RESEARCH

Combining a prioritization strategy and functional studies nominates 5′UTR variants underlying inherited retinal disease

Alfredo Dueñas Rey[1,2], Marta del Pozo Valero[1,2,3†], Manon Bouckaert[1,2†], Katherine A Wood[4],
Filip Van den Broeck[5,6], Malena Daich Varela[7,8], Huw B Thomas[4], Mattias Van Heetvelde[1,2], Marieke De Bruyne[1,2],
Stijn Van de Sompele[1,2], Miriam Bauwens[1,2], Hanne Lenaerts[1,2], Quinten Mahieu[1,2], Dragana Josifova[9],
Genomics England Research Consortium, Carlo Rivolta[11,12,13], Raymond T O'Keefe[4], Jamie Ellingford[4,10,14],
Andrew R Webster[7,8], Gavin Arno[7,8], Carmen Ayuso[3,15], Julie De Zaeytijd[5,6], Bart P Leroy[1,5,6,16],
Elfride De Baere[1,2] and Frauke Coppieters[1,2,17*]

Abstract

Background 5′untranslated regions (5′UTRs) are essential modulators of protein translation. Predicting the impact of 5′UTR variants is challenging and rarely performed in routine diagnostics. Here, we present a combined approach of a comprehensive prioritization strategy and functional assays to evaluate 5′UTR variation in two large cohorts of patients with inherited retinal diseases (IRDs).

Methods We performed an isoform-level re-analysis of retinal RNA-seq data to identify the protein-coding transcripts of 378 IRD genes with highest expression in retina. We evaluated the coverage of their 5′UTRs by different whole exome sequencing (WES) kits. The selected 5′UTRs were analyzed in whole genome sequencing (WGS) and WES data from IRD sub-cohorts from the 100,000 Genomes Project ($n = 2397$ WGS) and an in-house database ($n = 1682$ WES), respectively. Identified variants were annotated for 5′UTR-relevant features and classified into seven categories based on their predicted functional consequence. We developed a variant prioritization strategy by integrating population frequency, specific criteria for each category, and family and phenotypic data. A selection of candidate variants underwent functional validation using diverse approaches.

Results Isoform-level re-quantification of retinal gene expression revealed 76 IRD genes with a non-canonical retina-enriched isoform, of which 20 display a fully distinct 5′UTR compared to that of their canonical isoform. Depending on the probe design, 3–20% of IRD genes have 5′UTRs fully captured by WES. After analyzing these regions in both cohorts, we prioritized 11 (likely) pathogenic variants in 10 genes (*ARL3*, *MERTK*, *NDP*, *NMNAT1*, *NPHP4*, *PAX6*, *PRPF31*, *PRPF4*, *RDH12*, *RD3*), of which 7 were novel. Functional analyses further supported the pathogenicity of three variants. Mis-splicing was demonstrated for the *PRPF31*:c.-9+1G>T variant. The *MERTK*:c.-125G>A variant, overlapping a transcriptional start site, was shown to significantly reduce both luciferase mRNA levels and activity. The *RDH12*:c.-123C>T variant was found *in cis* with the hypomorphic *RDH12*:c.701G>A (p.Arg234His) variant in 11 patients. This

---

†Marta del Pozo Valero and Manon Bouckaert are second co-authors.

*Correspondence:
Frauke Coppieters
frauke.coppieters@ugent.be
Full list of author information is available at the end of the article

5'UTR variant, predicted to introduce an upstream open reading frame, was shown to result in reduced RDH12 protein but unaltered mRNA levels.

Conclusions This study demonstrates the importance of 5'UTR variants implicated in IRDs and provides a systematic approach for 5'UTR annotation and validation that is applicable to other inherited diseases.

Background

Among the numerous non-coding regulatory regions found across the human genome, 5′ untranslated regions (5'UTRs) are major determinants of post-transcriptional control and translation efficiency [1, 2]. These regions, with an average length of about 200 nucleotides in humans, are located immediately upstream from protein-coding sequences and include the Kozak consensus sequence around the AUG start codon [3]. 5'UTRs harbor numerous *cis*-regulatory elements such as internal ribosomal entry sites (IRES) and upstream open reading frames (uORFs), which can recruit scanning ribosomes and initiate translation [4]. In particular, uORFs, defined by an upstream start codon (uAUG) in-frame with a stop codon preceding the end of the primary open reading frame, can decrease downstream protein expression up to 80% [5]. Furthermore, 5'UTRs can serve as platforms for the formation of secondary and tertiary mRNA structures like stem-loops, hairpins, and RNA G-quadruplexes which further influence mRNA translation [6, 7]. Given the critical importance of 5'UTRs as modulators of protein expression, genetic variation within these regions can contribute to disease pathogenesis, as shown by several examples in a wide range of inherited diseases [8–11].

With the expanding implementation of massively parallel sequencing technologies in clinical practice, the rare disease field has witnessed a true paradigm shift enabling improved molecular diagnosis for many affected individuals [12–15]. Despite these advancements, and although this varies greatly across individuals with diverse clinical indications, a significant fraction of cases with suspected Mendelian disorders remains unsolved [12, 13, 16–18].

To date, 5'UTR variants are often not evaluated because of the multiple challenges they entail. Firstly, detection relies on inclusion of these non-coding regions in targeted resequencing panels or whole genome sequencing, which has been limited to date [19]. Secondly, interpretation of the effect is more difficult due to different mechanisms [8] and our relative lack of understanding of the functional consequences of 5'UTR variants. Finally, functional evidence is necessary to support their pathogenicity and hence confirm the molecular diagnosis. Recently,

to close the annotation and interpretation gap, several *in silico* tools [20–26] and guidelines [27] have been developed, although their specific applicability to the accurate and comprehensive interpretation of the diverse pathogenic mechanism of 5'UTR variants is yet to be fully established.

A clear instance of both the effectiveness and remaining challenges of clinical large-scale genome analyses can be found in inherited retinal diseases (IRDs) [28–31]. IRDs comprise a genetically and phenotypically diverse constellation of visually debilitating conditions affecting over 2 million people globally [32, 33]. With over 280 known disease genes [34], diverse modes of inheritance [29], intersecting phenotypes [35, 36], and available automated approaches to variant interpretation, establishing a genetic diagnosis in IRD patients can be challenging. Even after the use of unbiased approaches such as whole exome (WES) and whole genome sequencing (WGS), up to 50% of cases remain unsolved [15, 28, 37–39], and hence without the possibility of potential clinical intervention including gene therapy-based treatments [40, 41].

Recent studies have revealed an important contribution of non-coding variation to IRDs, particularly with the identification of deep-intronic mis-splicing variants [39, 42–51]. Moreover, there is a growing body of evidence supporting the emerging role of genetic variation affecting *cis*-regulatory regions in the molecular pathogenesis of IRDs [46, 49, 52–56]. However, thus far, only a few 5'UTR variants have been reported to be implicated in ocular diseases, including IRDs [49, 53, 54, 57–59], but no large-scale studies have been performed yet.

In this work, we conducted a systematic evaluation of 5'UTR variation in IRD genes using WGS and WES data derived from two IRD cohorts comprising 2417 participants of the 100,000 Genomes Project [12] and 1682 local cases, respectively. Firstly, we obtained a comprehensive selection of 5'UTRs of the most abundant canonical as well as non-canonical protein-coding IRD gene isoforms by performing transcript-level re-quantification of retinal expression data. We then screened these regions for variants in both IRD cohorts and developed a prioritization strategy to identify candidate pathogenic 5'UTR variants.

Dueñas Rey *et al. Genome Medicine*   (2024) 16:7

Page 3 of 21

This allowed us to identify 11 potentially causative 5'UTR variants, 10 of which were found in unsolved cases. Functional validation of the predicted pathogenetic mechanism was performed for five of these, further supporting the potential implication in disease for three variants. Overall, we show that 5'UTRs represent understudied targets of non-coding variation that can provide novel molecular diagnoses in IRDs and demonstrate the importance of reassessing these regions in existing exome and genome sequencing data.

Methods

Re-analysis of retinal RNA-seq data and selection of retina-enriched isoforms of IRD genes

We retrieved paired-end FASTQ files (GSE115828) derived from human postmortem retina samples characterized by Ratnapriya et al. (2019) [60]. Only samples derived from donor retinas showing no features of age-related macular degeneration were evaluated ($n = 102$, 51 females and 51 males, ages ranging from 55 to 94 years, average at 74). Transcripts were quantified through pseudoalignment by *Kallisto* [61] (v.0.46.1) using default parameters. Abundance estimates in transcripts per million (TPM) for all annotated transcripts (*Ensembl* human release 107) were retrieved. A custom R [62] (v.4.0.2) script was used for annotating each transcript with its corresponding gene and biotype and flagging them as canonical and/or belonging to the Matched Annotation from NCBI and EMBL-EBI (MANE) and MANE Plus Clinical set (version 1.0) [63]. For each group of isoforms belonging to each gene, we computed the average of abundance estimates across all samples and the isoform exhibiting the highest average among its corresponding group was deemed retina-enriched. This dataset was further filtered by retaining only protein-coding isoforms derived from selected genes, namely IRD genes listed in either the *Retinal disorders panel* (v2.195) from Genomics England PanelApp [64] or RetNet (https://sph. uth.edu/retnet/) (Additional file 1: Table S1). Additionally, we integrated cap analysis gene expression sequencing (CAGE-seq) data derived from fetal and adult retina (FANTOM5 [65] Robust Peak Set using an expression RLE normalized threshold >1) to evaluate the confidence of the annotated transcription start sites (TSS) of the selected transcripts in retina.

To aid tiering of variants, the Genomics England PanelApp [64] gene color-coded rating system was used: genes with diagnostic-grade rating, borderline evidence and research candidates were flagged with green, amber, and red, respectively. To further assess the retina-enrichment of specific isoforms, their loci were inspected using an integration of multiple publicly available multi-omics datasets derived from human retina (Additional file 1: Table S2).

Comparison of 5'UTRs of canonical and non-canonical isoforms and selection of 5'UTR variant search space

The exact genomic coordinates of the start and end positions of the 5'UTRs of all genes were downloaded from *Ensembl biomart* (Human Ensembl Genes 107, GRCh38. p13) and filtered to only include the transcripts defined above. We first assessed the number of exons composing these 5'UTRs to identify which IRD genes have spliced 5'UTRs. For each IRD gene for which a retina-enriched non-canonical isoform was identified, we compared the canonical and non-canonical 5'UTRs and computed their respective overlap. The calculated overlaps were then used to classify these genes into three different categories, namely genes with transcripts displaying: (i) fully distinct 5'UTRs, (ii) partly overlapping 5'UTRs, and (iii) fully overlapping 5'UTRs.

For each 5'UTR exon of all selected transcripts, we defined a near-splice region by including 25-bp intronic sequence up- and downstream of its respective splice donor and acceptor sites (excluding promoter and coding sequences). The resulting coordinates were annotated with their corresponding transcript identifier and gene name and stored into a sorted BED file (herein after referred to as *5'UTR analysis file*) for downstream variant assessment.

Evaluation of 5'UTR capture by whole-exome sequencing

We assessed the performance of exome captures of the selected 5'UTRs based on the designs provided by the kits which were mostly used for the generation of our in-house WES data, namely the SureSelect Human All Exon V6 and SureSelect Human All Exon V7 (Agilent Technologies), as well as a selection of the most recent versions of commonly used kits from 4 different providers: SureSelectXT Human All Exon V8 (Agilent Technologies), KAPA HyperExome V2 (Roche), Twist Exome 2.0 (Twist), and Illumina Exome Panel v1.2 (Illumina). BED files containing the genomic coordinates of the capture regions were downloaded from the corresponding design catalogs and intersected with the coordinates of the 5'UTRs of interest using *bedtools intersect* (v2.26.0) [66] with default parameters. For uniformity, only the captured regions were used to compare between the different designs. Additionally, for the SureSelect Human All Exon V6 and SureSelect Human All Exon V7 kits (Agilent Technologies), both the strict union of all regions covered by baits and a version padded by ±50bp extending into intronic regions were considered for the intersections. A custom Python (v.3.6.8) script was then used to compute

Dueñas Rey *et al. Genome Medicine*   (2024) 16:7

Page 4 of 21

for each IRD gene the length proportion (%) of its 5'UTR captured by these kits.

Search of 5'UTR variants in IRD genes submitted to ClinVar

The ClinVar database (ClinVar) was downloaded in a tab-delimited format [67] directly from the FTP site (https://ftp.ncbi.nlm.nih.gov/pub/clinvar/; version from 2023-03-18) and pre-filtered to keep only entries from the GRCh38.p13 build. The resulting file was further filtered to retrieve variants located within the regions defined in the *5'UTR analysis file.* Large copy number gain/loss variants extending into coding regions as well as variants with a protein-altering or synonymous annotation for the canonical transcript were removed from this analysis.

Patient cohorts

To assess the contribution of 5'UTR genomic variation to IRDs, individuals from two different cohorts were selected for this study: (i) 2397 participants (2100 probands) from a sub-cohort of the Rare Disease arm of the 100,000 Genomes Project (Genomics England—GE—cohort) affected by posterior segment abnormalities (Additional file 1: Table S3): more specifically, we created this sub-cohort by accessing the GE 100,000 Genomes dataset (version 16) and retrieving participants belonging to the *Normalised Disease Sub Group* "Posterior segment abnormalities" using the RLabKey API [68], (ii) a WES dataset from 1682 IRD cases (1030 probands) that consented to standard-of-care genetic testing at the Center for Medical Genetics Ghent (CMGG cohort). In both cases, sequencing data aligned to GRCh38 build were included.

The 100,000 Genomes Project Protocol has ethical approval from the HRA Committee East of England—Cambridge South (REC Ref 14/EE/1112). This study was registered with Genomics England within the *Hearing and sight* domain under Research Registry Projects 465. This study was approved by the ethics committee for Ghent University Hospital (B6702021000312) and performed in accordance with the tenets of the Helsinki Declaration and subsequent reviews.

Sequencing and variant analysis

We interrogated WGS and WES data from the two cohorts described above to detect germline single-nucleotide variants (SNVs) and small insertions and deletions (indels) overlapping the regions defined in the *5'UTR analysis file*. The sequencing and bioinformatic pipelines used for processing genome data derived from the participants of the GE cohort have been described previously [12].

Samples from the CMGG cohort were tested using WES with the SureSelect Human All Exon V6, SureSelect Human All Exon V7 (Agilent Technologies) or HyperExome (Roche) enrichment kits and sequenced on HiSeq 3000 or NovaSeq 6000 instruments (paired-end 150 cycles) (Illumina). Reads were aligned to the human reference genome (GRCh38 build) with *BWA* (v0.7.15) [69] and *GATK HaplotypeCaller* (v3.82) [70] was used for calling SNVs and indels. Resulting variant call format (VCF) files from both cohorts were subsequently parsed based on the regions of interest using *BCFtools* (v1.9) [71]. Only variants satisfying the filter criteria for sequencing depth (DP>10) and genotype quality (GQ>15) were retrieved.

Variant annotation and prioritization

From each cohort, a file containing unique variants was created and formatted so that it could be annotated with the *Ensembl Variant Effect Predictor* [72] (VEP, release 107). Apart from gene and transcript information, each variant was annotated with frequency data retrieved from gnomAD [73] (genomes, v3.1.2) allele frequencies, splicing (dbscSNV [74], SpliceAI [23], MaxEntScan [75]), pathogenicity predictions (EVE [76], CADD [77]), and regulatory data [78] (all plugin versions corresponding to VEP release 107). The recently developed *UTRannotator* [20] tool was used to annotate variants that create or disrupt uORFs. The classical [79] and retinal [80] Kozak consensus sequences were considered altered when variants were identified within the position range −1 to −10 relative to the main AUG (canonical start codon). In addition, nucleotide frequency plots corresponding to the Kozak sequences of the retina-enriched and not retina-enriched transcripts evaluated in this study were generated with *WebLogo* [81] (v.3.7.12). Annotation data tables containing predicted changes in translational efficiency [82] and secondary structure minimum free energy affecting double-stranded RNA or G4 quadruplex structures [83] were downloaded (*5utr ['suter']*: https://github.com/leklab/5utr) and queried for the variant positions using *tabix* [84] (v.1.7-2); further evaluation of changes in mRNA secondary structure was performed using the *Ufold* [85] (v.1.2), *REDfold* [86] (v1.14.alpha), and *MXfold2* [87] (v0.1.2) tools. To assess whether variants were located within TSS relevant to retinal gene expression, we made use of the CAGE-seq data described above. Translation initiation-related feature data from the Human Internal Ribosome Entry Sites (IRES) Atlas [88] was also queried to annotate variants found within these regions involved in cap-independent translation initiation.

Annotated 5'UTR variants were classified into the following 7 categories: (i) introduction of an upstream start codon creating an uORF (uAUG gained), (ii) change in existing uORF, (iii) alteration of classical or retinal primary Kozak context, (iv) splicing, (v) change in

Dueñas Rey *et al. Genome Medicine*   (2024) 16:7

Page 5 of 21

translational efficiency (TE), (vi) change in secondary structure minimum free energy, and (vii) overlapping a retinal TSS and/or an IRES. Variants were first selected when their minor allele frequency (MAF) was lower than 2% in all populations. We decided to use a less stringent MAF threshold (typically set at less than 1% for rare variants) in view of the high frequency of hypomorphic and founder alleles reported for IRDs [33, 89, 90]. The resulting variants were further filtered using specific criteria for each category: (i) uAUG created in a strong or moderate Kozak context, (ii) natural uAUG loss, (iii) variant located in positions -3, -4, -5, -6, or -9, (iv) any SpliceAI Delta Score (DS_AG, DS_AL, DS_DG, DS_DL) higher than 0.2 for variants within ±25bp of intron-exon boundaries, following the original recommendation from the developers and supported by benchmarking studies [23, 91], (v) |log2FC|≥0.5, corresponding to a ~1.5-fold increase/reduction of translation with respect to the wild-type allele, (vi) fold-change |FC| ≥1.5, corresponding to a 1.5-fold increase/decrease of minimum free energy compared to the wild-type allele (threshold set

to be higher than the average of all computed changes in minimum free energy), and (vii) variant located within a TSS or IRES. Variants for which the inheritance pattern and phenotype of the patient were compatible with the gene in which the 5'UTR variant was identified were selected as candidates. This prioritization procedure is depicted in Fig. 1. For comparing the proportion of variants for each category between cohorts, the statistical analysis was performed in R using the χ^2 test. Finally, for each candidate variant identified in unsolved cases, an additional screening was performed to discard (likely) pathogenic variants, both SNVs and structural variants (SVs), in other IRD genes that could provide an alternative molecular diagnosis. For the cases from the GE cohort in which the identified 5'UTR variant remained as candidate, we requested to have a clinical collaboration with Genomics England.

The pathogenicity of missense variants was assessed using the ensemble predictor *REVEL* [92], a widely-used meta-predictor that incorporates scores derived from 13 individual *in silico* tools, for which we employed the
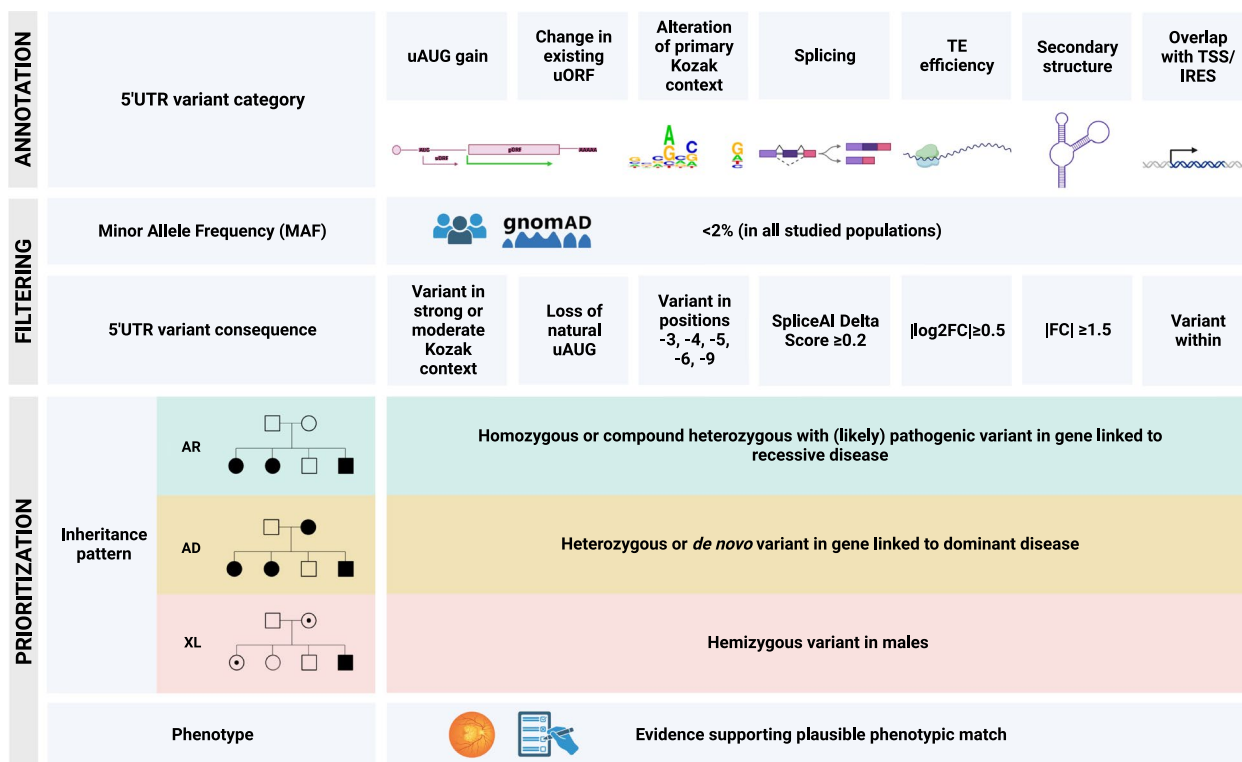

Fig. 1 Functional annotation of 5'UTR variants, filtering, and prioritization strategy followed in this study. A combination of *in silico* tools (*see "Methods"*) was used to annotate 5'UTR variants, which were then classified into 7 functional categories. For each of these categories, specific criteria were established for prioritizing variants with a more likely functional impact (*bottom*). Only variants with a minor allele frequency (MAF) <2% were further studied. The following information was then reviewed: inheritance pattern of the family (AR including sporadic cases; AD; XL) and clinical features. For the selection of candidate variants, both had to be in agreement with the reported mode of inheritance and phenotype associated with the gene in which the 5'UTR variant was found. AD autosomal dominant, AR autosomal recessive, FC fold change, IRES internal ribosomal entry site, TE translational efficiency, TSS transcription start site, uAUG upstream AUG, uORF upstream open reading frame

Dueñas Rey *et al. Genome Medicine*   (2024) 16:7

Page 6 of 21

pathogenicity threshold ranges given by Pejaver et al. (2022) [93].

Cell culture

ARPE-19 (ATCC, CRL-2302™) and HEK-293T cells (ATCC CRL-3216™) cells were grown in either DMEM:F12 (Gibco) medium supplemented with 10% fetal bovine serum (Gibco), 1% penicillin-streptomycin (Gibco), 1% non-essential amino acid solution (Gibco), and 0.1% amphotericin B (Gibco) or Dulbecco's minimal essential medium (DMEM) with phenol red (Gibco), supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin-streptomycin (Gibco), respectively. Cells were cultured at 37°C and 5% CO_2 and tested for mycoplasma contamination prior to use.

To perform functional studies of the identified 5'UTR *ARL3* variant, lymphocytes from affected carriers (*n* = 2) were isolated from EDTA blood using Lymphoprep (STEMCELL technologies). For each sample, two cultures were started in RPMI medium with 10% fetal bovine serum and substituted with interleukin-2 and phytohaemagglutinin. One of both cultures was treated for 4 h with puromycin (200 μg/mL) prior to RNA extraction to suppress nonsense-mediated decay.

Cloning and mutagenesis

In order to evaluate the functional effect using dual luciferase reporter assays of three selected 5'UTRs variants (*RDH12*:c.-123C>T, *MERTK*:c.-125G>A, *PAX6*:c.-44T>C), we cloned the wild-type 5'UTR of interest (IDT gBlock) into a psiCHECK™-2 dual luciferase vector (Promega) using the Cold Fusion cloning kit (Sanbio BV). The recombinant vectors were then amplified in One Shot TOP10 Chemically Competent *E. coli* cells (Invitrogen) and purified using the NucleoBond Xtra Midi kit (Filter Service S.A). For the generation of the overexpression construct to further assess the *RDH12*:c.-123C>T variant, we designed a gBlocks™ fragment comprising the wild-type 5'UTR and coding sequence (CDS) of *RDH12*; both sequences were modified to include downstream Myc and FLAG in-frame tags to evaluate the translation of the uORF introduced by the *RDH12*:c.-123C>T variant and RDH12 protein levels, respectively. These fragments were then cloned into a pcDNA™3.1$^{(+)}$ (Invitrogen) vector by restriction-ligation cloning and the recombinant vectors were amplified and purified as described above. For all constructs, 5'UTRs variants were introduced using the Q5 Site-Directed Mutagenesis Kit (NEB) using variant-specific primers designed with the *NEBaseChanger* tool. The sequence of each insert was confirmed by Sanger sequencing using the BigDye Terminator v3.1 kit (Life Technologies) and/or long-read

whole plasmid sequencing (Plasmidsaurus). A schematic overview of these constructs is shown in Additional file 2: Figure S1. All primer sequences can be found in Additional file 1: Table S4.

For the splicing minigene assay, a 284bp fragment of the *PRPF31* (NM_015629.3) 5'UTR was amplified from commercially available reference genomic DNA (Promega) using either wild-type or variant-specific primers (Additional file 1: Table S4) using Phusion High Fidelity DNA Polymerase (NEB) according to manufacturer's recommendations. The wild-type and mutant *PRPF31* fragments were cloned into the SK3 plasmid (a derivative of the pSpliceExpress minigene splice reporter vector, gifted from Stefan Stamm, Addgene #32485) using the Gibson method. Constructs were transformed into competent bacteria, candidate colonies were cultured, and vector DNA was isolated using the GenElute Plasmid Miniprep kit (Sigma). Sequences of the minigene vector constructs were verified using Sanger sequencing by Eurofins Genomics.

Dual luciferase assays

ARPE-19 cells were seeded in a 24-well plate (Greiner Bio-One BVBA) at a density of 50,000 cells/well in 1 mL of medium without antibiotics to reach 70–90% confluence at transfection (18–24h after plating). Cells were transfected at a 3:1 reagent to plasmid DNA ratio using the TransIT-X2® Dynamic Delivery System (Mirus Bio) according to the manufacturer's instructions. After 24 h, cells were lysed, and luciferase activity was detected using the Dual-Glo® Luciferase Assay System (Promega) in a Glomax 96-Microplate Luminometer (Promega). Each transfection was performed in triplicate and each experiment was repeated at least three times to ensure reproducibility. For each well, the ratio of *Renilla* luciferase activity was normalized to *Firefly* luciferase activity. A custom R (v4.3.0) script was used to evaluate the effect of each variant on luciferase activity through a linear mixed effects model (implemented in the *lme4* package [94]) having set the luciferase vector as fixed effect and the biological replicate as random effect.

Splicing minigene assays

Human K562 cells (ATCC, CCL-243™) were plated in 2ml RPMI-1640 (Sigma) supplemented with 10% fetal bovine serum (Sigma) at approximately 80% confluency in 6-well tissue culture plates (Corning) on the day of transfection. Cells were transiently transfected with 2.5μg *PRPF31* plasmid DNA using Lipofectamine 3000 (Invitrogen) following the manufacturer's protocol and incubated for 20–24 h at 37°C with 5% CO_2.

Dueñas Rey *et al. Genome Medicine*   (2024) 16:7

Page 7 of 21

RNA isolation

Total RNA was extracted using either the RNeasy Mini kit® (Qiagen) (ARPE-19 and HEK-293T cells), TRIzol reagent (Invitrogen) (K562 cells), the Maxwell RSC simply RNA kit (Promega) (cultured lymphocytes), or the PAXgene Blood System kit (Qiagen) (whole blood) according to the manufacturer's instructions. Isolated RNA was cleaned up using the RNeasy Mini RNA Isolation Kit (Qiagen) and underwent either an on-column DNA digestion step (K562 cells) or DNase treatment (ArcticZymes, Tromsø, Norway) prior to cDNA synthesis with the iScript cDNA Synthesis (Bio-Rad Laboratories), the High Capacity RNA to cDNA (Applied Biosystems), or the Superscript IV with random hexamers (Invitrogen) kits according to supplier's recommendations.

Quantitative polymerase chain reaction (qPCR)

For each cDNA sample, qPCR assays were prepared using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad Laboratories) and run on LightCycler 480 System (Roche). Data were analyzed with *qbase+* [95] (CellCarta, v.3.4) and normalized either to a set of housekeeping genes (*YWHAZ*, *HPRT1*, *HMBS*, *SDHA*) or *Firefly* luciferase for normalization of *Renilla* luciferase mRNA levels. All primer sequences can be found in Additional file 1: Table S4. Statistical analyses were performed in R using the Wilcoxon rank sum test.

Reverse transcription PCR (RT-PCR)

RT-PCR reactions were performed with Phusion High Fidelity DNA Polymerase (NEB) according to the manufacturer's recommendations, using cDNA as a template and primer pairs listed in Additional file 1: Table S4. PCR products were visualized using agarose gel electrophoresis supplemented with SafeView nucleic acid stain (NBS Biologicals) and a 1kb Plus ladder (NEB) as a size standard. PCR products were purified using the PureLink Quick Gel Extraction kit (Invitrogen) and sequenced by Eurofins Genomics.

Overexpression and immunoblotting

To perform immunoblotting of RDH12 and the predicted peptide encoded by the uORF introduced by the *RDH12*:c.-123C>T variant, ARPE-19 cells were seeded in 12-well plates (Greiner Bio-One BVBA) at a density of 100,000 cells/well, allowed to settle overnight, and then transfected with the wild-type and mutant *RDH12* overexpression vectors using the TransIT-X2® Dynamic Delivery System (Mirus Bio) according to the manufacturer's instructions. The pcDNA™3.1(+) (Invitrogen) backbone vector was transfected as negative control. Four hours prior to RNA and protein isolation, cells were treated with 10μM MG-132 proteasome inhibitor (Merck

Life Science). For total protein extraction, cells were lysed with RIPA Buffer (Sigma-Aldrich) including protease inhibitory cocktail (Roche Diagnostics), phosphatase inhibitory cocktail 2, and phosphatase inhibitory cocktail 3 (Sigma-Aldrich). Protein concentrations were measured using the Pierce™ BCA Protein Assay kit (Fisher Scientific). After centrifugation and reduction with 1M DTT (Sigma-Aldrich), protein lysates were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis using either NuPAGE™ 4–12% Bis-Tris (RDH12 blot) or Novex™ 16% Tricine (uORF blot) Protein Gels (Fisher Scientific) with a ladder (Precision Plus Protein All Blue Standards, Bio-Rad Laboratories). Proteins were then transferred to a nitrocellulose membrane using the iBlot 2 Dry Blotting System (Thermo Fisher Scientific). Membranes were blocked for 2 h at room temperature in 2% ECL™ Blocking Agent (Cytiva Amersham) and incubated at 4°C overnight with anti-FLAG (1:1000, F1804, Merck Life Science) or anti-Myc (1:1000, ab9106, Abcam) primary antibodies. Membranes were subsequently incubated for 2 h at room temperature with the appropriate horseradish-peroxidase-conjugated secondary antibody (1:2500, 7076S or 7074S, Cell Signaling Technologies) and revealed with the SuperSignal™ West Dura Extended Duration Substrate (Fisher Scientific). Membranes were scanned with an Amersham Imager 680 system (GE Healthcare Life Sciences). Protein quantification was performed by firstly stripping the membranes with Restore™ PLUS Western Blot Stripping Buffer (Thermo Scientific), incubating them for 1 h at room temperature with a primary antibody against β-tubulin (1:2500, ab6046, Abcam) and for 2 h with a horseradish-peroxidase-conjugated secondary antibody (1:2500, 7074S, Cell Signaling Technologies). RDH12 (FLAG) signal intensity quantification was achieved using *ImageJ* (NIH, v.1.50i) and normalized to the amount of β-tubulin.

To further assess the translation of the peptide encoded by the uORF introduced by the *RDH12*:c.-123C>T variant, transfected HEK-293T cells were lysed as previously described, and protein lysates incubated at 4°C overnight with Dynabeads Protein A (Thermo Fisher Scientific), to which the anti-Myc primary antibody (1:1000, ab9106, Abcam) was bound. Bound proteins were subsequently eluted and subjected to anti-Myc Western Blot analysis as described above. Statistical analyses were performed in R using the Student's *t*-test.

Screening of RDH12: c.-123C>T variant

The *RDH12*:c.-123C>T 5'UTR variant was shared with members of the European Retinal Disease Consortium (ERDC: https://www.erdc.info/) to evaluate its pathogenicity and identify additional carriers of the *RDH12*:c.701G>A (p.Arg234His) missense variant to

confirm its *cis*-configuration. Primer sequences for PCR and Sanger sequencing are listed in Additional file 1: Table S4.

Results

Isoform-level re-quantification of retinal gene expression reveals differential 5'UTRs of IRD genes

To obtain a relevant selection of 5'UTRs for downstream variant screening, we performed a transcript-level re-analysis of RNA-seq data to identify protein-coding non-canonical isoforms with relevant retinal expression, i.e., higher than its respective canonical isoform, to be retained in addition to the canonical isoforms. A total of 454 canonical and non-canonical transcripts belonging to 378 IRD genes were thus selected (with maximum of 2 isoforms per gene) (Additional file 1: Table S5), from which their 5'UTRs were retrieved, resulting in 638 genomic regions (Additional file 1: Table S6). Considering these 454 transcripts, the 5'UTRs of approximately 62% of IRD genes (233/378) are only part of the first coding exon while 33% of IRD genes (126/378) have exclusively transcripts with spliced 5'UTRs, i.e., their 5'UTR comprise additional full non-coding exons. A minor fraction (19/378) of genes has isoforms with both spliced and non-spliced 5'UTRs (Fig. 2A; Additional file 1: Table S7). Of the 76 IRD genes for which a retina-enriched non-canonical isoform was identified, 20 display a fully distinct 5'UTR compared to the one of their corresponding canonical isoforms (Fig. 2B; Additional file 1: Table S8). For six of these, the non-canonical TSS is further supported by CAGE-seq data derived from both adult and fetal retina (Additional file 1: Table S5; Additional file 2: Figure S2). Two remarkable examples are *CRB1* and *RIMS2*, for which the non-canonical

isoforms (ENST00000681519.1: 48.38±14.97 TPM; ENST00000436393.6: 41.58±12.35 TPM) were found to be more abundant in retina compared to their respective canonical isoforms (ENST00000367400.8: 2.54±3.88 TPM; ENST00000696799.1: 12.19±3.85 TPM). These findings were also supported by the integration of multiple multi-omics datasets derived from human retina (Additional file 2: Figure S3; Additional file 1: Table S2).

A substantial fraction of 5'UTRs can be captured by whole-exome sequencing

In view of the large volume of existing WES data, we evaluated the performance of six recent and commonly used commercial exome capture designs from four different providers on the selected 5'UTRs of IRD genes. The kits considered in this analysis were found to display a variable performance with an average coverage of the selected 5'UTRs ranging from 7% (Twist Exome 2.0) to 39% (Illumina Exome Panel v1.2) (Fig. 3A; Additional file 1: Table S9). Besides, the fraction of IRD genes with 5'UTRs fully captured ranged from 3% (10/378) to 20% (73/378). Regarding the kits that were mostly used in-house (SureSelect Human All Exon V6 and SureSelect Human All Exon V7), although a slightly higher 5'UTR capture was observed for the kit with a higher amount of probes (SureSelect Human All Exon V6) (Additional file 2: Figure S4; Additional file 1: Table S10), we found that approximately 15% of IRD genes (57/378) have the selected 5'UTRs fully captured. This fraction increased up to 39% (148/378) (Fig. 3B; Additional file 1: Table S10) if a padded bait design, including all regions that can be confidently genotyped, was considered.
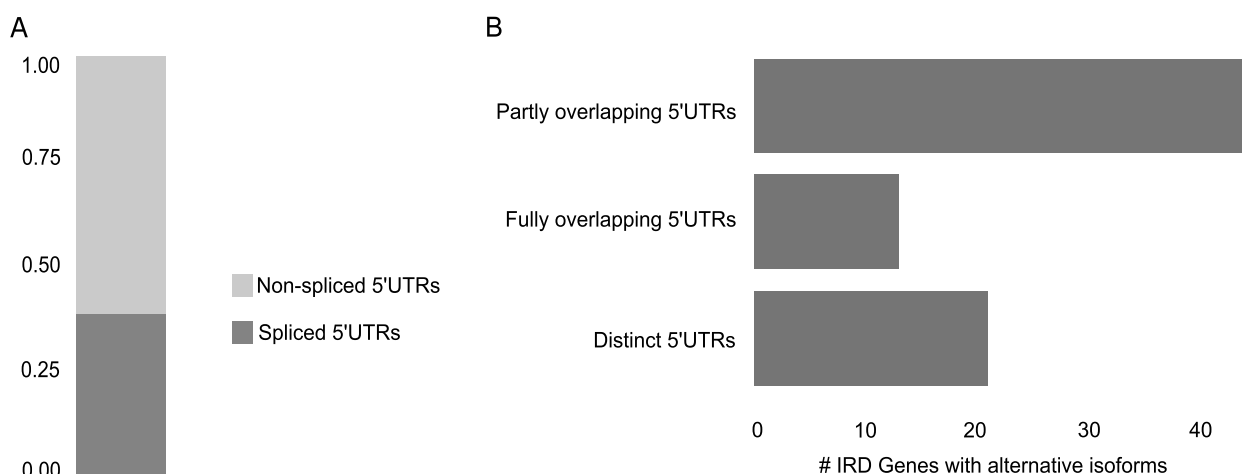

Fig. 2 Characterization of 5'UTRs of IRD genes. **A** Representation of the proportion of IRD gene isoforms based on the structure of their 5'UTRs. **B** Classification of IRD genes with a retina-enriched isoform differing to the canonical one based on the comparison of their respective 5'UTRs
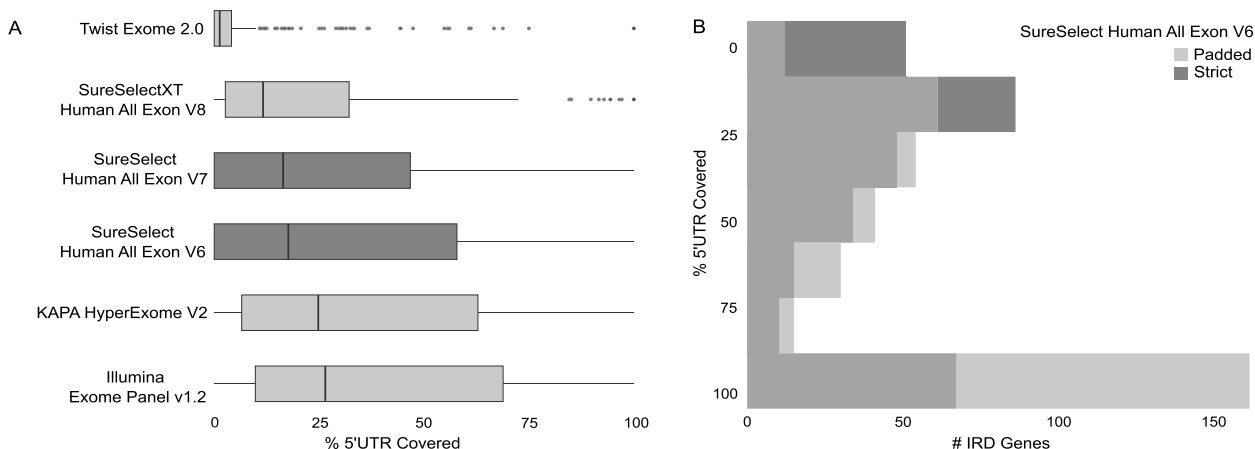
Fig. 3 Evaluation of 5'UTR capture by whole-exome sequencing (WES). **A** Boxplots showing the capture performance (*x-axis*) of commonly used commercial exome capture designs on the selected 5'UTRs. The kits that were mostly used for the generation of our in-house WES data are highlighted in *darker gray*. **B** Histogram representing the portions of 5'UTRs (*y-axis*) of the selected IRD genes which are captured by the SureSelect Human All Exon V6 kit (Agilent Technologies) considering a *strict* or a *padded* design (*see "*Methods*"*)

A retrospective analysis of 5'UTR variants reported in IRD genes reveals a majority classified as variants of uncertain significance (VUS)

A total of 1547 5'UTR variants in IRD genes have been submitted to the ClinVar database thus far (Additional file 1: Table S11). All variants except four have been clinically interpreted. Only 2% (31/1547) of the variants have been classified as pathogenic or likely pathogenic, of which 32% (10/31) and 16% (5/31) were found in the *PAX6* and *NMNAT1* genes, respectively. On the other hand, 34% (527/1547) of the 5'UTR variants have been classified as benign or likely benign, of which no single gene was found to account for more than 5% of these variants. While 5% (81/1547) of the variants had conflicting interpretations of pathogenicity, the largest fraction comprising 58% (904/1547) of all variants have been classified as VUS.

Analysis of 5'UTRs in WGS and WES data of two IRD cohorts reveals rare and ultra-rare variants predicted to affect 5'UTR function

To systematically assess the contribution of 5'UTR genomic variation to IRDs, we performed a variant analysis within the selected 638 genomic regions in affected individuals from two different IRD sub-cohorts, namely the GE (*n* = 2397 WGS) and CMGG (*n* = 1682 WES) cohorts. We identified a total of 2898 and 381 distinct 5'UTR variants within the GE and CMGG cohorts, respectively. The majority of these variants (2637/2898 and 334/381) had a MAF lower than 2%

in all populations and a substantial fraction (506/2898 and 92/381) was found to be absent from all reference population public databases. To aid the interpretation of these 5'UTR variants, we classified them into seven (non-mutually exclusive) categories according to their *in silico* predicted functional consequences (Fig. 1). A summarized overview of the number of variants that remained after classification and category-specific filtering can be found in Table 1. Out of the 1450 remaining variants (Additional file 3: Dataset S1), 336 were present in more than one category, with the majority (230/336) predicted to both overlap a retinal TSS or an IRES and change the secondary structure (Additional file 1: Table S12). Of the remaining variants assigned to a single category, the bulk corresponded to variants overlapping a retinal TSS or an IRES (761/1450; 221 -TSS-, 488 -IRES-, 52 -TSS & IRES-), followed by variants with a predicted change in secondary structure minimum free energy (237/1450). When comparing the number of variants in each category between the two cohorts, we observed statistically significant differences in the number of variants overlapping a retinal TSS and/or IRES, for which the proportion of variants was higher for the WGS-based GE cohort compared to the WES-based CMGG cohort (~55% and ~35% respectively, *p* < 0.05). These differences are most likely due to the expanded capture of regions including TSS allowed by the use of WGS in the GE cohort. Additionally, it is noteworthy that for the IRD genes with 5'UTRs harboring IRES (92/378), these elements were found to span on average 61% of the 5'UTR and even its entire length for certain genes (12/92) (Additional file 1: Table S13).

Table 1 Summarized overview of the number of 5'UTR variants that remained after functional classification and filtering

	GE cohort	CMGG cohort	Total	%
uAUG gained	30	5	35	2.41
Change in existing uORF	17	1	18	1.24
Alteration of classical or retinal primary Kozak context	16	7	23	1.59
Splicing	29	10	39	2.69
Change in translational efficiency	1	0	1	0.07
Change in secondary structure minimum free energy	194	43	237	16.34
Overlapping a retinal transcription start site or an IRES	692	69	761	52.48
> 1 category	276	60	336	23.17
Total	1255	195	1450	

CMGG Center for Medical Genetics Ghent, *GE* Genomics England, *IRES* internal ribosomal entry site, *uORF* upstream open reading frame

Systematic variant evaluation allows the prioritization of candidate pathogenic 5'UTR variants

To enhance the likelihood of identifying potential pathogenic 5'UTR variants, we prioritized the ones with higher predicted impact based on specific criteria for each category (Fig. 1) and for which the inheritance patterns and reported phenotypes fitted the genes in which the 5'UTR variants had been identified. The application of this prioritization procedure is illustrated in Additional file 1: Table S14. We prioritized 11 candidate pathogenic 5'UTR variants. Of these, 10 were found in unsolved cases and 7 have not been reported before. Of note, six of these variants are covered by the WES approaches used in the CMGG cohort. An overview of the genetic, phenotypic, and *in silico* prediction details of these variants is shown in Table 2. The pedigrees of these families including variant segregation are shown in Additional file 2: Figure S5.

Four of these variants were found to alter splicing in genes linked to autosomal dominant (*PRPF31*), autosomal recessive (*NMNAT1*), and X-linked (*NDP*) disease. The *NDP*:c.-70G>A and *PRPF31*:c.-9+1G>A variants have recently been described by Daich Varela et al. (2023) [49]. In particular, the authors of this study also predicted a deleterious effect on splicing for these variants, which could not be functionally validated as it was not possible to obtain fresh samples from the patients for further analysis. We identified a *de novo* variant (*PRPF31*:c.-9+1G>T) in the same 5'UTR position of *PRPF31* in a sporadic case with rod-cone dystrophy in which bi-allelic *GRM6* variants had been identified, one of them reported as likely benign (Additional file 1: Table S15). Variants in *GRM6* cause autosomal recessive congenital stationary night blindness [99], which is different from the clinical diagnosis in this proband. The *NMNAT1*:c.-57G>A variant was found in an individual with macular dystrophy that carries *in trans* an extremely rare missense variant (Table 2) for which *in silico* predictions (*REVEL* score = 0.797) support a pathogenic effect.

Changes in secondary structure were predicted for four variants in genes linked to autosomal dominant (*ARL3, PAX6*) and autosomal recessive (*MERTK, RD3*) disease (Additional file 2: Figure S6). Given the inheritance pattern and *in silico* predictions, a gain-of-function effect was hypothesized for the *ARL3*:c.-88G>A and *PAX6*:c.-44T>C variants. Both the *MERTK*:c.-125G>A and *RD3*:c.-394G>A variants were identified in homozygous state and found to overlap their respective TSS.

The introduction of an uORF was predicted for two variants in the *NPHP4* and *RDH12* genes. In contrast to the uORF created by the *RDH12*:c.-123C>T variant, the one introduced by the *NPHP4*:c.-21C>T variant is out-of-frame and overlaps the CDS. Although no segregation could be established for this case with non-syndromic rod-cone dystrophy, an ultra-rare missense *NPHP4* variant of uncertain significance (*REVEL* score = 0.328) was found exclusively in this individual (Table 2).

Only one variant in a gene linked to dominant disease (*PRPF4*) was predicted to affect the primary Kozak consensus sequence. The *PRPF4*:c.-6C>T variant results in a transition to a less frequent nucleotide at this position within the Kozak consensus sequences derived from the IRD gene isoforms selected in this study (Additional file 2: Figure S7). This variant was found in a sporadic case with rod-cone dystrophy and clinical signs of Bardet-Biedl syndrome in which 2 likely pathogenic variants have been identified in the *IFT140* gene (Additional file 1: Table S15); segregation analysis was not possible to establish the phase of any of these alleles.

Downstream functional analyses support the pathogenicity of the MERTK:c.-125G>A, PRPF31:c.-9+1G>T, and RDH12:c.-123C>T variants

Five of the variants (*ARL3*:c.-88G>A, *MERTK*:c.-125G>A, *PAX6*:c.-44T>C, *PRPF31*:c.-9+1G>T, *RDH12*:c.-123C>T) were selected for functional validation to showcase a representative selection of different experimental approaches

Dueñas Rey *et al. Genome Medicine*   (2024) 16:7

Page 11 of 21

Table 2 Summarized overview of the 5'UTR variants that were prioritized as candidates

ID	Cohort	Retinal phenotype	Inheritance pattern	Gene	GRCh38 coordinates	Variant	Zygosity	MAF	5'UTR variant interpretation	Segregation	Captured by WES[b]	Reference
F1[a]	GE	Macular dystrophy	Sporadic (AR)	RDH12	chr14:67722520C>T	NM_152443.3: c.-123C>T	Heterozygous	0.00008544	uAUG gain: uORF with a stop codon in 5'UTR	Yes	No	Thompson et al., 2005 [96]
					chr14:67729233G>A	NM_152443.3: c.701G>A (p.Arg234His)	Heterozygous	0.00008544	-		-	Thompson et al., 2005 [96]
					chr14:67729258_6772926del	NM_152443.3: c.735_743del (p.Cys245_Leu247del)	Heterozygous	Absent	-		-	Ba-Abbad et al., 2020 [97]
F2	GE	Rod-cone dystrophy	Sporadic (AR)	NPHP4	chr1:5986310G>A	NM_015102.5: c.-21C>T	Heterozygous	0.00001972	uAUG gain: uORF out of frame and overlapping with CDS	Not available	Yes (if padded)	This study
					chr1:5961845C>G	NM_015102.5: c.622G>C (p.Val208Leu)	Heterozygous	Absent	-		-	
F3	GE	Macular dystrophy	Sporadic (AR)	NMNAT1	chr1:9943515G>A	NM_022787.4: c.-57G>A	Heterozygous	Absent	Splice donor site loss	Yes	No	This study
					chr1:9982341G>C	NM_022787.4: c.480G>C (p.Leu160Phe)	Heterozygous	0.000006573	-		-	
F4	GE	LCA or Early-Onset Severe Retinal Dystrophy	AD	PRPF31	chr19:54115798G>A	NM_015629.4: c.-9+1G>A	Heterozygous	Absent	Splice donor site loss/gain	Yes	No	Daich Varela et al., 2023 [49]
F5	GE	Rod-cone dystrophy	Sporadic (AD)	PRPF31	chr19:54115798G>T	NM_015629.4: c.-9+1G>T	Heterozygous	Absent	Splice donor site loss/gain	De novo	No	Liu et al., 2008 [98]
F6	GE	Familial exudative vitreoretinopathy	Sporadic (X-linked)	NDP	chrX:43958715C>T	NM_000266.4: c.-70G>A	Hemizygous	Absent	Splice donor site gain	Yes	Yes	Daich Varela et al., 2023 [49]
F7	GE	Rod-cone dystrophy	Sporadic (AR)	RD3	chr1:211492150C>T	NM_001164688.2: c.-394G>A	Homozygous	Absent	Change in secondary structure and/or retinal TSS	Yes	No	This study

Table 2 (continued)

ID	Cohort	Retinal phenotype	Inheritance pattern	Gene	GRCh38 coordinates	Variant	Zygosity	MAF	5'UTR variant interpretation	Segregation	Captured by WES[b]	Reference
F8	GE	LCA or Early-Onset Severe Retinal Dystrophy	Sporadic (AR)	MERTK	chr2:111898611G>A	NM_006343.3: c.-125G>A	Homozygous	0.00002632	Change in secondary structure and/or retinal TSS	Yes	Yes	This study
F9	GE	Cone Dysfunction Syndrome	AD	PAX6	chr11:31806455A>G	NM_001368894.2: c.-44T>C	Heterozygous	Absent	Change in secondary structure	Yes	Yes	This study
F10	CMGG	Retinal dystrophy with macular involvement	AR	ARL3	chr10:102714363C>T	NM_004311.4: c.-88G>A	Heterozygous	Absent	Change in secondary structure	Yes	Yes (if padded)	This study
F11	CMGG	Rod-cone dystrophy/ BBS	Sporadic (AR)	PRPF4	chr9:113275738C>T	NM_001244926.2: c.-6C>T	Heterozygous	Absent	Alteration of primary Kozak consensus	Not available	Yes	This study

AD autosomal dominant, *AR* autosomal recessive, *BBS* Bardet - Biedl syndrome, *CMGG* Center for Medical Genetics Ghent, *GE* Genomics England, *LCA* Leber congenital amaurosis, *MAF* minor allele frequency, *TSS* transcription start site, *uORF* upstream open reading frame.

[a] The status of this case in the GE Research Environment was classified as "solved" for *RDH12*-associated disease before this study

[b] In-house capture kits (SureSelect Human All Exon V6 and SureSelect Human All Exon V7)

Dueñas Rey *et al. Genome Medicine*   (2024) 16:7

Page 13 of 21

that can be used to validate 5'UTR variation. Depending on the tissue-specific gene expression and availability of patient material, we conducted either *in vitro* evaluation in ARPE-19 (dual luciferase assays or overexpression) and K562 cells (splicing minigene assays), or mRNA expression analysis in patient-derived material (whole blood or lymphocyte cultures) (Fig. 4A).

The *RDH12*:c.-123C>T and *MERTK*:c.-125G>A variants were found to result in a significant decrease in luciferase activity (Fig. 4B). To further elucidate the underlying mechanism, *Renilla* mRNA expression analysis was performed. Relative *Renilla* luciferase mRNA levels remain unchanged for *RDH12*:c.-123C>T, whereas a significant decrease was observed for *MERTK*:c.-125G>A (Fig. 4C), hence suggesting a translational and a transcriptional effect for these variants, respectively.

Taking advantage of the expression of *ARL3* and *PRPF31* in accessible tissues, we could interrogate functionally the identified variants in these genes using patient-derived material. In particular, we conducted qPCR-based quantification of *ARL3* mRNA abundance in lymphocyte cDNA from the two affected siblings in whom we identified the *ARL3*:c.-88G>A variant and six healthy controls. No significant differences were observed in *ARL3* mRNA abundance between the affected carriers and controls (Fig. 4D; Additional file 2: Figure S8). Interestingly, *ARL3* expression levels were slightly higher in the patient samples derived from the lymphocyte cultures treated with puromycin, a translation inhibitor used to suppress nonsense-mediated mRNA decay, compared to the corresponding untreated samples and controls. Therefore, although the *ARL3*:c.-88G>A variant was not predicted to affect splicing, we also performed Sanger sequencing on the cDNA derived from the puromycin-treated and untreated patient lymphocyte cultures. Neither splicing defects nor allele-specific expression were observed (*data not shown*).

Regarding the *PRPF31*:c.-9+1G>T variant, RT-PCR on RNA isolated from patient and control human blood and K562 cell lines revealed abnormal splicing patterns for the patient compared to the control samples (Additional file 2: Figure S9A). In addition to the wild-type transcript, the patient sample contained a smaller product lacking the 3'end of exon 1. This is in line with the *SpliceAI in silico* predictions suggesting activation of a cryptic splice donor site at position -44 relative to the variant (Additional file 3: Dataset S1). These findings were also confirmed using an *in vitro* minigene splicing assay (Additional file 2: Figure S9B). In view of the negative results obtained for the *PAX6*:c.-44T>C and *ARL3*:c.-88G>A variants, the phenotypes of patients F9 and F10 were re-evaluated. Neither anterior segment abnormalities nor other clinical presentations compatible with a *PAX6*-related disease were observed in patient F9 or her affected relatives. Regarding F10, re-evaluation of one of the affected siblings revealed an acquired vascular ocular condition instead of an IRD.

The RDH12:c.-123C>T 5'UTR variant is found always in cis with the p.Arg234His hypomorphic allele and results in reduced RDH12 protein levels

We identified the *RDH12*:c.-123C>T variant in one solved case from the GE cohort characterized by bi-allelic *RDH12* coding variants (c.[701G>A];[c.735_743del], p.[Arg234His]; [Cys245_Leu247del]). The 5'UTR variant, reported as VUS in ClinVar, was found in *cis* with the p.Arg234His allele (*REVEL* score = 0.555). Further evaluation of this 5'UTR variant in 10 additional *RDH12* bi-allelic patients from 8 families carrying the p.Arg234His variant revealed that the c.-123C>T and p.Arg234His variants always form a complex allele (Additional file 1: Table S16). This was also shown for 7 carriers of the p.Arg234His variant affected by other pathologies (*data not shown*).

Here, we report the first patient who is homozygous for the *RDH12*: c.[-123C>T];[701G>A] complex allele (Additional file 1: Table S16). Compared to other patients

(See figure on next page.)

Fig. 4 Functional evaluation of candidate pathogenic 5'UTR variants in the *ARL3*, *MERTK*, *PAX6*, and *RDH12* genes. **A** Various approaches were used for functionally evaluating candidate variants, including *in vitro* studies (dual luciferase reporter assays and overexpression) and experiments with clinically-accessible tissues (expression analysis in patient-derived lymphocytes). **B** Results from the luciferase assays for the *MERTK*:c.-125G>A, *PAX6*:c.-44T>C, and *RDH12*:c.-123C>T variants. The bar plot shows, for each variant, the fold change (FC) of the luciferase reporter level relative to the level of their corresponding wild-type (WT) construct luciferase vector (FC = 1)**.** The *RDH12*:c.-123C>T and *MERTK*: c.-125G>A variants resulted in significant ($p < 0.001$) decrease in luciferase activity (~92% and ~99%, respectively). **C** Relative *Renilla* luciferase mRNA levels were significantly decreased (~42%, p<0.01) for the *MERTK*:c.-125G>A variant while they remain the same for the *RDH12*:c.-123C>T variant when normalized to mRNA of *Firefly* luciferase and compared to their corresponding wild-type (WT) construct luciferase vectors. **D** qPCR quantification of *ARL3* mRNA abundance in lymphocyte cDNA of two affected siblings carrying the *ARL3*:c.-88G>A novel variant and six healthy controls. No significant differences were observed in *ARL3* mRNA abundance between the affected carriers and controls. Data presented as the mean of biological replicates within each group ± their corresponding standard deviation. **E**–**F** Using an overexpression setting, the *RDH12*:c.-123C>T variant was shown to result in **E** unaltered mRNA levels but **F** significantly reduced (~73%, $p < 0.01$) RDH12 protein levels
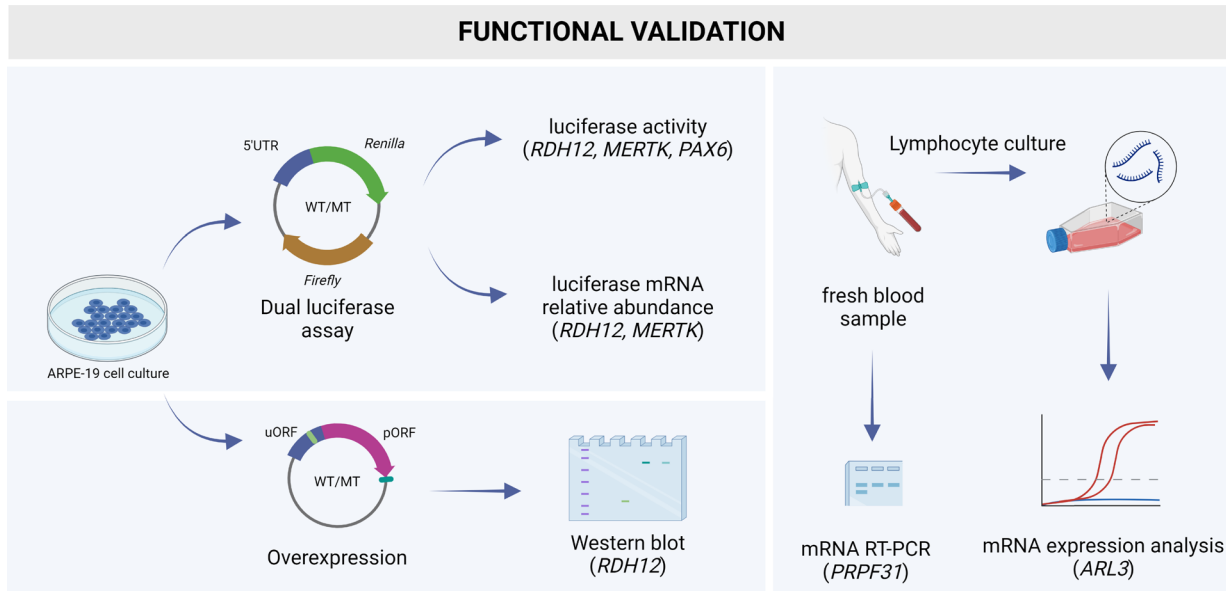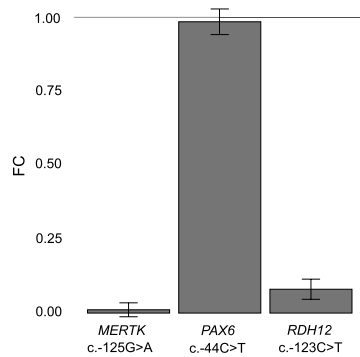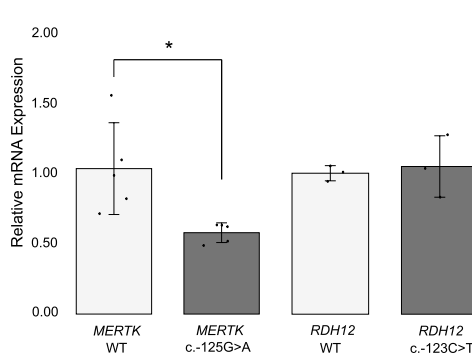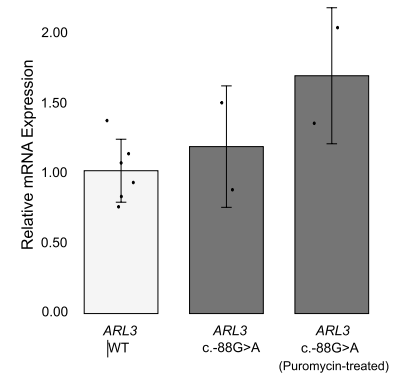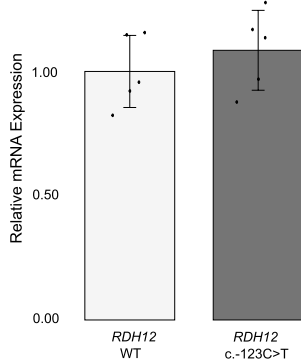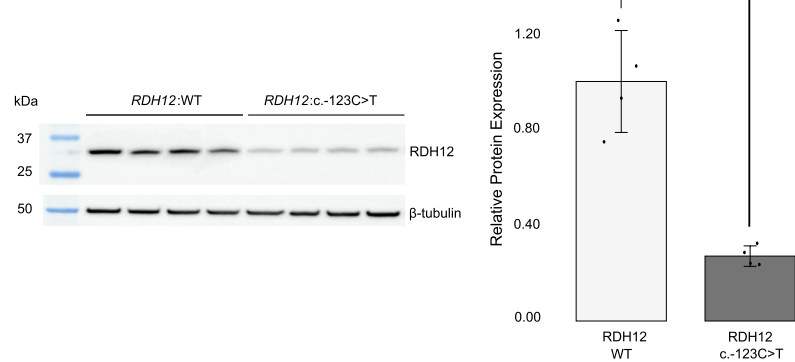
Fig. 4 (See legend on previous page.)

carrying the 5'UTR variant *in trans* with a null *RDH12* allele [100], patient F21 presents with a milder phenotype (mild decreased acuity (20/25 OD and OS) in fourth

decade of life, foveal sparing maculopathy with a circumscribed area of atrophy within the vascular arcades, without a nasal component.

The c.-123C>T variant is predicted to introduce an uAUG into a strong Kozak consensus sequence that is in-frame with a stop codon located 75 nucleotides downstream in the 5'UTR. In view of the results obtained in the luciferase assays and the predicted uORF-introducing effect of the 5'UTR variant, we further inspected the potentially exclusive effect at the translational level by assessing RDH12 mRNA abundance and protein levels in an overexpression setting. Although we could not confirm translation of the 25-amino-acid-long peptide predicted to be encoded by the uORF by co-immunoprecipitation (*data not shown*), we observed unaltered mRNA but significantly ($p < 0.05$) reduced RDH12 protein levels, thereby providing further evidence for a post-transcriptional or translational effect (Fig. 4E, F) which was already suggested by the mRNA evaluation in the luciferase assays.

Discussion

Given the emerging role of non-coding variation underlying IRDs [46, 49, 52–55] and the essential regulatory function of 5'UTRs [2, 6], we set out to evaluate their contribution to this heterogeneous group of disorders by analyzing, systematically annotating, filtering, and prioritizing 5'UTR variants in two large IRD cohorts in combination with different experimental approaches for functional validation.

Thus far, only a few studies have implicated 5'UTR variation in the molecular pathogenesis of IRD cases [49, 53, 54]. Our retrospective analysis of 5'UTR variants in IRD genes listed in the ClinVar database [67] revealed that as many as 58% of all variants have been classified as VUS. A large fraction of the underrepresentation of (likely) pathogenic 5'UTR variants in clinical databases can be explained by their exclusion from downstream tiering pipelines, which prioritize variants with protein-altering consequences and hence neglect the potential impact of 5'UTR variants [101]. Furthermore, the identification of these variants is not always feasible by means of commercial exome enrichment platforms, in particular when variants are found within 5'UTR exons that are not part of or proximal to the first protein-coding exon [19]. Here, we show that the enriched 5'UTR fraction is highly dependent on the WES capture kit used and ranges from 7 to 39%. Overall, as previous studies have also argued [102–104], our analysis underscores the diagnostic value of re-analyzing exome data, in which causative variants might already be present but previously disregarded.

Furthermore, the identification of the genetic defects underlying IRDs is greatly impacted by the inherent complexity of the retinal transcriptional landscape. Variants in retina-enriched isoforms and tissue-specific mis-splicing have been shown to be important molecular mechanisms underpinning disease pathogenesis and phenotypic heterogeneity [105–109]. As gene isoforms can display differential 5'UTRs that can result in differential translational efficiencies [110], we performed a transcript-level re-analysis of retinal expression data to obtain a relevant selection of 5'UTRs of IRD genes for downstream variant analysis. We identified 76 IRD genes with alternative isoforms exhibiting retinal expression levels higher than their respective canonical isoforms, of which 20 displayed a fully distinct 5'UTR. This analysis also found the recently identified photoreceptor-specific non-canonical *CRB1* isoform, which bears a unique 5'UTR exon [111]. Similarly, we confirmed the retinal enrichment of an alternative *RIMS2* isoform containing an unconventional 5'UTR exon, which has been shown to be photoreceptor-specific and functionally conserved in mouse (Del Pozo-Valero et al., unpublished data). By revealing that a significant fraction of IRD genes express alternative retinal isoforms, our analysis highlights the importance of isoform-aware variant annotation for adequate interpretation.

To aid the assessment of 5'UTR variant pathogenicity, *in silico* tools have recently been developed, with an emphasis put on uORF-perturbing variants [20, 26]. A systematic characterization of this class of 5'UTR variants showed that they are subject to strong negative selection, which could even be equivalent to that observed against missense variants [11]. However, thus far, these tools do not provide comprehensive annotations of all possible effects 5'UTR variants can exert [8]. Here, we designed an integrative annotation of 5'UTR-informative features and built a prioritization strategy combining family and phenotypic data to increase the probability of identifying potential pathogenic 5'UTR variants. Although we performed these analyses in IRD cohorts, the features evaluated and the methods presented in this study can be extrapolated and applied in any setting involving clinical interpretation of 5'UTR variation.

We prioritized 11 candidate pathogenic 5'UTR variants. The great majority of these variants (7/11) have not been reported before, most likely due to their exclusion from routine prioritization pipelines and subsequent clinical interpretation [101]. These 11 variants were assigned to 5 of the 7 categories we defined to classify their potential functional consequences, thus reflecting the diversity in predicted functional effect. Of note, the majority (4/11) was predicted to have an effect on splicing which, in line with our analysis

revealing that almost one third of all IRD genes have their 5'UTRs spliced, supports that these regions are also susceptible to disease-causing splicing defects. In this regard, we cannot rule out other potential spliceogenic 5'UTR variants not covered in this study, such as those affecting splicing regulatory elements [112] or those located further from the defined near-splice sites. Notably, we also provided a set of filtered annotated 5'UTR variants with *in silico* predictions pointing to a pathogenic effect that were not prioritized as candidates due to inconsistent genotypes or phenotypes in the patients evaluated. Therefore, we cannot dismiss the possibility that these variants can have a pathogenic effect when present in the right context, e.g., *in trans* with a second pathogenic allele in a gene linked to recessive disease. Additionally, we cannot discard that other 5'UTR variants identified in these patients might also affect 5'UTR function through mechanisms that were not included in the proposed strategy, such as the disruption of DNA/RNA binding protein motifs or other transcription regulatory elements found within these regions [113].

Given the particular importance of providing functional evidence to support the pathogenicity of variants located within non-coding regions, we evaluated the predicted functional consequence of five candidate variants. Despite the extreme rarity and disease co-segregation of the *PAX6*:c.-44T>C variant, which was predicted to affect its secondary structure, both the lack of functional evidence and the unusual phenotype established upon clinical re-evaluation indicate that this variant is not explaining the phenotype. Although predictions did not point to a splicing effect, aberrant splicing cannot be ruled out using our luciferase approach.

Likewise, we could not confirm the predicted effect of the *ARL3*:c.-88G>A variant using accessible patient-derived material. The sibling of the proband was considered not affected by an IRD during the course of this study, indicating that the variant is either not causative, incompletely penetrant, or simply a recessive allele. Interestingly, a phenotype similar to that of the proband has been associated with a pathogenic variant in this gene [114], which could suggest that this variant exerts a retina-specific effect that cannot be evaluated with the current experimental setting. Besides, although our predictions did not point to a post-transcriptional effect for this variant, we cannot rule out this possibility. Nevertheless, it is important to note that in this case, both siblings have been studied by WES, but were not tested for known deep-intronic and structural variants. Therefore, and given the eventual lack of disease co-segregation, before proceeding with further functional characterization of the *ARL3*:c.-88G>A variant, WGS could be performed to confirm or rule out such variants. Altogether, these results highlight the need and potential limitations of experimental assays to validate the functionality of non-coding variants.

The use of patient-derived material, however, allowed us to support the splicing defect exerted by the *PRPF31*:c.-9+1G>T variant, which had also been reported by Liu et al. (2008) [98]. Here, this *de novo* variant provided a molecular diagnosis in accordance with the clinical presentation in a patient with previous inconclusive genetic testing results.

We could provide another novel molecular diagnosis in a patient with severe early-onset retinal dystrophy that remained unsolved after WGS screening. This individual was found to be homozygous for the extremely rare c.-125G>A variant in *MERTK*. This variant is an example of the 336 variants that were assigned to more than one functional category. In particular, as a large fraction of these variants, both a change in the secondary structure and an overlap with a transcription start site were predicted. Interestingly, these sites have been shown to exhibit a high propensity to form stable secondary structures [115], which could explain the overrepresented combination of these two categories that support an effect at the transcriptional level. Indeed, a drastic reduction of luciferase activity accompanied by decreased relative *Renilla* luciferase mRNA levels confirmed the hypothesized transcriptional effect of the variant, which is consistent with a loss of function of *MERTK*. To our knowledge, this is the first reported disease-causing 5'UTR variant in *MERTK*.

One of the candidate pathogenic 5'UTR variants was found in a case solved with bi-allelic *RDH12* coding variants, namely the p.Arg234His hypomorphic missense [97, 100, 116] and the p.Cys245_Leu247del in-frame variants. The identified c.-123C>T 5'UTR variant was found *in cis* with p.Arg234His, both in this case as well as in 10 additional *RDH12* bi-allelic patients screened for this 5'UTR variant, with one of them being homozygous for this complex allele. The pathogenicity of p.Arg234His has been questioned in view of functional assays revealing RDH12 protein levels and catalytic activity comparable to the wild-type or polymorphic alleles, respectively [96]. We therefore hypothesized a pathogenic effect exerted by the *RDH12*:c.-123C>T variant as part of this complex allele. This 5'UTR variant was predicted to introduce an upstream start codon (uAUG) into a strong Kozak consensus, thereby creating an uORF. Given that ribosomes would first encounter and start translating from this
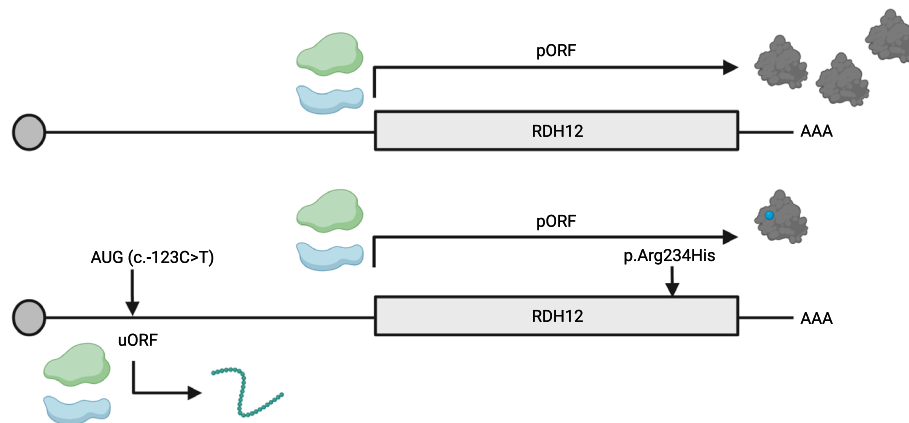
Fig. 5 Proposed pathogenetic mechanism of the *RDH12*: c.[-123C>T];[701G>A] complex allele. Graphical representation of the *RDH12* wild-type allele (*top*) and the *RDH12* complex allele harboring the c.701G>A (p.Arg234His) and the c.-123C>T variants (*bottom*). The 5'UTR variant introduces an upstream start codon into a strong Kozak sequence which can be recognized by ribosomes. Either initiation or active translation of the introduced upstream open reading frame (uORF) results in lower translational efficiency of the primary open reading frame (pORF). As a result, there is a decreased level of RDH12 protein with the arginine-to-histidine amino acid substitution at position 234 (*blue point*)

gained uAUG, we hypothesized that this variant would decrease normal RDH12 translation and thus result in lower protein levels. Furthermore, this variant has been argued to affect the function of an alternative promoter of *RDH12* based on the observed decreased activity in a luciferase reporter assay [52]. Importantly, this experimental setup did not include comparison between mRNA levels and luciferase protein activity and was therefore unable to pinpoint at which level the effect occurs. Here, using a dual luciferase assay with the 5'UTR instead of the genomic sequence, followed by downstream evaluation at both the mRNA and protein level, we confirmed the hypothesized uORF-mediated effect. We tested this hypothesis further using an overexpression setting and demonstrated unaltered mRNA but significantly reduced RDH12 protein levels, hence confirming the translational effect of *RDH12*:c.-123C>T variant. Although we could not obtain experimental evidence supporting the translation of the uORF-encoded peptide, which might suggest its instability, we cannot rule out alternative mechanisms such as ribosome stalling [9]. However, further studies are needed to confirm this hypothesis. Overall, our results indicate that not p.Arg234His alone, but rather the combination of p.Arg234His and c.-123C>T, or even c.-123C>T by itself, is disease-causing (Fig. 5).

Conclusions

5'UTRs are essential modulators of post-transcriptional and translational control. Even though they can be captured by WES to a large extent, variants within these regions are typically overlooked in both research and routine clinical settings, mainly due to their challenging interpretation. In this study, we developed a systematic strategy to comprehensively annotate and prioritize 5'UTR variants in IRD genes. This strategy combined with functional studies allowed us to pinpoint 5'UTR variants providing either novel molecular diagnoses or the full picture explaining the pathogenicity of previously reported hypomorphic variants. Overall, in this study, we highlight the importance of multi-layered annotation and validation of non-coding variation potentially underlying disease and provide a 5'UTR interpretation approach that could be extrapolated to other rare diseases.

Abbreviations

5'UTR	5' Untranslated region
AD	Autosomal dominant
AR	Autosomal recessive
BBS	Bardet-Biedl syndrome
CAGE-seq	Cap analysis gene expression sequencing
CMGG	Center for Medical Genetics Ghent
CDS	Coding sequence
DMEM	Dulbecco's minimal essential medium
ERDC	European Retinal Disease Consortium
FC	Fold change
GE	Genomics England
IRD	Inherited retinal disease
IRES	Internal ribosomal entry site
MAF	Minor allele frequency
OD/OS	*Oculus dexter/sinister* (right/left eye)
qPCR	Quantitative polymerase chain reaction
SNV	Single-nucleotide variant
SRE	Splicing regulatory element
SV	Structural variant
TSS	Transcription start site
TPM	Transcripts per million
TE	Translational efficiency
uORF	Upstream open reading frame
VUS	Variant of uncertain significance
WES/WGS	Whole exome/genome sequencing

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-023-01277-1.

Additional file 1: Table S1. Custom diagnostic gene panel comprising all IRD genes listed in either the Retinal disorders panel (v2.195) from Genomics England PanelApp or RetNet. **Table S2.** Publicly available multi-omics datasets derived from human retina used in this study. **Table S3.** Overview of the IRD sub-cohort from the Rare Disease arm of the 100,000 Genomes Project (Genomics England). **Table S4.** Primer sequences used in this study. **Table S5.** Overview of selected canonical and non-canonical protein-coding transcripts of IRD genes and their expression in adult human retina. **Table S6.** Genomic coordinates (GRCh38) of the regions in which variants were searched in this study (*5'UTR analysis file*). **Table S7.** Overview of the structure of 5'UTRs of the transcripts included in this study. **Table S8.** Comparison of 5'UTRs of non-canonical transcripts with respect to their corresponding canonical isoforms. **Table S9.** Overview of the performance of commercial exome capture designs on the 5'UTRs screened in this study. **Table S10.** Overview of the performance of commercial exome capture kits mostly used for the generation of our in-house WES data considering strict and padded designs. **Table S11.** Overview of all 5'UTR variants reported in IRD genes submitted to the ClinVar database. **Table S12.** Summarized overview of the distribution of 5'UTR variants assigned to more than one functional category. **Table S13.** Overview of the span of internal ribosomal entry sites (IRES) contained in the 5'UTR of IRD genes. **Table S14.** Illustration of the designed prioritization strategy applied to a selection of the identified variants. **Table S15.** Overview of additional variants reported for patients F5, F8, F11. **Table S16.** Overview of the 11 patients carrying the *RDH12*:c.701G>A (p.Arg234His) variant found *in cis* with *RDH12*:c.-123C>T.

Additional file 2: Figure S1. Schematic overview of constructs used for functional studies. **Figure S2.** Assessment of transcription start site (TSS) confidence by CAGE-seq in retina. **Figure S3.** Retina-enriched non-canonical isoforms of the *CRB1* and *RIMS2* genes. **Figure S4.** Evaluation of 5'UTR capture by the kits mostly used for generating our in-house WES data. **Figure S5.** Pedigrees of families carrying the 11 candidate 5'UTR variants and segregation analysis. **Figure S6.** Secondary structure prediction for wild-type and mutated 5'UTR sequences. **Figure S7.** Kozak consensus sequence of all IRD gene isoforms evaluated in this study. **Figure S8.** Sample overview of qPCR-based analysis of *ARL3* levels in *ARL3*:c-88G>A heterozygotes. **Figure S9.** Characterization of the 5'UTR splice variant c.-9+1G>T in *PRPF31*.

Additional file 3: Dataset S1. Annotation of the 1,450 5'UTRs variants that remained after functional classification and filtering. For each cohort (GE or CMGG) there is one tab per functional category.

Authors' contributions

ADR: Conception and project design, acquisition of data, analysis and interpretation of data, and drafting and revising the manuscript. MDPV: Acquisition of data, analysis and interpretation of data, and drafting and revising the manuscript. MB: Acquisition of data, analysis and interpretation of data, and revising the manuscript. KAW: Acquisition of data, analysis and interpretation of data, and revising the manuscript. FVB: Acquisition of data and revising the manuscript. MDV: Acquisition of data and revising the manuscript. HBT: Acquisition of data and revising the manuscript. MVH: Acquisition of data and revising the manuscript. MDB: Acquisition of data and revising the manuscript. SVS: Acquisition of data and revising the manuscript. MBW: Acquisition of data and revising the manuscript. HL: Acquisition of data and revising the manuscript. QM: Acquisition of data and revising the manuscript. DJ: Acquisition of data and revising the manuscript. GER: Acquisition of data and revising the manuscript. CR: Acquisition of data and revising the manuscript. RTO: Acquisition of data and revising the manuscript. JE: Acquisition of data and revising the manuscript. AW: Acquisition of data and revising the manuscript. GA: Acquisition of data and revising the manuscript. CA: Acquisition of data and revising the manuscript. JDZ: Acquisition of data and revising the manuscript. BL: Acquisition of data and revising the manuscript. EDB: Project supervision, acquisition of data, and revising the manuscript. FC: Conception and project supervision, acquisition of data, analysis and interpretation of data, and revising the manuscript. All authors read and approved the final manuscript.

Availability of data and materials

The data that support the findings of this study are available within the Genomics England (protected) Research Environment but restrictions apply to the availability of these data, as access to the Research Environment is limited to protect the privacy and confidentiality of participants. Likewise, healthcare and genomic data derived from individuals included in the Center for Medical Genetics Ghent (CMGG) cohort are not publicly available to comply with the consent given by those participants. Data as well as analysis scripts are available from the corresponding author (frauke.coppieters@ugent.be) who will, upon reasonable request, ensure secure data sharing in a reasonable timeframe after a request has been made. Extended data generated in this study are available in the supplementary materials.

Declarations

Ethics approval and consent to participate

The participants from the GE sub-cohort consented to participate in the Rare Disease arm of the 100,000 Genomes Project. The participants from the CMGG sub-cohort consented to genetic testing at the Center for Medical Genetics Ghent. The 100,000 Genomes Project Protocol has ethical approval from the HRA Committee East of England – Cambridge South (REC Ref 14/EE/1112). This study was registered with Genomics England within the *Hearing and sight domain* under Research Registry Projects 465. This study was approved by the ethics committee for Ghent University Hospital (B6702021000312). The studies have been conducted in accordance with the tenets of the Helsinki Declaration and subsequent reviews.

Consent for publication

Not applicable. We present only de-identified data.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Center for Medical Genetics Ghent (CMGG), Ghent University Hospital, Ghent, Belgium. [2]Department of Biomolecular Medicine, Ghent University, Corneel Heymanslaan 10, Ghent 9000, Belgium. [3]Department of Genetics, Instituto de Investigación Sanitaria-Fundación Jiménez Díaz, University Hospital, Universidad Autónoma de Madrid (IIS-FJD, UAM), Madrid, Spain. [4]Division

Dueñas Rey *et al. Genome Medicine*   (2024) 16:7

Page 19 of 21

of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicines and Health, University of Manchester, Manchester, UK. [5]Department of Ophthalmology, Ghent University Hospital, Ghent, Belgium. [6]Department of Head & Skin, Ghent University, Ghent, Belgium. [7]UCL Institute of Ophthalmology, University College London, London, UK. [8]Moorfields Eye Hospital, London, UK. [9]Clinical Genetics Department, Guy's Hospital, London, UK. [10]Genomics England, London, UK. [11]Department of Ophthalmology, University of Basel, Basel, Switzerland. [12]Institute of Molecular and Clinical Ophthalmology Basel (IOB), Basel, Switzerland. [13]Department of Genetics and Genome Biology, University of Leicester, Leicester, UK. [14]Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester, UK. [15]Center for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, Madrid, Spain. [16]Division of Ophthalmology, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [17]Department of Pharmaceutics, Ghent University, Ghent, Belgium.

References

1. Jackson RJ, Hellen CUT, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol. 2010;11(2):113–27. https://doi.org/10.1038/nrm2838.
2. Araujo PR, Yoon K, Ko D, et al. Before it gets started: regulating translation at the 5′UTR. Comp Funct Genomics. 2012;2012. https://doi.org/10.1155/2012/475731.
3. Kozak M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc Natl Acad Sci U S A. 1986;83(9):2850–4. https://doi.org/10.1073/PNAS.83.9.2850.
4. Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell. 2009;136(4):731. https://doi.org/10.1016/J.CELL.2009.01.042.
5. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci U S A. 2009;106(18):7507. https://doi.org/10.1073/PNAS.0810916106.
6. Leppek K, Das R, Barna M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat Rev Mol Cell Biol. 2018;19(3):158–74. https://doi.org/10.1038/nrm.2017.103.
7. Bugaut A, Balasubramanian S. 5′-UTR RNA G-quadruplexes: translation regulation and targeting. Nucleic Acids Res. 2012;40(11):4727–41. https://doi.org/10.1093/nar/gks068.
8. Steri M, Idda ML, Whalen MB, Orrù V. Genetic variants in mRNA untranslated regions. Wiley Interdiscip Rev RNA. 2018;9(4):e1474. https://doi.org/10.1002/WRNA.1474.
9. Barbosa C, Peixeiro I, Romão L. Gene expression regulation by upstream open reading frames and human disease. PLOS Genet. 2013;9(8):e1003529. https://doi.org/10.1371/JOURNAL.PGEN.1003529.
10. Soukarieh O, Meguerditchian C, Proust C, et al. Common and rare 5′UTR variants altering upstream open reading frames in cardiovascular genomics. Front Cardiovasc Med. 2022;9:542. https://doi.org/10.3389/FCVM.2022.841032.
11. Whiffin N, Karczewski KJ, Zhang X, et al. Characterising the loss-of-function impact of 5′ untranslated region variants in 15,708 individuals. Nat Commun. 2020;11(1):1–12. https://doi.org/10.1038/s41467-019-10717-9.
12. Smedley D, Smith KR, Martin A, et al. 100,000 genomes pilot on rare-disease diagnosis in health care — preliminary report. N Engl J Med. 2021;385(20):1868–80. https://doi.org/10.1056/nejmoa2035790.
13. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nat Rev Genet. 2018;19(5):253–68. https://doi.org/10.1038/nrg.2017.116.
14. Fernández-Marmiesse A, Gouveia S, Couce ML. NGS technologies as a turning point in rare disease research, diagnosis and treatment. Curr Med Chem. 2018;25(3):404. https://doi.org/10.2174/0929867324666170718101946.
15. Carss K, Arno G, Erwood M, et al. Comprehensive rare variant analysis via whole-genome sequencing to determine the molecular pathology of inherited retinal disease. Am J Hum Genet. 2017;100(1):75–90. https://doi.org/10.1016/J.AJHG.2016.12.003.
16. Retterer K, Juusola J, Cho MT, et al. Clinical application of whole-exome sequencing across clinical indications. Genet Med. 2016;18(7):696–704. https://doi.org/10.1038/gim.2015.148.
17. Lionel AC, Costain G, Monfared N, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. Genet Med. 2018;20(4):435–43. https://doi.org/10.1038/GIM.2017.119.
18. Dai P, Honda A, Ewans L, et al. Recommendations for next generation sequencing data reanalysis of unsolved cases with suspected Mendelian disorders: a systematic review and meta-analysis. Genet Med. 2022;24(8):1618–29. https://doi.org/10.1016/J.GIM.2022.04.021.
19. Wright CF, Quaife NM, Ramos-Hernández L, et al. Non-coding region variants upstream of MEF2C cause severe developmental disorder through three distinct loss-of-function mechanisms. Am J Hum Genet. 2021;108(6):1083. https://doi.org/10.1016/J.AJHG.2021.04.025.
20. Zhang X, Wakeling M, Ware J, Whiffin N. Annotating high-impact 5′untranslated region variants with the UTRannotator. Bioinformatics. 2021;37(8):1171–3. https://doi.org/10.1093/BIOINFORMATICS/BTAA783.
21. Wang J, Gribskov M. IRESpy: an XGBoost model for prediction of internal ribosome entry sites. BMC Bioinformatics. 2019;20(1):1–15. https://doi.org/10.1186/S12859-019-2999-7.
22. Chatterjee S, Pal JK. Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. Biol Cell. 2009;101(5):251–62. https://doi.org/10.1042/BC20080104.
23. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. Cell. 2019;176(3):535-548.e24. https://doi.org/10.1016/J.CELL.2018.12.015.
24. Zhao J, Wu J, Xu T, Yang Q, He J, Song X. IRESfinder: Identifying RNA internal ribosome entry site in eukaryotic cell using framed k-mer features. J Genet Genomics. 2018;45(7):403–6. https://doi.org/10.1016/J.JGG.2018.07.006.
25. Lim Y, Arora S, Schuster SL, et al. Multiplexed functional genomic analysis of 5′ untranslated region mutations across the spectrum of prostate cancer. Nat Commun. 2021;12(1):1–18. https://doi.org/10.1038/s41467-021-24445-6.
26. Filatova A, Reveguk I, Piatkova M, et al. Annotation of uORFs in the OMIM genes allows to reveal pathogenic variants in 5′UTRs. Nucleic Acids Res. 2023;51(3):1229–44. https://doi.org/10.1093/nar/gkac1247.
27. Ellingford JM, Ahn JW, Bagnall RD, et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. Genome Med. 2022;14(1):1–19. https://doi.org/10.1186/S13073-022-01073-3/FIGURES/3.
28. Pontikos N, Arno G, Jurkute N, et al. Genetic Basis of inherited retinal disease in a molecularly characterized cohort of more than 3000 families from the United Kingdom. Ophthalmology. 2020;127(10):1384–94. https://doi.org/10.1016/j.ophtha.2020.04.008.
29. Ratnapriya R, Swaroop A. Genetic architecture of retinal and macular degenerative diseases: the promise and challenges of next-generation sequencing. Genome Med. 2013;5(10). https://doi.org/10.1186/GM488.
30. O'sullivan J, Mullaney BG, Bhaskar SS, et al. A paradigm shift in the delivery of services for diagnosis of inherited retinal disease. J Med Genet. 2012;49(5):322–6. https://doi.org/10.1136/jmedgenet-2012-100847.
31. Dockery A, Whelan L, Humphries P, Farrar GJ. Next-generation sequencing applications for inherited retinal diseases. Int J Mol Sci. 2021;22(11):5684. https://doi.org/10.3390/ijms22115684.
32. Berger W, Kloeckener-Gruissem B, Neidhardt J. The molecular basis of human retinal and vitreoretinal diseases. Prog Retin Eye Res. 2010;29(5):335–75. https://doi.org/10.1016/J.PRETEYERES.2010.03.004.
33. Hanany M, Rivolta C, Sharon D. Worldwide carrier frequency and genetic prevalence of autosomal recessive inherited retinal diseases. Proc Natl Acad Sci U S A. 2020;117(5):2710–6. https://doi.org/10.1073/PNAS.1913179117.
34. RetNet - Retinal Information Network. Accessed 6 April 2023. https://web.sph.uth.edu/RetNet/.
35. Tatour Y, Ben-Yosef T. Syndromic inherited retinal diseases: genetic, clinical and diagnostic aspects. Diagnostics (Basel, Switzerland). 2020;10(10). https://doi.org/10.3390/DIAGNOSTICS10100779.
36. Ellingford JM, Hufnagel RB, Arno G. Phenotype and genotype correlations in inherited retinal diseases: population-guided variant

Dueñas Rey *et al. Genome Medicine*   (2024) 16:7

Page 20 of 21

interpretation Variable Expressivity and Incomplete Penetrance. Genes (Basel). 2020;11(11):1–4. https://doi.org/10.3390/GENES11111274.

37. Sharon D, Ben-Yosef T, Goldenberg-Cohen N, et al. A nationwide genetic analysis of inherited retinal diseases in Israel as assessed by the Israeli inherited retinal disease consortium (IIRDC). Hum Mutat. 2020;41(1):140–9. https://doi.org/10.1002/HUMU.23903.

38. Perea-Romero I, Gordo G, Iancu IF, et al. Genetic landscape of 6089 inherited retinal dystrophies affected cases in Spain and their therapeutic and extended epidemiological implications. Sci Rep. 2021;11(1). https://doi.org/10.1038/S41598-021-81093-Y.

39. Fadaie Z, Whelan L, Ben-Yosef T, et al. Whole genome sequencing and in vitro splice assays reveal genetic causes for inherited retinal diseases. NPJ Genomic Med. 2021;6(1). https://doi.org/10.1038/S41525-021-00261-1.

40. Lenassi E, Clayton-Smith J, Douzgou S, et al. Clinical utility of genetic testing in 201 preschool children with inherited eye disorders. Genet Med. 2020;22(4):745–51. https://doi.org/10.1038/S41436-019-0722-8.

41. Lam BL, Leroy BP, Black G, Ong T, Yoon D, Trzupek K. Genetic testing and diagnosis of inherited retinal diseases. Orphanet J Rare Dis. 2021;16(1):1–9. https://doi.org/10.1186/S13023-021-02145-0.

42. Liquori A, Vaché C, Baux D, et al. Whole USH2A gene sequencing identifies several new deep intronic mutations. Hum Mutat. 2016;37(2):184–93. https://doi.org/10.1002/HUMU.22926.

43. Jamshidi F, Place EM, Mehrotra S, et al. Contribution of noncoding pathogenic variants to RPGRIP1-mediated inherited retinal degeneration. Genet Med. 2019;21(3):694–704. https://doi.org/10.1038/S41436-018-0104-7.

44. Holtan JP, Selmer KK, Heimdal KR, Bragadóttir R. Inherited retinal disease in Norway - a characterization of current clinical and genetic knowledge. Acta Ophthalmol. 2020;98(3):286–95. https://doi.org/10.1111/AOS.14218.

45. Reurink J, Weisschuh N, Garanto A, et al. Whole genome sequencing for USH2A-associated disease reveals several pathogenic deep-intronic variants that are amenable to splice correction. HGG Adv. 2023;4(2). https://doi.org/10.1016/J.XHGG.2023.100181.

46. Bauwens M, Garanto A, Sangermano R, et al. ABCA4-associated disease as a model for missing heritability in autosomal recessive disorders: novel noncoding splice, cis-regulatory, structural, and recurrent hypomorphic variants. Genet Med. 2019;21(8):1761–71. https://doi.org/10.1038/S41436-018-0420-Y.

47. Sangermano R, Garanto A, Khan M, et al. Deep-intronic ABCA4 variants explain missing heritability in Stargardt disease and allow correction of splice defects by antisense oligonucleotides. Genet Med. 2019;21(8):1751–60. https://doi.org/10.1038/S41436-018-0414-9.

48. Khan M, Cornelis SS, Del Pozo-Valero M, et al. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. Genet Med. 2020;22(7):1235–46. https://doi.org/10.1038/S41436-020-0787-4.

49. Daich Varela M, Bellingham J, Motta F, et al. Multidisciplinary team directed analysis of whole genome sequencing reveals pathogenic non-coding variants in molecularly undiagnosed inherited retinal dystrophies. Hum Mol Genet. 2023;32(4):595–607. https://doi.org/10.1093/HMG/DDAC227.

50. Qian X, Wang J, Wang M, et al. Identification of deep-intronic splice mutations in a large cohort of patients with inherited retinal diseases. Front Genet. 2021;12:276. https://doi.org/10.3389/FGENE.2021.647400.

51. Weisschuh N, Sturm M, Baumann B, et al. Deep-intronic variants in CNGB3 cause achromatopsia by pseudoexon activation HHS Public Access. Hum Mutat. 2020;41(1):255–64. https://doi.org/10.1002/humu.23920.

52. Cherry TJ, Yang MG, Harmin DA, et al. Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease. Proc Natl Acad Sci U S A. 2020;117(16):9001–12. https://doi.org/10.1073/PNAS.1922501117/-/DCSUPPLEMENTAL.

53. Coppieters F, Todeschini AL, Fujimaki T, et al. Hidden genetic variation in LCA9-associated congenital blindness explained by 5′UTR mutations and copy-number variations of NMNAT1. Hum Mutat. 2015;36(12):1188. https://doi.org/10.1002/HUMU.22899.

54. Ruberto FP, Balzano S, Namburi P, et al. Heterozygous deletions of noncoding parts of the PRPF31 gene cause retinitis pigmentosa via reduced gene expression. Mol Vis. 2021;27:107.

55. Van de Sompele S, Small KW, Cicekdal MB, et al. Multi-omics approach dissects cis-regulatory mechanisms underlying North Carolina macular dystrophy, a retinal enhanceropathy. Am J Hum Genet. 2022;109(11):2029–48. https://doi.org/10.1016/J.AJHG.2022.09.013.

56. Small KW, DeLuca AP, Whitmore SS, et al. North Carolina Macular Dystrophy is caused by dysregulation of the retinaltranscription factor PRDM13. Ophthalmology. 2016;123(1):9. https://doi.org/10.1016/J.OPHTHA.2015.10.006.

57. Filatova AY, Vasilyeva TA, Marakhonov AV, et al. Upstream ORF frameshift variants in the PAX6 5′UTR cause congenital aniridia. Hum Mutat. 2021;42(8):1053–65. https://doi.org/10.1002/HUMU.24248.

58. Zuercher J, Neidhardt J, Magyar I, et al. Alterations of the 5′untranslated region of SLC16A12 lead to age-related cataract. Invest Ophthalmol Vis Sci. 2010;51(7):3354. https://doi.org/10.1167/IOVS.10-5193.

59. Zhang W, Kassels AC, Barrington A, et al. Macular corneal dystrophy with isolated peripheral Descemet membrane deposits. Am J Ophthalmol Case Rep. 2019;16. https://doi.org/10.1016/J.AJOC.2019.100571.

60. Ratnapriya R, Sosina OA, Starostik MR, et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. Nat Genet. 2019;51(4):606–10. https://doi.org/10.1038/S41588-019-0351-9.

61. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5):525–7. https://doi.org/10.1038/NBT.3519.

62. Team RC. R: a language and environment for statistical computing. R Found Stat Comput: Published online; 2021.

63. Morales J, Pujar S, Loveland JE, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. Nature. 2022;604(7905). https://doi.org/10.1038/S41586-022-04558-8.

64. Martin AR, Williams E, Foulger RE, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet. 2019;51(11):1560–5. https://doi.org/10.1038/S41588-019-0528-2.

65. Forrest ARR, Kawaji H, Rehli M, et al. A promoter-level mammalian expression atlas. Nature. 2014;507(7493):462–70. https://doi.org/10.1038/NATURE13182.

66. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinforma Appl NOTE. 2010;26(6):841–2. https://doi.org/10.1093/bioinformatics/btq033.

67. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44(D1):D862–8. https://doi.org/10.1093/NAR/GKV1222.

68. Research Environment User Guide - Genomics England Research Environment - Genomics England Research Environment. Accessed December 1, 2023. https://research-help.genomicsengland.co.uk/display/GERE/Research+Environment+User+Guide.

69. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. 2009;25(14):1754-1760. https://doi.org/10.1093/bioinformatics/btp324.

70. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303. https://doi.org/10.1101/GR.107524.110.

71. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2). https://doi.org/10.1093/GIGASCIENCE/GIAB008.

72. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1). https://doi.org/10.1186/S13059-016-0974-4.

73. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434–43. https://doi.org/10.1038/S41586-020-2308-7.

74. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res. 2014;42(22):13534–44. https://doi.org/10.1093/NAR/GKU1206.

75. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11(2–3):377–94. https://doi.org/10.1089/1066527041410418.

76. Frazer J, Notin P, Dias M, et al. Disease variant prediction with deep generative models of evolutionary data. Nature. 2021;599(7883):91–5. https://doi.org/10.1038/s41586-021-04043-8.

77. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310–5. https://doi.org/10.1038/NG.2892.

78. Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. The ensembl regulatory build. Genome Biol. 2015;16(1):56–56. https://doi.org/10.1186/S13059-015-0621-5.

79. Kozak M. The scanning model for translation: an update. J Cell Biol. 1989;108(2):229–41. https://doi.org/10.1083/JCB.108.2.229.

80. McClements ME, Butt A, Piotter E, Peddle CF, Maclaren RE. An analysis of the Kozak consensus in retinal genes and its relevance to gene therapy. Mol Vis. 2021;27:233.

81. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188–90. https://doi.org/10.1101/GR.849004.

82. Sample PJ, Wang B, Reid DW, et al. Human 5′ UTR design and variant effect prediction from a massively parallel translation assay. Nat Biotechnol. 2019;37(7):803–9. https://doi.org/10.1038/s41587-019-0164-5.

83. Murat P, Marsico G, Herdy B, Ghanbarian A, Portella G, Balasubramanian S. RNA G-quadruplexes at upstream open reading frames cause DHX36- and DHX9-dependent translation of human mRNAs. Genome Biol. 2018;19(1):1–24. https://doi.org/10.1186/S13059-018-1602-2.

84. Li H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics. 2011;27(5):718. https://doi.org/10.1093/BIOINFORMATICS/BTQ671.

85. Fu L, Cao Y, Wu J, Peng Q, Nie Q, Xie X. UFold: fast and accurate RNA secondary structure prediction with deep learning. Nucleic Acids Res. 2022;50(3):e14–e14. https://doi.org/10.1093/NAR/GKAB1074.

86. Chen CC, Chan YM. REDfold: accurate RNA secondary structure prediction using residual encoder-decoder network. BMC Bioinformatics. 2023;24(1):122. https://doi.org/10.1186/S12859-023-05238-8/FIGURES/3.

87. Sato K, Akiyama M, Sakakibara Y. RNA secondary structure prediction using deep learning with thermodynamic integration. Nat Commun. 2021;12(1):1–9. https://doi.org/10.1038/s41467-021-21194-4.

88. Yang TH, Wang CY, Tsai HC, Liu CT. Human IRES Atlas: an integrative platform for studying IRES-driven translational regulation in humans. Database. 2021;2021:1–16. https://doi.org/10.1093/DATABASE/BAAB025.

89. Sharon D, Banin E. Nonsyndromic retinitis pigmentosa is highly prevalent in the Jerusalem region with a high frequency of founder mutations. Mol Vis. 2015;21.

90. Zernant J, Lee W, Collison FT, et al. Frequent hypomorphic alleles account for a significant fraction of ABCA4 disease and distinguish it from age-related macular degeneration. J Med Genet. 2017;54(6). https://doi.org/10.1136/jmedgenet-2017-104540.

91. Rowlands C, Thomas HB, Lord J, et al. Comparison of in silico strategies to prioritize rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. Sci Rep. 2021;11(1). https://doi.org/10.1038/s41598-021-99747-2.

92. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99(4). https://doi.org/10.1016/j.ajhg.2016.08.016.

93. Pejaver V, Byrne AB, Feng BJ, et al. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. Am J Hum Genet. 2022;109(12). https://doi.org/10.1016/j.ajhg.2022.10.013.

94. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48. https://doi.org/10.18637/JSS.V067.I01.

95. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 2007;8(2):R19. https://doi.org/10.1186/gb-2007-8-2-r19.

96. Thompson DA, Janecke AR, Lange J, et al. Retinal degeneration associated with RDH12 mutations results from decreased 11- cis retinal synthesis due to disruption of the visual cycle. Hum Mol Genet. 2005;14(24):3865–75. https://doi.org/10.1093/HMG/DDI411.

97. Ba-Abbad R, Arno G, Robson AG, et al. Macula-predominant retinopathy associated with biallelic variants in RDH12. Ophthalmic Genet. 2020;41(6):612–5. https://doi.org/10.1080/13816810.2020.1802763.

98. Liu JY, Dai X, Sheng J, et al. Identification and functional characterization of a novel splicing mutation in RP gene PRPF31. Biochem Biophys Res Commun. 2008;367(2). https://doi.org/10.1016/j.bbrc.2007.12.156.

99. Dryja TP, McGee TL, Berson EL, et al. Night blindness and abnormal cone electroretinogram ON responses in patients with mutations in the GRM6 gene encoding mGluR6. Proc Natl Acad Sci U S A. 2005;102(13). https://doi.org/10.1073/pnas.0501233102.

100. De Zaeytijd J, Van Cauwenbergh C, De Bruyne M, et al. Isolated maculopathy and moderate rod-cone dystrophy represent the milder end of the RDH12-related retinal dystrophy spectrum. Retina. 2021;41(6):1346–55. https://doi.org/10.1097/IAE.0000000000003028.

101. Tiering (Rare Disease) - Genomics England Research Environment - Genomics England Confluence. Accessed 14 April 2023. https://cnfl.extge.co.uk/pages/viewpage.action?pageId=113194832.

102. Schmitz-Abe K, Li Q, Rosen SM, et al. Unique bioinformatic approach and comprehensive reanalysis improve diagnostic yield of clinical exomes. Eur J Hum Genet. 2019;27(9):1398–405. https://doi.org/10.1038/s41431-019-0401-x.

103. Wenger AM, Guturu H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. Genet Med. 2017;19(2):209–14. https://doi.org/10.1038/GIM.2016.88.

104. de Bruijn SE, Rodenburg K, Corominas J, et al. Optical genome mapping and revisiting short-read genome sequencing data reveal previously overlooked structural variants disrupting retinal disease-associated genes. Genet Med. 2023;25(3). https://doi.org/10.1016/J.GIM.2022.11.013.

105. Del Pozo-Valero M, Martin-Merida I, Jimenez-Rolando B, et al. Expanded phenotypic spectrum of retinopathies associated with autosomal recessive and dominant mutations in PROM1. Am J Ophthalmol. 2019;207:204–14. https://doi.org/10.1016/J.AJO.2019.05.014.

106. Moreno-Leon L, West EL, O'Hara-Wright M, et al. RPGR isoform imbalance causes ciliary defects due to exon ORF15 mutations in X-linked retinitis pigmentosa (XLRP). Hum Mol Genet. 2021;29(22):3706–16. https://doi.org/10.1093/HMG/DDAA269.

107. Vig A, Poulter JA, Ottaviani D, et al. DYNC2H1 hypomorphic or retina-predominant variants cause nonsyndromic retinal degeneration. Genet Med. 2020;22(12):2041–51. https://doi.org/10.1038/S41436-020-0915-1.

108. Riazuddin SA, Iqbal M, Wang Y, et al. A splice-site mutation in a retina-specific exon of BBS8 causes nonsyndromic retinitis pigmentosa. Am J Hum Genet. 2010;86(5):805–12. https://doi.org/10.1016/J.AJHG.2010.04.001.

109. Mairot K, Smirnov V, Bocquet B, et al. CRB1-related retinal dystrophies in a cohort of 50 patients: a reappraisal in the light of specific Müller cell and photoreceptor CRB1 isoforms. Int J Mol Sci. 2021;22(23):12642. https://doi.org/10.3390/IJMS222312642.

110. Weber R, Ghoshdastider U, Spies D, et al. Monitoring the 5′UTR landscape reveals isoform switches to drive translational efficiencies in cancer. Oncogene. 2023;42(9):638–50. https://doi.org/10.1038/S41388-022-02578-2.

111. Ray TA, Cochran K, Kozlowski C, et al. Comprehensive identification of mRNA isoforms reveals the diversity of neural cell-surface molecules with roles in retinal development and disease. Nat Commun. 2020;11(1). https://doi.org/10.1038/S41467-020-17009-7.

112. Grodecká L, Buratti E, Freiberger T. Mutations of pre-mRNA splicing regulatory elements: are predictions moving forward to clinical diagnostics? Int J Mol Sci. 2017;18(8). https://doi.org/10.3390/ijms18081668.

113. Ryczek N, Łyś A, Makałowska I. The functional meaning of 5′UTR in protein-coding genes. Int J Mol Sci. 2023;24(3). https://doi.org/10.3390/ijms24032976.

114. Ratnapriya R, et al. A novel ARL3 gene mutation associated with autosomal dominant retinal degeneration. Front Cell Dev Biol. 2021;9:720782. https://doi.org/10.3389/fcell.2021.720782.

115. Szlachta K, Thys RG, Atkin ND, Pierce LCT, Bekiranov S, Wang YH. Alternative DNA secondary structure formation affects RNA polymerase II promoter-proximal pausing in human. Genome Biol. 2018;19(1). https://doi.org/10.1186/s13059-018-1463-8.

116. Wang J, Wang Y, Li S, et al. Clinical and genetic analysis of RDH12-associated retinopathy in 27 Chinese families: a hypomorphic allele leads to cone-rod dystrophy. Invest Ophthalmol Vis Sci. 2022;63(9):24. https://doi.org/10.1167/iovs.63.9.24.

Publisher's Note