for human RetMICs and RetLONGs. Annotations for all exons can be downloaded from VastDB. Events IDs are provided in column 1. C1 and C2 correspond to the exons upstream and downstream the alternative exon (A).
**Dataset 5 - Structural analysis of RetMIC-containing proteins.** 32 microexon events are listed with their relative gene name, Uniprot ID, predicted ORF disruption (yes/no). Structural model information as PDB identifier, title and molecule are given at columns 5-7. Modbase sub-models used in this study can be identified by their template structure and modelling quality respectively at column 5 and 6. Four proximal residues with relative solvent accessibility and secondary structure are listed at columns 8, 9 and 10. Column 11 indicates the upstream (C1) or downstream (C2) exons used to localise the target RetMIC within the structure.

**Dataset 6 - Regulation of retina-enriched exons by MSI1 and SRRM3/4 in cell lines.** Inclusion of RetMICs, RetLONGs (27nt < length ≤ 50nt) and RetLONGs (length >50 nt) among all the cell lines and experiments included in this study. Columns 1-4: name, event VastID, event genomic coordinate, exon length. Columns 5-14: for each comparison between the control and cells expressing MSI1 (red header), SRRM3 (purple) or SRRM4 (blue) ectopically, ΔPSI (PSI_OE - PSI_Cont) is provided. NA indicates that the control and/or the experimental condition did not have sufficient read coverage to reliably estimate inclusion levels. Conditional formatting for ΔPSI values has been added from -100 (red) to 100 (green).

**Dataset 7 - Source data for the OKR test.** - OKR recording data for WT, srrm3 HET and srrm3 MUT fish. For each fish, six values are shown, corresponding to the eye speed recorded under six different contrasts (100%, 70%, 40%, 20%, 10% and 5% Maximum Contrast). For each genotype, average, standard deviation (sd), total number of fish (n) and standard error of the mean (sem) are provided in the yellow boxes.

**Dataset 8 - Differentially expressed genes in srrm3 MUT eyes at different time points.** Differentially expressed genes in srrm3 MUT eyes respect to the WT siblings (|log2FC(MUT/WT)| > 1) at all the time-points considered in the study (48 hpf, 60 hpf, 72 hpf and 5 dpf) are listed, including Ensembl ID, gene symbol, expression levels (cRPKM) in WT and srrm3 MUT eyes, and log2 fold change (log2FC(MUT/WT)). The genes with log2FC(MUT/WT) < -1.5 are highlighted in blue. Genes that are lowly expressed or supported by few reads have been excluded, following the default parameters of vast-tools compare_expr.
Dataset 9 - Enriched GO terms for *srrm3* MUT downregulated genes at 5 dpf (source data for Fig. 4C). GO enrichment analysis for the downregulated genes at 5 dpf, considering only the genes with log2FC(MUT/WT) < -1.5 and performed using the software ClueGO. The table includes: GO ID, GO Term, Pvalue, Pvalue corrected with Bonferroni step down, GO Group, % of associated genes and number of genes. The different GO groups are highlighted in blue, green and grey as represented in Fig. 4C.

Dataset 10 - Information about microexons, longer exons and genes for all the developmental stages. For microexons and longer exons: PSIs and changes in PSI between WT and *srrm3* MUT eyes (ΔPSI(MUT-WT)) at at 48 hpf, 60 hpf, 72 hpf and 5 dpf for RetMICs or RetLONGs, or microexons or long exons mis-spliced at 5 dpf (ΔPSI(MUT-WT<-25); “MICs DOWN at 5 dpf” and “LONGs DOWN at 5 dpf”, respectively). PSI values at 24 hpf heads are also included. From columns 1 to 4, general information for each exon: gene name, event VastID, event genomic coordinate and exon length. For genes: cRPKMs and changes in expression between WT and *srrm3* MUT eyes (log2FC(MUT/WT)) at at 48 hpf, 60 hpf, 72 hpf and 5 dpf for RetGENES or genes downregulated at 5 dpf (log2FC(MUT/WT<1.5); “Genes DOWN at 5 dpf”). Expression values at 24 hpf heads are also included. From columns 1 to 4, general information for each exon: gene name, event VastID, event genomic coordinate and exon length.

Dataset 11 - Source data for the ZPR-1/ZPR-3, thickness and caspase-3 analyses. The analysis of ZPR-3 and ZPR-1 positive areas, calculated as Integrated Density/Area, is shown in the first (“ZPR-1”) and second sheets (“ZPR-3”). Time-points and genotypes are highlighted in green. Outer Nuclear Layer (ONL), Inner Nuclear Layer (INL), Ganglion Cell Layer (GCL) and Inner Plexiform Layer (IPL) thickness measurements (μm) for *srrm3* MUT and WT retinas are listed in the third sheet (“layer thickness”). Time-points and layer types are highlighted in yellow. For each fish, three values are shown, corresponding to the layer thickness measured in three different areas of the retina along the apical-basal axis. The analysis of the Caspase3 staining for *srrm3* MUT and WT retinas is shown in the fourth sheet (“caspase3”). For each fish, the number of apoptotic cells across the entire ONL is provided. The last sheet (“eMIC injection thickness”) includes the analysis of ONL thickness (μm) for *srrm3* MUT and WT retina injected with the eMIC domain.
Dataset 12 - Source data for the mitochondrial analysis. EM analysis data for WT and mutant fish. The first sheet (“mitochondria area”) includes the measurements of mitochondrial area for WT, srrm3 MUT and DMUT fish. It also includes the data about the averaged mitochondrial area per field, the total mitochondria area per field and the number of mitochondria per field. The second sheet (“mitochondria perimeter”) includes the measurements of mitochondrial perimeter for WT, srrm3 MUT and DMUT fish, with the data about the averaged mitochondrial perimeter per field. The third sheet (“ipr space area”) includes the measurements of interphotoreceptor (ips) space area for WT, srrm3 MUT and DMUT fish.

Dataset 13 - Source data for the ERG recordings. b-wave ERG recordings data for WT and mutant fish. The first sheet (“srrm3 MUT”) includes the recordings for WT, srrm3 MUT (MUT) and srrm3 heterozygous (HET) fish. The second sheet (“srrm3 srrm4 DMUT”) includes the recordings from in-crosses of srrm4-/-;srrm3+/- fish. The labels (WT, MUT, HET) thus refer to the srrm3 genotype with a srrm4-/- background. For each fish, three values are shown (1 %, 10 % and 100% light intensity). For each genotype, average, standard deviation (sd), total number of fish (n) and standard error of the mean (sem) are provided in the yellow boxes.

Dataset 14 - Source data for GSEA. Results of GSEA performed on log2 fold changes between the zebrafish srrm3 MUT and the relative WT at 48 hpf, 60 hpf, 72 hpf and 5 dpf. Each table includes the GO category (pathway), the raw Pvalue (pval), the BH-adjusted Pvalue (padj), the enrichment score (ES), the normalized enrichment score (NES), the number of times a random gene set had a more extreme enrichment score value (nMoreExtreme) and the number of genes belonging to each GO category (size). The results are ordered by padj.

Dataset 15 - Config file for calling tissue-enriched microexons. Config files used to run Get_Tissue_Specific_AS.pl to derive sets of tissue-enriched microexons from PSI tables, for all the species considered in the manuscript. Column 3 “Excluded” contains the tissue groups (e.g. Retina, Muscle, Heart) excluded from the pool of compared tissues. This column must be empty to include all the tissues in the comparisons.

Dataset 16 - Primer sequences used in this study. It contains two sheets, one for human and another one for zebrafish, indicating primer name, sequence and application in the study.
**Dataset 17 - RNA-seq data used in this study.** For each species, SRA identifiers, read number, read length and source are provided for each public RNA-seq file used in this study. For RNA-seq samples generated in this study, mapping RNA statistics are provided in the last sheet.